

# HExpoChem: a systems biology resource to explore human exposure to chemicals

Olivier Taboureau<sup>1,2,†</sup>, Ulrik Plesner Jacobsen<sup>1,†</sup>, Christian Kalhauge<sup>1</sup>, Daniel Edsgård<sup>1</sup>, Olga Rigina<sup>1</sup>, Ramneek Gupta<sup>1</sup> and Karine Audouze<sup>1,\*</sup>

<sup>1</sup>Department of Systems Biology, Center for Biological Sequence Analysis, Technical University of Denmark, 2800 Lyngby, Denmark and <sup>2</sup>UMR-S973, MTi, University Paris Diderot, F-75013 Paris, France

Associate Editor: Jonathan Wren

## ABSTRACT

**Summary:** Humans are exposed to diverse hazardous chemicals daily. Although an exposure to these chemicals is suspected to have adverse effects on human health, mechanistic insights into how they interact with the human body are still limited. Therefore, acquisition of curated data and development of computational biology approaches are needed to assess the health risks of chemical exposure. Here we present HExpoChem, a tool based on environmental chemicals and their bioactivities on human proteins with the objective of aiding the qualitative exploration of human exposure to chemicals. The chemical–protein interactions have been enriched with a quality-scored human protein–protein interaction network, a protein–protein association network and a chemical–chemical interaction network, thus allowing the study of environmental chemicals through formation of protein complexes and phenotypic outcomes enrichment.

**Availability:** HExpoChem is available at <http://www.cbs.dtu.dk/services/HExpoChem-1.0/>.

**Contact:** karine@cbs.dtu.dk

**Supplementary information:** Supplementary data are available at *Bioinformatics* online.

Received on December 4, 2012; revised on February 6, 2013; accepted on February 28, 2013

## 1 INTRODUCTION

Humans, in their daily lives, are exposed to diverse hazardous chemicals through their environment, and it has been recognized as an important component in the interplay between genes, lifestyle/diet, natural environmental exposure and disease etiology (Judson *et al.*, 2012; Rappaport, 2012). Hazardous chemicals are present in everything from food to consumer products to industrial effluents. However, knowledge about their mechanisms of action within the human body is still limited.

As large-scale databases on chemical toxicity, i.e. ACToR (Judson *et al.*, 2012) and the Comparative Toxicogenomics Database (CTD; Davis *et al.*, 2011), are being assembled, investigation of the relationship between environmental chemicals and human diseases can be investigated much more systematically, not solely at the molecular level but also at the cellular and systems level. For example, integration of chemical toxicology within

systems biology using protein–protein interaction (PPI) and protein–protein association network (PPAN) has been suggested to allow previously uncharacterized connections between chemical compounds and diseases to be revealed (Audouze *et al.*, 2010; Audouze and Grandjean, 2011).

Here, we present HExpoChem, a server with the aim to explore chemicals exposure and their impact on human health. HExpoChem is distinct from other tools, as (i) it provides information about the source of the chemical exposure, (ii) it identifies proteins interacting with a chemical and consequently suggests chemical–chemical interactions and qualitative cumulative risk of chemical mixtures and (iii) it integrates PPI and PPAN data into biological outcomes and environmental diseases using computational biology approaches.

## 2 IMPLEMENTATION

### 2.1 Exposure resources

Five different human environmental exposures are considered in HExpoChem, four of which are exogenous sources: medicinal drugs, cosmetics, foods and industrial chemicals such as pesticides and plasticizers. Human endogenous metabolites are chosen as the endogenous source, as these might interfere with the xenobiotics. Description of the data source can be found in additional Supplementary Material.

For each source of exposure, we developed a chemical–protein repository by manual curation, using interactions from three databases, i.e. STITCH 3.1 (Kuhn *et al.*, 2012), CTD and DrugBank (Knox *et al.*, 2011). We integrated 10 183 unique chemicals with 19 483 human proteins for a total of 284 533 interactions (details provided in Supplementary Table S1). An interaction could either be binding data from STITCH and DrugBank (chemical A is binding to protein B) or gene expression information from CTD (chemical A is affecting the gene expression of protein B). The effect of a small molecule on a protein, i.e. agonist or antagonist, is included when available. We decided to not include all chemical–protein interactions provided by STITCH, but instead only consider interactions that reached at least a confidence score of 0.5 (a little superior to a medium confidence according to STITCH).

### 2.2 Connecting chemical biology to diseases and biological outcomes

Once the relationships between chemicals and proteins are characterized, the impact of such associations is evaluated as a

\*To whom correspondence should be addressed.

<sup>†</sup>The authors wish it to be known that, in their opinion, the first two authors should be regarded as joint First Authors.

complex system. Diverse biological outcomes of interest such as (i) human diseases: OMIM (Hamosh *et al.*, 2005) and GeneCards (Safran *et al.*, 2010), (ii) pathways: KEGG (Kanehisa *et al.*, 2010) and Reactome (Croft *et al.*, 2011) and (iii) Gene Ontology terms: GO (Camon *et al.*, 2004) are evaluated. Four output styles are generated to assist the user: (i) PPI, (ii) PPAN, (iii) topological protein clustering (TPC) and (iv) chemical-chemical interactions (CCI). Globally, for each PPI and PPAN output, a small biological network representing a protein complex is depicted. Pointing the cursor to one of the annotations highlights the corresponding proteins. A corrected *P*-value is included to evaluate the correlation of the protein complex to a biological outcome (the lower the value, the stronger is the correlation). The outputs for TPC are lists of proteins corresponding to a cluster and overlapping proteins. Finally, CCIs are based on the assumption that two bio-active chemicals (i.e. agonist-agonist or antagonist-antagonist) for the same set of proteins could have a more significant cumulative effect than only showing an activity on one protein.

