Epigenetic Target Profiler: a web server to predict epigenetic targets of small molecules

Norberto Sánchez-Cruz[1,*] and José L. Medina-Franco[1,*]

[1]DIFACQUIM research group, Department of Pharmacy, School of Chemistry, Universidad Nacional Autónoma de México, Mexico City 04510, Mexico

*To whom correspondence should be addressed.

Abstract

Motivation: The identification of protein targets of small molecules is essential for drug discovery. With the increasing amount of chemogenomic data in the public domain, multiple ligand-based models for target prediction have emerged. However, these models are generally biased by the number of known ligands for different targets, which involves an underrepresentation of epigenetic targets. Epigenetic drug discovery is of increasing importance but there are no open tools for epigenetic target prediction.

Results: We introduce Epigenetic Target Profiler (ETP), a freely accessible and easy-to-use web application for the prediction of epigenetic targets of small molecules. For a query compound, ETP predicts its bioactivity profile over a panel of 55 different epigenetic targets. To that aim, ETP uses a consensus model based on two binary classification models for each target, relying on support vector machines and built on molecular fingerprints of different design. A distance-to-model parameter related to the reliability of the predictions is included to facilitate their interpretability and assist the identification of small molecules with potential epigenetic activity.

Availability: Epigenetic Target Profiler is freely available at http://www.epigenetictargetprofiler.com/

Contact: norberto.sc90@gmail.com (N.S-C.); medinajl@unam.mx (J.L.M-F.)

1. Introduction

The identification of protein targets for small molecules plays a key role in multiple areas of drug discovery, since it allows the prioritization of compounds for the discovery of novel inhibitors against one or a set of therapeutic targets, as well as the estimation of their off-target effects, which can be useful for the study of the polypharmacology of compounds (Anighoro *et al.*, 2014) and drug repurposing (Oprea *et al.*, 2011). The increase in the publicly available chemogenomic data over the years have led to the construction of multiple ligand-based models to predict the protein targets of small molecules, with some of these models available as web-based tools (Sam and Athri, 2019).

Current ligand-based target prediction methods assign the targets for a given small molecule based on the known targets of the most similar ligands in their training datasets, which represents a target-bias over specific protein families. Considering that chemogenomic data related to epigenetics is minimal in comparison to protein families such as kinases, G protein-coupled receptors, and ion channels (Zdrazil *et al.*, 2020), epigenetic targets are underrepresented in the currently available models. Moreover, despite the increasing relevance of these targets across several different therapeutic areas (Sessions *et al.*, 2020), there are no web-based tools to support epigenetic drug discovery.

Herein, we present Epigenetic Target Profiler (ETP), an easy-to-use and free web application for the prediction of the bioactivity profile of small molecules over a panel of 55 epigenetic targets. ETP implements the best performing model for epigenetic target prediction, as identified from a systematic comparison of machine learning models built on molecular fingerprints (FPs) of different design described in a separate work.

2. Approach

In a separate study, we performed a comprehensive comparison of fifteen machine learning models, derived from the combination of three different molecular FPs as compound representations, and five different machine learning algorithms, for binary classification of compounds over 55 epigenetic targets, using a quantitative measure of biological activity cutoff of 10 µM (IC_{50}, EC_{50}, K_i or K_d). We found support vector machines trained on Morgan FPs of radius 2 (Morgan::SVM) and on RDK FPs (RDK::SVM) as the two best performing models for this task. We built a consensus model by combining the predictions of these two models and examined the performance of the individual models and the derived consensus model on a distance-to-model (DM) basis by classifying each prediction into four quartiles (Q1 - Q4) according to the mean Jaccard's distance of the compound to the corresponding training set. The consensus model showed higher precision than the individual models for the prediction of active compounds for all distance quartiles, with a mean precision of 0.896, ranging from 0.923 for compounds in Q1 to 0.810 for compounds in Q4.

The three models were tested for epigenetic target prediction on ten assembled samples containing the same number of active compounds for each of the epigenetic targets. As a result, the consensus model also showed a superior performance for the correct identification of epigenetic targets, with mean precisions ranging from 0.952 for compounds in Q1 to 0.773 for compounds in Q4. The practical applicability of these model was shown by the retrospective identification of the epigenetic targets of two recently reported epigenetic inhibitors (Wilson *et al.*, 2020; Chen *et al.*, 2020). Supported on these findings, we implemented the consensus model as an easy-to-use web application, described in the following section.

