The Performance of Gene Expression Signature-Guided Drug-Disease Association in Different Categories of Drugs and Diseases

by

Xiguang Qi

Bachelor of Medicine, Sun Yat-Sen University, 2018

Submitted to the Graduate Faculty of the

School of Pharmacy in partial fulfillment

of the requirements for the degree of

Master of Science

University of Pittsburgh

2020

UNIVERSITY OF PITTSBURGH

SCHOOL OF PHARMACY

This thesis was presented

by

Xiguang Qi

It was defended on

March 26, 2020

and approved by

Dr. Lirong Wang, Assistant Professor, Department of Pharmaceutical Sciences

Dr. Levent Kirisci, Professor, Department of Pharmaceutical Sciences

Dr. Xinghua Lu, Professor, Department of Biomedical Informatics

Dr. Junmei Wang, Associate Professor, Department of Pharmaceutical Sciences

Thesis Advisor: Dr. Lirong Wang, Assistant Professor, Department of Pharmaceutical Sciences

The Performance of Gene Expression Signature-Guided Drug-Disease Association in Different Categories of Drugs and Diseases

Xiguang Qi, BM

University of Pittsburgh, 2020

Gene expression signature (GES) is a group of genes that shows a unique expression profile as a result of transcriptional machinery-related perturbations by drugs, genetic modification or diseases. The comparisons between GES profiles have been used to investigate the relationships between drugs, their targets and diseases with some successful cases reported. The rationale behind GES-guided drug-disease association is that if a medication can induce an opposite GES profile against that of a disease, it should possess the ability to reverse the gene expressions caused by the disease, and can be considered as a potential treatment of that disease. In this study, we data-mined the crowd extracted expression of differential signatures (CREEDS) database to evaluate the similarity of GES profiles between FDA approved drugs and their indicated diseases. The aim of our study is to explore the application domains of GES-guided drug-disease associations, that is, through the analysis of the similarity of GES profiles on known pairs of drug-disease associations, we can identify subgroups of drugs/diseases that are suitable for GES-guided drug-disease association for repositioning drugs. Our results suggest that GES-guided drug-disease association method is better suited for only some subgroups or pathways of drugs/diseases, such as drugs and diseases associated with immune system, non-chemotherapy drugs or mTOR signaling pathway, which showed significant higher correlations between their GES profiles.

Table of Contents

List of Tables

List of Figures

Preface

I always believe rules hide in data. I believe a massive amount of cases and experiments speak the truth. If not, that is because the application domain or method of this data is wrong. When I was doing a GES-guided drug repositioning research during the first year of my master study, I continually got unreasonable results. With the same method, someone succeeded, others failed. At that time, I thought that maybe this GES-guided drug repositioning method was not suitable for every kind of drug and disease. That is why I did this research, to validate the performance of gene expression signature-guided drug-disease association in different categories of drugs and diseases, which is the major content of this thesis.

Firstly, I would like to thank Dr. Xiang-Qun (Sean) Xie and Dr. Lirong Wang who gave me the opportunity to study in University of Pittsburgh as a master student in the past two years. And I am very glad to further appreciate Dr. Lirong Wang again as my adviser for his patiently guidance, encouragement and support. I would also like to thank Dr. Levent Kirisci for his statistical technique support and suggestions for my thesis. And Dr. Sweet Robert with his valuable suggestions in revising this thesis. Thank all my fellow students, friends and families for their help throughout my whole master study. Thank all the committee members Dr. Lirong Wang, Dr. Levent Kirisci, Dr. Xinghua Lu and Dr. Junmei Wang for the suggestions to this thesis and the attendance to the defense.

1.0 Introduction and Backgrounds

1.1 Drug Repositioning Technology and Application

Drug discovery is a costly and time-consuming process. With traditional drug discovery strategies, it can cost over 10 billion dollars to develop a new drug approved by the US Food and Drug Administration (FDA)[1]. Meanwhile, the productivity of new drugs does not catch up with the trend of the increasing fund spent in new drug development since 1990[1-6]. On the other hand, the long research and development time and low average success rate for developing an entirely new drug also impair the efficiency of drug discovery. As the success rate of inventing a newly approved small molecular drug is only around 2%[7, 8], drug repositioning, which aims at finding new indications for existing drugs, is undisputedly a low-risk and high-reward strategy comparing to traditional drug discovery methods. Taking the advantage of improving bioinformatics methods with biology big data and known data for existing drugs, the time cost in finding a new indication for an already approved drug through drug repositioning decreases remarkably. It only spends an average of 8 years and 8.4 million dollars on reposition a drug[1, 9]. Additionally, the repositioned drugs have already passed all phases of the clinical trials which further increased the success rate of drug discovery through repurposing.

