Digital Personalized Health and Medicine
L.B. Pape-Haugaard et al. (Eds.)

doi:10.3233/SHTI200427

Linking Genome and Exposome: Computational Analysis of Human Variation in Chemical-Target Interactions

Liana BRUGGEMANN[a], Christopher HAWTHORNE[b], Ram SAMUDRALA[a] and Guillermo H. LOPEZ-CAMPOS[b,1]
[a] *Department of Biomedical Informatics, State University of New York (SUNY), Buffalo, NY, USA*
[b] *Wellcome-Wolfson Institute for Experimental Medicine, Queen's University Belfast, Belfast, UK*

Abstract. The growing amount of available public data repositories containing a plethora of rich chemical and biomedical information is enabling new in silico research avenues. In this project we aim to link human genome variations and the exposome applying in silico biomedical informatics approaches to analyse the potential effects of those variants in the interactions with different chemicals.

Keywords. Translational bioinformatics, Protein modelling, exposome, single nucleotide variant, toxicogenomics.

1. Introduction

Nowadays, human societies are living in an environment where individuals are exposed to an increasing number of xenobiotic chemicals. The acknowledgment of the importance of these and other exposures led to the development of the concept of the exposome defined as the whole set of exposures of an individual (including, chemical, biological, physical or and social exposures) [1]. The exposome could affect the human phenotype along with the genome and therefore its understanding could be important in the development of precision medicine.

Two decades ago, toxicogenomics was developed combining toxicology with genomics and post-genomics approaches to study the effects of the exposures to chemicals and chemical toxicants. Biomedical informatics has played a key role in this discipline, providing tools and methods to analyse, store and integrate the results from these "omics" techniques. More recently, the development of modelling and simulation tools and methods designed to predict the potential interactions of newly designed chemicals with biological targets (mostly hormone and other cellular receptors) has been another important area of development. Toxicogenomics can be considered as the toxicological relative of pharmacogenomics. An important area of research in the latter is the analysis of the effects of the genetic variants in the response to different drugs, and this has led to different applications and biomedical informatics resources. The analysis

[1] Corresponding Author, Wellcome-Wolfson Institute for Experimental Medicine, Queen's University Belfast, 97 Lisburn road, Belfast, BT9 7BL, United Kingdom; E-mail: g.lopezcampos@qub.ac.uk.

of individual genetic variation is an aspect that has not been explored in detail in the area of toxicogenomics. Therefore, in this work we propose the development of a computational analytical pipeline designed to integrate the increasing amounts of information related with human genetic variation and the chemical component of the exposome analyzing the effects of different coding variants in the interactions with chemical compounds.

2. Methods and Results

As a proof of concept for our approach we have chosen to analyse the effects of the different coding genetic variants in the human Estrogen Receptor 1 (ESR1) and how those would affect the interactions with bisphenol-A (BPA), a chemical compound known to link with ESR1 and a potential endocrine disruptor due to its actions as a xeno-estrogen.

For the analyses we selected the wild type (WT) sequence and five single nucleotide variants (SNV) (R269C, R269H, M315V, E330K and R477Q) in the ESR1 gene. These variants were chosen as they were condign variants, they were not known to be pathogenic, were not located in the catalytic/binding site of the protein and had some of the higher minor allele frequencies for the variants in this gene.

We first used I-Tasser [2] to predict the ESR1 structure beyond the resolved region available and then we modelled our five variants using SCWRL [3]. Finally, we used those models to simulate the protein/ligand docking and calculate the binding affinities for BPA and estradiol for each of the five mutants and the wild type structures using CANDOCK [4]. As expected the results showed (table 1) that there was a higher affinity in this case for the natural ligand (estradiol) than for the xenobiotic (BPA) as well as that the effect in the reduction of the binding activity was more intense for the natural ligand than for BPA in all the analysed variants.

Table 1. Table showing the binding affinities (Kcal) for the wild type and the five variants analysed.

	Wild type	**R269C**	**R269H**	**M315V**	**E330K**	**R477Q**
Estradiol	-64.2911	-48.8941	-48.8198	-46.3230	-50.8965	-47.6796
BPA	-47.4722	-36.5059	-39.6272	-41.1594	-39.5739	-41.2424

With this approach we brought together genomics and exposomics applying in silico approaches. This strategy can be extended for the analyses of other targets and variants for the different chemicals to which individuals are exposed daily

References

[1] C.P. Wild. Complementing the genome with an "exposome": the outstanding challenge of environmental exposure measurement in molecular epidemiology. *Cancer Epidemiol Biomarkers Prev*. **8** (2005), 1847-50

[2] A. Roy, A. Kucukural, Y. Zhang. I-TASSER: a unified platform for automated protein structure and function prediction. *Nature Protocols*, **5** (2010), 725-738

[3] Q. Wang, A.A. Canutescu, R.L. Jr. Dunbrack. SCWRL and MolIDE: computer programs for side-chain conformation prediction and homology modeling. *Nat Protoc*.**12** (2008), 1832-47

[4] J.A. Fine, J. Konc, R. Samudrala, G Chopra. CANDOCK: Chemical atomic network based hierarchical flexible docking algorithm using generalized statistical potentials. *BioRxiv*,(2018) 442897