3. The ETP web interface

ETP is freely accessible as a web application at http://www.epigenetictargetprofiler.com/ and all row data to reproduce it is available free of charge at figshare repository (10.6084/m9.figshare.13524368). ETP was implemented using Flask version 1.1.1 for Python as web framework and its graphical user interface (GUI) was written in HTML5, CSS and JavaScript with all major browsers supported. The GUI of ETP starts with an initial page wherein the user can either draw a query molecule using the JavaScript based JSME molecular editor (Bienfait and Ertl, 2013) and generate its corresponding SMILES by clicking on the “Get SMILES” button or directly paste the SMILES of a query compound in the cell provided for that purpose (Figure 1).

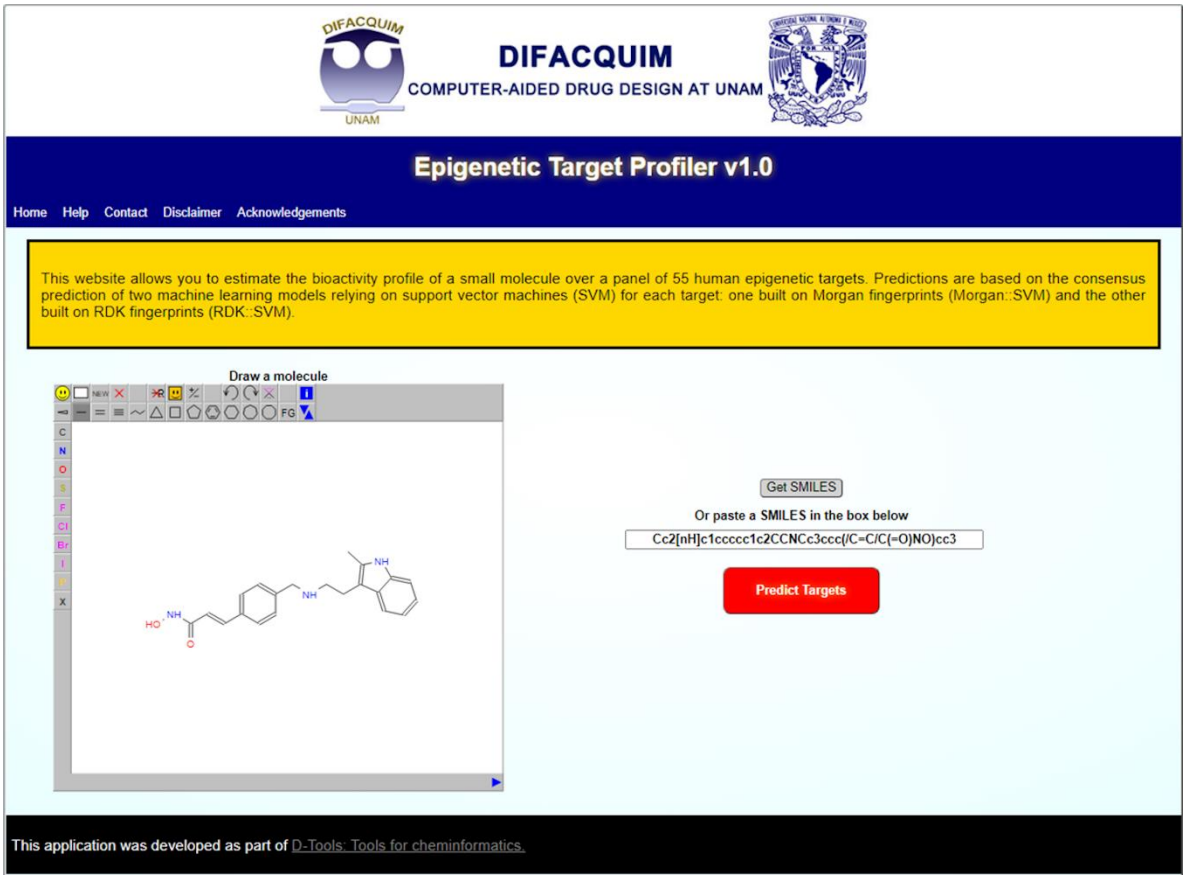

Figure 1. *Graphical User Interface of Epigenetic Target Profiler. Panobinostat is shown in the JSME molecular editor and its SMILES is shown in the cell on the left.*

Following the entry of a query SMILES, the target prediction can be initiated by clicking on the “Predict Targets” button and the user will be directed to the results page. The server standardizes the input compound according to the same process described in Supplementary Section 1 and generates its Morgan and RDK FPs. For each target, if the compound is part of the target-associated compound dataset, no further processing is done and its known association is returned, otherwise the sever computes its mean Jaccard distance to the dataset using Morgan FP, classify it into a quartile accordingly (Supplementary Table S1) and performs the prediction using the Morgan::SVM and RDK::SVM models described in the previous section. The predicted targets for the query compound are those predicted by both models. This process is illustrated in Figure 2. Details on the hyperparameters for each machine learning model and their performances for each target and distance quartile in a 10-fold cross-validation are included in Supplementary Tables S2-S4.

