Novel computational tools based on bioinformatic and chemoinformatic data to complement zebrafish embryo teratogenicity test

Xabier Cendoya1, Celia Quevedo2, Ángel Rubio3, Maitane Ipiñazar4, Arantxa Muriana2 and Francisco J. Planes1

1CEIT and Tecnun, University of Navarra, San Sebastian, Spain. galanes@tecnun.es
2BIOBIDE; San Sebastian, Spain. quevedo@biobide.es

INTRODUCTION
Approximately 3% of newborns present congenital anomalies and around 5-10% of those are caused by exposure to teratogenic agents. For this reason, regulatory organisms and the industry demand for effective methods to test the developmental toxicity of drugs, industry chemicals or wastes products. The use of the zebrafish embryotoxicity test is an attractive strategy to minimize in vivo assays and animal models. Overall, this assay has a good predictability; however, the outcome is based on morphologic evaluation, which is subjective and subtle effects might be neglected. With the increasing amount of molecular databases, the development of in silico tools that complement experimental assays is promising. In this work, we present an in silico platform that makes use of bioinformatics and chemoinformatics data, as well as machine learning methods, in order to predict the teratogenic potential of a particular compound. First, we show a combined systems biology and metabolomics study in order to identify metabolic biomarkers that improve the sensitivity of the zebrafish embryotoxicity test. Second, a learning algorithm using structural information is evaluated and compared using publicly available data, analyzing their complementarity with the zebrafish embryotoxicity test and metabolic biomarkers with newly generated proprietary data.

METHODS
This work consists of an in silico platform (TERATOOL) that integrates diverse sources of data to allow bioinformatics and chemoinformatics analysis (Figure 1). The information that was integrated included:

- A database of approximately 400 compounds with labels for their risk of teratogenicity. 290 of them were obtained from Enoch et al [1].
- Chemical structures and properties, together with biological target data, were obtained from CHEMBL [2], DrugBank [3] and HMDB [4].
- A number of publicly available transcriptomics experiments of the zebrafish embryotoxicity test (40 compounds) [5,6,7,8,9,10].
- A metabolic network reconstruction of the zebrafish was obtained from Bekaaert et al [11].

RESULTS
Identifying metabolic biomarkers
The reporter metabolites algorithm [12] was used with the transcriptomics data and the zebrafish metabolic network to search for metabolites that indicated highly altered regions of the metabolic network, by establishing an integrative score based on differential expression analysis of neighbor genes (Figure 2). We found several metabolites potentially reporting teratogenic action with a substantially higher redundancy than gene biomarkers, and these are currently being analyzed experimentally.

Prediction of teratogenicity with machine learning
Molecular fingerprints were obtained and compared for similarity using the Tanimoto metric. Assuming that similar molecules tend to have similar properties, new targets were annotated for each molecule when they appeared recurrently as targets of similar molecules. The new related molecules were evaluated computationally using Autodock [13] whenever protein structures where available, obtaining good affinities (< -6 kcal/mol).

Using the molecular fingerprints and annotated genes, a machine learning algorithm was developed. A variable selection was carried out, leaving out properties that had low variance or were not good predictors of teratogenicity. We used a support vector machine with a polynomial kernel of degree 1, obtained the cost with tuning function, and obtained good predictive power (85% in training, 81% in validation) for the molecular characteristics of the molecules in the database (Figure 3).

CONCLUSIONS
Bioinformatic and chemoinformatic methods seem to be capable of complementing current experimental methods in the testing of teratogenicity by proposing biomarkers and consistent, data-driven approach to the prediction of the teratogenic potential of a certain molecule.

REFERENCES