OpenRiskNet

RISK ASSESSMENT E-INFRASTRUCTURE

Case Study

Toxicogenomics-based prediction and mechanism identification [**TGX**]

SUMMARY **2**

DESCRIPTION **3**

Implementation team 3

Case Study objectives 3

Risk assessment framework 3

DEVELOPMENT **4**

Databases and tools 4

Technical implementation 4

OUTCOMES **5**

First top-down approach 5

Second top-down approach 5

Related resources 6

REFERENCES **7**

SUMMARY

In this case study a transcriptomics-based hazard prediction model for identification of specific molecular initiating events (MIE) was foreseen based on (A) top-down and (B) bottom-up approaches.

The MIEs can include, but are not limited to: (1) Genotoxicity (p53 activation), (2) Oxidative stress (Nrf2 activation), (3) Endoplasmic Reticulum Stress (unfolded protein response), (4) Dioxin-like activity (AhR receptor activation), (5) HIF1 alpha activation and (6) Nuclear receptor activation (e.g. for endocrine disruption).

This case study focussed on two top-down approaches for genotoxicity prediction. The first approach resulted in the creation of a Nextflow-based workflow from the publication "A transcriptomics-based in vitro assay for predicting chemical genotoxicity in vivo" by Magkoufopoulou *et al.* (2012), thereby reproducing their work as proof of principle. The workflow for one of the prediction models described in the publication is shown in Figure 1.

The Nextflow-based workflow has been translated into a more generic approach, especially for step 1, forming the basis of the second top-down approach. In this approach transcriptomics data together with toxicological compound information were collected from multiple toxicogenomics studies and used for building a metadata genotoxicity prediction model.

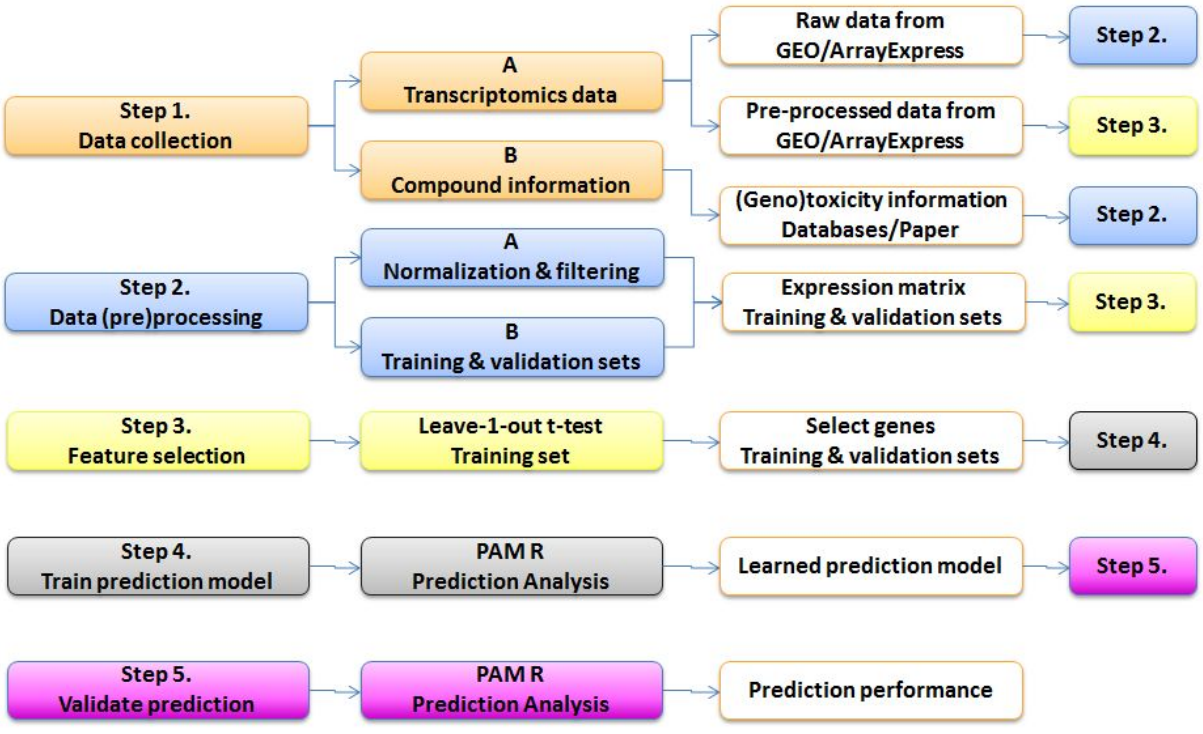

Figure 1. Workflow for genotoxicity prediction using whole genome transcriptomics data.

DESCRIPTION

Implementation team

Coordination:

- Danyel Jennen, Maastricht University, Department of Toxicogenomics

Other members:

- Jumamurat Bayjanov, Maastricht University, Department of Toxicogenomics
- Evan Floden, CRG

Case Study objectives

- Creation of prediction models based on differentially regulated genes (top-down approach);
- Using knowledge of stress response pathways to integrate data sets for their activation or inhibition (bottom-up approach).