The key of drug repositioning method is to identify potential new drug-disease indication relationships. Among all drug repositioning methods, the computational approaches have the lowest cost and fewer resource requirements[10]. Within these years, large varieties of databases such as DrugBank (https://www.drugbank.ca/)[11], Kyoto encyclopedia of genes and genome (KEGG, https://www.genome.jp/kegg/)[12] and Gene Expression Omnibus (GEO,

https://www.ncbi. nlm.nih.gov/geo/)[13] have been established, and they can facilitate the development of new computational approaches. Among drug repositioning studies we found[14-25], protein-protein interaction (PPI) network is the most popular method and gene expression signature (GES, see **section 1.2** for more details) receives a relatively less attention. The aim of this study is to identify the reasons for the low productivity of GES-guided drug repositioning method. Through conducting a simulated drug repositioning study, we evaluated the performance of this method in different drug and disease categories.

1.2 Gene Expression Signature Technology and Application

Gene expression signature (GES) is a set of comprehensive gene expression profiles that can reveal the difference of gene expressions between stimulated and normal cell states[26]. This concept was initially created for distinguishing different types of diffuse large B-cell lymphoma[26], and it's currently applied in cancer-related areas for classifying disease genotype and predicting clinical outcomes [27-46]. For example, Ramaswamy, S. et al. had created a GES database for diagnosing and categorizing the tumour type with an accuracy rate of 78%[27]. Wright, G. *et al.* developed a Bayesian rule-based algorithm to classify diffuse large B cell lymphoma into two subgroups which had a significant difference in 5-year survival rate[28]. Chen, H.-Y. *et al.* selected a five-gene signature which served as an independent predictor of relapse and survival rate in non-small-cell lung cancer [32]. Theoretically, the GES-guided method can reveal the associations (or in another word, similarity) among cell stages under disease conditions and drug interventions, thus it can be utilized as a drug repositioning strategy. As a matter of fact, some

successful cases of application on drug development are also reported in recent years [47-54] (See below **Table 1**).

1.3 Gene Expression Signature-Guided Drug Repositioning

There are two strategies for applying GES analysis on drug development: drug-drug based, and drug-disease based. Drug-drug based method is to compare the similarity between GES induced by a drug of interest and GES of drugs with known mechanisms to study the mechanisms of that drug. If two different drugs could initiate similar GES profiles, they are considered to have "functional similarity"[55]. Drug-disease based method is to compare the similarity between a GES of drugs and that of a disease to identify medications with new therapeutic potentials. If the two GES profiles from a drug-disease pair have opposite expression patterns (a reversed similarity), this drug is considered to have potential therapeutic effects for this disease. Studies aimed at drug repurposing or repositioning based on GES analysis usually use one or combine both strategies [47-55]. In addition, some studies tried to combine the GES method with other methods like machine learning to increase the accuracy of compound-indication prediction[55]. However, those kinds of studies usually reported the successful predictions only. For example, in the study of Sirota et al. [48], among 164 most significant compounds which they believed have an undiscovered potential novel indications, only one compound (cimetidine) was validated by *in vitro* and *in vivo* rodent experiment (**Table 1**), and the true accuracy of this method remains to be assessed.

Table 1. Results of reported GES-guided drug repositioning studies

Drug candidates*	Drugs reported and validated	Validation method	Disease	Reference
57**	Fasudil	In vitro (human)	Neurodegenerative Disorders	[47]
164	Cimetidine	In vitro and in vivo (rodent)	Lung Adenocarcinoma	[48]
20**	Ursolic acid	In vitro and in vivo (rodent)	Muscle Atrophy	[50]
50**	Chlorpromazine and trifluoperazine	In vitro and in vivo (rodent)	Hepatocellular carcinoma	[51]
18	Fluphenazine	In vivo (rodent)	Alopecia	[52]
5	Phenoxybenzamine	In vivo (rodent)	Osteoarthritic Pain	[53]
5**	Vorinostat	In vitro (human)	Gastric Cancer	[54]

*Drugs with GES evidence to have potential novel indications according to the article.

**Only the top X GES-scored compounds were showed in this article

Due to the different and complex mechanisms of disease processes, the hypothesis of an "inverse pattern of GES between drugs and diseases for therapeutic effects" may not hold or at least may not be suitable for all categories of drugs and diseases. In other words, GES can be best used for certain diseases but not for others. According to our knowledge, the performance of GES-guided drug-disease associations stratifying by drug and disease categories haven't been reported yet. Herein, we conducted a study to validate the power of GES-guided drug repositioning method and to further explore which specific subgroups of drug-disease pairs are more suitable (have higher true positive rates) for this method.

2.0 Method and Material

2.1 Gene Signature Data Collection and Filtering

In this study, all gene signature information was collected from a well-calibrated GES repository, Crowd Extracted Expression of Differential Signatures (CREEDS)[56] database. The CREEDS database is maintained by the Ma'ayan Lab at Icahn School of Medicine, Mount Sinai. CREEDS utilized GEO2Enrichr[57] to extract GES profiles from GEO database maintained by the National Center for Biotechnology Information (NCBI) and applied Characteristic Direction (CD) model[58] to identify differentially expressed genes. This database V1.0 includes 10,797 single-gene perturbations, 2,258 disease signatures, and 5,516 drug perturbation gene signatures. Among these signatures, 2,176 single-gene perturbations, 828 disease signatures, and 875 drug perturbation signatures were manually calibrated, and they are more accurately compared with the automatically generated GES by machine learning method. The CREEDS database allows users to compare the similarity between the user-specified GES and the GESs processed and stored in the CREEDS.

