Supplementary Material for Unveiling new biological relationships using shared hits of chemical screening assay pairs

Xueping Liu1,2 and Monica Campillos1,2,*
1Institute of Bioinformatics and Systems Biology and 2German Center for Diabetes Research, Helmholtz Center Munich, 85764, Neuherberg, Germany.

Received on XXXXX; revised on XXXXX; accepted on XXXXX

Associate Editor: XXXXXXX

1 METHODS

1.1 Calculation of hit similarity

We represented the hit profile of an assay using a binary fingerprint of the chemical activity of all compounds (“1” indicates that the compound is a hit; “0” indicates that compound has been tested, but inactive. Missing value signifies that the compound has not been tested).

To calculate the hit profile similarity between two assays we used the continuous Tanimoto coefficient (Tc) (Willett et al., 1998) equation (equation (1)),

\[
T_c = \frac{\sum_{i=1}^{n} (w_i x_i)(w_i y_i)}{\sum_{i=1}^{n} (w_i x_i)^2 + \sum_{i=1}^{n} (w_i y_i)^2 - \sum_{i=1}^{n} (w_i x_i)(w_i y_i)}
\]

Where \( n \) is the total number of compounds tested in both assays, \( I \) iterates over all compounds, \( w_i \) is the promiscuity (ratio of the number of assays where the compound is active and the number of assays where the compound was tested) of the compound at ChemBank database level including 1,640 assays. \( x_i, y_i \) are the activity values (1 or 0) of the compound.

1.2 Similarity of assay projects by applying EXCERBT

EXtraction of Classified Entities and Relations from Biomedical Texts (EXCERBT) (Mewes et al., 2011) is a free-to-use biomedical text-mining system based on all available abstracts from PubMed (http://www.ncbi.nlm.nih.gov/pubmed/), full-text articles from PubMed Central (http://www.ncbi.nlm.nih.gov/pmc/) and articles in OMIM (http://www.ncbi.nlm.nih.gov/omim). As EXCERBT is case-sensitive we searched for different combinations of lower and upper case of keywords available in EXCERBT. For example, for the keyword “Wnt”, EXCERBT provides different forms such as “wnt”, “WNT” and “Wnt”. For “histone deacetylase”, the terms “HDAC”, “HISTONE DEACETYLASE”, “Histone deacetylase”, and “Histone Deacetylase” are available. Furthermore, the search result depends on the order of two keywords. For instance, there are four evidences linking “HDAC” and “Wnt” and only two for the search of “Wnt” and “HDAC”. Thus, we searched every combination of keywords of two projects and selected the search that retrieved the highest number of occurrences.

1.3 Manual annotation for projects

Out of the possible GO terms describing the biological activity (i.e. molecular activity GO term for target-based assays and biological process GO term for phenotypic assays) measured in a project, we selected the most specific one. We assigned one GO term to those projects measuring single molecular activity or biological process and several GO terms to those projects where several biological processes were measured.

For instance, in the project “Wnt And Lithium Modulators” two biological processes are measured, namely, the modulation of Wnt pathway and of Lithium-mediated signal transduction. Consistently, we assigned two GO terms to this project: “GO:0016055: Wnt receptor signaling pathway” and “GO:0071285 cellular response to lithium ion”. In contrast, in the project of “Wnt Inhibitors” only chemical modulators of one biological processes (Wnt pathway) are sought, therefore only one GO term “GO:0016055: Wnt receptor signaling pathway” was assigned to it.

For the annotation of projects using keywords, we selected the most specific keywords (from EXCERBT database) describing the biological activity measured in the projects and assigned them to the projects. Due to the existence of synonyms keywords, normally, more keywords than GO terms are needed for characterizing a project. For instance, “cell growth” and “cellular growth” were the keywords assigned to the project “Bacterial Viability”, while only “GO:0016049 cell growth” was assigned to this project.

The number of manual annotation terms for projects is listed in Table S1.

REFERENCES

Mewes,H.W. et al. (2011) MIPS: curated databases and comprehensive secondary data resources in 2010. Nucleic Acids Res., 39, D220–224.
Willett,P. et al. (1998) Chemical Similarity Searching. J. Chem. Inf. Comput. Sci., 38, 983–996.

SUPPLEMENTARY FIGURES AND TABLES

Supplementary Figures S1-S3 and Tables S1-S2 are located in the end of this document.
Chemical hit similarity comparison of the assay pairs after application of filter F0 and F1. Hit similarity was calculated by continuous Tc. We randomly chose 10,000 assay pairs and compared the hit similarity after the filter F0 and F1. F0 contains all the initial compounds; F1 includes the compounds that are active in more than one project.

