DrugSig: A resource for computational drug repositioning utilizing gene expression signatures

Hongyu Wu1,2*, Jinjiang Huang1*, Yang Zhong1, Qingshan Huang1,2*

1 School of Life Sciences, Fudan University, Shanghai, China, 2 Shanghai High-Tech United Biotechnological R&D Co., Ltd., Shanghai, China

* These authors contributed equally to this work.
* qshuang@fudan.edu.cn

Abstract

Computational drug repositioning has been proved as an effective approach to develop new drug uses. However, currently existing strategies strongly rely on drug response gene signatures which scattered in separated or individual experimental data, and resulted in low efficient outputs. So, a fully drug response gene signatures database will be very helpful to these methods. We collected drug response microarray data and annotated related drug and targets information from public databases and scientific literature. By selecting top 500 up-regulated and down-regulated genes as drug signatures, we manually established the DrugSig database. Currently DrugSig contains more than 1300 drugs, 7000 microarray and 800 targets. Moreover, we developed the signature based and target based functions to aid drug repositioning. The constructed database can serve as a resource to quicken computational drug repositioning. Database URL: http://biotechlab.fudan.edu.cn/database/drugsig/.

Introduction

Over the past decades, to develop a de novo drug often takes billions of dollars and about 9–12 years [1]. New drug discovery has grown to be time-consuming and costly. This directly resulted in small quantity and high price of new drugs on the market. Drug repositioning, by exploring new clinical indications for those existing drugs has become an increasingly important strategy for drug development resulted from their proved drug safety and the abridged process of drug discovery and preparation [1–9]. However, traditional drug repurposing is mostly through serendipity or explored from a better understanding of the drugs’ mechanism of action. The efficacy of these methods is very low. When the drug-related and genome-wide data initiatives grew quickly, the mode for computational drug repositioning has been changed.

By integrating data from various sources, like pharmacological, genetic, chemical or clinical data, a set of new computational repositioning strategies and techniques has emerged [3,10,11]. Especially, the Connectivity Map (CMap) [12,13] project which produced large-scale drug response gene expression profiles lead to the establishment and development of methods
of ‘gulit-by-association’ and ‘signature reversion’ for computational drug repurposing [14].
With these methods, Sirota et al. and Dudley et al. had found that an antiulcer drug and an
antiepileptic drug can be reused for lung cancer and inflammatory bowel disease by comparing
each of these disease signatures to each of the gene expression signatures for 164 drugs from
CMap [15,16]. Obviously, the quantity and quality of drug response gene signatures is the core
for these computational approaches. But these data still scattered in separated or individual
experimental data, it brought about low efficient outputs. So, a database archiving enough
drug response gene signatures will be very helpful to computational drug repurposing.

Based on above observation, we collected most of drug response microarray data from
GEO or scattered in the separated database to develop the DrugSig database. Moreover, we
manually inspected targets information for each drug extracted from microarray data and
archived them into DrugSig. Finally, we implemented two functions for repositioning old
drugs using signature or target based drug repositioning method respectively. The constructed
database will serve as a resource to quicken computational drug repositioning.

Results and discussion

Database description

DrugSig was created as a resource for computational drug repositioning utilizing gene expres-
sion signatures. As a web based database, DrugSig provides a user-friendly web interface for
users to easily query and retrieve information on drug signatures. All the data in DrugSig can
be accessed and retrieved directly from the web browser. Fig 1 describes the schema of the cre-
ation of DrugSig. All raw data were manually collected from literate and public databases. We
processed these microarrays, drugs, targets and literate data into seven Mysql tables, such as
drugs, instances, drugsig, platform, targets, drug_target and papers table. On the basis of these
data, we developed tools for signature based and target based computational drug repurposing
functions and recorded the computational history into uses table.

