Exposome and health: Characterization and network-based exploration of diverse environmental chemical spaces

By

Janani R

LIFE10201604004

The Institute of Mathematical Sciences
Chennai

A thesis submitted to the
Board of Studies in Life Sciences
In partial fulfillment of requirements
for the Degree of

DOCTOR OF PHILOSOPHY
of
HOMI BHABHA NATIONAL INSTITUTE

December 2021

Homi Bhabha National Institute

Recommendations of the Viva Voce Committee

As members of the Viva Voce Committee, we certify that we have read the dissertation prepared byJanani R entitled: "Exposome and health: Characterization and network-based exploration of diverse environmental chemical spaces" and recommend that it may be accepted as fulfilling the dissertation requirement for the Degree of Doctor of Philosophy.

Date: 22/06/2022

Chair - Prof. Rahul Siddharthan

Date: 22/06/2022

Supervisor/Convener - Prof. Areejit Samal

Date: 22/06/2022

Member 1 - Prof. Sitabhra Sinha

Date: 22/06/2022

Member 2 - Prof. Satyavani Vemparala

Date: 22/06/2022

Member 3 - Prof. Dhiraj Kumar

Date: 22/06/2022

External Examiner - Prof. Nagasuma Chandra

Final approval and acceptance of this dissertation is contingent upon the candidate's submission of thefinal copies of the dissertation to HBNI.

I hereby certify that I have read this dissertation prepared under my direction and recommend that it may be accepted as fulfilling the dissertation requirement.

Date: 22/06/2022

Place: CHENNAI

Supervisor

Statement by Author

This dissertation has been submitted in partial fulfillment of requirements for an advanced degree at Homi Bhabha National Institute (HBNI) and is deposited in the Library to be made available to borrowers under rules of the HBNI.

Brief quotations from this dissertation are allowable without special permission, provided that accurate acknowledgement of source is made. Requests for permission for extended quotation from or reproduction of this manuscript in whole or in part may be granted by the Competent Authority of HBNI when in his or her judgement the proposed use of the material is in the interests of scholarship. In all other instances, however, permission must be obtained from the author.

Janani R

Declaration

I, hereby declare that the investigation presented in this thesis has been carried out by me. The work is original and has not been submitted earlier as a whole or in part for a degree or diploma at this or any other Institution or University.

Janani R

List of Publications arising from the thesis

Journals

Published

1. *A curated knowledgebase on endocrine disrupting chemicals and their biological systems-level perturbations*, B.S. Karthikeyan[†], **J. Ravichandran**[†,⋆], K. Mohanraj, R.P. Vivek-Ananth and A. Samal[⋆], Science of the Total Environment, 692: 281-296 (2019). https://doi.org/10.1016/j.scitotenv.2019.07.225
2. *DEDuCT 2.0: An updated knowledgebase and an exploration of the current regulations and guidelines from the perspective of endocrine disrupting chemicals*, B.S. Karthikeyan[†], **J. Ravichandran**[†,⋆], S.R. Aparna and A. Samal[⋆], Chemosphere, 267: 128898 (2021). https://doi.org/10.1016/j.chemosphere.2020.128898
3. *ExHuMId: A curated resource and analysis of Exposome of Human Milk across India*, B.S. Karthikeyan[†], **J. Ravichandran**[†,⋆], S.R. Aparna and A. Samal[⋆], Chemosphere, 271: 129583 (2021). https://doi.org/10.1016/j.chemosphere.2021.129583
4. *NeurotoxKb: compilation, curation and exploration of a knowledgebase of environmental neurotoxicants specific to mammals*, **J. Ravichandran**[†], B.S. Karthikeyan[†], P. Singla, S.R. Aparna and A. Samal[⋆], Chemosphere, 278: 130387 (2021). https://doi.org/10.1016/j.chemosphere.2021.130387
5. *Network biology approach to human tissue-specific chemical exposome*, **J. Ravichandran**[†], B.S. Karthikeyan[†], S.R. Aparna and A. Samal[⋆], The Journal of Steroid Biochemistry and Molecular Biology, 214: 105998 (2021). https://doi.org/10.1016/j.jsbmb.2021.105998
6. *An atlas of fragrance chemicals in children's products*, **J. Ravichandran**[†], B.S. Karthikeyan[†], J. Jost and A. Samal[⋆], Science of the Total Environment, 818: 151682 (2022). https://doi.org/10.1016/j.scitotenv.2021.151682
7. *Investigation of a derived adverse outcome pathway (AOP) network for endocrine-mediated perturbations*, **J. Ravichandran**, B.S. Karthikeyan and A. Samal[⋆], Sci-

ence of the Total Environment, 826: 154112 (2022). https://doi.org/10.1016/j.scitotenv.2022.154112

[† Joint-first authors; ⋆ Corresponding author(s)]

Copyright

1. *DEDuCT - Database of Endocrine Disrupting Chemicals and their Toxicity Profiles*
   Authors: A. Samal, **J. Ravichandran**, B.S. Karthikeyan, M. Karthikeyan and R.P. Vivek Ananth.
   Copyright granted to The Institute of Mathematical Sciences by the Copyright office, Government of India, with the Dairy Number 16429/2018-CO/L.

Oral or Poster presentations

1. Oral presentation titled *ExHuMId: A curated resource and analysis of Exposome of Human Milk across India* at the Young Scientists' conference of the India International Science Festival (IISF-2020) held from December 22-24, 2020.
2. Poster presentation titled *A curated knowledgebase on endocrine disrupting chemicals enabling mechanistic insights into systems-level perturbations upon exposure* at the German Conference on Bioinformatics (GCB-2019) held at Heidelberg, Germany from September 16-19, 2019.
3. Poster presentation titled *A curated knowledgebase on endocrine disrupting chemicals enabling mechanistic insights into systems-level perturbations upon exposure* at the Horizons in Molecular Biology held at the Max Planck Institute for Biophysical chemistry, Göttingen, Germany from September 9-12, 2019.
4. Seminar titled *Network Science perspective on a chemical space harmful to humankind* at The Institute of Mathematical Sciences (IMSc), Chennai on March 11, 2019.
5. Poster presentation titled *Development of knowledgebase for Systems Toxicology and Toxicogenomics* at the Young Scientists' conference of the India International

Science Festival (IISF-2018) held at Lucknow from October 5-6, 2018.

6. Poster presentation titled *Development of knowledgebase for Systems Toxicology and Toxicogenomics* at the International Conference on Bioinformatics (INCoB 2018) held at the Jawaharlal Nehru University (JNU), New Delhi from September 26-28, 2018.

Research visits and seminars

1. Seminar in the Laboratory of Prof. Oliver Ebenhöh, Institut für Quantitative und Theoretische Biologie, Heinrich Heine Universität, Düsseldorf, Germany on September 20, 2019.
2. Seminar in the Laboratory of Dr. Tiago C. Alves, Institute for Clinical Chemistry and Laboratory Medicine, Dresden, Germany on September 13, 2019.

Janani R

This thesis is dedicated

TO MY PARENTS

For their unconditional love and support

Acknowledgements

At the outset, I would like to thank my thesis supervisor, Prof. Areejit Samal, for his continued guidance and support. I have greatly benefited from his guidance and he has always inspired me to aim high in my career. His constructive criticism and suggestions have helped me in moulding myself into a strong personality and a better researcher. His hardworking nature and willingness to assist in all aspects of the research, regardless of time, has truly inspired me, and is something I hope to carry forward throughout my career. I am grateful for his efforts and support throughout the pandemic, during which he ensured that the work was never hindered. Further, I am immensely thankful to him for his assistance during all of my conference participation and research visits.

I am extremely grateful to all of my co-authors with whom I have collaborated on various research projects. These collaborations have allowed me to pick up new skills from each of them. A special thanks to BS Karthikeyan and SR Aparna, who have co-authored the majority of my publications reported in this thesis, for their unwavering support in designing projects, writing manuscripts, and revising them. My heartfelt thanks to M Karthikeyan, from whom I have learned database development and data visualization skills during his time in IMSc. He never failed to lend a hand whenever I needed help, even after he left for his doctoral studies in Germany. Many thanks to RP Vivek-Ananth for his assistance with Cheminformatics analysis reported in Chapter 2 and DEDuCT revisions. I also thank our intern, Palak Singla, for working sincerely even during the pandemic, which led to a significant publication reported in Chapter 5 of this thesis. I thank Prof. Jürgen Jost for his valuable insights for the publication reported in Chapter 7 of this thesis.

I thank all my doctoral committee members - Prof. Rahul Siddharthan, Prof. Sitabhra Sinha, Prof. Satyavani Vemparala, and Prof. Dhiraj Kumar for their critical comments during the doctoral committee meetings. I would also like to thank all the faculties and researchers - Prof. Gautam Menon, Prof. Satyavani Vemparala, Prof. Sitabhra Sinha,

Prof. Rahul Siddharthan, Prof. Areejit Samal, Prof. S Krishnaswamy, Dr. Vasudharani Devanathan, Dr. Nivedita Chatterjee, Prof. Vijayalakshmi Mahadevan, Dr. Grace Chongloi, Dr. P Varuni, who have taught various courses during the course-work period of my PhD.

I would like to thank Mr. B Raveendra Reddy for his assistance in setting up the web server described in the thesis chapters.

I am very grateful for the intellectual discussions I had with Prof. Oliver Ebenhöh, Prof. Martin J. Lercher, and Dr. Tiago C. Alves during my academic visit to Germany.

My sincere thanks to Prof. Sanjay Jain, Prof. Amit Singh, Prof. Dhiraj Kumar, Prof. Vinay K. Nandicoori for their thoughtful comments and suggestions on my research projects. Further, I thank Dr. Kushi Anand and other lab members of Prof. Amit Singh for sharing their research experiences and knowledge with me during my visit to IISc.

I thank IMSc for their funding and support during my PhD. A special thanks to Mrs. R Indra, who coordinated and assisted me during my participation at both national and international conferences. I would like to thank all administrative staff members and computer committee members for doing the needful whenever necessary. A huge thanks to everyone who works in the canteen, housekeeping, civil and electrical departments for making IMSc a pleasant place to work. A special thanks to Mrs. Mahalakshmi, the housekeeping staff, for her kindness and care, especially during my illness.

At this time, I would also like to thank all the editors and reviewers of all my publications, for their critical comments, which were really helpful in improving the quality of the work.

I am extremely grateful to all of my friends from IMSc who have helped and supported me in a variety of circumstances - Garima, Semanti, Ajjath, Pooja, Sruthy, Rakesh, Deepika, Vignesh, Savitha, Sourav, Ankita. Sincere gratitude to Semanti, one of the best roommates in my hostel life, who has mostly bought me food and medicines while I was sick. A heartfelt thanks to all the past and current lab members - Roshani, Van-

dana, Sreejith, Gayathri, Subathra, Sudharsan, Meena, Pavithra, Kavyaa, Tamil Maran, Divya, Murugesan, Jyotsna, Sudharsan V, Nithin, Vishalini, Evanjalee, Ajaya, Ajay, and Yasharth, for their support. Many thanks to Vandana who is currently a PhD student at IISc, Bangalore, for hosting me during my academic visits. A very special thanks to Garima, Gayathri, Subathra, Sudharsan, and Roshani, who have always been there for me in times of difficulty and happiness. Thank you for the short outings, dinners, and leisure walks, which will be treasured memories for the rest of my life.

I also thank all my friends from school (Divya, Dhana Laksmi, Mogana), BTech (Vinothini, Deva Priya, Sreemol, Gayathri), and Masters (Aravind, Jeffy, Divya Selvaraju, Nisha), for staying in touch and encouraging me throughout my PhD. A big thanks to Divya Selvaraju for her care and support, who visited from Vienna during my academic visit to Dresden. Furthermore, I thank Ashreya, a friend from IMSc, for her time and support while we were in Heidelberg. Even though our time together was limited, they were always memorable.

I am extremely grateful to my brother Kavin, mother-in-law, and Vedavalli aunty, all of whom reside in Chennai and never fail to make me feel at home. I am forever indebted to my aunts Valarmathi, Amudha, Kala, my uncles Madhavan, Chandrasekar, Manivannan, my cousins Geetha Priya, Padma Priya, Yogesh, and my brother Barath for their love and consistent motivation.

Lastly, I would like to express my deep gratitude to my dear parents Rani and Ravichandran, and my husband Bala Kumar who entrusted and supported all my life decisions. In particular, I thank my parents for tolerating and understanding my emotional overload at times, as well as ensuring a happy and positive environment to do my research peacefully. A big thanks to mom, dad, and Bala for believing in me and supporting me to pursue my dreams. They are the pillars of strength in my life, without whom I would not be the person I am today.

Janani R

Contents

List of Figures

List of Tables

Abstract

Humans are exposed to environmental chemicals in their everyday life and such exposure can contribute to the incidence of several chronic diseases. Characterization, monitoring and regulation of the ever-increasing space of environmental chemicals for their potential adverse health effects is both necessary and challenging. In other words, characterization of the chemical exposome from a health perspective is necessary for human well-being. To this end, there has been growing interest in characterizing the human exposome along with the genome to better understand the environmental factors crucial for human health and disease.

In this thesis, we focus on environmental chemicals that have gained significant attention from scientists, regulatory authorities, and the general public, due to their potential health concerns. In order to link chemical exposomes to health effects, we have undertaken a systematic compilation, curation and exploration of the existing information contained in published toxicological studies on diverse groups of environmental chemicals. Specifically, we focus on five groups of chemicals with toxicological relevance, namely endocrine disrupting chemicals (EDCs), environmental neurotoxicants, human milk contaminants, fragrance chemicals in children's products, and exogenous chemicals detected in human tissues.

Furthermore, there is recent recognition of the need to leverage network science and systems biology approaches in characterizing the chemical exposome. Therefore, we extensively employ these approaches on the compiled toxicological information for the five groups of environmental chemicals studied in this thesis. Specifically, we investigated similarity networks of these environmental chemicals based on similarity in chemical structures or similarity of target genes. Further, we constructed bipartite networks of environmental chemicals and their target genes, and tripartite networks of environmental chemicals, their target genes and associated diseases, to reveal perturbed pathways and potential disease comorbidities related to chemical exposure. Moreover, we derive a

comprehensive adverse outcome pathway (AOP) network for endocrine-mediated perturbations, and thereafter, employ graph-theoretic measures to identify the critical biological events associated with endocrine disruption upon chemical exposure.

To further demonstrate the utility of our research for chemical risk assessment, we perform a comparative study using several chemical lists that are a part of inventories, guidelines or regulations to assess the regulatory status and source of the diverse groups of environmental chemicals considered in this thesis. These analyses reveal that several environmental chemicals of concern are part of everyday exposures, and moreover, many of these chemicals are found to be produced in high volume.

In sum, the curated resources and multi-pronged analyses of diverse environmental chemical spaces described in this thesis will facilitate research in toxicology and human exposome.

Chapter 1

Introduction

1.1 Motivation

> *Our state of health or disease is really a reflection of the environment we all live in. And the environment we perceive.*
>
> - Darnell Houston

In the last century, industrial advances have resulted in the rapid synthesis and commercialization of myriad chemicals. As of October 2021, more than 86000 such chemicals have been registered with the United States Environmental Protection Agency (US EPA) under the Toxic Substances Control Act [1]. Further, based on an estimate from the United States National Toxicology Program report of 2017 [2], around 2000 new commercial chemicals are introduced into the market every year. However, only a small fraction of these chemicals released into the environment have been tested for safety or toxicity concerns to date [3, 4]. Humans are exposed to many of these environmental chemicals in their daily life in the form of consumer products including personal care products, pharmaceuticals, food additives, pesticides and insecticides [5–8]. Such exposure to environmental chemicals contribute significantly to the incidence of several chronic diseases [9–12]. In short, the ever-increasing rate of new chemicals released into the environment and the subsequent global prevalence of chronic diseases underline the urgent

need for the characterization and prioritization of environmental chemicals of concern to human health [9–11, 13–17].

To capture the diverse environmental factors influencing health and disease starting from the prenatal period, Wild [18] introduced the concept of "exposome". Subsequently, others have both expanded and refined the definition of the exposome. Rappaport *et al.* [19] included the body's internal chemical environment in the definition of the exposome. Miller *et al.* [20] expanded the definition of the exposome to include the behavioral aspects of human beings, including social interactions and emotional stressors. In sum, the human exposome captures a variety of environmental factors, both internal and external, among which the assessment of external stressors in the form of environmental contaminants or toxicants and the resulting impact on human health is gaining momentum among researchers [13, 18–20].

To improve the risk assessment of environmental chemicals, there is a need for systematic characterization and better understanding of the human health impact of such chemical exposures. Simply stated, there is immense interest in characterizing this chemical exposome. In this direction, two approaches have been undertaken to characterize the chemical exposome: "bottom-up" or "top-down" [21–23]. Using a "bottom-up" approach, the different classes of chemicals present in the external environment such as food, air, and water, can be evaluated and monitored for their potential health effects. This approach also enables the identification of exogenous exposures along with their sources in the environment. In contrast, a "top-down" approach involves the characterization of both exogenous and endogenous chemicals within the biological samples such as blood, urine, breast milk, and adipose tissue, of an individual. This approach does not provide any information on the source of the exogenous chemicals identified in the biological samples [21–23]. In short, the above-mentioned two approaches can be used to capture an individual's overall exposome. The characterization of an individual's exposome over their lifetime, however, remains a challenging task. Figure 1.1 is an illustration of the various environmental exposure sources contributing to the chemical exposome of the

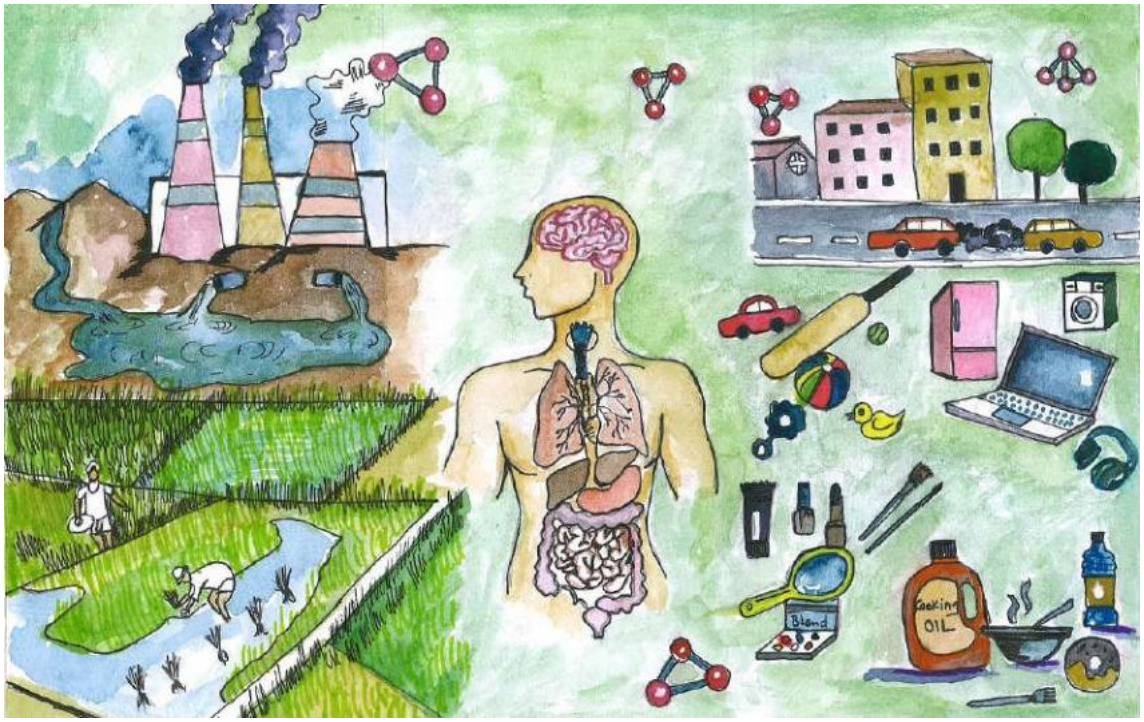


Figure 1.1: An overview of the various environmental exposure sources contributing to the chemical exposome of humankind.

humankind. In this thesis, we have employed both approaches to identify and characterize certain groups of (exogenous) chemicals in the environment that have potential to cause adverse health effects in various populations. In particular, we have studied prominent groups of chemicals of concern such as endocrine disruptors and neurotoxicants, that have received significant attention from scientists, regulatory agencies and the public due to their potential health hazards.

In recent times, several initiatives have been undertaken to establish large-scale exposome resources using bottom-up or top-down approaches, and these resources enable the regulatory authorities to prioritize environmental chemicals with potential to cause adverse effects. The Exposome-Explorer database [24], which compiles biomarkers of exposure to dietary and environmental risk factors for diseases, is one of the largest exposome resources established to date. The Human Indoor Exposome Database [25] is another manually curated exposome resource dedicated to risk factors identified in indoor dust from human exposure studies. T3DB [26] is a toxic exposome database that

contains information about toxic compounds and their target interactions. The database of intentionally added food contact chemicals (FCCdb) [27] compiles a list of chemicals used in food contact materials or food contact articles. There have also been initiatives to create exposome databases tailored to specific biological tissues or biospecimens, such as the Blood Exposome Database [28] and Saliva Exposome [29]. Moreover, Comparative Toxicogenomics Database (CTD) [30] also compiles information on environmental chemicals detected in different biospecimens. Specific to potential health impact of environmental factors on both mothers and infants, there have been a few initiatives such as the Human Early Life Exposome study in Europe [31] and the Drugs and Lactation Database (LactMed) [32, 33] of the US National Library of Medicine. Additionally, some non-profit organizations also compile information on common chemicals and exposure concerns to help mothers better understand their possible health effects on infants [34].

In this thesis, we focus on certain groups of environmental chemicals that have gained significant attention from scientists, regulatory authorities, and the general public due to their potential health concerns. Specifically, we aim to highlight the links between chemical exposome and human health. For this purpose, a systematic compilation, curation and exploration of the existing information derived from toxicological studies can aid in assessing the biological response to environmental chemical exposure. As a first step toward establishing a link between chemical exposome and human health, we identify and compile at least five groups of chemicals with toxicological relevance from published experimental studies, namely endocrine disrupting chemicals (EDCs) [35–37], environmental neurotoxicants [38], human milk contaminants [39], fragrance chemicals in children's products [40], and exogenous chemicals detected in human tissues [41]. We have employed both the bottom-up and top-down approaches to characterize the above-mentioned groups of environmental chemicals that have a potential to cause adverse health effects in humans. Furthermore, there is a growing interest in using network science and systems biology approaches to characterize the chemical exposome in order to better understand the links between environmental exposures and human biology [13, 42]. As a result, in

this thesis, we have extensively utilized network science and systems biology approaches to shed light on biological perturbations associated with exposure to diverse groups of environmental chemicals using the compiled toxicological information in our compiled resources. In addition, we have studied the exposure sources, regulatory status and the nature of compiled chemical spaces using computational approaches.

The subsequent sections of this chapter will provide an overview of the different groups of environmental chemicals studied here and a description of various analyses presented in this thesis.

1.2 Compilation and curation of diverse groups of environmental chemicals of concern

We have undertaken a systematic compilation and curation of the existing information contained in published toxicological studies on certain groups of environmental chemicals, which include endocrine disrupting chemicals (EDCs) [35–37], environmental neurotoxicants [38], human milk contaminants [39], fragrance chemicals in children's products [40], and exogenous chemicals detected in human tissues [41].

To begin, we consider the EDCs [35,36] present in the environment that are capable of interfering with the normal functioning of the human endocrine system. Binding of EDCs to the native hormonal receptors interferes with the normal endocrine signalling mechanism leading to adverse health effects related to reproduction, development, metabolism, immune system, neurological system, liver or hormone-related cancers [4, 8, 43, 44]. Notably, the estimated annual cost of disease burden and impact on healthcare due to EDCs is $340 billion in the USA and €163 billion in the European Union (EU) [43, 45]. While there have been previous attempts such as the World Health Organization (WHO) report [8], The Endocrine Disruption Exchange (TEDX) [46], EDCs Databank [47, 48] and Endocrine Disruptor Screening Program (EDSP) [49] of the United States Environmental Protection Agency (US EPA), to compile the list of potential EDCs, the earlier efforts

have not assessed the weight of evidence of endocrine disruption from existing literature, as highlighted by Solecki *et al*. [45] and the scientific statements from the Endocrine Society [43, 50, 51]. Further, none of the earlier resources on EDCs compiled the adverse health effects associated with chemical exposure that can facilitate the mechanistic understanding of endocrine disruption. In Chapter 2, we present a systematic workflow for identifying and compiling potential EDCs in the environment along with their adverse effects, from published experimental studies. In Chapter 3, we explore the current regulations and guidelines from the perspective of EDCs, which can aid in the better risk assessment. In Chapter 4, we build a comprehensive adverse outcome pathway (AOP) network relevant to endocrine disruption which can aid in understanding the systems-level endocrine-mediated perturbations resulting from exposure to EDCs.

Subsequently, we explore environmental neurotoxicants [38] whose exposure can cause a variety of neurological illnesses and neurotoxic consequences that can manifest at any stage of human life, from infancy to old age [52, 53]. The human nervous system is both complex and sensitive to environmental exposures [54, 55]. When nervous system is exposed to these chemicals, such exposure have the potential to cause permanent or irreversible damage, which can lead to a decline in brain function [55–57]. In particular, toxic chemical exposure during pregnancy or childhood has a detrimental effect on neurodevelopment and neurobehavioral processes [57]. Despite an increase in the number of chemicals introduced into commerce, only a minuscule proportion of them have been assessed for neurotoxicity [58,59]. Although there have been some efforts [57,58,60–62] to compile the list of potential neurotoxicants identified in the published literature, there was no dedicated online resource on environmental neurotoxicants specific to mammals prior to our work. In Chapter 5, we present the first comprehensive online knowledgebase on non-biogenic neurotoxicants along with their neurotoxic effects captured from published evidence specific to mammals.

Thereafter, we focus on the environmental chemicals that have potential to cause adverse health effects in children from two different perspectives. First, we explore several

environmental contaminants that are capable of entering human milk [39] and can have a potential impact on maternal health [63] and the early development of a child [64, 65]. These contaminants are mostly lipophilic, persistent and bioaccumulative in nature, and have a tendency to deposit in adipose tissue of women or mothers who are exposed to these chemicals [66, 67]. During lactation these chemicals can transfer to human milk primarily via passive diffusion [68–72]. In Chapter 6, we investigate these human milk contaminants and their potential health impact on infant and mothers. Second, we investigate fragrance chemicals in children's products to emphasize the importance of monitoring and regulating them. Exposure to fragrance chemicals can lead to asthma, contact dermatitis (irritant or allergic), dyschromia, photosensitivity, and migraine headaches [73–78]. Specifically, the exposure to hazardous chemicals is a significant health concern for children who have high metabolic rate, immature organ systems, thin skin, rapid growth and development of organs and tissues [79–81]. Despite being a subset of chemicals utilized in children's products, fragrance chemicals are either self-controlled or weakly regulated [75, 79, 81]. In Chapter 7, we present a knowledgebase on the fragrance chemicals in children's products and their potential health hazards.

Lastly, we investigate the environmental chemicals detected across different human tissues [41]. Human biomonitoring studies have enabled the measurement of these chemicals in various human biospecimens using analytical techniques [82–84]. The use of human tissues in the biomonitoring of environmental chemicals is considered the gold standard in the study of exposed populations, as they reflect the long-term exposure and bioaccumulation of environmental chemicals [85]. Existing resources [24, 28–30, 39, 86] compiling the chemicals detected in various human biospecimens do not provide a cohesive picture of chemical exposure-disease relationships specific to human tissues. In Chapter 8, we study this chemical component of the external exposome, specific to human tissues, and explore the possible exposure-disease associations.

In sum, the compilations of the above-mentioned environmental chemicals led to the development of five highly curated knowledgebases containing relevant toxicological in-

formation associated with these environmental chemicals, which can facilitate chemical risk assessment [35, 36, 38–41].

1.3 Linking exposome and health using network science approach

The growing number of chemicals in commerce necessitates the use of computational and high-throughput techniques to prioritise the subset of chemicals linked to serious health consequences [13, 87]. Data-driven exploration using published toxicological studies can facilitate the identification of biological consequences of environmental chemical exposures [87]. To comprehend the environmental and biological components of the exposome, however, a systems approach to the "paradigm of biological complexity" is necessary [87]. Network-centric techniques can aid in understanding the organizing principles of complex biological systems [88]. Furthermore, there is a recent interest to leverage network science and systems biology approaches in characterizing the chemical exposome. The use of networks, in particular, might provide a conceptual framework for capturing the intricate relationship between the environment and human health [13, 42]. In this thesis, we leverage the compiled toxicological information associated with the five groups of environmental chemicals to capture the different components of the biological system such as perturbed genes, receptors or pathways, as well as disease outcomes as a result of environmental chemical exposure (Figure 1.2). Specifically, we extensively apply network science and systems biology approaches to investigate the links between chemical exposome and human health.

Bipartite network of environmental chemicals and target genes

The U.S. Environmental Protection Agency's Toxicity Forecaster (ToxCast) [89] has screened more than 9000 chemicals using high-throughput assay experiments to capture the molecular or cellular level changes that occur as a result of individual chemical expo-

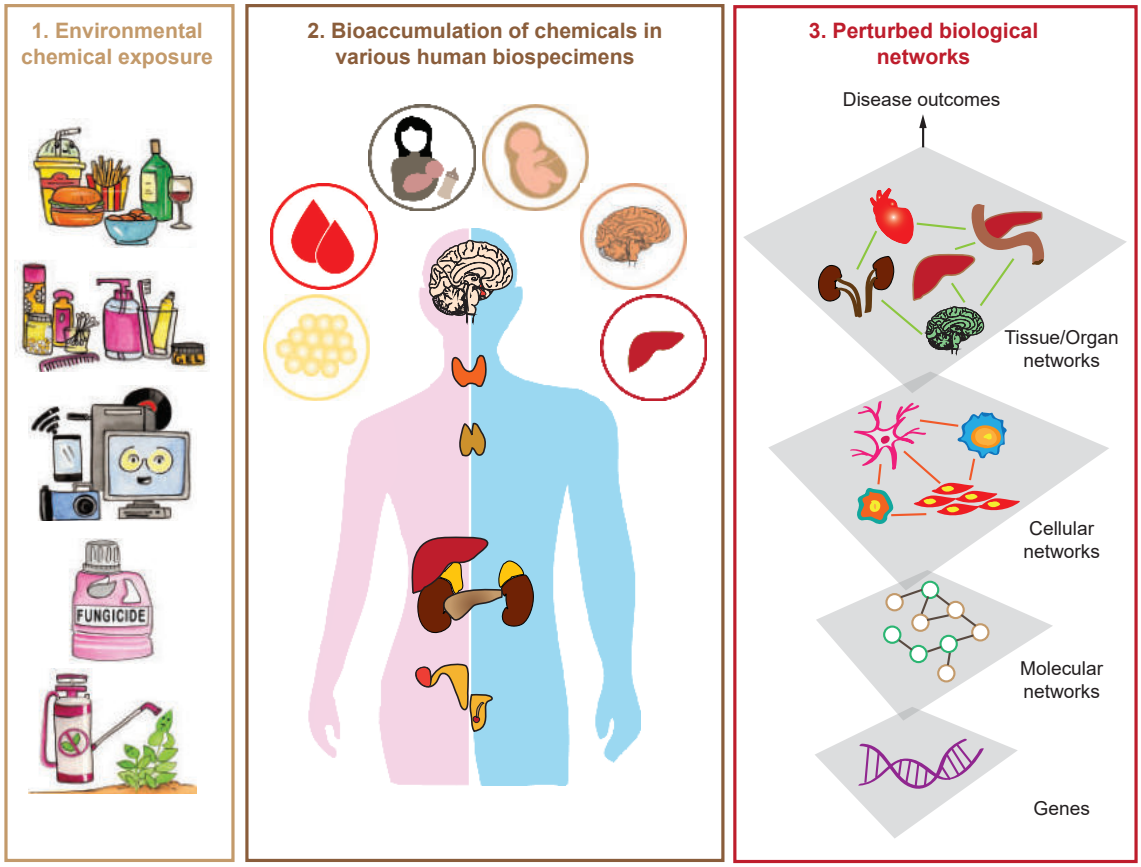


Figure 1.2: A figure depicting the complex interplay of environmental chemical exposure and perturbed biological networks at various levels of organization, which can result in disease outcomes.

sure. This data can be leveraged to prioritize chemicals using computational toxicology approaches. Apart from ToxCast, CTD [30] provides a manually curated list of chemical-gene associations compiled from the existing literature. In a toxicological context, chemicals do not affect the function of a single gene or protein, but rather they affect multiple genes or proteins at the same time. Thus, in order to better understand the aetiology of several chronic diseases, it is necessary to gather information on multiple target genes that are perturbed as a result of chemical exposure [90]. Studying the chemical-gene networks can be further helpful in understanding the various receptor-mediated processes and the potential pathways that get perturbed upon chemical exposure. Furthermore, information on molecular interactions can throw light on network-level perturbations such as in protein-protein interaction network, metabolic network, and gene regulatory network, enabling us to capture the cellular behavior at systems-scale in response to environmental exposures [88,90]. In this thesis, we have studied bipartite networks of these environmental chemicals and their target genes wherein the interactions were identified based on the *in vitro* human assays in ToxCast.

Visualizing 'Toxicity pathways' as 'Adverse Outcome Pathways'

In 2007, the U.S. National Research Council issued a vision report titled 'Toxicity testing in the twenty-first century: a vision and a strategy' [91], which included several recommendations to enhance and expedite chemical toxicity testing. The report [91] urged the use of high-throughput screening technologies such as *in vitro* toxicology, *in silico* approaches, to accomplish rapid, efficient, and cost-effective screening of chemicals [92]. In addition, the report [91] emphasized the importance of the notion of 'toxicity pathways' for the purpose of chemical risk assessment. These toxicity pathways are described as a set of cellular processes that were found to mediate toxicant-induced adverse effects [93–98]. Ankley *et al*. [99] suggested a similar framework, "Adverse Outcome Pathways (AOPs)", to gather mechanistic information on documented adverse effects in humans or wildlife following chemical exposure. AOPs can serve as a basis for In-

tegrated Approaches to Testing and Assessment (IATA), and they have the potential to identify and fill knowledge gaps, prioritize chemicals, and support regulatory decision-making [100, 101].

An AOP is defined as: "the conceptual construct that portrays existing knowledge concerning the linkage between a direct molecular initiating event and an adverse outcome at a biological level of organization relevant to risk assessment" [99] (Figure 1.3A). The Organization for Economic Cooperation and Development (OECD) established an international programme in 2012 to standardize the development and evaluation of AOPs. Following that, several studies reported the development of specific AOPs [101–103] and their applications in risk assessment, human- and eco-toxicology [97, 104–112]. Each AOP consists of two components, namely, key events (KEs) and key event relationships (KERs). A KE in an AOP is defined as: "a measurable change in biological state that is essential, but not necessarily sufficient for the progression from a defined biological perturbation toward a specific adverse outcome" [105] (Figure 1.3A). Among KEs, Molecular Initiating Events (MIEs) capture the initial molecular level interactions between chemicals or stressors and their target receptor(s), while, Adverse Outcomes (AOs) capture perturbations at the organ or higher levels of biological organization such as changes in morphology or physiology [105] (Figure 1.3A). A KER is a directed interaction between any two KEs in an AOP [97, 105, 106].

In 2014, OECD initiated AOP knowledge base (AOP-KB) [113] for the collaborative development of AOPs. AOP-Wiki [114] is an actively maintained module within AOP-KB that receives real-time updates and serves as a central repository for AOPs in various stages of development. The sharing of KEs within AOP-Wiki can result in the development of 'AOP networks'. An AOP network is defined as: "an assembly of 2 or more AOPs that share one or more KEs, including specialized KEs such as MIEs and AOs" [107] (Figure 1.3B). Recent studies [107, 110, 115–118] have highlighted the potential applicability of such AOP networks in exploring specific toxicology-related questions. The use of graph-theoretic techniques [88] to analyze such derived AOP networks can high-

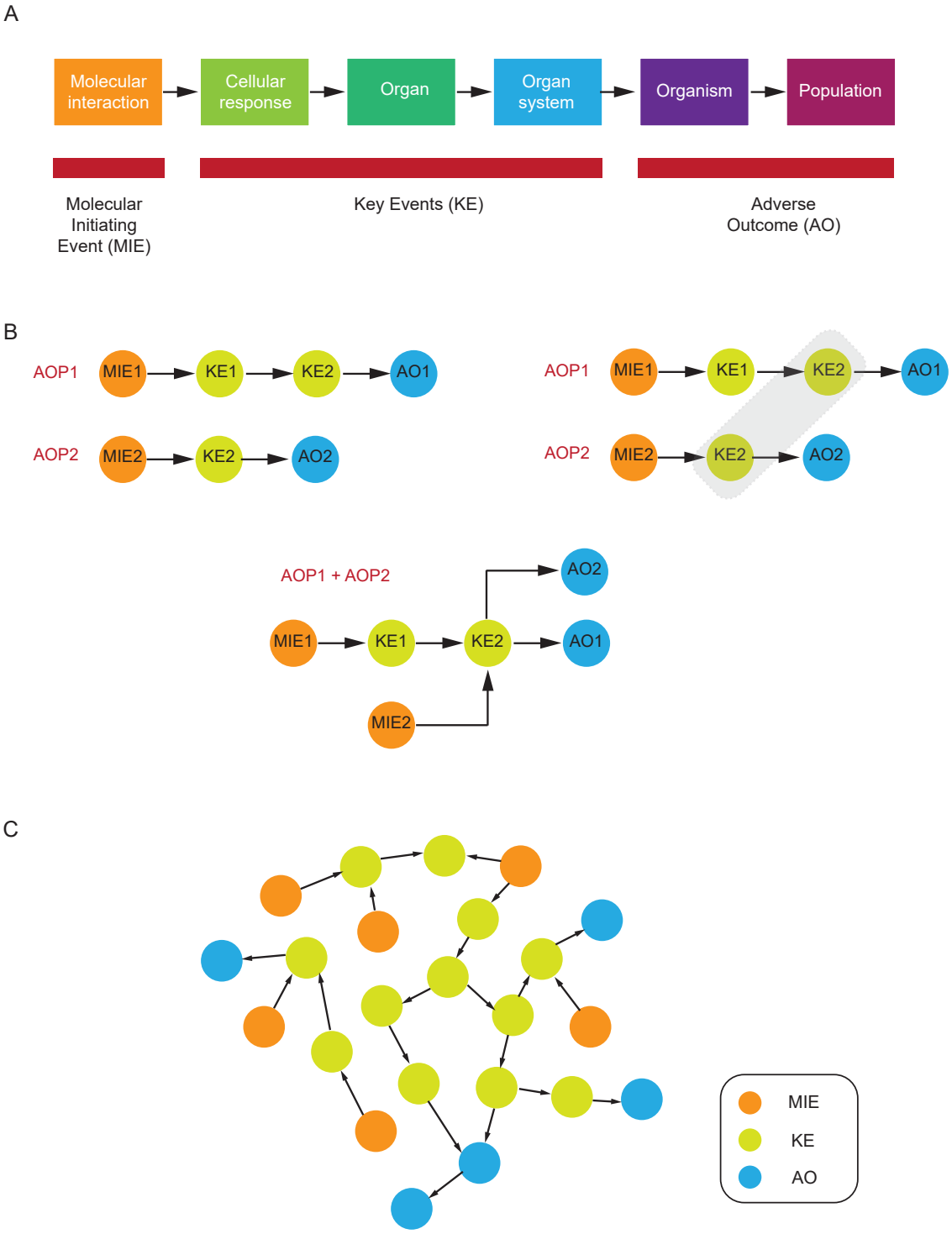


Figure 1.3: (A) Schematic representation of Adverse Outcome Pathways (AOPs) that comprise of Molecular Initiating Events (MIEs), Key Events (KEs) and Adverse Outcomes (AOs) spanning across different levels of biological organization. (B) Two AOPs can be assembled together based on shared KEs to form an AOP network. (C) An illustration of an AOP network built from existing information in AOP-Wiki, which can then be derived to study a specific research question.

light important topological features, critical paths, and relationships among individual AOPs [107, 110]. In Chapter 4 of this thesis, we develop and analyze a comprehensive AOP network relevant to endocrine disruption based on the existing information available in AOP-Wiki.

Exposome-disease associations

In human biomonitoring studies, analytical techniques like high-resolution mass spectrometry is used to assess the chemicals accumulated in diverse human biospecimens [82–84]. These biomonitoring techniques help in exposure assessment, specifically linking chemical exposures to health effects [24–33]. Existing exposome databases offer information on chemical exposures in a variety of human biospecimens, including biological fluids (such as blood, human milk, urine, and saliva) and biological non-fluids (such as the brain, placenta, and liver). Among the human biospecimens, human tissues are considered the 'gold standard' in the exposure assessment, as they reflect long-term exposure and body burden of environmental contaminants [85]. To comprehend the complexities of human exposure, it is critical to characterize tissue-specific exposomes, which can offer insight on exposure-effect correlations. The use of data-driven computational approaches, in particular, can aid in a better understanding of the interconnections, mechanistic linkages, and patterns concerning the influence of chemical exposure on human health. Recent studies have well documented the tissue-specificity of diseases [119], as well as tissue-specific gene-disease interactions relevant to cancer [120] and respiratory disorders [121]. Similarly, it is vital to establish the exposure-disease associations of chemicals detected across human tissues. Some studies have further established the effect of environmental chemicals on human biological systems and their relationship to diseases [122, 123]. However, these studies typically do not consider tissue-specific exposome data. In Chapter 8 of this thesis, we explore the relationships between tissue-specific chemical exposome and human diseases using network biology approaches.

1.4 Characterization of environmental chemical spaces

In silico or computational toxicology was originally developed for drug development. But, in recent years, it has been employed for toxicological research and risk assessment in the environmental chemical space [124]. In particular, *in silico* approaches are being employed to predict or model the toxicological mechanisms, adverse outcomes or systems-level behaviour [124]. *In silico* approaches in this direction include databases, data mining, read-across, different kinds of quantitative structure-activity relationship (QSAR) methods, molecular modelling, and network-based approaches [124, 125]. Several of these computational approaches are based on the similarity principle, which assumes that structurally similar chemicals will have similar toxicological effects [126, 127]. In particular, chemical categorization and read-across methods are widely used for risk assessment of chemicals.

Structure-based similarity analysis can aid in the understanding of the diversity of the investigated environmental chemical space. Any chemical space can be characterized by a multi-dimensional space of descriptors such as hydrophobicity, chemical connectivity, presence or absence of particular substructures, and these features can be measured experimentally or obtained computationally [128]. For this, each chemical structure is represented in the form of binary fingerprints that capture different aspects such as hydrophobicity, chemical connectivity, presence or absence of particular substructures [126, 128]. Similarity between any two chemicals is quantified using distance measures such as Tanimoto index, Dice index, Cosine coefficient and Soergel distance [126]. These distance measures typically give the chemical similarity value in the range between 0 and 1, with 0 representing no resemblance and 1 representing strong similarity. Some of the widely-used molecular fingerprints for similarity quantification include the extended connectivity fingerprints (ECFP4) [129], the MACCS keys fingerprints [130], and the Daylight-like fingerprints. Visualisation and analysis of a particular environmental chemical space by constructing chemical similarity networks (CSNs) can provide insight into the diversity

of the compiled chemical spaces [131]. In CSN, the nodes are the chemicals, and there is an edge between two nodes (chemicals) if they share certain level of structural similarity. To this end, we have constructed CSNs for various groups of environmental chemicals studied in this thesis, and further, have evaluated the structural diversity of associated chemical spaces.

In addition to the chemical structure similarity, we have leveraged the predicted chemical classification, predicted physicochemical properties, and predicted absorption, distribution, metabolism, and excretion (ADME) properties to characterize the compiled environmental chemical spaces studied in this thesis.

1.5 Regulatory assessment of environmental chemicals

To address vast inventories of existing chemicals as well as emerging new chemicals in commerce, rapid and effective chemical risk assessment is required [132]. Concerns about chemicals in various items have spurred proposals for a reform of the laws that govern toxic substances. As a result, the European Union and the United States of America have recently enacted legislation to increase regulation of toxic chemicals [133]. Following that, many regulatory bodies have been established to address the hazard assessment of chemicals related to various exposure sources including dietary exposures [134–140], skin-related products [141–144], children-related exposures [145–148], or occupational exposures [149]. For example, the US Department of Labor Occupation Safety and Health Administration (OSHA) has identified toxic and highly reactive hazardous chemicals that are of concern under the Occupational Safety and Health Standards [149]. Moreover, the Organisation for Economic Cooperation and Development (OECD) High Production Volume (HPV) list [150], the United States High Production Volume (USHPV) database [151] and REACH registered substances [152] provide a list of high production volume (HPV) chemicals depending on the quantity of a chemical manufactured or imported annually. To draw a list of high priority chemicals, it is important to evaluate

the publicly available scientific and regulatory sources of toxicity information [153]. The presence of diverse groups of environmental chemicals in the existing chemical lists representing the current chemical regulations, guidelines or inventories can also reflect the gaps in the current regulation across various exposure sources. To this end, comparative studies for food, food additives and food contact compounds have been performed [154, 155], and these studies have revealed inadequacies in current regulation that lead to the inclusion of substances of concern in food-related products.

In this direction, we have compiled the publicly available chemical lists representing current regulations, guidelines or inventories in this thesis, and thereafter, classified the chemical lists according to various exposome categories. Thereafter, we have explored the presence of the five groups of environmental chemicals studied in this thesis, across the chemical lists representing current regulations, guidelines or inventories, in order to assess the current regulatory status of the different groups of environmental chemicals.

1.6 Thesis organization

The remaining chapters of this thesis are organized as follows:

Chapter 2 presents a detailed workflow designed to identify EDCs with supporting evidence of endocrine disruption in published experiments in humans or rodents. Importantly, we have also collated the observed adverse effects or endocrine-specific endpoints along with dosage information, for the potential EDCs from the supporting published experiments. In order to enable future research based on this compiled information on potential EDCs, we have built an online knowledgebase, Database of Endocrine Disrupting Chemicals and their Toxicity profiles (DEDuCT 1.0), accessible at: https://cb.imsc.res.in/deduct/ [35]. In this chapter, we also describe the network-centric analysis of the chemical space and the associated biological space of target genes of EDCs. **The work reported in this chapter is contained in the published manuscript [35].**

Chapter 3 presents an overview of the updated knowledgebase DEDuCT 2.0, and an investigation of the current regulations and guidelines from the perspective of EDCs. In this chapter, we sought to understand how scientific knowledge from academic research could be used to improve chemical regulation, with an emphasis on EDCs. We expand our comparative analysis with various chemical lists and classifying them based on an influential report commissioned by the European Parliament [156]. To understand the scale of exposure and the related hazard potential, we analyze which of these potential EDCs in human use are produced in large volumes. Lastly, we also demonstrate how the compiled information in curated knowledgebases like DEDuCT 2.0 can aid in the risk assessment of EDCs using an example. **The work reported in this chapter is contained in the published manuscript [36].**

Chapter 4 presents the steps involved in the characterization, development and investigation of an adverse outcome pathway (AOP) network derived to capture the endocrine-mediated perturbations resulting from environmental exposure. In this chapter, we assess the quality and completeness of information of each AOP compiled in AOP-Wiki [114], and thereafter, identify high-confidence AOPs relevant to endocrine disruption (ED-AOPs). The identified ED-AOPs were used to construct an ED-AOP network by assembling the information on shared KEs and KERs among them. We further utilize a graph-theoretic approach to study the ED-AOP network and identify critical biological events perturbed upon endocrine disruption. Besides, we also study the systems-level perturbations caused by endocrine disruption, emergent paths, and stressor-event associations. **The work reported in this chapter is contained in the manuscript [37].**

Chapter 5 presents a detailed workflow to identify and compile potential non-biogenic neurotoxicants with evidence specific to mammals from published literature. This compilation led to the creation of environmental Neurotoxicants Knowledgebase NeurotoxKb 1.0, which is accessible at: https://cb.imsc.res.in/neurotoxkb. In this chapter, we also explore the possible source or route of human exposure to environmental neurotoxicants using different analyses. For instance, we analyze the presence of

compiled neurotoxicants in various chemical lists representing regulations, guidelines or inventories. We also characterize the associated chemical space by constructing a chemical similarity network. **The work reported in this chapter is contained in the published manuscript [38].**

Chapter 6 describes the detailed steps involved in the creation of **Ex**posome of **Hu**man **M**ilk across **Ind**ia (ExHuMId) version 1.0, an India-specific repository compiling environmental contaminants detected experimentally in human milk samples across various Indian states. ExHuMId 1.0 is accessible at: https://cb.imsc.res.in/exhumid/. In this chapter, motivated by Vasios *et al.* [72], we also explore the propensity of the compiled environmental contaminants to transfer into human milk based on the physicochemical properties. We also analyze the potential effect of the human milk contaminants on the lactation pathway and cytokine signalling and production pathway, using a systems biology approach. **The work reported in this chapter is contained in the published manuscript [39].**

Chapter 7 presents a detailed overview on the repository of Fragrance Chemicals in Children's Products (FCCP) that compiles fragrance chemicals from published experimental studies. FCCP is accessible at: https://cb.imsc.res.in/fccp/. Since the fragrance chemicals in children's products are known to be poorly regulated, we sought to explore the current regulatory status of these chemicals and the potential health effects in children upon exposure in this chapter. Further, we analyze the structural diversity of the space of compiled fragrance chemicals and banned allergenic fragrance chemicals in EU Toy Safety Directive [145]. **The work reported in this chapter is contained in the published manuscript [40].**

Chapter 8 describes a Human **T**issue-specific **Ex**posome **A**tla**s** (TExAs), a compilation of environmental chemicals detected across different human tissues in published studies. TExAs is accessible at: https://cb.imsc.res.in/texas. In this chapter, we explore the patterns in the associations between tissue-specific chemical exposures and human diseases using a network biology approach. We analyze the source and route of

human exposures to environmental chemicals detected in human tissues, as well as the current status of their monitoring and regulation. Further, we propose a priority list of potentially hazardous chemicals based on a comparative analysis of TExAs with SVHC REACH regulation [157] and high production volume chemicals. **The work reported in this chapter is contained in the published manuscript [41].**

Chapter 9 concludes this thesis with a brief summary of the research reported across different chapters. The chapter also discusses the future prospects and the scope of our efforts in identifying, compiling and characterizing different classes of environmental chemicals, and linking them to potential health hazards in humans.

Chapter 2

DEDuCT 1.0: A curated knowledgebase on endocrine disrupting chemicals and their biological systems-level perturbations

In this chapter, we focus on a prominent group of chemicals of concern in the environment, namely, Endocrine disrupting chemicals (EDCs). EDCs interfere with the normal functioning of the human endocrine system and can lead to adverse effects related to reproduction, development, metabolism, immune system, neurological system, liver or hormone-related cancers [8, 44, 45]. EDC exposure can alter hormonal imbalance in humans through different mechanisms. For example, EDCs can mimic the natural hormones and bind to their respective nuclear receptors either as an agonist or an antagonist [4, 43]. So far there is a lack of biological systems or pathway level understanding of the different mechanisms via which specific EDCs alter the hormonal homeostasis.

For the risk assessment of EDCs, an important limitation is the lack of availability of validated test systems for their identification [43, 45]. This has hampered both researchers

and policymakers to reach a consensus agreement on identification of EDCs and the characterization of their endocrine disruption mechanisms [43, 45]. In this direction, Solecki *et al*. [45] have outlined a detailed consensus statement on the scientific principles that can form a basis for the identification of EDCs and their disruption mechanism. Furthermore, the scientific statements by the endocrine society [43, 50, 51] provide principles for better understanding of disruption mechanisms by EDCs.

Given the potential risk from EDCs in our environment, there have been multiple efforts towards their compilation which include the World Health Organization (WHO) report [8], The Endocrine Disruption Exchange (TEDX) [46] and EDCs Databank [47, 48] and Endocrine Disruptor Screening Program (EDSP) [49] of United States Environmental Protection Agency (US EPA). However, these existing resources on potential EDCs consider evidence for endocrine disruption upon exposure from disparate types of published studies. Specifically, the WHO report and TEDX contain manually curated information on EDCs based on published literature evidence including *in vivo*, *in vitro*, *in silico*, environmental monitoring and epidemiological studies while EDCs Databank compiles EDCs from the TEDX and the EU list of potential endocrine disruptors followed by PubMed [158] search to associate literature evidence with EDCs. Another important limitation of these existing resources on potential EDCs is the lack of systematic effort to compile the observed adverse effects specific to endocrine disruption in supporting published experiments.

In this chapter, we describe our curated knowledgebase namely, **D**atabase of **E**ndocrine **D**isr**u**pting **C**hemicals and their **T**oxicity profiles (DEDuCT), which compiles 686 potential EDCs that were identified using a detailed four-stage workflow from published experimental evidence for endocrine disruption in humans or rodents [35]. **The work reported in this chapter is contained in the published manuscript [35].**

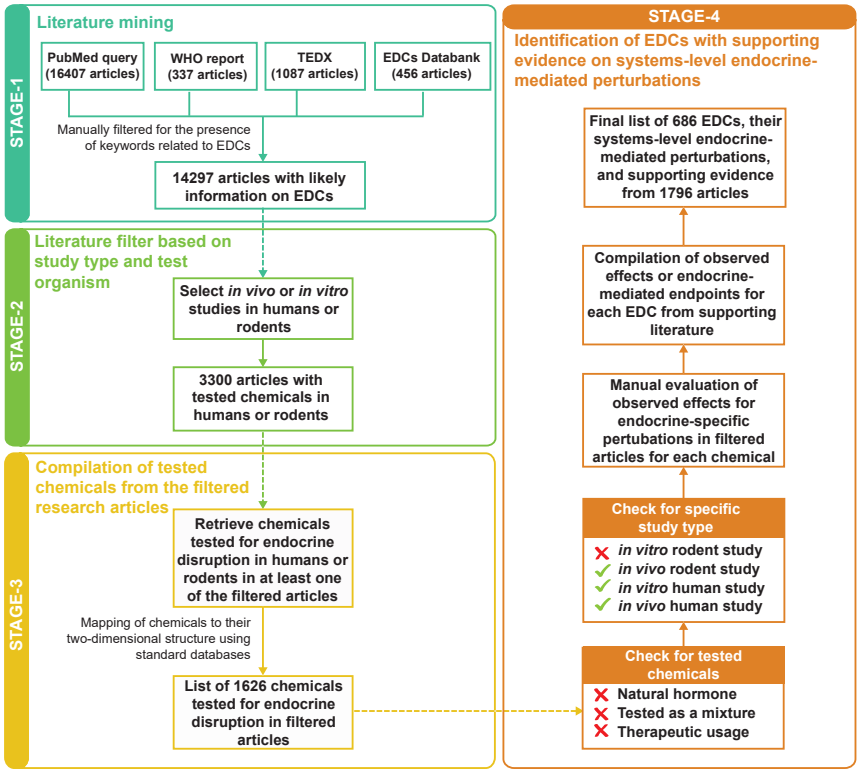


Figure 2.1: Detailed workflow with four stages to identify potential EDCs from published research articles containing supporting experimental evidence of systems-level endocrine-mediated perturbations in humans or rodents.

2.1 Workflow for the identification of EDCs

Based on the consensus statement by Solecki *et al.* [45] and the scientific statement by the endocrine society [43, 50, 51], we have developed a detailed flowchart to identify EDCs from published research articles containing supporting experimental evidence of systems-level endocrine-mediated perturbations in humans or rodents (Figure 2.1). Our workflow for the identification of EDCs can be divided into four stages which are described below [35].

2.1.1 Literature mining

In stage 1, we performed an extensive literature search to compile 14297 published research articles which are likely to contain information on EDCs (Figure 2.1).

Firstly, we mined PubMed [158] using the following keyword search:

"EDCs" OR "EDC" OR ("endocrine" AND "disrupt") OR ("disrupt" AND "endocrine") OR "endocrine disruptors" OR "endocrine-disruptors" OR "endocrine disruptor" OR "endocrine-disruptor" OR "endocrine disrupters" OR "endocrine-disrupters" OR "endocrine disruption" OR "endocrine-disruption" OR "endocrine disruptive" OR "endocrine-disruptive" OR "endocrine disrupting" OR "endocrine-disrupting" OR "endocrine disrupter"

The above query was designed to filter abstracts on EDCs from PubMed, and this keyword search in February 2018 led to 16407 research articles. Secondly, we compiled research articles from three existing resources on EDCs, namely, the WHO report [8], TEDX [46] and EDCs Databank [47, 48]. Specifically, the WHO report, TEDX and EDCs Databank captured information from 337, 1087 and 456 research articles, respectively.

Subsequently, we manually filtered the compiled abstracts from PubMed query, WHO report, TEDX and EDCs Databank for the presence of keywords such as endocrine disruptors or endocrine disrupters or endocrine disrupting or endocrine disrupting chemicals or EDC or EDCs. In particular, we check that the acronym EDC in a filtered abstract refers to endocrine disrupting chemicals. For example, we found that the acronym EDC in certain abstracts may refer to irrelevant terms such as electric dynamic catathermometer or expected delivery cesarean or endothelium-derived contracting. This manual filtration of abstracts based on presence of keywords relevant to endocrine disruption studies led to 14297 research articles at the end of the stage 1 (Supplementary Table S2.1). Of these 14297 research articles at the end of stage 1, 12879 are not captured in existing resources, namely, WHO report, TEDX or EDCs Databank [35].

2.1.2 Literature filter based on study type and test organism

In stage 2, we screened the 14297 research articles from stage 1 to select studies based on *in vivo* or *in vitro* experiments in humans or rodents (Figure 2.1). Here, we have excluded

published studies where receptor-based binding assays or *in silico* methods are employed to infer the potential endocrine disruption by a chemical using binding affinity or bioactivity information. Such binding affinity or bioactivity values do not provide sufficient information on whether chemical exposure can actually lead to adverse effects due to endocrine disruption [159]. We have also excluded human epidemiological studies due to insufficient mechanistic evidence linking observed adverse effects to potential endocrine disruption upon chemical exposure [160, 161]. The filtration based on study type and test organism led to a subset of 3300 research articles at the end of stage 2 (Supplementary Table S2.2). Of these 3300 research articles at the end of stage 2, 2394 are not captured in existing resources, namely, WHO report, TEDX or EDCs Databank [35].

In this work, we do not include information from two existing resources on EDCs, namely, the Endocrine Disruptor Knowledge Base (EDKB) [162] and Endocrine Disruptor Screening Program (EDSP) of the United States Environmental Protection Agency (US EPA). EDKB compiles EDCs based on multiple receptor binding assays and *in silico* QSAR studies, and such evidence is ignored in our workflow to identify EDCs (Figure 2.1). EDSP screens chemicals based on several hormonal assays in test organisms such as human, rat, fish and amphibians to determine its potency to interact with the human endocrine system. EDSP identifies a chemical to be an EDC if the chemical displays consistent evidence of endocrine disruption across all hormonal assays carried out by them. As highlighted by Zoeller *et al*. [43], the weight of evidence used by EDSP to identify EDCs is too stringent which leads to omission of several chemicals with significant endocrine-specific effects. Specifically, in the EDSP Tier 1 screening of 52 chemicals, 18 were determined to have conclusive evidence for endocrine disruption while 34 have inconclusive evidence. However, a closer inspection of the 34 chemicals determined by EDSP to have inconclusive evidence finds well-known EDCs such as Chlorpyrifos and 2,4-Dichlorophenoxyacetic acid highlighted by the WHO report and the Endocrine society [163]. Thus, we decided not to include information from EDSP in our resource.

2.1.3 Compilation of tested chemicals from the filtered research articles

In stage 3, we gathered the set of chemicals tested for potential endocrine disruption in any of the 3300 research articles from stage 2. Moreover, we also gathered information on the two-dimensional (2D) structure of each tested chemical using PubChem [86] and Chemical Abstracts Service (CAS) [164] databases (Figure 2.1). Note that we have omitted any tested chemical in the 3300 research articles which could not be mapped to a chemical identifier in standard chemical databases. At the end of stage 3, we compiled 1626 chemicals along with their 2D structures that were tested for endocrine disruption in humans or rodents in at least one of the filtered research articles from stage 2 (Supplementary Table S2.3) [35].

2.1.4 Identification of potential EDCs with supporting evidence for systems-level endocrine-mediated perturbations

In stage 4, we identify potential EDCs among the 1626 chemicals compiled in stage 3 by assessing the significance of observed effects for endocrine disruption upon exposure in published experiments in humans or rodents (Figure 2.1).

Prior to this assessment of supporting evidence for endocrine disruption upon chemical exposure, we excluded a tested chemical or its published experiment based on the following criteria (Figure 2.1):

1. Chemical is a natural hormone.

2. Chemical was tested as part of a mixture in the published experiment. This criterion reflects our choice to include chemicals which as single entities can cause endocrine disruption upon exposure.

3. Chemical was tested for therapeutic relevance in the published experiment.

Moreover, we excluded published experiments which contain evidence for endocrine dis-

ruption upon chemical exposure in an *in vitro* rodent system. Since the observed effects in an *in vitro* rodent system do not adequately reflect the complexities observed in humans, the last criterion omits such evidence in the published literature (Figure 2.1). For the next phase of the workflow, we filtered chemicals and their associated literature which pass the above-mentioned criteria.

For each chemical which passed the above-mentioned criteria, we next evaluated the level of supporting evidence for endocrine disruption in humans or rodents upon exposure based on published experiments contained in the filtered research articles. For this evaluation, we manually compiled the observed effects upon exposure of each chemical in associated published experiments in humans or rodents. A published experiment in humans or rodents is considered as strong supporting evidence for endocrine disruption by a chemical if the chemical upon exposure leads to observed effects or endpoints related to endocrine-specific perturbations such as changes in morphology, physiology, growth, reproduction, development and lifespan [8]. Thereafter, if a chemical has at least one published experiment with strong supporting evidence for endocrine disruption upon exposure, then it is identified as a potential EDC in stage 4 of the workflow. At the end of stage 4, we identified 686 potential EDCs with supporting evidence of endocrine-mediated perturbations in published literature spanning 1796 research articles (Supplementary Table S2.4) [35].

2.1.5 Compilation of endocrine-mediated endpoints and their classification into systems-level perturbations

For the identification of EDCs, we have manually compiled the observed effects or endpoints related to endocrine-specific perturbations reported in published experiments on chemical exposure in humans or rodents (Figure 2.1). This compiled list of observed effects or endpoints was then used to assess the level of supporting evidence for endocrine disruption upon chemical exposure. In order to standardize the reported evidence for

an EDC, we undertook an extensive manual effort to unify the biological terms used to describe the observed effects or endpoints related to endocrine-specific perturbations in published experiments upon chemical exposure.

This standardization effort led to a comprehensive list of 514 endocrine-mediated endpoints which refer to the adverse effects such as changes in morphology, physiology, growth, reproduction, development and lifespan that may be observed in experiments after the administration or ingestion of a tested chemical (Supplementary Table S2.5). For the 686 EDCs, we have also compiled the observed adverse effects in terms of these 514 endocrine-mediated endpoints from published experiments in supporting literature [35].

EDCs perturb the normal functioning of the human endocrine system which consists of several glands that secrete hormones which in turn regulate diverse biological functions such as development, growth, reproduction, metabolism, immunity and behaviour [165, 166]. Hence, exposure to EDCs can have adverse effects in several biological processes regulated by the human endocrine system (Figure 2.2). In addition, the endocrine-related processes perturbed by EDCs can also induce cancer in humans [8, 50, 51]. Motivated by the major biological processes controlled by the human endocrine system, we have classified the 514 endocrine-mediated endpoints into 7 systems-level perturbations which are:

1. Reproductive endocrine-mediated perturbations (RT)
2. Developmental endocrine-mediated perturbations (DT)
3. Metabolic endocrine-mediated perturbations (MT)
4. Immunological endocrine-mediated perturbations (IT)
5. Neurological endocrine-mediated perturbations (NT)
6. Hepatic endocrine-mediated perturbations (HT)
7. Endocrine-mediated cancer (CT)

In Supplementary Table S2.5, we list the 514 endocrine-mediated endpoints and their categorization into 7 systems-level endocrine-mediated perturbations in DEDuCT 1.0 [35]. Figure 2.3A shows the occurrence of these 7 systems-level perturbations in the support-

ing published experiments for the 686 EDCs in DEDuCT 1.0 [35]. Among the 686 EDCs in DEDuCT 1.0 [35], it is seen that 535 have supporting evidence for reproductive perturbations and 315 for metabolic perturbations (Figure 2.3A). Thus, majority of EDCs in DEDuCT 1.0 have supporting evidence for adverse effects on the reproductive system followed by metabolism [35].

We highlight that future studies and toxicological databases can leverage our comprehensive list of endocrine-mediated endpoints and their categorization into 7 systems-level perturbations while reporting or documenting the adverse effects related to endocrine disruption from experiments related to chemical exposure. Hence, our work also contributes towards development of a unified biological vocabulary to describe toxicity profiles of chemicals.

2.1.6 Compilation of dosage information for observed endocrine-mediated endpoints

In stage 4 of the workflow, we have also compiled the dosage values for each EDC at which the endocrine-mediated endpoints are observed in the published experiments (Figure 2.1). Firstly, we have gathered the test dosage values for each EDC in appropriate units from the published experiments. Secondly, we have identified the effective dosage value among the test dosage values at which a particular endocrine-mediated endpoint is observed upon EDC exposure in the published experiment. Thirdly, the published experiments with supporting evidence for endocrine disruption by EDCs employ different units to report the test and effective dosage values. Thus, we undertook a significant effort to convert and express the test and effective dosage values taken from published experiments on EDCs in a uniform format wherever possible.

Based on this effort, we realized that the different units used to report the test and effective dosage values of EDCs in published experiments can be classified into two broad categories:

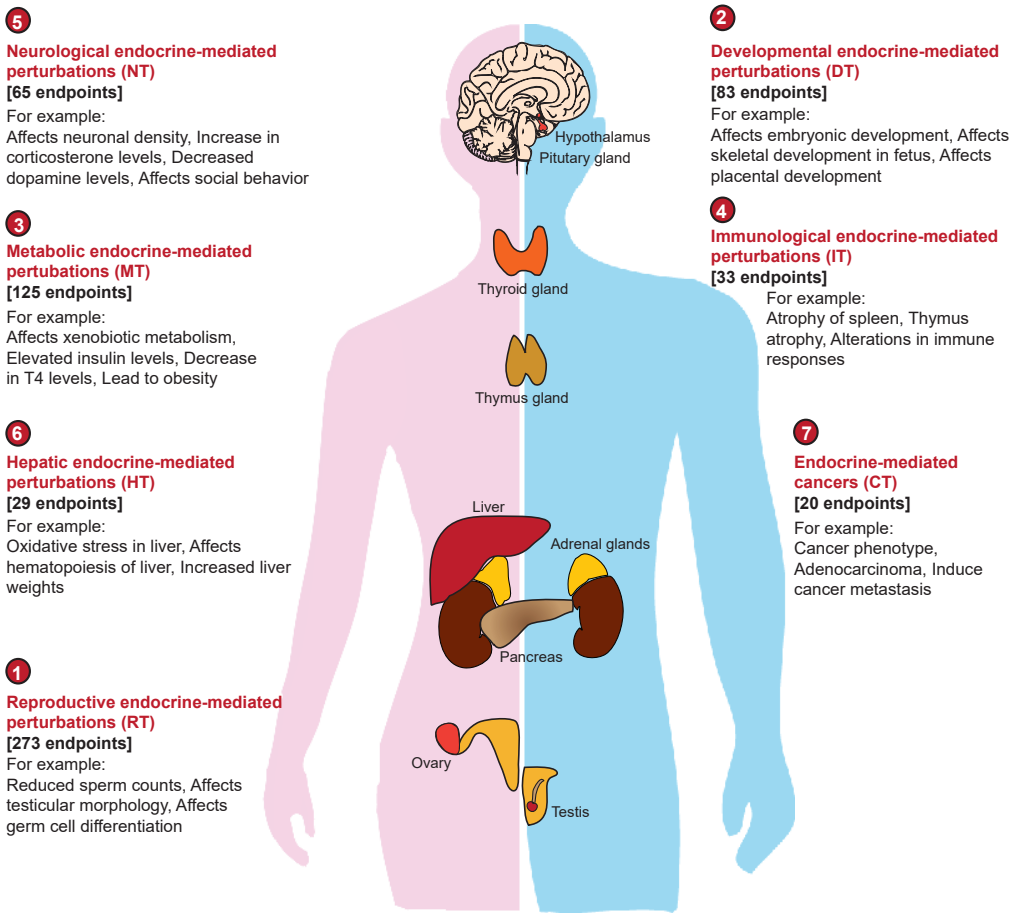


Figure 2.2: Schematic figure depicting the classification of the 514 endocrine-mediated endpoints into 7 systems-level perturbations in DEDuCT 1.0. Note that this classification of endpoints into systems-level perturbations is overlapping, that is, a given endpoint may fall into more than one systems-level perturbations.

1. Dose which gives the amount of chemical that is administered directly to the test organism in the experiment.

2. Concentration which gives the amount of chemical present in another substance such as food, soil or water that is administered to the test organism in the experiment.

Moreover, only a fraction of the published experiments on EDCs report dosage values normalized by the body weight of the individual test organism and duration of exposure [167]. For example, if a published experiment on EDC reports the dosage value in the unit mg/kg/day then this gives the amount of chemical administered per kg of the body weight of the test organism per day.

Due to the above-mentioned limitations, we were able to convert the different units used in published experiments to report the dosage values of EDCs into 19 standardized units. Supplementary Table S2.6 lists these 19 standardized units which were used to compile the dosage values of EDCs specific to endocrine-mediated endpoints from published experiments. For each EDC, we have compiled the test and effective dosage values specific to endocrine-mediated endpoints in standardized units, and this information is readily available via the DEDuCT webserver.

NOAEL and LOAEL information for EDCs

Natural hormones in human body can carry out their physiological functions at very low concentration. EDCs are known to interfere with the endocrine system by mimicking the natural hormones. Thus, it is important for risk assessment of EDCs to understand the adverse effects caused by their low dose exposure [168–170]. In this direction, our compilation of the test and effective dosage values for EDCs in DEDuCT 1.0 from published experiments can be leveraged to elucidate such low dose effects. Specifically, we have used the test and effective dosage values for EDCs in DEDuCT 1.0 to determine the following dose-response measures [51, 168]:

1. No Observed Adverse Effect Level (NOAEL) gives the highest dose of an EDC at which no observed effects or endocrine-mediated endpoints are seen in the published ex-

periments.

2. Low Observed Adverse Effect Level (LOAEL) gives the lowest dose of an EDC at which any one of the observed effects or endocrine-mediated endpoints are seen in the published experiments.

Note that the supporting evidence for the EDCs in DEDuCT 1.0 has been compiled from three different types of published experiments, namely, *in vivo* or *in vitro* experiments in humans or *in vivo* experiments in rodents. In cases where the supporting evidence for an EDC comes from more than one type of published experiment, we determine the NOAEL and LOAEL values for the EDC separately for different types of published experiments (Supplementary Table S2.7). Moreover, the supporting evidence for an EDC in DEDuCT 1.0 may come from published experiments employing different units to specify test and effective dosage values. In such cases, we determine the NOAEL and LOAEL values for the EDC separately for different standardized units across the published experiments (Supplementary Table S2.7). Note that we did not compile information on the route and duration of EDC exposure from published experiments in DEDuCT. Supplementary Table S2.7 lists the NOAEL and LOAEL values for EDCs in DEDuCT 1.0.

2.1.7 Classification of EDCs

Based on the type of supporting evidence in published experiments

We have classified the 686 EDCs in DEDuCT 1.0 into 4 categories based on the type of supporting evidence in published experiments. EDCs in category I have supporting evidence from *in vivo* human experiments, category II from *in vivo* rodent and *in vitro* human experiments but not from *in vivo* human experiments, category III from only *in vivo* rodent experiments, and category IV from only *in vitro* human experiments (Supplementary Table S2.8). Thus, potential EDCs in category I have the highest level of supporting evidence in published experiments followed by category II, III and IV, respectively. Of the 686 EDCs in DEDuCT 1.0, 7, 142, 367 and 170 are in category I, II, III and IV, respec-

tively (Supplementary Table S2.8). These 142, 367 and 170 potential EDCs in categories II, III and IV, respectively, in DEDuCT 1.0 require additional experimentation and further risk assessment for their potential risk to humankind [35].

We then compared potential EDCs in each category (I-IV) to the safer chemical ingredients list (SCIL) developed and released by the US EPA as part of its safer choice program [171]. US EPA has identified 931 chemicals in SCIL to be 'safe' based on their functional use categories. In SCIL, US EPA has labelled chemicals of low concern by green circle, chemicals of low concern for which additional data is required by green half-circle, chemicals satisfying safer choice criteria only for a particular functional use while possibly displaying hazardous profile in other uses by yellow triangle, and chemicals unsuitable for use in consumer products by grey square. We have compared the subset of 930 SCIL chemicals labelled by green circle or green half-circle or yellow triangle with the 686 potential EDCs in DEDuCT 1.0.

We find that 10 out of the 686 potential EDCs in DEDuCT 1.0 to be also in the SCIL (Figure 2.3B). None of these 10 potential EDCs in SCIL are listed under category I EDCs in DEDuCT 1.0 with supporting evidence for endocrine disruption from *in vivo* human experiments. Of these 10 potential EDCs, 1, 7 and 2 are in category II, III and IV, respectively. Benzyl salicylate is the only chemical in SCIL that is listed as category II EDC in DEDuCT 1.0 with supporting evidence for endocrine disruption from *in vivo* rodent and *in vitro* human experiments while lacking evidence from *in vivo* human experiments. As Benzyl salicylate is labelled by yellow triangle in SCIL based on the functional use category of fragrances, this suggests that this chemical may have potential to display hazardous profile in other use categories. For improved risk assessment, there is need to further evaluate and gather additional evidence for potential EDCs listed in the SCIL [35].

We have also compared the list of 3312 inactive ingredients used in US Food and Drug Administration (FDA) approved drug products from inactive ingredient database [172] with 686 potential EDCs in DEDuCT 1.0 [35]. Inactive ingredients in a drug are the chemicals that do not have any pharmacological effect and these include colorants, drug preser-

vatives and flavouring agents. We find that 44 of the 686 potential EDCs in DEDuCT 1.0 are used as inactive ingredients in FDA approved drugs (Figure 2.3B). None of these 44 potential EDCs are listed under category I EDCs in DEDuCT 1.0. Of 44 potential EDCs in FDA inactive ingredients list, 7 chemicals (Caffeine, Trichloroethylene, Diethyl phthalate, Butyl p-hydroxybenzoate, Methyl p-hydroxybenzoate, Ethyl p-hydroxybenzoate, Butylated hydroxyanisole) are in category II, 30 in category III, and 7 in category IV of DEDuCT 1.0. For better risk assessment, these 44 potential EDCs in FDA inactive ingredients list require additional evidence from *in vivo* human experiments considering the effective dosage, route of exposure, and duration of exposure [35].