**Figure S2** Distribution of cell-free, cell-based and microorganism assays for projects.
Figure S3 Different cut-offs on the number of shared hits in F3 of ChemBank assay pairs. Hit similarity was calculated by continuous Tc. (a) Relationships indicated by GO terms and (b) relationships indicated by text-mining.
Table S1 Number of manual annotation terms for projects

| Gene Ontology (GO) terms annotation | Keywords for text-mining annotation |
|-------------------------------------|-----------------------------------|
| Number of GO terms | Number of projects | Number of keywords | Number of projects |
| 1 | 168 | 1 | 79 |
| 2 | 5 | 2 | 68 |
| 3 | 3 | 3 | 19 |
| 4 | 5 | 4 | 6 |
| 7 | 1 | 5 | 2 |
| Unknown | 8 | 11 | 1 |
| Unknown | 15 | | |
| M.Tuberculosis Macrophage | GSI Synthetic Lethal | Shared 7 chemical hits |
|--------------------------|----------------------|------------------------|
|                          | GSI Synthetic Lethal |                        |
|                          | Adipocyte Differentiation | Share 8 chemical hits |
|                          | GSI Synthetic Lethal |                        |
|                          | Unfolded Protein Response | Share 6 chemical hits |
|                          | Wnt Inhibitors | Share 14 chemical hits |
|                          | Histone Modification | Stem Cell Differentiation | Share 15 chemical hits |
|                          | Genotype Specific Inhibitors NSCLC | Share 10 chemical hits |
|                          | Histone Modification | Stem Cell Differentiation | Share 9 chemical hits |
|                          | Genotype Specific Inhibitors NSCLC | Share 195 chemical hits |
|                          | Glioblastoma Modulators |               |

- **Table S2 Shared hits with predicted targets between two assays**

Below each assay combination, the shared ChemBank compound IDs along with HitPick predicted targets (precision >50%) are shown. The other hits without target information are not shown.

| M.Tuberculosis Macrophage | GSI Synthetic Lethal | Shared 7 chemical hits |
|--------------------------|----------------------|------------------------|
|                          | GSI Synthetic Lethal |                        |
|                          | Adipocyte Differentiation | Share 8 chemical hits |
|                          | GSI Synthetic Lethal |                        |
|                          | Unfolded Protein Response | Share 6 chemical hits |
|                          | Wnt Inhibitors | Share 14 chemical hits |
|                          | Histone Modification | Stem Cell Differentiation | Share 15 chemical hits |
|                          | Genotype Specific Inhibitors NSCLC | Share 10 chemical hits |
|                          | Histone Modification | Stem Cell Differentiation | Share 9 chemical hits |
|                          | Genotype Specific Inhibitors NSCLC | Share 195 chemical hits |
|                          | Glioblastoma Modulators |               |

- **Table S2 Shared hits with predicted targets between two assays**

Below each assay combination, the shared ChemBank compound IDs along with HitPick predicted targets (precision >50%) are shown. The other hits without target information are not shown.

| M.Tuberculosis Macrophage | GSI Synthetic Lethal | Shared 7 chemical hits |
|--------------------------|----------------------|------------------------|
|                          | GSI Synthetic Lethal |                        |
|                          | Adipocyte Differentiation | Share 8 chemical hits |
|                          | GSI Synthetic Lethal |                        |
|                          | Unfolded Protein Response | Share 6 chemical hits |
|                          | Wnt Inhibitors | Share 14 chemical hits |
|                          | Histone Modification | Stem Cell Differentiation | Share 15 chemical hits |
|                          | Genotype Specific Inhibitors NSCLC | Share 10 chemical hits |
|                          | Histone Modification | Stem Cell Differentiation | Share 9 chemical hits |
|                          | Genotype Specific Inhibitors NSCLC | Share 195 chemical hits |
|                          | Glioblastoma Modulators |               |
### Glioblastoma Modulators

| Wnt Inhibitors | Glia Inhibitors | Share 114 chemical hits |
|----------------|----------------|-------------------------|
| 1054556        | CA1;CA2;       |                         |
| 1134906        | ADORA3;        |                         |
| 1241281        | ALDOA;         |                         |
| 1247906        | HTR6;          |                         |
| 1393431        | CNR2;          |                         |