All the manually calibrated GES profiles were then filtered by following criterion:

1. Assays must be from human tissues or human cell lines; and
2. Drugs have been approved by the FDA.

Each GES profile includes a list of up-regulated genes and a list of down-regulated genes. The Signed Jaccard Index (SJI)[56] (see below), a measurement for the similarity between two GES profiles from the paired drug-disease, is calculated. When a drug or a disease has multiple GES profiles, we calculated the SJIs of all the possible combinations. And an overall score for

each unique drug-disease pair is calculated from the average of all scores from pairs sharing the same drug-disease combination. All the disease signatures and drug perturbation signatures were requested through the application program interface (API) provided by CREEDS. A GES profile will be removed if it was generated from the same assay but was labelled as from both a drug and a disease because this may cause exceptional similarity. Then under the criteria that (a) the GES profiles must come from assays of human cells/tissues and (b) drugs must be approved by FDA, the remained signatures were paired within drugs and diseases according to the indication associations. Signatures without at least one indicated drug-disease relationship were also excluded from the further analysis.

2.2 Similarity score calculation

As our study aims at validating the power of GES-guided drug repositioning method through a simulated GES-guided drug repositioning study, we believe an unranked scoring method is more suitable for this study to find out which kind of drug-disease pairs has larger probability to present a reverse pattern. As such, in our analysis, SJI, which is based on the Jaccard similarity coefficient[59], was used to calculate the similarity between paired GES profiles of drugs and diseases. The Jaccard similarity coefficient is a statistic used to gauge the similarity between two sample sets. It is defined as the size of the intersection divided by the size of the union of two sample sets. The Jaccard similarity coefficient of two given gene sets is calculated as follows:

$$Jaccard\ Similarity\ Coefficient(G_1, G_2) = \frac{SAME}{ALL}$$

G_1 and G_2 stand for two lists of differential expressed gene sets, "SAME" represents the number of same genes between two given gene sets, and "ALL" stands for all the unique genes appeared in the two gene sets. Signed Jaccard index or SJI, which combines Jaccard similarity coefficient with gene regulation direction is calculated as follows:

$$Signed\ Jaccard\ index(G_1, G_2) = \frac{J(G_1^{up}, G_2^{up}) + J(G_1^{down}, G_2^{down}) - J(G_1^{up}, G_2^{down}) - J(G_1^{down}, G_2^{up})}{2}$$

Where J means Jaccard similarity coefficient, G^{up} and G^{down} are up- or down-regulated genes in the given gene set G, respectively. The SJI ranges from 1 to -1, where 1 and -1 indicate a completely same pattern and inverse pattern of two gene sets, respectively. And 0 indicates that these two sets have no associations, or the same part is cancelled out by the inverse part. The CREEDS API offers the function to calculate the SJI automatically. However, we found the API could not calculate the SJI correctly when two GES profiles are highly overlapped. All the SJIs in this study were re-calculated.

2.3 Subgroup Classification

In our analysis, we assessed the following factors that might influence the power of the GES-guided drug repositioning method:

i. Disease classifications
ii. Drug target subfamilies
iii. The relationship between the drugs' main therapeutic targets and human TFs
iv. The drug is a chemotherapy drug or not

v. The drug's therapeutic category

Specifically, the five categories of subgroups mentioned above were defined as:

i. Disease classifications: A subgroup was assigned to a disease in a drug-disease pair according to the ICD-11-level 1 code of the disease.

ii. Drug target: Subgroups are divided by the main therapeutic target of each drug. To avoid groups split too small, some targets from same subfamily are grouped as one. For example, "Beta-1 adrenergic receptor", "Beta-2 adrenergic receptor" and "Beta-3 adrenergic receptor" are grouped in the same subgroup "Beta adrenergic receptors".

iii. TF level: A TF level was assigned according to the relationship between the drugs' main therapeutic targets and human TF. Drugs with main therapeutic targets that can directly interact with at least one TF were labelled as "directly". Drugs with main therapeutic targets which are human DNA structures or human proteins but not TFs were labelled as "not-directly". Drugs interacting with proteins or structures of non-human (for example, from virus or bacterial) as main therapeutic targets were labelled as "non-Human".

iv. Chemotherapy: Drug with main therapeutic targets as "DNA cross-linking/alkylation", "DNA/ligase", "DNA/methyltransferase", "DNA/polymerase". "DNA/topoisomerase-human", "micro-tubules", "nucleotide synthesis" or "Thymidylate synthase" were defined as chemotherapy drugs.

v. ATC classification: Subgroups were divided according to the Anatomical Therapeutic Chemical (ATC) Classification system, level 3. Drugs with multiple

classifications caused by different administration routes were unified to systematic use.