Based on the environmental source

Based on the environmental source of EDCs, we have classified the 686 EDCs into 7 broad categories, namely, 'Agricultural and farming', 'Consumer products', 'Industry', 'Intermediates', 'Medicine and health care', 'Natural sources', and 'Pollutant' (Figure 2.4). Furthermore, the 7 broad categories of EDCs were further classified into 48 subcategories (Figure 2.4). Note that this environmental source-based classification of EDCs is overlapping, that is, a given EDC may belong to multiple broad or sub-categories. Majority of EDCs in DEDuCT 1.0 are used in 'Consumer products' (Figure 2.4).

Based on chemical structure

We have employed the web-based application ClassyFire [173, 174] to obtain a chemical classification of the 686 EDCs in DEDuCT 1.0. Note that ClassyFire [174] gives a non-overlapping hierarchical chemical classification based on the structure and composition of the molecule. Using ClassyFire, the 686 EDCs in DEDuCT 1.0 were classified into two chemical kingdoms, namely, organic and inorganic compounds (Figure 2.5). Moreover, the EDCs in the organic kingdom can be further classified into 19 super-classes while those in the inorganic kingdom fall into 3 super-classes (Figure 2.5). Of the 686 EDCs in DEDuCT 1.0, 646 are organic and 40 are inorganic (Figure 2.5A). Among the 646 organic EDCs in DEDuCT 1.0, the largest fraction belongs to the chemical super-class

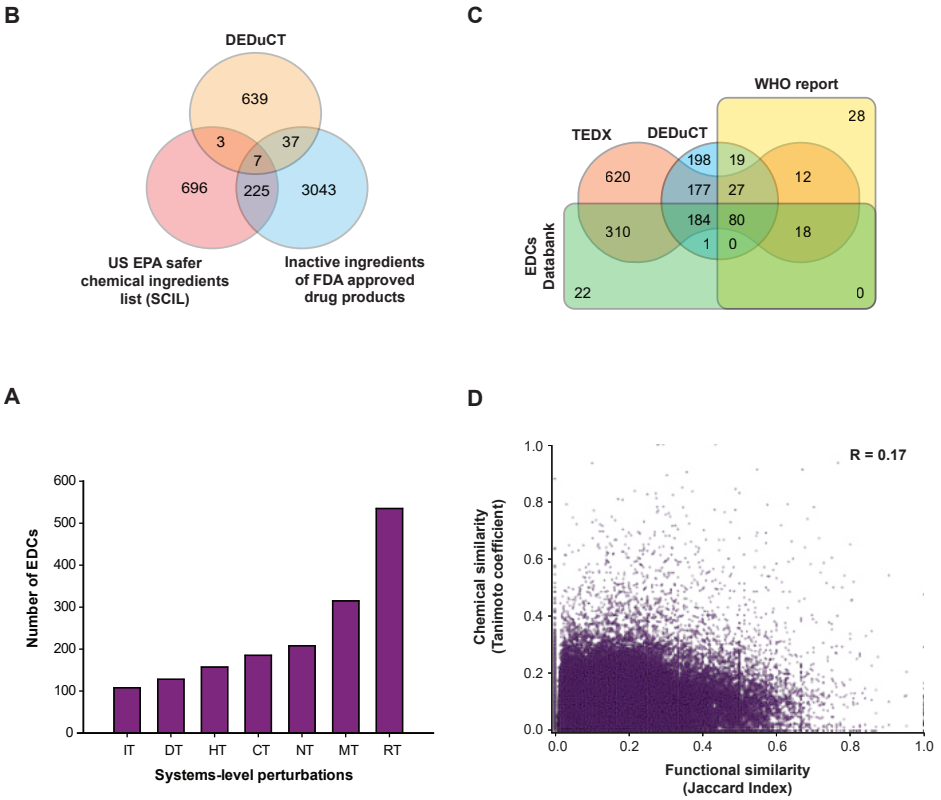


Figure 2.3: (A) Histogram shows the occurrence of 7 systems-level perturbations in the supporting evidence compiled from published experiments for the 686 EDCs in DEDuCT 1.0. Majority of EDCs in DEDuCT 1.0 have adverse effects on the reproductive system followed by metabolism. (B) Comparison of the 686 EDCs in DEDuCT 1.0 with the US EPA SCIL and the FDA inactive ingredients list. 10 EDCs are present in the SCIL while 44 EDCs are present in FDA inactive ingredients list. (C) Comparison of the 686 EDCs in DEDuCT 1.0 with those in the WHO report, TEDX and EDCs Databank. From the Venn diagram, it is seen that 198 EDCs in DEDuCT 1.0 are not captured in the three other existing resources. (D) Scatter plot of target similarity versus chemical structure similarity between pairs of EDCs. Here chemical structure similarity was computed using Tanimoto coefficient with ECFP4 fingerprint. We find no significant correlation (Pearson correlation coefficient $R = 0.17$) between the structural and target similarity of EDCs.

Figure 2.4: Classification of the 686 EDCs in DEDuCT 1.0 into 7 broad categories and 48 sub-categories based on their source in the environment. In this figure, the number of EDCs in each category or sub-category is reported within the parenthesis.

Benzenoids (Figure 2.5A). In Figure 2.5B, we show the chemical structure of a representative EDC in each chemical super-class with at least 10 potential EDCs in DEDuCT 1.0 [35].

2.1.8 Physicochemical properties and molecular descriptors

For the 686 EDCs in DEDuCT 1.0, we obtained the 2D chemical structure from Pubchem and CAS databases. Thereafter, Balloon [175, 176] and Open Babel [177, 178] with Merck Molecular Force Field (MMFF94) were used to generate the lowest energy three-dimensional (3D) structure of the EDCs. RDKit [179] and Open Babel [177, 178] were used to compute the basic physicochemical properties of the EDCs. In addition, we have also computed the one-dimensional (1D), 2D and 3D molecular descriptors using PaDEL [180, 181], RDKit [179] and Pybel [182]. For each EDC, PaDEL, RDKit and Pybel gave 1875, 213 and 14 descriptors, respectively. For each EDC, we have made its 2D and 3D chemical structure, physicochemical properties and molecular descriptors readily available via the DEDuCT 1.0 webserver, and this information can aid future efforts to develop computational toxicity models based on structure-activity relationships.

2.1.9 Predicted ADMET properties

Absorption, Distribution, Metabolism, Excretion and Toxicity (ADMET) properties can be utilized for the toxicity assessment of chemicals. Thus, several computational tools have been developed to predict the ADMET properties of chemicals such as admetSAR 2.0 [183], pkCSM [184], ProTox [185], SwissADME [186], Toxtree 2.6.1 [187] and vNN server [188]. We have employed these tools to predict the ADMET properties of the 686 potential EDCs in DEDuCT 1.0.

Absorption properties of a chemical reflect its ability to be absorbed from intestine to bloodstream. The predicted absorption properties for EDCs include Caco-2 permeability, human intestinal absorption (HIA), human oral bioavailability and skin permeability (log

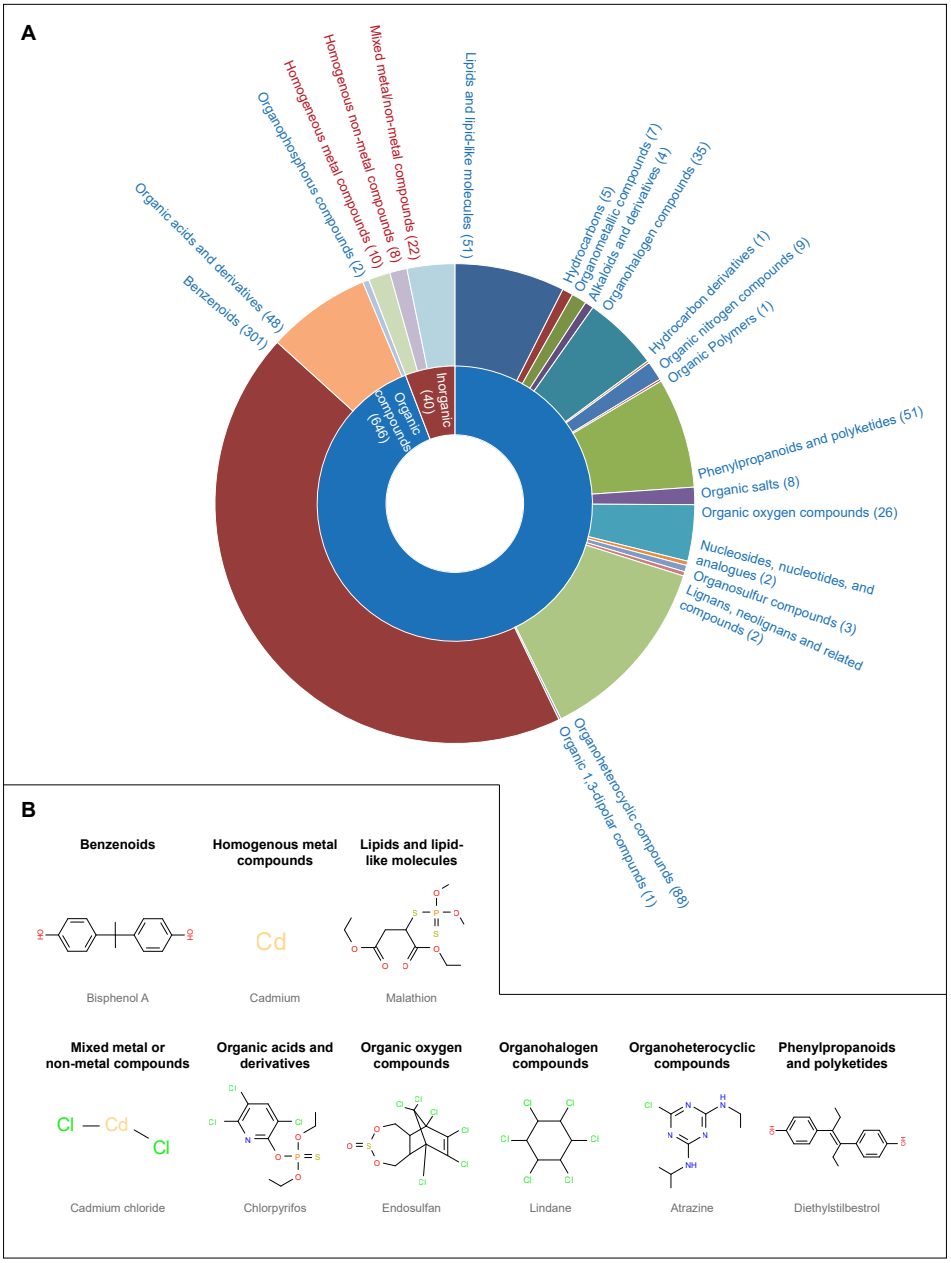


Figure 2.5: Classification of the 686 EDCs in DEDuCT 1.0 into chemical kingdoms and chemical super-classes using ClassyFire. (A) Of the 686 EDCs, 646 are organic and 40 are inorganic compounds. The 646 organic EDCs can be further classified into 19 super-classes while the 40 inorganic EDCs fall into 3 super-classes. The number of EDCs in each super-class is reported within the parenthesis. (B) The chemical structure of a representative EDC in each super-class with more than 10 EDCs is shown here. For instance, the super-class Benzenoids contains 301 EDCs including Bisphenol A shown here.

Kp). Distribution properties of a chemical shed light on its availability in other parts of the body after being absorbed into the bloodstream. The predicted distribution properties for EDCs include blood-brain barrier (BBB), CNS permeability, fraction unbound in human, P-glycoprotein inhibitor, P-glycoprotein substrate, plasma protein binding, steady state volume of distribution (VDss) and subcellular localization. Metabolism properties of a chemical describe its conversion into metabolites through enzymatic breakdown prior to elimination from the human body. The predicted metabolism properties for EDCs include assessment to act as a substrate or inhibitor of CYP450 enzymes, human bile salt export pump (BSEP), human liver microsomal (HLM) stability assay, human multidrug and toxin extrusion (MATE) transporter, organic anion-transporting polypeptides (OATP) and UDP-glucuronosyltransferases (UGT) catalysis. The predicted excretion properties for EDCs include total clearance rate and the ability to inhibit or act as a substrate for renal organic cation transporter 2 (OCT2). The predicted toxicological properties for EDCs include biodegradation capacity, carcinogenicity, Cramer's rule, cytotoxicity, hepatotoxicity, hERG inhibitors, maximum recommended tolerated dose (MRTD), mitochondrial membrane potential (MMP), rat oral toxicity and skin sensitization. Supplementary Table S2.9 lists the predicted ADMET properties by different tools used here.

2.2 Web interface of DEDuCT

We have created an online resource, Database of Endocrine Disrupting Chemicals and their Toxicity profiles (DEDuCT) version 1.0 [35], which contains detailed information on the 686 potential EDCs with supporting evidence compiled from 1796 published research articles. Importantly, DEDuCT 1.0 compiles the above-mentioned information on the 686 EDCs such as the endocrine-mediated endpoints, systems-level endocrine-mediated perturbations, dosage value specific to endpoints, type of supporting evidence based classification, environmental source-based classification, 2D and 3D chemical structures, chemical classification, physicochemical properties, molecular descriptors, predicted ADMET properties and target genes. DEDuCT 1.0 is accessible at:

https://cb.imsc.res.in/deduct/.

The web interface of DEDuCT 1.0 was created using PHP [189], HTML, CSS, Bootstrap 4, and jQuery [190]. To facilitate interactive visualization, we have used Google Charts [191], D3.js [192], Cytoscape.js [193] and JSmol [194] in the web interface. The compiled database on EDCs is stored using MariaDB [195], and the information from the database is retrieved using Structured Query Language (SQL). DEDuCT 1.0 website is hosted on Apache [196] webserver running on Debian 9.4 Linux Operating System.

Using the Browse section in the web interface of DEDuCT, users can view the EDCs based on their type of supporting evidence or environmental source or chemical classification or systems-level perturbations (Figure 2.6). Using the Simple search option in DEDuCT, users can search for individual EDCs using chemical name or standard identifier (Figure 2.6). Using the Physicochemical filter option in DEDuCT, users can also filter EDCs based on their physicochemical properties such as molecular weight, number of hydrogen bond donors or acceptors, and number of rotatable bonds (Figure 2.6). By clicking the chemical name of any EDC in DEDuCT, users can view the entire compiled information including supporting evidence and dosage information.

To better expose the utility of DEDuCT, let us consider the well-known EDC, Atrazine, as an example. Based on environmental source, DEDuCT 1.0 classifies Atrazine into the broad categories 'Agriculture and farming' and 'Pollutant', and sub-categories 'Environmental Pollutant', 'Fertilizer', 'Fungicide', 'Herbicide' and 'Pesticide'. Based on chemical classification, Atrazine is an 'Organic' compound belonging to super-class 'Organoheterocyclic compounds' and class 'Triazines'. In DEDuCT 1.0, Atrazine is a potential EDC with supporting experimental evidence from 40 research articles and falls into category II based on the type of supporting evidence. Based on compiled evidence in DEDuCT 1.0, Atrazine exposure can lead to any of the 7 systems-level perturbations and users can view the compiled dosage information corresponding to the observed endocrine-mediated endpoints in the web interface.

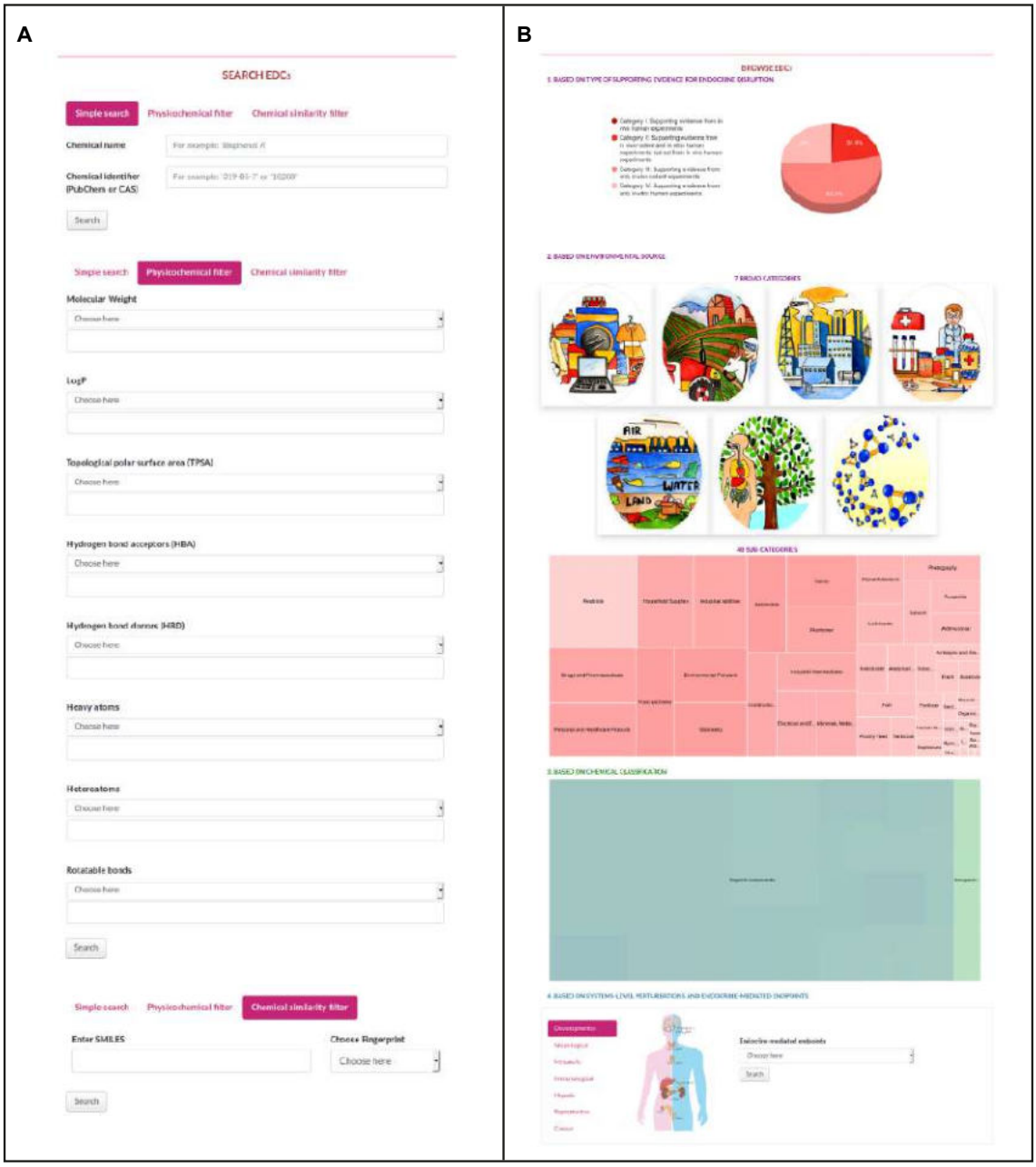


Figure 2.6: The web interface of DEDuCT. (A) The screenshot shows the different search options in our resource to obtain information on EDCs. Simple search option in DEDuCT can be used to search for individual EDCs using the chemical name or standard identifier. Physicochemical filter option in DEDuCT can be used to also filter EDCs based on their physicochemical properties such as molecular weight, number of hydrogen bond donors or acceptors, number of rotatable bonds. Chemical similarity filter gives the top 10 structurally similar EDCs in DEDuCT in comparison to the query molecule. (B) The Browse section in the web interface of DEDuCT can be used to view the EDCs based on the type of supporting evidence or their environmental source or chemical classification or systems-level perturbations and endocrine-mediated endpoints.

2.3 Comparison of DEDuCT 1.0 with existing resources on EDCs

In addition to extensive PubMed mining to identify published experiments on EDCs, DEDuCT integrates information from three existing resources, WHO report, TEDX and EDCs Databank (Figure 2.1). We find that 198 out of the 686 potential EDCs (28.9%) and 1294 out of the 1796 associated published research articles (72.0%) containing supporting experimental evidence in DEDuCT 1.0 are not captured in any of the three existing resources (Figure 2.3C; Table 2.1). Unlike DEDuCT, the supporting evidence for compiled EDCs in the three existing resources are not limited to *in vivo* or *in vitro* studies in humans and *in vivo* studies in rodents (Figure 2.1). Note that we were unable to find supporting evidence for endocrine disruption upon exposure in published experiments on humans or rodents for several chemicals listed as EDCs in the WHO report or TEDX or EDCs Databank, and thus, such chemicals are not contained in DEDuCT 1.0 (Figure 2.3C). Importantly, in contrast to the three existing resources, DEDuCT 1.0 compiles the observed endocrine-mediated endpoints and systems-level perturbations upon EDC exposure from published experiments (Table 2.1). Moreover, in contrast to the three existing resources, DEDuCT compiles the dosage information at which endocrine-mediated endpoints were observed upon EDC exposure from published experiments (Table 2.1).

2.4 Network view on the chemical space of EDCs

2.4.1 Chemical similarity network

Chemical similarity networks (CSNs) can shed insights on the extent of scaffold diversity in the associated chemical space [197–199]. We constructed the chemical similarity network (CSN) of the 686 EDCs in DEDuCT 1.0 as follows. In the CSN, nodes are EDCs and the edge weights reflect the extent of chemical similarity between pairs of EDCs.

Among the metrics for chemical similarity, Tanimoto [126, 200] and Dice [201] coefficients were determined to be the best choices [126]. In addition, while computing the Tanimoto or Dice coefficient, there are several choices of molecular fingerprints such as the extended connectivity fingerprints (ECFP4) [129], the MACCS keys fingerprints [130] and the Daylight-like (DLL) fingerprints, and ECFP4 has been shown to outperform other widely-used fingerprints [126, 202]. Thus, there are multiple choices based on similarity metrics and molecular fingerprints to specify the edge weights in the CSN, and in this work, we have explored six possible choices, namely, Tanimoto with ECFP4, Tanimoto with MACCS, Tanimoto with DLL, Dice with ECFP4, Dice with MACCS, and Dice with DLL which were computed using RDKit [179]. By exploring these six possible choices to construct CSN, we show that the broad conclusions from the analysis of CSN are robust to choices of similarity metrics and molecular fingerprints.

Since both Tanimoto coefficient and Dice coefficient for any pair of chemicals is in the range 0 to 1, the edge weights in the six CSNs are in the same range. To visualize the high similarity backbone of the CSN, we decided to omit edges with weights below a chosen threshold value signifying poor chemical similarity. Rather than choosing an arbitrary threshold value to construct this high CSN, we have investigated the size of the largest connected component (LCC) of the CSN as a function of the increasing threshold value for omitting edges (Figure 2.7). Note that the size of the LCC reflects the overall connectivity of the network. By identifying the threshold value at which there is a sharp decrease in the size of the LCC of the CSN, we have obtained the threshold value to construct the high CSN (Figure 2.7).

We find that this threshold value to construct the high CSN differs based on the six choices to assign edge weights, and it is found to be 0.45 for Tanimoto with ECFP4, 0.66 for Tanimoto with MACCS, 0.56 for Tanimoto with DLL, 0.62 for Dice with ECFP4, 0.80 for Dice with MACCS, and 0.72 for Dice with DLL (Figure 2.7A-F). Interestingly, we find that the size and composition of the LCC of the high CSNs depend on the choice of the molecular fingerprints rather than the similarity metric. That is, the size and composi-

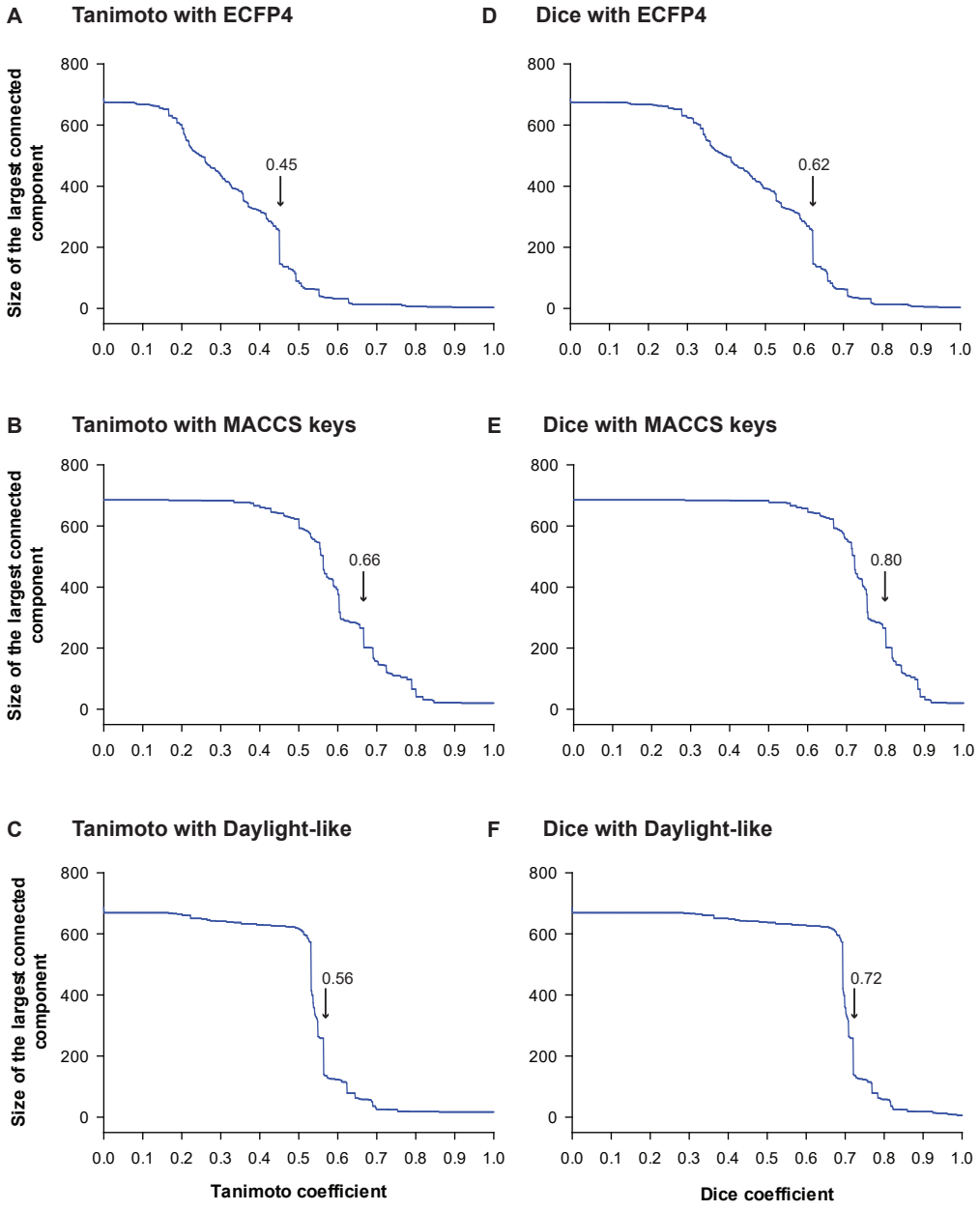


Figure 2.7: The size of the largest connected component (LCC) of the chemical similarity network (CSN) of EDCs as a function of the increasing threshold for omitting edges. (A) Tanimoto with ECFP4. (B) Tanimoto with MACCS. (C) Tanimoto with Daylight-like (DLL). (D) Dice with ECFP4. (E) Dice with MACCS. (F) Dice with Daylight-like (DLL).

tion of the LCC for the high CSNs constructed using Tanimoto with ECFP4 or Dice with ECFP4 are same with 255 EDCs, Tanimoto with MACCS or Dice with MACCS are same with 266 EDCs, and Tanimoto with DLL or Dice with DLL are same with 258 EDCs. Furthermore, we find more than 75% overlap between EDCs contained in LCCs corresponding to any pair of the six high CSNs [35]. Thus, we have chosen to show only the high CSNs constructed using Tanimoto with ECFP4, Tanimoto with MACCS and Tanimoto with DLL (Figure 2.8; Figure 2.9). Moreover, we have chosen to report the detailed analysis of the high CSN constructed using Tanimoto with ECFP4 (Figure 2.8; Supplementary Table S2.10) as the combination of Tanimoto coefficient and ECFP4 fingerprints was earlier found to be the best choice for chemical similarity computations [126, 202].

Since EDCs are believed to cause endocrine disruption by mimicking the hormones in human body [8, 50, 203], it is worthwhile to investigate the chemical properties shared by EDCs. Based on the chemical classification of the 686 EDCs in DEDuCT 1.0, we find that EDCs can be either organic or inorganic compounds, and moreover, are spread across diverse chemical classes (Figure 2.8). Still, 301 of the 686 EDCs (43.9%) in DEDuCT 1.0 belong to a single chemical super-class Benzenoids (Figure 2.8). We further investigate this chemical space by analyzing the CSN for the 686 EDCs in DEDuCT 1.0.

In Figure 2.8A, it is seen that the high CSN has a LCC of 255 EDCs, 8 small components with 5 to 14 EDCs, 44 small components with 2 to 4 EDCs and many isolated EDCs. In order to reveal the finer clustering of EDCs within the LCC, we have employed Louvain modularity [204] as implemented in the network visualization tool Gephi [205] to identify 14 modules within the LCC of the high CSN (Figure 2.8A). Moreover, a closer inspection revealed that 210 out of the 255 EDCs in the LCC belong to the chemical super-class Benzenoids. This observation inspired us to investigate the number of benzene rings contained in each EDC (Figure 2.8A) [35].

Interestingly, we find that 254 out of the 255 EDCs in the LCC contain at least 1 benzene ring. Furthermore, 42 out of the 43 EDCs in the largest module of the LCC have 2 benzene rings (Module 1 in Figure 2.8A). Similarly, 29 out of the 31 EDCs in the

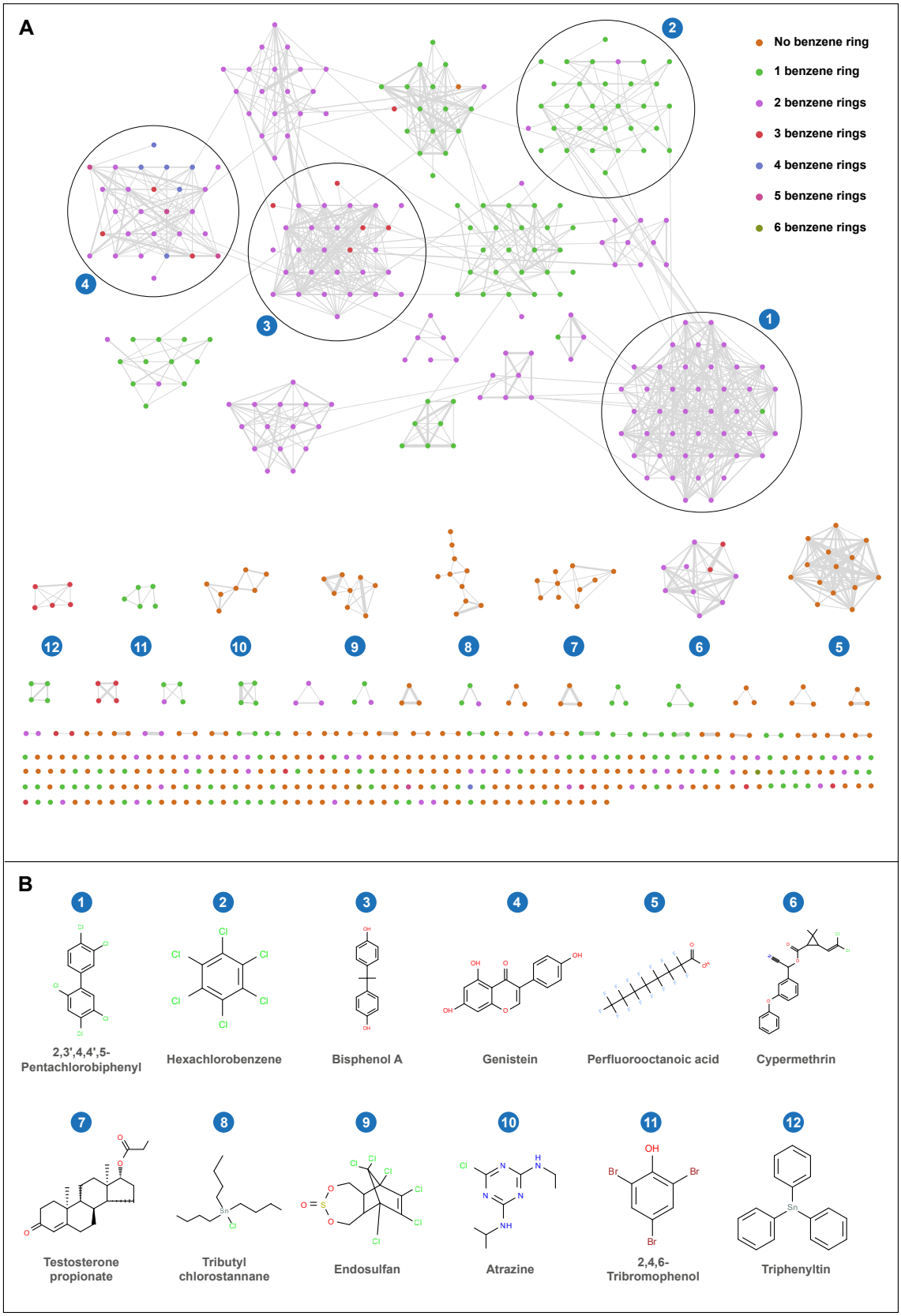

A
No benzene ring
1 benzene ring
2 benzene rings
3 benzene rings
4 benzene rings
5 benzene rings
6 benzene rings
1
2
3
4
5
6
7
8
9
10
11
12
B
1
2,3',4,4',5-Pentachlorobiphenyl
2
Hexachlorobenzene
3
Bisphenol A
4
Genistein
5
Perfluorooctanoic acid
6
Cypermethrin
7
Testosterone propionate
8
Tributyl chlorostannane
9
Endosulfan
10
Atrazine
11
2,4,6-Tribromophenol
12
Triphenyltin

Figure 2.8 ***(previous page)*:** Network visualization of the high chemical similarity network (CSN) of 686 EDCs in DEDuCT 1.0. (A) High CSN of 686 EDCs where nodes represent EDCs and edges represent chemical similarity between pairs of EDCs quantified using Tanimoto coefficient with ECFP4 fingerprints. Here, the edge thickness reflects the extent of chemical similarity between two EDCs, and the node colour is based on the number of benzene rings in its chemical structure. Moreover, Louvain modularity within the network visualization tool Gephi was employed to identify 14 modules within the LCC. The four largest modules in LCC and 8 smaller connected components with 5 to 14 EDCs have been prominently labelled in this figure. (B) The chemical structure of a representative EDC in each of the labelled modules or connected components in (A) is shown here.

second largest module of the LCC have 1 benzene ring (Module 2 in Figure 2.8A) and 24 out of the 29 EDCs in the third largest module of the LCC have 2 benzene rings (Module 3 in Figure 2.8A). These observations suggest a striking pattern within larger modules of the LCC in terms of the number of constituent benzene rings of EDCs. In contrast to Modules 1, 2 and 3 of the LCC, the fourth largest module contains 28 EDCs of which 16, 3, 6 and 4 EDCs have 2, 3, 4 and 5 benzene rings, respectively (Figure 2.8A). In Figure 2.8B, we also show the chemical structure of a representative EDC contained in the 4 largest modules of the LCC and 8 smaller components or clusters with 5 to 14 EDCs. For example, Bisphenol A is a well-known EDC contained in Module 3 of the LCC (Figure 2.8B).

Furthermore, a visual inspection of the 8 smaller components with 5 to 14 EDCs finds that 5 of these components (Cluster 5, 7, 8, 9 and 10 in Figure 2.8A) consist solely of EDCs with no benzene rings. For instance, Cluster 5 has 14 EDCs which are fluorinated linear chain hydrocarbon compounds (e.g., Perfluorooctanoic acid), Cluster 7 has 10 EDCs which are structurally similar to steroids and their derivatives (e.g., the drug testosterone propionate), and Cluster 8 has 10 EDCs which have linear hydrocarbon chains with or without metals (e.g., Tributylchlorostannane) with no benzene rings (Figure 2.8). In contrast, Cluster 11 has 5 EDCs including 2,4,6-Tribromophenol whose structures have 1 brominated benzene ring (Figure 2.8). Note that Module 2 in LCC and Cluster 11 primarily consist of EDCs with 1 benzene ring, however, a likely explanation for their separation into different connected components is the presence of brominated benzene ring in EDCs

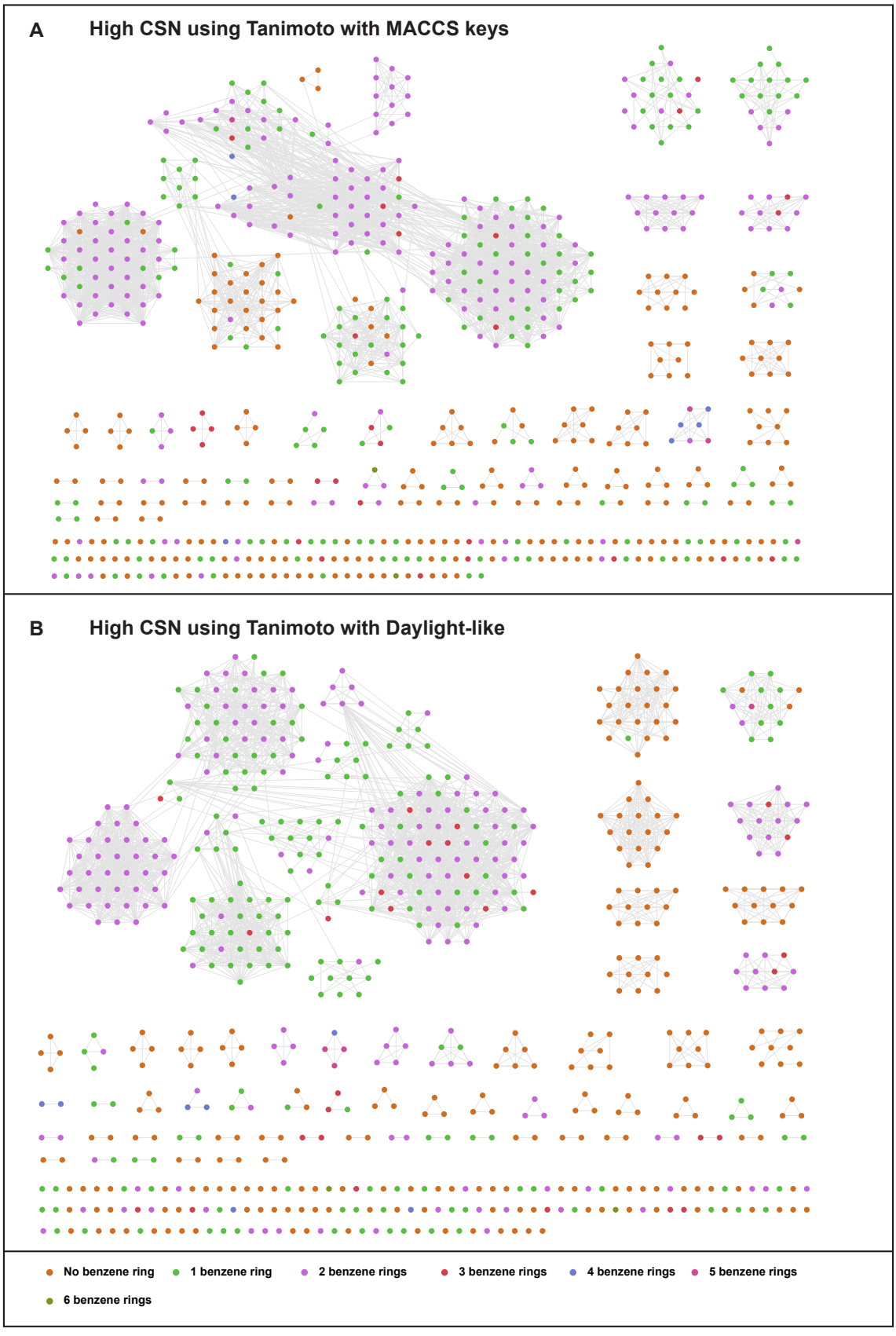

A
High CSN using Tanimoto with MACCS keys
B
High CSN using Tanimoto with Daylight-like
No benzene ring
1 benzene ring
2 benzene rings
3 benzene rings
4 benzene rings
5 benzene rings
6 benzene rings

Figure 2.9 ***(previous page)*:** Network visualization of the high chemical similarity network (CSN) of 686 EDCs in DEDuCT 1.0. (A) High CSN where chemical similarity is quantified by Tanimoto coefficient with MACCS keys fingerprints. (B) High CSN where chemical similarity is quantified by Tanimoto coefficient with Daylight-like (DLL) fingerprints. In this figure, the edge thickness reflects the extent of chemical similarity between two EDCs, and the node colour is based on the number of benzene rings in its chemical structure. Moreover, Louvain modularity within the network visualization tool Gephi was employed to identify modules within the LCC.

of Cluster 11 in contrast to the presence of chlorinated benzene ring in EDCs of Module 2 (Figure 2.8). In summary, this analysis of the high CSN reveals on the one hand the diversity of the chemical space of EDCs and on the other hand leads to modules or clusters of EDCs which can be explained by distinct chemical features [35].

2.4.2 Target genes of EDCs based on ToxCast assays

To better understand the molecular events leading to adverse effects or endocrine-specific perturbations upon EDC exposure, it is important to characterize the target genes of EDCs. EDCs sharing target genes are likely to have adverse effects or functional perturbations in common. Hence, we gathered information on the target genes of EDCs that can elucidate molecular initiating events leading to adverse effects upon chemical exposure. ToxCast [89] uses high-throughput assays designed to screen toxic chemicals based on perturbation of biological activities upon exposure. To date, ToxCast has screened more than 9000 chemicals using more than 900 high-throughput assays. We used the ToxCast invitroDB3 dataset released in October 2018 [206] to obtain the list of perturbed genes upon EDC exposure.

The assay summary information file (Assay_Summary_180918.csv) contains the detailed annotation of the ToxCast assays including assay type, assay component, assay component endpoint, assay target information, cell lines used for the assay, and assay citation. Using the assay component endpoint of a ToxCast assay, one can obtain the observed biological effect such as changes in gene expression upon chemical exposure. In practice, the assay component endpoint of a ToxCast assay may correspond to one or

more target genes. The assay activity information file (hitc_Matrix_180918.csv) provides a list of active or inactive chemicals based on the potency of the chemical to produce a significant biological effect captured via 1504 assay component endpoints of different ToxCast assays. In this work, we restrict to ToxCast assays and their corresponding assay component endpoints that are specific to humans. If a tested chemical is active for a particular assay component endpoint of a ToxCast assay, then the corresponding gene is assigned to be the target of the chemical.

Of the 686 potential EDCs in DEDuCT 1.0, we found target genes for 383 EDCs based on 1228 ToxCast assay component endpoints specific to humans. Supplementary Table S2.11 gives the target genes of these 383 EDCs based on ToxCast assay component endpoints specific to human [35]. We remark that it is possible to expand this information on target genes of EDCs using toxicological databases such as CTD [30], however, CTD compiles target information from both experiments and computational predictions.

2.4.3 Target similarity network

To reveal the target similarity between EDCs, we next investigated the target similarity network (TSN) of EDCs. For the 383 EDCs with information on target genes from ToxCast assays, we have constructed a target similarity network (TSN) based on shared target genes between pairs of EDCs. In the TSN, nodes are EDCs and edge weights signify the target similarity between pairs of EDCs. To quantify the similarity between two sets of target genes corresponding to a pair of EDCs, we use the standard measure, Jaccard index [207], given by the ratio of the number of elements in the intersection over the number of elements in the union of the two sets of target genes. By construction, Jaccard index is in the range 0 to 1. Jaccard index between two EDCs is 0 if they have no target genes in common, and it is 1 if they have all target genes in common.

To visualize the high similarity backbone of the TSN, we decided to omit edges with weights below a chosen Jaccard index value signifying poor target similarity between

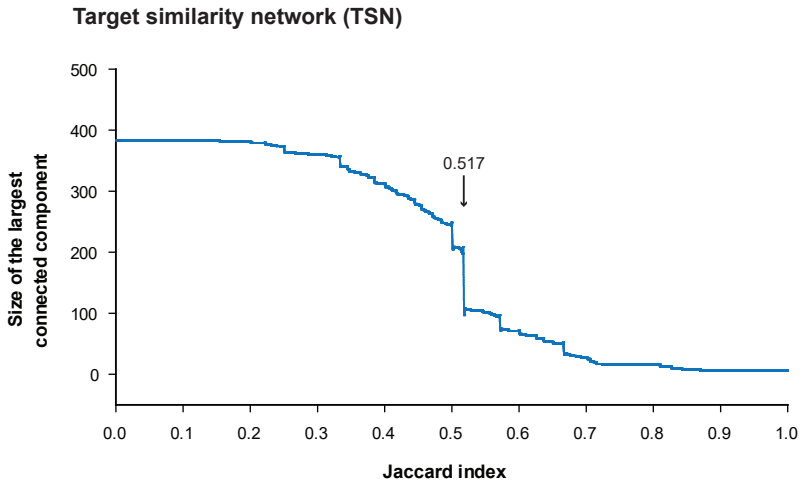


Figure 2.10: The size of the largest connected component (LCC) of the target similarity network (TSN) of EDCs as a function of the increasing Jaccard index for omitting edges.

pairs of EDCs. Rather than choosing an arbitrary Jaccard index value to construct this high TSN, we have investigated the size of the LCC of the TSN as a function of the increasing Jaccard index value for omitting edges (Figure 2.10). Based on this investigation, we find that there is a sharp decrease in the size of the LCC of the TSN obtained after omitting edges below the Jaccard index of 0.517 (Figure 2.10). Subsequently, we used this threshold Jaccard index of 0.517 to construct the high TSN of the 383 EDCs (Figure 2.11; Supplementary Table S2.12).

In Figure 2.11, it is seen that the high TSN has a LCC of 199 EDCs, 13 smaller components of 2 to 6 EDCs and 145 isolated EDCs. We have also employed Louvain modularity [204] to partition the LCC of the high TSN into 6 modules (Figure 2.11). The sizes of nodes in the high TSN reflect the weighted degree of EDCs, and the top 2 hubs are well-known EDCs, o,p'-DDT (CID:13089) and 4-Octylphenol (CID:15730), that belong to the largest module within the LCC of the high TSN (Figure 2.11). Based on the TSN constructed using limited information on target genes from ToxCast assays, we conclude that EDCs can have very different set of target genes [35].

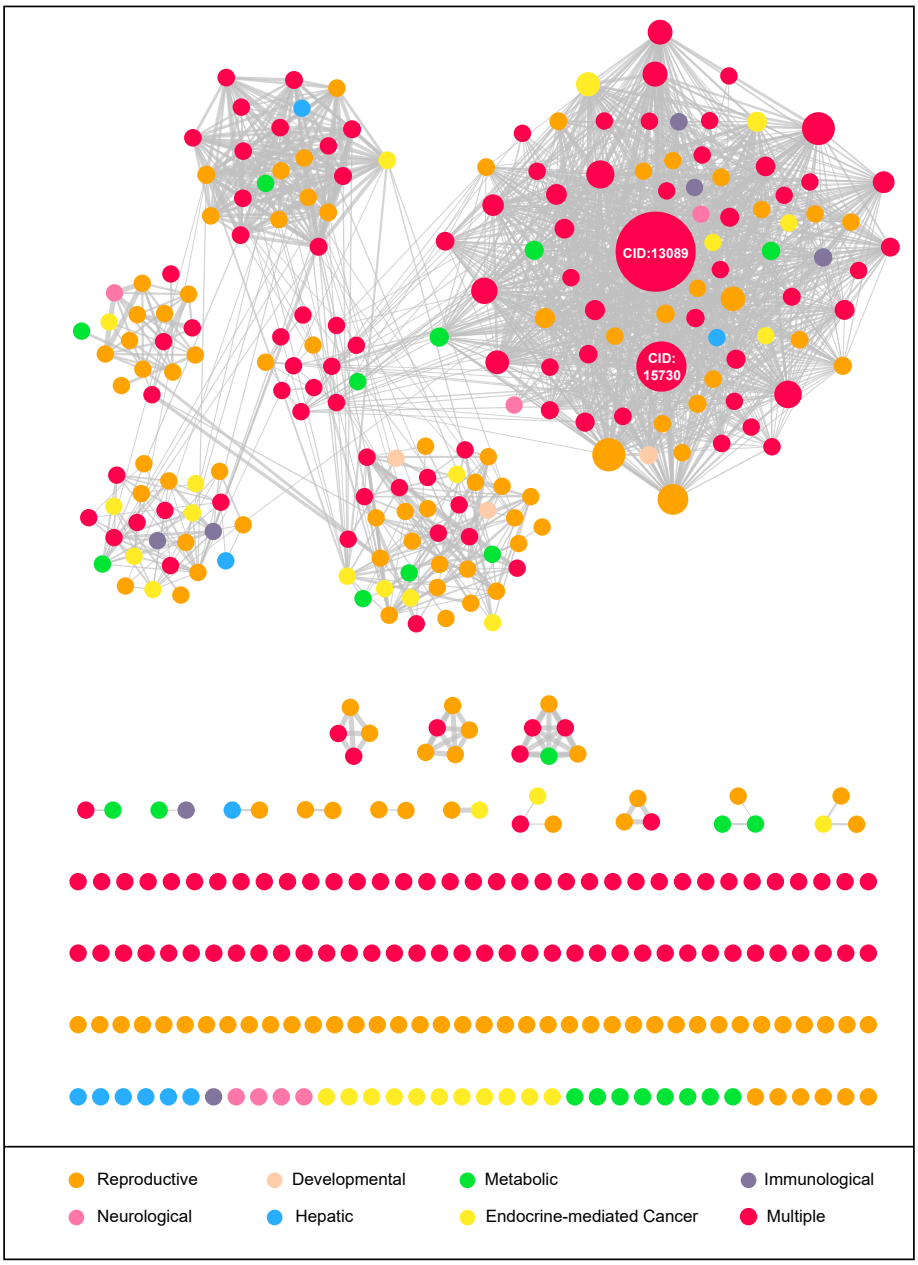


Figure 2.11: Network visualization of high target similarity network (TSN) of 383 EDCs. The high TSN was constructed for 383 EDCs which have information on their target genes from ToxCast assays. The legend at the bottom of this figure gives the colour code for nodes or EDCs in TSN which is based on the 7 systems-level perturbations, namely, Reproductive (RT), Developmental (DT), Metabolic (MT), Immunological (IT), Neurological (NT), Hepatic (HT) and Endocrine-mediated cancer (CT), associated with EDCs in DEDuCT 1.0. Note that if an EDC is associated with more than one systems-level perturbations then its colour is given by Multiple. Moreover, the sizes of the nodes in the high TSN reflect their weighted degree in the network and the thicknesses of the edges in the high TSN reflect their weights given by Jaccard index. In addition, we have labelled the top 2 hubs, namely, o,p'-DDT (CID:13089) and 4-Octylphenol (CID:15730), based on the weighted degree of nodes in this network.

2.5 Lack of correlation between chemical structure and target genes of EDCs

We next investigated whether there is any relationship between structural similarity and target similarity of EDCs. Recall that the structural similarity between two EDCs is quantified using six possible choices of two similarity metrics (Tanimoto or Dice coefficient) and three molecular fingerprints (ECFP4, MACCS or DLL) while the target similarity or commonality between the sets of target genes for two EDCs is quantified using the Jaccard index. In Figure 2.3D, we plot this structural similarity computed using Tanimoto with ECFP4 versus the target similarity for pairs of EDCs within the subset of 383 EDCs with information on target genes from ToxCast assays, and we find no significant correlation between structural similarity and target similarity of EDCs. Figure 2.12A-F also displays this plot for the six choices to compute chemical similarity between EDCs, and it can be seen that our observation of no significant correlation between structural similarity and target similarity is independent of the choice of chemical similarity metric used for computations.

These observations underscore the challenge in developing computational models to predict adverse effects of EDCs. Since traditional computational toxicity models based on quantitative structure activity relationship (QSAR) use chemical similarity and bioactivity information for their predictions, our results based on high CSN and high TSN suggest that such models to predict adverse effects of EDCs are unlikely to have high predictive power. Alternatively, computational systems toxicity models leveraging information in DEDuCT 1.0 on chemical structure, dosage information, set of target genes and systems-level perturbations of EDCs may have better predictive power [35].

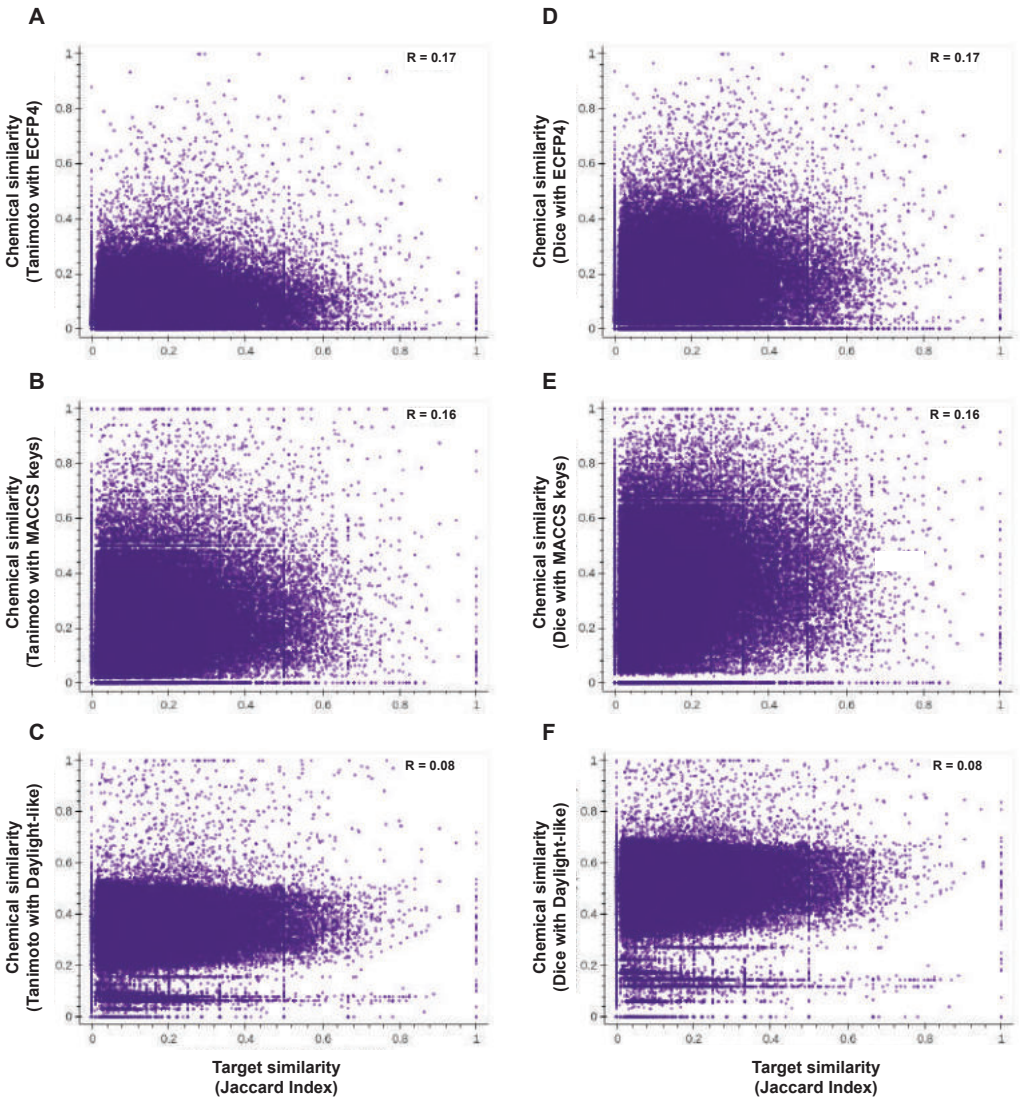


Figure 2.12: Scatter plots of target similarity versus chemical structure similarity between pairs of EDCs. In this figure, we explore six combinations of two similarity metrics and three molecular fingerprints to compute the chemical similarity between pairs of EDCs. (A) Tanimoto coefficient with ECFP4 fingerprints. (B) Tanimoto coefficient with MACCS keys fingerprints. (C) Tanimoto coefficient with Daylight-like (DLL) fingerprints. (D) Dice coefficient with ECFP4 fingerprints. (E) Dice coefficient with MACCS keys fingerprints. (F) Dice coefficient with Daylight-like (DLL) fingerprints. In each figure, we report the Pearson correlation coefficient R between structural and target similarity of EDCs. Regardless of the choice of metric to compute the chemical similarity, we find no significant correlation between the structural and target similarity of EDCs.

2.6 Evaluation of the sensitivity of toxicity predictors using compiled experimental evidence in DEDuCT 1.0

Several computational toxicity predictors such as admetSAR 2.0 [183], pkCSM [184], ProTox [185], SwissADME [186], Toxtree 2.6.1 [187] and vNN server [188] have been developed for risk assessment of chemicals. We have used these tools to predict the ADMET properties of the 686 EDCs, and this information is readily available from DEDuCT 1.0 webserver (Supplementary Table S2.9). Since DEDuCT 1.0 compiles experimentally observed toxicity profiles or endocrine-mediated endpoints for the 686 EDCs from supporting literature, we decided to utilise this compiled experimental evidence as a positive dataset to evaluate the sensitivity of computational toxicity prediction tools.

In DEDuCT 1.0, 157 EDCs have experimental evidence to cause hepatic endocrine-mediated perturbations. Among the toxicity predictors, admetSAR 2.0, pkCSM and vNN server can predict the hepatotoxicity of chemicals. Of these 157 EDCs, admetSAR 2.0, pkCSM and vNN server gave correct prediction for 60, 23 and 41 EDCs, respectively. Thus, the sensitivity for predicting hepatotoxicity of EDCs by admetSAR 2.0, pkCSM and vNN server are 0.382, 0.146 and 0.261, respectively, based on our dataset.

In DEDuCT 1.0, 185 EDCs have experimental evidence to cause endocrine-mediated cancer. Among the toxicity predictors, admetSAR 2.0 and Toxtree 2.6.1 can predict the carcinogenicity of chemicals. Of these 185 EDCs, admetSAR 2.0 predicted 56 while Toxtree 2.6.1 predicted none to be carcinogens. Thus, the sensitivity for predicting carcinogenicity of EDCs by admetSAR 2.0 and Toxtree 2.6.1 is 0.302 and 0.0, respectively, based on our dataset.

admetSAR 2.0 predicted 127 out of the 185 EDCs with experimental evidence to cause cancer in DEDuCT 1.0 to be non-carcinogens, and we have compared these 127 EDCs with the potential carcinogens released by the International Agency for Research on Cancer (IARC) Monographs [208, 209] and the Report on Carcinogens (RoC) by the

National Toxicology Program [210]. Based on this comparison, we found 9 of the 127 EDCs predicted as non-carcinogens by admetSAR 2.0 were listed as potential carcinogens in IARC Monographs and RoC. Notably, 3 of the 127 EDCs, namely, benzo[a]pyrene, diethylstilbesterol and pentachlorophenol are categorized as group 1 potential carcinogens for human by IARC Monographs.

Overall, this evaluation of the computational toxicity tools for prediction of hepatotoxicity and carcinogenicity of EDCs based on the compiled experimental evidence in DEDuCT 1.0 suggests lack of significant predictive power. A possible interim solution towards increasing the predictive power of the existing tools will be to update their positive training dataset with experimental information on EDCs from DEDuCT [35].

2.7 Discussion

EDCs are a group of chemicals of emerging concern which are omnipresent in our environment. Since endocrine disruption mechanism is a special form of toxicity, the risk assessment and identification of EDCs remains challenging [8]. In this chapter, we have developed a detailed workflow which was employed to identify 686 potential EDCs from 1796 research articles with supporting evidence for endocrine disruption from published experiments in humans or rodents. Further, we have compiled, unified and standardized the observed adverse effects upon EDC exposure in published experiments into 514 unique endocrine-mediated endpoints which were further classified into 7 systems-level perturbations. DEDuCT 1.0 compiles additional information including the dosage information, environmental source classification, classification based on supporting evidence, chemical structure, physicochemical properties, predicted ADMET properties, and target genes for the 686 potential EDCs, and this information is accessible at: https://cb.imsc.res.in/deduct/ (Figure 2.13).

Furthermore, we have employed a network-centric approach to understand the link between the chemical space of EDCs and their biological target space. Here, we have

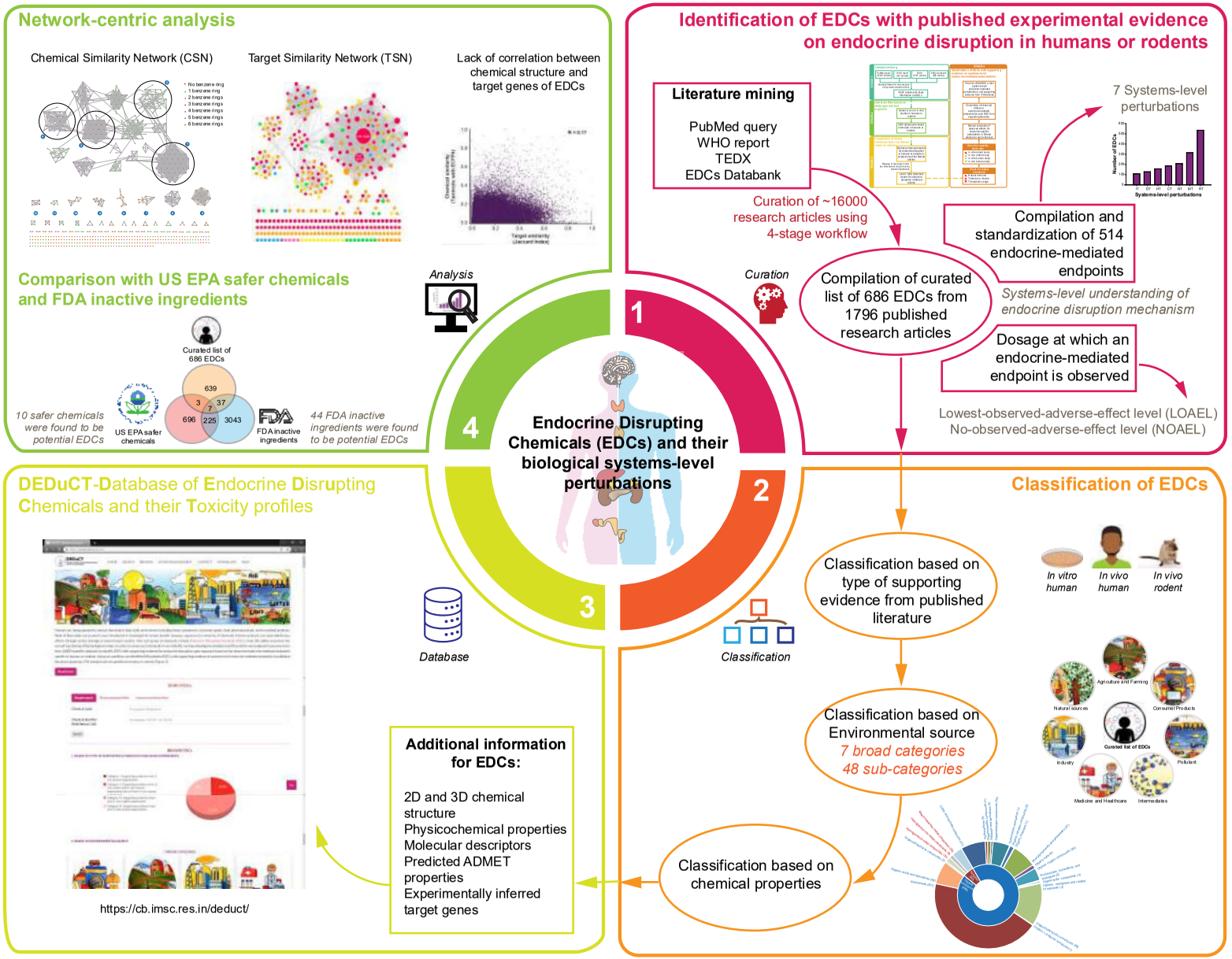


Figure 2.13: Schematic diagram summarizing DEDuCT 1.0 on endocrine disruptors.

constructed and analysed two different networks of EDCs, namely, the chemical similarity network (CSN) and the target similarity network (TSN). Based on CSN, we infer that EDCs are diverse in their chemical structure and can be grouped into modules with distinct chemical features. Based on TSN, we infer that EDCs can have very different set of target genes. Subsequent investigation of the relationship between the chemical structure and biological (gene) targets of EDCs found no correlation. These observations on the lack of correlation between chemical structure and target genes of EDCs raises potential challenges in developing structure-based computational models to predict adverse effects of EDCs (Figure 2.13). Lastly, the compiled experimental evidence for EDCs in DEDuCT 1.0 was used to evaluate the predictive power of existing computational toxicity tools. Such an evaluation using our compiled dataset suggests that the existing tools for predicting hepatotoxicity and carcinogenicity of chemicals lack significant predictive power. In the near future, toxicity predictors can integrate experimental evidence from DEDuCT to improve their predictive power.

An important aspect of EDCs is their ability to exert adverse effects even at low dosage values [168–170]. Our compilation of dosage information at which endocrine-mediated endpoints were observed in published experiments upon individual EDC exposure will further help researchers to understand the low dose exposure effects of EDCs. Also, our large-scale compilation of the observed effects or endpoints along with the systems-level perturbations upon EDC exposure can be visualized as a tripartite network with nodes as EDCs, endocrine-mediated endpoints and systems-level perturbations. Future exploration of this tripartite network will enhance systems-level understanding of perturbed biological pathways upon EDC exposure.

After publication [35], DEDuCT has received coverage in national and international media including India Science Wire, Chemistry and Engineering News (c&en) [211] of the American Chemical Society, Hindustan Times, Chemical Watch, and European Trade Union Institute. Importantly, DEDuCT has been well received by scientific peers. To highlight, the French Agency for Food, Environmental and Occupational Health & Safety

(ANSES) has come up with a list of substances to be further included in their assessment program as part of the Second French National Endocrine Disruptor Strategy (SNPE 2). To draw their list of priority substances, ANSES has utilized DEDuCT 1.0 as one of their primary resources after assessing 27 existing initiatives on EDCs worldwide. According to this ANSES report [212], the robust approach followed in DEDuCT 1.0 to identify EDCs meets the SNPE 2 criteria for the inclusion of priority substances. In sum, DEDuCT is an important resource on EDCs that will enable delivery of safer consumer products.

Supplementary Information

Supplementary Tables S2.1-S2.12 associated with this chapter are available for download from the GitHub repository: https://github.com/asamallab/PhDThesis-Janani_R/blob/main/SI/ST_Chapter2.xlsx.

Feature	DEDuCT 1.0	EDCs Databank	TEDX	WHO report
Number of EDCs	686	615	1428	184
Web interface	Yes	Yes	Yes	No
Compilation of endocrine-mediated endpoints for EDCs from published experiments on endocrine disruption in humans or rodents	Yes	No	No	No
Dosage information specific to endocrine-mediated endpoints for EDCs from published experiments on endocrine disruption in humans or rodents	Yes	No	No	No
Systems-level perturbations for EDCs based on observed endocrine-mediated endpoints in published experiments on endocrine disruption in humans or rodents	Yes	No	No	No
Categorization of EDCs based on the type of supporting evidence	Yes	No	No	No
Categorization of EDCs based on environmental source	Yes	No	No	No
Categorization of EDCs based on their use	Yes	Yes	Yes	No
Chemical classification of EDCs	Yes	No	No	No
Availability of 2D structure for EDCs	Yes	Yes	No	No
Availability of 3D structure for EDCs	Yes	Yes	No	No
Downloadable formats for 2D and 3D structure of EDCs	SDF, MOL2, PDB, PDBQT	SDF, MOL2, PDB, PDBQT	No	No
Chemical identifiers of EDCs	PubChem or CAS	PubChem or CAS	CAS	No
Physicochemical properties of EDCs	Yes	Yes	No	No
Molecular descriptors for EDCs	Yes	No	No	No
Predicted ADMET properties of EDCs	Yes	No	No	No
Chemical-gene association based on experimental assays	Yes	No	No	No
Chemical similarity filter	Yes	Yes	No	No

Table 2.1: Comparison of the information on EDCs in DEDuCT with three existing resources, namely, EDCs Databank, TEDX and WHO report.

Chapter 3

DEDuCT 2.0: An updated knowledgebase and an exploration of the current regulations and guidelines from the perspective of endocrine disrupting chemicals

Due to the hazardous potential of EDCs, their adverse health effects on humans and wildlife have been studied for more than three decades, and this information is documented in scientific literature, including published research articles, toxicological reports, and regulatory guidelines [8, 213]. Despite the increasing research interest, several limitations and uncertainties challenge the risk assessment and regulation of EDCs [3, 213]. Importantly, a standard (consensus) definition for EDCs can dictate the evidence needed for its identification among environmental chemicals [3, 43, 45].

In this direction, several definitions have been proposed and adopted by various regulatory agencies. However, clarity and standardization are yet to be achieved in EDCs

research [3]. This is also reflected in a recent comprehensive study commissioned by the European Parliament on endocrine disruptors and the current EU regulations on the subject [156]. In particular, the report found gaps in the definition of EDCs, test requirements and guidelines for authorization of products in a number of categories such as cosmetics, drinking water and workers' regulations [156]. Another challenge to the regulation of EDCs is the wide range of factors to be considered in developing risk assessment criteria. In addition to defining the adverse effects, factors such as source and dosage of exposure need to be considered, all of which are aspects studied and documented in peer-reviewed articles in scientific journals. However, it is unknown to what extent this scientific literature is consulted during the development of risk assessment criteria and testing standards for EDCs. In fact, toxicity test guidelines have received criticism for having omitted several relevant endpoints which are captured in academic research [214].

The above-mentioned two observations, namely, the growth in the volume of scientific knowledge surrounding EDCs, and the perceived presence of gaps in the risk assessment and regulation of EDCs, have prompted the comparative analysis reported in this chapter. In this chapter, we explore how academic research leading to curated knowledgebases can inform current chemical regulations on EDCs. To this end, we present in this chapter an updated knowledgebase DEDuCT 2.0, and thereafter, studied the distribution of potential EDCs across several chemical lists that reflect guidelines for use or regulations [36]. **The work reported in this chapter is contained in the published manuscript [36].**

3.1 DEDuCT 2.0 and growing research effort on EDCs

As described in chapter 2, we have built a unique knowledgebase, DEDuCT version 1.0, containing information on 686 potential EDCs with supporting evidence from 1796 research articles [35]. In this chapter, we will use this knowledgebase to highlight the growing research effort in the academia on EDCs over the past decades.

To create DEDuCT 1.0 [35], we had mined and curated more than 16000 research articles published until February 2018 to finally obtain a corpus of 1796 articles containing supporting experimental evidence specific to humans or rodents for 686 potential EDCs. An analysis of this corpus of 1796 articles published until February 2018 found that the number of articles with supporting evidence on potential EDCs has significantly increased over the last three decades (Figure 3.1A) [36]. The continuous growth of literature on EDCs (Figure 3.1A) and community interest in DEDuCT 1.0 [211] served as motivation to perform a substantial update of our knowledgebase to include published scientific literature until January 2020.

Here, we have built an updated knowledgebase, DEDuCT version 2.0, with information on 792 potential EDCs with supporting experimental evidence from 2218 published research articles (Supplementary Tables S3.1-S3.2). In order to achieve the updated database DEDuCT 2.0, we had to mine and curate additional 3396 research articles on EDCs which were published until January 2020. Essentially, we followed the four staged workflow used to create DEDuCT 1.0 [35] as described in chapter 2, to create the updated database DEDuCT 2.0 (Figure 3.2). The compiled information on 792 potential EDCs and additional information including supporting literature, systems-level perturbations, observed endocrine-mediated endpoints and corresponding dosage information is accessible via DEDuCT 2.0 webserver at: `https://cb.imsc.res.in/deduct` [35,36].

A chronological analysis of the corpus of 2218 published articles which form the supporting evidence for 792 potential EDCs in DEDuCT 2.0 finds that there are 1181 articles published in the period 2011-2020, followed by 696 articles in the period 2001-2010, followed by 192 articles in the period 1991-2000 (Figure 3.1A). We remark that the corpus of 2218 research articles in DEDuCT 2.0 is likely to be a lower estimate of the accumulated scientific knowledge to date on EDCs; nevertheless, it is evident from Figure 3.1A that there has been significant growth in research on EDCs in the past three decades.

