Computational method to predict drug pathway profiles

The approach was based on the Shannon-Entropy Descriptor (SHED) concept. In the conventional SHED approach (Gregori-Puigjané and Mestres, 2006) the chemical structure is converted into a 2D topological graph in which the nodes correspond to the atoms in the drug and edges connecting two nodes indicate the existence of a chemical bond. From this graph the shortest path length between each and every residue pair (characterized by their atom centered features) is calculated and stored as a function of feature pair. The Shannon-entropy quantifies the variability of feature-pair distributions in the molecule. Here, we have considerably extended the list of atom-centered features: (Lipophylic (L), positively charged (P), negatively charged (N), Hydrogen-Bond Donor (D), Hydrogen-Bond Acceptor (A) and simultaneous Hydrogen-Bond Donor/Acceptor (AD). A given atom is thus described via a 6-letter string consisting of (0,1), where 1 indicates the presence of a given feature. For example, a sp3-carbon atom is given by (100000), while a carboxylic oxygen reads as (001010) indicating the negative charge and the hydrogen bond acceptor capability. In total 25 feature pairs are used. Drug information is stored as a 1D Shannon entropy vector and thus allows for large-scale applications. Pairwise drug similarities were quantified by calculating the Euclidean distance. To obtain the biochemical pathway profile of a query molecule we screened the DrugBank database (Wishart et al., 2008), a repository of approved drugs and their experimentally verified protein targets. In a first step drug analogs from the DrugBank database are identified using the Shannon entropy vector (euclidean similarity cutoff: 0.25). Next, the experimentally verified protein targets for the identified DrugBank analogs are used to derive information about the

involved KEGG biochemical pathways. KEGG pathways are quantified by enumerating how often they are found in the list of identified DrugBank analogs. The numbers are normalized to the number of the most prevalent KEGG pathway (typically metabolic pathway or neuroactive ligand-receptor interaction pathway). Finally, the normalized pathway numbers are referenced (difference) to the statistical abundance of KEGG pathway obtained for the entire DrugBank dataset. The statistical abundance of the KEGG pathways was obtained by predicting the KEGG pathway profile (as described above) for all of the drugs in the DrugBank database.

The resulting ranked profile signatures were then used to score corresponding genes of azelastine-HCL and hydroxychloroquine. These genes were identified and described for homo sapiens using the HGNC Database, HUGO Gene Nomenclature Committee (HGNC) of the European Bioinformatics Institute (EMBL-EBI) in the biomaRt package (Durinck et al., 2009). For azelastine-HCL 201 and for hydroxychloroquine 185 genes were identified with an overlap of 129 genes. For azelastine-HCL, 72 genes were exclusively identified and 56 genes were exclusive for hydroxychloroquine (Supplementary Material 4).

Predicted pathway profiles of drugs active against SARS-CoV and/or SARS-CoV-2

We employed pathway information we predicted for drugs shown to be active against SARS-CoV and/or SARS-CoV-2. Three experimentally verified and characterized compounds, hydroxychloroquine (Yao et al., 2020), SSAA09E2 {N-

[[4-(4-methylpiperazin-1-yl)phenyl]methyl]-1,2-oxazole-5-carboxamide} and SSAA09E3 {N-(9,10-dioxo-9,10-dihydroanthracen-2-yl)benzamide} (Adedeji et al., 2013) were employed. Hydroxychloroquine reduces endosomal acidification, SSAA09E2 acts by blocking early interactions of SARS-CoV with its receptor, the angiotensin converting enzyme 2 (ACE2), shared by SARS-CoV-2 and SSAA09E3 prevents fusion of the viral membrane with the host cellular membrane. For all three selected ligands, the pathway profiles were calculated, and the 50 highest scoring pathways were considered for the analysis

References

Adedeji, A. O., Severson, W., Jonsson, C., Singh, K., Weiss, S. R., and Sarafianos, S. G. (2013). Novel Inhibitors of Severe Acute Respiratory Syndrome Coronavirus Entry That Act by Three Distinct Mechanisms. *Journal of Virology* 87, 8017–8028. doi:10.1128/jvi.00998-13.

Durinck, S., Spellman, P. T., Birney, E., and Huber, W. (2009). Mapping identifiers for the integration of genomic datasets with the R/ Bioconductor package biomaRt. *Nature Protocols* 4, 1184–1191. doi:10.1038/nprot.2009.97.

Gregori-Puigjané, E., and Mestres, J. (2006). SHED: Shannon entropy descriptors from topological feature distributions. *Journal of Chemical Information and Modeling* 46, 1615–1622. doi:10.1021/ci0600509.

Wishart, D. S., Knox, C., Guo, A. C., Cheng, D., Shrivastava, S., Tzur, D., et al. (2008). DrugBank: A knowledgebase for drugs, drug actions and drug targets. *Nucleic Acids Research* 36, D901. doi:10.1093/nar/gkm958.

Yao, X., Ye, F., Zhang, M., Cui, C., Huang, B., Niu, P., et al. (2020). In Vitro Antiviral Activity and Projection of Optimized Dosing Design of Hydroxychloroquine for the Treatment of Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2). *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America* 71, 732–739. doi:10.1093/cid/ciaa237.