2.4 Drug-related information collection

Drug targets information was collected from DrugBank[11, 60] Release Version 5.1.4[61] (https://www.drugbank.ca/releases/latest#external-links). Only the targets with the main therapeutic effect in mechanism of action section were included. For example, for drug "Atorvastatin", it has five targets labelled in DrugBank (3-hydroxy-3-methylglutaryl-coenzyme A reductase, Dipeptidyl peptidase 4, Aryl hydrocarbon receptor, Histone deacetylase 2, and Nuclear receptor subfamily 1 group I member 3). Only "3-hydroxy-3-methylglutaryl-coenzyme A reductase" is labeled as pharmacological active. Others are labeled as pharmacological action unknown. In this case the "3-hydroxy-3-methylglutaryl-coenzyme A reductase" will be the main therapeutic target for "Atorvastatin".

The human transcription factor (TF) list was collected from literature published by Samuel A. Lambert et al.[62].

Drugs' Anatomical Therapeutic Chemical (ATC) Classifications were collected from the WHO official website (https://www.whocc.no/atc_ddd_index/). The ATC 1st levels are main anatomical groups, the ATC 2nd levels are pharmacological subgroups, and the level 3 is pharmacological and therapeutic subgroup. The 3-level which defined by the function of the drug is used in this study.

The drugs' indications were collected from "indications and usage" section of FDA label on FDA website (https://labels.fda.gov/). For example, the FDA "indications and usage" label

section of metformin is “GLUMETZA (brand name of metformin) is a biguanide indicated as an adjunct to diet and exercise to improve glycemic control in adults with type 2 diabetes mellitus.” (https://www.accessdata.fda.gov/drugsatfda_docs/label/2011/021748s010lbl.pdf) So, its indication relationship is “metformin-type 2 diabetes mellitus”.

Disease Classification was assigned to each disease based on the International Classification of Diseases 11th Revision (ICD-11), level 1 (https://icd.who.int/en) which separate disease by system level. For example, “childhood type dermatomyositis” is labeled as “Diseases of the immune system”.

2.5 Statistical analysis and pathway analysis

Random control group is generated by calculating the average SJI of all possible drug-disease pairs without indicating associations to imitate a GES-guided drug repositioning screening.

A t-test[63] was applied to quantify the mean differences of the SJI between drug-indicated-disease pairs and random control.

For subgroup analysis, generalized linear model[64] (GLM) least squares mean partitions F tests function is applied to estimate the mean difference between indicated group and control group under multiple factors (In our case, different subgroups in a same category) since the data is unbalanced. A significant result of a certain subgroup indicated that the average SJI of this subgroup is significantly different between two indication levels (Yes/No). False discovery rate (FDR) q-value of Benjamini–Hochberg procedure[65] is controlled to 0.05 to avoid inflated experiment-wise type I error rate caused by multiple comparisons among all subgroups.

Data processing and statistical analysis (student t-tests, GLM, FDR calculation) were conducted using R studio 3.6.1[66] and SAS software version 9.4 (Copyright © 2019 SAS Institute Inc. Cary, NC, USA).

Differentially reversed expression genes (Top 5% negative score according to the relatively reverse percentage) from the most significant subgroup will be chosen as examples to conduct biological pathway enrichment analysis.

The relatively reverse percentage is calculated as:

$$Relatively\ expression\ probability\ of\ a\ gene(G^{I-R\%}) = D^{I}\% - D^{R}\%$$

$D^{I}\%$ and $D^{R}\%$ stand for the percentage of the gene which is differentially expressed in all assays of indicated/random drug-diseases pairs. It is calculated as:

$$D\% = \frac{NS - NR}{Total\ assays\ pairs}$$

NS and NR represent the times of this gene showed a same or reverse regulation direction between assays of drugs and diseases among all drug-disease assays pairs.

The $G^{I\text{-}R\%}$ ranges from 100% to -100%. A higher positive score indicates that this gene is more likely to be expressed in the same direction in indicated drug-disease assays compared to random drug-disease assays. Likewise, a lower negative score indicates that this gene has a higher probability to express reversely between indicated drug-disease assays compared to random drug-disease assays.

Biological pathway enrichment analysis was conducted by Ingenuity pathway analysis (IPA, QIAGEN Inc., https://www.qiagenbioinformatics.com/products/ingenuitypathway-analysis).

3.0 Results

3.1 GES Profiles Enrollment and Drug-disease pairs

GSE10432, GSE7036, GSE6264, GSE38713, GSE31773, GSE11393, GSE8157, GSE13887, and GSE11223 were found sourced from the same assays but constructed signatures of both drugs and diseases. All drug parts of these assays were removed (including CREEDS IDs, drug:3292, drug:3064, drug: 3289, drug:3194, drug:3195, drug:2485, drug:3401, drug:3196, drug:2796, drug:3181, drug:3294, drug:3287) except CREEDS IDs of dz:297 and drug:2772. In this case, the disease "acne" was mismatched with its assay GSE10432. So, the disease GES profile (CREEDS IDs: dz:297) was removed. Two GES profiles from mouse, drug:3288 and dz:724, were mis-specified as human and they were also excluded. The relationship between these GEO Series (GSE) and CREEDS IDs is showed in **Table 2**.