In addition, we leverage the 792 potential EDCs along with the associated supporting

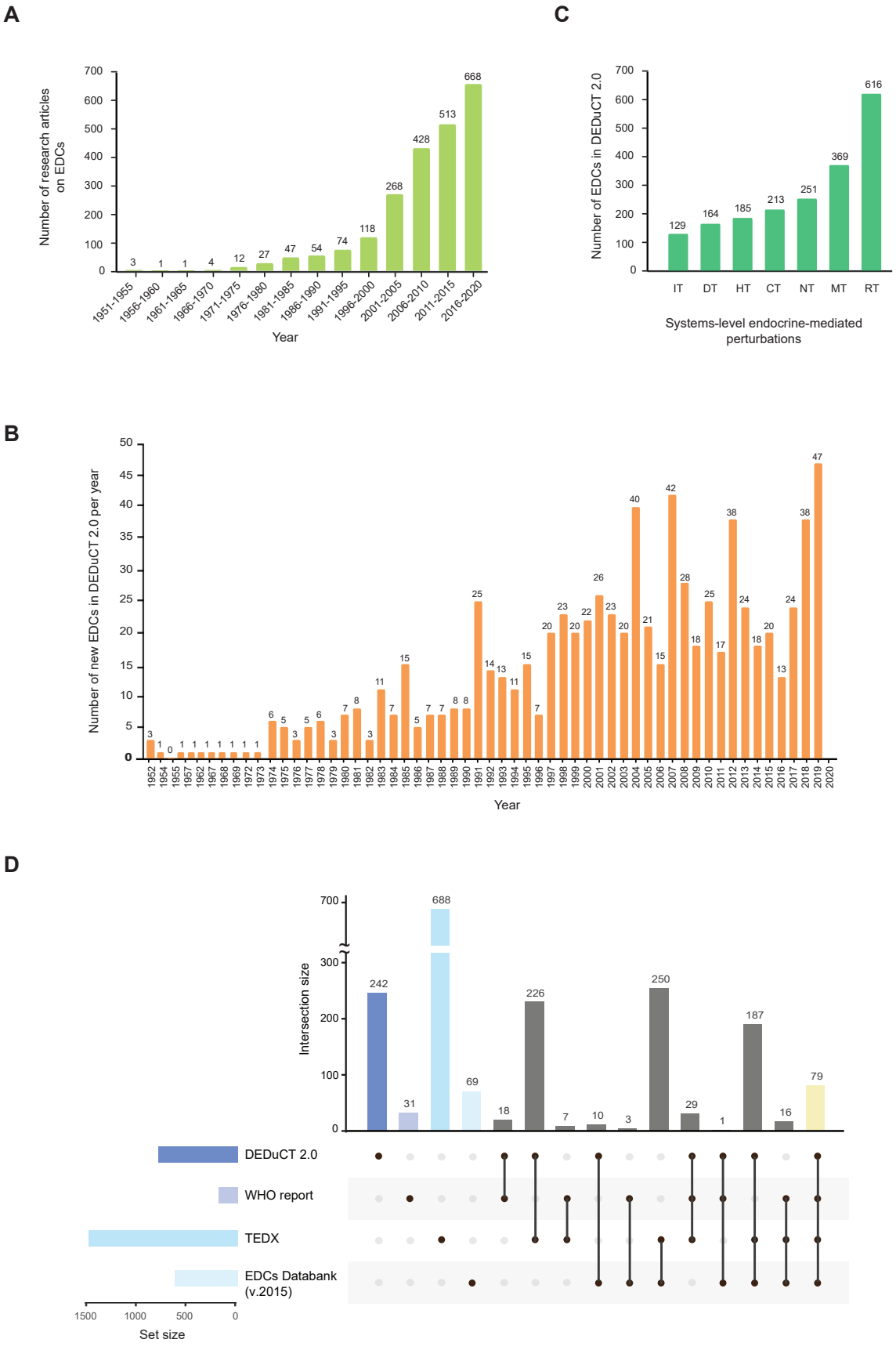

A
Number of research articles on EDCs
Year
1951-1955
1956-1960
1961-1965
1966-1970
1971-1975
1976-1980
1981-1985
1986-1990
1991-1995
1996-2000
2001-2005
2006-2010
2011-2015
2016-2020
3
1
1
4
12
27
47
54
74
118
268
428
513
668
C
Number of EDCs in DEDuCT 2.0
IT
DT
HT
CT
NT
MT
RT
129
164
185
213
251
369
616
Systems-level endocrine-mediated perturbations
B
Number of new EDCs in DEDuCT 2.0 per year
Year
D
Intersection size
242
31
688
69
18
226
7
10
3
250
29
1
187
16
79
DEDuCT 2.0
WHO report
TEDX
EDCs Databank (v.2015)
Set size

Figure 3.1 ***(previous page)*****:** (A) A chronological analysis of the corpus of 2218 published articles which form the supporting evidence for 792 potential EDCs in DEDuCT 2.0. (B) A plot of the number of new EDCs identified in published literature per year based on information compiled in DEDuCT 2.0. (C) Evidence for seven different systems-level perturbations from published experiments across 792 potential EDCs compiled in DEDuCT 2.0. (D) Comparison of the list of EDCs captured in DEDuCT 2.0 with three other resources. From the UpSetR plot, it is seen that 242 out of 792 potential EDCs in DEDuCT 2.0 are not captured in any other resource.

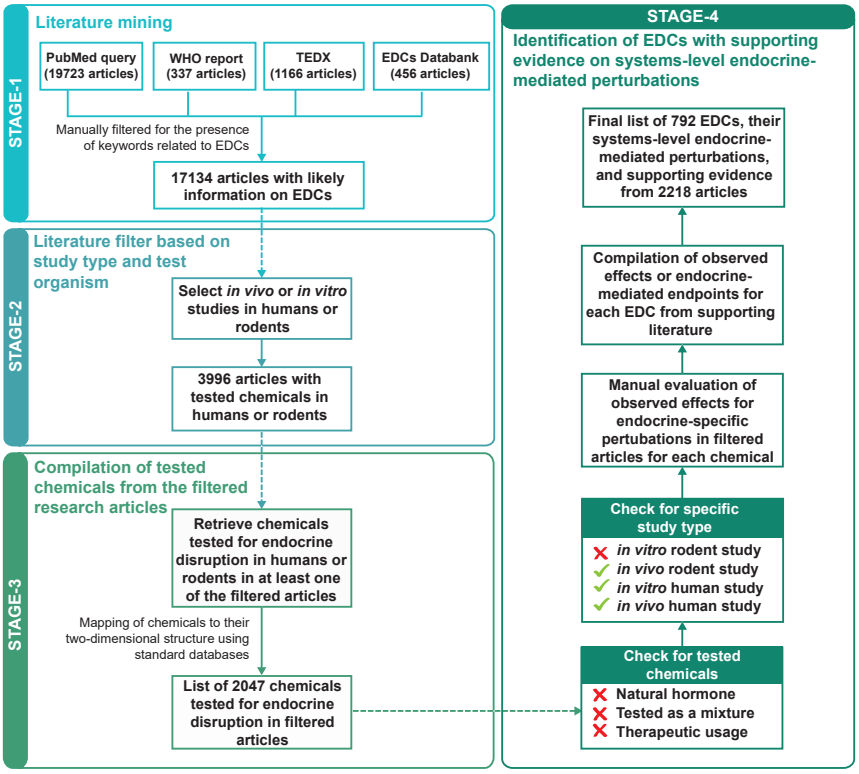


Figure 3.2: Detailed workflow for the compilation of potential EDCs and creation of the updated knowledgebase DEDuCT 2.0.

literature of 2218 research articles, to study the identification of new EDCs in the past decades. In Figure 3.1B, we show the number of new EDCs reported in published literature over the last 70 years. For this analysis, we consider a potential EDC captured in DEDuCT 2.0 to be identified for the first time in a particular year, if the earliest supporting experimental evidence for that EDC is from a research article published in that year. From Figure 3.1B, it is seen that the number of new EDCs identified in the scientific literature has slowly but surely increased on average over the past decades. These observations also align with the observed growth in scientific literature on EDCs [36].

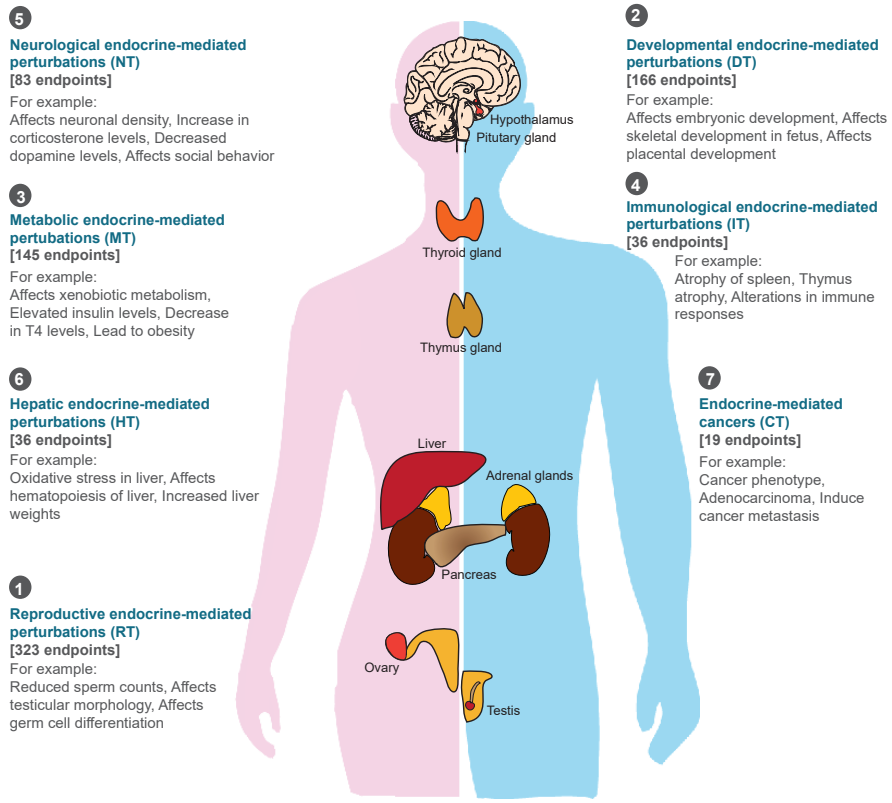


Figure 3.3: Schematic figure depicting the classification of the 609 endocrine-mediated endpoints into 7 systems-level perturbations in DEDuCT 2.0.

A unique feature of our resource, DEDuCT 2.0, on EDCs is the compilation of observed 609 unique endocrine-mediated endpoints and their classification into 7 systems-level perturbations from supporting literature (Figure 3.3) [35]. We have also studied the available evidence for any of the 7 different systems-level perturbations across the 792 potential EDCs in DEDuCT 2.0 (Figure 3.1C). Of the 792 potential EDCs in DEDuCT 2.0, 616 EDCs have evidence for reproductive endocrine-mediated perturbations, 369 EDCs for metabolic perturbations and 251 EDCs for neurological perturbations (Figures 3.1C and 3.3). This reflects that reproductive effects followed by metabolic effects may have been the main focus of the scientific investigations on EDCs [36].

Since DEDuCT compiles potential EDCs with supporting evidence specific to humans or rodents [35], we also considered three other resources on EDCs, namely, the WHO report [8], TEDX and the EDCs Databank [48] for the subsequent analysis. Figure 3.1D also gives an overview of unique and overlapping EDCs across the four resources.

Figure 3.4: Classification of the 792 potential EDCs in DEDuCT 2.0 into 7 broad categories and 48 sub-categories based on their source in the environment. In this figure, the number of EDCs in DEDuCT 2.0 contained in each category or sub-category is reported within the parenthesis.

Specifically, 242 EDCs in DEDuCT 2.0 are not captured in any of the other three resources. In subsequent sections, we compare chemical lists pertaining to guidelines or regulations with the union of EDCs across these four resources which add up to 1856 potential EDCs (Figure 3.1D) [36].

Additional information on EDCs in DEDuCT 2.0

In addition to experimental evidence, DEDuCT 2.0 also compiles diverse information for the 792 potential EDCs including 2D and 3D chemical structure, physicochemical properties, predicted ADMET properties, molecular descriptors, and experimentally inferred target genes from ToxCast database version August 2019 [215]. We also provide a classifica-

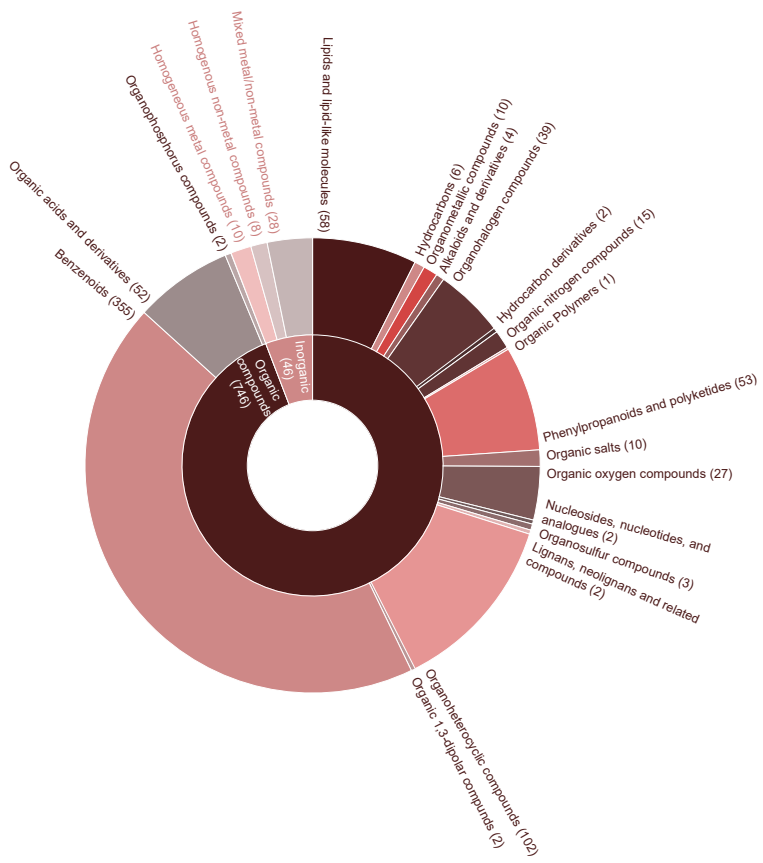


Figure 3.5: Classification of the 792 EDCs in DEDuCT 2.0 into chemical kingdoms and chemical super-classes using ClassyFire. Of the 792 EDCs, 746 are organic and 46 are inorganic compounds. The 746 organic EDCs can be further classified into 19 super-classes while the 46 inorganic EDCs fall into 3 super-classes. The number of EDCs in each super-class is reported within the parenthesis.

tion of the potential EDCs based on their environmental source into 7 broad categories and 48 sub-categories (Figure 3.4). We also provide a hierarchical classification of the 792 potential EDCs based on their chemical structure information using ClassyFire [174] (Figure 3.5). Moreover, the final list of 792 potential EDCs were classified into 4 categories (I-IV) based on the type of supporting evidence for endocrine disruption in published experiments specific to humans or rodents (Supplementary Table S3.2). All the compiled information in DEDuCT 2.0 is accessible at: https://cb.imsc.res.in/deduct/ [35, 36]. In sum, the expanded list of potential EDCs in DEDuCT 2.0 can assist academia, industry, and regulatory agencies in developing safer consumer products.

3.2 Compilation of chemical lists that are a part of inventories, regulations and guidelines

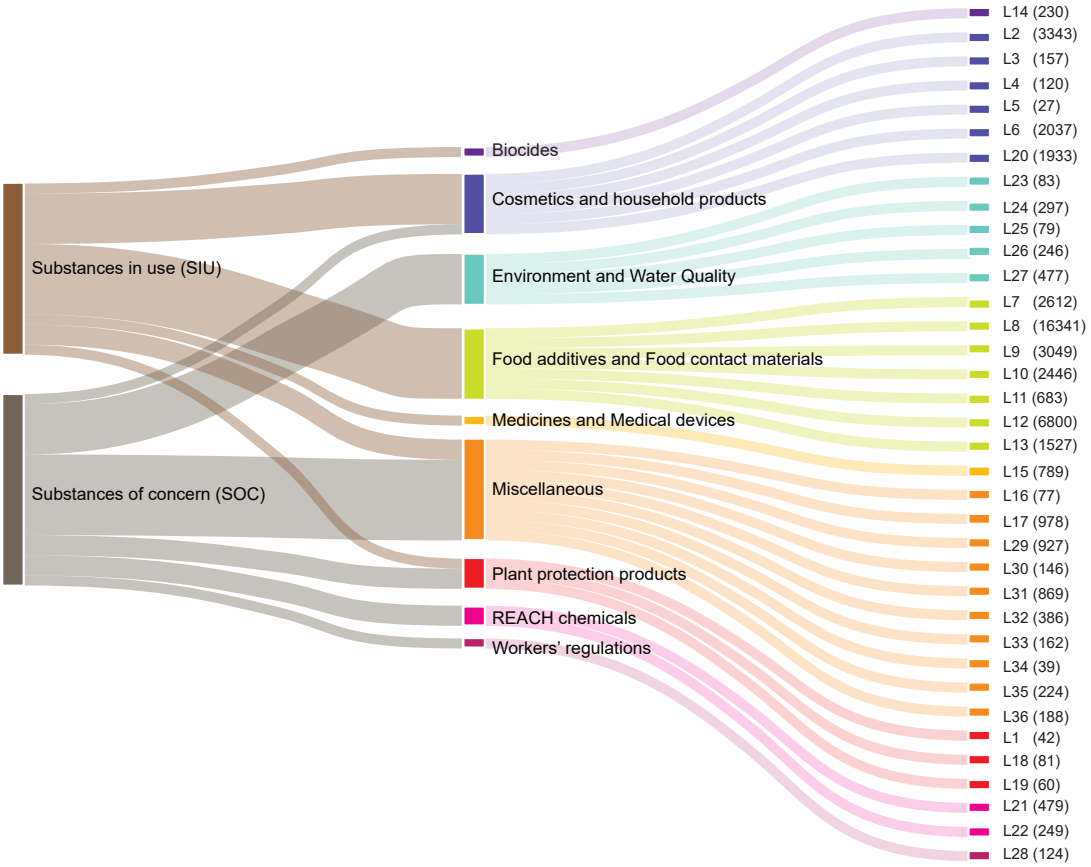
To explore the extent to which current knowledge on EDCs in scientific literature is reflected in guidelines on chemical use or regulations worldwide, we systematically compiled such lists of chemicals that are part of inventories, regulations and guidelines from public resources. For this work, we were able to compile 36 chemical lists which were broadly classified into two categories, namely, 'Substances in use (SIU)' and 'Substances of concern (SOC)' (Figure 3.6; Supplementary Table 3.3). In Supplementary Table 3.3, we provide a detailed description of these 36 chemical lists (L1-L36).

Apart from the broad classification into SIU or SOC, we have also organized the 36 chemical lists into 9 categories based on the recent report commissioned by the European Parliament [156]. These 9 categories include Plant protection products, Cosmetics and household products, Food additives and Food contact materials, Biocides, Medicines and Medical devices, REACH chemicals, Environment and Water Quality, Workers' regulations, and Miscellaneous (Figure 3.6; Supplementary Table S3.3). Note that we were able to find from public resources both SIU and SOC lists for only 3 out of these 9 categories (Figure 3.6; Supplementary Table S3.3).

For unequivocal analysis of chemicals in these 36 chemical lists representing guidelines or regulation, their respective CAS [164] identifiers were used throughout this chapter.

3.2.1 Substances in use (SIU) lists

A list is considered a SIU list if it fulfills one of the following criteria: (a) It is an inventory of substances generally found to be in use in a certain product category; (b) It is a part of a guideline document, issued either by a government agency or an independent body, for safer product formulation; (c) It is a list of substances permitted for use in a certain



Substances in use (SIU)

L1	Active ingredients allowed in minimum risk pesticide products
L2	IFRA transparency list
L3	EU list of colorants allowed in cosmetic products
L4	EU list of preservatives allowed in cosmetic products
L5	EU list of UV filters allowed in cosmetic products
L6	Consumer product ingredient database
L7	Substances added to food (EAFUS)
L8	FooDB
L9	The Joint FAO/WHO Expert Committee on Food Additives (JECFA) list
L10	EU food flavorings database
L11	EU plastic food packaging materials
L12	Pew list of food additives
L13	ESCO list of non-plastic food contact materials
L14	ECHA biocidal products
L15	US FDA inactive ingredient list
L16	Production of major chemicals year-wise in India
L17	US EPA safer chemical ingredients list

Substances of concern (SOC)

L18	List of banned pesticides in India
L19	List of banned and restricted pesticide products in China
L20	EU list of substances prohibited in cosmetic products
L21	Restricted substances under REACH
L22	SVHC under REACH
L23	NPI Australia
L24	Singapore list of controlled hazardous substances
L25	Ozone-depleting substances in India
L26	EWG tap water database
L27	Human Indoor Exposome database
L28	US OSHA list
L29	SIN List
L30	Toxic chemicals restricted to be imported or exported in China
L31	IARC monographs on carcinogens
L32	Schedule 1 hazardous chemical list in India
L33	Schedule 3 hazardous chemical list in India
L34	NZ EPA priority chemical list
L35	ECHA list of chemicals in Annex I
L36	PACSs list Japan

Figure 3.6 ***(previous page)*:** Sankey plot showing the classification of 36 chemical lists that are part of inventories, guidelines and regulations obtained from public resources. The 36 chemical lists were broadly classified into two categories, namely, 'Substances in use (SIU)' and 'Substances of concern (SOC)'. Based on chemical use or environmental source, the 36 chemical lists are further organized into 9 categories, namely, Plant protection products, Cosmetics and household products, Food additives and Food contact materials, Biocides, Medicines and Medical devices, REACH chemicals, Environment and Water Quality, Workers' regulations, and Miscellaneous. In this figure, the number of chemicals in each list is reported in parenthesis besides each list.

product category, by a regulatory authority. Note that though inventories, regulations and guidelines, from where the 17 SIU lists were compiled, may have followed their own criteria to define the specific chemical lists, it is evident that the chemicals captured in these 17 SIU lists are in use in various consumer and industrial products.

Further the 17 SIU lists were classified into 6 categories including Plant protection products, Cosmetics and household products, Food additives and Food contact materials, Biocides, Medicines and Medical devices, and Miscellaneous (Figure 3.6; Supplementary Table S3.3). Of the 17 SIU lists, the category 'Food additives and Food contact materials' has the maximum number of chemical lists (L7-L13), while 'Plant protection products', 'Biocides', and 'Medicines and Medical devices' contain only one chemical list in each of their category. Five SIU lists (L2-L6) fall under the 'Cosmetics and household products' category. Two lists namely, 'L16 - Production of major chemicals year-wise in India' and 'L17 - US EPA safer chemical ingredients list' were categorized under 'Miscellaneous' lists.

An example of SIU list is the 'L7 - Substances added to food (EAFUS)' which is an inventory developed by the US Food and Drug Administration (FDA), and this list was previously known as Everything Added to Foods in the United States (EAFUS) (Figure 3.6; Supplementary Table S3.3). The L7 list contains 2612 unique chemicals which are used as food additives, color additives and other substances approved for specific use in food by the US FDA (Figure 3.6; Supplementary Table S3.3).

3.2.2 Substances of concern (SOC) lists

A list is considered a SOC list if it fulfills one of the following criteria: (a) It is an inventory of substances considered toxic, published either by a government agency or an independent body; (b) It is a list of substances monitored, restricted or banned for import, export or manufacture by a regulatory authority, due to their hazard potential. Following the above criteria, we have compiled 19 SOC lists that are a part of chemical inventories, regulations or guidelines.

The SOC lists were further divided into 6 categories, namely, Plant protection products, Cosmetics and household products, REACH chemicals, Environment and Water Quality, Workers' regulations, and Miscellaneous. Of these 6 categories, REACH chemicals, Environment and Water Quality, and Workers' regulations were specific to SOC lists. The 'Plant protection products' category has two lists (L18-19) specific to banned/restricted pesticidal substances. The categories, 'Cosmetics and household products' and 'Workers' regulations', each constitute only one chemical list containing the substances that are prohibited in cosmetic products (L20) and the substances with potential occupational hazards (L28), respectively. Two lists, namely, 'L21 - Restricted substances under REACH' and 'L22 - SVHC under REACH' were categorized as 'REACH chemicals'. The category 'Environment and Water Quality' includes five lists (L23-L27) containing the list of substances that were monitored by the environmental agencies across different countries. Of 19 SOC lists, 8 chemical lists were categorized as 'Miscellaneous' that identified the substances of potential hazard.

An example of SOC list is the 'L24 - Singapore list of controlled hazardous substances' which is a chemical regulatory list compiled under the Schedule 2 of the Environmental Protection and Management Act of Singapore (Figure 3.6; Supplementary Table S3.3). The L24 list contains 297 hazardous substances (Figure 3.6; Supplementary Table S3.3).

3.3 Exploration of potential EDCs across chemical lists that are a part of inventories, regulations and guidelines

Following the compilation of potential EDCs from four resources and 36 chemical lists, we have performed a three step systematic analysis to understand how potential EDCs are distributed across SIU and SOC lists.

First, we tried to identify any chemical overlap between the SIU and SOC lists. Upon finding a large chemical overlap between these two classes, we split the chemicals from the SIU and SOC lists into 3 groups (I-III). Group I consists of chemicals that are present only in 17 SIU lists, and not in any of the 19 SOC lists. Group II represents the list of chemicals that are present both in 17 SIU and 19 SOC lists. Group III represents the list of chemicals that are present only in 19 SOC lists, and not in any of the 17 SIU lists. We found 23483, 1139 and 3223 chemicals in group I, II and III, respectively (Figure 3.7A).

Second, we compared the list of potential EDCs compiled from 4 resources, namely, DEDuCT 2.0, the WHO report, TEDX and EDCs Databank, with the group I chemicals. We refer to the list of potential EDCs in group I chemicals as group I EDCs or 'EDCs in use (EIU)' (Figure 3.7A). A similar comparison also led to group II EDCs and group III EDCs (Figure 3.7A). Based on the comparison, we find 242, 356 and 278 potential EDCs in groups I, II and III, respectively (Figure 3.7A; Supplementary Table S3.4) [36]. Note that group II which is the intersection of chemicals present in SIU and SOC lists, contains more EDCs than groups I or III.

Third, we compared the EIU list with the list of High Production Volume (HPV) chemicals to identify the potential EDCs in use which are produced or manufactured in high volume. For this analysis, we have compiled HPV chemicals from the union of two resources, namely, the United States High Production Volume (USHPV) database and the Organisation for Economic Co-operation and Development (OECD) High Production

Volume (OECD HPV) list last updated on 2004. The OECD HPV list contains 4712 chemicals that are produced more than 1000 tonnes per year in at least one OECD member country or region. The USHPV database compiles 4297 chemicals that are produced or imported in the United States in quantities of 1 million pounds or more per year. A similar comparison of the group II EDCs and group III EDCs was also performed with the HPV chemicals.

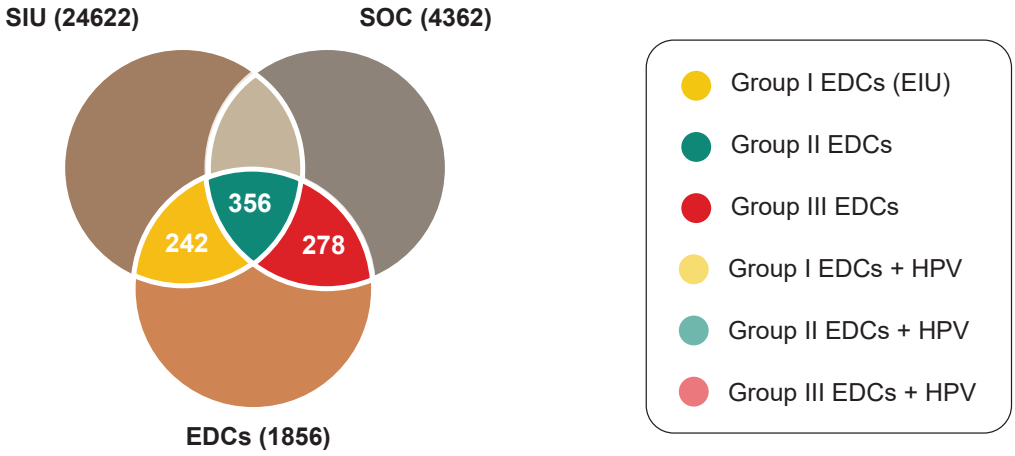
3.3.1 Potential EDCs across substances in use

We designate the 242 potential EDCs among group I chemicals as EDCs in use (EIU) (Figure 3.7A; Supplementary Table S3.4). These 242 EIU are distributed across 5 of the 9 categories of chemical lists, and thus pose a high risk of exposure (Figure 3.7A). Majority of EIU are found in 2 categories of chemical lists, namely, 'Food additives and Food contact materials' and 'Cosmetics and household products'. Minority of EIU are found in 3 categories of chemical lists, namely, 'Biocides', 'Medicines and Medical devices' and 'Miscellaneous' (Figure 3.7B; Supplementary Table S3.4). Of the 242 EIU, DEDuCT 2.0 captures 119 potential EDCs along with supporting experimental evidence (Supplementary Table S3.4). Lastly, 6 EIU, namely, 2,4,5,2',4',5'-Hexabromobiphenyl, Coumestrol, Daidzein, Genistein, Pendimethalin and Zearalenone are captured in all four resources on EDCs (Supplementary Table S3.4) [36].

3.3.2 EDCs in use and high production volume chemicals

EIU produced in high volume can pose significant risk as humans are readily exposed to them through use of commercial products. Figure 3.7B gives the distribution of 63 EIU produced in high volume across 5 different categories of chemical lists (Supplementary Table S3.4). While none of EIU produced in high volume are captured in all four resources on EDCs, 7 EIU produced in high volume, namely, 4,4'-Dihydroxybiphenyl, 4-Hydroxybenzoic acid, 4-sec-Butylphenol, Chlorocresol, Monosodium glutamate, N,N'-Diphenyl-4-phenylenediamine and Sodium fluoride, are captured in three of the four re-

A


SIU (24622)
SOC (4362)
356
242
278
EDCs (1856)
Group I EDCs (EIU)
Group II EDCs
Group III EDCs
Group I EDCs + HPV
Group II EDCs + HPV
Group III EDCs + HPV

B

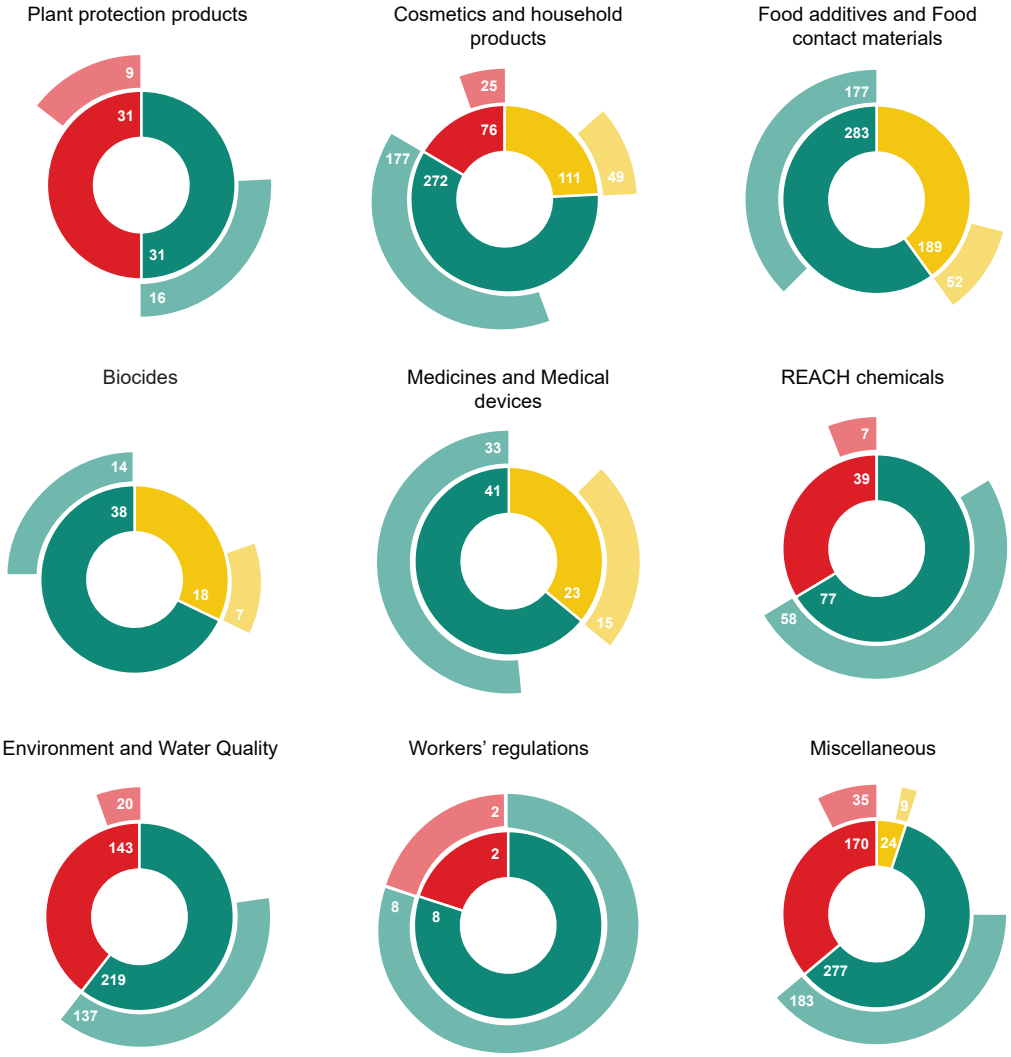

Plant protection products
9
31
31
16
Cosmetics and household products
25
76
177
272
111
49
Food additives and Food contact materials
177
283
189
52
Biocides
14
38
18
7
Medicines and Medical devices
33
41
23
15
REACH chemicals
7
39
77
58
Environment and Water Quality
20
143
219
137
Workers' regulations
2
2
8
8
Miscellaneous
35
9
170
24
277
183

Figure 3.7 ***(previous page)*****:** Distribution of potential EDCs from four resources, namely, DEDuCT 2.0, WHO report, TEDX and EDCs Databank, across 36 chemical lists that are part of inventories, guidelines and regulations. (A) Venn diagram displaying the intersections of group I, II and III chemicals with potential EDCs. (B) Sunburst plot showing the distribution of potential EDCs across 9 categories of chemical lists. Within each category in this plot, the inner ring gives the number of potential EDCs in group I, II and III, and the outer ring gives the number of potential EDCs in group I, II and III that are also high production volume (HPV) chemicals.

sources on EDCs. These 7 EIU produced in high volume are found in 4 categories of chemical lists, namely, 'Biocides', 'Cosmetics and household products', 'Food additives and Food contact materials' and 'Medicines and Medical devices' (Figure 3.7B; Supplementary Table S3.4). Finally, 31 of the 63 EIU produced in high volume are captured in DEDuCT 2.0 (Supplementary Table S3.4) [36].

From this analysis, it is evident that several EDCs in commercial use are also produced in high volume. The risk of exposure and associated hazard potential warrant an evaluation of these EIU produced in high volume, and framing appropriate risk assessment criteria will help such efforts. Later in this chapter, we illustrate how our knowledgebase, DEDuCT 2.0, on EDCs can aid in risk assessment.

3.3.3 Potential EDCs across group II and III chemicals

There are 356 group II EDCs (Figure 3.7A) of which 211 are also HPV chemicals. Among the 356 group II EDCs, 46 are captured in all four resources on EDCs (Supplementary Table S3.4). Of these 46 group II EDCs, 28 are also produced in high volume. These 28 group II EDCs produced in high volume are distributed across 6 categories of chemical lists, namely, 'Plant protection products', 'Cosmetics and household products', 'Food additives and Food contact materials', 'Environment and Water Quality', 'REACH chemicals' and 'Miscellaneous' (Supplementary Table S3.4) [36]. Given the volume of production and their possible presence in commercial products, the risk of human exposure to these potential EDCs is a concern.

We next analyzed group III chemicals which are only present in SOC lists and found

278 potential EDCs among them (Figure 3.7A). Of these 278 group III EDCs, 5 chemicals, namely, Simazine, Linuron, Acetochlor, Vinclozolin, and Prochloraz, were found to be produced in high volume and captured in all four resources on EDCs (Supplementary Table S3.4). These 5 group III EDCs are distributed across 4 categories of SOC lists, namely, 'Plant protection products', 'Cosmetics and household products', 'Environment and Water Quality', and 'Miscellaneous' (Supplementary Table S3.4). These 5 potential EDCs in SOC lists need better monitoring as they are produced in high volume in spite of known concern [36].

We further analyzed the distribution of HPV chemicals within potential EDCs in group II or III across the 9 categories of chemical lists (Figure 3.7B). Interestingly, we find that 33 out of 41 group II EDCs within the 'Medicines and Medical devices' category are produced in high volume. Also, all group II or III EDCs within 'Workers' regulations' category are produced in high volume indicating the risk of occupational exposure. Note that we were able to obtain only a single SOC list with 124 chemicals in the category of 'Workers' regulations' of which 10 are potential EDCs produced in high volume (Figure 3.7B), and this analysis reveals the current gap in regulation of occupational exposure to hazardous chemicals. Moreover, 3 potential EDCs namely, formaldehyde, ethylene oxide and methyl bromide, in the SOC list in 'Workers' regulations' category are captured in three out of four resources on EDCs. Also, formaldehyde and ethylene oxide are present in 8 SIU lists suggesting potential risk of exposure from use of common products [36]. In sum, a thorough evaluation of potential EDCs in 36 chemical lists (L1-L36), and incorporation of diverse information captured in scientific literature can improve safety assessment and regulation of EDCs.

3.4 A case study of DEDuCT 2.0 in risk assessment of EDCs

To better understand how diverse information in a curated knowledgebase such as DEDuCT 2.0 [35, 36] can aid in chemical regulation, we present a case study for a potential EDC. We focused on 28 group II EDCs produced in high volume and captured in all four resources on EDCs including DEDuCT 2.0. Of these 28 group II EDCs, 'Dibutyl phthalate (CAS: 84-74-2)' is a potential EDC present in 6 SIU lists and 7 SOC lists which are distributed across 5 categories, namely, 'Cosmetics and household products', 'Food additives and Food contact materials', 'REACH chemicals', 'Environment and Water Quality', and 'Miscellaneous'. We next discuss the utility of DEDuCT 2.0 in risk assessment of chemicals using Dibutyl phthalate as an example.

According to the United States National Academy of Sciences, risk assessment involves four steps, namely, Hazard identification, Dose-response assessment, Exposure assessment, and Risk characterization [216]. Among the four resources on EDCs, notably, DEDuCT has compiled the observed endocrine-mediated endpoints and the dosage at which endpoints are observed, from published experiments specific to humans and rodents [35, 36], and this information can aid in risk assessment process. DEDuCT 2.0 compiles supporting evidence on endocrine disruption upon Dibutyl phthalate exposure from *in vivo* experiments in rodents and *in vitro* experiments in humans which were published in 35 research articles.

For the first step in risk assessment, we used DEDuCT 2.0 to identify health hazards posed by Dibutyl phthalate. For Dibutyl phthalate exposure, DEDuCT 2.0 has compiled 81 endocrine mediated endpoints spanning 7 systems-level perturbations, namely, reproductive, developmental, metabolic, immunological, neurological, hepatic, and endocrine-mediated cancer (Figure 3.3). For the second step in risk assessment, one can use the dosage information compiled in DEDuCT 2.0 for 81 endpoints observed upon Dibutyl

phthalate exposure. In particular, we have analyzed the dosage information for Dibutyl phthalate compiled in DEDuCT 2.0 specific to endpoints observed in *in vivo* rodent studies using dosage unit as mg/kg/day (Supplementary Table S3.5). In these published *in vivo* rodent studies on Dibutyl phthalate, the test concentration range for different endpoints is 0.01-1000 mg/kg/day across compiled studies in DEDuCT 2.0, the lowest dose at which an adverse effect is observed in any of these studies is 0.01 mg/kg/day, and the highest dose at which no adverse effects are observed in any of the studies is 125 mg/kg/day (Supplementary Table S3.5). We remark that the compiled dosage information for Dibutyl phthalate in DEDuCT 2.0 is compatible with previous reports suggesting possible non-monotonic dose response for this chemical [217].

The third step of exposure assessment involves the identification of routes, frequency and duration of exposure at the population level. Though DEDuCT 2.0 compiles information on environmental sources of potential EDCs, it does not capture their duration and routes of exposure. A possible expansion of the knowledgebase to include biomonitoring and epidemiological information for EDCs from published literature will further aid in exposure assessment and risk characterization; however, such an update of DEDuCT 2.0 requires significant effort beyond the current scope of our work.

3.5 Discussion

The number of chemicals introduced into the market for commercial purposes continues to be high. Adequate risk assessment strategies are needed now, more than ever, to cope with the increasing demand for safe product formulations. In general, regulatory standards and criteria differ across countries and this lack of standardization applies to the regulation of EDCs as well [218, 219]. The regulatory assessment of EDCs is complex as there are several challenges and limitations associated with these substances [3, 218]. In recent years there has been a rapid increase in endocrine disruption studies and the accumulation of knowledge surrounding EDCs (Figure 3.1A,B). However, regulatory

assessments fall short due to the limitations and uncertainties in the risk assessment of EDCs [3, 218, 220]. This may be also due to the lack of knowledge transfer from academic research to the regulatory assessment of EDCs.

The presence of potential EDCs in the compiled chemical lists is a concern as humans are exposed to these potential EDCs via the use of industrial and consumer products. Similar investigations have previously been conducted for food, food additives, and food contact chemicals [154, 155], and these studies have revealed regulatory gaps that contribute to the inclusion of substances of concern in food and associated products. However, these studies were not specific to EDCs, and were also limited to a single category of substances. Hence, there is a need to incorporate endocrine disruption as a standard criterion in chemical risk assessment. Despite scientific efforts to evaluate the risks that EDCs pose, there is a gap in the transfer of knowledge to the policy planning level [214]. Focused systematic review of these lists by regulatory agencies and non-governmental chemical advocacy groups, coupled with better incorporation of research data compiled in academic resources may help improve and strengthen chemical regulations and guidelines, and consequently, improve the safety of our products as well.

Based on the extent and variety of information necessary for building regulatory standards, the utility of the WHO report, TEDX, and EDCs Databank in regulatory assessment may be limited. These resources lack the systematic compilation of observed adverse effects specific to endocrine disruption from published literature. The compilation of endocrine-mediated adverse effects along with dosage information in DEDuCT 2.0 may prove valuable in the risk assessment and regulation of EDCs as demonstrated using a case study for Dibutyl phthalate in this chapter. Additional information including species, strain, sex, route, and duration of exposure for the compiled EDCs from published literature will aid in better risk assessment of chemicals. Moreover, a possible update of DEDuCT to include biomonitoring and epidemiological studies for the compiled EDCs from published literature can also aid in exposure assessment and risk characterization. However, such an update of DEDuCT will also require an intensive manual curation effort.

To this end, experimental evidence of endocrine disruption for potential EDCs compiled in knowledgebases could help in the early identification of hazardous substances, so that regulatory bodies can then streamline the process for safety testing, and in turn improve chemical safety standards.

Supplementary Information

Supplementary Tables S3.1-S3.5 associated with this chapter are available for download from the GitHub repository: https://github.com/asamallab/PhDThesis-Janani_R/blob/main/SI/ST_Chapter3.xlsx.

Chapter 4

Derivation, characterization and analysis of an Adverse Outcome Pathway network relevant for endocrine disruption

Chemical regulatory risk assessment is based on *in vivo* methods, which are time consuming, costly, and necessitate the use of a large number of animals for testing [221,222]. To improve and accelerate chemical toxicity testing, the US National Research Council published a vision report in 2007 titled 'Toxicity testing in the 21st century: a vision and a strategy' recommending the implementation of high-throughput screening methods such as *in vitro* toxicology or *in silico* approaches [93,94,96,98]. In this context, 'toxicity pathways' were proposed to capture the perturbed biological events that occur as a result of chemical exposure and can be utilised to predict the observed adverse effects [93–96,98]. Later, the concept of Adverse Outcome Pathways (AOPs) was suggested to organize available mechanistic knowledge on observed adverse effects in humans or wildlife following chemical exposure [99]. Subsequently, several studies have reported the development of specific AOPs and their applications in risk assessment [97,104–106,108,111,112,223].

In 2012, the OECD launched an international program to formalize the development and evaluation of AOPs. This has led to a series of OECD guidance documents [101–103] and primary literature [97, 104–106, 108, 109, 111, 112, 223] for the development of AOPs and their potential applications in human- and eco-toxicology. AOP-Wiki [114], an actively maintained module within AOP-KB created by OECD serves as a central repository of AOPs at various stages of development [105, 106].

Application of network-based approaches can aid in unraveling the organizing principles of complex biological systems [88]. A primary goal of the emerging discipline, computational systems toxicology, is to harness network and systems biology approaches in building predictive toxicological models through heterogeneous data integration across diverse levels of biological organization [224–227]. The AOP framework has an inherent modular structure which enables sharing of KEs and KERs between individual AOPs, and this sharing of KEs leads to emergence of 'AOP networks' [107, 110, 228]. Knapen *et al.* [107] have defined an 'AOP network' as: "an assembly of 2 or more AOPs that share one or more KEs, including specialized KEs such as MIEs and AOs". To date, 9 AOP networks have been derived from AOP-Wiki to address specific toxicity problems related to reproduction [115, 116], development [115], nervous system [229, 230], liver [231], metabolism [107, 110] and immune system [232].

On similar lines, the AOP framework is ideal for organizing the existing knowledge and providing a pathway perspective on diverse modes of endocrine disruption by EDCs [233–235]. Moreover, the development and analysis of an AOP network relevant to endocrine disruption has the potential to reveal key events, critical paths, and unexpected links between individual AOPs capturing varied adverse effects [107, 110]. Previously, there have been few efforts to construct AOP networks for disruption specific to a single hormone, namely, androgen [116], thyroid, or thyroxine [107, 110]. Due to the focus on specific hormones, the constructed AOP networks in these studies do not provide a comprehensive picture of all endocrine disruption mechanisms captured within AOP-Wiki. In this chapter, we first aim to build a comprehensive derived AOP network for endocrine

disruption by curating and organizing existing toxicological information from AOP-Wiki. Second, we aim to utilize this derived AOP network for endocrine disruption to better understand the perturbed biological events involving multiple systems that occur when exposed to environmental chemicals. Finally, we use graph-theoretic measures to identify critical biological events, emergent new paths, chemical stressors associated with the events, and possible adverse outcomes following EDC exposure. Such information can aid in the development of new endpoints or assays for better risk assessment of environmental chemicals. **The work reported in this chapter is contained in the published manuscript [37].**

4.1 Derived AOP network relevant for endocrine disruption

4.1.1 Compilation of AOP dataset from AOP-Wiki

The aim of this study is to develop a derived AOP network relevant to endocrine disruption based on information in AOP-Wiki. From the Project Downloads section (https://aopwiki.org/downloads) of the AOP-Wiki, we have downloaded the XML archive as on 03 January 2021. This XML archive from AOP-Wiki was parsed using the xml2 package in R to obtain information on AOPs, Key Events (KEs), Key-Event Relationships (KERs), and stressors. To construct this AOP network relevant to endocrine disruption or 'ED-AOP network', we have compiled detailed information on 316 AOPs, 1131 KEs and 1363 KERs from AOP-Wiki. Due to continuous development of AOP-Wiki, some AOPs may have incomplete information at any particular time (Figure 4.1).

For each AOP in AOP-Wiki, we have retrieved information including the AOP identifier, AOP title, OECD status, and Society for the Advancement of AOPs (SAAOP) status. For each KE in an AOP, we have gathered information including the KE identifier, KE type, level of biological organization and taxonomy. The KE type can be either molecular

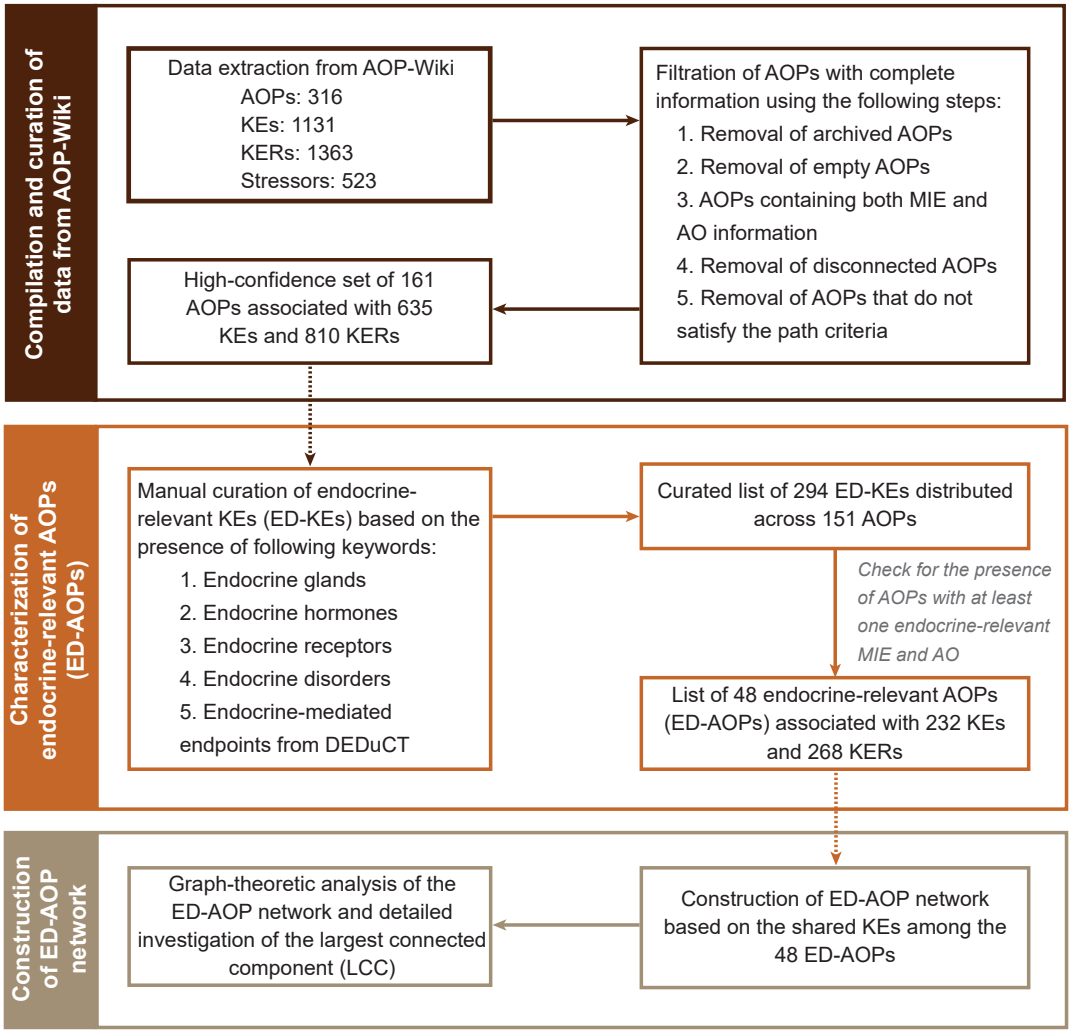


Figure 4.1: Detailed workflow for the development, characterization and analysis of an adverse outcome pathway (AOP) network for endocrine disruption.

initiating event (MIE), key event (KE) or adverse outcome (AO). For each KER in an AOP, we have gathered information including the KER identifier, upstream KE, downstream KE, the weight of evidence (WoE), adjacency information, and the quantitative understanding score (OECD, 2018). Lastly, we have compiled the chemical stressors linked to KEs in different AOPs along with their structure information such as the CAS identifier [164], DSSTOX identifier [236] and InChIKey. Note that the AOP-Wiki also contains information on non-chemical stressors such as genetic or environmental factors.

We remark that each AOP can be viewed as a directed graph or network wherein the nodes are KEs and directed edges are KERs linking upstream KEs with downstream KEs. In this directed graph representation of an AOP, it is straightforward to determine the existence of a directed path between any pair of KEs.

4.1.2 Filtration of high-confidence AOPs from AOP-Wiki

Since AOP-Wiki is under continuous development, some AOPs may have incomplete information [101]. Therefore, it is important to evaluate the quality and completeness of information in each AOP before their selection for the derived AOP network construction [110]. We have assessed the quality and completeness of information in each AOP obtained from AOP-Wiki as follows (Figure 4.1).

Firstly, we have removed the 'archived AOPs' based on SAAOP status as these are no longer under active development. This led to the removal of 6 AOPs. Secondly, we have removed 'empty AOPs', which are AOP pages created in AOP-Wiki but lack a KE or a KER [228]. After removing 'archived AOPs' and 'empty AOPs', we have 218 AOPs that remain under consideration. Thirdly, we have removed any AOP which does not contain at least one MIE and at least one AO. After this step, we have 182 AOPs with both MIE and AO that remain under consideration. Fourthly, we have computed the number of (weakly) connected components in each AOP because the presence of more than one component in an AOP may indicate AOPs in the early stages of development

[228]. This led to the identification of 3 disconnected AOPs that have more than one connected component. After the removal of 3 disconnected AOPs, we have 179 AOPs that remain under consideration.

Fifthly, we have computed directed paths from different MIEs to different AOs in each AOP to filter out incomplete AOPs. Since an AOP can have both multiple MIEs and multiple AOs, we have computed the directed paths between each pair of MIE and AO in an AOP to impose this path criterion. We have retained an AOP only if it satisfies the following path criteria:

(a) Every MIE in an AOP has at least one (outgoing) path to at least one AO in the same AOP.
(b) Every AO in an AOP has at least one (incoming) path from at least one MIE in the same AOP.
(c) Every KE in an AOP (other than MIEs and AOs) has at least one incoming path from at least one MIE in the same AOP and at least one outgoing path to at least one AO in the same AOP.

After removing AOPs that do not satisfy the path criteria, we arrive at a high-confidence set of 161 AOPs which are associated with 635 KEs and 810 KERs (Figure 4.1; Supplementary Table S4.1). Next, these 161 high-confidence AOPs were considered for the identification of AOPs relevant for endocrine disruption.

4.1.3 Curated subset of endocrine-relevant AOPs

To build the AOP network specific to endocrine disruption, it is important to identify the subset of endocrine-relevant AOPs (ED-AOPs) among the 161 high-confidence AOPs. To identify ED-AOPs, we have manually curated the endocrine-relevant KEs (ED-KEs) among the 635 KEs associated with the 161 high-confidence AOPs.

A KE was identified as an ED-KE if the KE contains keywords relevant to the endocrine system. Keywords relevant to endocrine system were identified based on: (a)

List of endocrine glands, (b) List of endocrine hormones, (c) List of endocrine receptors where hormones can bind, (d) List of endocrine disorders in MeSH [237], and (e) List of endocrine-specific endpoints in DEDuCT [35, 36]. All of the data used for filtering ED-KEs in the aforementioned criteria are specific to humans or rodents (which are commonly used animal models for human endocrine disruption) [238]. This process led to a curated subset of 294 ED-KEs (Supplementary Table S4.2). Afterwards, we retained 151 AOPs among the 161 high-confidence AOPs that contain at least one ED-KE. Furthermore, we consider an AOP to be an ED-AOP if it contains at least one MIE which is an ED-KE and at least one AO which is an ED-KE. This filtration led to a curated subset of 48 ED-AOPs which are associated with 232 KEs and 268 KERs (Table 4.1; Figure 4.1; Supplementary Table S4.3). Due to the use of humans or rodents-specific data to filter the ED-KEs, the majority of these ED-AOPs contain KEs relevant for humans or rodents.

Subsequently, we have studied the enrichment of ED-KEs across these 48 ED-AOPs by computing the fraction of ED-KEs among KEs in an ED-AOP. Among the curated subset of 48 ED-AOPs, we find that 11 ED-AOPs are such that 100% (all) of their KEs are ED-KEs, and moreover, 45 ED-AOPs are such that at least 50% of their KEs are ED-KEs. Note that the minimum fraction of ED-KEs in an ED-AOP among the 48 ED-AOPs is found to be 37.5% (Table 4.1). Furthermore, we have computed a cumulative weight of evidence (cumulative WoE) score for each of the 48 ED-AOPs based on the weight of evidence (WoE) scores given by AOP-Wiki to the associated 268 KERs. For each KER, the AOP-Wiki gives one of the following values namely, 'high', 'moderate', 'low' or 'not specified' as the WoE score, and this value is a measure of the strength of empirical evidence supporting the causal relationship between the pair of KEs connected by a KER. Note that the WoE scores assigned to KERs by AOP-Wiki can change with updates in the resource [101, 228]. Also, different KERs in any AOP can differ in their WoE scores. Therefore, we propose the following cumulative WoE score for each ED-AOP based on the WoE scores given by AOP-Wiki to associated KERs.

For each ED-AOP, we compute the fraction of KERs with different values of the WoE

score namely, ‘high’, ‘moderate’, ‘low’ or ‘not specified’. For example, the fraction of KERs in an ED-AOP with WoE score ‘high’ can be computed from the ratio of the number of KERs in the AOP with WoE score ‘high’ and the total number of KERs in the AOP, and this quantity for an ED-AOP is denoted by F(‘high’). Similarly, it is straightforward to compute the quantities F(‘moderate’), F(‘low’) and F(‘not specified’) for an ED-AOP. For each of the 48 ED-AOPs, we have computed the quantities F(‘high’), F(‘moderate’), F(‘low’) and F(‘not specified’) from the WoE scores of the associated KERs (Supplementary Table S4.4). Subsequently, we have assigned the cumulative WoE score to each ED-AOP as follows:

(i) If an ED-AOP has F(‘high’) ≥ 0.5, then the cumulative WoE score was assigned to ‘high’.

(ii) Else if an ED-AOP has F(‘high’) < 0.5 but has [F(‘high’) + F(‘moderate’)] ≥ 0.5, then the cumulative WoE score was assigned to ‘moderate’.

(iii) Else if an ED-AOP has [F(‘high’) + F(‘moderate’)] < 0.5 but has [F(‘high’) + F(‘moderate’) + F(‘low’)] ≥ 0.5, then the cumulative WoE score was assigned to ‘low’.

(iv) Else if an ED-AOP has [F(‘high’) + F(‘moderate’) + F(‘low’)] < 0.5, then the cumulative WoE score was assigned to ‘not specified’.

Based on this definition, we find that 18, 12, 1 and 17 ED-AOPs were assigned cumulative WoE score of ‘high’, ‘moderate’, ‘low’ and ‘not specified’, respectively (Table 4.1; Supplementary Table S4.4).

In Supplementary Table S4.5, we compile the biological domain information namely, taxonomic, sex and life stage applicability, for each ED-AOP from AOP-Wiki. For example, AOP:13 is ‘Chronic binding of antagonist to N-methyl-D-aspartate receptors (NMDARs) during brain development induces impairment of learning and memory abilities’. The taxonomic applicability information for AOP:13 indicates that the AOP is applicable to human, mouse, rat and monkey. The sex applicability information for AOP:13 indicates that the AOP is applicable to both sexes (male and female). The life stage applicabil-

ity information for AOP:13 indicates that the AOP is relevant during brain development (Supplementary Table S4.5). Similar to WoE scores for KERs in AOP-Wiki, the WoE information for taxonomic, sex, or life stage applicability for each AOP in AOP-Wiki can have one of the four values namely, 'high', 'moderate', 'low' or 'not specified'. Lastly, we have evaluated the information on taxonomic applicability of the 48 ED-AOPs from AOP-Wiki webpage (last accessed in April 2021) to assess the human applicability of each ED-AOP. We find that 14 out of the 48 ED-AOPs have evidence for human applicability in AOP-Wiki (Table 4.1; Supplementary Table S4.5). Of these 14 ED-AOPs with evidence for human applicability, 4, 4 and 6 ED-AOPs have WoE score for human applicability to be 'high', 'moderate' and 'low', respectively (Table 4.1; Supplementary Table S4.5). Note that if the WoE score for taxonomic applicability of an ED-AOP for Homo sapiens was 'not specified' in AOP-Wiki, we have assigned the WoE score for human applicability of that ED-AOP in Table 4.1 to 'low'.

Evidently, the cumulative WoE score and the WoE score for human applicability listed in Table 4.1 can be used to qualitatively assess the level of evidence for an ED-AOP and further filter the curated subset of 48 ED-AOPs. Nevertheless, we have not imposed any filters based on taxonomic, sex, or life stage applicability information in AOP-Wiki during the filtration of the 48 ED-AOPs for the subsequent construction of the derived AOP network. Note that these WoE scores are qualitative indicators representing the strength of evidence based on current knowledge compiled in AOP-Wiki, and they tend to vary over time. Hence, it is worthwhile to manually evaluate the evidence while applying filters based on these scores specific to research question. In addition, these scores indicate the knowledge gaps in the development of AOPs.

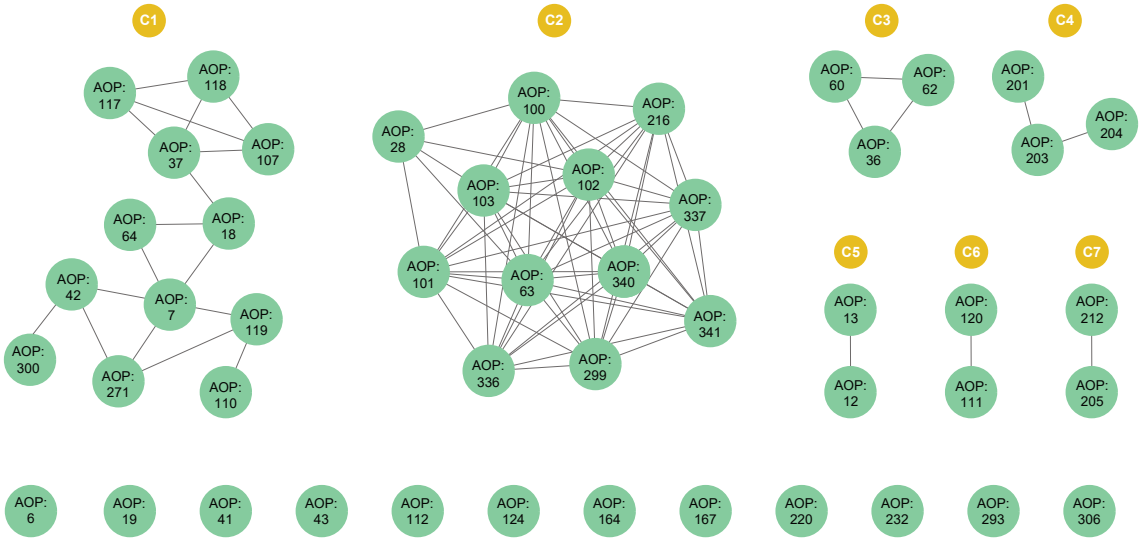


Figure 4.2: Visualization of the ED-AOP network based on shared KEs among the 48 ED-AOPs. Here, each node corresponds to an ED-AOP and there exists an edge between any two ED-AOPs if they have at least one shared KE. The network has 7 connected components (labeled C1-C7) with ≥ 2 ED-AOPs and 12 isolated ED-AOPs. The two largest connected components (LCCs) labeled by C1 and C2 contain 12 ED-AOPs each.

4.1.4 Construction of the ED-AOP network and its connected components

After filtration of the curated subset of 48 ED-AOPs, we have constructed the AOP network specific to endocrine disruption by assembling the information on shared KEs and KERs among the 48 ED-AOPs. We refer to this derived AOP network as 'ED-AOP network' (Figure 4.2). The ED-AOP network contains KEs and KERs across the 48 ED-AOPs, and thus, captures diverse biological perturbations related to endocrine system [107, 110]. The ED-AOP network can be visualized as an undirected graph of 48 nodes corresponding to the 48 ED-AOPs, and there exists an edge between any two nodes in this undirected graph if the two ED-AOPs have at least one shared KE (Figure 4.2).

In this chapter, we have performed a graph-theoretic analysis of the ED-AOP network to reveal important topological features [110]. To assess the overall connectivity of the ED-AOP network, we have computed the connected components using python package NetworkX [239]. A connected component is a subset of nodes in the graph wherein there exists at least one path between every pair of nodes in the induced subgraph. Note that a

completely connected network has a single connected component comprising all nodes in the graph. Based on this computation, we find that the ED-AOP network can be decomposed into 7 connected components with ≥ 2 ED-AOPs and 12 isolated ED-AOPs. These 7 connected components together comprise 36 ED-AOPs (Figure 4.2; Supplementary Table S4.6). Among these 7 connected components, the two largest connected components (LCCs) labeled by C1 and C2 in Figure 4.2 contain 12 ED-AOPs each, and the remaining 5 connected components contain ≤ 3 ED-AOPs each. The LCCs C1 and C2 comprise of 44 and 48 KEs, respectively, of which 19 and 20 KEs are shared among 2 or more ED-AOPs in C1 and C2, respectively (Figures 4.3 and 4.4).

To better understand the systems-level effects of AOs in the 7 components of the ED-AOP network, we have categorized AOs into 4 systems-level endocrine-mediated perturbations, namely, 'hepatic', 'metabolic', 'neurological' and 'reproductive', and this classification depends on the perturbed biological process corresponding to an AO (Table 4.2). For example, the AO titled 'Increase, hepatocellular adenomas and carcinomas' in AOP-Wiki was classified as 'hepatic' while the AO titled 'impaired, Fertility' as 'reproductive' (Table 4.2). This categorization of AOs in ED-AOPs into 4 systems-level perturbations follows a similar classification scheme for observed adverse effects upon exposure to endocrine disrupting chemicals (EDCs) in our previous work [35, 36] described in Chapter 2. We observe that majority of AOs in the ED-AOP network affect the 'reproductive' system (Table 4.2). Moreover, the AOs in C1 can affect 4 different systems, while all AOs in C2 affect solely the 'reproductive' system (Table 4.2).

4.2 Topological analysis of the largest components in the ED-AOP network

Since the two LCCs dominate the ED-AOP network, we decided to next focus on them. For a detailed analysis of each LCC in the ED-AOP network, we have constructed the corresponding directed network wherein nodes are KEs and each directed edge represents

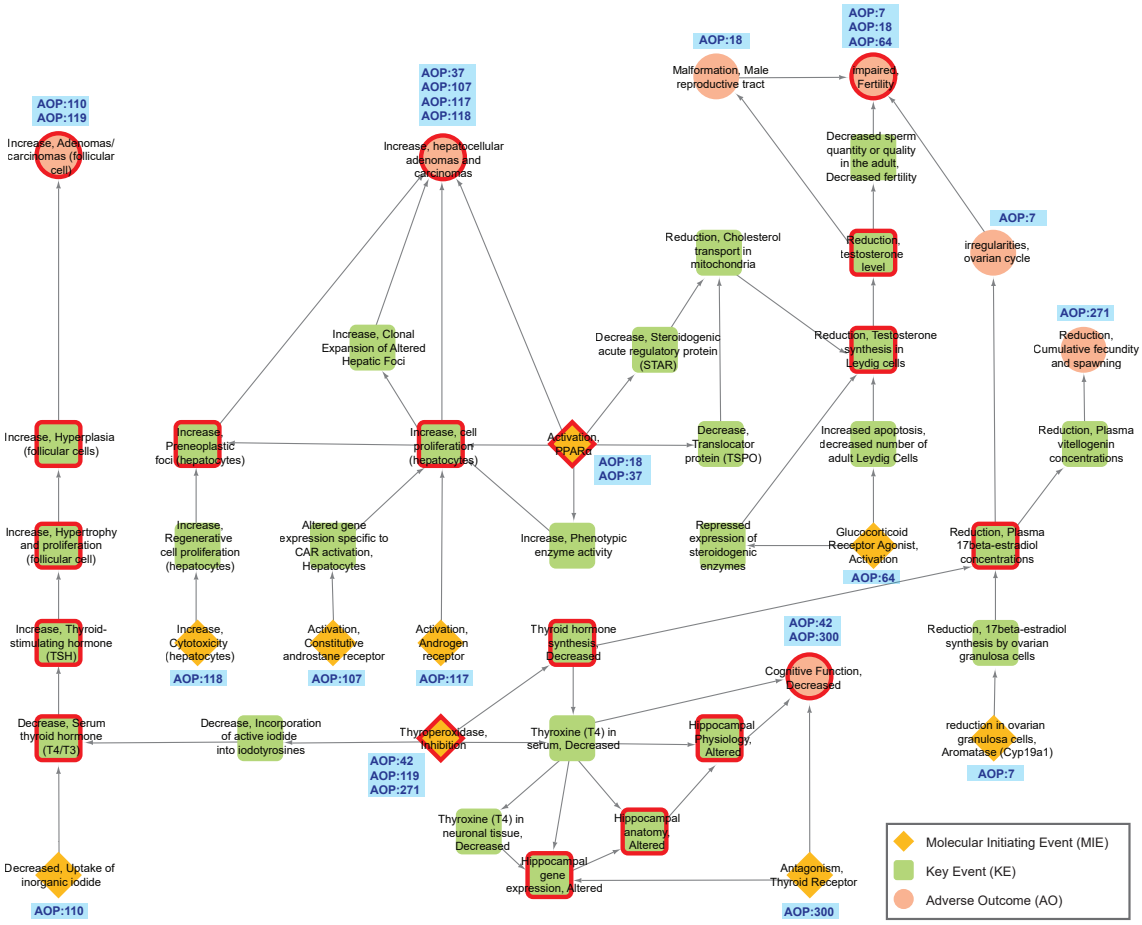


Figure 4.3: The directed network for LCC C1 in the ED-AOP network consisting of 44 KEs and 56 KERs. The 44 KEs in C1 can be categorized into 9 MIEs, 28 KEs and 7 AOs. MIEs, KEs and AOs are shown in distinct shapes namely, diamond, square and circle, respectively. The 19 shared KEs in C1 are marked in 'red'. For each MIE and AO, the corresponding AOP identifier is displayed in this figure.

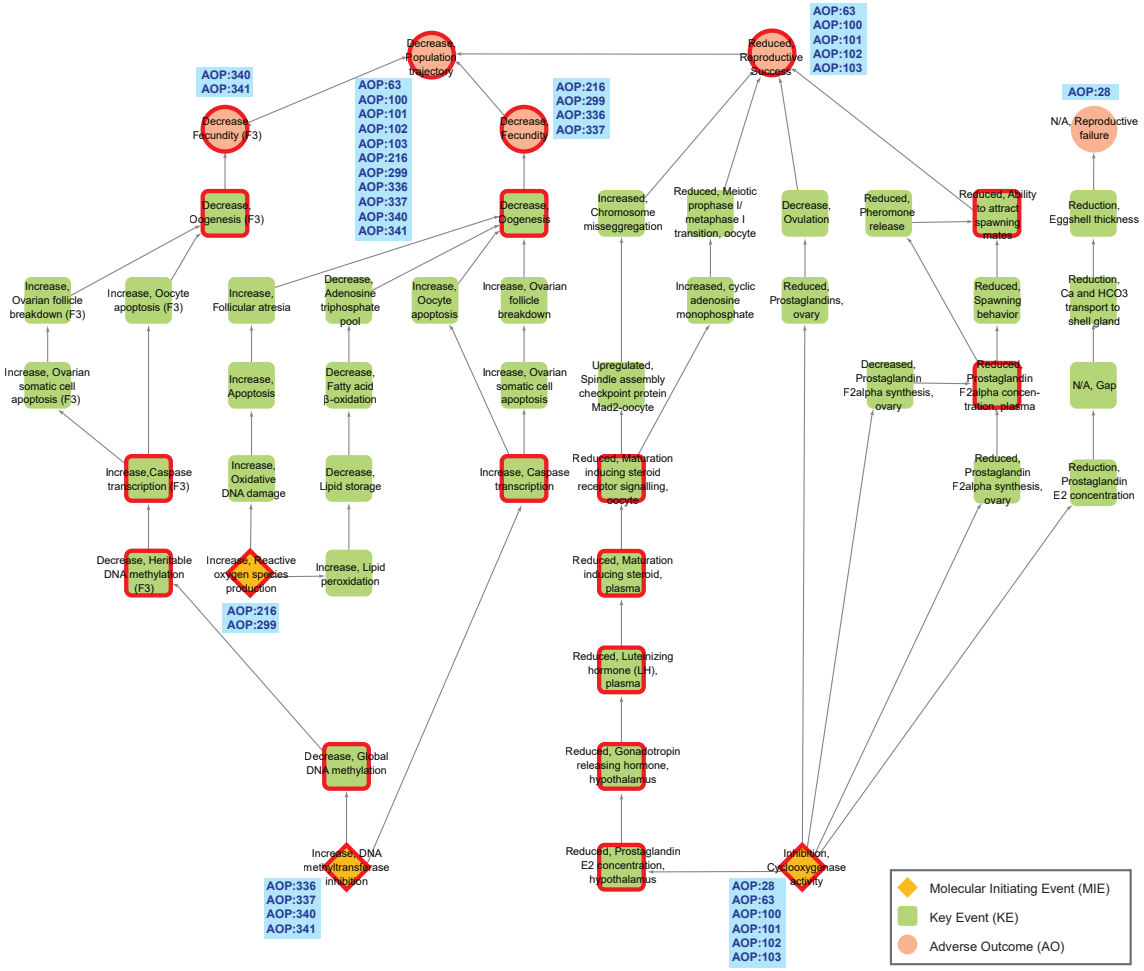


Figure 4.4: The directed network for LCC C2 in the ED-AOP network consisting of 48 KEs and 56 KERs. The 48 KEs in C2 can be categorized into 3 MIEs, 40 KEs and 5 AOs. MIEs, KEs and AOs are shown in distinct shapes namely, diamond, square and circle, respectively. The 20 shared KEs in C2 are marked in 'red'. For each MIE and AO, the corresponding AOP identifier is displayed in this figure.

a KER linking its upstream KE with its downstream KE. The directed network for C1 (Figure 4.3) has 44 KEs and 56 KERs while that for C2 (Figure 4.4) has 48 KEs and 56 KERs. Subsequently, we have studied four standard network measures namely, in-degree, out-degree, betweenness centrality and eccentricity, for KEs in the directed network corresponding to LCC, and these measures were computed using NetworkAnalyzer [240] in Cytoscape [241]. In the directed network, in-degree (respectively, out-degree) of a KE refers to the number of KEs immediately upstream (respectively, immediately downstream) of that KE [110]. Importantly, in-degree and out-degree of KEs can help identify points of convergence and divergence in the directed network. Further, betweenness centrality can help identify KEs crucial for the spread of biological perturbations, while eccentricity can help identify KEs which are farthest upstream or farthest downstream in the directed network [110, 230]. By applying network measures, we have studied the systems-level perturbations caused by endocrine-mediated events in the ED-AOP network upon chemical exposure. We have also investigated the ED-AOP network for possible emergence of new paths between pairs of MIE and AO that are both ED-KEs and belong to different ED-AOPs.

Firstly, we identified convergent and divergent events within the directed networks for C1 and C2 by assessing the in-degree and out-degree of each KE. A KE is considered to be 'convergent' if the in-degree is greater than (>) out-degree for the particular KE, while a KE is considered to be 'divergent' if the in-degree is less than (<) out-degree for the particular KE [110]. In C1, there are 13 convergent KEs and 12 divergent KEs. Among the 13 convergent KEs in C1, 2 KEs namely, 'Increase, cell proliferation (hepatocytes)' and 'Increase, hepatocellular adenomas and carcinomas', have the highest in-degree of 4. Among the 12 divergent KEs in C1, 2 KEs namely, 'Activation, PPARα' and 'Thyroxine (T4) in serum, Decreased', have the highest out-degree of 5, and in other words, these 2 divergent events lead to 5 other events in C1 (Figure 4.3; Supplementary Table S4.7). In C2, there are 6 convergent KEs and 7 divergent KEs. Among the 6 convergent KEs in C2, 2 KEs namely, 'Decrease, Oogenesis' and 'Reduced, Reproductive Success', have

the highest in-degree of 4. Among the 7 divergent KEs in C2, the KE 'Inhibition, Cyclooxygenase activity' has the highest out-degree of 5 (Figure 4.4; Supplementary Table S4.7).

Secondly, we have assessed the betweenness centrality of KEs in the directed networks for C1 and C2. The shared KE 'Reduction, Testosterone synthesis in Leydig cells' has the maximum betweenness centrality of 0.4 in C1 (Figure 4.5; Supplementary Table S4.7), while the shared KE 'Reduced, Maturation inducing steroid receptor signalling, oocyte' has the maximum betweenness centrality of 0.43 in C2 (Figure 4.6; Supplementary Table S4.7). Since these KEs with the highest betweenness centrality are on the shortest paths linking various nodes in C1 or C2, the events serve as significant control points in the ED-AOP network [242].

Thirdly, we have assessed the eccentricity of KEs in the directed networks for C1 and C2. The higher the eccentricity value for a node, the farther is the node located with respect to other nodes in the network, and thus, low eccentricity value for a node indicates its central location in the network [243]. In C1, the 2 shared KEs namely, 'Activation, PPARα' and 'Thyroperoxidase, Inhibition', have the maximum eccentricity value of 6 (Figure 4.7; Supplementary Table S4.7). In C2, the shared KE 'Reduced, Prostaglandin E2 concentration, hypothalamus' has the maximum eccentricity value of 8 (Figure 4.8; Supplementary Table S4.7).