DrugSig web interface

A concise navigational interface comprised of the Home, Browse, Search, Tools and Guide
options was designed to generate a clearly structured database layout that enables fast and easy
navigation (Fig 2A). The Browse interface allows users to navigate all drugs included in Drug-
Sig. The current database is composed of more than 1300 drugs. A click of each drug will dis-
play a results page with four sections: drug summary, drug signature, drug targets and links.
(Fig 2B). Drug summary section consists of drug name, chemical name, formula, CAS no,
description and drug indications. In addition, it also provides a link to DrugBank [17] for fur-
ther investigation. Drug signature section demonstrates its common signatures which are
comprised of top 50 up-regulated and down-regulated genes and its data source (list all related
microarray). For each microarray, there is a page to display its signatures. The drug targets sec-
tion consists of the drug targets and their expression value in cells treated by the drug and
other drugs. The prior expression level reflects the expression of the target gene response to
the drug while the latter expression level shows the potential of other drugs which inhibit or
stimulate the target. The Search interface can be used to retrieve specific information using
either a quick or advanced option (Fig 2C). A quick search only allows keywords field, while
the advanced search accepts the specification of up to six separate fields: Drug Id, Drug Name,
DrugBank Id, Disease, Target, and Signature Symbol. The user can query the database by
either one particular condition or a combination of various conditions. The prior five fields
search produces a results page which list all drugs meet the specified conditions. The final field
search produces a list of the gene expression of specified signature symbol if its expression
level lies in top 500 up-regulated or down-regulated genes. The Tools interface implements the signature based and target based drug repositioning functions (see ’Drug repurposing tools implemented in DrugSig’ section below for further details). The Guide interface provides detail instructions to potential users on how to use DrugSig.

**Drug repurposing tools implemented in DrugSig**

Tools of drug repurposing implemented in DrugSig consist of signature based and target based drug repositioning functions. The signature based drug repositioning function provides an interface to input user’s gene list to compute against DrugSig (Fig 3A). After submitting the gene list to DrugSig, user can click the start computing button to compute the scores which is the ratio of the number of common genes between user’s gene list and each gene signature to the number of user’s genes (Fig 3B). Once the computing finished, DrugSig will sent a notice email to user. The results can be accessed later from the email or by searching the task history with user’s email address. The computing result contains queried gene list, top 50 score drugs produced reverse gene list and top 50 score drugs produced similar gene list (Fig 4A). The
Fig 2. Overview of DrugSig. (A) Browsing the entire database by drugs. (B) A snapshot of the page of drug detail. (C) Searching the database by six options.

https://doi.org/10.1371/journal.pone.0177743.g002
reverse and similar drugs infer to potential indications. Each drug produces a page display the reverse gene list and similar gene list for further investigation (Fig 4B).

The target based drug repositioning function provides an interface to explore the specified target (Fig 5A) and its targeting drugs, as well as target gene expression level in cells after treated by drugs (Fig 5B). The targeting drugs may have similar indications for drug repurposing. The expression level of target gene partly infers the potential of the drug which either inhibits or enhances the target.

---

**Fig 3.** A case study for how to use signature based drug repositioning function. (A) The input interface. (B) The computing interface.

https://doi.org/10.1371/journal.pone.0177743.g003
### Queried gene list

**Up list (30)**
- PLN
- UGT2A3
- AQP4
- HOXB9
- ARJS
- LAMB4
- SSX4B
- ARS1
- MS4A4A
- CDS1
- PDE4DIP
- ABO
- RAG2
- WNT2B
- MST1L
- TNFRSF4
- LC3B
- CLCA3
- PDLIM5
- RPL15
- GP5
- SCHL12
- NF2
- PIN1
- PFKFB2
- ABCA8
- RG59
- BMP2K
- NPB2
- IL26
- CDR1

**Down list (30)**
- PRKD1
- ZNF407
- UTP14A
- ATRNL1
- SRI
- SLC4A7
- PLXNC1
- CTHZ
- ZEB2
- C8orf4
- LRR2B
- TNFRSF11B
- IGK
- IGKC
- UHRF1BP1
- LKTLG
- DPT
- NOVA2
- ZNF135
- FRZB
- SLC17A6
- CR1
- EPB41L3
- IGJ
- IGHV4-31
- DBT
- CDK11A
- C1
- TMEM47
- ZNF711
- RARRE1
- S1
- ALG13