Table 2 The GEO Series with CREEDS IDs Excluded

GEO Series	CREEDS IDs	Excluded CREEDS IDs
GSE10432	drug:2772, dz:297	dz:297
GSE7036	drug:3292, dz:181	drug:3292
GSE6264	drug:3064, dz:582	drug:3064
GSE38713	drug:3289, drug:3194, drug:3195, dz:810	drug:3289, drug:3194, drug:3195
GSE31773	drug:2485, dz:712, dz:713, dz:714, dz:715	drug:2485
GSE11393	drug:3401, drug:3196, dz:773, dz:267	drug:3401, drug:3196
GSE8157	drug:2796, dz:880	drug:2796
GSE13887	drug:3181, dz:450	drug:3181,
GSE11223	drug:3294, drug:3287, dz:590, dz:591, dz:593, dz:589, dz:588, dz:587, dz:586, dz:585	drug:3294, drug:3287
GSE10432	drug:2772, dz:297	dz:297
GSE7762	drug:3288	drug:3288
GSE3248	dz:724	dz:724

Then after applying the inclusion criteria, 230 manual disease signatures and 244 manual drug perturbation signatures were enrolled into the analysis. There were 2,929 pairs of known

drug-disease association (among 56120 total pairs of assays), among which 167 are unique pairs from 56 unique drugs and 71 unique diseases. In addition, 3809 (56*71-167) unique random drug-disease pairs were generated as the control group by calculating average SJI of drug-disease paired assays from rest 53191 (56120-3809) pairs without indicating associations.

3.2 Subgroups distribution

Among all these 56 drugs, 32 unique protein targets with 22 categories of ATC classification were assigned. Thirteen drugs are classified as chemotherapy drugs, and 44 drugs are not (Methotrexate is both a chemotherapy and a non-chemotherapy drug due to its different main therapeutic targets when against different diseases). For TF level, 12 drugs were labelled as "directly", 39 drugs were labelled as "not-directly", and 5 drugs were labelled as "non-Human". Also, seventy-one diseases were divided into 11 ICD-11 categories. Totally, 70 subgroups belonging to five categories were assigned (**Figure 1**, **Table 3** and **Table 4**).