Afterwards, we assessed the available information in AOP-Wiki for the two LCCs, C1 and C2. For C1, 21 out of the 44 KEs, i.e. nearly 50%, have evidence for human applicability in AOP-Wiki. For C2, however, 46 out of the 48 KEs do not have taxonomic applicability information in AOP-Wiki. Further, C2 contains two pairs of ED-AOPs namely, (i) AOP:336 and AOP:337, and (ii) AOP:340 and AOP:341, such that each pair of ED-AOPs contain the identical set of MIEs and AOs (Supplementary Table S4.6). Further, each pair of ED-AOPs is such that the two ED-AOPs have most of their KEs in common, and thus, it may be worthwhile to consider only one ED-AOP in each pair to avoid duplication of information in the ED-AOP network. Moreover, we find that AOP:28 of C2 contains KEs

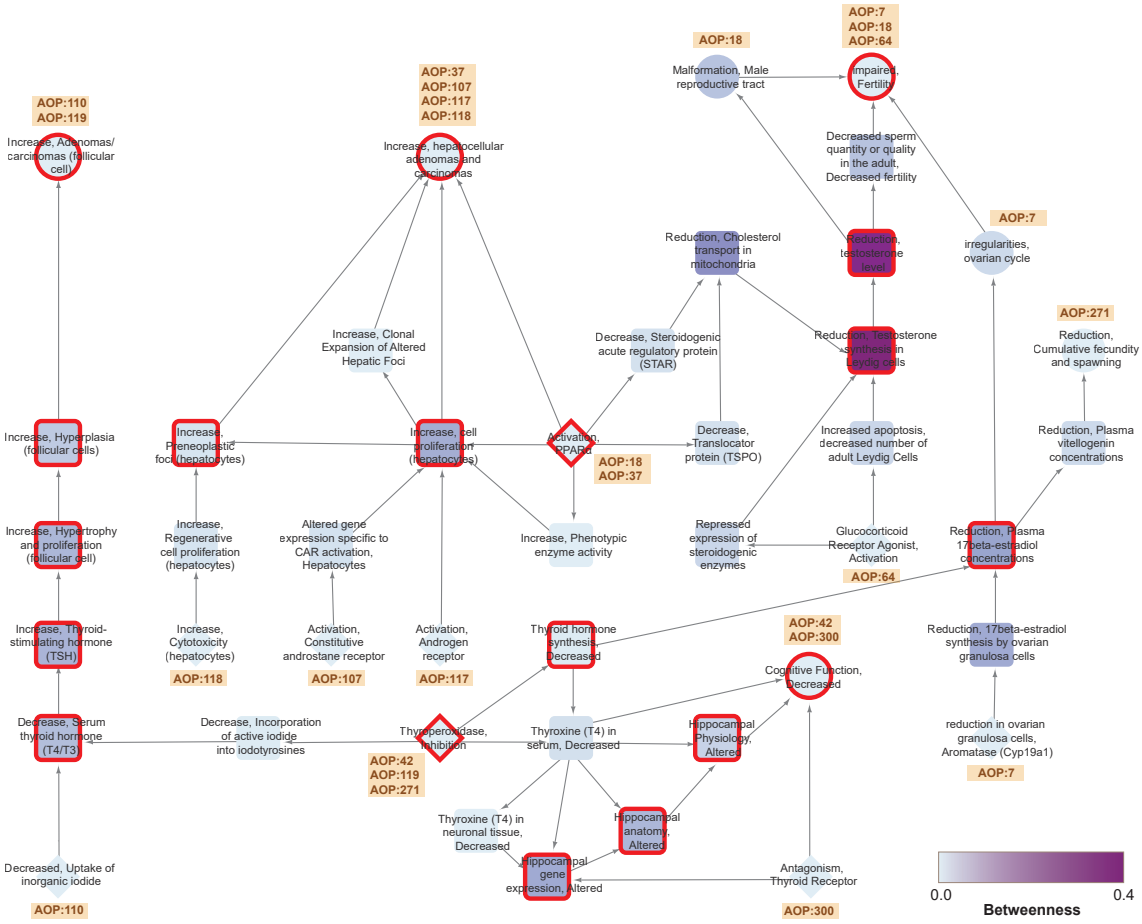


Figure 4.5: The directed network for LCC C1 wherein the KEs are colored based on their betweenness centrality values. MIEs, KEs and AOs are shown in distinct shapes namely, diamond, square and circle, respectively. The shared KEs are marked in 'red'. For each MIE and AO, the AOP identifier is displayed in this figure.

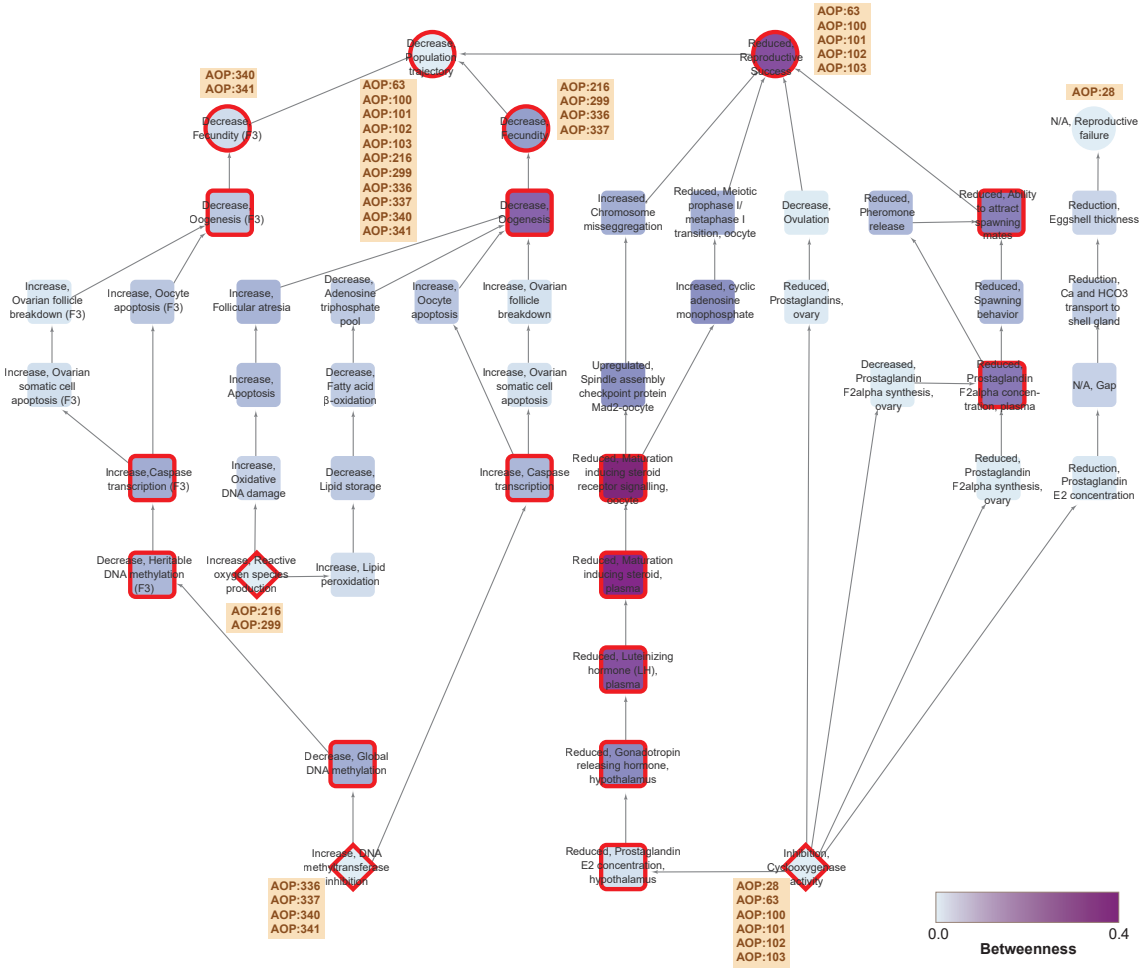


Figure 4.6: The directed network for LCC C2 wherein the KEs are colored based on their betweenness centrality values. MIEs, KEs and AOs are shown in distinct shapes namely, diamond, square and circle, respectively. The shared KEs are marked in 'red'. For each MIE and AO, the AOP identifier is displayed in this figure.

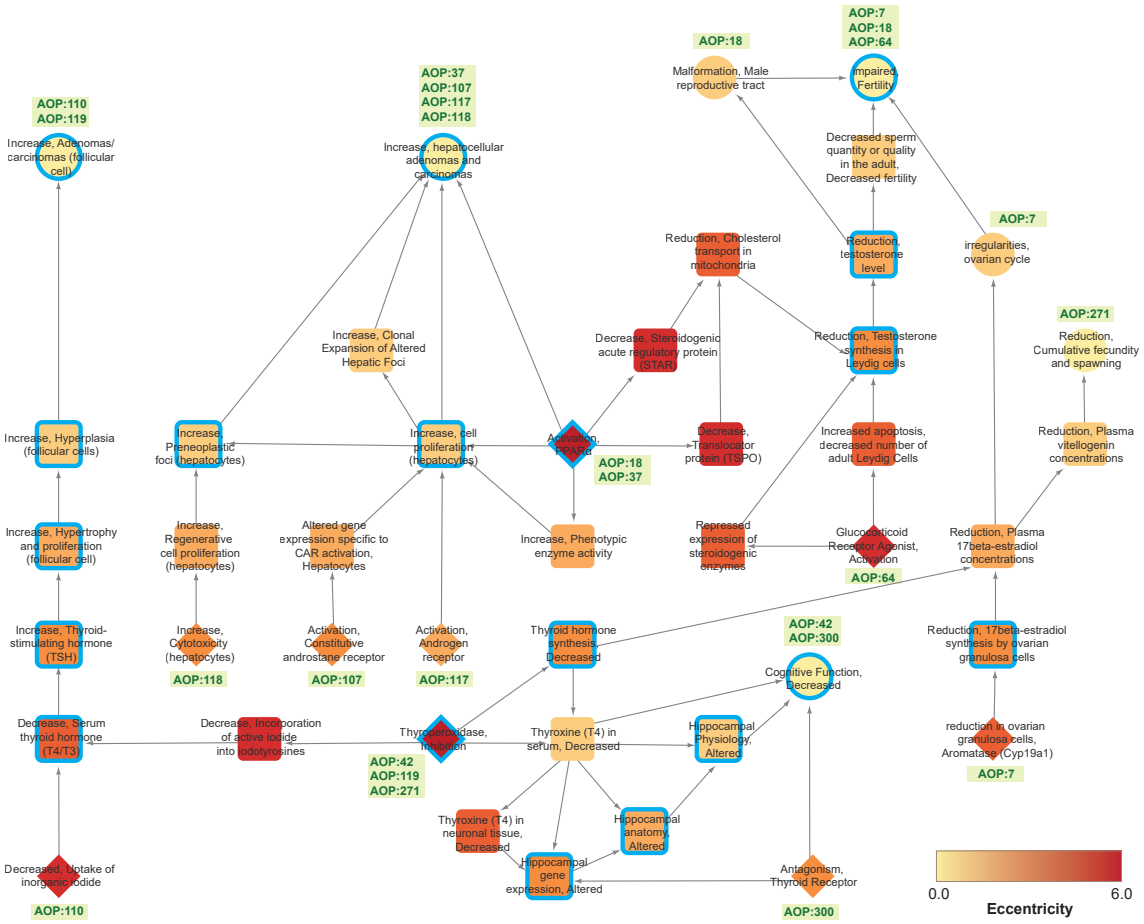


Figure 4.7: The directed network for LCC C1 wherein the KEs are colored based on their eccentricity values. MIEs, KEs and AOs are shown in distinct shapes namely, diamond, square and circle, respectively. The shared KEs are marked in 'red'. For each MIE and AO, the AOP identifier is displayed in this figure.

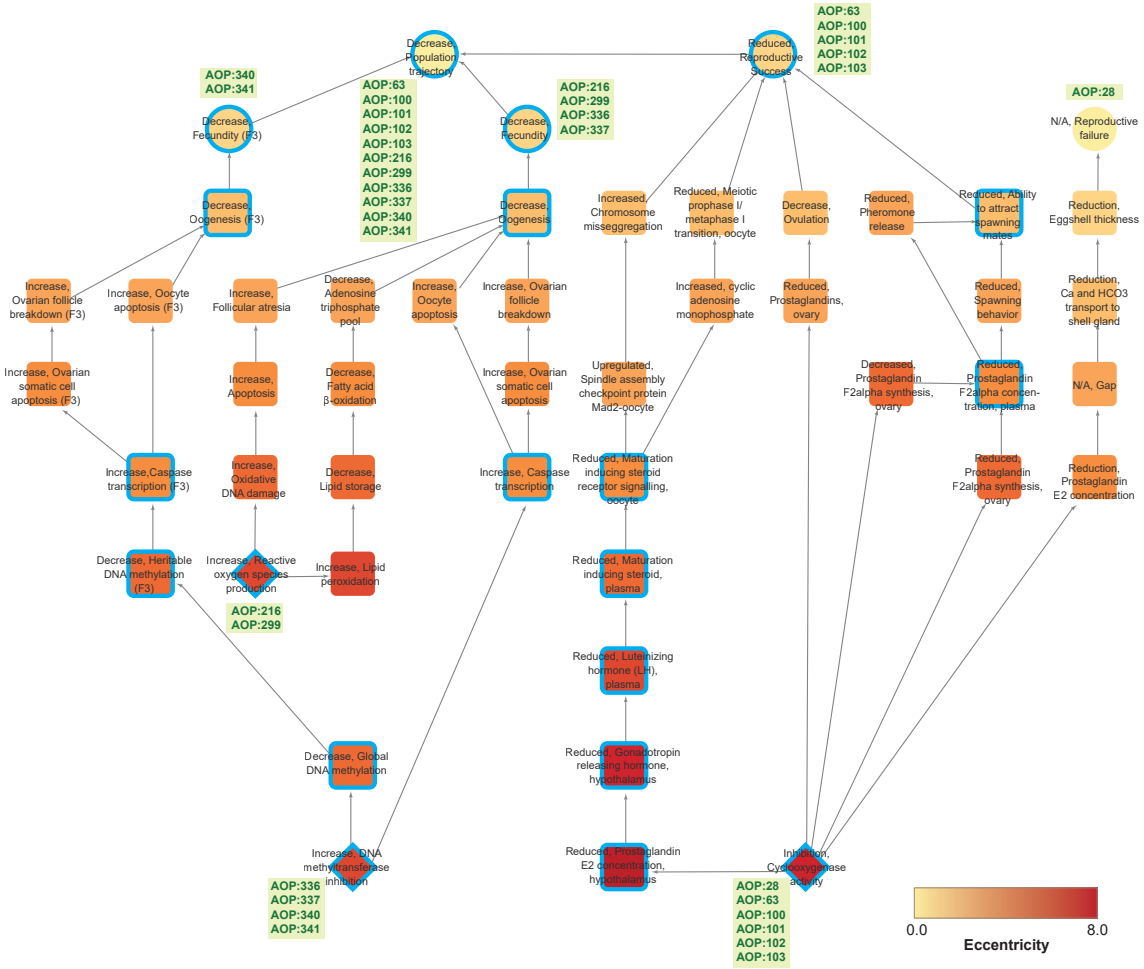


Figure 4.8: The directed network for LCC C2 wherein the KEs are colored based on their eccentricity values. MIEs, KEs and AOs are shown in distinct shapes namely, diamond, square and circle, respectively. The shared KEs are marked in 'red'. For each MIE and AO, the AOP identifier is displayed in this figure.

such as ‘N/A, Gap’ and ‘N/A, Reproductive failure’. Overall, this highlights disparity and gaps in available information across AOPs in AOP-Wiki. In sum, the available information is more comprehensive for the 12 ED-AOPs in C1 (in comparison to C2). As a result, the LCC C1 was further investigated to reveal the systems-level perturbations caused by endocrine-mediated events, the emergence of new paths linking MIEs and AOs, and the chemical stressors associated with KEs.

4.3 Systems-level perturbations caused by endocrine-mediated events in the largest component C1 of the ED-AOP network

Human exposure to EDCs can lead to endocrine disruption that in turn can affect various biological systems. Of late, there is concern regarding an increase in the incidence of endocrine-mediated disorders linked to reproduction, metabolism, development, nervous system and immunity in humans and wildlife [43, 50, 51, 244]. To better understand the systems-level perturbations upon EDC exposure, it is important to investigate the associated endocrine-mediated events leading to varied adverse outcomes. In this direction, we have investigated the systems-level perturbations caused by endocrine-mediated events captured in LCC C1 of the ED-AOP network.

In LCC C1, there are 44 KEs of which 9 are MIEs and 7 are AOs. Notably, 37 out of these 44 KEs (84%) in C1 were found to be ED-KEs. Depending on the perturbed cell types, organs or biological processes, we categorized the 44 KEs in C1 into 4 different systems-level endocrine-mediated perturbations, namely, ‘hepatic’, ‘metabolic’, ‘neurological’ and ‘reproductive’ (Figure 4.9; Supplementary Table S4.8). This categorization scheme for the 44 KEs in C1 is similar to the one used for AOs listed in Table 4.2. For example, the KE titled ‘Increase, Phenotypic enzyme activity’ in AOP-Wiki is associated with the cellular term ‘hepatocyte’, and thus, the KE is categorized as ‘hepatic’ in

our scheme (Figure 4.9; Supplementary Table S4.8). However, the information on the perturbed cell types, organs or biological processes is not available in AOP-Wiki for 3 MIEs in C1, namely, 'Antagonism, Thyroid Receptor', 'Activation, Androgen receptor', and 'Activation, Constitutive androstane receptor', and this prevented the categorization of these 3 MIEs into any of the 4 different systems-level perturbations (Figure 4.9; Supplementary Table S4.8). In addition, the OECD recommends generalizing some KEs in terms of their cell or tissue specificity so that they can be linked to different AOPs (OECD, 2018). Of the remaining 41 KEs in C1, 9, 10, 5, and 17 KEs were categorized as 'hepatic', 'metabolic', 'neurological' and 'reproductive' systems-level perturbations, respectively (Figure 4.9; Supplementary Table S4.8).

Thereafter, we analyzed the topology of LCC C1 by considering the categorization of the 41 KEs into 4 different systems-level perturbations (Figure 4.9). Specifically, we determined KERs in C1 that connect two KEs that differ in their categorization into systems-level perturbations. We find 8 such KERs in C1 of which 5 KERs connect KEs in metabolic and neurological systems, 2 KERs connect KEs in hepatic and reproductive systems, and 1 KER connects KEs in metabolic and reproductive systems (Figure 4.9). Among the KEs associated with these 8 KERs, 3 (divergent) KEs namely, 'Activation, PPARα', 'Thyroxine (T4) in serum, Decreased', and 'Thyroid hormone synthesis, Decreased', serve as points of divergence linking different systems in C1 (Figure 4.9). Specifically, the divergent KE titled 'Thyroxine (T4) in serum, Decreased' is categorized as 'metabolic' by our scheme, and this KE is immediately upstream of 5 KEs namely, 'Thyroxine (T4) in neuronal tissue, Decreased', 'Hippocampal gene expression, Altered', 'Hippocampal anatomy, Altered', 'Hippocampal Physiology, Altered', and 'Cognitive Function, Decreased', categorized as 'neurological' in C1 (Figure 4.9). In other words, this analysis of C1 reveals that the metabolic event 'Thyroxine (T4) in serum, Decreased' can lead to 5 neurological events, and interestingly, we were able to find independent supporting evidence for these particular associations between metabolic and neurological events in the published literature [245–248].

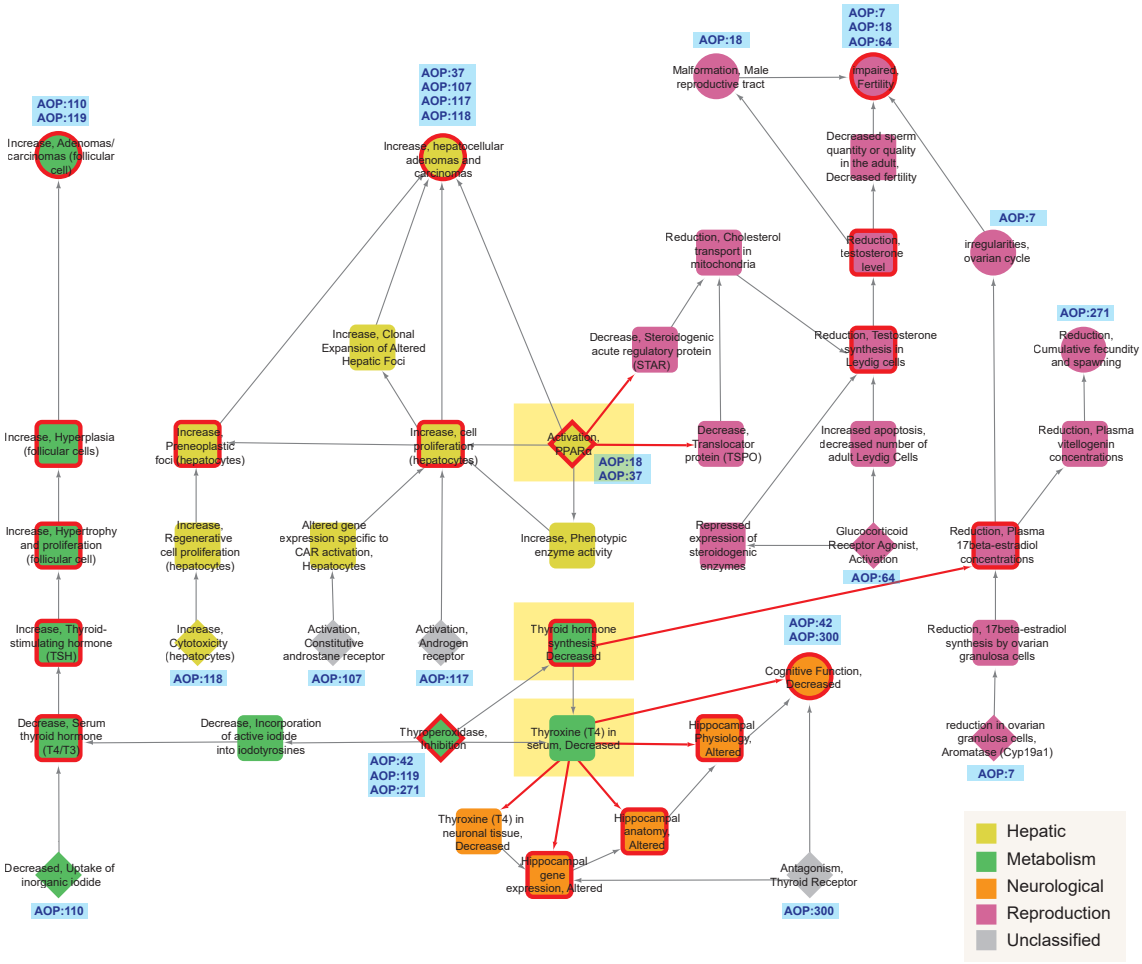


Figure 4.9: The directed network for LCC C1 in the ED-AOP network consisting of 44 KEs wherein the KEs are colored based on their categorization into 4 systems-level perturbations namely, hepatic, metabolic, neurological and reproductive. MIEs, KEs and AOs are shown in distinct shapes namely, diamond, square and circle, respectively. The 19 shared KEs in C1 are marked in 'red'. The 'red' edges highlight KERs that connect KEs categorized into different systems-level perturbations. The 'yellow rectangles' highlight 3 divergent KEs which serve as *point of divergence* from one system to another system.

Furthermore, the divergent KE titled 'Activation, PPARα' is categorized as 'hepatic', and this KE is immediately upstream of 2 KEs namely, 'Decrease, Steroidogenic acute regulatory protein (STAR)' and 'Decrease, Translocator protein (TSPO)', categorized as 'reproductive' in C1 (Figure 4.9), and there are supporting evidences for these particular associations between hepatic and reproductive events in the published literature [249–251]. Finally, the divergent KE titled 'Thyroid hormone production, Decreased' is categorized as 'metabolic', and this KE is immediately upstream of a KE titled 'Reduction, Plasma 17beta-estradiol concentrations' categorized as 'reproductive' in C1 (Figure 4.9), and there is supporting evidence for this particular association on the influence of thyroid levels on reproductive hormones [252]. Analysis of divergent KEs in the ED-AOP network can offer insights into links between different systems affected by endocrine disruption. Furthermore, these points of divergence tend to branch out into multiple downstream occurrences, reflecting a strong predictive utility and thus suggesting novel endpoints or assays that might be designed for better chemical risk assessment.

Lastly, we observed that 4 out of 12 ED-AOPs in C1 contain a shared AO titled 'Increase, hepatocellular adenomas and carcinomas' which is categorized as 'hepatic' systems-level perturbation. In C1, the shared KE titled 'Increase, cell proliferation (hepatocytes)' has maximum in-degree and is an important point of convergence leading to the above-mentioned AO. Further, this convergent KE is downstream of MIEs linked to activation of three hormonal receptors namely, Constitutive Androstane receptor (CAR), Androgen receptor (AR), and PPARα, highlighting the possibility of additive effects upon exposure to EDCs targeting multiple receptors (Figure 4.9). These convergent KEs reflect the points at which the effects of several stressors may converge, influencing downstream events, and so can serve as a framework for risk assessment of multiple stressors at the same time.

4.4 Emergent paths in the ED-AOP network

Since an AOP network contains multiple AOPs connected via shared KEs, new (directed) paths, other than those in individual AOPs, can emerge between MIEs and AOs belonging to different AOPs in the corresponding directed network of KEs and KERs. Such emergent paths from MIEs to AOs in an AOP network can also lead to the development of new stand-alone AOPs [110]. Here, we have investigated the possibility of such emergent paths between MIEs and AOs in the LCC C1 of the ED-AOP network consisting of 12 ED-AOPs. We have found 4 new paths in the LCC C1 that connect an endocrine-relevant MIE in one ED-AOP to an endocrine-relevant AO in another ED-AOP (Figure 4.3; Table 4.3).

Of the 4 new paths in C1 (Figure 4.3; Table 4.3), 2 new paths start from the shared MIE 'Thyroperoxidase, Inhibition' (in AOP:42, AOP:119, and AOP:271) and end at the 2 AOs namely, 'irregularities, ovarian cycle' (in AOP:7) and 'impaired, Fertility' (in AOP:7, AOP:18, and AOP:64). We find that previously published research supports the above links indicating the impact of thyroperoxidase on reproduction [253–256]. Another new path in C1 starts from the MIE 'reduction in ovarian granulosa cells, Aromatase (Cyp19a1)' (in AOP:7) and ends at the AO 'Reduction, Cumulative fecundity and spawning' (in AOP:271). The AO 'Reduction, Cumulative fecundity, and spawning' in this path describes the process of releasing eggs or sperms for aquatic animals like fishes. On the other hand, Aromatase appears to play a substantial role in egg release in both humans [257, 258] and fishes [259, 260] based on previous studies. Lastly, there is a new path in C1 starting from the MIE 'Glucocorticoid Receptor Agonist, Activation' (in AOP:64) and ending at the AO 'Malformation, Male reproductive tract' (in AOP:18). Previous research has shown that the glucocorticoid receptor has an effect on male reproduction [261, 262]. These emergent paths identified in LCC C1 of the ED-AOP network have potential to reveal unknown relationships between distant KEs and may represent toxicity pathways specific to endocrine disruption. Further, a closer inspection of these

emergent paths may also lead to prediction of unknown adverse effects upon specific EDC exposure, as well as guide future development of new AOPs.

4.5 Chemical stressors and the ED-AOP network

AOPs are known to be induced by one or multiple stressors. Linking chemical stressors to the biological events in an AOP network can reveal the possible adverse outcomes upon exposure, thereby facilitating regulatory decision-making or risk assessment. The stressors with incomplete information in AOP-Wiki were manually assigned to their structural identifiers including CAS, DSSTOX and InChIKey. Thus, we have analyzed the information on chemical stressors associated with KEs in LCC C1 from AOP-Wiki. Based on information in AOP-Wiki, 35 chemical stressors were found to be associated with different KEs in C1. By performing a comparative analysis of these 35 chemical stressors with the list of 792 potential EDCs in DEDuCT 2.0 [35, 36], we identified a subset of 16 chemical stressors associated with C1 that have strong supporting evidence of endocrine disruption (Supplementary Table S4.9). These 16 EDCs directly target at least one event among 5 MIEs, 9 KEs and 2 AOs in LCC C1. Among these 5 MIEs, MIE 'Thyroperoxidase, Inhibition' is directly linked to 7 EDCs, and MIE 'Activation, PPARα' is directly linked to 4 EDCs. Among the 16 EDCs, we find that the EDC '6-Propyl-2-thiouracil' directly targets 8 events in C1 (Supplementary Table S4.9).

Analyses of direct associations between chemical stressors and KEs in the ED-AOP network can reveal the diversity of biological mechanisms via which EDCs can cause different endocrine-mediated adverse effects. To aid ongoing efforts in risk assessment of EDCs, it will be worthwhile to undertake a future effort to associate all known EDCs, including 792 potential EDCs in DEDuCT 2.0, to different events in the ED-AOP network. In sum, a stressor-ED-AOP network can serve as a predictive model for EDCs and their adverse effects.

4.6 Discussion

An AOP is a systematic framework to encapsulate the existing toxicological information as a toxicity pathway to aid in risk assessment and chemical regulation [96, 97, 99, 104, 223]. Within AOP-Wiki, the up to date central repository of individual AOPs, AOP networks have emerged due to sharing of KEs and KERs across individual AOPs. Since AOP networks are expected to be the functional units for prediction in real-world scenarios, there is notable interest in the derivation and analysis of AOP networks tailored to address specific problems or applications [107, 110, 228].

The challenges in the risk assessment and regulation of EDCs partially stem from the existing knowledge gaps in linking chemical exposure to diverse adverse outcomes [3, 218, 244]. To address this challenge, a blueprint of the endocrine disruption mechanisms in the form of toxicity pathways spanning different levels of biological organization can be invaluable [234]. In this context, the development of a comprehensive AOP network relevant to endocrine disruption (i.e., an ED-AOP network) can aid ongoing research and policy framing surrounding EDCs. In this chapter, we have developed a detailed workflow (Figure 4.1) to leverage information in AOP-Wiki and construct a comprehensive ED-AOP network (Figure 4.2; Table 4.1). Ensuing graph-theoretic analysis of this ED-AOP network of 48 ED-AOPs, and in particular, its largest components C1 and C2 of 12 ED-AOPs each, reveals several mechanistic insights on endocrine-mediated perturbations upon chemical exposure.

Since AOP development is a continuous and iterative exercise, therefore the ED-AOP network constructed in this chapter is appreciably limited by the existing knowledge in AOP-Wiki. As AOPs are living documents, it will be important to maintain the ED-AOP network up to date with any expansion in AOP-Wiki. This could have an impact on the graph-theoretic analysis reflecting the bias of the existing data. For example, the key events with the highest betweenness value could reflect important control points in the ED-AOP network, as well as the most frequently investigated occurrences rather than a

biological reality. Another significant limitation is the choice of criteria for filtration of ED-KEs, where we used endocrine-relevant keywords such as glands, hormones, hormonal receptors, endocrine disorders, and endpoints specific to humans or rodents. As a result, the majority of ED-AOPs used to construct the ED-AOP network may be confined to these organisms. We expect that the detailed workflow in Figure 4.1 with a little or no modification can be used for any future update of the ED-AOP network. Moreover, the current information in AOP-Wiki on chemical stressors associated with events in the ED-AOP network is a small fraction of the existing knowledge on potential EDCs in the published literature [35, 36], and therefore, it will be important to invest future efforts towards developing a comprehensive stressor-ED-AOP network wherein all known EDCs are linked to different events in the ED-AOP network. In sum, ED-AOP network provides an overall landscape of potential adverse outcomes associated with EDC exposure, allowing for the identification of important biological events that are relevant for better risk assessment.

Supplementary Information

Supplementary Tables S4.1-S4.9 associated with this chapter are available for download from the GitHub repository: https://github.com/asamallab/PhDThesis-Janani_R/blob/main/SI/ST_Chapter4.xlsx.

S. No.	AOP identifier	AOP title	Fraction of ED-KEs	Cumulative WoE	Human WoE
1	6	Antagonist binding to PPARα leading to body-weight loss	62.5	High	High
2	7	Aromatase (Cyp19a1) reduction leading to impaired fertility in adult female	100	High	Low
3	12	Chronic binding of antagonist to N-methyl-D-aspartate receptors (NMDARs) during brain development leads to neurodegeneration with impairment in learning and memory in aging	87.5	Moderate	Low
4	13	Chronic binding of antagonist to N-methyl-D-aspartate receptors (NMDARs) during brain development induces impairment of learning and memory abilities	90	Low	High
5	18	PPARα activation in utero leading to impaired fertility in males	87.5	Moderate	Low
6	19	Androgen receptor antagonism leading to adverse effects in the male foetus (mammals)	100	-	-
7	28	Cyclooxygenase inhibition leading reproductive failure	66.7	Moderate	-
8	36	Peroxisomal Fatty Acid Beta-Oxidation Inhibition Leading to Steatosis	75	High	-
9	37	PPARα activation leading to hepatocellular adenomas and carcinomas in rodents	80	High	-
10	41	Sustained AhR Activation leading to Rodent Liver Tumours	100	High	-
11	42	Inhibition of Thyroperoxidase and Subsequent Adverse Neurodevelopmental Outcomes in Mammals	62.5	High	High
12	43	Disruption of VEGFR Signaling Leading to Developmental Defects	60	High	Moderate
13	60	NR1I2 (Pregnane X Receptor, PXR) activation leading to hepatic steatosis	58.3	High	-
14	62	AKT2 activation leading to hepatic steatosis	100	-	-
15	63	Cyclooxygenase inhibition leading to reproductive dysfunction	80	Moderate	Low
16	64	Glucocorticoid Receptor (GR) Mediated Adult Leydig Cell Dysfunction Leading to Decreased Male Fertility	100	-	-

17	100	Cyclooxygenase inhibition leading to reproductive dysfunction via inhibition of female spawning behavior	57.1	Moderate	-
18	101	Cyclooxygenase inhibition leading to reproductive dysfunction via inhibition of pheromone release	57.1	High	-
19	102	Cyclooxygenase inhibition leading to reproductive dysfunction via interference with meiotic prophase I /metaphase I transition	80	High	Low
20	103	Cyclooxygenase inhibition leading to reproductive dysfunction via interference with spindle assembly checkpoint	80	High	Low
21	107	Constitutive androstane receptor activation leading to hepatocellular adenomas and carcinomas in the mouse and the rat	80	High	-
22	110	Inhibition of iodide pump activity leading to follicular cell adenomas and carcinomas (in rat and mouse)	100	-	-
23	111	Decrease in androgen receptor activity leading to Leydig cell tumors (in rat)	100	-	-
24	112	Increased dopaminergic activity leading to endometrial adenocarcinomas (in Wistar rat)	100	-	-
25	117	Androgen receptor activation leading to hepatocellular adenomas and carcinomas (in mouse and rat)	75	-	-
26	118	Chronic cytotoxicity leading to hepatocellular adenomas and carcinomas (in mouse and rat)	75	-	-
27	119	Inhibition of thyroid peroxidase leading to follicular cell adenomas and carcinomas (in rat and mouse)	100	-	-
28	120	Inhibition of 5α-reductase leading to Leydig cell tumors (in rat)	100	-	-
29	124	HMG-CoA reductase inhibition leading to decreased fertility	83.3	-	-
30	164	Beta-2 adrenergic agonist activity leading to mesovarian leiomyomas in the rat and mouse	66.7	High	-
31	167	Early-life estrogen receptor activity leading to endometrial carcinoma in the mouse.	71.4	High	-
32	201	Juvenile hormone receptor agonism leading to male offspring induction associated population decline	50	-	-

33	203	5-hydroxytryptamine transporter inhibition leading to decreased reproductive success and population decline	37.5	-	-
34	204	5-hydroxytryptamine transporter inhibition leading to increased reproductive success and population increase	37.5	-	-
35	205	AOP from chemical insult to cell death	50	High	-
36	212	Histone deacetylase inhibition leading to testicular atrophy	66.7	Moderate	Moderate
37	216	Excessive reactive oxygen species production leading to population decline via follicular atresia	85.7	-	-
38	220	Cyp2E1 Activation Leading to Liver Cancer	80	High	Moderate
39	232	NFE2/Nrf2 repression to steatosis	87.5	-	-
40	271	Inhibition of thyroid peroxidase leading to impaired fertility in fish	80	High	-
41	293	Increased DNA damage leading to increased risk of breast cancer	66.7	Moderate	-
42	299	Excessive reactive oxygen species production leading to population decline via reduced fatty acid beta-oxidation	62.5	-	-
43	300	Thyroid Receptor Antagonism and Subsequent Adverse Neurodevelopmental Outcomes in Mammals	40	Moderate	High
44	306	Androgen receptor (AR) antagonism leading to short anogenital distance (AGD) in male (mammalian) offspring	100	High	Moderate
45	336	DNA methyltransferase inhibition leading to population decline (1)	57.1	Moderate	-
46	337	DNA methyltransferase inhibition leading to population decline (2)	62.5	Moderate	-
47	340	DNA methyltransferase inhibition leading to transgenerational effects (1)	62.5	Moderate	-
48	341	DNA methyltransferase inhibition leading to transgenerational effects (2)	66.7	Moderate	-

Table 4.1: The curated subset of 48 ED-AOPs among the 161 high-confidence AOPs filtered from AOP-Wiki. The table also gives the fraction of ED-KEs, the cumulative WoE score, and the WoE score for human applicability (Human WoE) for each of the 48 ED-AOPs.

S. No.	Component identifier	AO	Systems-level perturbation
1	C1	Increase, hepatocellular adenomas and carcinomas	Hepatic
2	C1	Increase, Adenomas/carcinomas (follicular cell)	Metabolic
3	C1	Cognitive Function, Decreased	Neurological
4	C1	impaired, Fertility	Reproductive
5	C1	irregularities, ovarian cycle	Reproductive
6	C1	Reduction, Cumulative fecundity and spawning	Reproductive
7	C1	Malformation, Male reproductive tract	Reproductive
8	C2	Decrease, Population trajectory	Reproductive
9	C2	Decrease, Fecundity	Reproductive
10	C2	Decrease, Fecundity (F3)	Reproductive
11	C2	N/A, Reproductive failure	Reproductive
12	C3	Increased, Liver Steatosis	Hepatic
13	C4	Increased, Male offspring	Reproductive
14	C4	Decreased, Reproductive Success	Reproductive
15	C4	Increased, Reproductive Success	Reproductive
16	C5	Impairment, Learning and memory	Neurological
17	C6	Increase, Leydig cell tumors	Reproductive
18	C7	Apoptosis	-
19	C7	Testicular atrophy	Reproductive

Table 4.2: The list of AOs in the 7 connected components of the ED-AOP network and their categorization into 4 systems-level endocrine-mediated perturbations, namely, 'hepatic', 'metabolic', 'neurological' and 'reproductive', depending on the perturbed biological processes.

S. No.	MIE	AO
1	Thyroperoxidase, Inhibition	irregularities, ovarian cycle
2	Thyroperoxidase, Inhibition	impaired, Fertility
3	reduction in ovarian granulosa cells, Aromatase (Cyp19a1)	Reduction, Cumulative fecundity and spawning
4	Glucocorticoid Receptor Agonist, Activation	Malformation, Male reproductive tract

Table 4.3: The table gives information on the starting MIE and the ending AO for each of the 4 new paths identified in the LCC C1 of the ED-AOP network.

Chapter 5

NeurotoxKb 1.0: compilation, curation and exploration of a knowledgebase of environmental neurotoxicants specific to mammals

Exposures to environmental neurotoxicants are of significant concern as they can cause permanent or irreversible damage to the nervous system [55, 56]. In the last few decades, several studies have documented the neurotoxic effects of heavy metals such as arsenic, lead, manganese and mercury, and other groups of environmental chemicals such as Polychlorinated biphenyls (PCBs), Perfluoroalkylated substances (PFAS) and Organotins [52, 53, 57, 58, 263]. In comparison to chemicals tested for neurotoxicity so far, the space of chemicals in commerce is huge. Specifically, there are over 100000 chemicals in commerce in the EU and USA, and only a tiny fraction of them have been tested for neurotoxicity to date [58, 59]. Some reasons for this gap in current knowledge on environmental neurotoxicants include the lack of systematic testing methods for neurotoxicity and the inherent complexity of neurotoxicological assessments [58, 263, 264].

Despite these limitations, there have been some efforts to compile potential neurotoxicants with evidence specific to mammals from published literature [57, 58, 60–62]. Although the lists of potential neurotoxicants compiled by Grandjean and Landrigan [58], Mundy *et al*. [61], and Aschner *et al*. [62] are available via the CompTox dashboard [265], there is no dedicated online resource to date on environmental neurotoxicants. In this chapter, we address this unmet need by building the first dedicated online knowledgebase, namely, NeurotoxKb 1.0 [38], which compiles 475 potential non-biogenic neurotoxicants with published evidence specific to mammals. **The work reported in this chapter is contained in the published manuscript [38].**

5.1 Building a knowledgebase of environmental neurotoxicants specific to mammals

We started building the curated knowledgebase on environmental neurotoxicants, namely NeurotoxKb 1.0 [38], with experimental evidence on neurotoxicity specific to mammals, by compiling potential neurotoxicants from four existing resources [57,58,60–62] in published literature as described in the following steps (Figure 5.1).

5.1.1 Compilation and filtration of potential non-biogenic neurotoxicants from existing resources

Firstly, we considered 802 potential neurotoxicants compiled in the US EPA report [60] published in 1976 on neurotoxic chemicals. From published literature, the US EPA report had compiled 802 chemicals tested for neurotoxic effects upon exposure on various living organisms including mammals and non-mammals [60]. Secondly, we have considered 214 potential neurotoxicants compiled by Grandjean and Landrigan [57, 58] to which humans are vulnerable upon exposure in early stages of development. For compiling their list, Grandjean and Landrigan [57, 58] had employed PubMed literature mining and toxicological resources such as TOXNET [266,267], TOXLINE [268] and Hazardous

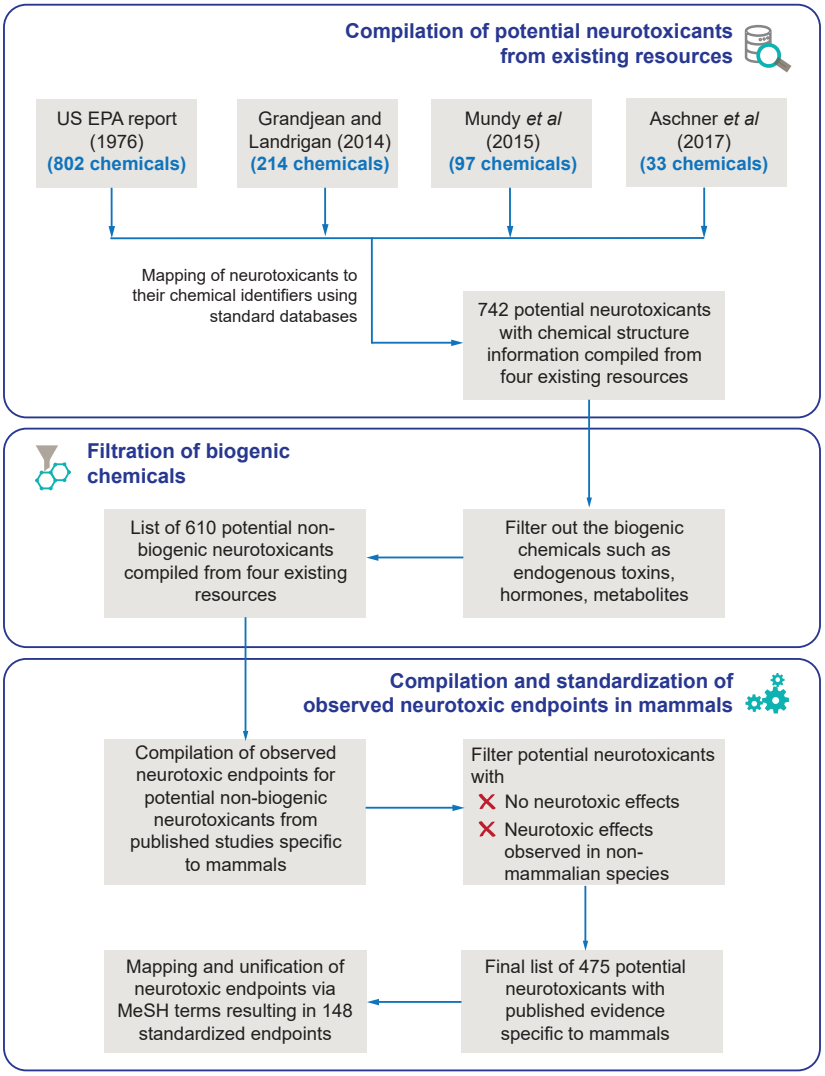


Figure 5.1: Schematic workflow describing the compilation of 475 potential non-biogenic neurotoxicants along with published evidence of observed neurotoxic endpoints specific to mammals.

Substances Data Bank (HSDB) [269]. Note that these toxicological resources have been integrated into other NLM databases since 2019 [267]. Note that Grandjean and Landrigan had first published a list of 201 potential human neurotoxicants in 2006 [58] which they subsequently expanded to 214 potential human neurotoxicants in 2014 [57]. Thirdly, we have considered the 97 potential neurotoxicants compiled by Mundy *et al.* [61, 270] that have demonstrated effects on neurodevelopment. Fourthly, we have considered the 33 potential neurotoxicants compiled by Aschner *et al.* [62, 271] that have evidence of triggering developmental neurotoxicity *in vivo*.

We remark that three of the above-mentioned four lists of potential neurotoxicants considered here (Figure 5.1), namely Grandjean and Landrigan [58], Mundy *et al.* [61] and Aschner *et al.* [62], are among the six lists of potential neurotoxicants captured by the CompTox chemistry dashboard [265]. Since our aim is to compile potential neurotoxicants specific to mammals, we have not considered three other lists of potential neurotoxicants captured by the CompTox chemistry dashboard. Specifically, we have not considered the list 'DNT Screening Library' [272] that compiles potential neurotoxicants with experimental evidence specific to Zebrafish. Similarly, we have not considered the two lists, namely 'Neurotoxicants from PubMed' [273] and 'NEURO: Neurotoxicants Collection from Public Resources' [274], as both lists gather potential neurotoxicants from literature without compiling information on the test organisms for neurotoxicity.

Next, we mapped the 802, 214, 97 and 33 potential neurotoxicants compiled from the US EPA report [60], Grandjean and Landrigan [57], Mundy *et al.* [61] and Aschner *et al.* [62], respectively, to chemical identifiers in standard databases such as PubChem [86], CAS [164] and CTD [30]. While mapping the potential neurotoxicants to their chemical structure, we have removed any potential neurotoxicant in the four lists that could not be mapped to a chemical identifier or represents a chemical mixture rather than individual chemical entity. This resulted in a non-redundant list of 742 potential neurotoxicants compiled from the four above-mentioned resources (Figure 5.1).

Next, we have removed any chemical from the non-redundant list of 742 potential

neurotoxicants compiled from the four above-mentioned resources that are of biological origin such as snake venoms, plant or microbial toxins, and hormones. This removal of potential biogenic neurotoxins is motivated by our exclusive focus on human-made environmental neurotoxicants. This resulted in a list of 610 potential non-biogenic neurotoxicants compiled from the four above-mentioned resources (Figure 5.1).

In summary, we have compiled from four existing resources, a curated list of 610 potential non-biogenic neurotoxicants along with their two-dimensional (2D) and three-dimensional (3D) chemical structure information via the above-mentioned steps in our workflow (Figure 5.1).

5.1.2 Compilation and standardization of observed neurotoxic endpoints for environmental neurotoxicants specific to mammals

In order to develop a comprehensive resource on environmental neurotoxicants, it is necessary to compile the observed neurotoxic endpoints (or adverse effects) upon exposure to neurotoxicants from the published literature. Although the four existing resources [57, 60–62] on potential neurotoxicants considered here compile observed neurotoxic endpoints upon chemical exposure, a lack of standardization in reporting of the adverse effects across the resources limit their utility for toxicological risk assessment. To address this unmet need and enable future research in neurotoxicity, we next compiled and manually curated the observed neurotoxic endpoints for the 610 potential non-biogenic neurotoxicants identified via the above-mentioned steps in our workflow (Figure 5.1).

Firstly, we have compiled from the USA EPA report [60], the observed neurotoxic endpoints for potential non-biogenic neurotoxicants along with the information on test organisms including mammals and non-mammals in the published experimental studies. Note that the USA EPA report [60] also compiles observations of no neurotoxic effects for potential neurotoxicants from published experimental studies.

Secondly, Mundy *et al*. [61] and Aschner *et al*. [62] have compiled potential develop-

mental neurotoxicants along with the information on their observed neurotoxic endpoints from published experimental studies in rodents and primates. However, the compilation of neurotoxic endpoints in Mundy *et al*. [61] and Aschner *et al*. [62] is much less detailed in comparison to the USA EPA report [60]. Specifically, Mundy *et al*. [61] have reported the neurotoxic endpoints from published studies after their broad categorization into 3 terms, namely, behaviour, morphology, and neurochemistry. Similarly, Aschner *et al*. [62] have reported the neurotoxic endpoints from published studies after their broad categorization into 40 terms. However, we believe that a detailed compilation of neurotoxic endpoints for potential neurotoxicants from published studies specific to mammals can render a valuable toxicological resource that can aid in early identification and regulation of hazardous chemicals. Therefore, we have performed a manual curation of the 287 published studies compiled by Mundy *et al*. [61] and Aschner *et al.* [62] to collect detailed neurotoxic endpoints for potential non-biogenic neurotoxicants covered by the two resources.

Thirdly, Grandjean and Landrigan [57, 58] have compiled a list of chemicals potentially toxic to the human nervous system from published literature. However, Grandjean and Landrigan [57, 58] have not compiled the observed neurotoxic endpoints for the potential neurotoxicants from associated published literature. Therefore, we have performed an extensive manual curation effort to compile the observed neurotoxic effects specific to humans from HSDB [269] for the potential neurotoxicants in the list by Grandjean and Landrigan [57,58]. Note that HSDB [269] (which has been integrated into PubChem [86]) was used by Grandjean and Landrigan [57, 58] to compile their list of 214 potential human neurotoxicants. During this manual curation effort, we were unable to gather experimental evidence specific to mammals from HSDB [269] for some of the 214 potential human neurotoxicants in the list by Grandjean and Landrigan [57, 58]. For such potential neurotoxicants in the list by Grandjean and Landrigan [57, 58] without any documented evidence of neurotoxicity in HSDB [269], we performed additional literature searches to gather any published evidence of neurotoxicity specific to mammals.

At the end of the above-mentioned steps to compile observed neurotoxic endpoints specific to mammals for 610 potential non-biogenic neurotoxicants from existing resources [57, 60–62], HSDB [269] and published literature, we were able to gather published experimental evidence specific to mammals for only 475 out of 610 potential non-biogenic neurotoxicants (Figure 5.1; Supplementary Table S5.1). These 475 potential non-biogenic neurotoxicants with experimental evidence specific to mammals from 835 published articles have been compiled in our environmental Neurotoxicants Knowledgebase, namely NeurotoxKb 1.0 [38], which is accessible at: http://cb.imsc.res.in/neurotoxkb.

Finally, we undertook an extensive manual curation effort to standardize the compiled information on detailed neurotoxic effects observed in 835 published studies specific to mammals for the 475 potential non-biogenic neurotoxicants in NeurotoxKb 1.0. For the unification and standardization of this compiled information on neurotoxic effects of the 475 potential neurotoxicants, we have leveraged Medical Subject Headings (MeSH) terms [237, 275]. For example, the observed neurotoxic effect, 'Lack of coordination', was mapped to the MeSH term 'Ataxia' and its corresponding MeSH identifier D001259. Through this exercise, we were able to map, unify and standardize a compiled list of 900 terms referring to observed neurotoxic effects from 835 published studies on 475 potential neurotoxicants to 148 standardized neurotoxic endpoints based on MeSH terms (Figure 5.1; Supplementary Table S5.2).

Of these 475 identified potential neurotoxicants in NeurotoxKb 1.0 [38], the US EPA report [60], Grandjean and Landrigan [57], Mundy *et al.* [61] and Aschner *et al.* [62] capture 292, 178, 88 and 26 potential neurotoxicants, respectively, with published evidence specific to mammals (Figure 5.2A). Notably, among the four existing resources, the US EPA report [60] contributes a unique set of 231 out of the 475 potential neurotoxicants (~ 50%) compiled in NeurotoxKb 1.0 with published evidence specific to mammals (Figure 5.2A). In other words, almost 50% of the potential neurotoxicants specific to mammals in NeurotoxKb 1.0 were solely identified due to our extensive manual effort

to digitize, compile, curate and organize the vast information on potential neurotoxicants captured in the US EPA report [60] published in 1976. Notably, the US EPA report [60] contributes a unique set of 414 out of the 835 published articles ($\sim$ 50%) compiled in NeurotoxKb 1.0 that provide mammalian-specific evidence on potential neurotoxicants.

5.1.3 Classification of neurotoxicants

Based on environmental source

Information on the major sources of exposure is vital for chemical regulation and monitoring by agencies. Therefore, we have compiled the environmental sources for the 475 potential neurotoxicants in NeurotoxKb 1.0. Specifically, NeurotoxKb 1.0 has classified the 475 potential neurotoxicants into 6 broad categories of environmental sources, namely, 'Agriculture and Farming', 'Consumer Products', 'Industry', 'Intermediates', 'Medicine and Healthcare', and 'Pollutant', and 41 sub-categories (Figure 5.3). It can be seen that majority of the 475 potential neurotoxicants are in the category 'Agriculture and Farming' which is followed by 'Industry' (Figure 5.3) [38].

Based on chemical structure

Furthermore, we have also classified the 475 potential neurotoxicants in NeurotoxKb 1.0 based on their chemical structure. Specifically, we have employed ClassyFire [173, 174] for a hierarchical chemical classification into kingdom, super-class, class and subclass. Of the 475 potential neurotoxicants, 430 are organic while 45 are inorganic (Figure 5.2B). Moreover, majority (100) of the 475 potential neurotoxicants belong to chemical super-class 'Benzenoids' (Figure 5.2B) [38]. Note that information on the chemical class of potential neurotoxicants can be used to draw inferences on their nature and behaviour.

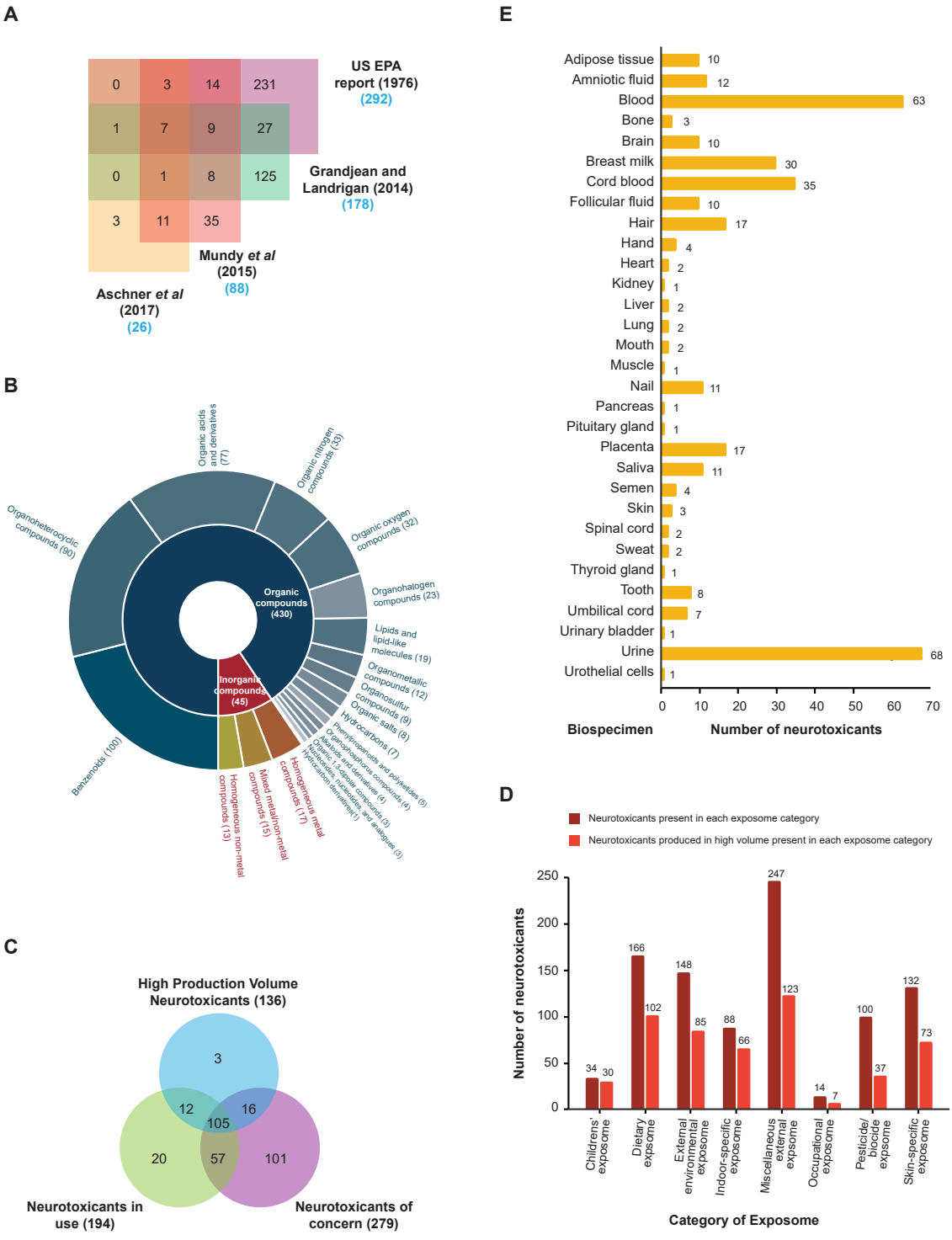

A
0
3
14
231
1
7
9
27
0
1
8
125
3
11
35
US EPA report (1976) (292)
Grandjean and Landrigan (2014) (178)
Mundy et al (2015) (88)
Aschner et al (2017) (26)
B
Organic compounds (430)
Inorganic compounds (45)
Organic acids and derivatives (77)
Organic nitrogen compounds (35)
Organic oxygen compounds (32)
Organohalogen compounds (23)
Lipids and lipid-like molecules (19)
Organometallic compounds (12)
Organosulfur compounds (9)
Organic salts (8)
Hydrocarbons (7)
Phenylpropanoids and polyketides (5)
Organoheterocyclic compounds (90)
Benzenoids (100)
Homogeneous non-metal compounds (13)
Mixed metal/non-metal compounds (15)
Homogeneous metal compounds (17)
C
High Production Volume Neurotoxicants (136)
3
12
16
105
20
57
101
Neurotoxicants in use (194)
Neurotoxicants of concern (279)
D
Neurotoxicants present in each exposome category
Neurotoxicants produced in high volume present in each exposome category
Number of neurotoxicants
250
200
150
100
50
0
34
30
166
102
148
85
88
66
247
123
14
7
100
37
132
73
Childrens' exposome
Dietary exposome
External environmental exposome
Indoor-specific exposome
Miscellaneous external exposome
Occupational exposome
Pesticide/ biocide exposome
Skin-specific exposome
Category of Exposome
E
Adipose tissue 10
Amniotic fluid 12
Blood 63
Bone 3
Brain 10
Breast milk 30
Cord blood 35
Follicular fluid 10
Hair 17
Hand 4
Heart 2
Kidney 1
Liver 2
Lung 2
Mouth 2
Muscle 1
Nail 11
Pancreas 1
Pituitary gland 1
Placenta 17
Saliva 11
Semen 4
Skin 3
Spinal cord 2
Sweat 2
Thyroid gland 1
Tooth 8
Umbilical cord 7
Urinary bladder 1
Urine 68
Urothelial cells 1
0 10 20 30 40 50 60 70
Biospecimen
Number of neurotoxicants

Figure 5.2 ***(previous page)*****:** (A) Venn diagram showing the occurrence of the 475 potential neurotoxicants compiled in NeurotoxKb 1.0 across four existing resources, namely, the US EPA report (1976), Grandjean and Landrigan (2014), Mundy *et al.* (2015), and Aschner *et al.* (2017). (B) Sunburst plot showing the hierarchical classification of the 475 potential neurotoxicants into 2 chemical kingdoms and 20 chemical super-classes. The number of potential neurotoxicants in each kingdom or super-class is indicated within parenthesis. (C) Venn diagram showing the overlap between the sets of potential neurotoxicants present in Substances in use (SIU) lists, Substances of concern (SOC) lists, and High production volume (HPV) lists. Here, the potential neurotoxicants present in SIU lists and SOC lists are labeled as 'Neurotoxicants in use' and 'Neurotoxicants of concern', respectively. (D) Presence of the 475 potential neurotoxicants across chemical lists categorized into 8 exposome categories, namely, Children's exposome, Dietary exposome, External environmental exposome, Indoor-specific exposome, Miscellaneous external exposome, Occupational exposome, Pesticide/biocide exposome, and Skin-specific exposome. This plot displays two bars for each exposome category wherein one bar gives the number of neurotoxicants present in that exposome while other bar gives the number of neurotoxicants that are produced in high volume present in that exposome. (E) The bar chart shows the occurrence of the 475 potential neurotoxicants in NeurotoxKb 1.0 across 31 different human biospecimens.

5.1.4 Physicochemical and ADMET properties of neurotoxicants

We have used cheminformatics software to compile physicochemical properties, molecular descriptors and predicted ADMET properties for the 475 potential neurotoxicants in NeurotoxKb 1.0. This information will assist both computational and experimental research on neurotoxicants in future. The physicochemical properties and the molecular descriptors for the 475 potential neurotoxicants were computed using RDKit [179], PaDEL [180, 181] and Pybel [182]. The ADMET properties for the 475 potential neurotoxicants were predicted using admetSAR 2.0 [183], pkCSM [184], SwissADME [186], Toxtree 2.6.1 [187] and vNN server [188].

5.2 Web interface of NeurotoxKb

NeurotoxKb 1.0 provides the compiled information on the 475 potential neurotoxicants via a user-friendly web interface (Figure 5.4). The web interface of NeurotoxKb 1.0 (Figure 5.4) has been created using an approach similar to that described in Section 2.2. The compiled database on the 475 potential neurotoxicants is stored and retrieved using MariaDB [195] and Structured Query Language (SQL), respectively. Interactive visualiza-

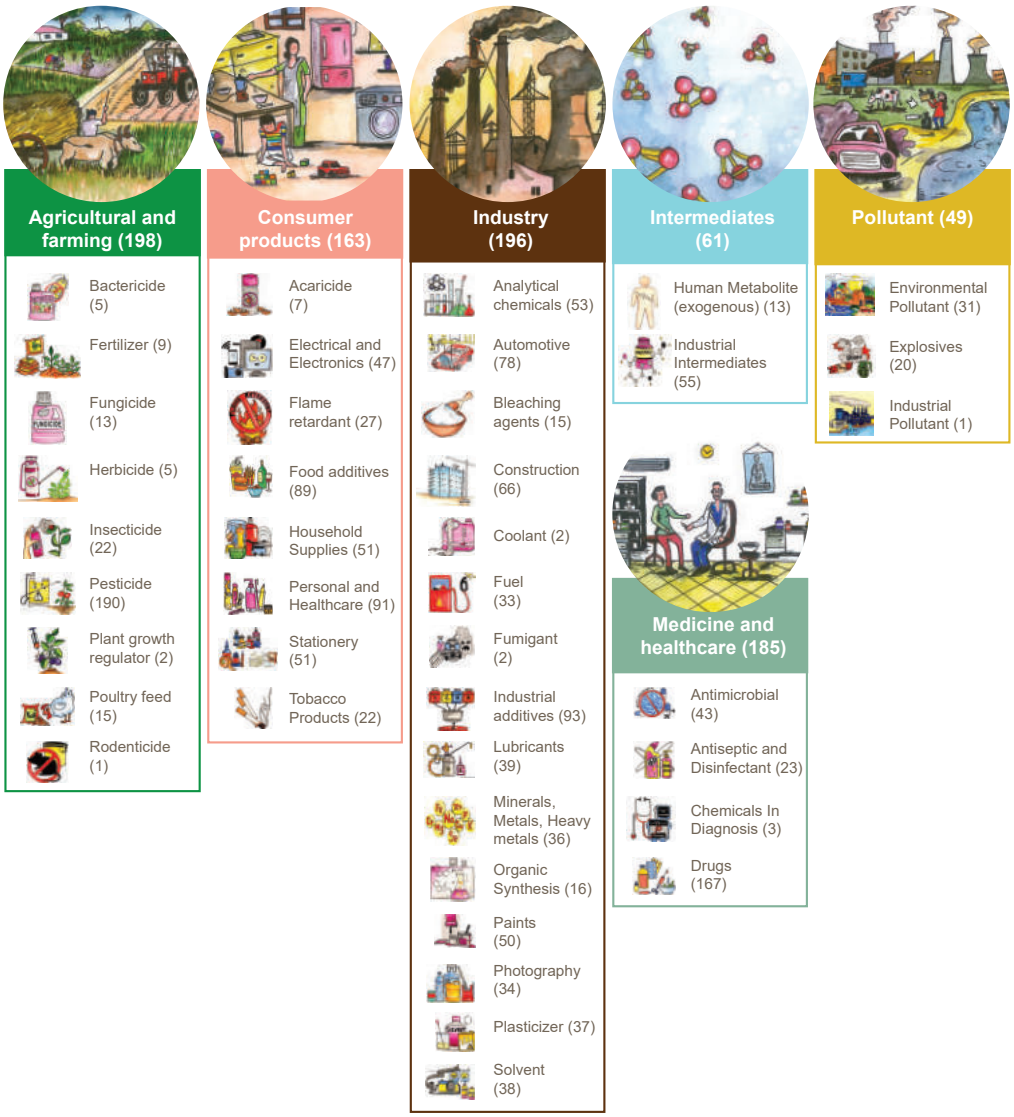


Figure 5.3: Classification of the 475 potential neurotoxicants in NeurotoxKb 1.0 into 6 broad categories and 41 sub-categories based on their environmental source. The number of potential neurotoxicants in each category or sub-category is mentioned besides the category or sub-category within parenthesis. Note that a potential neurotoxicant can belong to more than one category or sub-category of environmental sources.

tion of the compiled information in NeurotoxKb 1.0 is facilitated by Cytoscape.js [193], Google Charts [191] and Plotly [276]. NeurotoxKb 1.0 is hosted on an Apache [196] webserver running on Debian 9.4 Linux Operating System. Using the web interface of NeurotoxKb 1.0, users can access detailed information on any of the potential neurotoxicants via search or browse options (Figure 5.4).

5.3 Comparison of NeurotoxKb 1.0 with existing resources on neurotoxicants

Table 5.1 presents a comparison of our resource, NeurotoxKb 1.0, with the four existing resources, namely, the US EPA report [60], Grandjean and Landrigan [57], Mundy *et al*. [61] and Aschner *et al*. [62] on potential neurotoxicants. From this table, it is evident that NeurotoxKb 1.0 [38] will be a valuable resource for future research and monitoring of neurotoxicants due to several additional features in comparison to existing resources.

5.4 Exploration of potential neurotoxicants across chemical regulations and guidelines

Understanding the environmental sources and routes of exposure to neurotoxicants will be critical for monitoring and mitigation of their adverse effects on humankind. We have explored the presence of neurotoxicants in external exposomes via a comparative analysis with 55 publicly available chemical lists including inventories, regulations and guidelines (Figure 5.5; Supplementary Table S5.3). These 55 chemical lists were broadly classified into two categories, namely 'Substances in use (SIU)' and 'Substances of concern (SOC)' (Figure 5.5; Supplementary Table S5.3). SIU lists consist of chemicals that are permitted or found to be in regular use while SOC lists consist of chemicals that are marked hazardous, regulated or restricted by government or independent bodies across the world [36]. Based on the source or route of human exposure, the 55 chemical lists have further been

Figure 5.4: The web interface of NeurotoxKb. (A) The screenshot displays the home page of NeurotoxKb 1.0. NeurotoxKb 1.0 has options to search and retrieve information on potential neurotoxicants. (B) Simple search to retrieve potential neurotoxicants using their chemical names or identifiers. (C) Physicochemical filter to retrieve potential neurotoxicants based on their physicochemical properties. (D) Chemical similarity filter to retrieve potential neurotoxicants that are structurally similar to a query compound. NeurotoxKb 1.0 also has options to browse information on potential neurotoxicants based on their (E) Environmental source classification, (F) Chemical classification, (G) Presence in chemical regulation or guideline, and (H) Presence in human biospecimen.

classified into 8 categories of exposomes, namely, 'Children's exposome', 'Dietary exposome', 'External environmental exposome', 'Indoor-specific exposome', 'Occupational exposome', 'Pesticide/biocide exposome', 'Skin-specific exposome' and 'Miscellaneous external exposome' (Figure 5.5; Supplementary Table S5.3), and these contribute to the total external exposome of humans.

In this work, we have performed a comparative analysis for potential neurotoxicants with SIU and SOC lists similar to that performed for potential endocrine disruptors in our previous contribution [36]. Note that the presence of any potential neurotoxicant in SIU or SOC lists reflects its potential for human exposure. As highlighted by Grandjean and Landrigan [57, 58], several of the commercial chemicals which are produced in high volume across the world, have not been tested for their neurotoxic potential. In this direction, we have also explored the presence of 475 potential neurotoxicants in two publicly available lists of chemicals produced in high volume, namely, the United States High Production Volume (USHPV) database and the Organisation for Economic Cooperation and Development High Production Volume (OECD HPV) list which was last updated in 2004.

We find that 311 potential neurotoxicants in NeurotoxKb 1.0 are present in at least one of the 55 chemical lists (Supplementary Table S5.4). Figure 5.2C shows the distribution of these 311 potential neurotoxicants across SIU, SOC and HPV lists. Notably, 162 potential neurotoxicants are present in both SIU and SOC lists, and further, 105 of these 162 potential neurotoxicants are also produced in high volume (Figure 5.2C). Among the 311 potential neurotoxicants present in at least one of the 55 chemical lists, Ethylene oxide is present in the maximum number (24) of lists which includes both SIU and SOC lists (Supplementary Table S5.4) [38]. Published literature on Ethylene oxide has clearly documented experimental evidence on its neurotoxicity, and humans are mainly exposed to this neurotoxicant via occupational exposure [277, 278].

