The Importance of Biological Invariance in Drug Design

Oscar Álvarez and Fernández Martínez Juan Luis*

Department of Mathematics, Spain

***Corresponding author:** Fernández Martínez Juan Luis, Department of Mathematics, Spain*

ARTICLE INFO

ABSTRACT

Drug design and discovery is one of the forefront challenges within the pharmaceutical industries, the process which potential new chemicals are identified by means of pharmacology. Since understanding diseases and its potential drugs is an arduous task, the use of Artificial Intelligence has become an important asset. Biological invariance has arisen as a new paradigm in genomics and drug design, since the analysis of the genomics and potential compounds should be independent of the sampling methodology and the classifier utilized for their inference. Predominantly, phenotype prediction and the analysis of altered pathways has become an important discipline in drug discovery and precision medicine, that is, finding a set of genes that prospectively differentiates a specified phenotype disease with respect to a control sample. The problem is highly undetermined since the number of monitored genetic probes exceeds the number of samples, therefore, this creates an ambiguity in the characterization that must be overcome by the use of AI and deep sampling techniques, which ultimately help the drug repurposing and speeding up the drug selection process.

Introduction

Drug discovery comprises the screening of new compounds to rise some desired properties, while reducing the result of potential side effects and improving their efficacy towards a certain disease. All these assets should surge the chance of success in the clinical trials, since the drug development is a capital-intensive procedure. The average cost of developing a new medicine has been appraised to be 2.8 billion dollars [1]. The main reason behind this expenditure is a poor understanding of the disease expansion mechanisms, and consequently; being able to design and reposition drugs that optimally target the actionable genes, while minimizing side effects. Both are opened problems in pharmaco-genomics, which represent the forefront research in the pharmaceutical industry. To allow Precision Medicine, genomic kits are needed to easier diagnosis and adopt optimum decisions [2]. In this sense, the concept of biological invariance seems a promising asset to understand genomics, regardless the methodology employed, since the analysis of the altered pathways is independent of the sampling methodology utilized [3].

Genomics and the Phenotype Prediction Problem

Predicting and characterizing phenotypes from genomics is a complex and undetermined problem, due to the fact the number of monitored probes exceeds the number of samples. To solve this kind on problems the design of mathematical models, called classifiers, that link the genetic signatures to the classes in which the phenotype is divided. The associated uncertainty space to the classifier $L^*(g)$, $Mtol=\{g: O(g)<Etol\}$, is formed by the sets of high predictive networks with similar predictive accuracy, in other words, the sets of genes g whose prediction error, $O(g)$, is lower than a certain Etol. These sets are in flat curvilinear valleys, making the sampling and identification complex and arduous [4-6]. A wide set of algorithms could be utilized to tackle the prediction and classification of altered pathways involved in disease development, such as Nearest-Neighbor classifiers [7-8] Random Forest [9], Extreme Learning Machines [10], Support Vector Machines [11] etc. These algorithms, provided the fact that the noise and the data has been properly treated to minimize its impact in the posterior analysis of

the results, would lead to the same results, as reported by Cernea et al. [3]. Consequently, biological invariance, a concept that implies that the high discriminatory genetic networks are located within the neighborhood of Mtol and the altered pathways are independent of the classifier and the sampling method utilized to unravel them, would help in the drug design and repositioning process.

Drug Design and Repositioning

After a robust sampling of the genetic pathways, an optimum selection of potential drugs can be carried out. This methodology will dramatically increase the approval of drugs and reposition the existing ones. In the present days, there are a wide range of models utilized in drug design, from models based on large data sets of chemicals, compositions and disease-drug activity, known as perturbation models [12], to multiscale models that are capable of integrating different genomic, proteomic and metabolomic information [13]. Furthermore, a new paradigm known as de novo multiscale modelling has emerged, which allows the design of a new drug within the chemical subspace where it could be beneficious [14]. However, regardless of the methodology utilized, powerful AI tools are required in order to integrate information from different types of drugs to biological data and disease specific network studies that serve to identify new potential target [15]. In addition, AI shall be utilized to determine from complex genetic data and all the research information available the interaction a drug may have with its primary target (that tackle the disease) and secondary targets (that causes side-effects). This process is supposed to be, such as in the sampling of altered pathways, independent from the methodology.