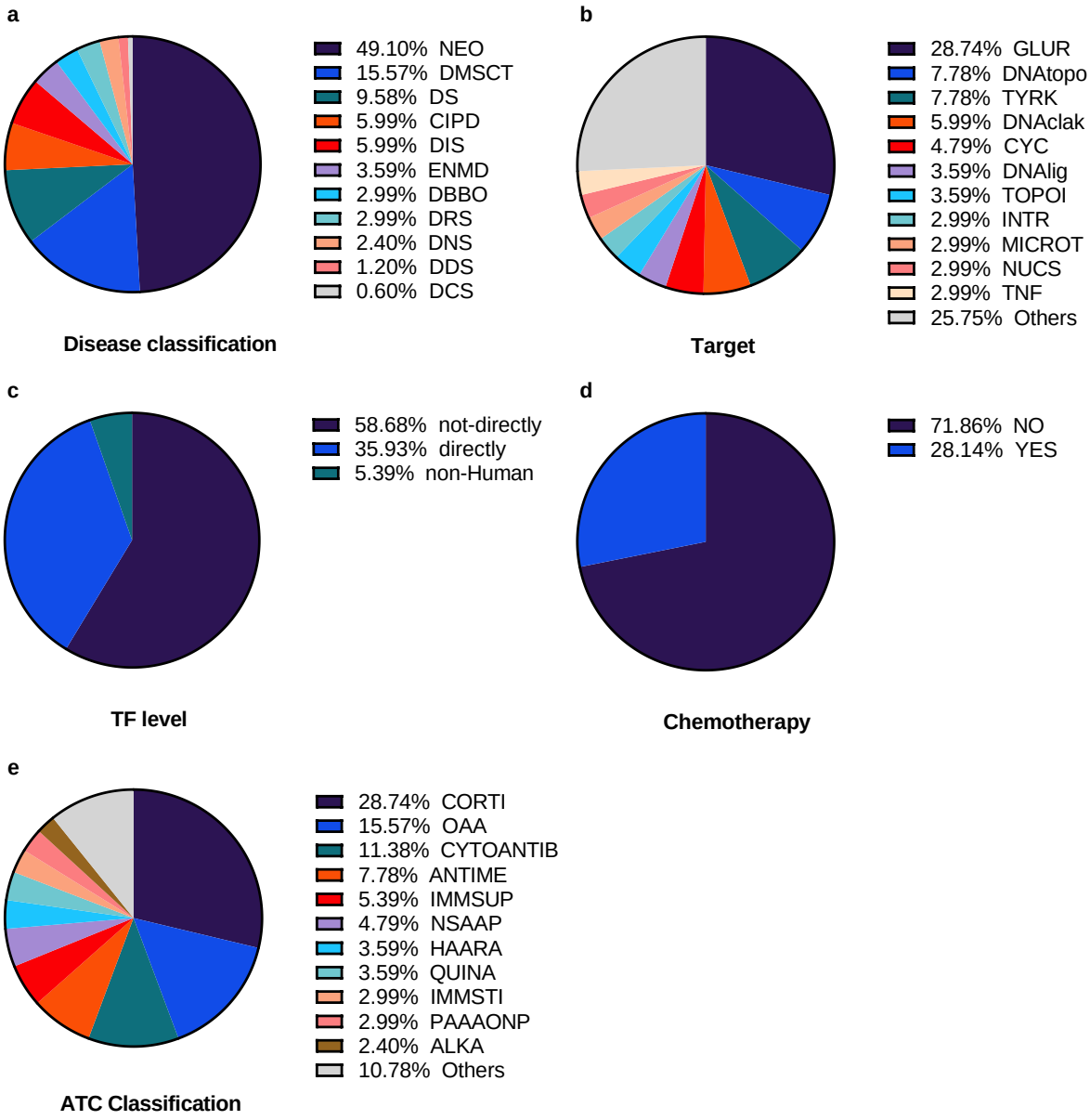

Figure 1 Subgroup distribition according to different catagories.

The subgroups proportion of 167 unique indicated drug-disease pairs of different categories: (A) Disease classification. NEO: Neoplasms, DMSCT: Diseases of the musculoskeletal system or connective tissue, DS: Diseases of the skin, CIPD: Certain infectious or parasitic diseases, DIS: Diseases of the immune system, ENMD: Endocrine, nutritional or metabolic diseases, DBBO: Diseases of the blood or blood-forming organs, DRS: Diseases of the respiratory system, DNS: Diseases of the nervous system, DDS: Diseases of the digestive system, DCS: Diseases of the circulatory system. (B) Drug Target. GLUR: Glucocorticoid receptor, DNAtopo: DNA/topoisomerase-human,

TYRK: Tyrosine kinase, DNAclak: DNA cross-linking/alkylation, CYC: Cyclooxygenase, DNAlig: DNA/ligase, TOPOI: Topoisomerase-non-human, INTR: Interferon receptor, MICROT: Microtubules, NUCS: Nucleotide synthesis, TNF: Tumor necrosis factor. (C) TF level. “directly” stands for drugs with main therapeutic targets that can directly interact with at least one TF. “not-directly” indicates drugs with main therapeutic targets as human DNA structures or human proteins but not TFs. “non-Human” represent drugs interacting with proteins or structures of non-human (for example, from virus or bacteria) as main therapeutic targets. (D) Chemotherapy. “YES” or “NO” indicates whether the drug is a chemotherapy drug or not. (E) ATC classification. CORTI: Corticosteroids for systemic use, plain, OAA: other antineoplastic agents, CYTOANTIB: Cytotoxic antibiotics and related substances, ANTIME: Antimetabolites, IMMSUP: Immunosuppressants, NSAAP: Anti-inflammatory and antirheumatic products, non-steroids, HAARA: Hormone antagonists and related agents, QUINA: Quinolone antibacterial, IMMSTI: Immunostimulants, PAAAONP: Plant alkaloids and other natural products, ALKA: Alkylating agents.

Table 3 Subgroup assignment according to drug properties

Unified drug name*	TF level	Target	Chemotherapy	ATC classification
Abiraterone	Not-directly	Cytochromes P450	No	Hormone antagonists and related agents
Actinomycin d	Not-directly	DNA/topoisomerase-human	Yes	Cytotoxic antibiotics and related substances
Aminolevulinic acid	Not-directly	Delta-aminolevulinic acid dehydratase	No	Other antineoplastic agents
Anastrozole	Not-directly	Cytochromes P450	No	Hormone antagonists and related agents
Atorvastatin	Not-directly	HMG-CoA reductase	No	Lipid modifying agents, plain
Azacitidine	Directly	DNA/methyltransferase	Yes	Antimetabolites
Bexarotene	Directly	Retinoic acid receptor	No	Other antineoplastic agents
Bicalutamide	Directly	Androgen receptor	No	Hormone antagonists and related agents
Bleomycin	Not-directly	DNA/ligase	Yes	Cytotoxic antibiotics and related substances
Bortezomib	Not-directly	Proteasome subunit beta	No	Other antineoplastic agents
Calcitriol	Directly	Vitamin D3 receptor	No	Vitamin a and d, incl. Combinations of the two
Carboplatin	Not-directly	DNA cross-linking/alkylation	Yes	Other antineoplastic agents
Celecoxib	Not-directly	Cyclooxygenase	No	Anti-inflammatory and antirheumatic products, non-steroids
Chlorambucil	Not-directly	DNA cross-linking/alkylation	Yes	Alkylating agents
Ciprofloxacin	Non-human	Topoisomerase-non-human	No	Quinolone antibacterials
Cisplatin	Not-directly	DNA cross-linking/alkylation	Yes	Other antineoplastic agents
Cytarabine	Not-directly	DNA/polymerase	Yes	Antimetabolites
Dasatinib	Not-directly	Tyrosine kinase	No	Other antineoplastic agents
Decitabine	Directly	DNA/methyltransferase	Yes	Antimetabolites
Dexamethasone	Directly	Glucocorticoid receptor	No	Corticosteroids for systemic use, plain
Diclofenac	Not-directly	Cyclooxygenase	No	Anti-inflammatory and antirheumatic products, non-steroids
Doxorubicin	Not-directly	DNA/topoisomerase-human	Yes	Cytotoxic antibiotics and related substances
Doxycycline	Non-human	16s ribosomal RNA	No	Tetracyclines
Estradiol	Directly	Estrogen receptor	No	Estrogens
Etanercept	Not-directly	Tumor necrosis factor	No	Immunosuppressants

