C-3-07

PoSSuM Updates and Integration With ChEMBL For Application of Drug Reuse

Kazuyoshi Ikeda[1,2,3] ikeda@level-five.jp
Jun-Ichi Ito[2] junichi@nibio.go.jp
Kenji Mizuguchi[2] kenji@nibio.go.jp
Kentaro Tomii[3] k-tomii@aist.go.jp

[1] Drug Discovery Informatics Group, System Solution Division, LEVEL FIVE Co.,Ltd., Shiodome Shibarikyu Bldg., 1-2-3 Kaigan, Minato-ku, Tokyo 105-0022, Japan.
[2] National Institute of Biomedical Innovation, 7-6-8 Saito Asagi, Ibaraki, Osaka, Japan.
[3] Computational Biology Research Center (CBRC), National Institute of Advanced Industrial Science and Technology (AIST), 2-4-7 Aomi, Koto-ku, Tokyo 135-0064, Japan.

Keywords: Ligand binding site, Pocket similarity, Open drug discovery data, Drug reuse, Database

PoSSuM is a large-scale database for finding similar ligand-binding sites in proteins [1], which is freely available via our web-interface (http://possum.cbrc.jp/PoSSuM). As of Aug. 2013, PoSSuM stores over four millions of data points including both known small molecule-binding pockets and putative ones. Detection of structurally similar sites of known small-molecule drugs gives us useful information for rational drug design such as drug-reuse and off-target prediction. For the purpose, we have integrated PoSSuM pocket similarity data with ChEMBL open drug discovery data [2], and have developed a new function for supporting the user to retrieve potential binding pockets for a given small-molecule drug, in addition to a number of improved web interfaces.

First, we selected orally bioavailable small-molecule drugs from the release of ChEMBL16, and matched the compound structures with the latest PDB ligand structures using IUPAC standard InChIKey. In PoSSuM, 4,282 sites where 193 unique small-molecule drugs bound were found. Second, we annotated 3D protein-structures in PDB using the ChEMBL target information for drugs and drug candidates, and classified the protein structures based on its drug development stage. For this purpose, we extracted 7,547 protein-targets from ChEMBL16 and joined those to the FDA-approved drug targets from DrugBank [3]. The protein-targets were classified into the following classes: A) Approved drug targets, B) Hit-to-lead stage, C) Bioactive targets and D) Unknown. Then, we combined the drug target classification with PoSSuM pocket similarity data, and built a pipeline for automatically identifying potential new drug targets having a pocket similar to known small-molecule drug binding sites.

We identified in total 154,191 binding sites, from 10,255 protein structures, which are structurally similar to the focused 4,282 binding sites, covering 182 unique small-molecule drugs deposited in ChEMBL16. Users can retrieve all of the identified similar sites and can visualize 3D superpositoin between them on our new web-interface which will be available soon. We also conducted a preliminary study to compare between PDB-scale pocket similarity and genomic-scale bioactive data for surveying relations between the similarity score and binding affinities.

In this presentation, we will show new functions of the PoSSuM database, utility for structure-based drug design, and discuss the possibility and limitation for application of drug discovery purpose.

[1] Ito, J., *et al*., *PoSSuM: a database of similar protein-ligand binding and putative pockets*, *Nucleic Acids Res*, **40**(Database issue):D541-D548. 2012.
[2] Gaulton, A., *et al*., *ChEMBL: a large-scale bioactivity database for drug discovery*, *Nucleic Acids Res*, **40**(Database issue):D1100–D1107. 2012.
[3] Knox C., *et al*., *DrugBank 3.0: a comprehensive resource for 'omics' research on drugs*, *Nucleic Acids Res*, **39**(Database issue):D1035-D1041. 2011.