Overall, a download option is available to obtain the output information in a flat format for further analysis. Details about these outputs can be found in the Supplementary Section. A workflow of the implementation is shown in Figure 1.

### 3 APPLICATIONS

As an example, we tested carvone, a natural compound found in essential oils from caraway and dill. This compound is used in food additives and cosmetics and has exhibited repellent and herbicide activities. HExpoChem automatically represented the chemical structure (using the Chemaxon applet) and seven proteins that are deregulated by this compound. ‘Cosmetic’, ‘food’ and ‘industrial’ were depicted as the possible sources for human exposure to it. PPI search retrieved protein interactions for six of the proteins, and enrichment to several biological outcomes were suggested. For example, the protein complex interacting with the protein C2 was associated with lupus erythematosus, a potential

reaction to carvone. The CCI search shows that many others chemicals also interact with the protein C2 and so can increase the impact on the biological outcomes described in the PPI. More examples are provided in the Supplementary Material.

### 4 CONCLUSION

In conclusion, HExpoChem assists in the exploration of underlying molecular mechanisms of the impact of environmental chemicals on human health. With the proposed computational systems biology approach, identification of chemicals with similar mode of action in a biological system is of great interest. For instance, chemicals present in food, cosmetics and in household products may share interacting proteins and pathways, and thus affect human health. Therefore, HExpoChem is complementary to traditional toxicology and epidemiology, and may help in further investigations of chemical risk assessment by providing new hypotheses. We expect to maintain and update HExpoChem regularly. For instance, integration of the biological activity measurement and description of the directionality of a PPI are already in the pipeline for implementation in the near future.

### ACKNOWLEDGEMENTS

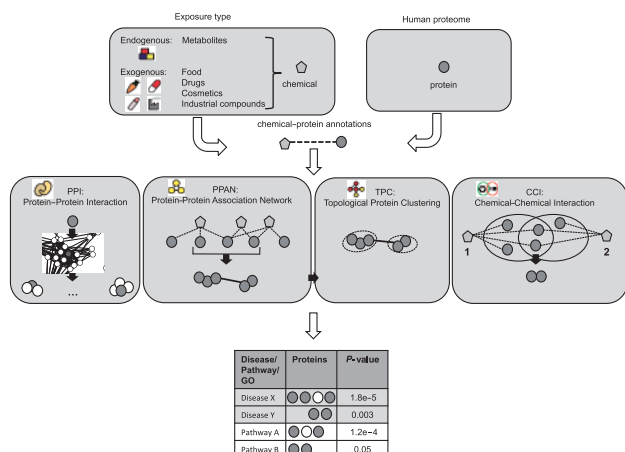
The authors would like to acknowledge the Innovative Medicines Initiative Joint Undertaking (eTOX), the EU project DEER, the Villum Kann Rasmussen foundation (VKR), the Danish Cancer Society and the GENDINOB project supported by the Danish Council for Strategic Research for supporting the work.

*Funding:* GENDINOB.

*Conflicts of Interest:* none declared.

### REFERENCES

- Audouze, K. *et al.* (2010) Deciphering diseases and biological targets for environmental chemicals using toxicogenomics networks. *PLoS Comput. Biol.*, **6**, e1000788.
- Audouze, K. and Grandjean, P. (2011) Application of computational systems biology to explore environmental toxicity hazards. *Environ. Health Perspect.*, **119**, 1754–1759.
- Camon, E. *et al.* (2004) The gene ontology annotation (GOA) database – sharing knowledge in Uniprot with gene ontology. *Nucleic Acids Res.*, **32**, D262–D266.
- Croft, D. *et al.* (2011) Reactome: a database of reactions, pathways and biological processes. *Nucleic Acids Res.*, **39**, D691–D697.
- Davis, A.P. *et al.* (2011) The comparative toxicogenomics database: update 2011. *Nucleic Acids Res.*, **39**, D1067–D1072.
- Hamosh, A. *et al.* (2005) Online Mendelian Inheritance in Man (OMIM), a knowledgebase of human genes and genetic disorders. *Nucleic Acids Res.*, **33**, D514–D517.
- Judson, R.S. *et al.* (2012) Aggregating data for computational toxicology applications: the U.S. Environmental Protection Agency (EPA) Aggregated Computational Toxicology Resource (ACToR) system. *Int. J. Mol. Sci.*, **13**, 1805–1831.
- Kanehisa, M. *et al.* (2010) KEGG for representation and analysis of molecular networks involving diseases and drugs. *Nucleic Acids Res.*, **38**, D355–D360.
- Knox, C. *et al.* (2011) DrugBank 3.0: a comprehensive resource for ‘omics’ research on drugs. *Nucleic Acids Res.*, **39**, D1035–D1041.
- Kuhn, M. *et al.* (2012) STITCH 3: zooming in on protein-chemical interactions. *Nucleic Acids Res.*, **40**, 876–880.
- Rappaport, S.M. (2012) Discovering environmental causes of disease. *J. Epidemiol. Community Health*, **66**, 99–102.
- Safran, M. *et al.* (2010) GeneCards version 3: the human gene integrator. *Database*, **2010**, article ID baq020; doi:10.1093/database/baq020.



**Fig. 1.** The HExpoChem implementation. Chemical data for five exposure sources are collected from various resources, and combined with their interacting proteins. Four types of protein complexes (PPI, PPAN, TPC and CCI) are generated, using different input data. diseases, pathways and Gene Ontology terms are integrated into these protein complexes, allowing prediction of potential biological outcomes of chemicals