Upon investigation of the presence of the 475 potential neurotoxicants across chemical lists categorized into 8 exposome categories revealed that 166 potential neurotoxicants in NeurotoxKb 1.0 are present in the dietary exposome, specifically as food additives,

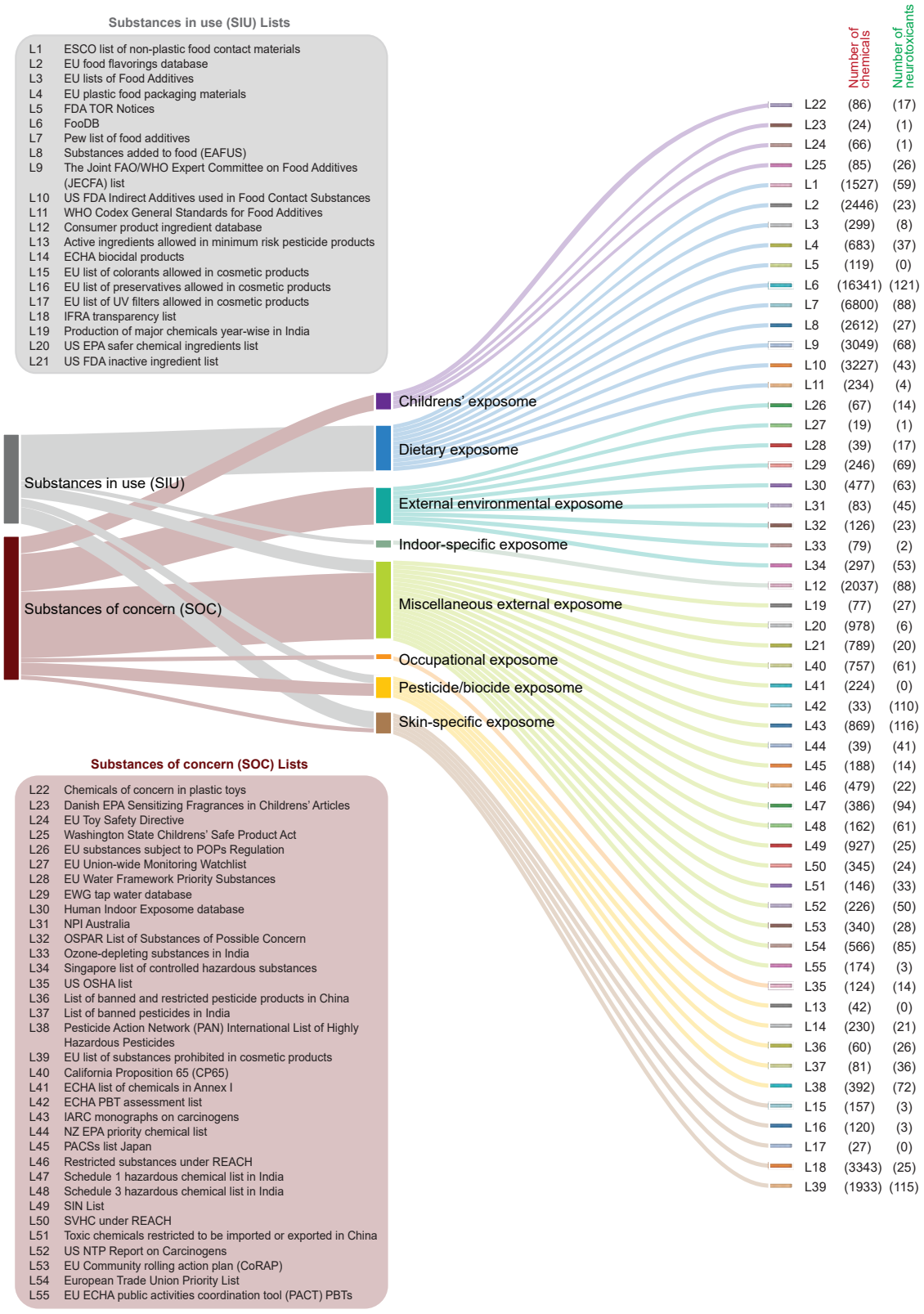

Substances in use (SIU) Lists
L1 ESCO list of non-plastic food contact materials
L2 EU food flavorings database
L3 EU lists of Food Additives
L4 EU plastic food packaging materials
L5 FDA TOR Notices
L6 FooDB
L7 Pew list of food additives
L8 Substances added to food (EAFUS)
L9 The Joint FAO/WHO Expert Committee on Food Additives (JECFA) list
L10 US FDA Indirect Additives used in Food Contact Substances
L11 WHO Codex General Standards for Food Additives
L12 Consumer product ingredient database
L13 Active ingredients allowed in minimum risk pesticide products
L14 ECHA biocidal products
L15 EU list of colorants allowed in cosmetic products
L16 EU list of preservatives allowed in cosmetic products
L17 EU list of UV filters allowed in cosmetic products
L18 IFRA transparency list
L19 Production of major chemicals year-wise in India
L20 US EPA safer chemical ingredients list
L21 US FDA inactive ingredient list
Substances in use (SIU)
Substances of concern (SOC)
Childrens' exposome
Dietary exposome
External environmental exposome
Indoor-specific exposome
Miscellaneous external exposome
Occupational exposome
Pesticide/biocide exposome
Skin-specific exposome
Substances of concern (SOC) Lists
L22 Chemicals of concern in plastic toys
L23 Danish EPA Sensitizing Fragrances in Childrens' Articles
L24 EU Toy Safety Directive
L25 Washington State Childrens' Safe Product Act
L26 EU substances subject to POPs Regulation
L27 EU Union-wide Monitoring Watchlist
L28 EU Water Framework Priority Substances
L29 EWG tap water database
L30 Human Indoor Exposome database
L31 NPI Australia
L32 OSPAR List of Substances of Possible Concern
L33 Ozone-depleting substances in India
L34 Singapore list of controlled hazardous substances
L35 US OSHA list
L36 List of banned and restricted pesticide products in China
L37 List of banned pesticides in India
L38 Pesticide Action Network (PAN) International List of Highly Hazardous Pesticides
L39 EU list of substances prohibited in cosmetic products
L40 California Proposition 65 (CP65)
L41 ECHA list of chemicals in Annex I
L42 ECHA PBT assessment list
L43 IARC monographs on carcinogens
L44 NZ EPA priority chemical list
L45 PACSs list Japan
L46 Restricted substances under REACH
L47 Schedule 1 hazardous chemical list in India
L48 Schedule 3 hazardous chemical list in India
L49 SIN List
L50 SVHC under REACH
L51 Toxic chemicals restricted to be imported or exported in China
L52 US NTP Report on Carcinogens
L53 EU Community rolling action plan (CoRAP)
L54 European Trade Union Priority List
L55 EU ECHA public activities coordination tool (PACT) PBTs
Number of chemicals
Number of neurotoxicants
L22 (86) (17)
L23 (24) (1)
L24 (66) (1)
L25 (85) (26)
L1 (1527) (59)
L2 (2446) (23)
L3 (299) (8)
L4 (683) (37)
L5 (119) (0)
L6 (16341) (121)
L7 (6800) (88)
L8 (2612) (27)
L9 (3049) (68)
L10 (3227) (43)
L11 (234) (4)
L26 (67) (14)
L27 (19) (1)
L28 (39) (17)
L29 (246) (69)
L30 (477) (63)
L31 (83) (45)
L32 (126) (23)
L33 (79) (2)
L34 (297) (53)
L12 (2037) (88)
L19 (77) (27)
L20 (978) (6)
L21 (789) (20)
L40 (757) (61)
L41 (224) (0)
L42 (33) (110)
L43 (869) (116)
L44 (39) (41)
L45 (188) (14)
L46 (479) (22)
L47 (386) (94)
L48 (162) (61)
L49 (927) (25)
L50 (345) (24)
L51 (146) (33)
L52 (226) (50)
L53 (340) (28)
L54 (566) (85)
L55 (174) (3)
L35 (124) (14)
L13 (42) (0)
L14 (230) (21)
L36 (60) (26)
L37 (81) (36)
L38 (392) (72)
L15 (157) (3)
L16 (120) (3)
L17 (27) (0)
L18 (3343) (25)
L39 (1933) (115)

Figure 5.5 ***(previous page)*:** Sankey plot displays the 55 chemical lists considered for comparative analysis that are a part of chemical inventories, regulations and guidelines. These lists were broadly classified into two categories, namely, Substances in use (SIU) and Substances of concern (SOC), based on the nature of substances. Further, these lists have also been classified into 8 categories of exposome, namely, Children's exposome, Dietary exposome, External environmental exposome, Indoor-specific exposome, Miscellaneous external exposome, Occupational exposome, Pesticide/biocide exposome, and Skin-specific exposome, based on the route or source of exposure. Besides each chemical list, the total number of chemicals and the number of potential neurotoxicants present in that list are shown within parenthesis.

food packaging materials and food contact substances (Figure 5.2D). For example, the Pew list of food additives (L7) contains 88 potential neurotoxicants (Figure 5.5; Supplementary Table S5.4). Further analysis of the SIU lists classified as Indoor-specific exposome, Pesticide/biocide exposome, Skin-specific exposome or Miscellaneous external exposome found the presence of several potential neurotoxicants compiled in NeurotoxKb 1.0 (Supplementary Table S5.4). In other words, we find that several potential neurotoxicants compiled in NeurotoxKb 1.0 are in regular use [38]. An analysis of the SOC lists classified as Children's exposome, Occupational exposome, Pesticide/biocide exposome, Skin-specific exposome, External environmental exposome or Miscellaneous external exposome found that several potential neurotoxicants compiled in NeurotoxKb 1.0 are also subject to chemical regulations worldwide [38].

To highlight the possible implications from this exploratory analysis of the presence of potential neurotoxicants across 55 chemical lists including inventories, regulations and guidelines, we next focus on chemical lists classified into a single category of external exposome, namely, Children's exposome. As neurotoxicants can cause permanent or irreversible damage to neuronal systems [55, 56], it is important to monitor and regulate their exposure to developing children. For this focused analysis, we considered 4 SOC lists namely, Chemicals of concern in plastic toys (L22), Danish EPA Sensitizing Fragrances in Children's Articles (L23), EU Toy Safety Directive (L24), and Washington State Children's Safe Product Act (L25), which contain chemicals prohibited or restricted in children related consumer products. We find that 34 potential neurotoxicants compiled

in NeurotoxKb 1.0 are present in the lists pertaining to Children's exposome, and of these, 30 potential neurotoxicants are also produced in high volume as they are present in HPV lists (Supplementary Table S5.4). Our observations are indicative of the extent to which these chemicals have been, or are currently being used, in children related products. These 30 potential neurotoxicants warrant further attention, and dedicated monitoring strategies to prevent exposure of children (Supplementary Table S5.4) [38].

5.5 Exploration of potential neurotoxicants in human biospecimens

Exposome refers to the totality of exposure during the lifetime of an individual and their associated health effects [13, 18–20]. Note that the presence of any potential neurotoxicant in a human biospecimen presents conclusive proof of human exposure and is also indicative of its potential to affect the nervous system. In this work, we have explored the presence of 475 potential neurotoxicants in human biospecimens using compiled data in two resources, namely, the Exposome-Explorer [24] and CTD [30].

Using literature mining, Exposome-Explorer [24] has compiled information on environmental chemicals detected in different human biospecimens from published literature based on dietary and pollution exposures. Similarly, 'Exposure – study associations' in CTD [30] can be used to retrieve compiled information from published literature on environmental chemicals detected in different human biospecimens. Importantly, the annotation of the human biospecimens is not uniform across Exposome-Explorer [24] and CTD [30]. Therefore, we have manually curated and unified the different human biospecimens captured in the two resources, Exposome-Explorer [24] and CTD [30], into 31 different types or exposomes (Figure 5.3; Supplementary Table S5.6). For example, we have grouped human biospecimens such as plasma, serum, blood proteins or blood cells into a single type 'blood' exposome in our work.

We find that 91 potential neurotoxicants were detected in at least one of the 31 human

biospecimens (Figure 5.2E; Supplementary Table S5.5). Among the 91 potential neurotoxicants detected in human biospecimens, Arsenic was detected in maximum number (16) of human biospecimens. Among the 31 human biospecimens, the 68 and 63 potential neurotoxicants were detected in urine and blood, respectively (Figure 5.2E) [38].

Human fetus is vulnerable to hazardous chemicals such as neurotoxicants [57, 58]. Several potential neurotoxicants were detected in human biospecimens related to fetal development or pregnancy. Specifically, we find that 35, 17, 12 and 7 potential neurotoxicants were detected in Cord blood, Placenta, Amniotic fluid and Umbilical cord, respectively (Figure 5.2E; Supplementary Table S5.5). Moreover, 30 potential neurotoxicants were also detected in Breast milk via which breastfed infants can be exposed to such chemicals (Figure 5.2E; Supplementary Table S5.5). Human brain is sensitive to neurotoxicants and the blood-brain barrier provides only partial protection against such chemicals [279]. We find that 10 potential neurotoxicants were detected in the brain (Figure 5.2E; Supplementary Table S5.5) [38].

We would like to highlight that well-known neurotoxicants including heavy metals such as Arsenic, Cadmium, Lead, Mercury, Nickel and Selenium, and Perfluoroalkyl substances such as Perfluorooctanesulfonic acid and Perfluorooctanoic acid, were detected in biospecimens related to fetal development, breast milk and brain. These observations underscore the omnipresence of well-known neurotoxicants in our environment, and invite further research and regular monitoring of these chemicals in daily use products and human exposome.

5.6 Prioritization of potential environmental neurotoxicants

An exploration of the current chemical regulations and guidelines enabled us to better understand the route and likelihood of human exposure to potential neurotoxicants in their lifetime. We next decided to explore the utility of our resource NeurotoxKb 1.0 in aiding

prioritization of potential neurotoxicants. For this purpose, we have analyzed the presence of the 475 potential neurotoxicants compiled in NeurotoxKb 1.0 across following lists:

1. Two lists of high production volume (HPV) chemicals, namely, the USHPV database and the OECD HPV list. These lists enable us to identify potential neurotoxicants that are extensively manufactured, and thus, have a high likelihood of human exposure.

2. List of substances of very high concern (SVHC) under Registration, Evaluation, Authorisation and Restriction of Chemicals (REACH) regulation of the European Union (EU). SVHC includes chemicals based on their potential to be: (i) Carcinogenic, Mutagenic, toxic to Reproduction (CMR), (ii) disruptive to the endocrine system, (iii) Persistent, Bioaccumulative and Toxic (PBT), and (iv) very Persistent and very Bioaccumulative (vPvB).

Table 5.2 gives the list of 18 potential neurotoxicants in NeurotoxKb 1.0 that are also present in both HPV and SVHC lists. Being registered as SVHC, these 18 chemicals are monitored and phased out where necessary, under stringent controls in the EU. These 18 chemicals are associated with multiple types of toxicity (Table 5.2). Overall, our analysis suggests the need for dedicated monitoring and worldwide prioritization of these 18 potential neurotoxicants. We remark that our analysis of the potential neurotoxicants produced in high volume is limited to HPV lists pertaining to EU and USA due to the lack of publicly available HPV lists for other countries. Regulatory bodies in other countries seeking to improve the prioritization of potential neurotoxicants can analyze NeurotoxKb in conjunction with country-specific data on chemical production volume and scale of use.

A common plasticizer, Bis(2-ethylhexyl) phthalate, is among the 18 potential neurotoxicants suggested for prioritization in this chapter. Bis(2-ethylhexyl) phthalate, also known as diethylhexyl phthalate or DEHP, is present in 22 out of the 55 chemical lists, of which 7 are SIU lists and 15 are SOC lists (Supplementary Table S5.4) [38]. These 22 chemical lists fall into 6 external exposome categories, namely, Children's exposome, Dietary exposome, External environmental exposome, Indoor-specific exposome, Skin-

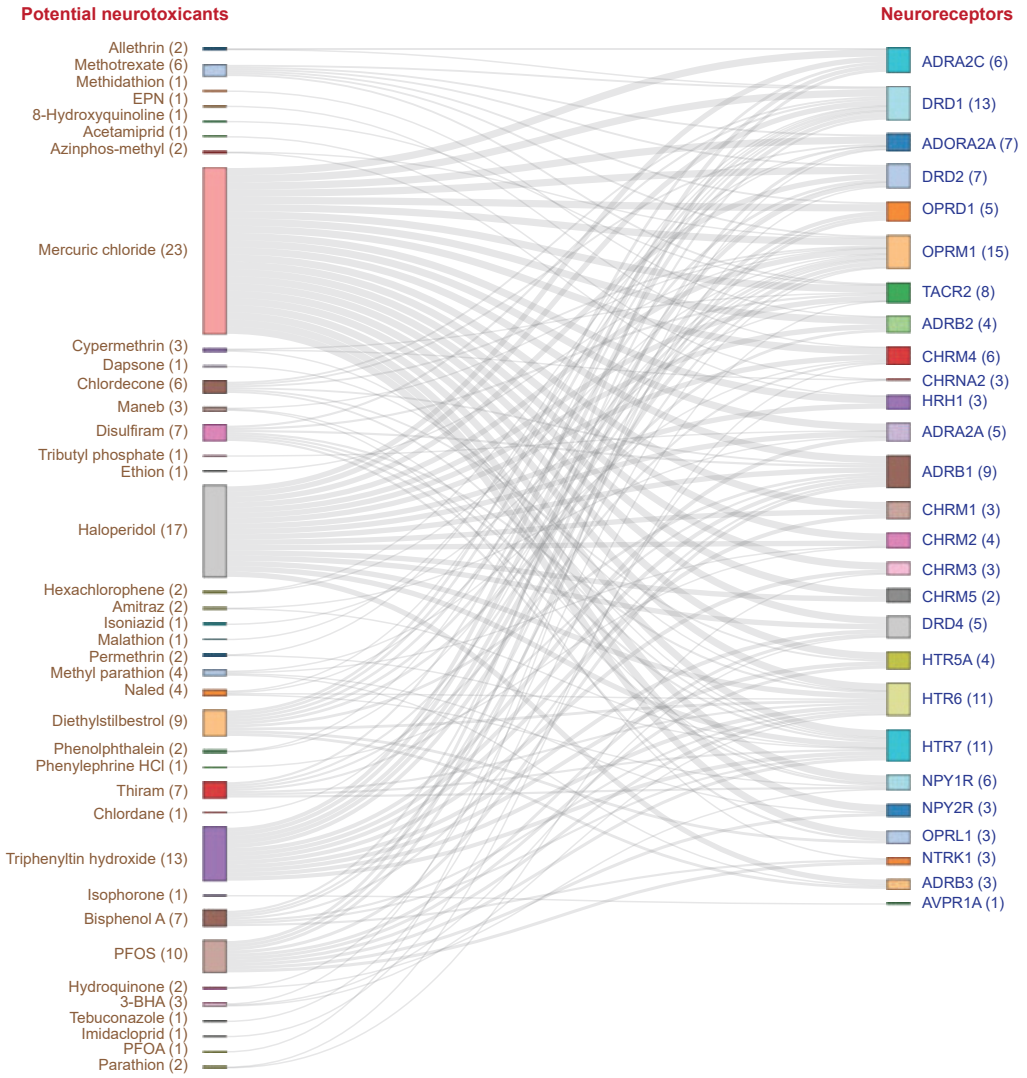


Figure 5.6: The bipartite network of 38 potential neurotoxicants in NeurotoxKb 1.0 that target 27 human neuroreceptors. Besides each potential neurotoxicant, the number of target neuroreceptors is indicated within parenthesis. Besides each neuroreceptor, the number of potential neurotoxicants targeting it is indicated within parenthesis.

specific exposome and Miscellaneous external exposome. Bis(2-ethylhexyl) phthalate has been found to impair learning and memory, and cause brain tissue damage in rodents and humans [280, 281]. In sum, our resource can aid and offer direction to monitoring organizations and regulatory agencies in identifying, prioritizing and improving the regulations around neurotoxicants.

5.7 Interaction of environmental neurotoxicants with neuroreceptors

Identification of target human genes or proteins of environmental neurotoxicants can shed light on complex molecular mechanisms via which these chemicals cause neurotoxicity. We have used ToxCast [89] to identify the target human genes or proteins of the 475 potential neurotoxicants in NeurotoxKb 1.0. To retrieve the list of target human genes perturbed by potential neurotoxicants, we have used ToxCast invitroDB version 3.2 dataset released in August 2019 [215]. We followed the method described in the Section 2.4.2 to extract from ToxCast the human target genes perturbed upon exposure to compiled neurotoxicants. Based on human-specific assays in ToxCast [89], we were able to obtain 255 target human genes for 220 out of the 475 potential neurotoxicants in NeurotoxKb 1.0 (Supplementary Table S5.6). Further investigation of the 255 target human genes of the 220 potential neurotoxicants revealed that 27 target genes correspond to neuroreceptors. We find that 38 potential neurotoxicants in NeurotoxKb 1.0 target at least one of these 27 neuroreceptors (Figure 5.6; Supplementary Table S5.6) [38]. Among these 38 potential neurotoxicants, 4 neurotoxicants namely, Mercuric chloride, Haloperidol, Triphenyltin hydroxide and Perfluorooctanesulfonic acid (PFOS), target 10 or more neuroreceptors (Figure 5.6). Among the 27 neuroreceptors which are targets of at least one potential neurotoxicant, the neuroreceptor OPRM1 (Opioid Receptor Mu 1) for endogenous opioids such as β-endorphin and endomorphin, was found to interact with 15 potential neurotoxicants. Other neuroreceptors which are targets of at least 10 potential neurotoxicants include the receptor DRD1 (Dopamine receptor D1) for neurotransmitter dopamine, and the receptors HTR6 (5-Hydroxytryptamine Receptor 6) and HTR7 (5-Hydroxytryptamine Receptor 7) for the neurotransmitter serotonin (Figure 5.6; Supplementary Table S5.6) [38]. In future, an in depth analysis of chemical-gene interactions will shed new insights on the molecular mechanisms via which the exposure to the 475 potential neurotoxicants in NeurotoxKb 1.0 can lead to documented neurotoxic

endpoints in mammals.

5.8 Chemical similarity network of environmental neurotoxicants

Chemical similarity approaches can aid in early identification of toxic chemicals [198, 199] including potential neurotoxicants. To construct the CSN of neurotoxicants, we have employed the similarity metric Tanimoto coefficient [200]. For any pair of chemicals, Tanimoto coefficient has a value in the range 0 to 1, wherein the level of chemical similarity between two molecules is directly proportional to the corresponding Tanimoto coefficient value. The computation of Tanimoto coefficient between pairs of chemicals can depend on the choice of chemical fingerprints used to represent the molecules. Here, we have chosen Extended Circular Fingerprints (ECFP4) [129] while computing Tanimoto coefficient between different pairs of potential neurotoxicants.

In the CSN of potential neurotoxicants in NeurotoxKb 1.0, there are 475 nodes corresponding to the 475 potential neurotoxicants, and there is an edge between any pair of nodes if the corresponding Tanimoto coefficient value is ≥ 0.5. The chosen cutoff of Tanimoto coefficient ≥ 0.5 to decide on significant structural similarity between pairs of chemicals was motivated by a similar choice made in previous studies [282–284].

We find that the CSN of 475 potential neurotoxicants is fragmented into 60 connected components with the number of neurotoxicants ≥ 2 and 286 isolated neurotoxicants (Figure 5.7). Moreover, the largest connected component consists of only 13 potential neurotoxicants (Figure 5.7). In Figure 5.7, we have coloured the nodes based on the number of aromatic rings in the corresponding neurotoxicant. It can be seen that neurotoxicants belonging to a connected component typically have the same number of aromatic rings. Altogether, this preliminary analysis of the CSN of potential neurotoxicants reveals a fragmented network, and thus, the associated toxicological space has high chemical diversity [38].

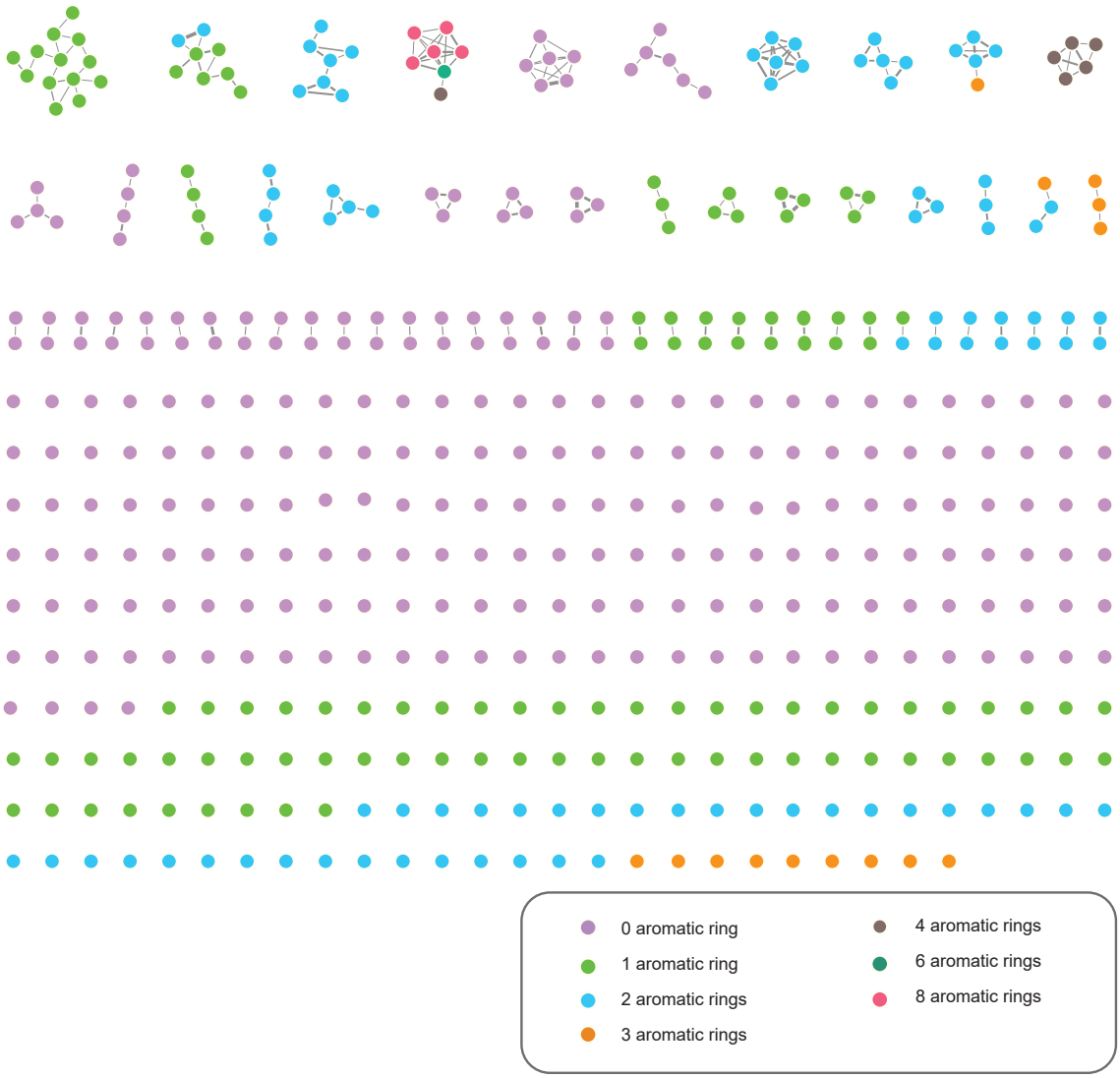


Figure 5.7: Chemical similarity network (CSN) of the 475 potential neurotoxicants in NeurotoxKb 1.0. In this figure, there are 475 nodes corresponding to the 475 potential neurotoxicants, and there is an edge between any pair of nodes if the corresponding Tanimoto coefficient value is $\geq$ 0.5. Further, nodes are coloured based on the number of aromatic rings present in the corresponding neurotoxicants, while the thickness of the edges indicate Tanimoto coefficient value between the corresponding neurotoxicants. Here, the connected components of the CSN are displayed in the decreasing order of the number of nodes in each component.

5.9 Discussion

The Swiss philosopher and poet, Henri-Frédéric Amiel (1821-1881), once stated that: "To repair is twenty times more difficult than to prevent". The quote is apt for the management of hazardous chemicals including environmental neurotoxicants. Since neurotoxicants can cause permanent or irreversible damage to the nervous system [52,55,56], early screening of environmental chemicals with potential to cause neurotoxicity is important for human well-being. In this direction, a comprehensive resource on potential neurotoxicants compiling published evidence specific to mammals, can aid in monitoring and regulation of human neurotoxicants. Here, we present such a comprehensive resource, NeurotoxKb 1.0, with compiled information on 475 potential non-biogenic neurotoxicants curated from 835 published studies specific to mammals. The entire compiled information on the 475 potential neurotoxicants in NeurotoxKb 1.0 can be easily accessed and retrieved via a user-friendly and interactive web interface (Figure 5.8).

Humans are exposed to environmental neurotoxicants via diverse sources (Figure 5.3). Firstly, a comparative analysis of NeurotoxKb 1.0 and 55 chemical lists which include inventories, regulations and guidelines, found that several potential neurotoxicants are both in regular use and produced in high volume (Figures 5.2C and 5.5). Secondly, a comparative analysis of NeurotoxKb 1.0 and chemicals detected in 31 different human biospecimens, found that several potential neurotoxicants have been detected in different biospecimens (Figure 5.2E). In other words, our comparative analysis with chemicals in regulatory lists or those detected in human biospecimens confirm the omnipresence of potential neurotoxicants in different categories of external exposomes (Figure 5.5). Furthermore, based on a comparative analysis of NeurotoxKb 1.0 with SVHC REACH regulation and HPV chemicals, we present a hazard priority list of 18 potential neurotoxicants (Table 5.2). In sum, NeurotoxKb 1.0 can be used for identification and prioritization of environmental neurotoxicants in human exposomes (Figure 5.8).

A unique feature of our resource on potential neurotoxicants is the compilation and

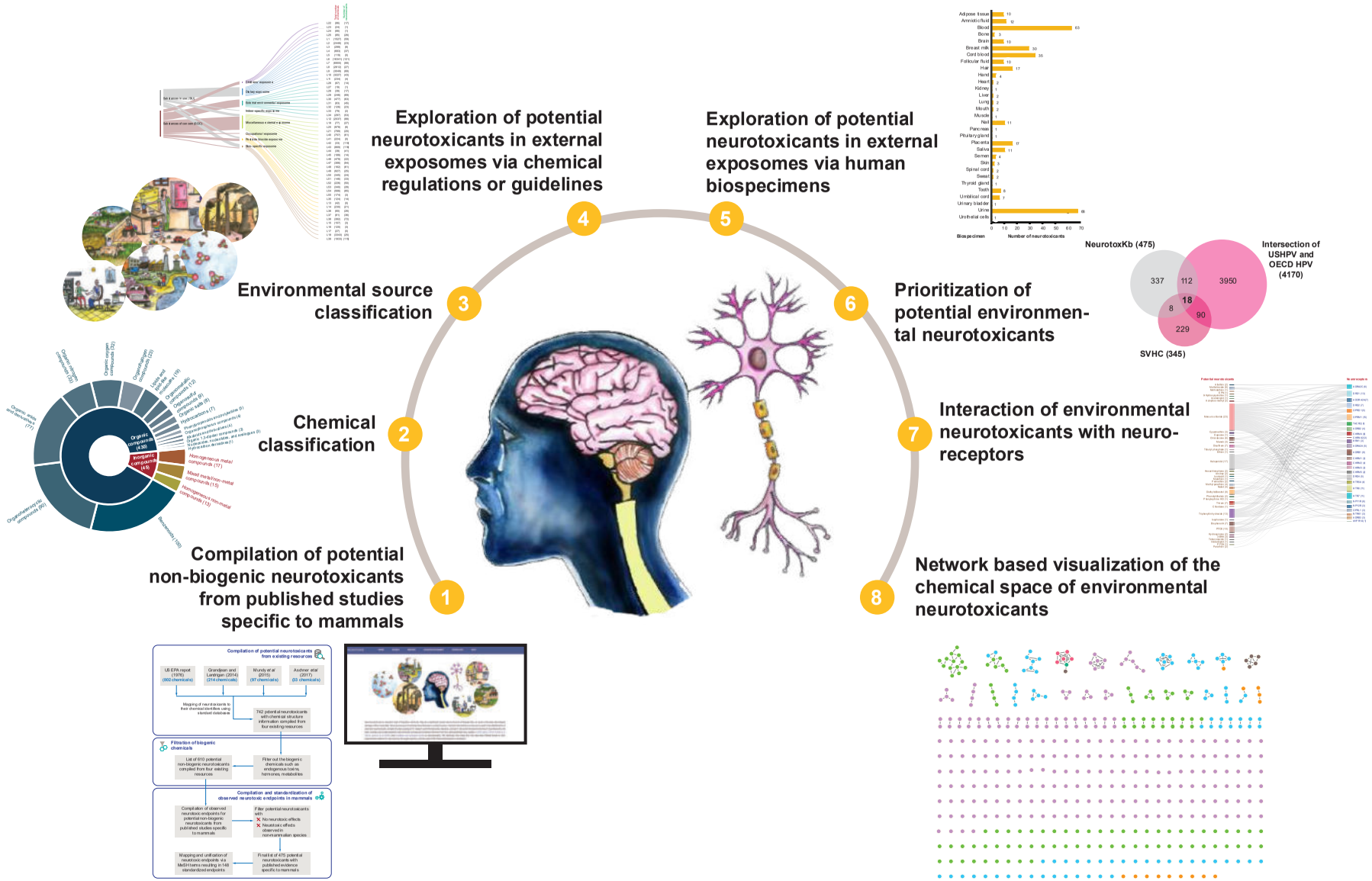


Figure 5.8: Schematic diagram summarizing NeurotoxKb 1.0 on environmental neurotoxicants.

standardization of neurotoxic endpoints from published studies specific to mammals. In future, it will be worthwhile to leverage this compiled information in NeurotoxKb 1.0 to develop adverse outcome pathways [99] for different neurotoxicants. We envisage that such an extension of our knowledgebase can further aid risk assessment of environmental chemicals.

Supplementary Information

Supplementary Tables S5.1-S5.6 associated with this chapter are available for download from the GitHub repository: https://github.com/asamallab/PhDThesis-Janani_R/blob/main/SI/ST_Chapter5.xlsx.

Feature	**NeurotoxKb 1.0**	**US EPA report (1976)**	**Grandjean and Landrigan (2014)**	**Mundy *et al.* (2015)**	**Aschner *et al.* (2017)**
Number of potential neurotoxicants	475	802	214	97	33
Web interface	Yes	No	No	Yes via CompTox	Yes via CompTox
Compilation of neurotoxic endpoints	Yes	Yes	No	Yes	Yes
Standardization of neurotoxic endpoints	Yes	No	No	No	No
Classification based on environmental source	Yes	No	Yes	Yes	Yes
Classification based on chemical structure	Yes	No	No	No	No
Presence in chemical regulation or guideline	Yes	No	No	Yes	Yes
Information on external exposomes	Yes	No	No	No	No
Presence in human biospecimen	Yes	No	No	No	No
Chemical identifiers	PubChem or CAS or MeSH	CAS	CAS	DSSTox substance identifier or CAS	DSSTox substance identifier or CAS
Download of 2D structure	SDF, MOL, MOL2	No	No	MOL	MOL
Download of 3D structure	SDF, MOL, MOL2, PDB, PDBQT	No	No	No	No
Physicochemical properties	Yes	No	No	Yes	Yes
Molecular descriptors	Yes	No	No	No	No
Predicted ADMET properties	Yes	No	No	Yes	Yes
Chemical-gene association	Yes	No	No	Yes	Yes
Chemical similarity filter	Yes	No	No	No	No

Table 5.1: Comparison of the features including compiled information captured in NeurotoxKb 1.0 for the potential neurotoxicants with respect to four existing resources.

Potential Neurotoxicant	Presence in USHPV	Presence in OECD HPV	Presence in SVHC	SVHC Criteria
Tributyltin oxide	Yes	Yes	Yes	PBT (Article 57d)
Lead	Yes	Yes	Yes	Toxic for reproduction (Article 57c)
N,N-Dimethylformamide	Yes	Yes	Yes	Toxic for reproduction (Article 57c)
Tetraethyllead	Yes	Yes	Yes	Toxic for reproduction (Article 57c)
Trichloroethylene	Yes	Yes	Yes	Carcinogenic (Article 57a)
Dinoseb	Yes	Yes	Yes	Toxic for reproduction (Article 57c)
Nitrobenzene	Yes	Yes	Yes	Toxic for reproduction (Article 57c)
Boric acid	Yes	Yes	Yes	Toxic for reproduction (Article 57c)
1-Bromopropane	Yes	Yes	Yes	Toxic for reproduction (Article 57c)
2-Methoxyethanol	Yes	Yes	Yes	Toxic for reproduction (Article 57c)
2,4-Dinitrotoluene	Yes	Yes	Yes	Carcinogenic (Article 57a)
Hydrazine	Yes	Yes	Yes	Carcinogenic (Article 57a)
Cadmium	Yes	Yes	Yes	Carcinogenic (Article 57a); Specific target organ toxicity after repeated exposure (Article 57(f) - human health)
Dibutyl phthalate	Yes	Yes	Yes	Toxic for reproduction (Article 57c); Endocrine disrupting properties (Article 57(f) - human health)
Propylene oxide	Yes	Yes	Yes	Carcinogenic (Article 57a); Mutagenic (Article 57b)
Acrylamide	Yes	Yes	Yes	Carcinogenic (Article 57a); Mutagenic (Article 57b)
Bisphenol A	Yes	Yes	Yes	Toxic for reproduction (Article 57c); Endocrine disrupting properties (Article 57(f) - environment); Endocrine disrupting properties (Article 57(f) - human health)
Bis(2-ethylhexyl) phthalate	Yes	Yes	Yes	Toxic for reproduction (Article 57c); Endocrine disrupting properties (Article 57(f) - environment); Endocrine disrupting properties (Article 57(f) - human health)

Table 5.2: List of 18 potential neurotoxicants in NeurotoxKb 1.0 suggested for prioritization. These 18 chemicals are considered to be substance of very high concern (SVHC) under REACH regulation, and moreover, are present in two lists of high production volume (HPV) chemicals, namely, United States High Production Volume (USHPV) database and Organisation for Economic Co-operation and Development High Production Volume (OECD HPV) list.

Chapter 6

ExHuMId: A curated resource and analysis of Exposome of Human Milk across India

The environmental exposure of women is a concern, especially during pregnancy and early motherhood [67]. A mother is exposed to a myriad of environmental chemicals through food, personal care products, household products, medicines, pollutants, or through her occupational environment [66, 285]. However, several environmental chemicals, which may affect the child, are capable of entering human milk [67, 285–287]. These chemicals are of concern due to the potential impact they can have on maternal health [63] and early development of a child [64, 65]. There is a need to monitor, regulate, and consciously avoid these chemicals wherever possible. Biomonitoring of human milk is therefore inevitable [64, 66, 67, 285–287].

Given that human milk is a biological matrix, whose monitoring is significant to healthcare and environmental safety, we believe it warrants a dedicated exposome database. The Exposome-Explorer contains a wide range of exposome detected in various biospecimens including blood, urine, plasma, and serum. It also includes the exposures

detected in human milk across different geographical regions [24]. Some studies have also compiled the list of chemicals detected in human milk, and these studies were published as research articles or scientific reports. A prominent example is the work of Lehmann *et al*. [286] that has compiled the human milk exposome from samples collected across the United States through literature mining and manual curation.

India is home to a population of nearly 1.33 billion [288] with extensive growth in agricultural and industrial sectors, contributing to the production and use of several commercial chemicals in everyday life [289]. Several studies have detected the presence of environmental contaminants in human milk and a few studies have also compiled the list of chemicals detected in human milk across India [290–292]. However, so far there has been no systematic effort towards the monitoring and compilation of these environmental contaminants in India, with the objective to aid chemical risk management and informing policy decisions [293]. For example, the reports by van den Berg *et al*. [294] and Sharma *et al*. [293] compile only the chemical component of the exposome [13], but lack the systematic compilation of maternal factors such as age, body weight, diet, and other factors which may affect the exposome.

In this chapter, we present a systematic approach to compile the **Ex**posome of **H**uman **M**ilk across **Ind**ia (ExHuMId) [39], through literature mining and manual curation of research articles that report experimentally detected environmental contaminants in breast milk in studies carried out across India. **The work reported in this chapter is contained in the published manuscript [39].**

6.1 Compilation of human milk contaminants specific to India

6.1.1 Literature mining and curation

We created the database, **Ex**posome of **Hu**man **M**ilk across **Ind**ia (ExHuMId) with the primary objective of bringing all the published knowledge surrounding human milk contaminants, specific to India, into a single knowledgebase [39]. In other words, ExHuMId compiles the list of human milk contaminants detected in published scientific studies involving samples collected across India.

As a first step, we performed an extensive literature search to identify relevant published research articles on PubMed [158] using the following keyword search:

(((breast OR human OR mother*) AND milk) OR breastmilk) AND India

This keyword search last performed on 24 August 2020, led to 1704 research articles. Subsequently, this set of 1704 articles was manually curated to obtain a subset of articles relevant to the study of human milk contaminants in India (Figure 6.1). Specifically, we retained only those articles pertaining to 'human milk' or 'breast milk', with samples collected solely from India. During the manual curation process, we excluded studies on samples collected from outside India, studies without specific geographical indication, review articles or conference abstracts, studies specific to essential nutrients, and articles promoting breastfeeding. This step resulted in a curated set of 36 research articles containing information about the environmental contaminants identified in human milk samples across India, using analytical techniques (Figure 6.1; Supplementary Table S6.1) [39].

From the curated list of 36 research articles, we have compiled the contaminants including their concentrations detected in human milk samples, geographical location, age, and other factors associated with the mothers from whom the milk samples were collected (Figure 6.1). For an unambiguous analysis, the data compiled in ExHuMId has

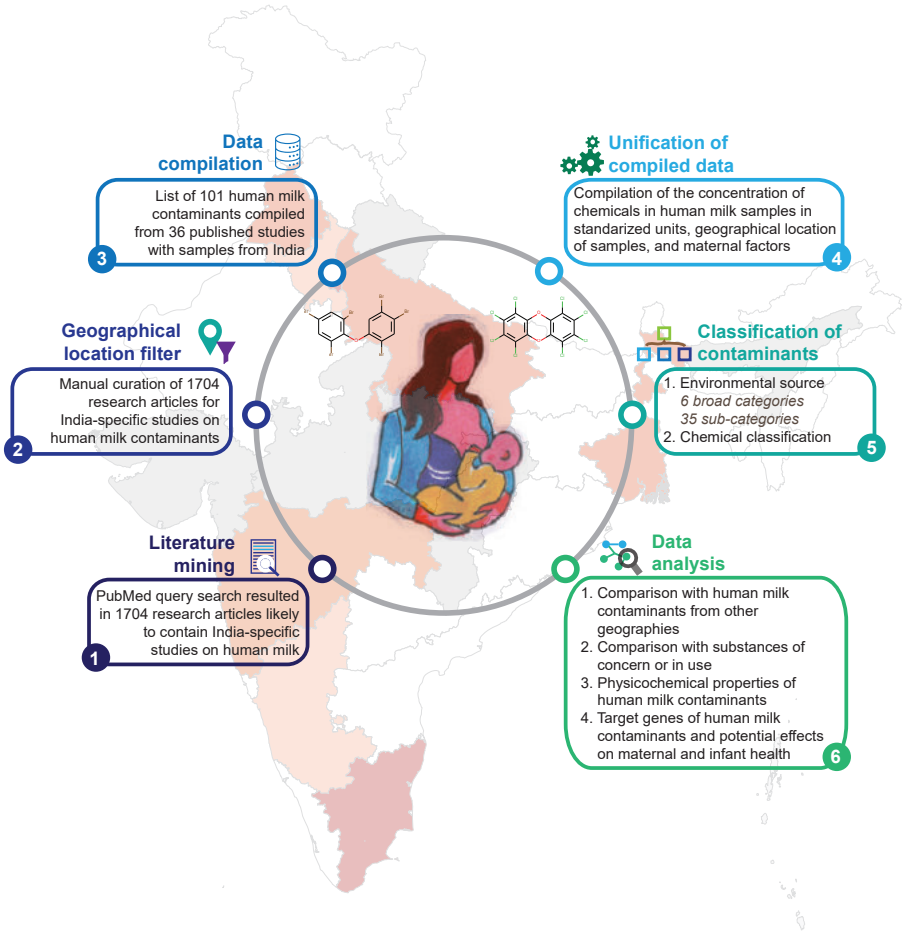


Figure 6.1: Schematic workflow describing the compilation, curation and analysis of the resource ExHuMId on Exposome of Human Milk across India.

been standardized and unified through the following steps.

The first step involved the standardization of the geographical locations from which human milk samples were collected in our curated set of 36 studies. The geographical locations of the study samples were mapped to their respective states in India (Figure 6.2A).

Our manually curated set of 36 studies also recorded a list of maternal factors that influence the presence or transfer of environmental chemicals into mothers' milk. The second step involved the unification of maternal factors that were compiled from the 36 research articles. We have compiled 23 maternal conditions associated with the human milk samples reported in the curated set of 36 published studies, and these maternal conditions include the body weight, food habits, societal factors, and other antenatal and postnatal conditions of the mothers. These maternal conditions were unified into 9 maternal factors, namely, body weight, food, gestational age, number of pregnancies (Primipara, Biparous and Multipara), occupation, phases of breast milk, residential area, social status, and types of birth (Figure 6.2D). Among these 9 maternal factors, the number of pregnancies is found to be highly distributed with many more contaminants (Figure 6.2D). Note that maternal factors are not available for all samples that have been compiled from the curated set of 36 published articles.

Next, the environmental chemicals detected in human milk across the curated set of 36 studies were mapped to standard chemical identifiers using PubChem [86], CAS, ChEMBL [295], and CTD [30] to obtain a set of 101 unique chemicals. The final step involved the manual unification of the units for the lowest concentration, highest concentration, mean, standard deviation and standard error associated with the measurement of each chemical in human milk samples in different studies. This step resulted in the unification of the compiled information in 12 different concentration units into 2 standardized concentration units, namely, μg/g lipid weight and μg/L lipid weight. Of 101 compiled human milk contaminants, we find 71 chemicals with concentration in standard unit μg/g lipid weight, 18 chemicals with concentration in standard unit μg/L lipid

weight, and 11 chemicals with concentrations in both the standard units [39]. Furthermore, we gathered information on their chemical structure including two-dimensional (2D) and three-dimensional (3D) structure (in SDF, MOL and MOL2 formats), canonical SMILES, InChI, and InChIKey.

6.1.2 Classification of human milk contaminants

Following the compilation and standardization of the data on human milk contaminants, we classified the human milk contaminants based on: (a) their environmental source, and (b) their chemical features [39].

Based on environmental source

Based on the classification of environmental sources, contaminants have been classified into the 6 broad categories: 'Agriculture and Farming', 'Consumer Products', 'Industry', 'Intermediates', 'Medicine and Healthcare', and 'Pollutant'. The majority of the chemicals compiled in ExHuMId fall under the category 'Pollutant' (Figure 6.2F). The above-mentioned 6 broad categories were further classified into 35 sub-categories based on their environmental sources.

Based on chemical structure

The human milk contaminants were structurally classified according to the taxonomy from ClassyFire [173, 174], a web-based application (Figure 6.2E). Upon classifying the 101 contaminants in ExHuMId based on their chemical class, we find that 96 are organic and 5 are inorganic (Figure 6.2E). Among the 96 organic chemicals in ExHuMId, the largest number (46 contaminants) belong to the super-class benzenoids (Figure 6.2E).

6.2 Web interface of ExHuMId

We believe, in agreement with many others in the science community [214], that scientific knowledge and experimental findings should be readily available to aid and spur fur-

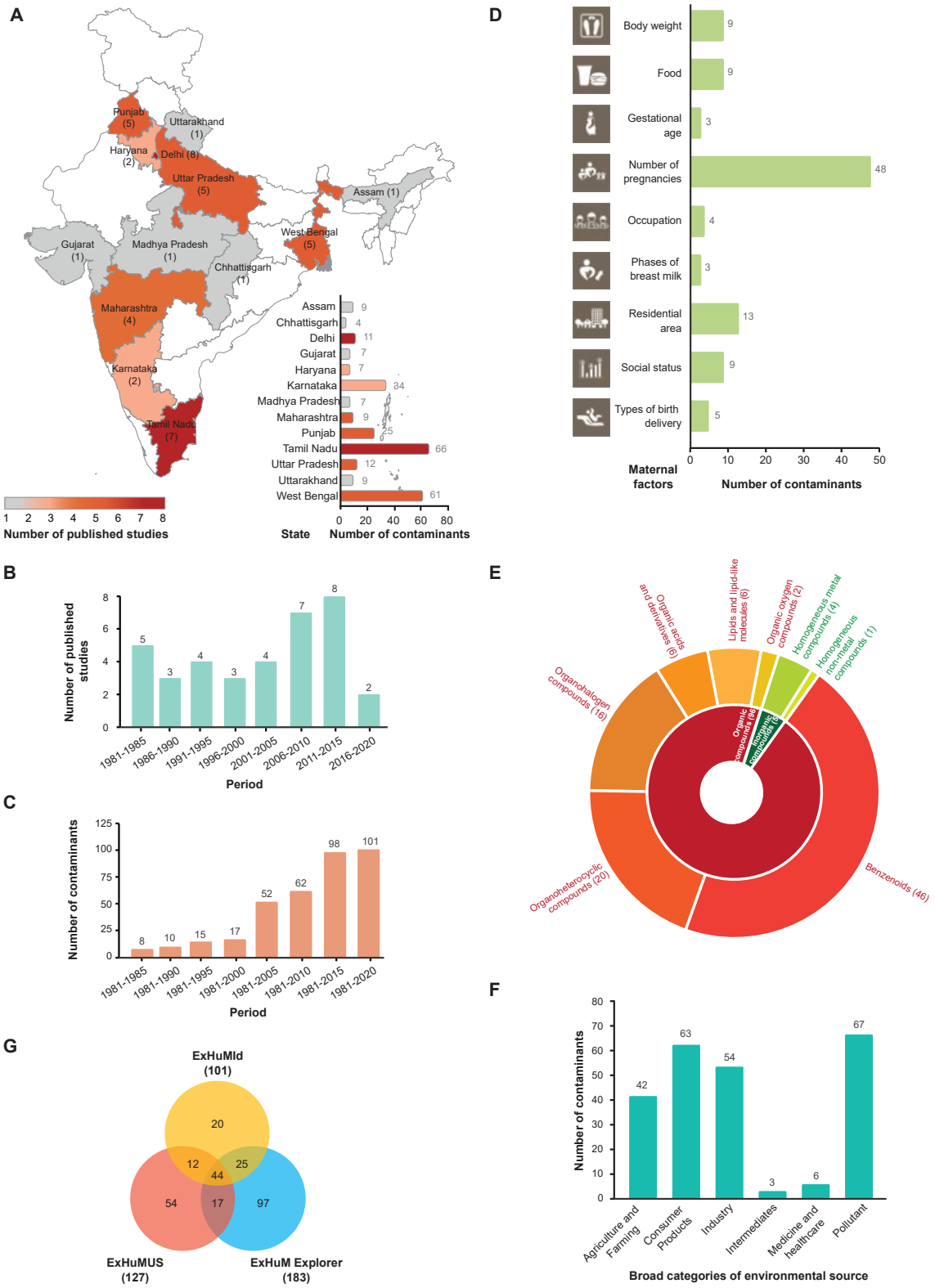

A
Punjab (5)
Uttarakhand (1)
Haryana (2)
Delhi (8)
Uttar Pradesh (5)
Assam (1)
West Bengal (5)
Gujarat (1)
Madhya Pradesh (1)
Chhattisgarh (1)
Maharashtra (4)
Karnataka (2)
Tamil Nadu (7)
1 2 3 4 5 6 7 8
Number of published studies
Assam 9
Chhattisgarh 4
Delhi 11
Gujarat 7
Haryana 7
Karnataka 34
Madhya Pradesh 7
Maharashtra 9
Punjab 25
Tamil Nadu 66
Uttar Pradesh 12
Uttarakhand 9
West Bengal 61
0 20 40 60 80
State
Number of contaminants
D
Body weight 9
Food 9
Gestational age 3
Number of pregnancies 48
Occupation 4
Phases of breast milk 3
Residential area 13
Social status 9
Types of birth delivery 5
0 10 20 30 40 50
Maternal factors
Number of contaminants
B
Number of published studies
5 3 4 3 4 7 8 2
1981-1985 1986-1990 1991-1995 1996-2000 2001-2005 2006-2010 2011-2015 2016-2020
Period
C
Number of contaminants
8 10 15 17 52 62 98 101
1981-1985 1981-1990 1981-1995 1981-2000 1981-2005 1981-2010 1981-2015 1981-2020
Period
E
Organic acids and derivatives (6)
Lipids and lipid-like molecules (6)
Organic oxygen compounds (2)
Homogeneous metal compounds (4)
Homogeneous non-metal compounds (1)
Organohalogen compounds (16)
Organic compounds (66)
Inorganic compounds (4)
Organoheterocyclic compounds (20)
Benzenoids (46)
F
Number of contaminants
42 63 54 3 6 67
Agriculture and Farming
Consumer Products
Industry
Intermediates
Medicine and healthcare
Pollutant
Broad categories of environmental source
G
ExHuMId (101)
20
12 25
44
54 17 97
ExHuMUS (127)
ExHuM Explorer (183)

Figure 6.2 ***(previous page)*:** (A) An India map displaying different states or geographical locations from where samples were obtained in the curated set of 36 published research articles in ExHuMId on human milk contaminants. The number besides each state in brackets gives the number of published articles reporting human milk samples from that state. The histogram shows the number of contaminants detected across samples obtained from each state. (B) A chronological analysis of the curated set of 36 published studies in ExHuMId. (C) A chronological analysis of the cumulative number of contaminants detected across published studies in different time periods. (D) Evidence across 9 maternal factors compiled from published articles associated with the human milk contaminants in ExHuMId. (E) Sunburst plot showing the chemical classification of 101 human milk contaminants in ExHuMId into 2 kingdoms and 8 super-classes as obtained from ClassyFire. (F) Distribution of 101 human milk contaminants in ExHuMId across 6 broad categories of environmental sources. (G) Comparison of 101 human milk contaminants in ExHuMId with those in two other resources, namely, ExHuMUS and ExHuM Explorer.

ther research, inform industry directions and policy decisions, especially when it comes to chemical usage and regulation. Knowledgebases make this possible, by serving as a platform for researchers, industry and regulatory authorities to access a range of useful information. This has motivated us to compile Exposome of Human Milk across India (ExHuMId) version 1.0, a curated resource on human milk contaminants specific to India.

ExHuMId is an online knowledgebase that compiles detailed information about the human milk contaminants detected in samples collected from India, with supporting evidence from 36 published scientific studies. This includes their chemical names, unique chemical identifiers, their concentrations as detected in our curated set of experiments, age and maternal factors of the donor of the sample, physicochemical properties, predicted ADMET properties, molecular descriptors, and target genes. Users can also access the identifiers, structural information including 2D and 3D structure for each substance. ExHuMId is accessible at: https://cb.imsc.res.in/exhumid.

The web interface of ExHuMId was created using an approach similar to that described in Section 2.2. Through the web interface (Figure 6.3), users can also access the identifiers, structural information including 2D and 3D structure for each human milk contaminant in ExHuMId. The users can navigate ExHuMId via either simple search or browse options (Figure 6.3).

A


ExHuMId
Exposome of Human Milk across India
HOME SEARCH BROWSE ACKNOWLEDGEMENT DOWNLOAD

B


CHEMICAL SEARCH
Simple search
Physicochemical filter
Chemical similarity filter
Chemical name
Chemical identifier (PubChem or CAS)
Search

C


Simple search
Physicochemical filter
Chemical similarity filter
Molecular Weight
LogP
TPSA
Hydrogen bond acceptors (HBA)
Hydrogen bond donors (HBD)
Heavy atoms
Heteroatoms
Rotatable bonds
Search

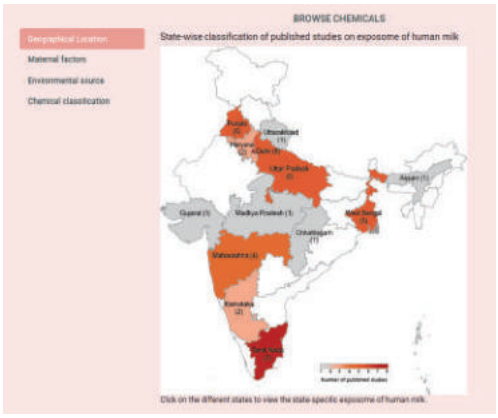
D


Simple search
Physicochemical filter
Chemical similarity filter
Enter SMILES
Choose Fingerprint
Search

E


Pb
Lead
Identification
Environmental source
Chemical classification
Experimental data
Chemical-gene interaction
Physicochemical properties
ADME properties
Toxicity properties
Descriptors

F


BROWSE CHEMICALS
Geographical Location
Maternal factors
Environmental source
Chemical classification
State-wise classification of published studies on exposome of human milk
Click on the different states to view the state specific exposome of human milk.

G


Geographical Location
Maternal factors
Environmental source
Chemical classification
Body weight
Food
Gestational age
Number of pregnancies
Occupation
Phases of breast milk
Residential area
Social status
Types of birth delivery

H


Geographical Location
Maternal factors
Environmental source
Chemical classification

I

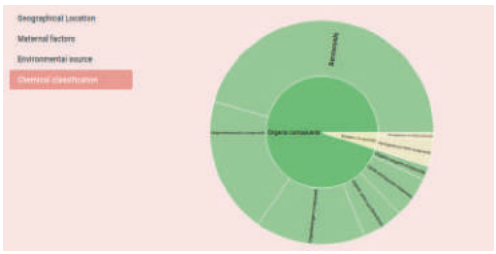

Geographical Location
Maternal factors
Environmental source
Chemical classification

Figure 6.3 ***(previous page)*****:** The web interface of ExHuMId. (A) A screenshot of the home page of ExHuMId. In Search section, there are three options available to search and obtain information on human milk contaminants compiled in ExHuMId. (B) Firstly, Simple search option can be used to search the chemicals using either chemical name or standard identifiers (CAS or PubChem). (C) Secondly, Physicochemical filter option can be used to filter the contaminants based on their physicochemical properties such as molecular weight, Log P, TPSA, number of hydrogen bond donors or number of hydrogen bond acceptors. (D) Thirdly, Chemical similarity filter can be used to filter the contaminants based on the structural similarity with respect to a query compound. (E) The screenshot shows the result page for an individual contaminant. For each contaminant, we can obtain information on structure identifiers, environmental source, chemical classification, experimental evidence, chemical-gene interaction, physicochemical properties, predicted ADMET properties and molecular descriptors. The Browse option in ExHuMId can be used to obtain the human milk contaminants based on: (F) Geographical location of samples, (G) Maternal factors associated with samples, (H) Environmental source classification, and (I) Chemical classification.

6.3 Geographical distribution of compiled chemicals in ExHuMId across Indian states

The distribution of samples collected in the 36 published studies compiled in ExHuMId across different states of India shows that Delhi accounts for the maximum number (8) of published studies followed by Tamil Nadu with 7 published studies (Figure 6.2A). An analysis of the number of human milk contaminants detected across the samples for each state reveals that the maximum number (66) of contaminants were detected in samples from Tamil Nadu followed by 61 contaminants detected in samples from West Bengal (Figure 6.2A). None of the 101 human milk contaminants were detected in each of the 13 states captured in ExHuMId. However, 2 of the 101 human milk contaminants namely, β-Hexachlorocyclohexane (CAS:319-85-7) and Lindane (CAS:58-89-9), were though detected in 12 out of the 13 states captured in ExHuMId [39]. Figure 6.2A shows the distribution of samples collected from each state in India across the curated set of 36 published articles (Supplementary Table S6.1), and number of chemicals or contaminants detected in each state across the samples.

6.4 Chronological analysis of published studies compiled in ExHuMId

Within the curated set of 36 published studies compiled in ExHuMId, the earliest study is from 1981 while the latest study is from 2018. Furthermore, Figure 6.2B presents a chronological analysis of the 36 published studies in five-year intervals. It is seen that the maximum number (8) of published studies are from the period 2011-2015 followed by 7 published studies from the period 2006-2010 (Figure 6.2B). Figure 6.2C displays a chronological analysis of the cumulative number of contaminants detected across published studies in different time periods [39]. It is seen that there is a significant increase in the cumulative number of contaminants from published studies after 2000 and 2010 (Figure 6.2C).

6.5 Comparison of ExHuMId with other resources on human milk exposome

The presence of environmental chemicals in human milk can cause infant exposure to these chemicals, and we here refer to these chemicals as the Exposome of Human Milk (ExHuM). In order to analyze the environmental chemicals found in human milk, we have considered data from 3 sources. The chemicals in our resource, 'ExHuMId' (Exposome of Human Milk across India), have been considered for their specificity to India. The chemicals studied by Lehmann *et al*. [286] have been considered for their specificity to USA, and we refer to this chemical space as 'ExHuMUS' (Exposome of Human Milk across USA). Several human milk contaminants are also compiled in Exposome-Explorer [24], and these are not specific to any geography, and we refer to this chemical space as 'ExHuM Explorer'. Notably, there are 127 and 183 chemicals, compiled from 44 and 31 published research articles, in ExHuMUS and ExHuM Explorer, respectively (Supplementary Table S6.2). Note that the data compiled in ExHuMUS and ExHuM Explorer are

not reflective of the entire US and global populations, respectively. However, given that they are the only compilations of human milk contaminants for geographies outside India, we have considered them in this work. The union of the above-mentioned three datasets gives us a list of environmental chemicals detected in human milk samples from various parts of the world, and we refer to this chemical space as 'Global ExHuM' (Supplementary Table S6.2). The intersection of ExHuMId, ExHuMUS and ExHuM Explorer (Figure 6.2G; Supplementary Table S6.2) contains 44 chemicals that are of potential concern in the Indian, USA and global scenarios, and we refer to this space of 44 chemicals as the 'Common ExHuM' (Figure 6.2G; Supplementary Table S6.2) [39].

Table 6.1 presents a detailed comparison of our resource ExHuMId with the other two resources on human milk contaminants. Note that the three resources, ExHuMId, ExHuMUS and ExHuM Explorer, do not have in common any published experimental evidence or literature as the resources compile data on different geographies. Further, the research article [286] on ExHuMUS provides the list of detected chemicals, their concentrations and the geographical location within USA from where the study samples were collected. However, the ExHuMUS publication is not accompanied by an online resource and the meta-analysis article offers limited information for the compiled list of human milk contaminants [286]. In contrast, ExHuM Explorer [24] contains detailed information on 183 contaminants which were detected in human milk samples collected across several countries. Specifically, ExHuM Explorer gives information on the 2D and 3D structures of the contaminants [24]. Notably, our resource ExHuMId compiles the different types of information in ExHuMUS and ExHuM Explorer on chemicals, and further, compiles the list of maternal factors that influence the transfer of the contaminants to human milk, their physicochemical properties, their target genes (including visualization of the chemical-gene or chemical-protein interactions), in comparison to the two other resources (Table 6.1). In sum, ExHuMId compiles information on human milk contaminants in the specific context of India, and further, makes the compiled information easily accessible to researchers via a user-friendly web interface.

6.6 Analysis of human milk contaminants with substances of concern or in use

A better understanding of the nature and exposure sources of the human milk contaminants will most likely help direct further research and regulatory efforts. We decided to perform a detailed analysis of the chemicals in ExHuM that are of potential concern across the world, with three categories of chemical substances in use or of concern, as described below. The lists of chemical substances employed for this analysis have been described in detail in our recent work [36] (Supplementary Table S6.3).

6.6.1 Hazardous substances in human milk

EDCs, carcinogenic substances, neurotoxins and prohibited substances have all been identified as hazards, and have been well-studied for their adverse effects. Mitigating the risk posed by these substances will involve identifying their common sources, monitoring and regulating them on a timely basis. Here, we focused on identifying substances in ExHuMId that are endocrine disruptors, carcinogens or neurotoxins. These three categories of chemicals are of particular concern due to their potential to affect development and leave behind long-term effects.

Specifically, we have considered four substance lists in this category for analysis of human milk contaminants. Firstly, to understand the presence of endocrine disruptors, we used the list of 792 potential EDCs from DEDuCT 2.0 [35, 36] for this analysis. Secondly, we considered the list of carcinogens from IARC monographs [296]. Thirdly, we considered two lists of neurotoxins from the CompTox chemistry dashboard [265] of US EPA, which are: (a) chemicals demonstrating effects on neurodevelopment (DNTEFFECTS) [61] and (b) chemicals triggering developmental neurotoxicity *in vivo* (DNTINVIVO) [62]. Fourthly, we have considered a chemical regulation, namely, the EU list of substances prohibited in cosmetic products [141]. In addition, we have also consid-

ered two lists of chemicals which are known to be produced in high volume: (a) United States High Production Volume (USHPV) database, and (b) Organisation for Economic Co-operation and Development (OECD) High Production Volume (OECD HPV) list last updated on 2004.

Comparing ExHuMId with resources for the above chemical categories revealed the following. We found that 43 potential EDCs are present in ExHuMId (Supplementary Table S6.3). The web interface of ExHuMId provides detailed information on environmental sources of these EDCs detected in human milk samples [62]. The IARC monographs classify carcinogenic substances into: (a) class 1 that are carcinogenic to humans, (b) class 2A that are probably carcinogenic to humans, (c) class 2B that are possibly carcinogenic to humans, and (d) class 3 that are not classifiable as to its carcinogenicity to humans [296]. Our comparative analysis revealed that 23 carcinogens were in ExHuMId of which 7 carcinogens belong to class 1, 4 to class 2A, 5 to class 2B and 7 to class 3. Six commonly found carcinogens listed by IARC were found in the Common ExHuM and have been detected in human milk samples from India, USA, and other parts of the world. Among these, there are 3 class 1 carcinogens, namely, 2,3,4,7,8-Pentachlorodibenzofuran, 3,4,5,3',4'-Pentachlorobiphenyl (PCB-126) and Lindane (Supplementary Table S6.3). Neurotoxins in human milk are a significant concern since they are capable of influencing neurodevelopment during the prenatal and postnatal stages [64]. We found 14 potential neurotoxins to be present in ExHuMId (Supplementary Table S6.3). Cosmetic products are a significant source of exposure to various substances, due to their ubiquitous nature and widespread use. On comparison, we found 16 prohibited cosmetic ingredients (under EU regulations) to be present in ExHuMId (Supplementary Table S6.3). Among these, 3 prohibited cosmetic ingredients, namely, Hexachlorobenzene, Chlorophenothane (DDT) and Lindane are also produced in high volume (Supplementary Table S6.3) [39].

6.6.2 Substances manufactured or regulated in India

We have built ExHuMId with the purpose of compiling and understanding the published data on human milk contaminants from samples specific to India. To obtain a deeper understanding of the contaminants in ExHuMId, we have considered lists that reflect either chemical regulation in India or chemical production scenario in India. Such an analysis is in line with the main focus of this work, that is, Exposome of Human Milk across India. Specifically, we have considered the following lists compiled by relevant departments of Government of India: (a) Production of major chemicals year-wise in India [297], (b) List of banned pesticides in India [298], (c) Schedule 1 hazardous chemicals list in India [299], and (d) Schedule 3 hazardous chemicals list in India [300]. A comparative analysis of ExHuMId with lists of chemicals manufactured in India and lists from Indian chemical regulations, can further clarify the status of human milk contamination in India [62].

Several major chemicals manufactured in India have been detected in ExHuMId. Apart from this, 15 substances identified as hazards in Indian chemical regulations are present in ExHuMId, of which 9 are produced in high volume (Supplementary Table S6.3). 3 of these 15 major chemicals, namely, Decabromobiphenyl ether, Chlorophenothane (DDT), Lindane, are also present in Common ExHuM (Supplementary Table S6.3). Further, 9 banned pesticides are also present in ExHuMId. 2 banned pesticides, namely, Chlorophenothane (DDT) and Lindane, are also present in the Common ExHuM, having been detected in human milk samples from USA and other parts of the world [24, 286] (Supplementary Table S6.3). Further monitoring on the regulatory front and research on the healthcare front may be necessary to mitigate the potential adverse effects of these substances to mother and infants [39].

6.6.3 Substances contaminating human milk through possible everyday exposure

Humans come into contact with a variety of substances in daily life, particularly via the usage or consumption of an increased number and variety of processed products in today's world. This is a significant factor in the case of a pregnant woman or breast-feeding mother, since several of these substances may find their way into the mother's milk [66, 67, 285]. A concern and consideration of this study was to better understand the scenario whereby chemicals encountered in everyday life make their way into human milk. For this, we have considered two lists of substances found in food: (a) FooDB [301], and (b) the Joint FAO/WHO Expert Committee on Food Additives (JECFA) list [140]. We found 12 food additives are present in ExHuMId (Supplementary Table S6.3) [39].

6.7 Analysis of physicochemical properties of human milk contaminants

Lipophilic chemicals can be transferred to human milk from maternal plasma via passive diffusion [68–72]. The Milk to Plasma (M/P) concentration ratio is generally used to identify the equilibrium concentration of chemicals in maternal plasma and breast milk [68, 71, 72], and can indicate propensity of the environmental contaminants to enter human milk. However, the M/P ratio, while easily available for drugs, is scarcely available for environmental contaminants [70]. There is substantial evidence suggesting that the transfer of xenobiotics into human milk is influenced by the physicochemical properties of the chemicals [68–72]. The key physicochemical properties that influence the transfer of environmental chemicals into human milk are the Log P, Topological Polar Surface Area (TPSA), the number of hydrogen bond donors (HBD), the number of hydrogen bond acceptors (HBA), the number of rotatable bonds, and molecular weight [68–70, 72]. Due to the unavailability of experimentally determined M/P ratio for the 101 chemicals

compiled in ExHuMId, we performed a comparative analysis of their physicochemical properties with those of chemicals for which the M/P ratio is available. Specifically, we considered the M/P ratios for a list of 375 chemicals compiled by Vasios *et al.* [72] from published literature, and compared the computed physicochemical properties of chemicals in ExHuM with those compiled by Vasios *et al.* The physicochemical properties of the chemicals in ExHuM or Vasios *et al.* were computed using RDKit [179].

Following Vasios *et al.* [72], we have considered the chemicals with M/P ratio ≥ 1.0 as high risk and chemicals with M/P ratio < 1 as low risk for transfer to human milk from maternal plasma. For a more detailed analysis, we have further divided the low risk compounds in Vasios *et al.* based on their M/P ratios into < 1, ≤ 0.75, ≤ 0.5 and ≤ 0.25 resulting in 249, 213, 170 and 114 chemicals, respectively. Thereafter, a comparison of the physicochemical properties was made across the sets of human milk contaminants in ExHuMId, ExHuMUS and ExHuM Explorer, high risk compounds in Vasios *et al.* [72] with M/P ratio ≥ 1, and low risk compounds in Vasios *et al.* [72] with M/P ratio < 1, ≤ 0.75, ≤ 0.5, ≤ 0.25 (Figure 6.4; Supplementary Table S6.4).

Figure 6.4 shows the mean and standard deviation of the distributions of 6 physicochemical properties, namely, Log P, TPSA, number of rotatable bonds, number of HBD, number of HBA and molecular weight, for chemicals in different sets. We report the mean, standard deviation, minimum value and maximum value for the 6 physicochemical properties for the sets of human milk contaminants in ExHuMId, ExHuMUS, ExHuM Explorer, high risk compounds in Vasios *et al.* [72] with M/P ratio ≥ 1, and low risk compounds in Vasios *et al.* [72] with M/P ratio < 1, ≤ 0.75, ≤ 0.5, and ≤ 0.25 (Supplementary Table S6.5). We find that the mean and standard deviation of the distributions of 6 physicochemical properties for human milk contaminants in ExHuMId are much closer to those for high risk compounds in Vasios *et al.* [72] with M/P ratio ≥ 1 [39]. Note that the high risk compounds in Vasios *et al.* [72] are capable of easily transferring to human milk if they are present in the lactating mother's body. Further, we observed the same trend for chemicals in ExHuMUS and ExHuM Explorer (Figure 6.4). Figure 6.4 also shows a

clear difference between the mean and standard deviation of the distributions of the above 6 physicochemical properties for the low risk compounds in Vasios *et al*. [72] in comparison to high risk compounds or human milk contaminants in ExHuMId, ExHuMUS and ExHuM Explorer.

Of the 6 computed physicochemical properties, the mean lipophilicity (Log P) of human milk contaminants is much higher than chemicals with low risk in Vasios *et al*. [72]. For example, the mean Log P of chemicals in ExHuMId is 5.9 ± 2.3 in comparison to 2.4 ± 3.1 for low risk chemicals with M/P ratio < 1 in Vasios *et al*. [72]. Moreover, the mean number of HBA, HBD, and rotatable bonds are much lower for human milk contaminants than chemicals with low risk in Vasios *et al*. [72]. Also, the mean TPSA of human milk contaminants is much lower than chemicals with low risk in Vasios *et al*. [72]. In contrast, there is no clear difference between mean molecular weight for human milk contaminants and chemicals with low risk in Vasios *et al*. [39]. In sum, our observations confirm previous observations [68–72] on physicochemical properties of chemicals with high risk of transfer to human milk from maternal plasma.

Overall, these results give insights into the effect physicochemical properties can have in the transfer of environmental chemicals into human milk, and further, can enable the prediction of such chemicals. While predicting the possible transfer of environmental chemicals into human milk based on physicochemical properties, it is important to bear in mind the due limitations of any such method that does not account for the influence of maternal factors, frequency of exposures, varying pharmacokinetic properties of contaminants, and the complexity of lactation pathways [66, 67, 286, 287, 302].

6.8 Analysis of potential effects of contaminants on maternal and infant health

Though the benefits of breastfeeding outweigh the risk of these environmental chemicals, the effect of these chemicals on mother and infant health remains poorly understood [66,

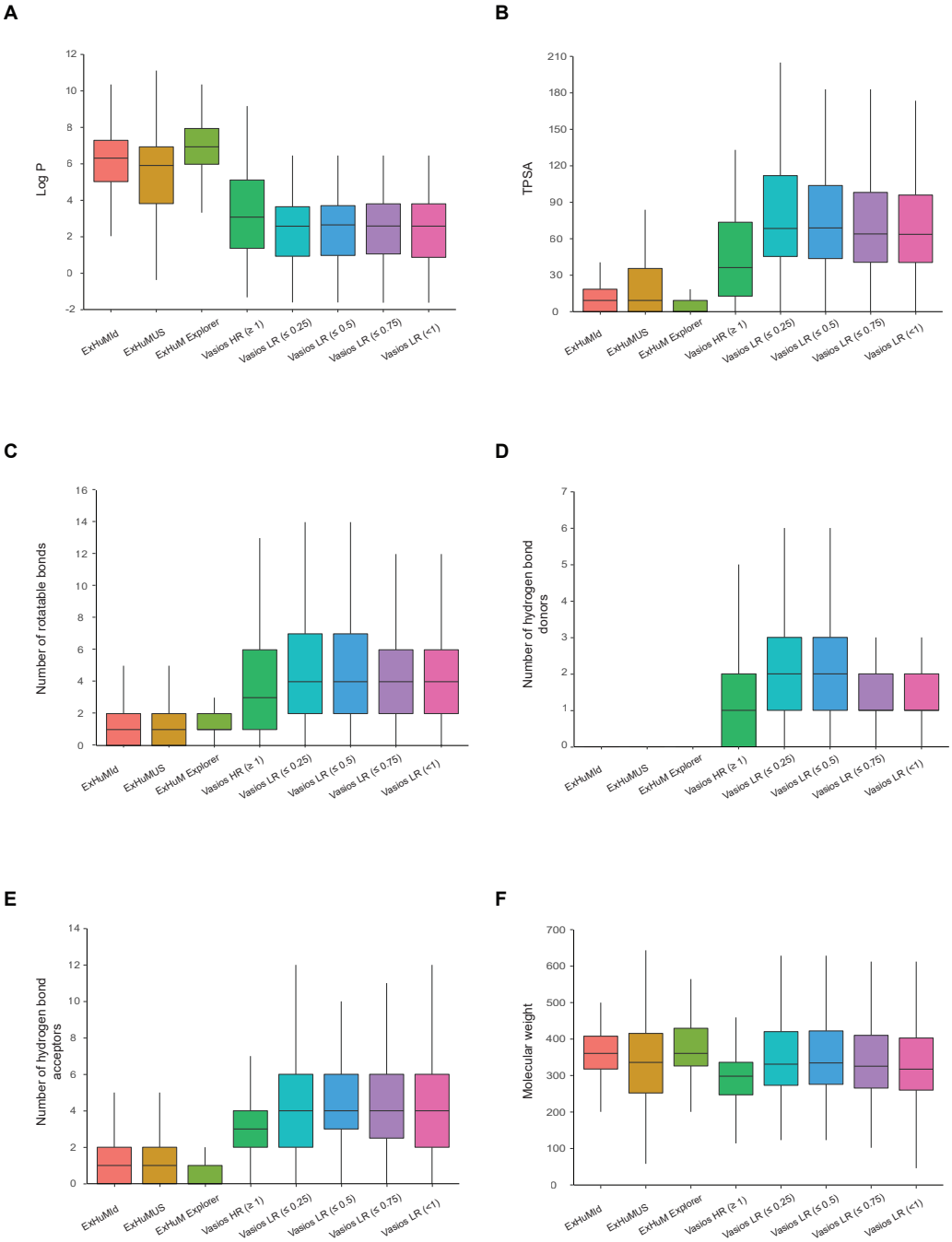


Figure 6.4: Box plots displaying the distributions of 6 physicochemical properties: (A) Log P, (B) TPSA, (C) number of rotatable bonds, (D) number of hydrogen bond donors (HBD), (E) number of hydrogen bond acceptors (HBA), and (F) molecular weight, for chemicals in 8 different sets, namely, human milk contaminants in ExHuMId, ExHuMUS, ExHuM Explorer, high risk compounds in Vasios *et al.* with M/P ratio ≥ 1 (Vasios HR ≥ 1), and low risk compounds in Vasios *et al.* with M/P ratio < 1 (Vasios LR < 1), M/P ratio ≤ 0.75 (Vasios LR ≤ 0.75), M/P ratio ≤ 0.5 (Vasios LR ≤ 0.5), and M/P ratio ≤ 0.25 (Vasios LR ≤ 0.25). Note that, the distributions for the number of HBD in subfigure (D) are not visible for chemicals in ExHuMId, ExHuMUS and ExHuM Explorer as the mean and standard deviation for each of the three sets is very close to 0.

67,285]. Hence, we were motivated to perform the following analysis to explore the effect of human milk contaminants on mother and child. Using systems biology approach, we provide another perspective from our analysis by predicting the effect of environmental contaminants on lactation, cytokine signalling and production pathways, and xenobiotic transporters with the help of existing large-scale toxicological resources such as ToxCast and CTD [30, 89].

6.8.1 Identifying the target genes of contaminants

To identify the target human genes or proteins of the chemicals in Global ExHuM, we have used two well-known toxicology resources, ToxCast [89] and CTD [30].