### Drugs produced reverse gene list

| Drug Name     | Array Id               | Up Score | Down Score | Total Score |
|---------------|------------------------|----------|------------|-------------|
| furosemide    | DS01B000622A01580      | 0.4(12/30)| 0.13(4/30) | 0.27(16/60) |
| chlorpropamide| DS01B000622A00144      | 0.23(7/30)| 0.27(8/30) | 0.25(15/60) |
| phenyl biguanide | DS01B0002400022     | 0.3(8/30) | 0.17(5/30) | 0.23(14/60) |
| rosiglitazone | DS01B000555A00430     | 0.27(8/30)| 0.2(6/30)  | 0.23(14/60) |
| valproic acid | DS01B00070A00467      | 0.33(10/30)| 0.13(4/30) | 0.23(14/60) |
| colchicine    | DS01B00112A00644      | 0.23(7/30)| 0.23(7/30) | 0.23(14/60) |

### Drugs produced similar gene list

| Drug Name | Array Id               | Up Score | Down Score | Total Score |
|-----------|------------------------|----------|------------|-------------|
| metformin | DS01B00001A00001       | 1(30/30) | 1(30/30)   | 1(60/60)    |
| metformin | DS01B00011A00002       | 0.57(17/30)| 0.57(17/30)| 0.57(34/60) |
| metformin | DS01B00011A00003       | 0.67(20/30)| 0.33(10/30)| 0.50(30/60) |
| metformin | DS01B00011A00004       | 0.67(20/30)| 0.33(10/30)| 0.50(30/60) |
| prinixic acid | DS01B00067A00487 | 0.3(9/30) | 0.23(7/30) | 0.27(16/60) |
| L-294002  | DS01B00071A00501       | 0.3(9/30) | 0.23(7/30) | 0.27(16/60) |

### Reversed gene list

**Up list VS Down-regulated genes of DrugSig(DS01B000622A01580)**

| Symbol | Gene Title          | Chromosome Location |
|--------|---------------------|---------------------|
| 1      | AQP4                | 18q11.2-q1          |
| 2      | LAMB4               | 7q31                |
| 3      | MS4A4A              | 11q12               |
| 4      | PDE4DIP             | 1q12                |
| 5      | RAG2                | 11p13               |

**Similar gene list**

**Up list VS Up-regulated genes of DrugSig(DS01B000622A01580)**

| Symbol | Gene Title          | Chromosome Location |
|--------|---------------------|---------------------|
| 1      | PLN                 | 6q22.1              |
| 2      | AQP4                | 18q11.2-q1          |
| 3      | PDE4DIP             | 1q12                |

**Down list VS Down-regulated genes of DrugSig(DS01B000622A01580)**

| Symbol | Gene Title          | Chromosome Location |
|--------|---------------------|---------------------|
| 1      | TNFRSF11B           | 8q24                |
| 2      | KITLG               | 12q22               |

---

**Fig 4. Results of signature based drug repositioning.** (A) The drug list for signature based drug repositioning. (B) The gene list for each calculated drug.