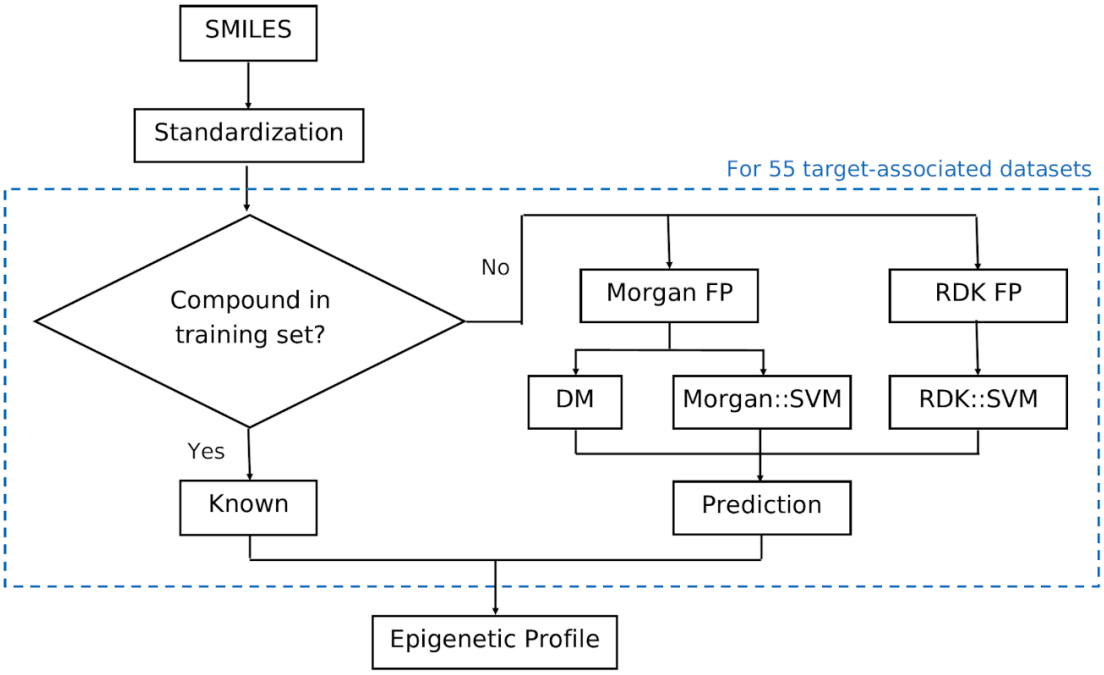

***Figure 2**. Schematic representation of the process performed by ETP.*

Once the predictions have been performed, the user is redirected to the results page, which contains two images at top and a table below them (Figure 3). The image on the left side shows the chemical structure of the query compound as interpreted from the SMILES

submitted by the user, while the image on the right side depicts the chemical structure of the standardized compound as processed by the sever. The results table shows the known and predicted targets for the query compound, including five columns with additional information. The first three columns contain the name of the targets and external links to ChEMBL and GeneCards (Name, ChEMBL ID, Gene) and the last two contain information about the predictions (Status and Quartile). The Status column indicates if the association is known or predicted, while the Quartile column indicates the distance quartile (Q1 – Q4) to which the query compound belongs for each of the predictions as an estimation on its reliability. The full list of results including the predictions from the individual models for all 55 targets can be downloaded by clicking on the “Download CSV” button below the table.