### Wnt And Lithium Modulators

| E-Cadherin Synthetic Lethal | Share 40 chemical hits |
|-----------------------------|-----------------------|
| 1021994 ESRGR               |                       |
| 1111942 SIGMAR1             |                       |
| 1112646 GRM5                |                       |
| 1118692 AVPR2               |                       |
| 1227426 PREP                 |                       |
| 1285366 SIGMAR1             |                       |
| 1311944 HRH3                |                       |
| 1446378 SIGMAR1             |                       |
| 1570064 F2                  |                       |
| 1733044 SCD                 |                       |
| 3065778 HSD11B1             |                       |
| 3178349 SCN10A              |                       |
| 3554525 FNTA                |                       |
| 3558623 NR1H4               |                       |
| 3558697 FNTA                |                       |
| 3558699 FNTA                |                       |
| 3559706 CYP1B1              |                       |
| 3559708 ATP1A1;SLCO1A2;SLC O4C1;CYP11A1; |   |
| 3557708 ATP1A1;HTR2C;       |                       |
| 3557710 ATP1A1;             |                       |
| 3557710 ATP1A1;             |                       |
| 3557735 ATP1A1;             |                       |
| 3557816 ATP2A1;             |                       |
| 3557837 ATP1A1;             |                       |
| 3558638 NR1H4;              |                       |
| 3558693 FNTA;               |                       |
| 3558697 FNTA;               |                       |
| 3558877 FNTA;               |                       |
| 3559060 MT-ND4;             |                       |
| 3559282 SLC5A2;             |                       |
| 3559706 CYP1B1;             |                       |
| 3559712 FNTA;               |                       |
| 3559755 UGT2B7;             |                       |
| 3559863 SLC18A2;            |                       |
| 3614483 DRD2;DRD4;          |                       |
| 3615241 MBL2;               |                       |
| 3616405 ATP1A1;             |                       |
| 3622930 HDAC1;              |                       |
| 3624592 CNR2;               |                       |
| 3625232 CETP;               |                       |
| 3625266 MDM2;               |                       |
| 3625359 FLI3;               |                       |
| 3625448 CNR2;               |                       |
| 3625502 HTR6;               |                       |
| 3625991 ADORA2A;            |                       |
| 3635064 HDAC6;HDAC1;        |                       |
| 3635080 HDAC3;HDAC2;HDAC6   |                       |
| 3635093 HDAC6;HDAC1;        |                       |
| 3635098 HDAC6;HDAC1;        |                       |
| 3644448 DRD4;               |                       |
| 3652467 CCR4;               |                       |
| 3652509 HDAC6;              |                       |
| 3652551 HDAC6;HDAC1;        |                       |
| 3652574 HTR2C;              |                       |
| 3544 DRD4;HTR7;HTR2A;SI     |                       |
| 2082296 CYP2D6;ORM1;KCNH2   |                       |
| :CHRM2;                   |                       |

### Beta-Catenin

| Histone Modification | Share 8 chemical hits |
|----------------------|-----------------------|
| 3214216 HDAC1;       |                       |
| 3214224 HDAC3;       |                       |
| 3214248 HDAC1;       |                       |
| 3214452 HDAC1;       |                       |

### Histone Modification

| Wnt And Lithium Modulators | E-Cadherin Synthetic Lethal | Share 40 chemical hits |
|----------------------------|-----------------------------|-----------------------|
| 3178349 SCN10A             |                            |                       |
| 3554525 FNTA               |                            |                       |
| 3558623 NR1H4              |                            |                       |
| 3558697 FNTA               |                            |                       |
| 3558699 FNTA               |                            |                       |
| 3559706 CYP1B1             |                            |                       |
| 3615060 TUBB1;TUBA4A;      |                            |                       |
| 3615241 MBL2;             |                            |                       |
| 3616341 KCNA3;            |                            |                       |
| 3616405 ATP1A1;           |                            |                       |
| 3616405 ATP1A1;           |                            |                       |
| 3622930 HDAC1;            |                            |                       |
| 3624592 CNR2;             |                            |                       |
| 3625232 CETP;             |                            |                       |
| 3625266 MDM2;             |                            |                       |
| 3625359 FLI3;             |                            |                       |
| 3625448 CNR2;             |                            |                       |
| 3625502 HTR6;             |                            |                       |
| 3625991 ADORA2A;          |                            |                       |
| 3635064 HDAC6;HDAC1;      |                            |                       |
| 3635080 HDAC3;HDAC2;HDAC6 |                            |                       |
| 3635093 HDAC6;HDAC1;      |                            |                       |
| 3635098 HDAC6;HDAC1;      |                            |                       |
| 3644448 DRD4;             |                            |                       |
| 3652467 CCR4;             |                            |                       |
| 3652509 HDAC6;            |                            |                       |
| 3652551 HDAC6;HDAC1;      |                            |                       |
| 3652574 HTR2C;            |                            |                       |
| 3544 DRD4;HTR7;HTR2A;SI   |                            |                       |
| 2082296 CYP2D6;ORM1;KCNH2 |                            |                       |
| :CHRM2;                  |                            |                       |
| Compound ID | Target |
|------------|--------|
| 3634663    | HDAC1  |
| 3634669    | HDAC1  |
| 3634737    | HDAC1  |