Conclusion

Further research and development of AI methods are required to confirm the hypothesis of Biological Invariance, however, it offers a tantalizing approach to handle this complex and difficult problem in the pharmaceutical industry. The mentioned concept would enable the design of affordable genetic tools that would enhance the precision medicine. However, further confirmation is required, being the absence of data one of the major drawback. Biological invariance suggests that any AI protocol to enhance drug design should be iterative and learn from experience. Algorithms and methodologies shall be kept simple and fast, since they would yield to the same results and with the same accuracy.

References

1. Di Masi JA, Grabowski HG, Hansen RW (2016) Innovation in the pharmaceutical industry: New estimates of R&D costs. J Health Econ 47: 20-33.

2. Cook D, Brown D, Alexander R, March R, Morgan P, et al. (2014) Lessons learned from the fate of AstraZeneca's drug pipeline: a five-dimensional framework. Nat Rev Drug Discov 13(6): 419-431.

3. Cernea A, Fernández Martínez JL, de Andrés Galiana EJ, Fernández Ovies FJ, Fernández Muñiz Z, et al. (2018) Comparison of different sampling algorithms for phenotype prediction. International Conference on Bioinformatics and Biomedical Engineering 2: 33-45.

4. Fernández Martínez JL, Fernández Muñiz Z, Tompkins MJ (2012) On the topography of the cost function in linear and nonlinear problems. Geophys 77(1): W1-W15.

5. Fernández Martínez JL, Pallero JLG, Fernandez Muñiz Z, Pedruelo Gonzalez LM (2013) From Bayes to Tarantola: New insights to understand uncertainty in inverse problems. J Appl Geophys 98: 62-72.

6. De Andrés Galiana EJ, Fernández Martínez JL, Sonis S (2016) Design of biomedical robots for phenotype prediction problems. J Comp Biol 23(8): 678-692.

7. Saligan LN, deAndrés Galiana EJ, Fernández Martínez JL, Sonis S (2014) Supervised classification by filter methods and recursive feature elimination predicts risk of radiotherapy related fatigue in patients with prostate cancer. Cancer Inform 13: 141-152.

8. NS Altman (1992) An introduction to kernel and nearest neighbor nonparametric regression. The American Statistician 46(3): 175-185.

9. L Breiman (2001) Random Forests. Machine Learning 45(1): 5-32.

10. Huang Guang Bin, Zhu Qin Yu, Siew Chee Kheong (2006) Extreme learning machine: theory and applications. Neurocomputing 70(1-3): 489-501.

11. Cortes, Corinna, Vapnik, Vladimir N (2017) Deep Learning with Tensor Flow.

12. Speck Planche A, Kleandrova VV, Luan F, Cordeiro MN (2015) Computational modeling in nanomedicine: prediction of multiple antibacterial profiles of nanoparticles using a quantitative structure-activity relationship perturbation model. Nanomedicine 10(2): 193-204.

13. Speck Planche A, Cordeiro MN (2014) Chemoinformatics for medicinal chemistry: in silico model to enable the discovery of potent and safer anti-cocci agents. Future Med Chem 6(18): 2013-2028.

14. Martinez Arzate SG, Tenorio Borroto E, Barbabosa Pliego A, Díaz Albiter HM, Vázquez Chagoyán JC, et al. (2017) PTML model for proteome mining of B-cell epitopes and theoretical experimental study of Bm86 protein sequences from Colima, Mexico. J Proteome Res 16(11): 4093-40103.

15. Berger SI, Iyengar R (2009) Network analyses in systems pharmacology. Bioinformatics 25(19): 2466-2472.