https://doi.org/10.1371/journal.pone.0177743.g004
Discussion

The purpose of establishing the DrugSig database is to aid drug repurposing. DrugSig server as not only a tool to reposition old drugs with user’s input but also an open resource for users to develop new computational approaches for drug repositioning. The current DrugSig contains more than 1300 drugs, 7000 microarray and 800 targets. DrugSig is different from the existing webservers or databases for drug repositioning. Although several drug repositioning related webservers or databases exist such as CMap, PREDICT [18], PROMISCUOUS [19], INDI [20] and Mantra [21], each has certain shortcomings, such as covering a limited collection of drug response microarray or only containing a computational framework. These shortcomings limit the accurate and scope of computational drug repurposing also increase the difficulties in using these data for scientists. Although main data in DrugSig had also been collected from CMap, DrugSig covered pertinent data from other experiments. Moreover, with pertinent experiments growth, DrugSig will contain more and more drugs and signatures.
Moreover, many projects focus on precision medicine will disclose insights between disease and genes. Recently, Rubio-Perez et al. developed an in silico drug prescription strategy based on driver alterations in each tumor and their druggability options and use it to identify druggable targets and promising repurposing opportunities [22]. When applied these insights on DrugSig, we can promptly verify these druggable targets and promising repurposing opportunities. So we constructed a gene list from TCGA Esophageal carcinoma (ESCA) data and submitted it to compute against DrugSig. Results showed that the top compounds predicted to be therapeutic for ESCA were acetylsalicylic acid, an anti-inflammatory drug had been reported to treat ESCA, and dizocilene, an antiepileptic drug not previously described to have efficacy for ESCA. Related validation work is in progress.

**Limitations and future prospects**

Currently DrugSig holds only 1300 drugs and 6000 plus signatures. Moreover, the functions and methods implemented in database are limited. In the future, we plan to updates the data continuously per half a year, and integrate some gene function analysis tools and other computational drug repurposing approaches into DrugSig to improve its interactivity with users and to increase functions to aid computational drug repositioning. In addition, we plan to develop open services convenient for researchers to get gene signatures applied to develop new computational approaches for drug repositioning.

**Conclusions**

DrugSig is a web accessible database for computational drug repurposing studies. The current version of DrugSig includes more than 1300 drugs, 6000 plus signatures and 800 available targets (till Jan, 2017). The database can be queried either by simply using keywords or by combinatorial conditions searches. DrugSig will not only aid in expanding our current understanding of drugs and their mechanisms of action but may have implications in the development of new indications for existed drugs. DrugSig now is available at [http://biotechlab.fudan.edu.cn/database/drugsig/](http://biotechlab.fudan.edu.cn/database/drugsig/).

**Methods**

**Data acquisition and storage**

The microarray data in DrugSig were obtained from the GEO [23] databases or individual scientific researches. The steps of the curation of DrugSig contained collecting, processing and computing (Fig 1). We first searched the scientific literature which contains drug response microarray experiments form PubMed using keywords like “human cell AND treatment AND (‘gene signature’ OR ‘expression profile’) AND (genechip OR microarray OR ‘gene expression’) AND English [la]” and collected available drug response microarray data from GEO database or special sources described in scientific literature. Finally, we obtained more than 7000 microarray raw data. We then read the data via RMA method of affy package in BioConductor [24] and constructed the drug induced signatures using two approaches depending on the quantity of raw data. When the replicates $< 3$ in raw data we computed the drug signatures by simple fold changes (FC $> 2.0$ or FC $< 0.5$) and when the replicates $>= 3$ we computed the drug signatures by Limma package of BioConductor program (FC $> 2.0$ or FC $< 0.5$ and P value $< 0.01$) which implemented the linear models to calculate the differently expression genes. In addition, if the number of calculated differential expressed probes $< 500$, we selected all differential expressed probes as signatures. After abstracted more than 1300 drugs from the related experiments, we investigated the drug related information from several public...
databases such as DrugBank [17,25], KEGG [26], CTD [27] and TTD [28]. Finally, 800 plus available targets were constructed according to descriptions from the literature. All of the collected information and computed data had been classified and filled into seven relational tables in MySQL. Moreover, we constructed the up list and down list file from DrugSig table and used them to implement signature based drug repurposing.

Database architecture and web interface
DrugSig was built on a 64 bit Windows (2008 R2) server running WAMPSERVER (V2.2d), which integrates the Apache HTTP Server (V2.2.21) with PHP (V5.3.10) and the MySQL Server (V5.5.20). All entries are stored in a MySQL database. The web application was coded in PHP, using the jQuery JavaScript Library (V1.6.2), the Highchart jQuery Plugin, and cascading Style Sheets (CSS) for the web design. Apache, MySQL, PHP, R and jQuery were preferred as they are open-source software and platform independent, making them suitable for academic use. The web server and all parts of the database are hosted at the Information Office of Fudan University, Shanghai, China.