We have used the ToxCast invitroDB3 dataset released in August 2019 [215] to retrieve the list of target genes or proteins of human milk contaminants in the Global ExHuM. We followed the method described in Section 2.4.2 to extract from ToxCast the human target genes perturbed upon exposure to human milk contaminants in the Global ExHuM. Thereafter, we also retrieved from CTD the list of target genes or proteins of chemicals in the Global ExHuM using specific filters. In CTD, we have considered only the chemical-gene or chemical-protein interactions specific to humans and those interactions which have at least one evidence in published scientific literature. Moreover, in CTD, we have considered only binary interactions involving one chemical and one gene [30], and thus, have filtered out complex interactions. In CTD, we have also not considered the interactions that contained the terms 'Chemical abundance' or 'Response to substance' based on their 'interaction actions'.

Of the 101 human milk contaminants in ExHuMId, information on target genes or proteins is currently available in ToxCast and CTD for 39 and 53 chemicals, respectively. The ExHuMId web interface provides this information on target genes or proteins for different human milk contaminants [39].

6.8.2 Identification of contaminants interacting with lactation relevant genes

Women exposed to environmental contaminants during early stages of pregnancy or lactation have been shown to preferentially store several persistent lipophilic chemicals in their adipose tissue [67], and subsequently during lactation such contaminants can transfer to infants via breastfeeding [67, 70, 286, 287, 294, 303]. In recent times, there have been significant advances in the understanding of lactation physiology and its pathways at a molecular level [304, 305] but the effect of environmental contaminants on physiology and health of mother and infant needs further attention [66, 67, 285]. Specifically, environmental chemicals are known to affect the lactation period [306] and the milk secretion [303] but the underlying molecular mechanisms by which these contaminants affect lactation physiology and milk secretion remains to be understood. These reported effects on lactation motivated us to investigate if any of the 101 human milk contaminants in ExHuMId can interfere with the genes involved in the pathways associated with lactation.

Prolactin [307] and oxytocin [308] are the major hormones responsible for lactation. Therefore, we have considered the signalling pathways associated with these hormones for this analysis. We compiled the set of genes involved in the prolactin and oxytocin signalling pathways in humans from NetPath [309–311] and Kyoto Encyclopedia of Genes and Genomes (KEGG) [312]. NetPath compiles a list of genes involved in prolactin and oxytocin signalling pathways in mammals, while the genes retrieved from KEGG are specific to humans. Further, we mapped these genes to their respective human NCBI Entrez identifiers. In this step, we obtained 181 and 237 genes involved in prolactin and oxytocin signalling pathways, respectively, from the above two resources. In addition to these pathways, we have included a set of 14 differentially expressed genes from Lemay *et al*. [304] that are involved in lactose synthesis pathways and important for milk production. Using ToxCast and CTD, we then identified chemicals from ExHuMId that may interact with these lactation relevant genes (Figures 6.5 and 6.6; Supplementary Table S6.6). More-

over, we have also performed the same analysis for chemicals in ExHuMUS and ExHuM Explorer (Supplementary Table S6.6).

We found 46 human milk contaminants compiled in ExHuMId target 83 genes out of 181 genes associated with the prolactin signalling pathway (Figure 6.5A; Supplementary Table S6.6). In the case of oxytocin signalling pathway, 118 out of 237 pathway-associated genes, are found to be the targets of 50 human milk contaminants compiled in ExHuMId (Figure 6.6A; Supplementary Table S6.6). Arsenic targets 48 genes of prolactin signalling pathway and 74 genes of oxytocin signalling pathway. The ESR1 (Estrogen Receptor 1), which is associated with both the oxytocin and prolactin signalling pathways, appears to be a common target, having interactions with the highest number of human milk contaminants (Figures 6.5A and 6.6A; Supplementary Table S6.6). Through the analysis of the genes responsible for the production of lactose, as reported by Lemay *et al*. [304], we find that arsenic perturbs lactose synthesis pathway via 4 genes, namely, GALK1, HK1, NME1-NME2 and SLC2A9 (Figure 6.5B; Supplementary Table S6.6) [39]. We have performed the same analysis for the chemicals compiled in ExHuMUS and ExHuM Explorer and their results are included in Supplementary Table S6.6.

6.8.3 Identification of contaminants interacting with cytokine signalling and production relevant genes

Environmental contaminants transferring to human milk were found to be potentially harmful to the development of newborns, due to their ability to disrupt the signalling pathways of infant development [64, 65, 313]. Here, we have investigated the effects of human milk contaminants on the immune system development in infants.

It is known that human milk contains several immunological factors including cytokines, chemokines, immunoglobulins, and other soluble receptors that can confer immunity in the lactating infants [314, 315]. Among these immunological factors, cytokines

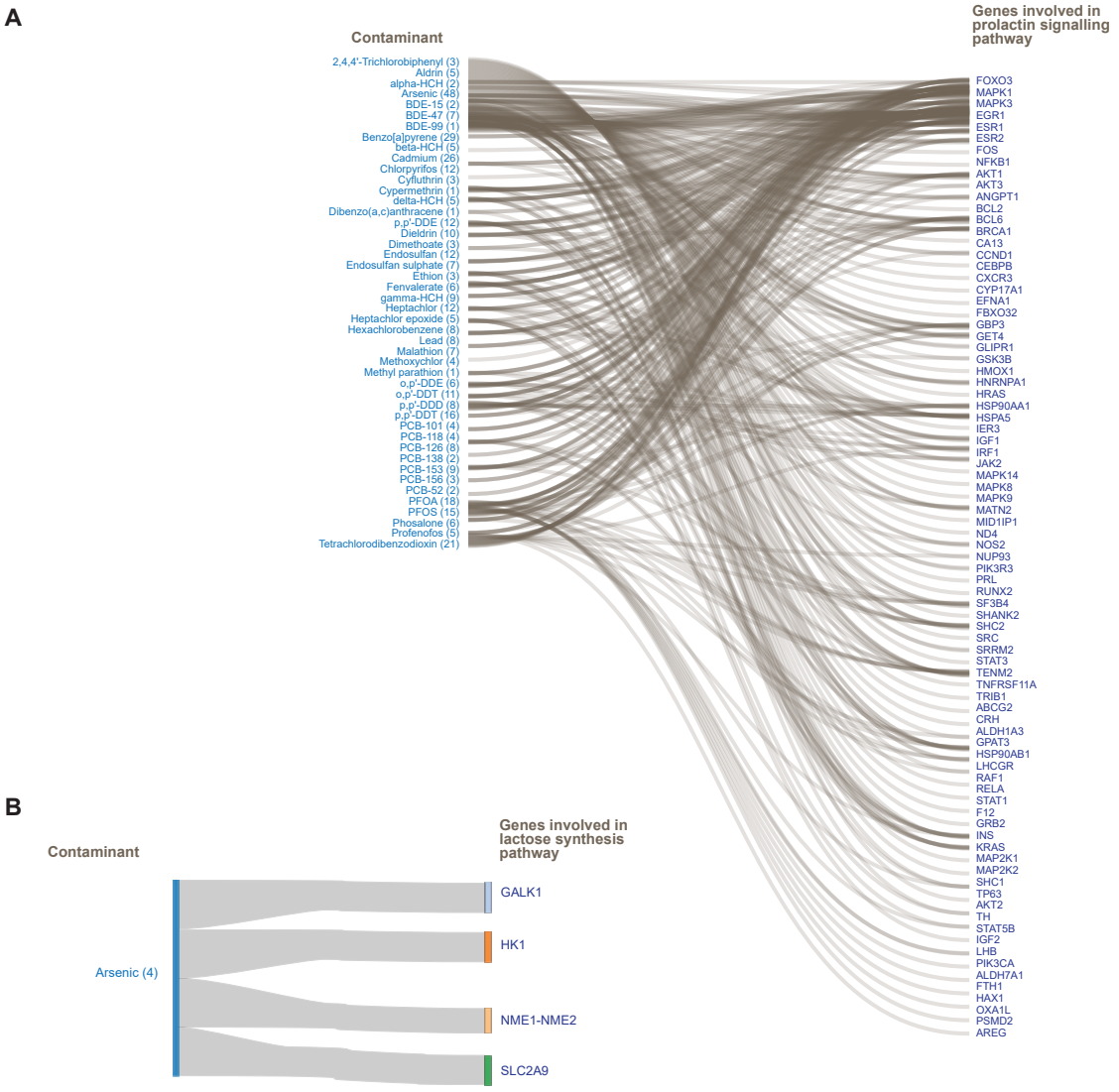


Figure 6.5: Sankey plots show the human milk contaminants in ExHuMId and their target genes or proteins involved in the pathways affecting lactation: (A) Prolactin signalling pathway, and (B) Lactose synthesis pathway. Besides each contaminant, the number of target genes is mentioned in parenthesis.

play a vital role in the regulation of specific and non-specific immune responses [302]. Cytokines bind to the cytokine receptors and trigger the production of cytokines or elicit the immune response via the activation of cytokine signalling pathways [316]. Notably, the presence of environmental contaminants in human milk can interfere with cytokine signalling and production [302, 317], thereby influencing the effective immune response in developing infants [64, 302, 313, 317]. Thus, we aimed to identify chemicals in the Global ExHuM that could potentially disrupt cytokine signalling pathways.

To this end, we first compiled the list of cytokine receptor genes from Cameron *et al*. [318], HGNC database [319, 320], KEGG BRITE database [312] and Guide to Pharmacology database [321]. In total, we have compiled 116 cytokine receptors for which the chemical-gene interactions were obtained from ToxCast and CTD. Finally, we have gathered the list of cytokines specific to the cytokine receptors that are known to interact with the human milk contaminants. This resulted in a tripartite network containing contaminants or chemicals, cytokine receptors, and cytokines (Figure 6.7; Supplementary Table S6.7).

On analyzing the list of 116 cytokine receptors with chemical interactions obtained from ToxCast and CTD, we found that 22 chemicals compiled in ExHuMId interact with 32 cytokine receptors, which in turn could interfere with signalling or production of 64 cytokines (Figure 6.7; Supplementary Table S6.7). These interactions are displayed in the form of a tripartite network in Figure 6.7. Among the chemicals in ExHuMId, arsenic targets the highest number of cytokine receptors (24 genes) followed by Benzo[a]pyrene (9 genes). Among the cytokine receptors, CD40 is perturbed by 17 contaminants compiled in ExHuMId, and the binding of these contaminants to the CD40 receptor could interfere with the signalling and production of CD40LG, a cytokine specific to CD40 (Figure 6.7; Supplementary Table S6.7) [39]. Thus, human milk contaminants targeting cytokine receptors could bind to these receptors and interfere with normal function of cytokines. For the chemicals compiled in ExHuMUS and ExHuM Explorer, we have also performed the same analysis, and found several contaminants in these resources to be capable of

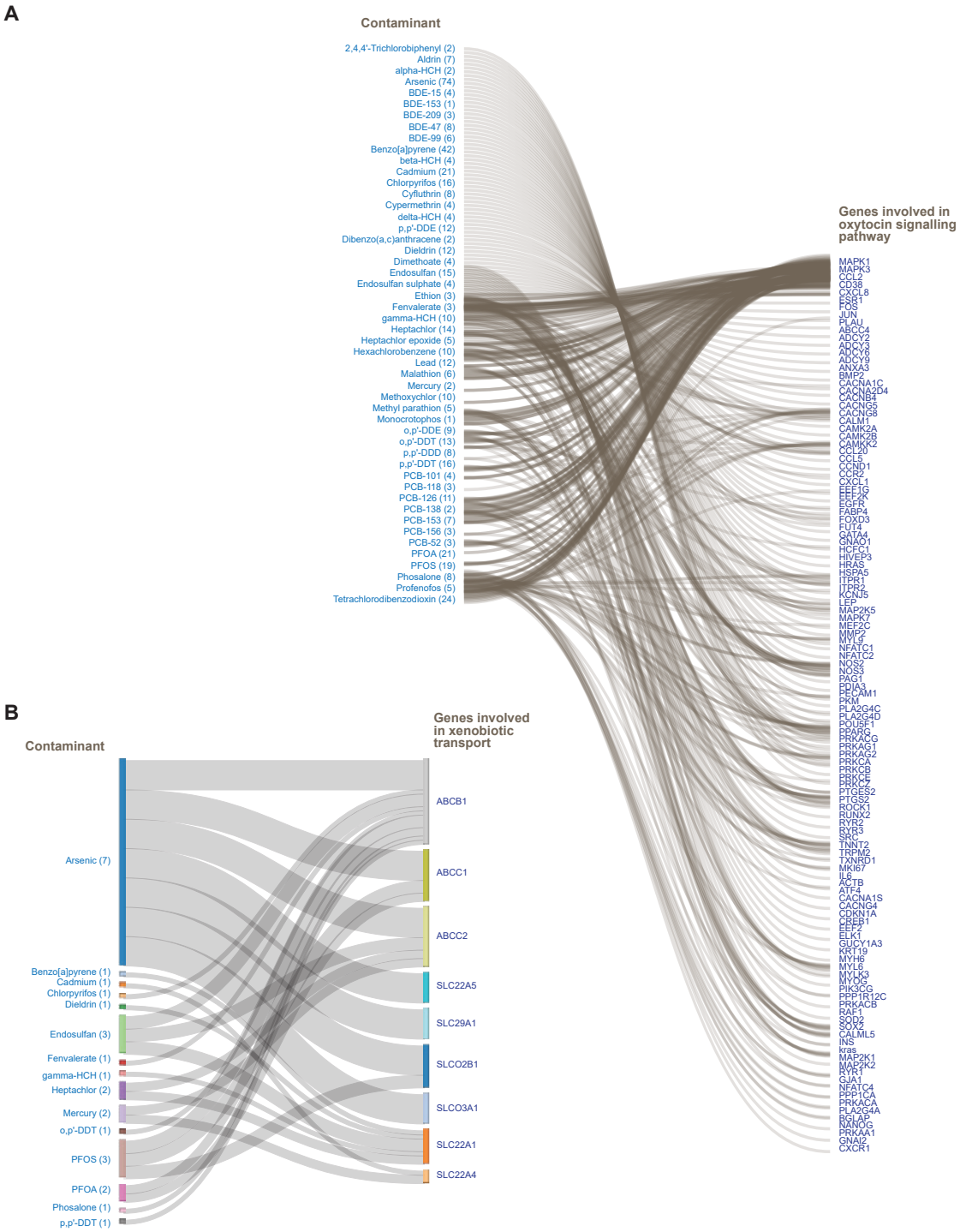


Figure 6.6: Sankey plots show the human milk contaminants in ExHuMId and their target genes or proteins involved in: (A) Oxytocin signalling pathway, and (B) Xenobiotic transporters. Besides each contaminant, the number of target genes is mentioned in parenthesis.

influencing cytokine signalling and production (Supplementary Table S6.7).

6.8.4 Identification of contaminants interacting with xenobiotic transporters

Drug or xenobiotic transporters are membrane proteins that play a major role in transfer of xenobiotics into human milk [322,323]. Some of these transporters have been found to be expressed in mammary gland during lactation [322–325]. From the study by Alcorn *et al*. [326], we compiled the list of 19 (out of 30) transporters that are expressed in the mammary gland during lactation based on their Real-Time Reverse Transcription-Polymerase Chain Reaction (RT-PCR) analysis. Thereafter, we have explored any potential interactions between the chemicals in Global ExHuM and these 19 transporters, using interaction data obtained from ToxCast and CTD (Figure 6.6; Supplementary Table S6.8).

The analysis of this dataset with chemical-gene interactions obtained from ToxCast and CTD revealed that 15 contaminants in ExHuMId target 9 transporters which are expressed during lactation (Figure 6.6B; Supplementary Table S6.8). Of these, there are two prominent transporter protein genes, namely, SLC22A1 and SLC22A4, which were found to be expressed 4-fold during lactation [326] (Figure 6.6B; Supplementary Table S6.8). Among the contaminants in ExHuMId, Arsenic targets 7 transporter genes. The ABCB1 transporter protein gene appears to be targeted by the maximum number of contaminants in ExHuMId (Figure 6.6B; Supplementary Table S6.8) [39]. We have also performed the same analysis for the chemicals compiled in ExHuMUS and ExHuM Explorer, and these results are included in Supplementary Table S6.8.

From the analysis reported in this section, it is evident that the human milk contaminant Arsenic can target several genes or proteins in lactation pathway, cytokine signalling and production pathway, and xenobiotic transporters (Figures 6.5, 6.6 and 6.7). Based on the compilation of studies in ExHuMId, Arsenic was detected in human milk samples collected from 3 states of India, namely, Chhattisgarh, Maharashtra and West Bengal.

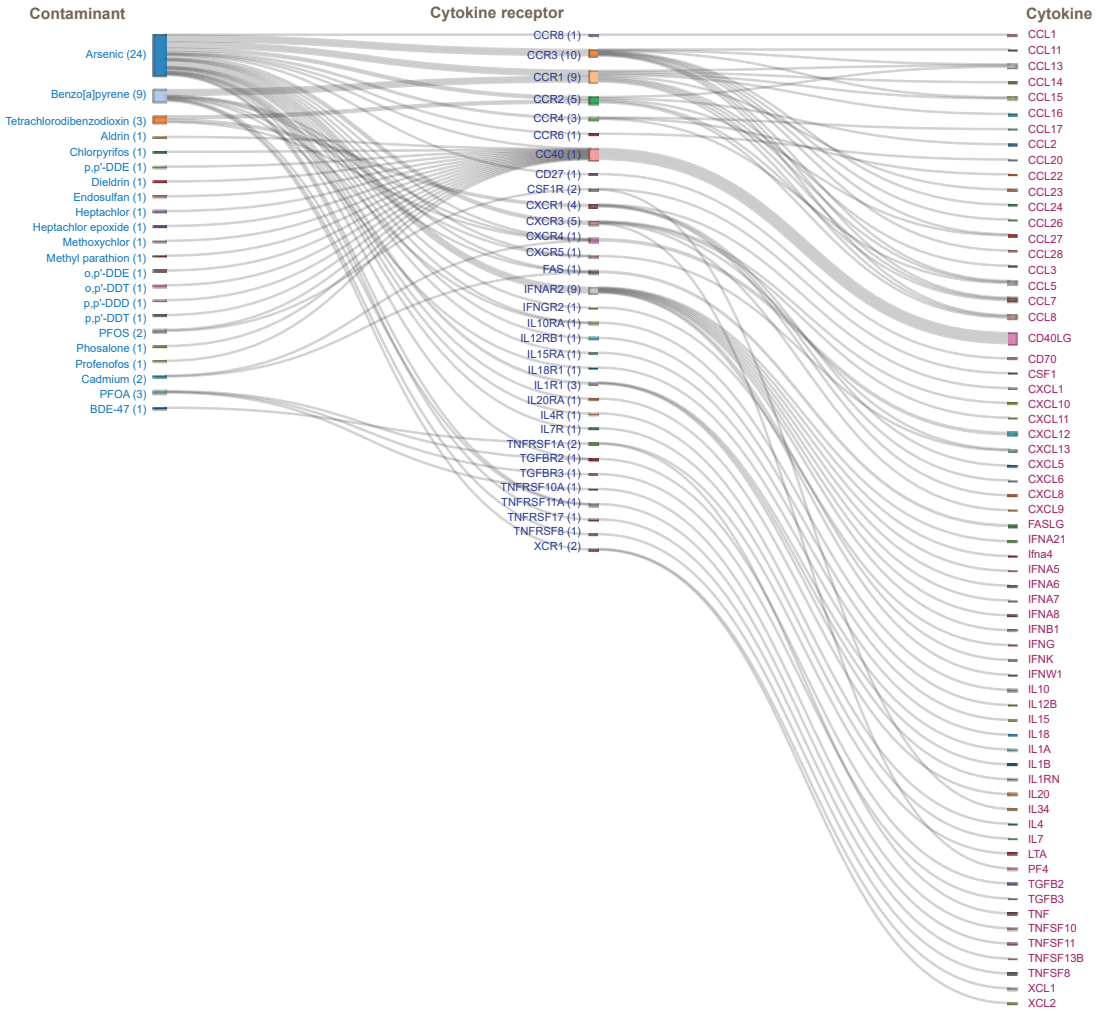


Figure 6.7: Sankey plot shows the tripartite network of human milk contaminants in ExHuMId, their target genes or proteins corresponding to cytokine receptors, and the cytokines regulated by the specific cytokine receptors. Besides each contaminant, the number of target cytokine receptors is mentioned in parenthesis, and similarly, besides each cytokine receptor, the number of cytokines regulated is mentioned in parenthesis.

Arsenic was also found in a human milk sample from the United States, as reported in Lehmann *et al*. [286]. From the evidence in scientific literature, Arsenic has been found to be present in many biospecimens from across the world [327]. Especially, the primary source of Arsenic is known to be ground water or drinking water [328, 329]. Moreover, there are several studies which have reported on Arsenic contamination in ground water and drinking water samples collected from several states in India [330–333]. Thus, it is not surprising that Arsenic has been found to be a human milk contaminant.

6.9 Discussion

Human milk is the sole source of nourishment for infants for the first few months of their lives, during which exposure to environmental contaminants is a concern. These contaminants may have an impact on maternal health and lactation as well. Understanding the effects of these environmental contaminants to maternal and infant health remains challenging [66, 67, 285]. In recent years there is an increased interest towards the development of an integrated approach in toxicology known as the exposome which captures all the environmental exposures of humans during their lifetime, their associated biological responses, and the implications of the exposures on their health [13, 18–20]. In this work we have developed a comprehensive resource on Exposome of Human Milk across India, ExHuMId version 1.0, through a systematic approach.

The development of a resource on human milk exposome specific to India is the first step in covering the wide range of information related to detected human milk contaminants, their concentrations, maternal factors, and other information which are dispersed across a large body of scientific literature. The determination of mean concentrations of contaminants or any established benchmarks like reference dose (RfD) or Tolerable Daily Intake (TDI) or Average Daily Dose (ADD) is not ventured into in this chapter, as the data compiled in this work is diverse in consonance with the breadth of the Indian population. It is important to highlight the availability of guidelines provided by the US EPA

on child-specific exposure scenarios examples [334] in the Indian context, which can help to estimate the above benchmarks specific to India. During our literature mining we also found thousands of research articles available in the corpus of PubMed [158], on the detection of environmental contaminants in human milk across the world. Thus, the expansion of human milk exposome resources worldwide, and the availability of experimentally determined M/P ratio for environmental contaminants can help in better risk assessment and management of human milk contaminants. Importantly, further studies are necessary to understand the influence of variable factors such as maternal factors [67, 71, 287], the pharmacokinetics of environmental contaminants [71, 286], and the complexity of lactation pathways and physiology [287, 313] in order to incorporate these variables in the risk estimation of human milk contaminants. We also note that there are several studies on detection of environmental contaminants in other specimens such as blood, plasma, serum, placenta, urine, saliva across India, and substantial manual effort is required to develop a comprehensive exposome resource specific to India which is beyond the scope of this work. In future, we would like to contribute further towards mapping the external exposomes specific to India.

Supplementary Information

Supplementary Tables S6.1-S6.8 associated with this chapter are available for download from the GitHub repository: https://github.com/asamallab/PhDThesis-Janani_R/blob/main/SI/ST_Chapter6.xlsx.

Feature	ExHuMId	ExHuMUS	ExHuM Explorer
Number of human milk contaminants	101	127	183
Number of published research articles covered	36	44	31
Web interface	Yes	No	Yes
Compilation of concentration of human milk contaminants	Yes	Yes	Yes
Compilation of maternal factors from the experimental data	Yes	No	No
Categorization of contaminants based on environmental source	Yes	No	No
Chemical classification of contaminants	Yes	No	Yes
Standard chemical identifiers of contaminants	Yes	No	Yes
Availability of 2D structure for contaminants	Yes	No	Yes
Availability of 3D structure for contaminants	Yes	No	Yes
Downloadable formats for 2D and 3D structure of contaminants	SDF, MOL, MOL2, PDB, PDBQT	No	MOL, SDF, PDB
Physicochemical properties of contaminants	Yes	No	No
Molecular descriptors for contaminants	Yes	No	No
Predicted ADMET properties of contaminants	Yes	No	No
Chemical-gene association network	Yes	No	No
Chemical similarity filter	Yes	No	No

Table 6.1: Comparison of the features including meta-information captured in ExHuMId with respect to two other resources, ExHuMUS and ExHuM Explorer, on human milk contaminants.

Chapter 7

FCCP: A repository of fragrance chemicals in children's products

Apart from breast milk, infants are also exposed to environmental chemicals in food, indoor air, child care products and toys, which are part of the external exposome of children [80, 81, 148, 335, 336]. Exposure to hazardous chemicals is a significant health concern for children who have high metabolic rate, immature organ systems, thin skin, rapid growth and development of organs and tissues [79–81]. Notably, children are exposed to chemicals in toys and different child care products related to feeding, diapering, bathing and clothing [81, 335, 337]. With respect to chemicals in children's products, the toxic effects of heavy metals, phthalates and brominated flame retardants have been well studied [80, 81, 148, 335, 336, 338]. There are also regulations in some parts of the world that limit the use of hazardous chemicals in children's products. However, fragrance chemicals which are a subset of chemicals used in children's products remain either self-regulated or poorly regulated [75, 79, 81]. Moreover, there is a lack of an overarching international approach for the global regulation of chemicals (including fragrances) in children's products [336].

Fragrance chemicals in terms of their chemical origin are either natural or synthetic compounds, and exposure to such chemicals can lead to asthma, contact dermatitis (ir-

ritant or allergic), dyschromia, photosensitivity, and migraine headaches [73–76, 78, 86]. Further, certain fragrance chemicals used in cosmetics or personal care products were found to be carcinogens, neurotoxicants, and linked to reproductive disorders [75, 78, 339–341]. Notably, fragrance chemicals have been detected in human samples of blood, adipose tissue and breast milk [75, 339]. Exposure to these fragrance chemicals can occur via direct skin contact, inhalation, or ingestion [342, 343]. For instance, when children are exposed to fragrance chemicals found in skin care products like moisturizing lotions, soaps, or baby diapers, such chemicals may penetrate through the skin, absorbed into the bloodstream, and subsequently, distributed to various organs [339]. Given the potential health risk posed by these fragrance chemicals in early childhood, there is a need to continuously monitor and regulate such chemicals to ensure safety of children's products. In the European Union (EU), the 'EU Toy Safety Directive' [145] and the 'Danish EPA Sensitizing Fragrances in Children's Articles' [146] are two regulations that limit the use of certain fragrance chemicals in children's products. Still, there is no dedicated online repository to date that compiles the inventory of fragrance chemicals used in children's products. In this chapter, we present a comprehensive resource of fragrance chemicals detected experimentally in children's products and several analyses of the associated chemical space to highlight the need and importance of monitoring and regulating the use of such chemicals in children's products. **The work reported in this chapter is contained in the published manuscript [40].**

7.1 Compiling an atlas of fragrance chemicals in children's products

7.1.1 Literature mining and curation

As a first step towards building the database, we performed literature mining to identify experimental published studies which report or detect fragrance chemicals used in chil-

dren's products. For this, we mined PubMed [158] using the following keyword search:

(perfume* OR "odor" OR "odour" OR odorant* OR "scent" OR "scented" OR fragrance* OR "fragrant") AND ("toys" OR "toy" OR ((child* OR "baby" OR "babies") AND ("products" OR "product")))

The above keyword search which was last performed on 23 March 2021 resulted in 306 research articles from PubMed. Further, we manually curated these 306 research articles to filter the relevant articles reporting the fragrance chemicals identified in children's products. Specifically, we retained experimental studies that reported fragrance or scented compounds detected across children's products. Moreover, studies that reported chemicals other than fragrance chemicals, as well as the ones that did not include any children's products were excluded. Finally, this manual curation led to the identification of 21 research articles that contain information on fragrance chemicals from children's products like toys, moisturizing creams, shampoos, infant milk formula, and baby diapers (Figure 7.1; Supplementary Table S7.1). Of these 21 research articles, 11 publications reported fragrance chemicals identified in 'toys' [40]. The steps involved in the filtration of the 306 research articles to compile experimental studies that have detected fragrance chemicals in children's products are described in a flowchart based on the preferred reporting items for systematic reviews and meta-analyses (PRISMA) [344] (Figure 7.1).

7.1.2 Compilation, unification and classification of fragrance chemicals

From the filtered set of 21 research articles, we next compiled the list of detected fragrance chemicals, along with the source or types of children's products in which the chemicals were identified. For unambiguous analysis of the fragrance chemicals compiled in this dataset, we further mapped the chemicals to their standard chemical identifiers using CAS [164] and PubChem [86]. This process led to the compilation of 153 unique fragrance chemicals from the filtered set of 21 research articles (Supplementary Table S7.2).

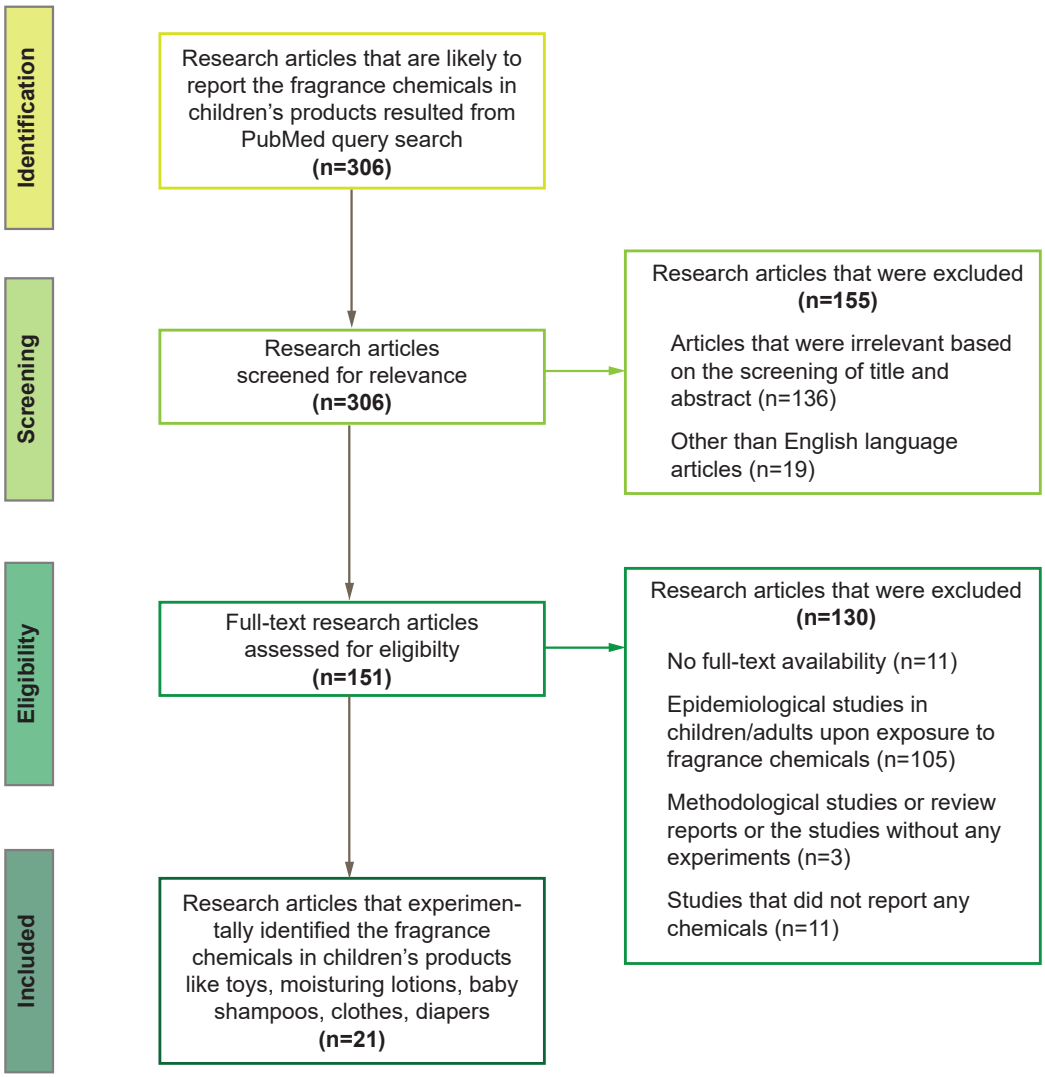


Figure 7.1: The flowchart depicting the steps involved in the selection of published research articles that are used to compile the fragrance chemicals experimentally detected in children's products.

Thereafter, using PubChem [86] database, we gathered two-dimensional (2D) and three-dimensional (3D) structure information, IUPAC name, canonical SMILES, InChI, and InChIKey for the 153 fragrance chemicals compiled in this dataset [40].

Subsequently, the 153 fragrance chemicals were classified based on: (a) chemical structure, (b) children's product source, and (c) chemical origin. Firstly, we used ClassyFire [173, 174] to classify the 153 fragrance chemicals based on their chemical structure (Figure 7.2A). ClassyFire [174] based chemical classification of the 153 fragrance chemicals in FCCP revealed that all fragrance chemicals in this resource are 'organic'. Further, among the 153 fragrance chemicals in FCCP, 50 are 'benzenoids' and 40 are 'organic oxygen compounds' according to ClassyFire (Figure 7.2A).

Secondly, we classified the children's product source information for the fragrance chemicals obtained from the associated literature, and this resulted in 8 broad categories and 19 sub-categories (Figure 7.2B). The 8 broad categories include 'Clothing and Accessories', 'Diapering', 'Diet and Feeding', 'Hair care', 'Miscellaneous products', 'Oral care', 'Skin care', and 'Toys'. We find that 5 chemicals namely, 'Benzyl alcohol', 'Benzyl benzoate', 'Citronellol', 'Hexyl cinnamic aldehyde', and 'Linalool' were present in 5 out of 8 broad categories of children's product source. 19 sub-categories represent the standardized term for children's products studied in the published literature. For example, sub-categories such as 'clay toys' and 'plastic toys' were grouped into the broad category of 'Toys'. Of the 153 fragrance chemicals in FCCP, 85 have their children's product source as 'Toys', and moreover, these chemicals belong to 9 different sub-categories of toys (Figure 7.2C).

Thirdly, we classified the fragrance chemicals based on their origin into either 'natural' or 'synthetic' (Figure 7.2C). Based on literature search, we determined whether a fragrance chemical is a natural product (i.e., produced by microbes, plants or animals) or a synthetic chemical (i.e., man-made or artificial). Several natural chemicals are being synthesized due to increased demand. However, if there is evidence that a fragrance chemical has a natural source (e.g., plants, animals, fungi, algae, bacteria), we label it as

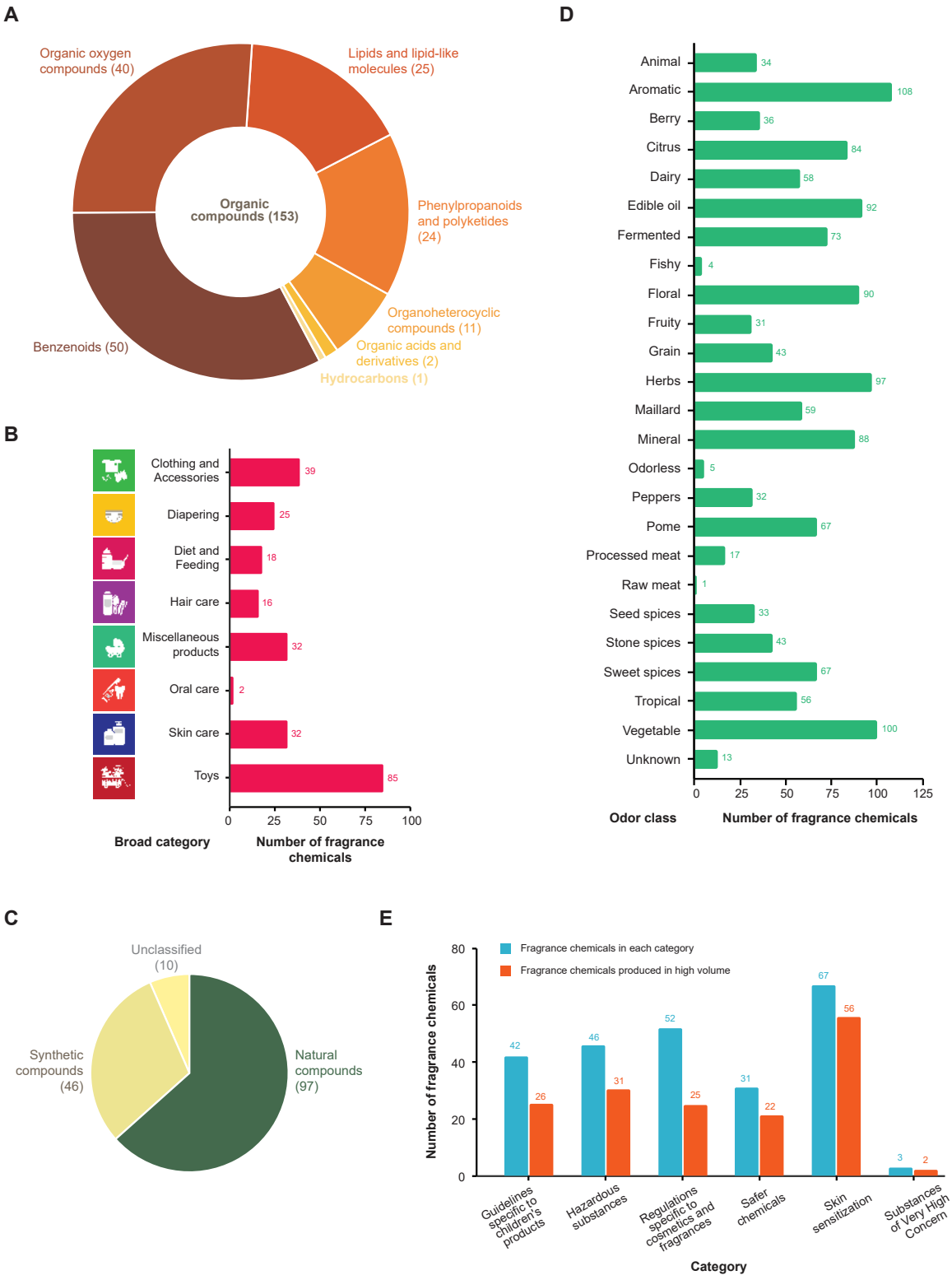

A
Organic oxygen compounds (40)
Lipids and lipid-like molecules (25)
Organic compounds (153)
Phenylpropanoids and polyketides (24)
Organoheterocyclic compounds (11)
Organic acids and derivatives (2)
Hydrocarbons (1)
Benzenoids (50)
B
Clothing and Accessories
Diapering
Diet and Feeding
Hair care
Miscellaneous products
Oral care
Skin care
Toys
39
25
18
16
32
2
32
85
0
25
50
75
100
Broad category
Number of fragrance chemicals
D
Animal
Aromatic
Berry
Citrus
Dairy
Edible oil
Fermented
Fishy
Floral
Fruity
Grain
Herbs
Maillard
Mineral
Odorless
Peppers
Pome
Processed meat
Raw meat
Seed spices
Stone spices
Sweet spices
Tropical
Vegetable
Unknown
34
108
36
84
58
92
73
4
90
31
43
97
59
88
5
32
67
17
1
33
43
67
56
100
13
0
25
50
75
100
125
Odor class
Number of fragrance chemicals
C
Unclassified (10)
Synthetic compounds (46)
Natural compounds (97)
E
Fragrance chemicals in each category
Fragrance chemicals produced in high volume
Number of fragrance chemicals
0
20
40
60
80
42
26
46
31
52
25
31
22
67
56
3
2
Guidelines specific to children's products
Hazardous substances
Regulations specific to cosmetics and fragrances
Safer chemicals
Skin sensitization
Substances of Very High Concern
Category

Figure 7.2 ***(previous page)*****:** (A) ClassyFire based classification of the 153 fragrance chemicals into 7 superclasses. The number of fragrance chemicals in each superclass is indicated within the parenthesis. (B) Histogram shows the distribution of the 153 fragrance chemicals across 8 broad categories of children's product source. (C) Classification of the 153 fragrance chemicals based on their chemical origin. The number of fragrance chemicals in each category is indicated within the parenthesis. (D) The column chart shows the distribution of the fragrance chemicals across 24 odor classes. (E) The graph shows the distribution of the 153 fragrance chemicals across different categories of chemical lists reflecting guidelines or regulations, namely, 'Guidelines specific to children's products', 'Hazardous substances', 'Regulations specific to cosmetics and fragrances', 'Safer chemicals', 'Skin sensitization', 'Substances of Very High Concern', and 'High Production Volume (HPV)' chemicals. This figure also gives the number of chemicals produced in high volume in each category.

'natural' in this compilation. According to the classification based on chemical origin, 97 fragrance chemicals in FCCP are natural compounds.

Furthermore, we compiled the odor information for the 153 fragrance chemicals from various resources including Flavornet [345, 346], FlavorDB [68, 347], The Good Scents Company Information System [348] and other published literature. Based on this compilation of the odor information, 102 odor types were known to be associated with 140 fragrance chemicals compiled in this dataset. Similar to Flavornet [346], these 102 odor types were further grouped into 24 odor classes (Figure 7.2D; Supplementary Table S7.3). Moreover, the odor profiling of the fragrance chemicals in FCCP showed that each chemical is associated with multiple odor classes (Supplementary Table S7.3). Of the 24 odor classes associated with the fragrance chemicals in FCCP, 'Aromatic' odor is found to be prevalent among 108 fragrance chemicals in FCCP, followed by the odor classes 'Vegetable' with 100 fragrance chemicals and 'Herbs' with 97 fragrance chemicals (Figure 7.2D; Supplementary Table S7.3).

7.2 Web interface of FCCP

To enable easy access to the list of 153 fragrance chemicals and associated information compiled from various sources, we created an online database, namely, FCCP, which is a repository of **F**ragrance **C**hemicals in **C**hildren's **P**roducts. FCCP is accessible online at:

https://cb.imsc.res.in/fccp [40].

The web interface of FCCP has been created using an approach similar to that described in Section 2.2. FCCP contains detailed information on the 153 fragrance chemicals and their chemical structures. Especially, users can readily download 2D and 3D structures of the fragrance chemicals in different formats such as MOL, MOL2, SDF, PDB, and PDBQT. In addition, we compiled physicochemical properties, molecular descriptors, and predicted ADMET properties for the 153 fragrance chemicals compiled in FCCP. To compute physicochemical properties and generate molecular descriptors of chemicals, we have used RDKit [179], PaDEL [180, 181] and Pybel [182]. For predicting ADMET properties of chemicals, we have used admetSAR 2.0 [183], pkCSM [184], SwissADME [186], Toxtree 2.6.1 [187] and vNN server [188]. In FCCP, users can obtain diverse information on a fragrance chemical, including 2D and 3D chemical structure, via the search and browse option in the user-friendly web interface (Figure 7.3).

7.3 Analysis of fragrance chemicals from regulatory perspective

To assess the current level of regulation of the fragrance chemicals compiled in FCCP, we performed a comparative analysis with 21 publicly available chemical lists which reflect chemical guidelines or regulations (Figure 7.2E; Supplementary Table S7.4). These chemical lists represent different categories including Guidelines specific to children's products, Regulations specific to cosmetics and fragrances, Substances of Very High Concern, Hazardous substances, Skin sensitization, and Safer chemicals.

Furthermore, we investigated the presence of High Production Volume (HPV) chemicals among the fragrance chemicals identified in the above mentioned categories. To do so, we considered 3 publicly available lists which include: (i) Organisation for Economic Co-operation and Development (OECD) High Production Volume (OECD HPV) list last updated on 2004 [150], (ii) United States High Production Volume (USHPV)

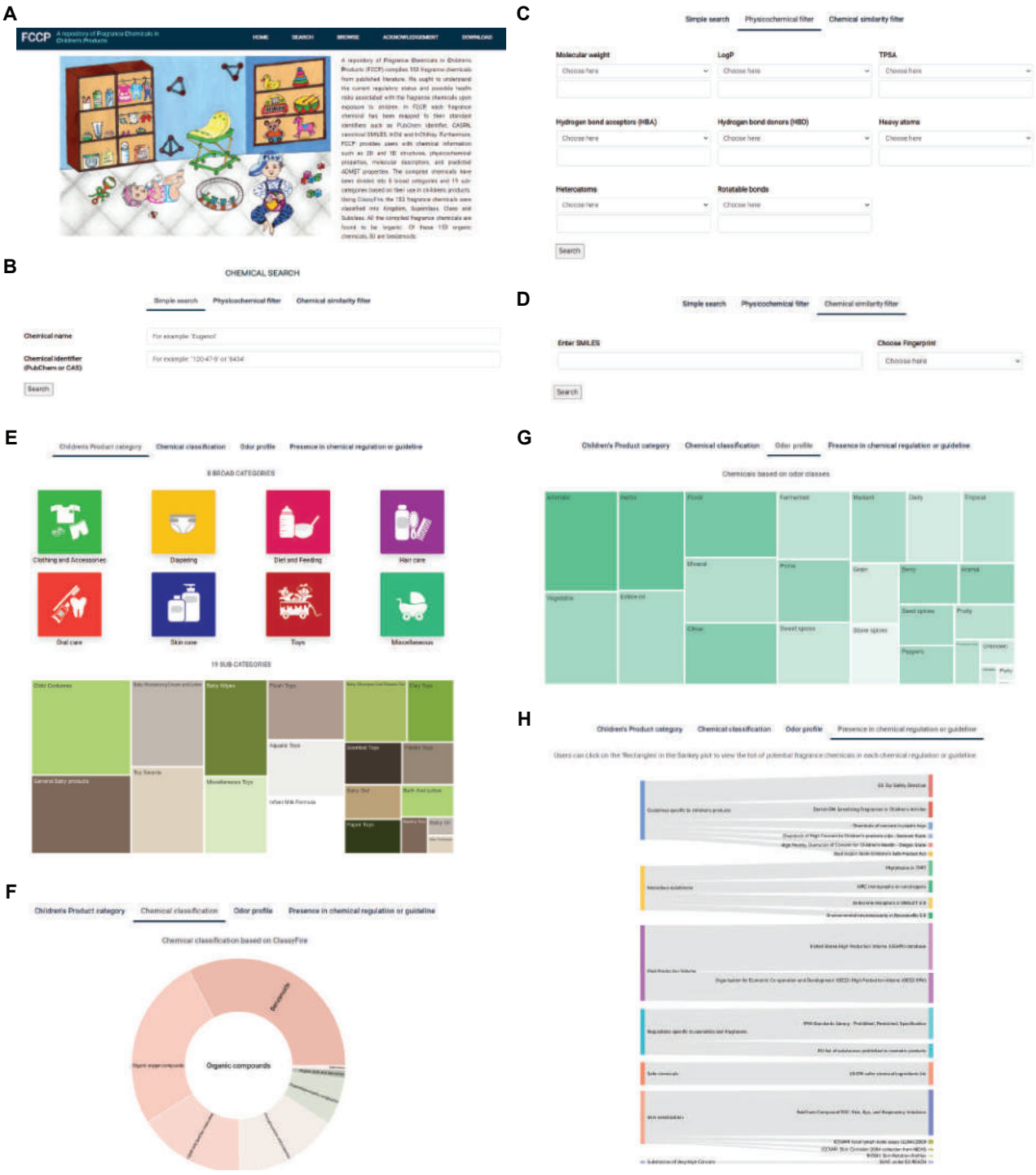


Figure 7.3: Screenshots from the web interface of FCCP. (A) Home page of FCCP. Users can retrieve the compiled fragrance chemicals using the following search options, namely, (B) Simple search, (C) Physicochemical filter, and (D) Chemical similarity filter. FCCP also provides a list of options to browse the compiled fragrance chemicals, namely, (E) Children's product source, (F) Chemical classification, (G) Odor profile, and (H) Presence in chemical regulation or guideline.

database [151], and (iii) REACH High Production Volume (REACH HPV) chemicals containing REACH registered substances as of 21 September 2021 with a tonnage range ≥ 1000 tonnes [152].

In the following subsections, we present a comparative analysis of compiled fragrance chemicals with chemical lists classified into various categories.

7.3.1 Guidelines specific to children's products

We considered 6 publicly available lists containing chemicals used in children's products that are subject to regulation. These lists contain chemicals that are restricted or prohibited for their use in children's products including toys and other child care products. The 6 chemical lists of concern to children include: (i) Chemicals of concern in plastic toys [148], (ii) Danish EPA Sensitizing Fragrances in Children's Articles [146], (iii) EU Toy Safety Directive [145], (iv) Washington State Children's Safe Product Act [147], (v) High Priority Chemicals of Concern for Children's Health - Oregon State [349], and (vi) Chemicals of High Concern to Children's products rule - Vermont State [350].

Based on comparison with the 6 chemical lists in the category 'Guidelines specific to children's products', we find that the 'EU Toy Safety Directive' list contains the highest number (31) of fragrance chemicals in FCCP (Figure 7.4; Supplementary Table S7.5). Of these 31 banned allergenic chemicals common to 'EU Toy Safety Directive' and FCCP, 3 fragrance chemicals namely, 'Methylparaben', 'Propylparaben', and 'Phenol' are also contained in 4 other chemical lists in the category 'Guidelines specific to children's products'. Interestingly, we also find that 18 out of 31 fragrance chemicals common to 'EU Toy Safety Directive' and FCCP are produced in high volume based on comparison with the three chemical lists of HPV chemicals (Figure 7.4; Supplementary Table S7.5). Notably, 14 fragrance chemicals common to FCCP and the chemical prioritization list 'Chemicals of concern in plastic toys' were also found to be present in the majority of the regulatory lists of concern investigated by Aurisano *et al.* [148]. Further, 13 out of these

14 fragrance chemicals are produced in high volume (Figure 7.4; Supplementary Table S7.5).

7.3.2 Regulations specific to cosmetics and fragrances

To better comprehend the regulation of compiled fragrance chemicals for their use in personal care products, we considered 2 publicly available lists that compile chemicals which are restricted or prohibited for their use in cosmetics or fragrance products. These 2 lists are: (i) EU list of substances prohibited in cosmetic products [141], and (ii) IFRA Standards Library - Prohibited, Restricted, Specification list [351].

Based on comparison with the 2 above chemical lists specific to cosmetics and fragrances, the 'IFRA Standards Library - Prohibited, Restricted, Specification' list contains 43 fragrance chemicals in FCCP, while the 'EU list of substances prohibited in cosmetic products' contains 19 fragrance chemicals in FCCP. Further, 10 fragrance chemicals in FCCP namely, '2-Heptenal', '2,4-Dihydroxy-3-methylbenzaldehyde', '4-Tert-Butylphenol', '7-Ethoxy-4-methylcoumarin', '7-Methoxycoumarin', '7-Methylcoumarin', 'Benzylideneacetone', 'Hexahydrocoumarin', 'Isophorone', and 'Lyral' are present in both chemical lists in the category 'Regulations specific to cosmetics and fragrances'. Moreover, of these 10 fragrance chemicals, 3 are also produced in high volume (Figure 7.4; Supplementary Table S7.5).

7.3.3 List of chemicals of very high concern

The European Union's Registration, Evaluation, Authorization, and Restriction of Chemicals (REACH) regulation (EC) No 1907/2006 includes a list of substances of very high concern (SVHC) [157]. Chemicals classified as SVHC have the potential to be: (i) Carcinogenic, Mutagenic, toxic to Reproduction (CMR), (ii) disruptive to the endocrine system, (iii) Persistent, Bioaccumulative and Toxic (PBT), and (iv) very Persistent and very Bioaccumulative (vPvB). The EU SVHC list was used to evaluate the chemicals of very

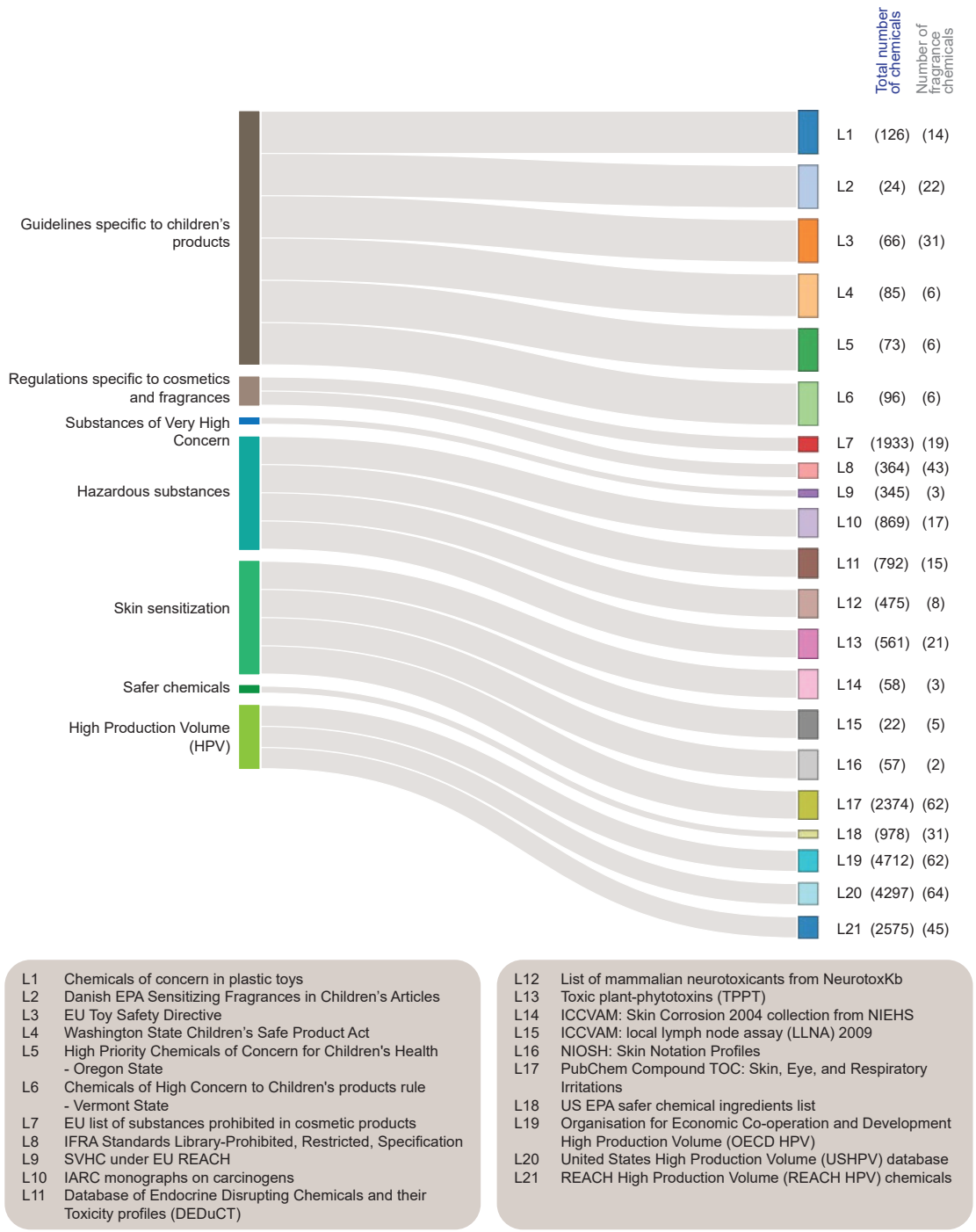


Figure 7.4: Sankey plot showing the presence of fragrance chemicals in FCCP across 21 chemical lists which reflect regulations or guidelines. Further, the 21 chemical lists have been classified into 7 categories which include Guidelines specific to children's products, Regulations specific to cosmetics and fragrances, Hazardous substances, Skin sensitization, Safer chemicals, Substances of Very High Concern, and High Production Volume (HPV) chemicals.

high concern among the compiled fragrance chemicals in FCCP.

Based on comparison with the only chemical list in the category 'Substances of Very High Concern', we find that 3 fragrance chemicals in FCCP are contained in 'SVHC under EU REACH' list. These 3 fragrance chemicals are '4-Tert-Butylphenol', 'Butylparaben', and 'Musk xylene', of which 2 fragrance chemicals are also produced in high volume (Figure 7.4; Supplementary Table S7.5).

7.3.4 List of hazardous chemicals

To analyze the fragrance chemicals in FCCP for known chemical hazards, we considered 4 publicly available lists which include: (i) IARC monographs on carcinogens [208], (ii) Database of Endocrine Disrupting Chemicals and their Toxicity profiles (DEDuCT) [35,36] (https://cb.imsc.res.in/deduct/), (iii) List of mammalian neurotoxicants from NeurotoxKb [37] (https://cb.imsc.res.in/neurotoxkb/), and (iv) Toxic plant-phytotoxins (TPPT) database [68, 352].

Based on comparison with the 4 chemical lists in the category 'Hazardous substances', 17, 15, 8, and 21 fragrance chemicals in FCCP are also carcinogens, endocrine disruptors, neurotoxicants and phytotoxins, respectively (Figure 7.4). The presence of these fragrance chemicals in consumer products for children increases the possibility of exposure, which may lead to potential health impacts in children. Carcinogens reported in IARC monographs have been categorized into one of the following groups: (i) Group 1 chemicals are human carcinogens, (ii) Group 2A chemicals are listed as 'probable' human carcinogens, (iii) Group 2B chemicals are possibly carcinogenic to humans, and (iv) Group 3 chemicals are not classifiable as human carcinogens [296]. Of the 17 fragrance chemicals in FCCP that are also carcinogens, 2, 1, 3 and 11 fragrance chemicals belong to Group 1, Group 2A, Group 2B and Group 3 based on IARC monographs classification. Further, 12 out of these 17 carcinogens in FCCP are also produced in high volume (Supplementary Table S7.5). A similar analysis revealed that 12, 8, and 10 fragrance

chemicals in FCCP which are endocrine disruptors, neurotoxicants and phytotoxins, respectively, are also produced in high volume, indicating the potential for adverse health effects in children when exposed to such chemicals (Supplementary Table S7.5). Notably, two fragrance chemicals in FCCP namely, 'Ethanol' and 'Acetaldehyde' are contained in 3 out of the 4 chemical lists in the category 'Hazardous substances' (Figure 7.4).

7.3.5 List of chemicals of concern to skin

Fragrance chemicals are known to induce skin sensitization [353]. It is worthwhile to investigate if the fragrance chemicals in FCCP are likely to cause skin sensitization. For this analysis, we considered 4 publicly available lists which include: (i) ICCVAM: Skin Corrosion 2004 collection from NIEHS [354], (ii) ICCVAM: Local lymph node assay (LLNA) 2009 [355], (iii) NIOSH: Skin Notation Profiles [356], and (iv) A list of chemicals that are known to cause Skin, Eye, and Respiratory Irritations compiled from PubChem Classification browser [86].

Based on comparison with the 4 chemical lists in the category 'Skin sensitization', we find that the chemical list 'PubChem Compound TOC: Skin, Eye, and Respiratory Irritations' contains 62 out of the 153 fragrance chemicals in FCCP (Figure 7.4; Supplementary Table S7.5). Further, 5 fragrance chemicals in FCCP namely, '2-Butoxyethanol', 'Citral', 'Eugenol', 'Lauric acid', and 'Phenol', are present in at least 2 out of the 4 chemical lists in the category 'Skin sensitization'. Moreover, all of these 5 fragrance chemicals are also produced in high volume (Supplementary Table S7.5).

7.3.6 Regulation for safer chemicals

The United States Environmental Protection Agency (US EPA) has released a list of chemicals that are considered to be among the safest for their intended functional use [171]. In other words, the chemicals in this list are safer alternatives for certain functional uses including chelating agents, colorants, polymers, preservatives, enzyme stabilizers,

perfumes, solvents, and surfactants. The US EPA considers a chemical to be a safer alternative for specific functional use category only if the chemical meets the Safer Choice Program criteria, which include the assessment of a wide range of potential toxicological effects such as carcinogenicity, mutagenicity, bioaccumulation, skin sensitization, allergenicity, and endocrine disruption. Further, US EPA gives the following classification of chemicals that indicates their safety status in each functional category: (i) 'Green circle' indicates the chemicals that are verified to be of low concern, (ii) 'Green half-circle' indicates the chemicals that are expected to be of low concern based on the available evidence, (iii) 'Yellow triangle' indicates the chemicals which have some evidence for hazardous nature though listed to be safe for certain functional-use, and (iv) 'Grey square' indicates the chemicals that are not acceptable for their use in some of the products and must be reformulated. We used this list to assess the fragrance chemicals in FCCP.

Based on this comparison, we find that 31 fragrance chemicals in FCCP are contained in the 'US EPA safer ingredients' list (Supplementary Table S7.5). Since the 'US EPA safer ingredients' list classifies the chemicals based on different use categories (like solvents, fragrances), we analyzed these 31 chemicals based on these categories. Of these 31 fragrance chemicals, we find that 25 were labeled as 'safer' for use as fragrance ingredients in consumer products, while the remaining 6 were not labeled as 'safer' for use as fragrance ingredients. Furthermore, analysis of these 25 (safer) fragrance chemicals in the 'US EPA safer ingredients' list based on the type of evidence revealed that 2, 3, and 20 fragrance chemicals belong to 'Green circle', 'Green half-circle', and 'Yellow triangle' categories, respectively (Figure 7.4). Of these 25 (safer) fragrance chemicals, we find that 4, 5, and 5 fragrance chemicals are present in 3 chemical lists that reflect guidelines specific to children's products namely, 'Chemicals of concern in plastic toys', 'Danish EPA Sensitizing Fragrances in Children's Articles', and 'EU Toy Safety Directive', respectively. Interestingly, we find that 22 out of the 25 (safer) fragrance chemicals are listed in 'IFRA Standards Library - Prohibited, Restricted, Specification'. By analyzing these 25 (safer) fragrance chemicals with chemical lists grouped in 'Hazardous

substances' category, we find that the chemicals 'Benzyl salicylate' and 'D-limonene' are class 3 carcinogen and endocrine disruptor, respectively. In addition, these two chemicals are also produced in high volume (Figure 7.4; Supplementary Table S7.5). Although these 25 chemicals were marked 'safer' for their use as fragrance ingredients by the US EPA, some of them are present in the different lists containing chemicals that display hazard profiles or suggested to be limited or prohibited in cosmetics or children's products.

Overall, these results highlight the disparities in the regulations or guidelines across countries, necessitating prioritization and risk assessment of fragrance chemicals used in children's products, as many of them have potency to cause health hazards in children [40].

7.4 Similarity network of fragrance chemicals in children's products

To better understand the space of fragrance chemicals in children's products, we compared the structural similarity of fragrance chemicals in our resource FCCP with the list of allergenic fragrance chemicals restricted or banned for their use in children's toys as compiled in the 'EU Toy Safety Directive' [145]. For this purpose, we constructed two chemical similarity networks (CSNs), one for the 153 fragrance chemicals in FCCP, and another for the 58 allergenic fragrance chemicals in the 'EU Toy Safety Directive'. Note that only 58 out of the 66 allergenic fragrance chemicals in the 'EU Toy Safety Directive' have chemical structure information available.

To build the CSNs, the Tanimoto coefficient [200] was computed using the Extended Circular Fingerprints (ECFP4) method [129] for each pair of chemicals between the two datasets. Tanimoto coefficient for any pair of compounds ranges from 0 to 1 with 1 signifying two compounds with identical structures. This led to two CSNs, one with 153 nodes for fragrance chemicals in FCCP, and another comprising 58 nodes for banned allergenic fragrance chemicals in the 'EU Toy Safety Directive'. Based on previous stud-

ies [283,357], a Tanimoto coefficient cut-off of 0.5 was used to determine if an edge exists between any pair of chemicals in the dataset, resulting in a high similarity network of fragrance chemicals. Moreover, we also computed the Tanimoto coefficient for each pair of a fragrance chemical in FCCP and a banned allergenic fragrance chemical in the 'EU Toy Safety Directive' (Supplementary Table S7.6). A detailed investigation of the two CSNs can help reveal the extent of structural similarities between chemicals in our resource and 'EU Toy Safety Directive'.

An analysis of the CSN of 153 fragrance chemicals in FCCP reveals that there are 16 connected components with ≥ 2 chemicals and 51 isolated nodes (chemicals), and this suggests a high structural diversity in the space of fragrance chemicals used in children's products (Figure 7.5A). Notably, the largest connected component in the CSN of 153 fragrance chemicals in FCCP consists of 25 fragrance chemicals (Figure 7.5A). In Figure 7.5A, the 31 fragrance chemicals common to FCCP and 'EU Toy Safety Directive' of banned allergenic chemicals are highlighted in green. We observed that the 31 banned allergenic chemicals are dispersed across different connected components in the CSN of 153 fragrance chemicals in FCCP, implying that both chemical spaces are structurally diverse. Furthermore, we computed the chemical similarity using the Tanimoto coefficient [200] between each chemical in FCCP and each banned allergenic chemical in 'EU Toy Safety Directive', and any fragrance chemical in FCCP with chemical similarity ≥ 0.7 to any of the banned allergenic chemicals in the 'EU Toy Safety Directive' are also highlighted in the CSN of 153 fragrance chemicals in FCCP (Figure 7.5A; Supplementary Table S7.6). Finally, we also built and visualized the CSN for the 58 banned allergenic chemicals in 'EU Toy Safety Directive' (Figure 7.5B). It is seen that the CSN of 58 banned allergenic chemicals in 'EU Toy Safety Directive' has 11 connected components with ≥ 2 chemicals and 26 isolated nodes (Figure 7.5B). Overall, an analysis of these CSNs reveals the structural diversity of the fragrance chemical space [40].

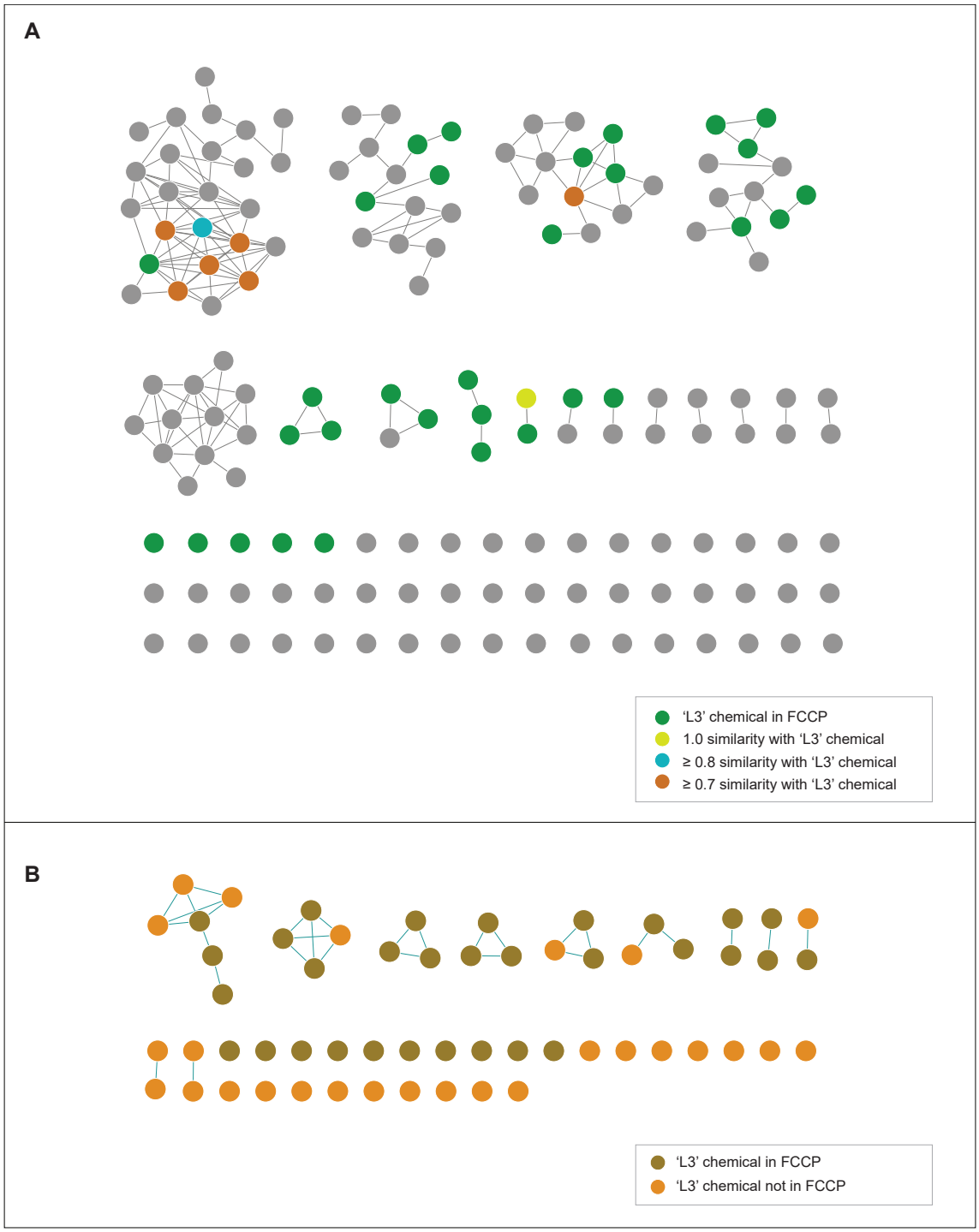

A
'L3' chemical in FCCP
1.0 similarity with 'L3' chemical
≥ 0.8 similarity with 'L3' chemical
≥ 0.7 similarity with 'L3' chemical
B
'L3' chemical in FCCP
'L3' chemical not in FCCP

Figure 7.5 ***(previous page)*:** Chemical similarity networks (CSNs) of fragrance chemicals. Here, nodes represent fragrance chemicals, and two nodes are connected by an edge if the corresponding chemicals have chemical similarity ≥ 0.5 based on Tanimoto coefficient. (A) CSN of the 153 fragrance chemicals in FCCP. Here, nodes corresponding to the 31 fragrance chemicals common to both FCCP and 'EU Toy Safety Directive' (L3) are highlighted in 'green', while the other nodes are colored based on their level of chemical similarity to the banned allergenic chemicals in L3. (B) CSN of the 58 allergenic fragrance chemicals in 'EU Toy Safety Directive' (L3). Note that only 58 out of the 66 allergenic fragrance chemicals in the 'EU Toy Safety Directive' have chemical structure information available. Here, nodes corresponding to the allergenic fragrance chemicals that are also present in FCCP have been highlighted.

7.5 Linking fragrance chemicals in children's products to their target genes

Olfactory receptors or odorant receptors are responsible for the olfactory perception of the fragrance molecules. These receptors are found in the olfactory sensory neurons of the olfactory epithelium within the nasal cavity [358, 359]. It is known that even slight modifications in the structure of fragrance molecules can alter the quality of olfactory perception [360]. Hence, existing information on odor receptors specific to fragrance chemicals can be used to better understand the mechanism of olfactory perception [358]. To compile the list of odor receptors that are known to bind experimentally to the fragrance chemicals in FCCP, we used Odor Molecules Database (OdorDB) [361, 362]. Olfactory Receptor Database (ORDB) [363, 364] compiles six classes of G-protein-coupled sensory chemoreceptors namely, olfactory receptor-like proteins (ORLs), *C. elegans* chemoreceptors (CCRs), vomeronasal receptors (VNRs), insect olfactory receptors (IORs), fungal pheromone receptors (FPRs) and taste papilla receptors (TPRs) [364]. Using OdorDB, we have compiled 54 odor receptors associated with 20 fragrance chemicals in FCCP (Figure 7.6A; Supplementary Table S7.7). OdorDB contains the list of ligands that can bind to the receptors compiled in the ORDB. Of these 20 fragrance chemicals in FCCP with odor receptor information, we find that 4 fragrance chemicals namely, 'Acetophenone', 'Coumarin', 'Cyclohexanone' and '2-Hepatanone', are known to bind to at least 10 different odor receptors. Among the 54 odor receptors to which at least one of the 20

fragrance chemicals in FCCP can bind, ORL2156, ORL2162, ORL1858, ORL1553 and ORL1138 are found to be targeted by at least 5 fragrance chemicals in FCCP. Additional information on the binding of fragrance chemicals in FCCP to different odor receptors can help better understand the mechanisms of olfactory perception [40].

Besides compiling the odor receptors, we also identified the target genes specific to humans of the fragrance chemicals in FCCP using ToxCast [89]. ToxCast provides information on the list of genes perturbed upon exposure to chemicals which were identified based on high-throughput experimental assays. To identify the human target genes for the fragrance chemicals in FCCP, we used ToxCast invitroDB3 dataset released in August 2019 [215]. We followed the method described in Section 2.4.2 to extract from ToxCast the human target genes perturbed upon exposure to fragrance chemicals in FCCP (Supplementary Table S7.8). Based on the ToxCast assays, we were able to compile 130 human genes which are targets of at least one of 102 fragrance chemicals in FCCP (Supplementary Table S7.8). Of these 102 fragrance chemicals in FCCP, 18 fragrance chemicals can target at least 20 human genes based on ToxCast assays. Specifically, 4 fragrance chemicals namely, 'Propylparaben', '2-Benzylideneheptanal', 'Oxacyclohexadecan-2-one', and 'Hexyl cinnamic aldehyde' can target more than 40 human genes based on ToxCast assays. Among the 130 human target genes of the 102 fragrance chemicals in FCCP, 14 human genes are targets of at least 20 fragrance chemicals in FCCP. An in-depth analysis of these target genes can shed light on shared toxicological mechanisms associated with fragrance chemicals in children's products [40].