These two use cases are relevant for the top-down approaches:

- Reproducing the prediction models published by Magkoufopoulou et al 2012;
- Advanced predictions using as much data as possible from the diXa data warehouse[1] and other repositories giving free access to the data.

Risk assessment framework

This case study is associated with all 3 tiers of the selected framework and in particular the following steps:

- Collection of support data;
- Identification of analogues / suitability assessment and existing data;
- Mode of Action hypothesis generation.

[1] http://wwwdev.ebi.ac.uk/fg/dixa/index.html

DEVELOPMENT

Databases and tools

Databases:

- diXa data warehouse (carcinoGENOMICs, Predict-IV), TG-GATEs, ArrayExpress/GEO, BioStudies.

Tools:

- Top-down: data normalisation tools, prediction tools such as Caret[2];
- Bottom-up: ToxPi.

Technical implementation

Integration with other case studies is needed. TGX acquires information and data from the DataCure case study as well as through the services ToxPlanet and ToxicoDB of the Implementation Challenge winners Toxplanet and UHH, respectively. The results of TGX can feed into SysGroup, AOPLink and ModelRX.

Currently available services:

- Nextflow
  - Service to run Nextflow pipelines
  - Service type: Service, Workflow, Software
- Transcriptomics data from human, mouse, rat in vitro liver models
  - Repository for transcriptomics data from multiple in vitro human, rat and mouse toxicogenomics projects
  - Service type: Database / data source

[2] http://topepo.github.io/caret/index.html

OUTCOMES

Outcome from this case study provides workflows for obtaining data which are suited for developing toxicity prediction models. This resulted in two top-down approaches for genotoxicity prediction.

First top-down approach

A workflow from the earlier publication "A transcriptomics-based in vitro assay for predicting chemical genotoxicity in vivo" by Magkoufopoulou *et al.* (2012) [1] was developed, thereby reproducing their work as proof of principle. The workflow was created for one of the three approaches that were described in the study. There were some minor differences between the newly developed workflow and the original study, but overall the results of the original study were reproduced.

The workflow created using the Snakemake workflow manager is available from a GitLab software repository[3], where every step is clearly described in Snakefile to reproduce the approach described by Magkoufopoulou *et al.* (2012) [1] and was used as reference for the transfer into an OpenRiskNet-based solution. The repository also includes required scripts as well as description of the steps necessary to reproduce the results.

The Snakemake-based workflow was converted to a Nextflow-based workflow[4], named nf-toxomix, in order to make use of the harmonization and interoperability of OpenRiskNet. The Nextflow version uses containerised steps, thus making it easier to deploy on any cloud infrastructure, and applicable to OpenRiskNet Virtual Environments. Furthermore, the Nextflow-based workflow has been translated into a more generic approach so that it can be applied to other toxicogenomics studies.

Second top-down approach

In this approach, transcriptomics data on human, mouse and rat *in vitro* liver cell models exposed to hundreds of compounds were collected from the diXa data warehouse, NCBI GEO and EBI's ArrayExpress using the workflow from the first top-down approach. The obtained datasets were merged per species. This was done manually because of differences in the description of the datasets, e.g. differences in used ontologies, different metadata file formats. For all the compounds used in the experiments genotoxic and carcinogenic information was gathered from literature and several databases, including ToxPlanet. After normalization of the transcriptomics data (per species) and gathering of the genotoxicity/carcinogenicity information, the data were ready to be fed into the prediction models of ModelRX.

The transcriptomics data from the human, mouse, rat *in vitro* liver cell models and the toxicological information are available through OpenRiskNet[5].

[3] https://gitlab.com/bayjan/openrisknet_magkoufopoulou
[4] https://github.com/openrisknet/nf-toxomix
[5] https://gitlab.com/bayjan/openrisknet_meta_analysis_data

Related resources

Use of Nextflow tool for toxicogenomics-based prediction and mechanism identification in OpenRiskNet e-infrastructure

Evan Floden

27 May 2019 | Webinar

OpenRiskNet Part II: Predictive Toxicology based on Adverse Outcome Pathways and Biological Pathway Analysis

Marvin Martens, Thomas Exner, Nofisat Oki, Danyel Jennen, Jumamurat Bayjanov, Chris Evelo, Tim Dudgeon, Egon Willighagen

28 August 2019 | Poster

Meta-analysis for genotoxicity prediction using data from multiple human in vitro cell models

Jumamurat R. Bayjanov Jos Kleinjans Danyel Jennen

12 Sep 2018 | Poster

Big Data in Toxicogenomics: Towards FAIR predictions

Danyel Jennen

26 Jul 2018 | Presentation ICCA 2018

REFERENCES

1. Magkoufopoulou C, Claessen SM, Tsamou M, Jennen DG, Kleinjans JC, van Delft JH. A transcriptomics-based in vitro assay for predicting chemical genotoxicity in vivo. Carcinogenesis. 2012 Jul;33(7):1421-9. doi:10.1093/carcin/bgs182.