Acknowledgments
We would like to thank all of our colleagues at School of Life Sciences, Fudan University and at Shanghai High-Tech United Bio-Technological R&D Co., Ltd., of China for their contributions in the literature search and discussions regarding this manuscript.

Author Contributions
Conceptualization: HW QH.
Data curation: HW JH.
Formal analysis: HW.
Funding acquisition: QH.
Investigation: HW JH.
Methodology: HW JH.
Resources: YZ QH.
Software: HW.
Supervision: YZ QH.
Writing – original draft: HW JH.
Writing – review & editing: HW YZ QH.

References
1. Ashburn TT, Thor KB. (2004) Drug repurposing: identifying and developing new uses for existing drugs. Nat Rev Drug Discov 3: 673–683. https://doi.org/10.1038/nrd1468 PMID: 15286734
2. Li J, Zheng S, Chen B, Butte AJ, Swamidass SJ, Lu Z. (2015) A survey of current trends in computational drug repurposing. Brief Bioinform 17: 2–12. https://doi.org/10.1093/bib/bbv020 PMID: 25832646
3. Strittmatter SM. (2014) Overcoming Drug Development Bottlenecks With Repurposing: Old drugs learn new tricks. Nature medicine 20: 590–591. https://doi.org/10.1038/nm.3595 PMID: 24901567
4. Jahchan NS, Dudley JT, Mazur PK, Flores N, Yang D, Palmerton A, et al. (2013) A drug repurposing approach identifies tricyclic antidepressants as inhibitors of small cell lung cancer and other
neuroendocrine tumors. Cancer Discov 3:1364–1377. https://doi.org/10.1158/2159-8290.CD-13-0183 PMID: 24078773

5. Corbett A, Williams G, Ballard C (2013) Drug repositioning: an opportunity to develop novel treatments for Alzheimer's disease. Pharmaceuticals (Basel) 6: 1304–1321.

6. Corbett A, Pickett J, Burns A, Corcoran J, Dunnett SB, Edson P, et al. (2012) Drug repositioning for Alzheimer's disease. Nat Rev Drug Discov 11: 833–846. https://doi.org/10.1038/nrd3869 PMID: 23123941

7. Sardana D, Zhu C, Zhang M, Gudivada RC, Yang L, Jegga AG. (2011) Drug repositioning for orphan diseases. Brief Bioinform 12: 346–356. https://doi.org/10.1093/bib/bbr021 PMID: 21504985

8. Harrison C (2011) Signatures for drug repositioning. Nat Rev Genet 12: 668.

9. Chong CR, Sullivan DJ Jr. (2007) New uses for old drugs. Nature 448: 645–646. https://doi.org/10.1038/448645a PMID: 17687303

10. Rastegar-Mojarad M, Ye Z, Kolesar JM, Hrebbrin SJ, Lin SM (2015) Opportunities for drug repositioning from phenotype-wide association studies. Nature biotechnology 33: 342–345. https://doi.org/10.1038/nbt.3183 PMID: 25850054

11. Wang Z-Y, Zhang H-Y (2013) Rational drug repositioning by medical genetics. Nature biotechnology 31: 1080–1082. https://doi.org/10.1038/nbt.2738 PMID: 24316641

12. Lamb J, Crawford ED, Peck D, Modell JW, Blat IC, Wrobel MJ, et al. (2006) The Connectivity Map: using gene-expression signatures to connect small molecules, genes, and disease. Science 313: 1929–1935. https://doi.org/10.1126/science.1132939 PMID: 17008526

13. Iorio F, Rittman T, Ge H, Menden M, Saez-Rodriguez J (2013) Transcriptional data: a new gateway to drug repositioning? Drug Discov Today 18: 350–357. https://doi.org/10.1016/j.drudis.2012.07.014 PMID: 22897878