7.6 ToxCast assays for skin sensitization

Since fragrance chemicals can trigger skin sensitivity [353], we decided to leverage *in vitro* ToxCast human assays [89] to identify the fragrance chemicals that have potential to cause skin sensitization. Motivated by Spinu *et al.* [230], we investigated the Adverse Outcome Pathways (AOPs) in AOP-Wiki [114] to determine the endpoints related to skin

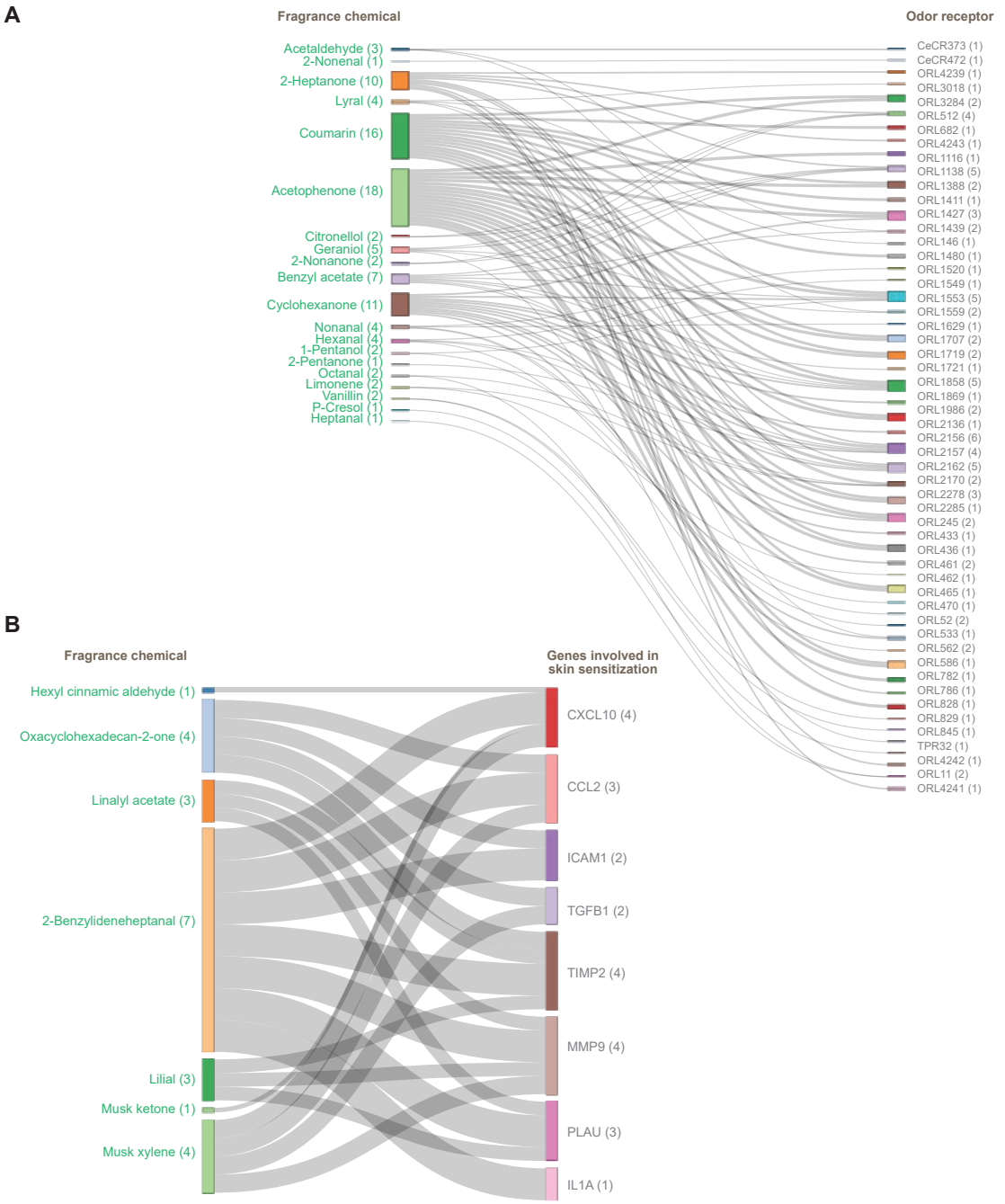


Figure 7.6: (A) Bipartite graph displaying the 20 fragrance chemicals in FCCP and their associated odor receptors identified using OdorDB. (B) Bipartite graph displaying the human target genes of 7 fragrance chemicals in FCCP which were identified to have potential to cause skin sensitization based on ToxCast *in vitro* human assays. Here, the number of odor receptors or target genes associated with each fragrance chemical is mentioned in parenthesis, and similarly, the number of fragrance chemicals associated with each odor receptor or target gene is also mentioned in parenthesis.

sensitization that can be used to select relevant ToxCast assays for skin sensitization. Within AOP-Wiki, AOP:40 describes the key events (KEs) that lead to skin sensitization, and these include chemical binding to skin proteins, activation of keratinocytes, dendritic cells, and T-cells. Among the KEs of AOP:40 for skin sensitization, we identified 'Activation, Keratinocytes' (KE:826) as a suitable endpoint for screening of skin sensitizing fragrance chemicals. Previous studies have also revealed that keratinocytes are useful in determining whether substances have the potential to cause skin sensitization [365, 366].

To select the list of relevant skin sensitization assays in ToxCast, we used the ToxCast invitroDB3 dataset released in August 2019 [215]. Firstly, we imposed a tissue-specific filter to only select ToxCast assays for human skin tissue. Two cell lines have been investigated among the shortlisted skin-specific ToxCast assays which are foreskin fibroblasts (hDFCGF) and co-culture of keratinocytes and foreskin fibroblasts (KF3CT). Secondly, we evaluated the ToxCast assays performed on KF3CT cell lines that have already been used to screen compounds for skin sensitization [367]. Thirdly, we selected only the reporter assays that were designed to analyze the regulation of gene expression in ToxCast. The above-mentioned filtration resulted in identification of human-specific skin sensitization assays from ToxCast which can be further used to test if a chemical has potency for skin sensitization. Note that each ToxCast assay constitutes multiple assay component endpoints which are designed to assess one or more target genes. Finally, if a fragrance chemical in FCCP has tested 'active' for the assay component endpoints specific to a selected human skin sensitization ToxCast assay, the corresponding gene is assigned as a target of that fragrance chemical in FCCP [35]. This process resulted in 16 assay component endpoints that are associated with the filtered set of skin sensitization assays in ToxCast [148]. Among the fragrance chemicals in FCCP, 7 fragrance chemicals have 10 out of the 16 assay component endpoints as 'active' upon exposure in the filtered set of skin sensitization assays in ToxCast (Supplementary Table S7.8). These 7 fragrance chemicals in FCCP namely, '2-Benzylideneheptanal', 'Hexyl cinnamic aldehyde', 'Linalyl acetate', 'Lilial', 'Musk ketone', 'Musk xylene', and 'Oxacyclohexadecan-2-one', have the poten-

tial to cause skin sensitization based on ToxCast assays, and moreover, the 7 fragrance chemicals are associated with 8 human target genes (Figure 7.6B).

Interestingly, we find that 5 out of these 7 fragrance chemicals in FCCP with skin sensitization potential based on ToxCast assays, are present in at least one of the 4 chemical lists in the category 'Skin sensitization'. Further, 3 out of these 7 fragrance chemicals are present in the 2 chemical lists namely, 'Danish EPA Sensitizing Fragrances in Children's Articles' and 'EU Toy Safety Directive'. Moreover, one of these 7 fragrance chemicals identified to have skin sensitization potential based on ToxCast assays namely, 'Oxacyclohexadecan-2-one', is not present in any of the chemical lists in the categories 'Skin sensitization' or 'Guidelines specific to children's products'. However, 'Oxacyclohexadecan-2-one' is a prohibited or restricted substance in cosmetics and fragrances according to 'IFRA Standards Library - Prohibited, Restricted, Specification' list (Supplementary Table S7.5).

7.7 Discussion

Exposure of children to hazardous chemicals via any route is a significant concern due to the potential impact on the growth and development during early childhood [18, 39, 80, 81, 148, 286, 287, 335, 336, 342]. Fragrance chemicals, a subset of chemicals used in children's products, are either self-regulated or poorly regulated [75, 79, 81]. The absence of a dedicated knowledgebase compiling the surrounding knowledge dispersed across scientific literature on fragrance chemicals in children's products may also hinder the risk assessment and regulatory decisions on such chemicals.

In this chapter [40], we present a manually curated knowledgebase FCCP that compiles 153 fragrance chemicals in children's products from 21 published experimental studies (Figure 7.7). The detailed information on fragrance chemicals in FCCP can be easily accessed via a user friendly web interface. Through a comparative analysis with 21 chemical lists reflecting current guidelines or regulations, we found that several fragrance

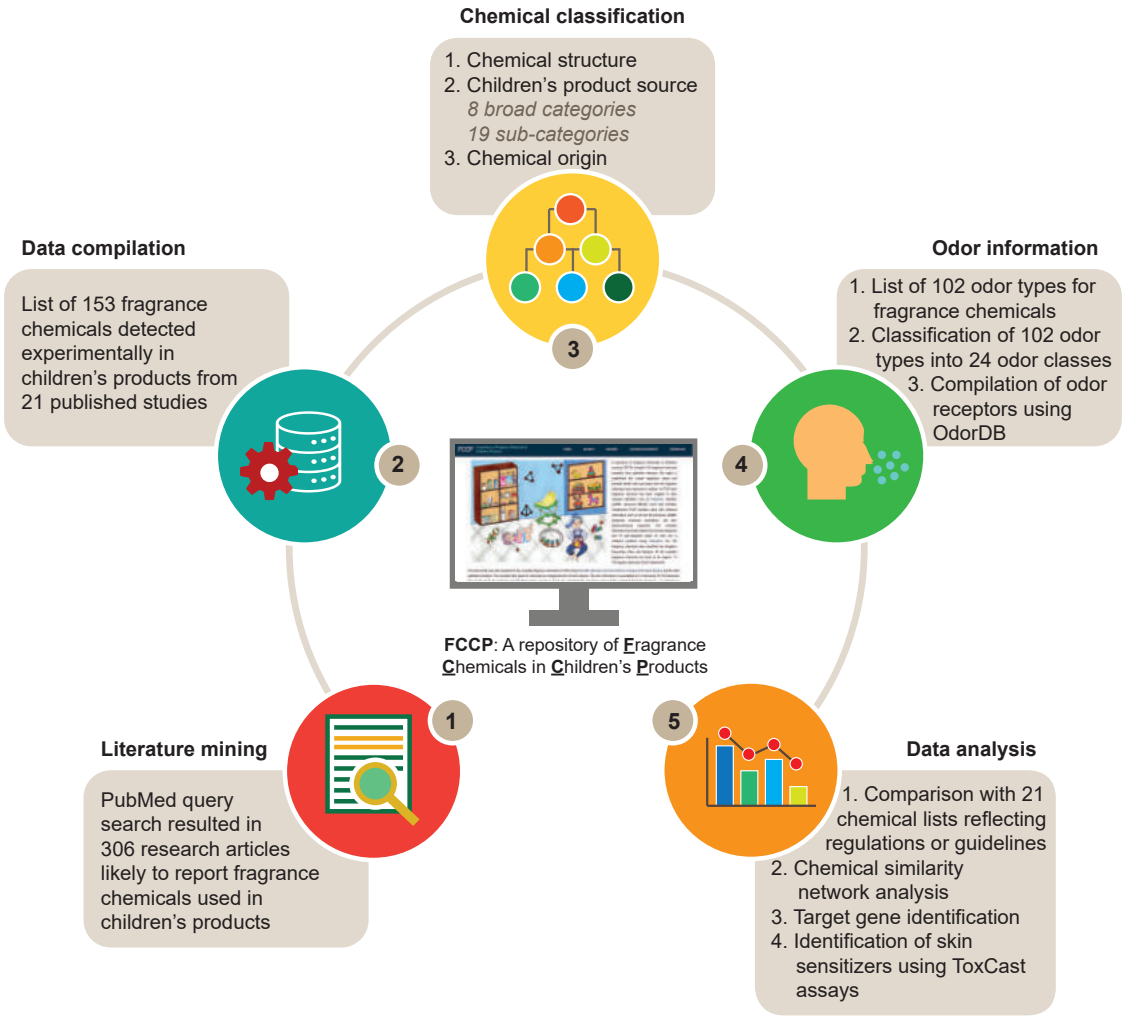


Figure 7.7: Schematic overview of the creation and analysis of the repository of Fragrance Chemicals in Children's Products (FCCP).

chemicals in FCCP are either banned allergenic chemicals, or are prohibited or restricted in cosmetics and fragrances. Further, this analysis revealed that several fragrance chemicals in FCCP are carcinogens, endocrine disruptors, neurotoxicants, phytotoxins and skin sensitizers, raising concerns about the potential health hazards in children. Notably, several fragrance chemicals in FCCP of potential concern are also produced in high volume. Next, we performed a similarity network based analysis of the fragrance chemicals in FCCP which revealed the structurally diverse nature of the associated chemical space. Then, we compiled and analyzed the odor receptors and human target genes for fragrance chemicals in FCCP. Lastly, we identified 7 skin sensitizing fragrance chemicals in FCCP using ToxCast *in vitro* human assays. In sum, our multipronged analysis of the atlas of fragrance chemicals in children's products underscores the need to monitor and regulate them (Figure 7.7).

Children can be exposed to fragrance chemicals through different routes including skin, respiration, or ingestion [75,339,342,343]. However, the main focus of safety testing of such chemicals by the fragrance industry has been skin related toxicity while ignoring other routes of exposure [339]. Therefore, additional studies are needed on toxicological or disease pathways associated with other routes of children exposure to fragrance chemicals. Further, some industries maintain secrecy on their fragrance ingredients or their composition, and this presents an additional challenge for researchers trying to understand the associated health impact on children upon exposure [79, 339, 340]. Thus, the information on fragrance chemicals compiled in FCCP can be used to better understand the health effects of exposure, enabling a better characterization of the external exposome of children. In conclusion, FCCP will facilitate future toxicological and exposome research, enabling risk assessment of fragrance chemicals, and thereby improving the safety of children's products.

Supplementary Information

Supplementary Tables S7.1-S7.8 associated with this chapter are available for download from the GitHub repository: https://github.com/asamallab/PhDThesis-Janani_R/blob/main/SI/ST_Chapter7.xlsx.

Chapter 8

Network-based exploration of a human tissue-specific chemical exposome atlas (TExAs)

Exposure to environmental chemicals such as pollutants or toxicants plays a major role in the burden of many chronic diseases [9–12]. To embark on research into the mechanistic aspects of chemical exposure-effect relationships, it is necessary to gather data on the presence of environmental chemicals in specific human biospecimens. Human biomonitoring studies have enabled the measurement of these chemicals in various human biospecimens using analytical techniques [82–84]. In particular, monitoring chemicals in human tissues is regarded as the gold standard in the study of exposed populations as it reflects long-term exposure and bioaccumulation of environmental chemicals. [85].

In this chapter, we aim to characterize the chemical component of the external exposome, specific to human tissues, and to explore ways to understand the health implications of these chemicals. For this purpose, we consider three resources namely, CTD [30], Exposome-Explorer [24] and PubChem [86], which have compiled chemicals detected across human tissues, based on exposure studies from published research articles. Since

we have chosen to focus on human tissues excluding biological fluids, comprehensive resources such as the Blood Exposome Database [28] pertaining to a biological fluid were not included in this chapter. The three resources [24, 30, 86] considered in this chapter, however, do not provide a cohesive picture of chemical exposure-disease relationships, specific to human tissues. In this chapter, we have explored exposure-disease relationships of the tissue-specific external exposome using network biology [13, 88] approaches. **The work reported in this chapter is contained in the published manuscript [41].**

8.1 Creation of a tissue-specific external exposome atlas

Biomonitoring is the measurement of environmental or toxic chemicals in biological specimens through analytical techniques [82]. We, therefore, consider the presence or detection of chemicals in human biological specimens to be an indication of human exposure to those chemicals [82, 83]. Our first step in developing a tissue-specific chemical exposome atlas is the compilation of chemicals detected in human tissues excluding biological fluids like blood, urine, and saliva (Figure 8.1). We consider three resources for this compilation, namely, CTD [30], Exposome-Explorer [24] and PubChem [86].

CTD has compiled a list of 1146 chemicals detected across non-biological and biological specimens from exposure studies published in scientific literature [30]. In CTD, the non-biological and biological specimens together are referred to as 'Mediums' in the database [30]. Exposome-Explorer is a comprehensive resource that compiles 'biomarkers' of dietary and environmental exposures that are risk factors for disease [24]. Although Exposome-Explorer compiles information on more than 1200 chemical biomarkers, we only considered the subset of 450 dietary and environmental chemicals in Exposome-Explorer with chemical structure information, after excluding entries that lack structure information or occur as chemical mixtures. PubChem, a comprehensive chemical database developed by the National Center for Biotechnology Information (NCBI), National Institutes of Health (NIH) of the United States, annotates information including

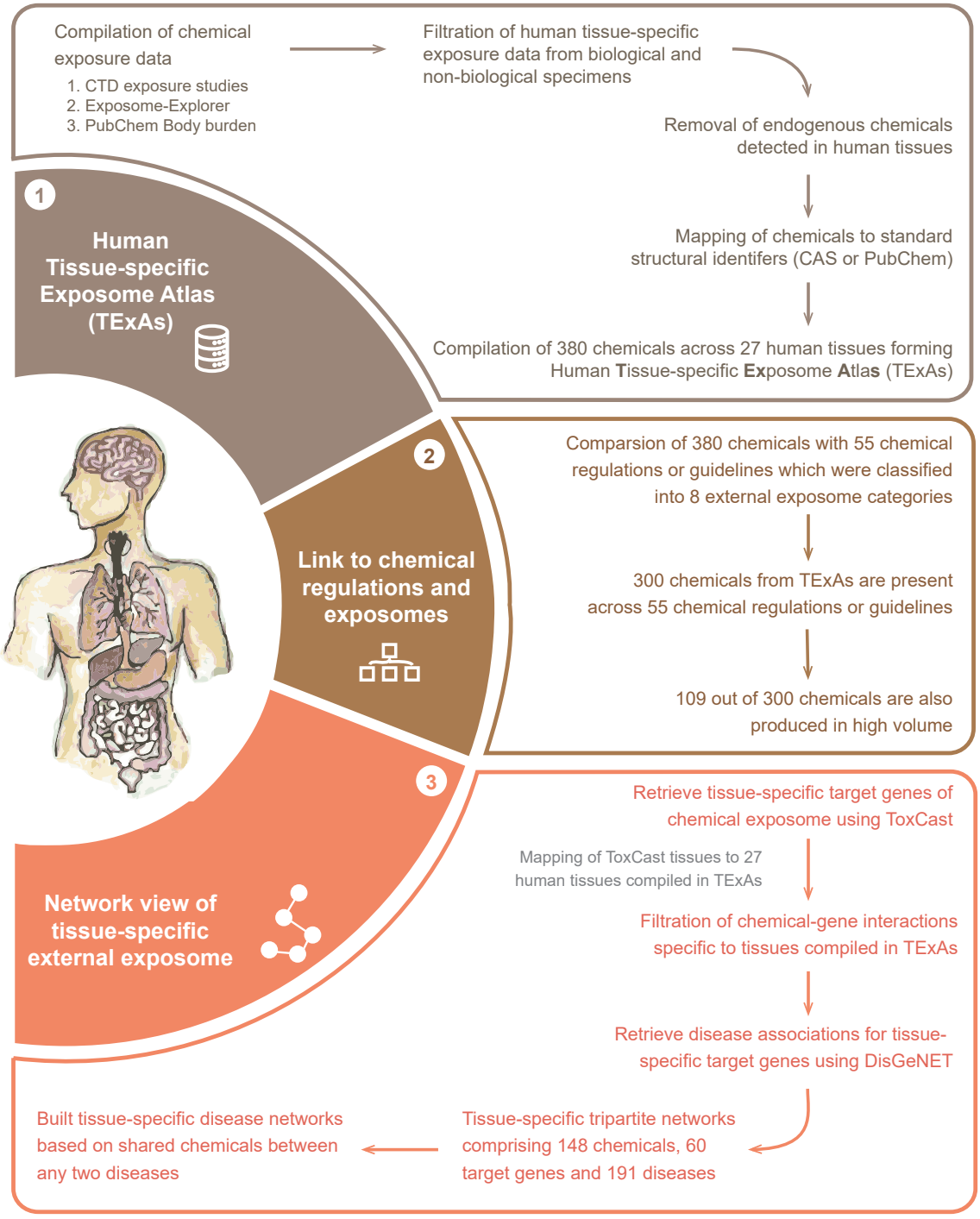


Figure 8.1: Detailed workflow describing the creation of Human Tissue-specific Exposome Atlas (TExAs) and downstream analysis of the compiled list of 380 environmental chemicals detected across 27 human tissues.

toxicological and exposure information for the chemicals compiled in the resource [86]. A list of 844 chemicals is available separately through PubChem Classification Browser under the hierarchy 'Body Burden'. These 844 chemicals have been annotated as chemicals detected across environmental samples and biological specimens in published scientific studies. To standardize the exposure and biospecimen data compiled from the three resources, we have manually unified the information on mediums and biospecimens to a standard vocabulary. Note that the above-mentioned three resources also give the references to the published literature evidence associated with exposure and biospecimen data. To build the human tissue-specific exposome atlas, we perform the following two steps [41].

8.1.1 Collection and filtration of human tissues

In the first step, the list of 467 mediums compiled from the three resources, CTD, Exposome-Explorer and PubChem, were manually filtered to 199 biological mediums. For example, non-biological mediums such as air, water or other environmental samples have been removed in this step. In the second step, we have filtered 61 human biospecimens from 199 biological mediums, which include both biological fluids, such as blood and sweat, and biological non-fluids, such as adipose tissue. In the last step, we have filtered 27 human tissues from the list of 61 human biospecimens to develop a human tissue-specific chemical exposome resource (Figure 8.1). In this work, we do not consider environmental chemicals detected in human biospecimens corresponding to biological fluids like blood, urine and saliva, and therefore, we have not gathered information from comprehensive resources such as the Blood Exposome Database [28].

8.1.2 Collection of chemicals detected across human tissues

We have considered the chemicals detected across all 61 human biospecimens from the three resources, CTD, Exposome-Explorer and PubChem. A set of 1510 chemicals have been detected across 61 human biospecimens (including biological fluids and biolog-

ical non-fluids). Endogenous chemicals do not constitute the external environmental exposures of a human being. We therefore manually filtered and considered only non-endogenous chemicals for further analysis. We then mapped the filtered chemicals to standard chemical identifiers such as Chemical Abstract Service (CAS) and PubChem [86] to compile a unified list of environmental chemicals. Note that chemical classes and mixtures were also removed in this step. At this stage, we filtered 380 unique environmental chemicals which have been detected across 27 human tissues (excluding biological fluids), from our initial compilation of 1510 chemicals (Figure 8.1; Supplementary Table S8.1). Among 27 human tissues in our compiled dataset, the maximum number of 240 environmental chemicals were detected in adipose tissue, followed by 120 chemicals in placenta. Figure 8.2B shows the number of environmental chemicals detected across the 27 human tissues in our compiled dataset.

While compiling the curated dataset of environmental chemicals detected in human tissues, we have manually evaluated the compiled evidence from more than 200 published research articles that are associated with the exposure and biospecimen data in the three resources: CTD, Exposome-Explorer and PubChem. This evaluation resulted in the classification of associated literature evidence into three classes: Level 1, Level 2, and Level 3. 'Level 1' indicates significant experimental evidence in the associated literature for chemical detection in human tissues. For example, if the associated literature evidence reports a chemical in a particular human tissue based on the experiments including gas chromatography/mass spectrometry (GC/MS), then the evidence is classified to be 'Level 1'. Similarly, 'Level 2' indicates evidence obtained from correlation studies, and 'Level 3' indicates limited or probable evidence (Supplementary Table S8.1) [41].

A hierarchical classification of the 380 environmental chemicals was obtained based on their chemical structures using ClassyFire [173, 174]. Based on this chemical classification, 339 chemicals are labelled as organic and 41 as inorganic (Figure 8.2C). Among the 339 organic chemicals, 150 belong to the super-class benzenoids, which is the largest among the chemical super-classes (Figure 8.2C).

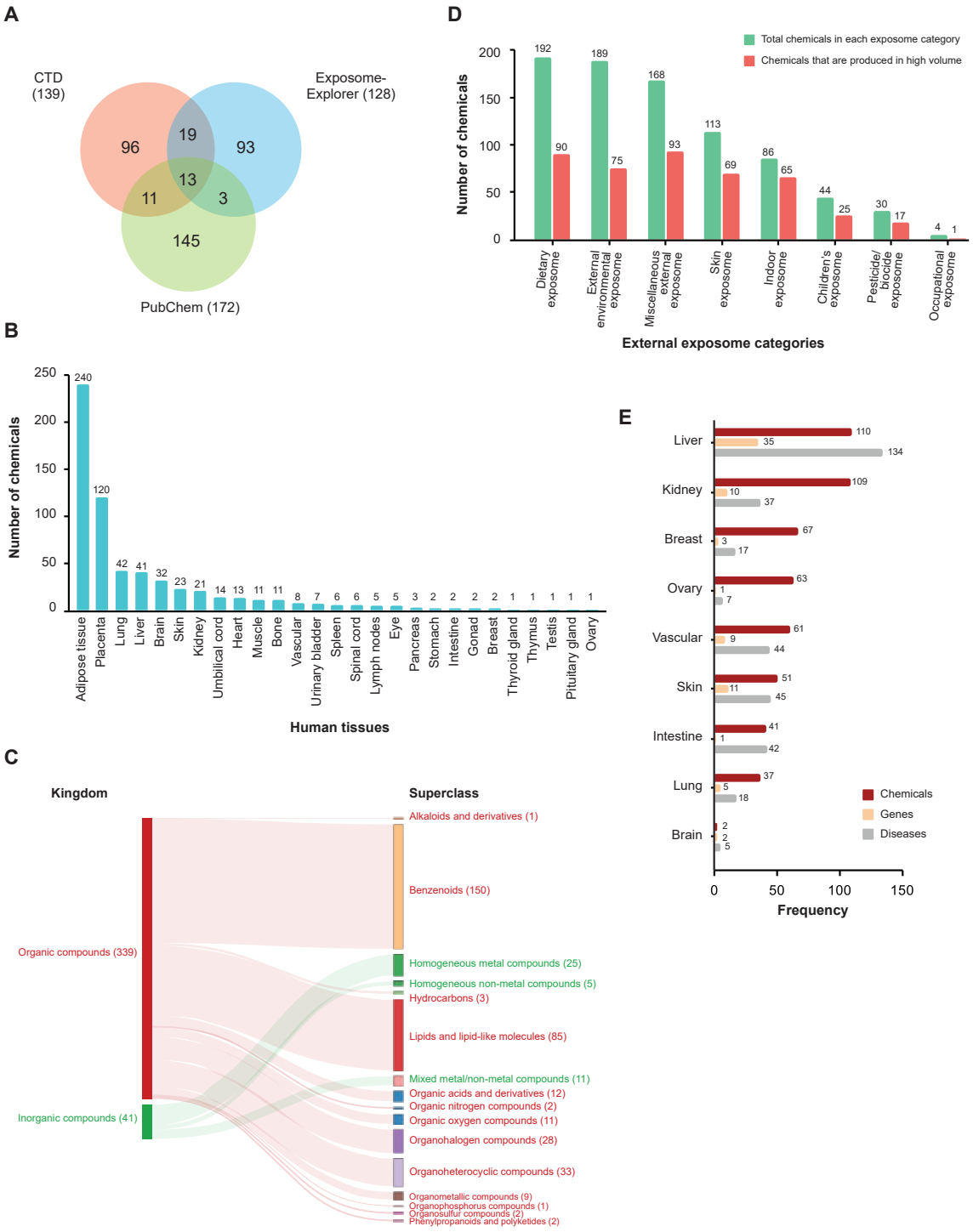

A
CTD (139)
Exposome-Explorer (128)
PubChem (172)
96
19
93
13
11
3
145
D
Total chemicals in each exposome category
Chemicals that are produced in high volume
Number of chemicals
External exposome categories
Dietary exposome
External environmental exposome
Miscellaneous external exposome
Skin exposome
Indoor exposome
Children's exposome
Pesticide/ biocide exposome
Occupational exposome
B
Number of chemicals
Human tissues
Adipose tissue
Placenta
Lung
Liver
Brain
Skin
Kidney
Umbilical cord
Heart
Muscle
Bone
Vascular
Urinary bladder
Spleen
Spinal cord
Lymph nodes
Eye
Pancreas
Stomach
Intestine
Gonad
Breast
Thyroid gland
Thymus
Testis
Pituitary gland
Ovary
E
Liver
Kidney
Breast
Ovary
Vascular
Skin
Intestine
Lung
Brain
Chemicals
Genes
Diseases
Frequency
C
Kingdom
Superclass
Organic compounds (339)
Inorganic compounds (41)
Alkaloids and derivatives (1)
Benzenoids (150)
Homogeneous metal compounds (25)
Homogeneous non-metal compounds (5)
Hydrocarbons (3)
Lipids and lipid-like molecules (85)
Mixed metal/non-metal compounds (11)
Organic acids and derivatives (12)
Organic nitrogen compounds (2)
Organic oxygen compounds (11)
Organohalogen compounds (28)
Organoheterocyclic compounds (33)
Organometallic compounds (9)
Organophosphorus compounds (1)
Organosulfur compounds (2)
Phenylpropanoids and polyketides (2)

Figure 8.2 ***(previous page)*****:** (A) The Venn diagram shows the presence of 380 environmental chemicals compiled in TExAs across the three resources, namely, CTD, Exposome-Explorer, and PubChem database. (B) The histogram shows the distribution of 380 environmental chemicals detected across 27 human tissues. (C) The Sankey plot shows the chemical classification of 380 environmental chemicals into 2 kingdoms and 16 super-classes based on ClassyFire. The number of chemicals in each classification is indicated within the parenthesis. (D) The bar plot shows the distribution of 300 environmental chemicals present in at least one of the 55 chemical lists (corresponding to chemical inventories, regulations, and guidelines), across 8 external exposome categories. For each external exposome category, one bar represents the total number of chemicals and the other represents the number of chemicals produced in high volume. (E) The grouped bar plot gives the number of environmental chemicals, target genes and diseases associated with each of the 9 human tissues.

8.2 Web interface of TExAs

To enable better access to this compilation of environmental chemicals detected in human tissues, we have developed a web interface, Human **T**issue-specific **Ex**posome **A**tla**s** (TExAs) [41] which includes detailed information for the 380 chemicals. For each chemical in TExAs, we have compiled the 2D (two-dimensional) and 3D (three-dimensional) structure information, canonical SMILES, InChI, and InChIKey. The compiled 2D and 3D structures can be downloaded in formats such as SDF, MOL, MOL2, PDB and PDBQT. Furthermore, we have computed the physicochemical properties for the chemicals using RDKit [179]. Users can navigate TExAs via either simple search or browse options through the web interface (Figure 8.3). The web interface of TExAS has been created using an approach similar to that described in Section 2.2.

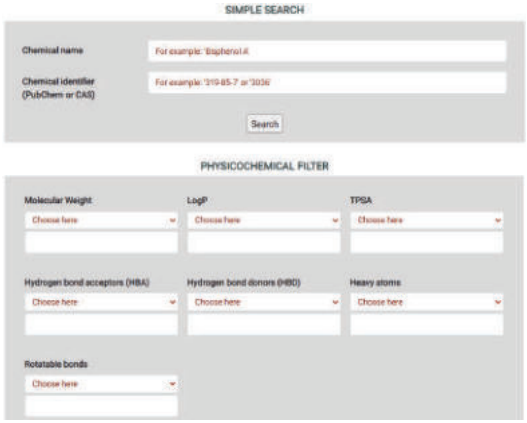
8.3 Mapping of chemicals to different exposome categories

The presence or detection of chemicals of concern in biological specimens is proof of human exposure [82], and thus, warrants further attention from the monitoring and regulatory perspectives to avoid future human exposure. We, therefore, sought to understand the source and nature of the environmental chemicals in TExAs through a comparative

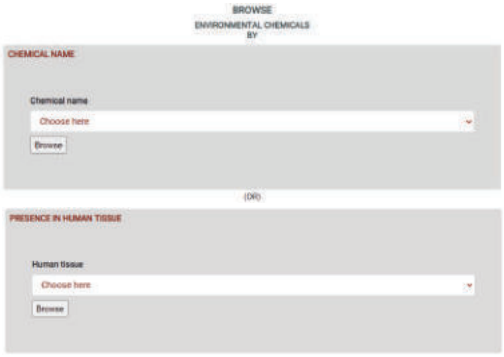
A


TExAs
Human Tissue-specific
Exposome Atlas
HOME
SEARCH
BROWSE
ACKNOWLEDGEMENT
DOWNLOAD
Humans are exposed to a variety of chemicals from birth, throughout their lives. Both endogenous and exogenous chemicals can affect our biological processes. There has been an increase in the usage of chemicals in various household and industrial applications. These environmental chemicals, constituting the external exposome, have been associated with a significant disease burden. Preventing exposure to hazardous chemicals in the external exposome, and mitigating the effects of such exposure, will require a clear understanding of the effects of these chemicals on the body. In order to characterize the external exposome and its effects, we compile tissue-specific chemical exposure data into this resource, Human Tissue-specific Exposome Atlas (TExAs). The chemicals in TExAs have further been integrated with potential gene targets and possible disease associations.
TExAs compiles 380 environmental chemicals detected across 27 human tissues. The chemicals are compiled from three resources that provide tissue-specific data from exposure studies - CTD Exposome Explorer and PubChem. The chemicals from these resources have been filtered to retain only those that have been detected in

B


SIMPLE SEARCH
Chemical name
For example: 'Bisphenol A'
Chemical identifier (PubChem or CAS)
For example: '319-85-7 or '3036'
Search
PHYSICOCHEMICAL FILTER
Molecular Weight
LogP
TPSA
Choose here
Hydrogen bond acceptors (HBA)
Hydrogen bond donors (HBD)
Heavy atoms
Rotatable bonds

C


BROWSE
ENVIRONMENTAL CHEMICALS
BY
CHEMICAL NAME
Chemical name
Choose here
Browse
(OR)
PRESENCE IN HUMAN TISSUE
Human tissue
Choose here
Browse

D

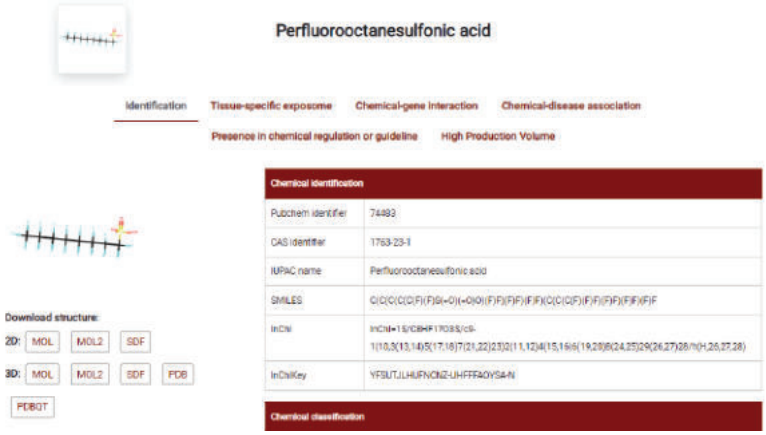

Perfluorooctanesulfonic acid
Identification
Tissue-specific exposome
Chemical-gene interaction
Chemical-disease association
Presence in chemical regulation or guideline
High Production Volume
Chemical identification
Pubchem identifier
74483
CAS identifier
1763-23-1
IUPAC name
Perfluorooctanesulfonic acid
SMILES
InChI
InChIKey
YFSUTJLHUFNCNZ-UHFFFAOYSA-N
Chemical classification
Download structure:
2D:
MOL
MOL2
SDF
3D:
MOL
MOL2
SDF
PDB
PDBQT

Figure 8.3 ***(previous page)*****:** The web interface of TExAs. (A) Screenshot of the TExAs home page. (B) The search page facilitates search for chemicals in two ways: Chemical search and Physicochemical filter. In the Chemical search option, a chemical can be searched using the chemical name or standard identifiers (CAS or PubChem). Using Physicochemical filter, the chemicals can be searched using physicochemical properties such as molecular weight, LogP, TPSA, number of rotatable bonds, number of hydrogen bond donors, or acceptors. (C) On the browse page, the chemical(s) can be obtained using either chemical name or based on their presence in 27 human tissues. (D) Screenshot showing the result page for each chemical compiled in TExAs. From the result page, chemical information including the structural identifiers, tissue-specific exposome, chemical-gene interaction, chemical-disease association, presence in chemical regulation or guideline, and presence of chemical in high production volume (HPV) lists can be obtained for each chemical.

analysis with 55 publicly available chemical inventories, regulations, and guidelines (Supplementary Table S8.2). Based on the nature of human exposure, these 55 chemical lists were classified into 8 external exposome categories such as 'Children's exposome', 'Dietary exposome', 'External environmental exposome', 'Indoor exposome', 'Occupational exposome', 'Pesticide/biocide exposome', 'Skin exposome' and 'Miscellaneous external exposome' (Supplementary Table S8.2). We find that 300 out of the 380 environmental chemicals in TExAs were also part of at least one of 55 chemical lists corresponding to chemical inventories, regulations, and guidelines (Supplementary Table S8.3). Further based on classification of these 55 chemical lists into various categories of the external exposome, we found the majority of environmental chemicals in TExAs belong to 'Dietary exposome' (192 chemicals) followed by 'External environmental exposome' (189 chemicals) (Figure 8.2D; Supplementary Table S8.3). The least number of environmental chemicals in TExAs belong to 'Occupational exposome' (4 chemicals), which may be due to data being limited to only one chemical regulatory list within this category (Figure 8.2D) [41].

Further to understand the scale at which humans are exposed to these chemicals, we have also compared against chemicals produced in high volume as compiled in the Organisation for Economic Cooperation and Development High Production Volume (OECD HPV) list which was last updated in 2004 and the United States High Production Volume (USHPV) database. We find that 109 of 300 environmental chemicals detected in hu-

man tissues and present in at least one of the 55 chemical lists, are also produced in high volume as per the OECD HPV list and USHPV database. Figure 8.2D shows the distribution of these 300 environmental chemicals across 8 exposome categories along with the HPV chemicals in each exposome category. The above-mentioned 109 environmental chemicals produced in high volume have been detected in at least one of 27 human tissues [41].

The high production volume of these chemicals also indicates their potential to cause severe or widespread exposure. We, therefore, sought to understand their hazard potential by comparing them with the substances of very high concern (SVHC) list under Registration, Evaluation, Authorisation and Restriction of Chemicals (REACH) regulation of the European Union (EU) [157]. The chemicals in SVHC have been identified as bioaccumulative, carcinogenic, mutagenic, or linked to serious health effects. Table 8.1 gives the list of 13 potentially hazardous chemicals in TExAs that have also been included in the SVHC list along with the information about the human tissues in which they have been detected. The table also provides the criteria for their inclusion under the SVHC candidate list. These 13 potential hazardous chemicals fall into 7 external exposome categories namely 'Children's exposome', 'Dietary exposome', 'External environmental exposome', 'Indoor exposome', 'Skin exposome' 'Pesticide/biocide exposome' and 'Miscellaneous external exposomes' (Table 8.1). Of these 13 chemicals listed under SVHC, 3 are carcinogens, 4 are endocrine disruptors and 5 are known to cause reproductive toxicity (Table 8.1). Notably, these 13 chemicals have been detected across 13 out of 27 human tissues in TExAs which include the brain, breast, kidney, liver, lung, pancreas and placenta. These findings highlight the various possible routes of human exposure, potential health concerns, and the implications for global monitoring and regulation of these 13 hazardous chemicals in the future.

8.4 Linking diseases to the tissue-specific external exposome

Previous studies have suggested linkages between exposures, genes and gene expression, and disease origins [368]. Earlier studies have also shown tissue specificity in the expression and interaction of genes, corresponding to the tissue-specific manifestation of diseases [119]. Network biology [88] approaches can help in identifying mechanistic links between the chemical spaces and their biological outcomes upon exposure [13]. Such analysis may also shed light on the tissue-specificity of the targets of the chemicals, which can further help in the risk assessment of potential hazardous chemicals. Thus, we construct a tripartite chemical-gene-disease network (considering only human tissue-specific genes) to understand the effect of these environmental chemicals detected across 27 human tissues (Figure 8.1). We do so through the following steps.

8.4.1 Tissue-specific target genes of chemical exposome

To retrieve tissue-specific target genes of the environmental chemicals detected in human tissues, we have used ToxCast [89] invitroDB3 dataset released in August 2019 [215] for our analysis. Although there are resources like Human Protein Atlas (HPA) [369] which provide the list of proteins expressed in different tissues, ToxCast [89] is the only resource that can provide tissue-specific chemical-gene associations based on experimental assays performed on human cell lines across different tissues. The assay summary file Assay_Summary_190708.csv from ToxCast invitroDB3 dataset [215] contains a detailed annotation of assay type, assay component, assay component endpoint and their corresponding tissue-specific target information for tested chemicals across different cell lines. To get the human tissue-specific target genes for the tested chemicals, we have excluded ToxCast assays which are not specific to humans or lack tissue-specific gene information. The ToxCast assay activity information file hitc_Matrix_190708.csv provides

data on whether a tested chemical is active or inactive for a particular assay component endpoint, corresponding to specific target genes. If a tested chemical is active for a particular assay component endpoint, then the corresponding tissue-specific target gene is assigned to the tested chemical. In total, ToxCast invitroDB3 dataset [215] compiles information based on various assays for 6623 tested chemicals that can target 138 genes present across 13 human tissues. Importantly, 9 out of the 13 human tissues for which information is compiled in ToxCast were mapped to the set of 27 human tissues compiled in TExAs. ToxCast provides tissue-specific chemical-gene interaction data for 13 human tissues, and we were able to map 9 out of the 27 human tissues in TExAs to their equivalent tissue names in ToxCast. For subsequent analysis, we have considered the chemicals in TExAs for which target gene information, across these 9 human tissues, is available in ToxCast. The chemical-gene interaction network built as a result of this analysis shows that 158 chemicals from TExAs interact with 121 gene targets, corresponding to 9 human tissues. Among these 9 tissues, only kidney, liver, lung, skin and vascular tissues have chemical-gene interaction information for 10 or more targets (Supplementary Table S8.4) [41].

8.4.2 Tissue-specific gene-disease associations of chemical exposome

To construct the tissue-specific gene-disease association network, we have used the curated gene-disease associations dataset in DisGeNET [370], which was compiled from PsyGeNET [371], UniProt [372], OrphaNet [373], CGI [374], CTD (human data) [30], ClinVar [375], and the Genomics England PanelApp [376]. DisGeNET also gives different scores which can be used to rank the compiled associations such as the gene-disease associations (GDA) score, Disease Specificity Index (DSI), and Evidence Index (EI) which range from 0 to 1 [370]. In our study, we first filtered high confidence gene-disease associations from DisGeNET using the GDA score cut-off of > 0.5. Note that the GDA score considers the level of curation, data source, test organisms and the number of associated publications [370]. Next, we filtered the resulting data using the EI cut-off

of > 0.5, which implies that at least 50% of the publications supporting the gene-disease associations are validated. Lastly, we chose only the gene-disease associations in which disease types are classified as 'disease'. After applying the above-mentioned filters in DisGeNET, we have retrieved the list of gene-disease associations for the target genes compiled in the previous step.

8.4.3 Network view of the relationships between tissue-specific chemical exposome and human diseases

The manifestation of human diseases is affected by the interplay of multiple tissue-specific genes [119], and therefore, multiple interactions between the environmental chemicals in the human exposome and their biological targets [368]. We employ a network biology [88] approach to better understand the interaction patterns of the environmental chemicals detected in human tissues with their tissue-specific gene targets, and to draw insights into the mechanistic linkages of chemical exposure and disease relationships [13]. Specifically, we have constructed a tissue-specific chemical-gene-disease network for the environmental chemicals compiled in TExAs using ToxCast [89] and DisGeNET [370] based on the shared genes. This ultimately resulted in a tripartite chemical-gene-disease network comprising 148 environmental chemicals, 60 target genes, and 191 associated diseases across 9 tissues (Figure 8.2E; Supplementary Table S8.4).

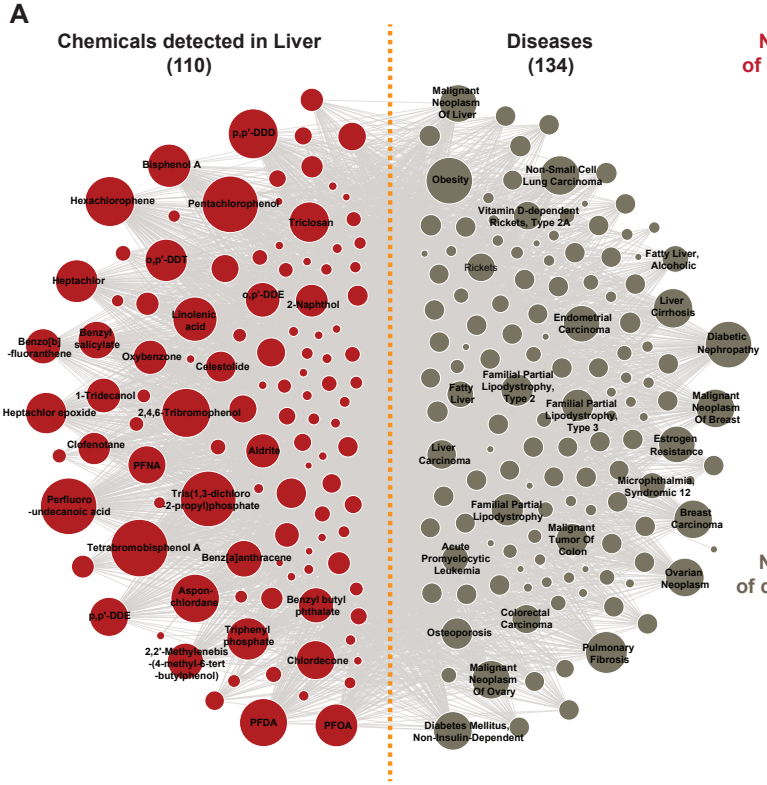
The liver is the human tissue with the largest number of linkages, consisting of 110 environmental chemicals targeting 35 genes which are associated with 134 diseases. Among these chemicals, Tetrabromobisphenol A is predicted to be associated with the maximum number (107) of diseases (Figure 8.4A; Supplementary Table S8.5). An inspection of the external exposome categories of these 110 environmental chemicals shows that a majority of them (81 chemicals) fall under the 'External environmental exposome' category (Supplementary Table S8.3). The 'External environmental exposome' category consists of 9 chemical lists including substances which are labelled hazardous, regulated, or re-

stricted for human exposure, and present as water or environmental contaminants. This result highlights the role and burden on the liver with regard to the environmental exposures of humans. We further discuss the health implications of this chemical burden on the liver [41].

Among the 134 diseases linked to the liver via chemical exposure, obesity and diabetic nephropathy are found to be associated with the maximum number (84) of the environmental chemicals detected in the liver (Figure 8.4A; Supplementary Table S8.5). Due to the shared chemical linkages amongst the diseases associated with the liver, we sought to understand possible connections and co-occurrences among them. We construct a liver-specific disease-disease network based on these shared chemicals. Analysis of such disease-disease networks could also give insights on commonalities in the biological mechanisms of diseases associated with shared chemicals. To get the most significant disease associations, we have computed the overlap score for each pair of diseases. The overlap score is the ratio of the number of chemicals shared between two diseases and the total number of chemicals detected in the tissue. Thus, the strength of the association between two disease pairs is proportionate to the overlap score, which ranges from 0 to 1. Here, we have used an overlap score ≥ 0.5 as the cut-off, to retrieve the most significant disease associations based on the shared chemicals.

Upon analysis of this disease-disease network, we found obesity to be associated with 12 other diseases, affecting different organs and biological systems such as the endocrine system, kidney, liver, and lung (Figure 8.4B; Supplementary Table S8.6). Notably, obesity is found to be associated with other liver diseases including liver cirrhosis and malignant neoplasm of the liver (Figure 8.4B), which describes the collective form of liver cancer or hepatocellular carcinoma [377]. Previous studies also show that obesity shares common biological mechanisms with liver cirrhosis and liver cancer [378–381]. We note that 48 environmental chemicals are shared among obesity, liver cirrhosis, and malignant neoplasm of the liver. Of these 48 chemicals, PFOA, DDT, DDE, bisphenol A are known obesogens [381–383].

A



Number of diseases	Chemicals with >70 associated diseases
107	Tetrabromobisphenol A
105	Pentachlorophenol
105	Perfluoroundecanoic acid
103	Tris(1,3-dichloro-2-propyl) -phosphate
91	Hexachlorophene
91	p,p'-DDD
89	2,4,6-Tribromophenol
88	Aspon-chlordane
87	PFDA
79	Linolenic acid
77	Bisphenol A
76	Heptachlor
76	PFOA
75	o,p'-DDT
75	Triphenyl phosphate
73	Heptachlor epoxide
71	Triclosan

Number of chemicals	Diseases associated with >60 chemicals
84	Diabetic Nephropathy
84	Obesity
74	Pulmonary Fibrosis
69	Liver Cirrhosis
69	Non-Small Cell Lung Carcinoma
67	Breast Carcinoma
67	Diabetes Mellitus, Non-Insulin-Dependent
67	Malignant Neoplasm Of Breast
67	Malignant Neoplasm Of Ovary
67	Ovarian Neoplasm
65	Malignant Neoplasm Of Liver
63	Endometrial Carcinoma
63	Estrogen Resistance

B

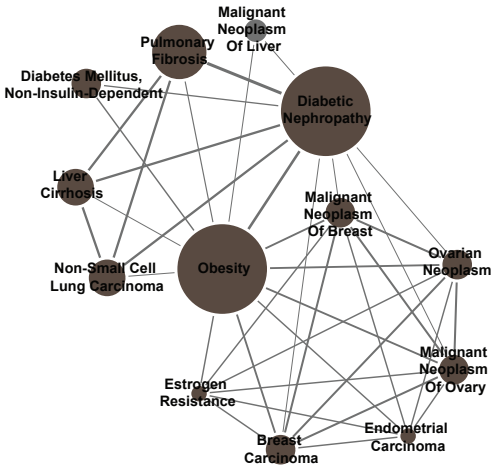


Figure 8.4 ***(previous page)*:** (A) The bipartite network of 110 chemicals detected in the liver and 134 associated diseases. In this network, the chemical nodes are colored in 'red' while the disease nodes are colored in 'grey'. The table (on the right) gives the list of chemicals detected in liver with more than 70 disease associations and diseases associated with more than 60 chemicals detected in liver. (B) Liver-specific disease-disease network built using the most significant disease-disease associations with an overlap score of ≥ 0.5. The overlap score is the ratio of the number of chemicals shared between any two diseases and the total number of chemicals detected in the tissue.

In summary, we present TExAs [41] that compiles a list of 380 environmental chemicals detected across 27 human tissues in published literature compiled in three existing resources. TExAs provides detailed information regarding the structures, chemical classification, and exposome categories for these 380 environmental chemicals. For the environmental chemicals in TExAs, we show the application of network biology approaches to explore chemical exposure-disease relationships in understanding the health burden of chemicals and the possibilities of disease comorbidities.

A large quantum of data regarding tissue-specific chemical exposures still remains dispersed in scientific literature, and a substantial effort is required to compile the complete information in published literature. Our compilation of the 380 environmental chemicals detected across 27 human tissues is limited to the compilation and curation of published literature captured in three resources (CTD, Exposome-Explorer and PubChem), rather than an extensive search for published studies in PubMed or Google Scholar. Another limitation of this analysis is the use of ToxCast assays in identifying tissue-specific gene targets of chemicals. As pointed out by Borrel *et al.* [123], the liver is the most represented tissue or organ type in ToxCast assays, and other tissues are not represented to the same extent. Nevertheless, ToxCast is the only resource available to date which contains chemical-gene interactions tested on assays specific to tissue types. Another barrier to a comprehensive understanding of tissue-specific exposure-disease relationships is the gap in the compilation of data surrounding the tissue-specific target genes of chemicals. For a better understanding of tissue-specific exposure-disease relationships, it is important to study the complete functional sub-network of genes (or disease modules) which are ex-

pressed within the particular tissue [119]. While the Human Protein Atlas (HPA) gives comprehensive information on the expression profiles of human genes in more than 50 tissue types [369], however, this presents only one side of the story as it is not linked to any chemical exposures.

This study is the first step towards the integration of data surrounding chemicals detected across human tissues into a single resource, which will help future exposome research. Systematic expansion of tissue-specific exposure data along with the integration of large-scale gene expression data will enable a better understanding of tissue-specific chemical-disease relationships and the impact of chemical combinations in tissues. From the perspective of chemical regulations, this expansion in data could guide the prioritization and regulation of environmental chemicals in the future. From the perspective of future research, several parallels and contrasts could be identified in chemical-disease associations when a chemical is present in different tissues. We believe the continued expansion, compilation, and standardization of exposure data, gene expression data, and gene-disease linkages are essential to understand the full impact of the external exposome on human health.

8.5 Discussion

We wish to note that our focus in this study has been to meaningfully integrate and explore the available data surrounding environmental chemicals and their tissue-specific disease associations, rather than to expand on the isolated compilation of environmental chemicals [41]. We obtain two important insights via our network-centric analysis. The first is the significant effect that environmental exposures can have on human health. The second is the interconnections and possible co-occurrence of diseases, specific to tissues. Such linkages between diseases have also been discussed in other studies [384]. This work could serve as a template for the development of similar network biology approaches to understand other exposure-disease relationships, character-

ize the effect of chemicals, and study exposome-related comorbidities [13]. The data integrations that led to these findings have been made available through a web interface (https://cb.imsc.res.in/texas) for use by the scientific community and the public alike.

Supplementary Information

Supplementary Tables S8.1-S8.6 associated with this chapter are available for download from the GitHub repository: https://github.com/asamallab/PhDThesis-Janani_R/blob/main/SI/ST_Chapter8.xlsx.

Chemical name	Presence in USHPV	Presence in OECD HPV	Presence in SVHC	SVHC Criteria
Decabromodiphenyl oxide	Yes	Yes	Yes	PBT (Article 57d); vPvB (Article 57e)
Bis (2-ethylhexyl)phthalate	Yes	Yes	Yes	Toxic for reproduction (Article 57c); Endocrine disrupting properties (Article 57(f) - environment); Endocrine disrupting properties (Article 57(f) - human health)
Anthracene	Yes	Yes	Yes	PBT (Article 57d)
Dechlorane plus	Yes	Yes	Yes	vPvB (Article 57e)
Octamethylcyclote-trasiloxane	Yes	Yes	Yes	PBT (Article 57d); vPvB (Article 57e)
Lead	Yes	Yes	Yes	Toxic for reproduction (Article 57c)
Cadmium	Yes	Yes	Yes	Carcinogenic (Article 57a); Specific target organ toxicity after repeated exposure (Article 57(f) - human health)
Arsenic acid	Yes	Yes	Yes	Carcinogenic (Article 57a)
Trichloroethylene	Yes	Yes	Yes	Carcinogenic (Article 57a)
Bisphenol A	Yes	Yes	Yes	Toxic for reproduction (Article 57c); Endocrine disrupting properties (Article 57(f) - environment); Endocrine disrupting properties (Article 57(f) - human health)
Musk xylene	Yes	Yes	Yes	vPvB (Article 57e)
Dibutyl phthalate	Yes	Yes	Yes	Toxic for reproduction (Article 57c); Endocrine disrupting properties (Article 57(f) - human health)
Benzyl butyl phthalate	Yes	Yes	Yes	Toxic for reproduction (Article 57c); Endocrine disrupting properties (Article 57(f) - human health)

Table 8.1: List of 13 chemicals detected in human tissues that are found to be produced in high volume by both OECD HPV list and USHPV database, and are also listed as 'substance of very high concern (SVHC)' by the European Chemicals Agency (ECHA).

Chapter 9

Summary and future outlook

In this thesis, we investigated five diverse groups of environmental chemicals including endocrine disrupting chemicals (EDCs) [35–37], environmental neurotoxicants [38], human milk contaminants [39], fragrance chemicals in children's products [40], and exogenous chemicals detected in human tissues [41]. Importantly, the research reported in this thesis highlights the possible links between chemical exposome and human health (Figure 9.1). By employing network science and systems biology approaches, we identified the perturbed target genes, perturbed pathways, and diseases associated with environmental chemical exposures (Figure 9.1). In the following section, we provide a summary of the research reported across different chapters of this thesis. Thereafter, we conclude with a short discussion of the possible future directions based on the research reported in this thesis.

9.1 Summary

DEDuCT 1.0: A curated knowledgebase on endocrine disrupting chemicals and their biological systems-level perturbations

EDCs are chemicals of emerging concern that have the potential to cause hormonal imbalance by interfering with the normal functioning of endocrine system [3, 4, 43]. In Chap-

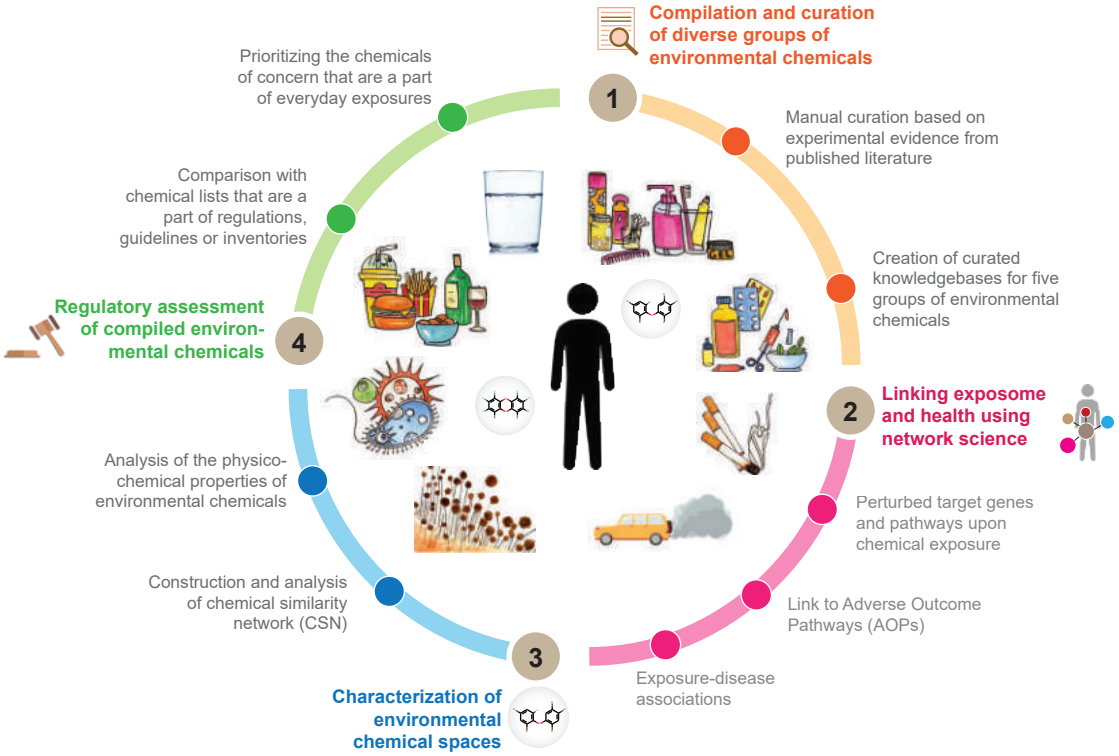


Figure 9.1: Summary of the research on compilation, curation and exploration of diverse groups of environmental chemicals reported in this thesis.

ter 2, we developed a detailed workflow (Figure 2.1) to identify potential EDCs from published research articles containing supporting experimental evidence for endocrine-specific perturbations in humans or rodents. In the initial stage of the workflow, we used extensive PubMed [158] literature mining and three existing resources, the WHO report, TEDX and EDCs Databank, to compile more than 16000 published research articles which are likely to contain information on EDCs. Subsequently, we process these articles using our workflow to manually compile 686 potential EDCs from 1796 published research articles containing supporting experimental evidence for endocrine-specific perturbations in humans or rodents. Of these 686 potential EDCs and 1796 research articles, 198 EDCs (28.9%) and 1294 articles (72.0%) are not captured in any of the three existing resources integrated in our workflow. A unique feature of our work is the compilation of the list of observed adverse effects or endocrine-specific perturbations from supporting published experiments for the 686 EDCs, and these observed effects were manually curated, unified and standardized into a list of 514 endocrine-mediated endpoints spanning

7 systems-level perturbations. Another unique feature of our work is the compilation and standardization of the dosage information at which endocrine-mediated effects were observed upon individual EDC exposure in published experiments. Moreover, the 686 EDCs were classified based on the type of supporting evidence in published experiments, their environmental source and their chemical classification. Lastly, we have also compiled additional detailed information for each EDC such as its two-dimensional (2D) and three-dimensional (3D) structure, physicochemical properties, molecular descriptors, predicted ADMET properties and experimentally inferred target genes. In order to widely share the compiled information on 686 potential EDCs and enable basic research towards the elucidation of systems-level perturbations caused by them, we have also created a webserver, DEDuCT 1.0, which is accessible at: https://cb.imsc.res.in/deduct/.

We employed network biology approaches [88, 385, 386] to gain a better understanding of the link between the underlying chemical space of EDCs and biological space of target genes or perturbed pathways [387, 388]. Specifically, we have constructed two networks of EDCs using our resource based on the similarity of chemical structures or target genes. Based on the chemical similarity network, we find that EDCs are diverse in their chemical structure and each module in the similarity network corresponds to distinct chemical features. Upon investigation of the target similarity network, we find that EDCs can have very different sets of target genes. Subsequent analysis revealed a lack of correlation between chemical structure and target genes of EDCs. These results highlight potential challenges in developing predictive models for the identification of EDCs. DEDuCT is a large-scale resource on potential EDCs compiling supporting evidence of endocrine-mediated perturbations and dosage information from published experiments in humans or rodents, and the compiled information will contribute to the future research in the field of computational systems toxicology.

DEDuCT 2.0: An updated knowledgebase and an exploration of the current regulations and guidelines from the perspective of endocrine disrupting chemicals

We next explored how knowledge on EDCs captured through academic research can help in risk and regulatory assessment of EDCs. This analysis was carried out in three steps, as described in Chapter 3. Firstly, we have analyzed the increase in research efforts and knowledge on EDCs in past decades, and have captured newly available information into our unique resource DEDuCT 2.0 (Figure 3.1). Thus, the updated knowledgebase, DEDuCT 2.0, compiles 792 potential EDCs along with 609 unique endocrine-mediated endpoints, spanning 7 systems-level perturbations. Secondly, we analyzed the distributions of 1856 potential EDCs compiled in DEDuCT 2.0 or three other resources, namely, WHO report, TEDX and EDCs Databank, across 36 chemical lists which are part of inventories, guidelines and regulations. Notably, we found several potential EDCs are distributed across diverse chemical lists, and further, some of these chemical lists with potential EDCs are in day-to-day product categories such as 'Food additives and Food contact materials' and 'Cosmetics and household products'. Moreover, we classified the chemicals in SIU and SOC lists into groups I, II and III containing 23483, 1139 and 3223 chemicals, respectively, of which 242, 356 and 278, respectively, are potential EDCs. Lastly, analysis of 242 group I EDCs with HPV chemicals found 63 group I EDCs in use which are also produced in high volume. Given the scale of exposure and the related hazard potential, an evaluation of these EDCs produced in large quantities is warranted, and developing adequate risk assessment criteria will aid in such efforts. We also described an example to demonstrate how the compiled information in curated knowledgebases like DEDuCT 2.0 can aid in the risk assessment of EDCs.

In sum, this chapter emphasizes the importance of bridging the gap between academic and regulatory aspects of chemical safety, as a step towards the better management of environment and health hazards such as EDCs. As ongoing scientific research will lead to new discoveries and a deeper understanding of the effects of chemical exposure, it will be

important to regularly monitor the substances permitted for use under various regulations, and substances generally found in use in products, through the same lens of scientific risk assessment, in order to restrict emerging substances of concern at the earliest. Inventories and independent guidelines of hazardous or toxic substances also need to be evaluated and brought under effective regulation. Information with a scientific basis is necessary to standardize criteria for this evaluation and risk assessment, especially in the case of a complex chemical class such as the EDCs.

Derivation, characterization and analysis of an Adverse Outcome Pathway network relevant for endocrine disruption

To understand the perturbed biological mechanisms upon exposure to EDCs, we developed a comprehensive adverse outcome pathway (AOP) network using existing knowledge compiled in AOP-Wiki [114]. In **Chapter 4**, we describe the steps involved in the characterization, development and investigation of an adverse outcome pathway (AOP) network derived to capture the endocrine-mediated perturbations resulting from environmental exposure [37]. In this work, we assess the quality and completeness of information of each AOP compiled in AOP-Wiki, and thereafter, identified 48 high-confidence AOPs relevant to endocrine disruption, i.e., 48 ED-AOPs. We proposed a cumulative weight of evidence score for these 48 ED-AOPs that is an indicator of the strength of empirical evidence for Key Events (KEs) and Key Event Relationships (KERs) in them. We evaluated the biological domain information extracted from AOP-Wiki for the 48 ED-AOPs, including taxonomic, sex, and life stage applicability. Subsequently, we constructed an ED-AOP network by assembling the information on shared KEs and KERs among 48 ED-AOPs capturing diverse biological perturbations related to the endocrine system.

Connectivity analysis of this ED-AOP network comprising 48 ED-AOPs reveals 7 connected components and 12 isolated AOPs. We performed a graph-theoretic analysis of the directed ED-AOP network corresponding to the two largest connected components (LCCs) to reveal important topological features using four standard measures namely, in-

degree, out-degree, betweenness centrality and eccentricity. These analyses lead to the identification of important events including points of convergence or divergence in the ED-AOP network. In particular, we focused on one of the LCCs of the ED-AOP network to better understand the series of biological events that lead to systems-level perturbations upon endocrine disruption. An in-depth analysis of the largest component in the ED-AOP network sheds light on the systems-level perturbations caused by endocrine disruption, emergent paths, and stressor-event associations. In sum, the derived ED-AOP network can be used to address the current knowledge gaps in the existing regulatory framework and aid in better risk assessment of environmental chemicals.

NeurotoxKb 1.0: compilation, curation and exploration of a knowledgebase of environmental neurotoxicants specific to mammals

Exposure to environmental chemicals can lead to various neurological disorders and neurotoxic effects which can manifest at any stage of human life, from infancy to old age [52,53]. In Chapter 5, we describe a detailed workflow (Figure 5.1) to identify potential non-biogenic neurotoxicants with evidence specific to mammals from published literature. We created the environmental Neurotoxicants Knowledgebase NeurotoxKb 1.0. An important limitation of the existing resources on neurotoxicants is in their compilation of observed neurotoxic effects using non-standardized vocabulary [60–62] or the complete lack thereof [57, 58]. To overcome this limitation of existing resources, we have performed an extensive manual curation effort to compile, unify and standardize the reported neurotoxic effects for potential neurotoxicants in published literature, into standardized neurotoxic endpoints. In a nutshell, we have identified here 475 potential neurotoxicants which are non-biogenic and have evidence of neurotoxicity specific to mammals from published studies. For these 475 potential neurotoxicants, our compilation includes observed neurotoxic effects in terms of 148 standardized neurotoxic endpoints curated from 835 published studies specific to mammals. For the 475 potential neurotoxicants, we have compiled additional information including chemical structures, chemical classifica-

tion, environmental sources, physicochemical properties, predicted ADMET properties, molecular descriptors and target human genes. The entire information compiled in NeurotoxKb 1.0, on the 475 potential neurotoxicants specific to mammals, is accessible at: https://cb.imsc.res.in/neurotoxkb.

To understand the current state of regulation and monitoring of environmental neurotoxicants through the perspective of exposomes, we analyzed the presence of potential neurotoxicants across 55 chemical lists which include inventories, regulations and guidelines. Notably, based on the source or route of exposure, we classified these 55 chemical lists into different categories of exposome. Thus, the presence of neurotoxicants in these 55 chemical lists is a clear indication of their presence in human exposome. As detection of environmental chemicals in biospecimens is a proof of their exposure, we also analyzed the presence of potential neurotoxicants among chemicals detected in different human biospecimens such as blood, urine, placenta and human milk [269]. Furthermore, based on comparative analyses with current chemical regulations and guidelines, we present a hazard priority list of 18 potential neurotoxicants. In short, we show the utility of our resource in aiding regulatory bodies worldwide in prioritization of hazardous chemicals, to streamline their monitoring and regulation.

We also constructed and analyzed a bipartite network of potential neurotoxicants in NeurotoxKb and their target human neuroreceptors. Moreover, we constructed a chemical similarity network which revealed that the space of potential neurotoxicants in NeurotoxKb is highly diverse. Overall, NeurotoxKb 1.0 is a comprehensive knowledgebase on potential environmental neurotoxicants specific to mammals which will enable future research in neurotoxicology.

ExHuMId: A curated resource and analysis of Exposome of Human Milk across India

Human milk is a significant biospecimen in the study of the mother exposome and a vital factor in a newborn's exposome. In this direction, we created **Ex**posome of **Hu**man

Milk across **Ind**ia (ExHuMId) version 1.0, an India-specific repository containing 101 human milk contaminants detected in milk samples from 13 Indian states, compiled from 36 published experimental studies. The detailed steps involved in this compilation of human milk contaminants is presented in Chapter 6. ExHuMId also compiles the detected concentrations of the contaminants, structural and physicochemical properties, and factors associated with the donor of the sample. In this chapter, we also considered human milk contaminants studied by Lehmann *et al*. [286] that are specific to USA (referred to as 'ExHuMUS'), and the human milk contaminants compiled in Exposome-Explorer [24] that are not specific to any geography (referred to as 'ExHuM Explorer').

We analyzed the human milk contaminants compiled in ExHuMId and two other resources from three perspectives. We first compared ExHuMId with the well-known chemical lists representing regulations and guidelines, to identify potential EDCs, carcinogens, neurotoxins or other hazardous chemicals. Of 101 human milk contaminants in ExHuMId, 43, 23 and 14 were found to be potential EDCs, carcinogens, and neurotoxicants, respectively. Similar analyses was performed on the human milk contaminants compiled in ExHuMUS and ExHuM Explorer [62], and several chemicals of concern produced in high volume were identified.

The second perspective of our analysis enables to better understand the structural features and properties which influence the transfer of environmental contaminants into human milk, and thus, provides a way to predict the risk of contaminant entering human milk. Due to the lack of experimental data on M/P ratios of human milk contaminants in ExHuMId, we considered the dataset reported by Vasios *et al*. [72] and performed a comparison of the physicochemical properties that have been widely reported to influence the transfer of contaminants or drugs into human milk. Through our analysis we observed that the distributions of physicochemical properties of contaminants in ExHuMId, ExHuMUS and ExHuM Explorer are close to the distributions of physicochemical properties of chemicals reported as highly likely to transfer to human milk in Vasios *et al*. [72].

The third aspect of our analysis predicts the effect of the human milk contaminants

on lactation pathway and cytokine signalling and production pathway, using a systems biology approach. Based on the interaction data obtained from ToxCast and CTD, we inferred that many of the human milk contaminants compiled in the above-mentioned 3 datasets can interact with genes associated with prolactin signalling, oxytocin signalling, lactose synthesis, cytokine signalling and xenobiotic transport. These observations need to be critically validated using experimental approaches, which should encompass various disciplines, to understand the influence of environmental contaminants on maternal and infant health [302]. In sum, from our systematic compilation and analysis of human milk contaminants, we observed there is a need for better chemical regulation and policy decisions to avoid these contaminants in human milk in India and globally.

FCCP: A repository of fragrance chemicals in children's products

Fragrance chemicals are either natural or synthetic compounds, and exposure to such chemicals can lead to asthma, contact dermatitis (irritant or allergic), dyschromia, photosensitivity, and migraine headaches [73–76, 78]. In **Chapter 7**, we present the repository of Fragrance Chemicals in Children's Products (FCCP) that compiles 153 fragrance chemicals from 21 published experimental studies. The fragrance chemicals in FCCP are classified based on their chemical structure, children's product source, chemical origin, and odor profile. Firstly, ClassyFire based classification revealed that all the compiled fragrance chemicals were 'Organic compounds'. Secondly, we find that 85 fragrance chemicals have their children's product source as 'Toys' based on the compiled information on children's product source for the fragrance chemicals. Thirdly, classification based on environmental source showed that 97 fragrance chemicals in FCCP are natural compounds. Fourthly, the odor profiling showed that 'Aromatic' odor is prevalent among the compiled fragrance chemicals in FCCP.

Since the fragrance chemicals in children's products are known to be poorly regulated, we sought to explore the current regulatory status of these chemicals and the potential health effects in children upon exposure. We analyzed the presence of the compiled

fragrance chemicals in different chemical lists that are a part of regulations and guidelines including the ones that are specific to children. We find that several fragrance chemicals in FCCP are either banned allergenic chemicals, or are prohibited or restricted in cosmetics and perfumes, based on a comparison with 21 chemical lists representing current guidelines or regulations. Specifically, the analysis revealed that 17, 15, 8, and 21 fragrance chemicals in FCCP are also carcinogens, endocrine disruptors, neurotoxicants and phytotoxins, respectively.

Further, we analyzed the structural diversity of the space of compiled fragrance chemicals and banned allergenic fragrance chemicals in EU Toy Safety Directive [145]. This similarity network-based analysis of the fragrance chemicals in FCCP revealed the diversity of the associated chemical space. We then identified the potential skin sensitizers among the compiled fragrance chemicals in children's products by leveraging ToxCast assays. The compiled information in FCCP can aid scientists, stakeholders and regulatory agencies in risk assessment and develop safer products for children. FCCP is accessible at: https://cb.imsc.res.in/fccp/.

Network-based exploration of a human tissue-specific chemical exposome atlas (TExAs)

The presence of chemicals in human tissues suggests long-term exposure and bioaccumulation of environmental contaminants [85]. In **Chapter 8**, we describe the steps involved in the compilation of environmental chemicals detected across human tissues. In this chapter, we explored the patterns in the associations between tissue-specific chemical exposures and human diseases using network biology approaches. For this purpose, we compile, filter and unify environmental chemicals that are detected across human tissues using information in CTD [30], Exposome-Explorer [24], and PubChem [86]. This resulted in the compilation of 380 environmental chemicals detected across 27 human tissues. We find that 240 environmental chemicals were detected in adipose tissue, followed by 120 chemicals in the placenta, among information for 380 chemicals across 27 human

tissues in our compiled dataset.

We also find that 300 out of the 380 environmental chemicals are present in at least one of 55 chemical lists that are part of global chemical regulations, guidelines, or inventories. Interestingly, we find that 109 of the 300 chemicals that are present in at least one of the 55 chemical lists, are also produced in high volume. Based on the classification of these 55 chemical lists into various external exposome categories, we find that 192 environmental chemicals belong to the 'Dietary exposome', followed by 189 chemicals that belong to the 'External environmental exposome'. Further, we propose a priority list of 13 potentially hazardous chemicals based on a comparative analysis of the compiled chemicals with SVHC REACH regulation [157] and high production volume chemicals. This analysis helps in understanding the environmental sources and routes of human exposure to environmental chemicals detected in human tissues, as well as the current status of their monitoring and regulation.

Subsequently, the compiled environmental chemicals have been linked to their potential gene targets using ToxCast assays, and to the associated diseases using DisGeNET [370]. This information was used to construct a tissue-specific chemical-gene-disease network. Specifically, we considered the role and burden of the liver towards the environmental exposures of humans. An analysis of the liver-specific disease network reveals the possibilities of disease comorbidities and demonstrates the application of network biology in unravelling complex exposure-disease associations. The entire information is compiled in Human **T**issue-specific **Ex**posome **A**tla**s** (TExAs), and accessible at https://cb.imsc.res.in/texas.

9.2 Future outlook

If we begin to diligently care for the environment, it will greatly improve human health.

- Lailah Gifty Akita

Human exposome is one of the promising areas of scientific research which aims to address human health issues caused by environmental exposures [389]. Ongoing research in exposome and toxicology is generating a large quantity of experimental data related to various environmental chemical exposures [42]. It is critical to mine and curate existing toxicological data in order to reveal significant and meaningful associations between environmental exposures and health impacts. In this direction, we present highly curated resources on diverse groups of environmental chemicals in this thesis. These knowledgebases will serve as one-stop resource for obtaining toxicological information and can aid in fundamental research on different groups of environmental chemicals. Specifically, in recent times, there is lot of interest in developing data-driven predictive models to identify toxicological effects upon exposure to certain chemicals [390–392]. Such models can be built using high-quality toxicological information compiled for a specific group of chemicals in the knowledgebases presented in this thesis. In future, the observed health effects and/or structural information compiled for different environmental chemicals in our resources can serve as a positive dataset for structure-activity relationship (SAR) studies, which rely on the quality of chemical and toxicological data in both training and testing datasets [390]. Further, chemical similarity networks or CSNs enable the visualization and characterization of the diverse biologically-relevant environmental chemical spaces, and can aid in analyzing the structural relationship between compounds having same or different biological activity.