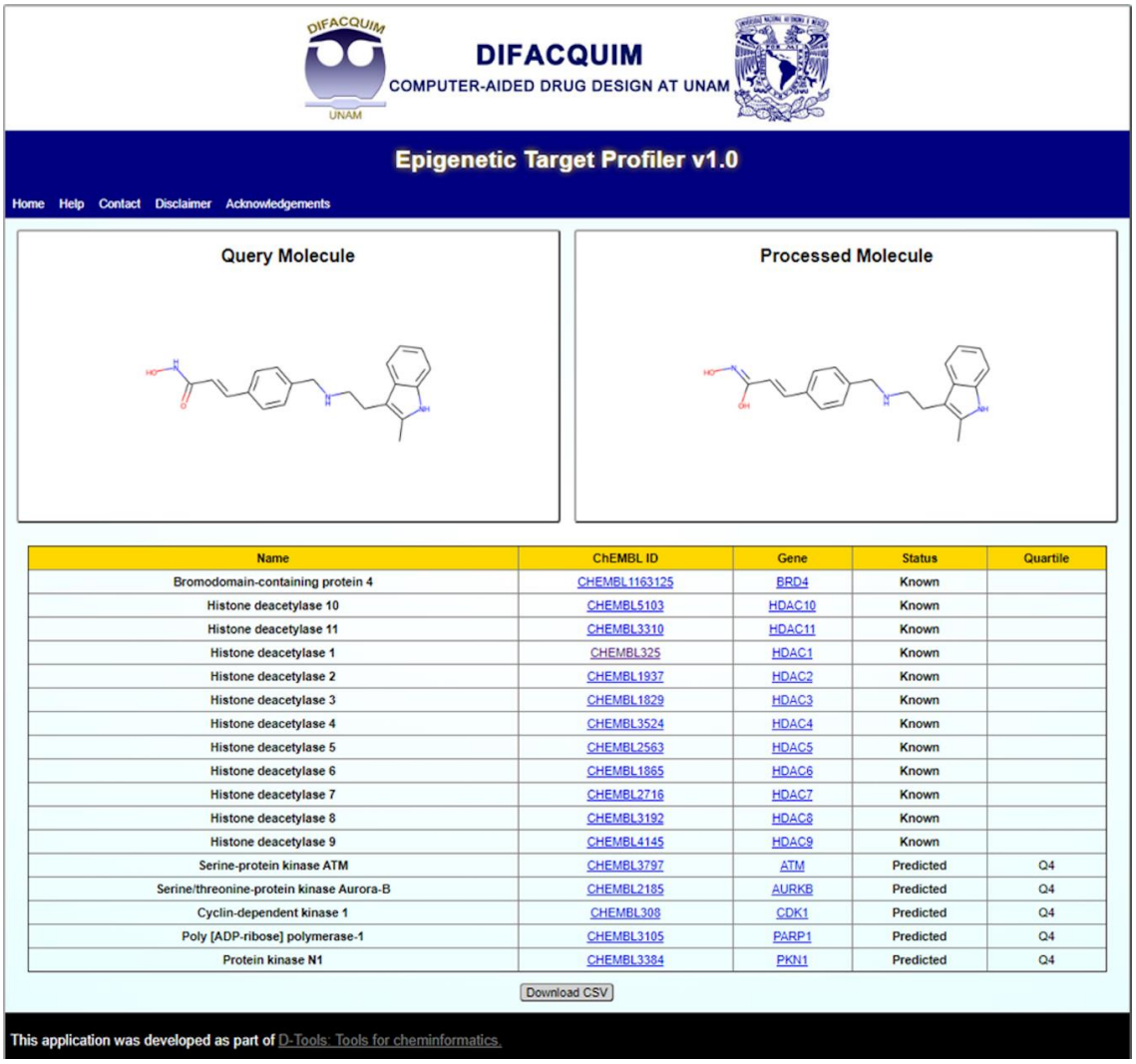

Figure S2. *Results page of Epigenetic Target Profiler. Known and predicted associations are shown for Panobinostat.*

4. Conclusion

ETP is an easy-to-use and free web-based tool to support epigenetic drug discovery projects.

Acknowledgements

NS-C is thankful to Consejo Nacional de Ciencia y Tecnología (CONACyT), Mexico, for the granted scholarship number 335997.

Funding

This work has been partially supported by the *Dirección General de Cómputo y de Tecnologías de Información y Comunicación* (DGTIC), UNAM that provided computational resources to use Miztli supercomputer with the project LANCAD-UNAM-DGTIC-335. The study was also supported by the *Consejo Nacional de Ciencia y Tecnología* (CONACyT), Mexico, grant 282785.

Conflict of Interest: none declared.

References

Anighoro,A. *et al.* (2014) Polypharmacology: Challenges and Opportunities in Drug Discovery. *J. Med. Chem.*, **57**, 7874–7887.

Bienfait,B. and Ertl,P. (2013) JSME: a free molecule editor in JavaScript. *J. Cheminform.*, **5**, 24.

Chen,J. *et al.* (2020) Discovery of selective HDAC/BRD4 dual inhibitors as epigenetic probes. *Eur. J. Med. Chem.*, **209**, 112868.

Oprea,T.I. *et al.* (2011) Drug repurposing from an academic perspective. *Drug Discov. Today Ther. Strateg.*, **8**, 61–69.

Sam,E. and Athri,P. (2019) Web-based drug repurposing tools: A survey. *Brief. Bioinform.*, **20**, 1–18.

Sessions,Z. *et al.* (2020) Recent progress on cheminformatics approaches to epigenetic drug discovery. *Drug Discov. Today*, **25**, 2268–2276.

Wilson,J.E. *et al.* (2020) Discovery of CPI-1612: A Potent, Selective, and Orally Bioavailable EP300/CBP Histone Acetyltransferase Inhibitor. *ACS Med. Chem. Lett.*, **11**, 1324–1329.

Zdrazil,B. *et al.* (2020) Moving targets in drug discovery. *Sci. Rep.*, **10**, 20213.