14. Sirota M, Dudley JT, Kim J, Chiang AP, Morgan AA, Sweet-Cordero A, et al. (2011) Discovery and preclinical validation of drug indications using compendia of public gene expression data. Sci Transl Med 3: 96ra77. https://doi.org/10.1126/scitranslmed.3001318 PMID: 21849665

15. Dudley JT, Sirota M, Shenoy M, Pai RK, Roedder S, Chiang AP, et al. (2011) Computational repositioning of the anticonvulsant topiramate for inflammatory bowel disease. Sci Transl Med 3: 96ra76. https://doi.org/10.1126/scitranslmed.3002648 PMID: 21849664

16. Law V, Knox C, Djoumbou Y, Jewison T, Guo AC, Liu Y, et al. (2014) DrugBank 4.0: shedding new light on drug metabolism. Nucleic Acids Res 42: D1091–1097. https://doi.org/10.1093/nar/gkt1068 PMID: 24203711

17. Gottlieb A, Stein GY, Ruppin E, Sharani R (2011) PREDICT: a method for inferring novel drug indications with application to personalized medicine. Mol Syst Biol 7: 496–504. https://doi.org/10.1038/msb.2011.26 PMID: 21654763

18. von Eichborn J, Murgueitio MS, Dunkel M, Koerner S, Bourne PE, Preissner R. (2011) PROMISCUOUS: a database for network-based drug-repositioning. Nucleic Acids Res 39: D1060–1066. https://doi.org/10.1093/nar/gkq1037 PMID: 21071407

19. Carrella D, Napolitano F, Rispoli R, Miglietta M, Carissimo A, Cutillo L, et al. (2014) Mantra 2.0: an online collaborative resource for drug mode of action and repurposing by network analysis. Bioinformatics 30: 1787–1788. https://doi.org/10.1093/bioinformatics/btu056 PMID: 24558125

20. Rubio-Perez C, Tamborero D, Schroeder MP, Antolin AA, Deu-Pons J, Perez-Llamas C, et al. (2015) In silico prediction of anticancer drugs to cohorts of 28 tumor types reveals targeting opportunities. Cancer cell 27: 382–396. https://doi.org/10.1016/j.ccel.2015.02.007 PMID: 25759023

21. Barrett T, Wilhite SE, Ledoux P, Evangelista C, Kim IF, Tomashevsky M, et al. (2013) NCBI GEO: archive for functional genomics data sets—update. Nucleic Acids Res 41: D991–995. https://doi.org/10.1093/nar/gks1193 PMID: 23193258

22. Reimers M, Carey VJ (2006) Bioconductor: an open source framework for bioinformatics and computational biology. Methods Enzymol 411: 119–134. https://doi.org/10.1016/S0076-6879(06)11008-3 PMID: 16959789

23. Knox C, Law V, Jewison T, Liu P, Ly S, Frolikis A, et al. (2011) DrugBank 3.0: a comprehensive resource for ‘omics’ research on drugs. Nucleic Acids Res 39: D1035–1041. https://doi.org/10.1093/nar/gkq1126 PMID: 21059682
26. Kanehisa M, Goto S, Kawashima S, Okuno Y, Hattori M (2004) The KEGG resource for deciphering the genome. Nucleic Acids Res 32: D277–280. https://doi.org/10.1093/nar/gkh063 PMID: 14681412

27. Davis AP, Grondin CJ, Lennon-Hopkins K, Saraceni-Richards C, Sciaky D, King BL, et al. (2015) The Comparative Toxicogenomics Database’s 10th year anniversary: update 2015. Nucleic Acids Res 43: D914–920. https://doi.org/10.1093/nar/gku935 PMID: 25326323

28. Yang H, Qin C, Li YH, Tao L, Zhou J, Yu CY, et al. (2016) Therapeutic target database update 2016: enriched resource for bench to clinical drug target and targeted pathway information. Nucleic Acids Res 44: D1069–1074. https://doi.org/10.1093/nar/gkv1230 PMID: 26578601