The ever-increasing rate at which new chemicals are introduced into the market necessitates regular monitoring of their possible health consequences. The presence of the different groups of environmental chemicals compiled in our resources across various product categories reflects the gap in the current chemical regulation. These results also highlight the need to bridge the gap between scientific research in academia and regulatory aspects of environmental chemicals of potential concern. Such analysis can aid in the early identification of hazardous compounds and chemical prioritization, allowing regulatory agencies to expedite the process of safety testing and, as a result, improving

chemical safety standards. Further investigation of experimentally derived dosage information for observed endocrine-mediated health effects compiled in DEDuCT [35,36] can enable identification of reference dose (RfD) or Tolerable Daily Intake (TDI) or Average Daily Dose (ADD) that can aid in regulatory risk assessment of chemicals [393]. Moreover, for risk assessment of chemicals of potential concern, it is worthwhile to consider the compilation of other toxicological information such as species, sex, route of administration, duration of exposure along with the observed effects upon exposure to environmental chemicals, which is one of the limitations of our compiled resources. In case of EDCs [35,36] or neurotoxicants [38], the inclusion of biomonitoring and epidemiological studies from published literature into our resources in future will broaden the scope of exposure assessment and risk categorization.

Network-based exploration of different spaces of environmental chemicals has helped us to gain insights into various perturbed biological events observed at different levels of biological organization upon chemical exposure. The use of network science and systems biology approaches can bring a new degree of understanding in decoding multifaceted environmental exposures and their associated health impacts [42]. Further integration of multi-omics data including genome, transcriptome, proteome, and metabolome can offer opportunities to measure the effects of the exposome [13, 394]. Such computational approaches can help with efficient chemical regulation while reducing the need for animal experimentation [42]. In this regard, we utilized the framework of AOPs that enable the organization of existing toxicological information to capture important biological events perturbed at the systems-level as a result of EDC exposure [37]. However, the derived ED-AOP network presented in Chapter 4 does not capture the entire complexity of endocrine disruption mechanisms since the construction is based on available information in AOP-Wiki. Such derived AOP networks can also be integrated with different layers of information such as sex, life-stage and species required to answer a specific research question [107,395]. As discussed in Chapter 8, the use of network biology approaches can also offer insights into potential exposure-disease relationships and diseases comorbidities

caused by environmental chemical exposures [40]. We believe that the work detailed in this thesis toward the characterization and compilation of environmental chemicals with potential human health hazards will aid basic research and regulatory bodies in improved risk assessment of such chemicals of concern. Overall, the work reported in this thesis is a step towards clean environment and healthy humankind.

References

[1] U.S. EPA. TSCA Chemical Substance Inventory. https://www.epa.gov/tsca-inventory/about-tsca-chemical-substance-inventory (2015).

[2] U.S. National Toxicology Program. https://ntp.niehs.nih.gov/about/index.html (2017).

[3] Futran Fuhrman, V., Tal, A. & Arnon, S. Why endocrine disrupting chemicals (EDCs) challenge traditional risk assessment and how to respond. *Journal of Hazardous Materials* **286**, 589–611 (2015).

[4] Schug, T. T. *et al.* Designing endocrine disruption out of the next generation of chemicals. *Green Chemistry* **15**, 181–198 (2013).

[5] Meeker, J. D. Exposure to environmental endocrine disrupting compounds and men's health. *Maturitas* **66**, 236–241 (2010).

[6] Mezcua, M. *et al.* Analysis of synthetic endocrine-disrupting chemicals in food: A review. *Talanta* **100**, 90–106 (2012).

[7] Muncke, J. Endocrine disrupting chemicals and other substances of concern in food contact materials: An updated review of exposure, effect and risk assessment. *The Journal of Steroid Biochemistry and Molecular Biology* **127**, 118–127 (2011).

[8] WHO/UNEP. State of the science of endocrine disrupting chemicals - 2012. In Bergman, Å., Heindel, J. J., Jobling, S., Kidd, K. & Zoeller, T. R. (eds.) *Summary for Decision-Makers* (World Health Organization, Geneva, 2013).

[9] Cui, Y. *et al.* The Exposome: Embracing the Complexity for Discovery in Environmental Health. *Environmental Health Perspectives* **124**, A137–A140 (2016).

[10] Landrigan, P. J. *et al.* Health Consequences of Environmental Exposures: Changing Global Patterns of Exposure and Disease. *Children's Health in a Changing Global Environment* **82**, 10–19 (2016).

[11] Shaffer, R. M. *et al.* Improving and Expanding Estimates of the Global Burden of Disease Due to Environmental Health Risk Factors. *Environmental Health Perspectives* **127**, 105001 (2019).

[12] Misra, B. B. The Chemical Exposome of Human Aging. *Frontiers in Genetics* **11**, 1351 (2020).

[13] Vermeulen, R., Schymanski, E. L., Barabási, A.-L. & Miller, G. W. The exposome and health: Where chemistry meets biology. *Science* **367**, 392–396 (2020).

[14] Praveena, S. M. *et al.* Recent updates on phthalate exposure and human health: a special focus on liver toxicity and stem cell regeneration. *Environmental Science and Pollution Research* **25**, 11333–11342 (2018).

[15] Praveena, S. M. *et al.* Phthalates exposure and attention-deficit/hyperactivity disorder in children: a systematic review of epidemiological literature. *Environmental Science and Pollution Research* **27**, 44757–44770 (2020).

[16] Sillé, F. C. M. *et al.* The exposome - a new approach for risk assessment. *ALTEX* **37**, 3–23 (2020).

[17] Misra, B. B. & Misra, A. The chemical exposome of type 2 diabetes mellitus: Opportunities and challenges in the omics era. *Diabetes & Metabolic Syndrome: Clinical Research & Reviews* **14**, 23–38 (2020).

[18] Wild, C. P. Complementing the Genome with an "Exposome": The Outstanding Challenge of Environmental Exposure Measurement in Molecular Epidemiology. *Cancer Epidemiology Biomarkers & Prevention* **14**, 1847–1850 (2005).

[19] Rappaport, S. M. & Smith, M. T. Environment and Disease Risks. *Science* **330**, 460–461 (2010).

[20] Miller, G. W. & Jones, D. P. The Nature of Nurture: Refining the Definition of the Exposome. *Toxicological Sciences* **137**, 1–2 (2014).

[21] Rappaport, S. M. Implications of the exposome for exposure science. *Journal of Exposure Science & Environmental Epidemiology* **21**, 5–9 (2011).

[22] Lioy, P. J. & Rappaport, S. M. Exposure Science and the Exposome: An Opportunity for Coherence in the Environmental Health Sciences. *Environmental Health Perspectives* **119**, a466–a467 (2011).

[23] van Tongeren, M. & Cherrie, J. W. An Integrated Approach to the Exposome. *Environmental Health Perspectives* **120**, a103–a104 (2012).

[24] Neveu, V. *et al.* Exposome-Explorer: a manually-curated database on biomarkers of exposure to dietary and environmental factors. *Nucleic Acids Research* **45**, D979–D984 (2017).

[25] Dong, T. *et al.* Human Indoor Exposome of Chemicals in Dust and Risk Prioritization Using EPA's ToxCast Database. *Environmental Science & Technology* **53**, 7045–7054 (2019).

[26] Wishart, D. *et al.* T3DB: the toxic exposome database. *Nucleic Acids Research* **43**, D928–D934 (2015).

[27] Groh, K. J., Geueke, B., Martin, O., Maffini, M. & Muncke, J. Overview of intentionally used food contact chemicals and their hazards. *Environment International* **150**, 106225 (2021).

[28] Barupal, D. K. & Fiehn, O. Generating the Blood Exposome Database Using a Comprehensive Text Mining and Database Fusion Approach. *Environmental Health Perspectives* **127**, 97008 (2019).

[29] Bessonneau, V., Pawliszyn, J. & Rappaport, S. M. The Saliva Exposome for Monitoring of Individuals' Health Trajectories. *Environmental Health Perspectives* **125**, 077014 (2021).

[30] Davis, A. P. *et al.* Comparative Toxicogenomics Database (CTD): update 2021. *Nucleic Acids Research* **49**, D1138–D1143 (2021).

[31] Vrijheid, M. *et al.* The Human Early-Life Exposome (HELIX): Project Rationale and Design. *Environmental Health Perspectives* **122**, 535–544 (2014).

[32] National Library of Medicine, U. *Drugs and Lactation Database (LactMed)[Internet]* (U.S. National Library of Medicine, Bethesda, MD, 2006).

[33] Fitzpatrick, R. B. LactMed: Drugs and Lactation Database. *Journal of Electronic Resources in Medical Libraries* **4**, 155–166 (2007).

[34] The Organization of Teratology Information Specialists (OTIS). MotherToBaby: Pregnancy & Breastfeeding Exposures. https://mothertobaby.org/fact-sheets/ (2017).

[35] Karthikeyan, B. S., Ravichandran, J., Mohanraj, K., Vivek-Ananth, R. P. & Samal, A. A curated knowledgebase on endocrine disrupting chemicals and their biological systems-level perturbations. *Science of the Total Environment* **692**, 281–296 (2019).

[36] Karthikeyan, B. S., Ravichandran, J., Aparna, S. R. & Samal, A. DEDuCT 2.0: An updated knowledgebase and an exploration of the current regulations and guide-

lines from the perspective of endocrine disrupting chemicals. *Chemosphere* **267**, 128898 (2021).

[37] Ravichandran, J., Karthikeyan, B. S. & Samal, A. Investigation of a derived adverse outcome pathway (AOP) network for endocrine-mediated perturbations. *Science of The Total Environment* **826**, 154112 (2022).

[38] Ravichandran, J., Karthikeyan, B. S., Singla, P., Aparna, S. R. & Samal, A. NeurotoxKb 1.0: Compilation, curation and exploration of a knowledgebase of environmental neurotoxicants specific to mammals. *Chemosphere* **278**, 130387 (2021).

[39] Karthikeyan, B. S., Ravichandran, J., Aparna, S. R. & Samal, A. ExHuMId: A curated resource and analysis of Exposome of Human Milk across India. *Chemosphere* **271**, 129583 (2021).

[40] Ravichandran, J., Karthikeyan, B. S., Jost, J. & Samal, A. An atlas of fragrance chemicals in children's products. *Science of The Total Environment* **818**, 151682 (2022).

[41] Ravichandran, J., Karthikeyan, B. S., Aparna, S. R. & Samal, A. Network biology approach to human tissue-specific chemical exposome. *The Journal of Steroid Biochemistry and Molecular Biology* **214**, 105998 (2021).

[42] Kalia, V., Jones, D. P. & Miller, G. W. Networks at the nexus of systems biology and the exposome. *Current Opinion in Toxicology* **16**, 25–31 (2019).

[43] Zoeller, R. T. *et al.* Endocrine-Disrupting Chemicals and Public Health Protection: A Statement of Principles from The Endocrine Society. *Endocrinology* **153**, 4097–4110 (2012).

[44] Swedenborg, E., Rüegg, J., Mäkelä, S. & Pongratz, I. Endocrine disruptive chemicals: mechanisms of action and involvement in metabolic disorders. *Journal of Molecular Endocrinology* **43**, 1–10 (2009).

[45] Solecki, R. *et al.* Scientific principles for the identification of endocrine-disrupting chemicals: a consensus statement. *Archives of Toxicology* **91**, 1001–1006 (2017).

[46] The Endocrine Disruption Exchange (TEDX). https://endocrinedisruption.org/.

[47] EDCs Databank. http://edcs.unicartagena.edu.co/.

[48] Montes-Grajales, D. & Olivero-Verbel, J. EDCs DataBank: 3D-Structure database of endocrine disrupting chemicals. *Toxicology* **327**, 87–94 (2015).

[49] Endocrine Disruptor Screening Program (EDSP). https://www.epa.gov/endocrine-disruption.

[50] Diamanti-Kandarakis, E. *et al.* Endocrine-Disrupting Chemicals: An Endocrine Society Scientific Statement. *Endocrine Reviews* **30**, 293–342 (2009).

[51] Gore, A. C. *et al.* EDC-2: The Endocrine Society's Second Scientific Statement on Endocrine-Disrupting Chemicals. *Endocrine Reviews* **36**, E1–E150 (2015).

[52] Caito, S. & Aschner, M. Chapter 11 - Neurotoxicity of metals. In *Handbook of Clinical Neurology*, vol. 131, 169–189 (Elsevier, 2015).

[53] Bjørklund, G., Mutter, J. & Aaseth, J. Metal chelators and neurotoxicity: lead, mercury, and arsenic. *Archives of Toxicology* **91**, 3787–3797 (2017).

[54] Koch, C. Complexity and the Nervous System. *Science* **284**, 96–98 (1999).

[55] Tshala-Katumbay, D., Mwanza, J.-C., Rohlman, D. S., Maestre, G. & Oriá, R. B. A global perspective on the influence of environmental exposures on the nervous system. *Nature* **527**, S187–S192 (2015).

[56] Claudio, L. An analysis of the U.S. Environmental Protection Agency neurotoxicity testing guidelines. *Regulatory Toxicology and Pharmacology* **16**, 202–212 (1992).

[57] Grandjean, P. & Landrigan, P. J. Neurobehavioural effects of developmental toxicity. *The Lancet Neurology* **13**, 330–338 (2014).

[58] Grandjean, P. & Landrigan, P. Developmental neurotoxicity of industrial chemicals. *The Lancet* **368**, 2167–2178 (2006).

[59] Vargas, R. & Ponce-Canchihuamán, J. Emerging various environmental threats to brain and overview of surveillance system with zebrafish model. *Toxicology Reports* **4**, 467–473 (2017).

[60] Office of Toxic Substances, U. E. Chemicals Which Have Been Tested for Neurotoxic Effects. Tech. Rep. EPA-560/1-76-005, U.S. Environmental Protection Agency, Washington, D.C. (1976).

[61] Mundy, W. R. *et al.* Expanding the test set: Chemicals with potential to disrupt mammalian brain development. *Neurotoxicology and Teratology* **52**, 25–35 (2015).

[62] Aschner, M. *et al.* Reference compounds for alternative test methods to indicate developmental neurotoxicity (DNT) potential of chemicals: example lists and criteria for their selection and use. *ALTEX* **34**, 49 (2017).

[63] Li, Z.-M., Albrecht, M., Fromme, H., Schramm, K.-W. & De Angelis, M. Persistent Organic Pollutants in Human Breast Milk and Associations with Maternal Thyroid Hormone Homeostasis. *Environmental Science & Technology* **54**, 1111–1119 (2020).

[64] Leibson, T., Lala, P. & Ito, S. Chapter 24 - Drug and Chemical Contaminants in Breast Milk: Effects on Neurodevelopment of the Nursing Infant. In Slikker, W., Paule, M. G. & Wang, C. (eds.) *Handbook of Developmental Neurotoxicology*, 275–284 (Academic Press, 2018).

[65] Council, N. R. (ed.) *Scientific frontiers in developmental toxicology and risk assessment* (National Academy Press, Washington, DC, 2000).

[66] Sonawane, B. R. Chemical contaminants in human milk: an overview. *Environmental Health Perspectives* **103**, 197–205 (1995).

[67] Mead, M. N. Contaminants in Human Milk: Weighing the Risks against the Benefits of Breastfeeding. *Environmental Health Perspectives* **116**, A426–A434 (2008).

[68] Agatonovic-Kustrin, S., Ling, L., Tham, S. & Alany, R. Molecular descriptors that influence the amount of drugs transfer into human breast milk. *Journal of Pharmaceutical and Biomedical Analysis* **29**, 103–119 (2002).

[69] Zhao, C. *et al.* Prediction of Milk/Plasma Drug Concentration (M/P) Ratio Using Support Vector Machine (SVM) Method. *Pharmaceutical Research* **23**, 41–48 (2006).

[70] Heinzow, B. Endocrine disruptors in human milk and the health-related issues of breastfeeding. In *Endocrine-Disrupting Chemicals in Food*, 322–355 (Woodhead Publishing, 2009).

[71] Anadón, A., Martínez-Larrañaga, M. R., Ramos, E. & Castellano, V. Transfer of drugs and xenobiotics through milk. In *Reproductive and Developmental Toxicology*, 57–71 (Academic Press, 2011).

[72] Vasios, G. *et al.* Simple physicochemical properties related with lipophilicity, polarity, molecular size and ionization status exert significant impact on the transfer of drugs and chemicals into human breast milk. *Expert Opinion on Drug Metabolism & Toxicology* **12**, 1273–1278 (2016).

[73] Rastogi, S. C. *et al.* Contents of fragrance allergens in children's cosmetics and cosmetic-toys. *Contact Dermatitis* **41**, 84–88 (1999).

[74] Bickers, D. R. *et al.* The safety assessment of fragrance materials. *Regulatory Toxicology and Pharmacology* **37**, 218–273 (2003).

[75] Klaschka, U. & Kolossa-Gehring, M. Fragrances in the Environment: Pleasant odours for nature? (9 pp). *Environmental Science and Pollution Research* **14**, 44–52 (2007).

[76] Nardelli, A., Drieghe, J., Claes, L., Boey, L. & Goossens, A. Fragrance allergens in 'specific' cosmetic products. *Contact Dermatitis* **64**, 212–219 (2011).

[77] Kim, J.-H. *et al.* Risk assessment to human health: Consumer exposure to ingredients in air fresheners. *Regulatory Toxicology and Pharmacology* **98**, 31–40 (2018).

[78] Pastor-Nieto, M.-A. & Gatica-Ortega, M.-E. Ubiquity, Hazardous Effects, and Risk Assessment of Fragrances in Consumer Products. *Current Treatment Options in Allergy* **8**, 21–41 (2021).

[79] Fisher, B. E. Scents and sensitivity. *Environmental Health Perspectives* **106**, A594–A599 (1998).

[80] World Health Organization. *Principles for evaluating health risks in children associated with exposure to chemicals* (World Health Organization, 2006).

[81] Becker, M., Edwards, S. & Massey, R. I. Toxic Chemicals in Toys and Children's Products: Limitations of Current Responses and Recommendations for Government and Industry. *Environmental Science & Technology* **44**, 7986–7991 (2010).

[82] Dennis, K. K. *et al.* Biomonitoring in the Era of the Exposome. *Environmental Health Perspectives* **125**, 502–510 (2017).

[83] Kalia, V., Barouki, R. & Miller, G. W. The Exposome: Pursuing the Totality of Exposure. In Jiang, G. & Li, X. (eds.) *A New Paradigm for Environmental Chemistry and Toxicology: From Concepts to Insights*, 3–10 (Springer, Singapore, 2020).

[84] Barr, D. B. *et al.* The use of dried blood spots for characterizing children's exposure to organic environmental chemicals. *Environmental Research* **195**, 110796 (2021).

[85] Sexton, K., L.Needham, L. & L.Pirkle, J. Human Biomonitoring of Environmental Chemicals: Measuring chemicals in human tissues is the "gold standard" for assessing people's exposure to pollution. *American Scientist* **92**, 38–45 (2004).

[86] Kim, S. *et al.* PubChem in 2021: new data content and improved web interfaces. *Nucleic Acids Research* **49**, D1388–D1395 (2021).

[87] Niedzwiecki, M. M. & Miller, G. W. The Exposome Paradigm in Human Health: Lessons from the Emory Exposome Summer Course. *Environmental Health Perspectives* **125**, 064502 (2017).

[88] Barabási, A.-L. & Oltvai, Z. N. Network biology: understanding the cell's functional organization. *Nature Reviews Genetics* **5**, 101–113 (2004).

[89] Dix, D. J. *et al.* The ToxCast Program for Prioritizing Toxicity Testing of Environmental Chemicals. *Toxicological Sciences* **95**, 5–12 (2007).

[90] Mattingly, C. J. *et al.* The Comparative Toxicogenomics Database: A Cross-Species Resource for Building Chemical-Gene Interaction Networks. *Toxicological Sciences* **92**, 587–595 (2006).

[91] Council, N. R. *Toxicity Testing in the 21st Century: A Vision and a Strategy* (The National Academies Press, Washington, DC, 2007).

[92] Hartung, T. On mapping the human toxome. *ALTEX* **28**, 83–93 (2011).

[93] Hartung, T. Toxicology for the twenty-first century. *Nature* **460**, 208–212 (2009).

[94] Krewski, D. *et al.* Toxicity Testing in the 21st Century: A Vision and a Strategy. *Journal of Toxicology and Environmental Health* **13**, 51–138 (2010).

[95] Kleensang, A. *et al.* Pathways of Toxicity. *ALTEX* **31**, 53–61 (2014).

[96] Edwards, S. W., Tan, Y.-M., Villeneuve, D. L., Meek, M. & McQueen, C. A. Adverse Outcome Pathways—Organizing Toxicological Information to Improve Decision Making. *Journal of Pharmacology and Experimental Therapeutics* **356**, 170 (2016).

[97] Vinken, M. *et al.* Adverse outcome pathways: a concise introduction for toxicologists. *Archives of Toxicology* **91**, 3697–3707 (2017).

[98] Krewski, D. *et al.* Toxicity testing in the 21st century: progress in the past decade and future perspectives. *Archives of Toxicology* **94**, 1–58 (2020).

[99] Ankley, G. T. *et al.* Adverse outcome pathways: A conceptual framework to support ecotoxicology research and risk assessment. *Environmental Toxicology and Chemistry* **29**, 730–741 (2010).

[100] Tollefsen, K. E. *et al.* Applying Adverse Outcome Pathways (AOPs) to support Integrated Approaches to Testing and Assessment (IATA). *Regulatory Toxicology and Pharmacology* **70**, 629–640 (2014).

[101] The Organisation for Economic Co-operation and Development (OECD). Users' Handbook Supplement to the Guidance Document for Developing and Assessing Adverse Outcome Pathways. Tech. Rep. 233, OECD Environment, Health and Safety Publications, Paris (2018).

[102] The Organisation for Economic Co-operation and Development (OECD). Revised Guidance Document on Developing And Assessing Adverse Outcome Pathways. Tech. Rep. 184, OECD Environment, Health and Safety Publications, Paris (2013).

[103] The Organisation for Economic Co-operation and Development (OECD). Guidance Document for the Use of Adverse Outcome Pathways in Developing Integrated Approaches to Testing and Assessment (IATA). Tech. Rep. 260, OECD Environment, Health and Safety Publications, Paris (2017).

[104] Vinken, M. The adverse outcome pathway concept: A pragmatic tool in toxicology. *Toxicology* **312**, 158–165 (2013).

[105] Villeneuve, D. L. *et al.* Adverse Outcome Pathway (AOP) Development I: Strategies and Principles. *Toxicological Sciences* **142**, 312–320 (2014).

[106] Villeneuve, D. L. *et al.* Adverse Outcome Pathway Development II: Best Practices. *Toxicological Sciences* **142**, 321–330 (2014).

[107] Knapen, D. *et al.* Adverse outcome pathway networks I: Development and applications: Advancing adverse outcome pathway networks. *Environmental Toxicology and Chemistry* **37**, 1723–1733 (2018).

[108] Sewell, F. *et al.* The future trajectory of adverse outcome pathways: a commentary. *Archives of Toxicology* **92**, 1657–1661 (2018).

[109] Sakuratani, Y., Horie, M. & Leinala, E. Integrated Approaches to Testing and Assessment: OECD Activities on the Development and Use of Adverse Outcome Pathways and Case Studies. *Basic & Clinical Pharmacology & Toxicology* **123**, 20–28 (2018).

[110] Villeneuve, D. L. *et al.* Adverse outcome pathway networks II: Network analytics. *Environmental Toxicology and Chemistry* **37**, 1734–1748 (2018).

[111] Aguayo-Orozco, A. *et al.* sAOP: linking chemical stressors to adverse outcomes pathway networks. *Bioinformatics* **35**, 5391–5392 (2019).

[112] Jornod, F. *et al.* AOP4EUpest: mapping of pesticides in adverse outcome pathways using a text mining tool. *Bioinformatics* **36**, 4379–4381 (2020).

[113] The Organisation for Economic Co-operation and Development (OECD). AOP knowledge base (AOP-KB). https://aopkb.oecd.org/.

[114] AOP-Wiki. https://aopwiki.org.

[115] Knapen, D., Vergauwen, L., Villeneuve, D. L. & Ankley, G. T. The potential of AOP networks for reproductive and developmental toxicity assay development. *43rd Annual Conference of the European Teratology Society* **56**, 52–55 (2015).

[116] Howdeshell, K. L., Hotchkiss, A. K. & Gray, L. E. Cumulative effects of antiandrogenic chemical mixtures and their relevance to human health risk assessment. *International Journal of Hygiene and Environmental Health* **220**, 179–188 (2017).

[117] Coady, K. *et al.* When Are Adverse Outcome Pathways and Associated Assays "Fit for Purpose" for Regulatory Decision-Making and Management of Chemicals? *Integrated Environmental Assessment and Management* **15**, 633–647 (2019).

[118] Hecker, M. & LaLone, C. A. Adverse Outcome Pathways: Moving from a Scientific Concept to an Internationally Accepted Framework. *Environmental Toxicology and Chemistry* **38**, 1152–1163 (2019).

[119] Kitsak, M. *et al.* Tissue Specificity of Human Disease Module. *Scientific Reports* **6**, 35241 (2016).

[120] Kim, P. *et al.* TissGDB: tissue-specific gene database in cancer. *Nucleic Acids Research* **46**, D1031–D1038 (2018).

[121] Maiorino, E. *et al.* Discovering the genes mediating the interactions between chronic respiratory diseases in the human interactome. *Nature Communications* **11**, 811 (2020).

[122] Taboureau, O., El M'Selmi, W. & Audouze, K. Integrative systems toxicology to predict human biological systems affected by exposure to environmental chemicals. *Toxicology and Applied Pharmacology* **405**, 115210 (2020).

[123] Borrel, A., Auerbach, S. S., Houck, K. A. & Kleinstreuer, N. C. Tox21BodyMap: a webtool to map chemical effects on the human body. *Nucleic Acids Research* **48**, W472–W476 (2020).

[124] Raunio, H. In Silico Toxicology – Non-Testing Methods. *Frontiers in Pharmacology* **2**, 33 (2011).

[125] Floris, M. *et al.* A generalizable definition of chemical similarity for read-across. *Journal of Cheminformatics* **6**, 39 (2014).

[126] Bajusz, D., Rácz, A. & Héberger, K. Why is Tanimoto index an appropriate choice for fingerprint-based similarity calculations? *Journal of Cheminformatics* **7**, 20 (2015).

[127] Ford, K. A. Refinement, Reduction, and Replacement of Animal Toxicity Tests by Computational Methods. *ILAR Journal* **57**, 226–233 (2016).

[128] Saldívar-González, F. I., Pilón-Jiménez, B. A. & Medina-Franco, J. L. Chemical space of naturally occurring compounds. *Physical Sciences Reviews* **4**, 20180103 (2019).

[129] Rogers, D. & Hahn, M. Extended-Connectivity Fingerprints. *Journal of Chemical Information and Modeling* **50**, 742–754 (2010).

[130] Durant, J. L., Leland, B. A., Henry, D. R. & Nourse, J. G. Reoptimization of MDL Keys for Use in Drug Discovery. *Journal of Chemical Information and Computer Sciences* **42**, 1273–1280 (2002).

[131] Lo, Y.-C. & Torres, J. Z. Chemical Similarity Networks for Drug Discovery. In Chen, T. & Chai, S. C. (eds.) *Special Topics in Drug Discovery* (InTechOpen, 2016).

[132] Egeghy, P. P., Vallero, D. A. & Cohen Hubal, E. A. Exposure-based prioritization of chemicals for risk assessment. *Environmental Science & Policy* **14**, 950–964 (2011).

[133] Service, R. F. A New Wave of Chemical Regulations Just Ahead? *Science* **325**, 692–693 (2009).

[134] European Union. Commission Regulation (EU) No 10/2011 on plastic materials and articles intended to come into contact with food. https://eur-lex.europa.eu/eli/reg/2011/10/oj (2011).

[135] European Union. EU lists of food additives. https://webgate.ec.europa.eu/foods_system/main/?sector=FAD&auth=SANCAS.

[136] European Union. EU food flavorings database. https://webgate.ec.europa.eu/foods_system/main/?sector=FFL&auth=SANCAS.

[137] U.S. FDA. FDA TOR Notices. https://www.cfsanappsexternal.fda.gov/scripts/fdcc/?set=TOR.

[138] U.S. FDA. US FDA Indirect Additives used in Food Contact Substances. https://www.cfsanappsexternal.fda.gov/scripts/fdcc/?set=IndirectAdditives.

[139] World Health Organization. WHO Codex General Standards for Food Additives. http://www.fao.org/gsfaonline/additives/index.html (2019).

[140] World Health Organization. The Joint FAO/WHO Expert Committee on Food Additives (JECFA) list. http://apps.who.int/food-additives-contaminants-jecfa-database/search.aspx.

[141] European Union. EU List of Substances Prohibited in Cosmetic Products. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex:02009R1223-20150416.

[142] European Chemicals Agency. Cosmetic ingredient database (cosing) - List of colorants allowed in cosmetic products. https://www.echa.europa.eu/regulations/biocidal-products-regulation/approval-of-active-substances/list-of-approved-active-substances.

[143] European Chemicals Agency. Cosmetic ingredient database (cosing) - List of preservatives allowed in cos-

metic products. https://data.europa.eu/euodp/en/data/dataset/cosmetic-ingredient-database-list-of-preservatives-allowed-in-cosmetic-products.

[144] European Chemicals Agency. Cosmetic ingredient database (cosing) - List of UV filters allowed in cosmetic products. https://data.europa.eu/euodp/en/data/dataset/cosmetic-ingredient-database-list-of-uv-filters-allowed-in-cosmetic-products.

[145] European Union. Directive 2009/48/EC of the European Parliament and of the Council of 18 June 2009 on the safety of toys. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:02009L0048-20181126 (2009).

[146] Danish EPA. Danish EPA Sensitizing Fragrances in Children's Articles. https://www2.mst.dk/udgiv/publications/2006/87-7052-018-6/pdf/87-7052-019-4.pdf (2006).

[147] Washington State Children's Chemicals of High Concern. https://ecology.wa.gov/Regulations-Permits/Reporting-requirements/Reporting-for-Childrens-Safe-Products-Act/Chemicals-of-high-concern-to-children.

[148] Aurisano, N., Huang, L., Canals, L. M. i., Jolliet, O. & Fantke, P. Chemicals of concern in plastic toys. *Environment International* **146**, 106194 (2021).

[149] US Occupational Safety and Health Standards (OSHA) List. https://www.osha.gov/laws-regs/regulations/standardnumber/1910/1910.119AppA.

[150] The Organisation for Economic Co-operation and Development (OECD). OECD High Production Volume (OECD HPV). https://www.oecd.org/chemicalsafety/risk-assessment/33883530.pdf (2004).

[151] U.S. EPA. The United States High Production Volume (USHPV) database. https://comptox.epa.gov/dashboard/chemical_lists/EPAHPV (2004).

[152] European Chemicals Agency. REACH High Production Volume (HPV) chemicals. https://echa.europa.eu/en/information-on-chemicals/registered-substances.

[153] Stone, A. & Delistraty, D. Sources of toxicity and exposure information for identifying chemicals of high concern to children. *Environmental Impact Assessment Review* **30**, 380–387 (2010).

[154] Neltner, T. G., Alger, H. M., Leonard, J. E. & Maffini, M. V. Data gaps in toxicity testing of chemicals allowed in food in the United States. *Reproductive Toxicology* **42**, 85–94 (2013).

[155] Geueke, B., Wagner, C. C. & Muncke, J. Food contact substances and chemicals of concern: a comparison of inventories. *Food Additives & Contaminants: Part A* **31**, 1438–1450 (2014).

[156] Demeneix, B. & Salma, R. Endocrine disruptors: from scientific evidence to human health protection policy. Tech. Rep., Policy Department for Citizen's Rights and Constitutional Affairs, European Parliament (2019).

[157] European Union. Candidate List of Substances of Very High Concern (SVHC) for Authorisation. https://echa.europa.eu/candidate-list-table.

[158] PubMed database. https://www.ncbi.nlm.nih.gov/pubmed/.

[159] Baker, V. Endocrine disrupters — testing strategies to assess human hazard. *Toxicology in Vitro* **15**, 413–419 (2001).

[160] Bliatka, D., Lymperi, S., Mastorakos, G. & Goulis, D. G. Effect of endocrine disruptors on male reproduction in humans: why the evidence is still lacking? *Andrology* **5**, 404–407 (2017).

[161] Hernández, A. F. & Tsatsakis, A. M. Human exposure to chemical mixtures: Challenges for the integration of toxicology with epidemiology data in risk assessment. *Food and Chemical Toxicology* **103**, 188–193 (2017).

[162] Ding, D. *et al.* The EDKB: an established knowledge base for endocrine disrupting chemicals. *BMC Bioinformatics* **11**, S5 (2010).

[163] Endocrine Society. Endocrine Disrupting Chemicals. https://www.endocrine.org/topics/edc.

[164] Chemical Abstracts Service (CAS) database. https://www.cas.org/.

[165] Foulds, C. E., Treviño, L. S., York, B. & Walker, C. L. Endocrine-disrupting chemicals and fatty liver disease. *Nature Reviews Endocrinology* **13**, 445–457 (2017).

[166] Monneret, C. What is an endocrine disruptor? *Comptes Rendus Biologies* **340**, 403–405 (2017).

[167] Sharma, V. & McNeill, J. H. To scale or not to scale: the principles of dose extrapolation. *British Journal of Pharmacology* **157**, 907–921 (2009).

[168] Vandenberg, L. N. *et al.* Hormones and Endocrine-Disrupting Chemicals: Low-Dose Effects and Nonmonotonic Dose Responses. *Endocrine Reviews* **33**, 378–455 (2012).

[169] Vandenberg, L. N. Low-Dose Effects of Hormones and Endocrine Disruptors. *Vitamins & Hormones* **94**, 129–165 (2014).

[170] Welshons, W. V. *et al.* Large effects from small exposures. I. Mechanisms for endocrine-disrupting chemicals with estrogenic activity. *Environmental Health Perspectives* **111**, 994–1006 (2003).

[171] U.S. EPA. US EPA Safer Chemical Ingredients List. https://www.epa.gov/saferchoice/safer-ingredients.

[172] U.S. FDA. US FDA Inactive Ingredients List. https://www.accessdata.fda.gov/scripts/cder/iig/index.cfm.

[173] ClassyFire. http://classyfire.wishartlab.com/.

[174] Djoumbou Feunang, Y. *et al.* ClassyFire: automated chemical classification with a comprehensive, computable taxonomy. *Journal of Cheminformatics* **8**, 61 (2016).

[175] Balloon. http://users.abo.fi/mivainio/balloon/.

[176] Vainio, M. J. & Johnson, M. S. Generating Conformer Ensembles Using a Multi-objective Genetic Algorithm. *Journal of Chemical Information and Modeling* **47**, 2462–2474 (2007).

[177] Open Babel. http://openbabel.org/.

[178] O'Boyle, N. M. *et al.* Open Babel: An open chemical toolbox. *Journal of Cheminformatics* **3**, 33 (2011).

[179] RDKit: Open-Source Cheminformatics Software. https://www.rdkit.org/.

[180] PaDEL-Descriptor. http://www.yapcwsoft.com/dd/padeldescriptor/.

[181] Yap, C. W. PaDEL-descriptor: An open source software to calculate molecular descriptors and fingerprints. *Journal of Computational Chemistry* **32**, 1466–1474 (2011).

[182] O'Boyle, N. M., Morley, C. & Hutchison, G. R. Pybel: a Python wrapper for the OpenBabel cheminformatics toolkit. *Chemistry Central Journal* **2**, 5 (2008).

[183] Yang, H. *et al.* admetSAR 2.0: web-service for prediction and optimization of chemical ADMET properties. *Bioinformatics* **35**, 1067–1069 (2019).

[184] Pires, D. E. V., Blundell, T. L. & Ascher, D. B. pkCSM: Predicting Small-Molecule Pharmacokinetic and Toxicity Properties Using Graph-Based Signatures. *Journal of Medicinal Chemistry* **58**, 4066–4072 (2015).

[185] Banerjee, P., Eckert, A. O., Schrey, A. K. & Preissner, R. ProTox-II: a webserver for the prediction of toxicity of chemicals. *Nucleic Acids Research* **46**, W257–W263 (2018).

[186] Daina, A., Michielin, O. & Zoete, V. SwissADME: a free web tool to evaluate pharmacokinetics, drug-likeness and medicinal chemistry friendliness of small molecules. *Scientific Reports* **7**, 42717 (2017).

[187] Patlewicz, G., Jeliazkova, N., Safford, R., Worth, A. & Aleksiev, B. An evaluation of the implementation of the Cramer classification scheme in the Toxtree software. *SAR and QSAR in Environmental Research* **19**, 495–524 (2008).

[188] Schyman, P., Liu, R., Desai, V. & Wallqvist, A. vNN Web Server for ADMET Predictions. *Frontiers in Pharmacology* **8**, 889 (2017).

[189] PHP. http://php.net/.

[190] jQuery. https://jquery.com/.

[191] Google Charts. https://developers.google.com/chart/.

[192] D3: Data-Driven Documents. https://d3js.org/.

[193] Cytoscape.js. http://js.cytoscape.org/.

[194] JSmol. http://jmol.sourceforge.net/.

[195] MariaDB Server: The open source relational database. https://mariadb.org/.

[196] Apache: HTTP Server Pro. https://httpd.apache.org/.

[197] González-Medina, M. *et al.* Scaffold Diversity of Fungal Metabolites. *Frontiers in Pharmacology* **8**, 180 (2017).

[198] Wassenaar, P. N., Rorije, E., Janssen, N. M., Peijnenburg, W. J. & Vijver, M. G. Chemical similarity to identify potential Substances of Very High Concern – An effective screening method. *Computational Toxicology* **12**, 100110 (2019).

[199] Wassenaar, P. N., Rorije, E., Vijver, M. G. & Peijnenburg, W. J. Evaluating chemical similarity as a measure to identify potential substances of very high concern. *Regulatory Toxicology and Pharmacology* **119**, 104834 (2021).

[200] Tanimoto, T. *An Elementary Mathematical Theory of Classification and Prediction* (International Business Machines Corporation, 1958).

[201] Sorensen, T. A. A method of establishing groups of equal amplitude in plant sociology based on similarity of species content and its application to analyses of the vegetation on Danish commons. *Biol. Skar.* **5**, 1–34 (1948).

[202] Bender, A. *et al.* How Similar Are Similarity Searching Methods? A Principal Component Analysis of Molecular Descriptor Space. *Journal of Chemical Information and Modeling* **49**, 108–119 (2009).

[203] Tabb, M. M. & Blumberg, B. New Modes of Action for Endocrine-Disrupting Chemicals. *Molecular Endocrinology* **20**, 475–482 (2006).

[204] Blondel, V. D., Guillaume, J.-L., Lambiotte, R. & Lefebvre, E. Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment* **2008**, P10008 (2008).

[205] Bastian, M., Heymann, S. & Jacomy, M. Gephi: an open source software for exploring and manipulating networks. In *Third international AAAI conference on weblogs and social media*, 1–2 (2009).

[206] Toxicology, E. N. C. F. C. ToxCast and Tox21 Summary Files (2018).

[207] Jaccard, P. The distribution of the flora in the Alphine zone.1. *New Phytologist* **11**, 37–50 (1912).

[208] World Health Organization. IARC Monographs on the Identification of Carcinogenic Hazards to Humans. https://monographs.iarc.who.int/agents-classified-by-the-iarc/.

[209] Loomis, D., Guha, N., Hall, A. L. & Straif, K. Identifying occupational carcinogens: an update from the IARC Monographs. *Occupational and Environmental Medicine* **75**, 593–603 (2018).

[210] U.S. National Toxicology Program. 14th Report on Carcinogens. https://ntp.niehs.nih.gov/whatwestudy/assessments/cancer/roc/index.html (2016).

[211] Venkatasubramanian, K. V. Database of endocrine disruptors focuses on experimental evidence. *Chemical & Engineering News* (2019).

[212] The French Agency for Food, E. & (ANSES), O. H. . S. Elaboration of a list of substances of interest as regards to a potential endocrine activity and prioritisation strategy for assessment. Tech. Rep. 2019-SA-0179, ANSES, France (2021).

[213] Darbre, P. D. The history of endocrine-disrupting chemicals. *Current Opinion in Endocrine and Metabolic Research* **7**, 26–33 (2019).

[214] Agerstrand, M. *et al.* An academic researcher's guide to increased impact on regulatory assessment of chemicals. *Environmental Science: Processes & Impacts* **19**, 644–655 (2017).

[215] Toxicology, E. N. C. F. C. ToxCast and Tox21 Summary Files (2019).

[216] Council, N. R. *Risk Assessment in the Federal Government: Managing the Process* (National Academies Press, Washington, D.C., 1983).

[217] Beausoleil, C. *et al.* Review of non-monotonic dose-responses of substances for human risk assessment. *EFSA Supporting Publications* **13**, 1027E (2016).

[218] Darbre, P. D. Chapter 16 - An Introduction to the Challenges for Risk Assessment of Endocrine Disrupting Chemicals. In Darbre, P. D. (ed.) *Endocrine Disruption and Human Health*, 289–300 (Academic Press, Boston, 2015).

[219] Clahsen, S. C. S. *et al.* Why Do Countries Regulate Environmental Health Risks Differently? A Theoretical Perspective: Why Do Countries Regulate Environmental Health Risks Differently? *Risk Analysis* **39**, 439–461 (2019).

[220] Mihaich, E. M. *et al.* Challenges in assigning endocrine-specific modes of action: Recommendations for researchers and regulators: Assigning Endocrine-Specific Modes of Action. *Integrated Environmental Assessment and Management* **13**, 280–292 (2017).

[221] Jeong, J. & Choi, J. Use of adverse outcome pathways in chemical toxicity testing: potential advantages and limitations. *Environmental Health and Toxicology* **33**, e2018002 (2017).

[222] Ankley, G. T. & Edwards, S. W. The adverse outcome pathway: A multifaceted framework supporting 21st century toxicology. *Current Opinion in Toxicology* **9**, 1–7 (2018).

[223] Vinken, M. Taking adverse outcome pathways to the next level. *Toxicology in Vitro* **50**, A1–A2 (2018).

[224] Hartung, T. *et al.* Systems toxicology. *ALTEX* **29**, 119–128 (2012).

[225] Sturla, S. J. *et al.* Systems Toxicology: From Basic Research to Risk Assessment. *Chemical Research in Toxicology* **27**, 314–329 (2014).

[226] Hartung, T. *et al.* Systems Toxicology: Real World Applications and Opportunities. *Chemical Research in Toxicology* **30**, 870–882 (2017).

[227] Aguayo-Orozco, A., Taboureau, O. & Brunak, S. The use of systems biology in chemical risk assessment. *Current Opinion in Toxicology* **15**, 48–54 (2019).

[228] Pollesch, N. L., Villeneuve, D. L. & O'Brien, J. M. Extracting and Benchmarking Emerging Adverse Outcome Pathway Knowledge. *Toxicological Sciences* **168**, 349–364 (2019).

[229] LaLone, C. A. *et al.* Weight of evidence evaluation of a network of adverse outcome pathways linking activation of the nicotinic acetylcholine receptor in honey bees to colony death. *Science of the Total Environment* **584-585**, 751–775 (2017).

[230] Spinu, N. *et al.* Development and analysis of an adverse outcome pathway network for human neurotoxicity. *Archives of Toxicology* **93**, 2759–2772 (2019).

[231] Arnesdotter, E. *et al.* Derivation, characterisation and analysis of an adverse outcome pathway network for human hepatotoxicity. *Toxicology* **459**, 152856 (2021).

[232] Villeneuve, D. L. *et al.* Representing the Process of Inflammation as Key Events in Adverse Outcome Pathways. *Toxicological Sciences* **163**, 346–352 (2018).

[233] Carvaillo, J.-C., Barouki, R., Coumoul, X. & Audouze, K. Linking Bisphenol S to Adverse Outcome Pathways Using a Combined Text Mining and Systems Biology Approach. *Environmental Health Perspectives* **127**, 047005 (2019).

[234] Browne, P., Van Der Wal, L. & Gourmelon, A. OECD approaches and considerations for regulatory evaluation of endocrine disruptors. *Molecular and Cellular Endocrinology* **504**, 110675 (2020).

[235] Rugard, M., Coumoul, X., Carvaillo, J.-C., Barouki, R. & Audouze, K. Deciphering Adverse Outcome Pathway Network Linked to Bisphenol F Using Text Mining and Systems Toxicology Approaches. *Toxicological Sciences* **173**, 32–40 (2020).

[236] Distributed Structure-Searchable Toxicity (DSSTox) Database. https://www.epa.gov/chemical-research/distributed-structure-searchable-toxicity-dsstox-database.

[237] MeSH Browser. https://meshb.nlm.nih.gov/.

[238] Patisaul, H. B., Fenton, S. E. & Aylor, D. Animal models of endocrine disruption. *Best Practice & Research Clinical Endocrinology & Metabolism* **32**, 283–297 (2018).

[239] NetworkX. https://networkx.org/.

[240] NetworkAnalyzer. https://apps.cytoscape.org/apps/networkanalyzer.

[241] Assenov, Y., Ramírez, F., Schelhorn, S.-E., Lengauer, T. & Albrecht, M. Computing topological parameters of biological networks. *Bioinformatics* **24**, 282–284 (2008).

[242] Leydesdorff, L. Betweenness centrality as an indicator of the interdisciplinarity of scientific journals. *Journal of the American Society for Information Science and Technology* **58**, 1303–1319 (2007).

[243] Takes, F. W. & Kosters, W. A. Determining the diameter of small world networks. In *CIKM '11: Proceedings of the 20th ACM international conference on Information and knowledge management*, 1191–1196 (ACM Press, Glasgow, Scotland, UK, 2011).

[244] Bergman, A. *et al.* The Impact of Endocrine Disruption: A Consensus Statement on the State of the Science. *Environmental Health Perspectives* **121**, A104–A106 (2013).

[245] Bernal, J. Thyroid hormones in brain development and function. *Endotext [Internet]* (2000).

[246] Volpato, S. *et al.* Serum thyroxine level and cognitive decline in euthyroid older women. *Neurology* **58**, 1055–1061 (2002).

[247] Tunc-Ozcan, E., Ullmann, T. M., Shukla, P. K. & Redei, E. E. Low-Dose Thyroxine Attenuates Autism-Associated Adverse Effects of Fetal Alcohol in Male Offspring's Social Behavior and Hippocampal Gene Expression. *Alcoholism: Clinical and Experimental Research* **37**, 1986–1995 (2013).

[248] Cooke, G. E., Mullally, S., Correia, N., O'Mara, S. M. & Gibney, J. Hippocampal Volume Is Decreased in Adults with Hypothyroidism. *Thyroid* **24**, 433–440 (2014).

[249] Corton, J. C. & Lapinskas, P. J. Peroxisome Proliferator-Activated Receptors: Mediators of Phthalate Ester-Induced Effects in the Male Reproductive Tract? *Toxicological Sciences* **83**, 4–17 (2005).

[250] Latini, G., Scoditti, E., Verrotti, A., De Felice, C. & Massaro, M. Peroxisome Proliferator-Activated Receptors as Mediators of Phthalate-Induced Effects in the Male and Female Reproductive Tract: Epidemiological and Experimental Evidence. *PPAR Research* **2008**, 359267 (2008).

[251] Batarseh, A. & Papadopoulos, V. Regulation of translocator protein 18kDa (TSPO) expression in health and disease states. *Molecular and Cellular Endocrinology* **327**, 1–12 (2010).

[252] Saran, S. *et al.* Effect of hypothyroidism on female reproductive hormones. *Indian Journal of Endocrinology and Metabolism* **20**, 108 (2016).

[253] Jahnke, G. D., Choksi, N. Y., Moore, J. A. & Shelby, M. D. Thyroid toxicants: assessing reproductive health effects. *Environmental Health Perspectives* **112**, 363–368 (2004).

[254] Vissenberg, R. *et al.* Pathophysiological aspects of thyroid hormone disorders/thyroid peroxidase autoantibodies and reproduction. *Human Reproduction Update* **21**, 378–387 (2015).

[255] Chen, C.-W., Huang, Y.-L., Tzeng, C.-R., Huang, R.-L. & Chen, C.-H. Idiopathic Low Ovarian Reserve Is Associated with More Frequent Positive Thyroid Peroxidase Antibodies. *Thyroid* **27**, 1194–1200 (2017).

[256] Wang, X., Ding, X., Xiao, X., Xiong, F. & Fang, R. An exploration on the influence of positive simple thyroid peroxidase antibody on female infertility. *Experimental and Therapeutic Medicine* **16**, 3077–3081 (2018).

[257] Erickso, G. F., Hsueh, A., Quigley, M., Rebar, R. & Yen, S. Functional Studies of Aromatase Activity in Human Granulosa Cells from Normal and Polycystic Ovaries. *The Journal of Clinical Endocrinology & Metabolism* **49**, 514–519 (1979).

[258] Garzo, V. & Dorrington, J. Aromatase activity in human granulosa cells during follicular development and the modulation by follicle-stimulating hormone and insulin. *American Journal of Obstetrics and Gynecology* **148**, 657–662 (1984).

[259] Scholz, S. 17-α-ethinylestradiol affects reproduction, sexual differentiation and aromatase gene expression of the medaka (Oryzias latipes). *Aquatic Toxicology* **50**, 363–373 (2000).

[260] Sun, L., Zha, J., Spear, P. A. & Wang, Z. Toxicity of the aromatase inhibitor letrozole to Japanese medaka (Oryzias latipes) eggs, larvae and breeding adults. *Comparative Biochemistry and Physiology Part C: Toxicology & Pharmacology* **145**, 533–541 (2007).

[261] Hazra, R. *et al.* In Vivo Actions of the Sertoli Cell Glucocorticoid Receptor. *Endocrinology* **155**, 1120–1130 (2014).

[262] Silva, E. J., Queiróz, D. B., Honda, L. & Avellar, M. C. W. Glucocorticoid receptor in the rat epididymis: Expression, cellular distribution and regulation by steroid hormones. *Molecular and Cellular Endocrinology* **325**, 64–77 (2010).

[263] Iqubal, A. *et al.* Environmental neurotoxic pollutants: review. *Environmental Science and Pollution Research* **27**, 41175–41198 (2020).

[264] Council, N. R. *Environmental Neurotoxicology* (National Academies Press, Washington, D.C., 1992).

[265] Williams, A. J. *et al.* The CompTox Chemistry Dashboard: a community data resource for environmental chemistry. *Journal of Cheminformatics* **9**, 61 (2017).

[266] Fonger, G. C., Stroup, D., Thomas, P. L. & Wexler, P. Toxnet: A computerized collection of toxicological and environmental health information. *Toxicology and Industrial Health* **16**, 4–6 (2000).

[267] National Library of Medicine, U. TOXNET Update: New Locations for TOXNET Content. Tech. Rep. 431, NLM Tech Bulletin (2019).

[268] Schultheisz, R. J. TOXLINE: Evolution of an online interactive bibliographic database. *Journal of the American Society for Information Science* **32**, 421–429 (1981).

[269] Fonger, G. C., Hakkinen, P., Jordan, S. & Publicker, S. The National Library of Medicine's (NLM) Hazardous Substances Data Bank (HSDB): Background, recent enhancements and future plans. *Toxicology* **325**, 209–216 (2014).

[270] NEURO: Chemicals Demonstrating Effects on Neurodevelopment. https://comptox.epa.gov/dashboard/chemical-lists/DNTEFFECTS.

[271] NEURO: Chemicals Triggering Developmental Neurotoxicity In Vivo. https://comptox.epa.gov/dashboard/chemical_lists/DNTINVIVO.

[272] NEURO: DNT Screening Library. https://comptox.epa.gov/dashboard/chemical_lists/DNTSCREEN.

[273] NEURO: Neurotoxicants from PubMed. https://comptox.epa.gov/dashboard/chemical_lists/LITMINEDNEURO.

[274] NEURO: Neurotoxicants Collection from Public Resources. https://comptox.epa.gov/dashboard/chemical_lists/NEUROTOXINS.

[275] Rogers, F. B. Medical subject headings. *Bulletin of the Medical Library Association* **51**, 114–116 (1963).

[276] Plotly. https://plotly.com/javascript/.

[277] Estrin, W. J. *et al.* Evidence of Neurologic Dysfunction Related to Long-term Ethylene Oxide Exposure. *Archives of Neurology* **44**, 1283–1286 (1987).

[278] Estrin, W. J., Bowler, R. M., Lash, A. & Becker, C. E. Neurotoxicological evaluation of hospital sterilizer workers exposed to ethylene oxide. *Journal of Toxicology: Clinical Toxicology* **28**, 1–20 (1990).

[279] Zheng, W., Aschner, M. & Ghersi-Egea, J.-F. Brain barrier systems: a new frontier in metal neurotoxicological research. *Toxicology and Applied Pharmacology* **192**, 1–11 (2003).

[280] Miodovnik, A., Edwards, A., Bellinger, D. C. & Hauser, R. Developmental neurotoxicity of ortho-phthalate diesters: Review of human and experimental evidence. *NeuroToxicology* **41**, 112–122 (2014).

[281] Tang, J. *et al.* Neurobehavioral changes induced by di(2-ethylhexyl) phthalate and the protective effects of vitamin E in Kunming mice. *Toxicology Research* **4**, 1006–1015 (2015).

[282] Jasial, S., Hu, Y., Vogt, M. & Bajorath, J. Activity-relevant similarity values for fingerprints and implications for similarity searching. *F1000Research* **5**, 591 (2016).

[283] Mohanraj, K. *et al.* IMPPAT: A curated database of Indian Medicinal Plants, Phytochemistry And Therapeutics. *Scientific Reports* **8**, 4329 (2018).

[284] Vivek-Ananth, R. P., Sahoo, A. K., Kumaravel, K., Mohanraj, K. & Samal, A. MeFSAT: a curated natural product database specific to secondary metabolites of medicinal fungi. *RSC Advances* **11**, 2596–2607 (2021).

[285] Landrigan, P. J., Sonawane, B., Mattison, D., McCally, M. & Garg, A. Chemical contaminants in breast milk and their impacts on children's health: an overview. *Environmental Health Perspectives* **110**, A313–315 (2002).

[286] Lehmann, G. M. *et al.* Environmental Chemicals in Breast Milk and Formula: Exposure and Risk Assessment Implications. *Environmental Health Perspectives* **126**, 096001 (2018).

[287] LaKind, J. S. *et al.* Infant Dietary Exposures to Environmental Chemicals and Infant/Child Health: A Critical Assessment of the Literature. *Environmental Health Perspectives* **126**, 096002 (2018).

[288] Statista. https://www.statista.com/.

[289] Galli, A. *et al.* Assessing the global environmental consequences of economic growth through the Ecological Footprint: A focus on China and India. *Ecological Indicators* **17**, 99–107 (2012).

[290] Ramakrishnan, N., Kaphalia, B., Seth, T. & Roy, N. Organochlorine Pesticide Residues in Mother's Milk: a Source of Toxic Chemicals in Suckling Infants. *Human Toxicology* **4**, 7–12 (1985).

[291] Devanathan, G. *et al.* Persistent organochlorines in human breast milk from major metropolitan cities in India. *Environmental Pollution* **157**, 148–154 (2009).

[292] Devanathan, G. *et al.* Brominated flame retardants and polychlorinated biphenyls in human breast milk from several locations in India: Potential contaminant sources in a municipal dumping site. *Environment International* **39**, 87–95 (2012).

[293] Sharma, B. M., Bharat, G. K., Tayal, S., Nizzetto, L. & Larssen, T. The legal framework to manage chemical pollution in India and the lesson from the Persistent Organic Pollutants (POPs). *Science of the Total Environment* **490**, 733–747 (2014).

[294] van den Berg, M. *et al.* WHO/UNEP global surveys of PCDDs, PCDFs, PCBs and DDTs in human milk and benefit–risk evaluation of breastfeeding. *Archives of Toxicology* **91**, 83–96 (2017).

[295] Gaulton, A. *et al.* The ChEMBL database in 2017. *Nucleic Acids Research* **45**, D945–D954 (2017).

[296] Samet, J. M. *et al.* The IARC Monographs: Updated Procedures for Modern and Transparent Evidence Synthesis in Cancer Hazard Identification. *JNCI: Journal of the National Cancer Institute* **112**, 30–37 (2020).

[297] Indian Ministry of Chemicals & Fertilizers. Production of major chemicals year-wise in India. https://data.gov.in/catalogsv2?format=json&offset=0&limit=9&filters%5Bfield_ministry_department%3Aname%5D=Department+of+Chemicals+and+Petrochemicals&sort%5Bogpl_module_domain_name%5D=asc&sort%5Bcreated%5D=desc.

[298] Indian Ministry of Agriculture & Farmers Welfare. List of Banned Pesticides in India. http://ppqs.gov.in/divisions/cib-rc/registered-products.

[299] Indian Ministry of Environment & Forests. Schedule 1 hazardous chemical list in India. http://moef.gov.in/wp-content/uploads/2019/08/SCHEDULE-I.html.

[300] Indian Ministry of Environment & Forests. Schedule 3 hazardous chemical list in India. http://moef.gov.in/wp-content/uploads/2019/08/SCHEDULE-3.html.

[301] Computer software, Canada: The Metabolomics Innovation Centre. The Metabolomics Innovation Centre: FooDB (version 1.0). https://foodb.ca/ (2017).

[302] Pajewska-Szmyt, M., Sinkiewicz-Darol, E. & Gadzała-Kopciuch, R. The impact of environmental pollution on the quality of mother's milk. *Environmental Science and Pollution Research* **26**, 7405–7427 (2019).

[303] Neville, M. C. & Walsh, C. T. Effects of xenobiotics on milk secretion and composition. *The American Journal of Clinical Nutrition* **61**, 687S–694S (1995).

[304] Lemay, D. G. *et al.* RNA Sequencing of the Human Milk Fat Layer Transcriptome Reveals Distinct Gene Expression Profiles at Three Stages of Lactation. *PLoS ONE* **8**, e67531 (2013).

[305] Maningat, P. D. *et al.* Gene expression in the human mammary epithelium during lactation: the milk fat globule transcriptome. *Physiological Genomics* **37**, 12–22 (2009).

[306] Rogan, W. J. *et al.* Polychlorinated biphenyls (PCBs) and dichlorodiphenyl dichloroethene (DDE) in human milk: effects on growth, morbidity, and duration of lactation. *American Journal of Public Health* **77**, 1294–1297 (1987).

[307] Hill, P. D., Chatterton, R. T. & Aldag, J. C. Serum Prolactin in Breastfeeding: State of the Science. *Biological Research For Nursing* **1**, 65–75 (1999).

[308] Uvnäs-Moberg, K. & Eriksson, M. Breastfeeding: physiological, endocrine and behavioural adaptations caused by oxytocin and local neurogenic activity in the nipple and mammary gland. *Acta Paediatrica* **85**, 525–530 (1996).

[309] Chatterjee, O. *et al.* An overview of the oxytocin-oxytocin receptor signaling network. *Journal of Cell Communication and Signaling* **10**, 355–360 (2016).

[310] Kandasamy, K. *et al.* NetPath: a public resource of curated signal transduction pathways. *Genome Biology* **11**, R3 (2010).

[311] Radhakrishnan, A. *et al.* A pathway map of prolactin signaling. *Journal of Cell Communication and Signaling* **6**, 169–173 (2012).

[312] Kanehisa, M. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Research* **28**, 27–30 (2000).

[313] Rebelo, F. M. & Caldas, E. D. Arsenic, lead, mercury and cadmium: Toxicity, levels in breast milk and the risks for breastfed infants. *Environmental Research* **151**, 671–688 (2016).

[314] Dawod, B. & Marshall, J. S. Cytokines and Soluble Receptors in Breast Milk as Enhancers of Oral Tolerance Development. *Frontiers in Immunology* **10**, 16 (2019).

[315] Jackson, K. M. & Nazar, A. M. Breastfeeding, the immune response, and long-term health. *Journal of Osteopathic Medicine* **106**, 203–207 (2006).

[316] Bagley, C. J., Woodcock, J. M., Stomski, F. C. & Lopez, A. F. The Structural and Functional Basis of Cytokine Receptor Activation: Lessons From the Common β Subunit of the Granulocyte-Macrophage Colony-Stimulating Factor, Interleukin-3 (IL-3), and IL-5 Receptors. *Blood* **89**, 1471–1482 (1997).

[317] Cohen, M. Symposium overview: alterations in cytokine receptors by xenobiotics. *Toxicological Sciences* **48**, 163–169 (1999).

[318] Cameron, M. J. & Kelvin, D. J. Cytokines, chemokines and their receptors. In *Madame Curie Bioscience Database [Internet]* (Landes Bioscience, 2013).

[319] HGNC: HUGO Gene Nomenclature Committee. www.genenames.org.

[320] Braschi, B. *et al.* Genenames.org: the HGNC and VGNC resources in 2019. *Nucleic Acids Research* **47**, D786–D792 (2019).

[321] Armstrong, J. F. *et al.* The IUPHAR/BPS Guide to Pharmacology in 2020: extending immunopharmacology content and introducing the IUPHAR/MMV Guide to Malaria Pharmacology. *Nucleic Acids Research* **48**, D1006–D1021 (2020).

[322] Ito, S. & Alcorn, J. Xenobiotic transporter expression and function in the human mammary gland. *Advanced Drug Delivery Reviews* **55**, 653–665 (2003).

[323] García-Lino, A. M., Álvarez Fernández, I., Blanco-Paniagua, E., Merino, G. & Álvarez, A. I. Transporters in the Mammary Gland—Contribution to Presence of Nutrients and Drugs into Milk. *Nutrients* **11**, 2372 (2019).

[324] Montalbetti, N., Dalghi, M. G., Albrecht, C. & Hediger, M. A. Nutrient Transport in the Mammary Gland: Calcium, Trace Minerals and Water Soluble Vitamins. *Journal of Mammary Gland Biology and Neoplasia* **19**, 73–90 (2014).

[325] Ventrella, D., Forni, M., Bacci, M. L. & Annaert, P. Non-clinical Models to Determine Drug Passage into Human Breast Milk. *Current Pharmaceutical Design* **25**, 534–548 (2019).

[326] Alcorn, J., Lu, X., Moscow, J. A. & McNamara, P. J. Transporter Gene Expression in Lactating and Nonlactating Human Mammary Epithelial Cells Using Real-Time Reverse Transcription-Polymerase Chain Reaction. *Journal of Pharmacology and Experimental Therapeutics* **303**, 487–496 (2002).

[327] Mandal, B. & Suzuki, K. T. Arsenic round the world: a review. *Talanta* **58**, 201–235 (2002).

[328] World Health Organization. Arsenic: Fact sheets. https://www.who.int/news-room/fact-sheets/detail/arsenic (February 2018).

[329] Smith, A. H. & Smith, M. M. Arsenic drinking water regulations in developing countries with extensive exposure. *Toxicology* **198**, 39–44 (2004).

[330] Bhattacharya, P., Chatterjee, D. & Jacks, G. Occurrence of Arsenic-contaminated Groundwater in Alluvial Aquifers from Delta Plains, Eastern India: Options for Safe Drinking Water Supply. *International Journal of Water Resources Development* **13**, 79–92 (1997).

[331] Borah, K. K., Bhuyan, B. & Sarma, H. P. Lead, arsenic, fluoride, and iron contamination of drinking water in the tea garden belt of Darrang district, Assam, India. *Environmental Monitoring and Assessment* **169**, 347–352 (2010).

[332] Sharma, C., Mahajan, A. & Garg, U. K. Assessment of arsenic in drinking water samples in south-western districts of Punjab—India. *Desalination and Water Treatment* **51**, 5701–5709 (2013).

[333] Kumar, M., Rahman, M. M., Ramanathan, A. & Naidu, R. Arsenic and other elements in drinking water and dietary components from the middle Gangetic plain of Bihar, India: Health risk index. *Science of the Total Environment* **539**, 125–134 (2016).

[334] U.S. EPA. Child-Specific Exposure Scenarios Examples (Final Report). Tech. Rep. EPA/600/R-14-217F, U.S. Environmental Protection Agency, Washington, DC (2014).

[335] Landrigan, P. J. & Goldman, L. R. Children's Vulnerability To Toxic Chemicals: A Challenge And Opportunity To Strengthen Health And Environmental Policy. *Health Affairs* **30**, 842–850 (2011).

[336] Negev, M. *et al.* Regulation of chemicals in children's products: How U.S. and EU regulation impacts small markets. *Science of the Total Environment* **616-617**, 462–471 (2018).

[337] Brod, B. A., Treat, J. R., Rothe, M. J. & Jacob, S. E. Allergic contact dermatitis: Kids are not just little people. *Clinics in Dermatology* **33**, 605–612 (2015).

[338] Högberg, J. *et al.* Phthalate Diesters and Their Metabolites in Human Breast Milk, Blood or Serum, and Urine as Biomarkers of Exposure in Vulnerable Populations. *Environmental Health Perspectives* **116**, 334–339 (2008).

[339] Bridges, B. Fragrances and health. *Environmental Health Perspectives* **107**, A340 (1999).

[340] Pinkas, A., Gonçalves, C. L. & Aschner, M. Neurotoxicity of fragrance compounds: A review. *Environmental Research* **158**, 342–349 (2017).

[341] Krowech, G. *et al.* Identifying Chemical Groups for Biomonitoring. *Environmental Health Perspectives* **124**, A219–A226 (2016).

[342] Bridges, B. Fragrance: emerging health and environmental concerns. *Flavour and Fragrance Journal* **17**, 361–371 (2002).

[343] Aurisano, N., Fantke, P., Huang, L. & Jolliet, O. Estimating mouthing exposure to chemicals in children's products. *Journal of Exposure Science & Environmental Epidemiology* (2021).

[344] Preferred Reporting Items for Systematic Reviews and Meta-analyses (PRISMA). http://www.prisma-statement.org/.

[345] Flavornet. http://www.flavornet.org/flavornet.html.

[346] Arn, H. & Acree, T. Flavornet: a database of aroma compounds based on odor potency in natural products. *Developments in food science* **40**, 27–28 (1998).

[347] FlavorDB. https://cosylab.iiitd.edu.in/flavordb/.

[348] The Good Scents Company Information System. http://www.thegoodscentscompany.com/.

[349] Oregon Health Authority. High Priority Chemicals of Concern for Children's Health. https://www.oregon.gov/oha/ph/healthyenvironments/healthyneighborhoods/toxicsubstances/pages/childrens-chemicals-of-concern.aspx (2015).

[350] Vermont Department of Health. Chemicals of High Concern to Children's products rule. https://www.healthvermont.gov/sites/default/files/documents/pdf/Env_CDP_chemicals_high_concern_childrens_products_rule.pdf (2020).

[351] The International Fragrance Association (IFRA). IFRA Transparency List. https://ifrafragrance.org/priorities/ingredients/ifra-transparency-list.

[352] NORMAN: Toxic Plant Phytotoxin (TPPT) Database. https://comptox.epa.gov/dashboard/chemical_lists/PHYTOTOXINS.

[353] Drechsel, D. A. *et al.* Skin Sensitization Induction Potential From Daily Exposure to Fragrances in Personal Care Products. *Dermatitis* **29**, 324–331 (2018).

[354] U.S. National Toxicology Program. ICCVAM: Skin Corrosion 2004 collection from NIEHS. https://comptox.epa.gov/dashboard/chemical_lists/ICCVAMSKIN (2004).

[355] U.S. National Toxicology Program. ICCVAM: Local Lymph Node Assay (LLNA) 2009. https://comptox.epa.gov/dashboard/chemical_lists/ICCVAMLLNA (2009).

[356] National Institute for Occupational Safety and Health. NIOSH: Skin Notation Profiles. https://comptox.epa.gov/dashboard/chemical_lists/NIOSHSKIN (2009).

[357] Baldi, P. & Nasr, R. When is Chemical Similarity Significant? The Statistical Distribution of Chemical Similarity Scores and Its Extreme Values. *Journal of Chemical Information and Modeling* **50**, 1205–1222 (2010).

[358] Farbiszewski, R. & Kranc, R. Olfactory receptors and the mechanism of odor perception. *Polish Annals of Medicine* **20**, 51–55 (2013).

[359] Gaillard, I., Rouquier, S. & Giorgi, D. Olfactory receptors. *Cellular and Molecular Life Sciences CMLS* **61**, 456–469 (2004).

[360] Genva, M., Kenne Kemene, T., Deleu, M., Lins, L. & Fauconnier, M.-L. Is It Possible to Predict the Odor of a Molecule on the Basis of its Structure? *International Journal of Molecular Sciences* **20**, 3018 (2019).

[361] Odor Molecules Database (OdorDB). https://senselab.med.yale.edu/odordb/.

[362] Crasto, C. J. The olfactory receptor database: web-based resources for the genomics, proteomics and function of olfactory receptors. *Flavour* **3**, O8 (2014).

[363] Olfactory Receptor Database (ORDB). http://ycmi.med.yale.edu/senselab/ordb/.

[364] Skoufos, E. Olfactory Receptor Database: a sensory chemoreceptor resource. *Nucleic Acids Research* **28**, 341–343 (2000).

[365] van de Sandt, J. *et al.* The Use of Human Keratinocytes and Human Skin Models for Predicting Skin Irritation: The Report and Recommendations of ECVAM Workshop 38 '. *Alternatives to Laboratory Animals* **27**, 723–743 (1999).

[366] Hennen, J. Keratinocytes improve prediction of sensitization potential and potency of chemicals with THP-1 cells. *ALTEX* **34**, 279–288 (2017).

[367] Kleinstreuer, N. C. *et al.* Phenotypic screening of the ToxCast chemical library to classify toxic and therapeutic mechanisms. *Nature Biotechnology* **32**, 583–591 (2014).

[368] Rappaport, S. M., Barupal, D. K., Wishart, D., Vineis, P. & Scalbert, A. The Blood Exposome and Its Role in Discovering Causes of Disease. *Environmental Health Perspectives* **122**, 769–774 (2014).

[369] Uhlén, M. *et al.* Tissue-based map of the human proteome. *Science* **347**, 1260419 (2015).

[370] Piñero, J. *et al.* The DisGeNET knowledge platform for disease genomics: 2019 update. *Nucleic Acids Research* **48**, D845–D855 (2020).

[371] Gutiérrez-Sacristán, A. *et al.* PsyGeNET: a knowledge platform on psychiatric disorders and their genes. *Bioinformatics* **31**, 3075–3077 (2015).

[372] The UniProt Consortium. UniProt: a hub for protein information. *Nucleic Acids Research* **43**, D204–D212 (2015).

[373] Rath, A. *et al.* Representation of rare diseases in health information systems: The orphanet approach to serve a wide range of end users. *Human Mutation* **33**, 803–808 (2012).

[374] Ma, X., Lee, H., Wang, L. & Sun, F. CGI: a new approach for prioritizing genes by combining gene expression and protein–protein interaction data. *Bioinformatics* **23**, 215–221 (2007).

[375] Landrum, M. J. *et al.* ClinVar: public archive of interpretations of clinically relevant variants. *Nucleic Acids Research* **44**, D862–D868 (2016).

[376] Martin, A. R. *et al.* PanelApp crowdsources expert knowledge to establish consensus diagnostic gene panels. *Nature Genetics* **51**, 1560–1565 (2019).

[377] Goodman, Z. D. Neoplasms of the liver. *Modern Pathology* **20**, S49–S60 (2007).

[378] Aleksandrova, K., Stelmach-Mardas, M. & Schlesinger, S. Obesity and Liver Cancer. In Pischon, T. & Nimptsch, K. (eds.) *Obesity and Cancer. Recent Results in Cancer Research, vol 208*, 177–198 (Springer International Publishing, Cham, 2016).

[379] Marchesini, G., Moscatiello, S., Di Domizio, S. & Forlani, G. Obesity-Associated Liver Disease. *The Journal of Clinical Endocrinology & Metabolism* **93**, s74–s80 (2008).

[380] Marengo, A., Rosso, C. & Bugianesi, E. Liver Cancer: Connections with Obesity, Fatty Liver, and Cirrhosis. *Annual Review of Medicine* **67**, 103–117 (2016).

[381] Holtcamp, W. Obesogens: an environmental link to obesity. *Environmental Health Perspectives* **120**, a62–a68 (2012).

[382] Valvi, D. *et al.* Prenatal concentrations of polychlorinated biphenyls, DDE, and DDT and overweight in children: a prospective birth cohort study. *Environmental Health Perspectives* **120**, 451–457 (2012).

[383] Gupta, R. *et al.* Endocrine disruption and obesity: A current review on environmental obesogens. *Current Research in Green and Sustainable Chemistry* **3**, 100009 (2020).

[384] Taboureau, O. & Audouze, K. Human Environmental Disease Network: A computational model to assess toxicology of contaminants. *ALTEX* **34**, 289–300 (2017).

[385] Barabási, A.-L., Gulbahce, N. & Loscalzo, J. Network medicine: a network-based approach to human disease. *Nature Reviews Genetics* **12**, 56–68 (2011).

[386] Zhou, X., Menche, J., Barabási, A.-L. & Sharma, A. Human symptoms–disease network. *Nature Communications* **5**, 4212 (2014).

[387] Dobson, C. M. Chemical space and biology. *Nature* **432**, 824–828 (2004).

[388] Lipinski, C. & Hopkins, A. Navigating chemical space for biology and medicine. *Nature* **432**, 855–861 (2004).

[389] Rager, J. E. *et al.* Review of the environmental prenatal exposome and its relationship to maternal and fetal health. *Reproductive Toxicology* **98**, 1–12 (2020).

[390] Helma, C., Kramer, S., Pfahringer, B. & Gottmann, E. Data quality in predictive toxicology: identification of chemical structures and calculation of chemical properties. *Environmental Health Perspectives* **108**, 1029–1033 (2000).

[391] Helma, C., Gottmann, E. & Kramer, S. Knowledge discovery and data mining in toxicology. *Statistical Methods in Medical Research* **9**, 329–358 (2000).

[392] McKinney, J. D. The Practice of Structure Activity Relationships (SAR) in Toxicology. *Toxicological Sciences* **56**, 8–17 (2000).

[393] U.S. EPA. Reference Dose (RfD): Description and Use in Health Risk Assessments. https://www.epa.gov/iris/reference-dose-rfd-description-and-use-health-risk-assessments (1993).

[394] Xue, J., Lai, Y., Liu, C.-W. & Ru, H. Towards Mass Spectrometry-Based Chemical Exposome: Current Approaches, Challenges, and Future Directions. *Toxics* **7**, 41 (2019).

[395] Leist, M. *et al.* Adverse outcome pathways: opportunities, limitations and open questions. *Archives of Toxicology* **91**, 3477–3505 (2017).