Table 3 (continued)

Unified drug name*	TF level	Target	Chemotherapy	ATC classification
Fluorouracil	Not-directly	Thymidylate synthase	Yes	Antimetabolites
Formoterol	Not-directly	Beta adrenergic receptor	No	Adrenergics, inhalants
Gatifloxacin	Non-human	Topoisomerase-non-human	No	Quinolone antibacterials
Gefitinib	Not-directly	Tyrosine kinase	No	Other antineoplastic agents
Hydrocortisone	Directly	Glucocorticoid receptor	No	Corticosteroids for systemic use, plain
Hydroxyzine	Not-directly	Histamine H1 receptor	No	Anxiolytics
Imatinib	Not-directly	Tyrosine kinase	No	Other antineoplastic agents
Insulin	Not-directly	Insulin receptor	No	Insulins and analogues
Interferon beta-1a	Not-directly	Interferon receptor	No	Immunostimulants
Interferon beta-1b	Not-directly	Interferon receptor	No	Immunostimulants
Interferon gamma-1b	Not-directly	Interferon receptor	No	Immunostimulants
Isotretinoin	Directly	Retinoic acid receptor	No	Anti-acne preparations for systemic use
Lapatinib	Not-directly	Tyrosine kinase	No	Other antineoplastic agents
Letrozole	Not-directly	Cytochromes P450	No	Hormone antagonists and related agents
Levofloxacin	Non-human	Topoisomerase-non-human	No	Quinolone antibacterials
Metformin	Not-directly	AMP-activated protein kinase	No	Blood glucose lowering drugs, excl. Insulins
Methotrexate	Not-directly	Nucleotide synthesis	Yes	Antimetabolites
Methotrexate	Not-directly	Aminoimidazole caboxamide ribonucleotide transformylase	No	Immunosuppressants
Metoprolol	Not-directly	Beta adrenergic receptor	No	Beta blocking agents
Natural alpha interferon	Not-directly	Interferon receptor	No	Immunostimulants
Nilotinib	Not-directly	Tyrosine kinase	No	Other antineoplastic agents
Paclitaxel	Not-directly	Microtubules	Yes	Plant alkaloids and other natural products
Pertuzumab	Not-directly	Tyrosine kinase	No	Other antineoplastic agents
Pimecrolimus	Not-directly	Kinase mTOR	No	Other dermatological preparations
Pioglitazone	Directly	Peroxisome proliferator-activated receptors	No	Blood glucose lowering drugs, excl. Insulins

Table 3 (continued)

Unified drug name*	TF level	Target	Chemotherapy	ATC classification
Ribavirin	Non-human	Inosine-5'-monophosphate dehydrogenase	No	Direct acting antivirals
Rituximab	Not-directly	CD20 antigen	No	Other antineoplastic agents
Rosiglitazone	Directly	Peroxisome proliferator-activated receptors	No	Blood glucose lowering drugs, excl. Insulins
Sorafenib	Not-directly	Tyrosine kinase	No	Other antineoplastic agents
Tamoxifen	Directly	Estrogen receptor	No	Hormone antagonists and related agents
Temozolomide	Not-directly	DNA cross-linking/alkylation	Yes	Alkylating agents
Trastuzumab	Not-directly	Tyrosine kinase	No	Other antineoplastic agents
Tretinoin	Directly	Retinoic acid receptor	No	Anti-acne preparations for systemic use
Vemurafenib	Not-directly	Serine/threonine-protein kinase B-Raf	No	Other antineoplastic agents

*Different names of the same drugs have been unified to a unique name. This may cause some difference between drug names in this table and **appendix table 1**.

Table 4 Subgroup assignment according to disease

Unified disease name*	Classification
Bacterial infectious disease	Certain infectious or parasitic diseases
Hepatitis C	Certain infectious or parasitic diseases
Septic shock	Certain infectious or parasitic diseases
Aplastic anemia	Diseases of the blood or blood-forming organs
Autoimmune thrombocytopenic purpura	Diseases of the blood or blood-forming organs
Diamond-Blackfan anemia	Diseases of the blood or blood-forming organs
Acute myocardial infarction	Diseases of the circulatory system
Ulcerative colitis	Diseases of the digestive system
Childhood type dermatomyositis	Diseases of the immune system
Chronic granulomatous disease	Diseases of the immune system
Dermatomyositis	Diseases of the immune system
Polymyositis	Diseases of the immune system
Pulmonary sarcoidosis	Diseases of the immune system
Sarcoidosis	Diseases of the immune system
Systemic lupus erythematosus	Diseases of the immune system
Ankylosing spondylitis	Diseases of the musculoskeletal system or connective tissue
Juvenile rheumatoid arthritis	Diseases of the musculoskeletal system or connective tissue
Osteoarthritis	Diseases of the musculoskeletal system or connective tissue
Osteoporosis	Diseases of the musculoskeletal system or connective tissue
Psoriatic arthritis	Diseases of the musculoskeletal system or connective tissue
Rheumatoid arthritis	Diseases of the musculoskeletal system or connective tissue
Multiple sclerosis	Diseases of the nervous system
Relapsing-remitting multiple sclerosis	Diseases of the nervous system
Allergic asthma	Diseases of the respiratory system
Asthma	Diseases of the respiratory system
Chronic obstructive pulmonary disease	Diseases of the respiratory system
Acne	Diseases of the skin
Actinic keratosis	Diseases of the skin
Allergic contact dermatitis	Diseases of the skin
Atopic dermatitis	Diseases of the skin
Discoid lupus erythematosus	Diseases of the skin
Psoriasis	Diseases of the skin
Urticaria	Diseases of the skin
Familial hypercholesterolemia	Endocrine, nutritional or metabolic diseases
Type 1 diabetes mellitus	Endocrine, nutritional or metabolic diseases
Type 2 diabetes mellitus	Endocrine, nutritional or metabolic diseases
Acute myeloid leukemia	Neoplasms
Anaplastic thyroid carcinoma	Neoplasms
Astrocytoma	Neoplasms

Table 3 (continued)

Unified disease name*	Classification
Breast cancer	Neoplasms
Chronic myeloid leukemia	Neoplasms
Colon cancer	Neoplasms
Ductal carcinoma in situ	Neoplasms
Esophagus adenocarcinoma	Neoplasms
Esophagus squamous cell carcinoma	Neoplasms
Gastrointestinal stromal tumor	Neoplasms
Glioblastoma multiforme	Neoplasms
Head and neck squamous cell carcinoma	Neoplasms
Hepatocellular carcinoma	Neoplasms
LGLL - Large granular lymphocytic leukemia	Neoplasms
Lung adenocarcinoma	Neoplasms
Lung large cell carcinoma	Neoplasms
Lung small cell carcinoma	Neoplasms
Lung squamous cell carcinoma	Neoplasms
Melanoma	Neoplasms
Multiple myeloma	Neoplasms
Myelodysplastic syndrome	Neoplasms
Nephroblastoma	Neoplasms
Ovarian cancer	Neoplasms
Ovarian serous carcinoma	Neoplasms
Pancreatic cancer	Neoplasms
Papillary thyroid carcinoma	Neoplasms
Precursor B lymphoblastic lymphoma/leukemia	Neoplasms
Prostate cancer	Neoplasms
Renal cell carcinoma	Neoplasms
Skin squamous cell carcinoma	Neoplasms
Squamous cell carcinoma of mouth	Neoplasms
Testicular cancer	Neoplasms
Testis seminoma	Neoplasms
Chronic lymphocytic leukemia	Neoplasms
Leukemia, chronic T-cell	Neoplasms

*Different names of the same diseases have been unified to a unique name. This may cause some difference between disease names in this table and **appendix table 2**.

3.3 Overall GES Similarity Scores of Drug-Indicted Disease Pairs against Random Drug-Disease Pairs

We observed a significantly lower similarity mean score (SJI) of drug-disease indication pairs than that of random drug-disease pairs (P-value two-side t-test, equals 0.02324). The average similarity score of indicated pairs is -0.00386 with a standard deviation of 0.01794 and that of random control pairs is -0.00072 with a standard deviation of 0.01750, indicating that GES similarity can reflect the therapeutic effects of the drugs.

3.4 Subgroup Scores of GES Similarity of Drug-Indicated Disease Pairs against Random Drug-Disease Pairs

More specifically, we compared drugs from five different categories of subgroups: (1) Disease classification; (2) Drug target; (3) TF level; (4) Chemotherapy; and (5) ATC classification. The results were listed in **Figure 2**, **Table 5** and **Table 6**. Subgroups with important or significant (FDR q-value lower than 0.05) results according to GLM least squares mean partitions F tests were listed in **Table 7**.

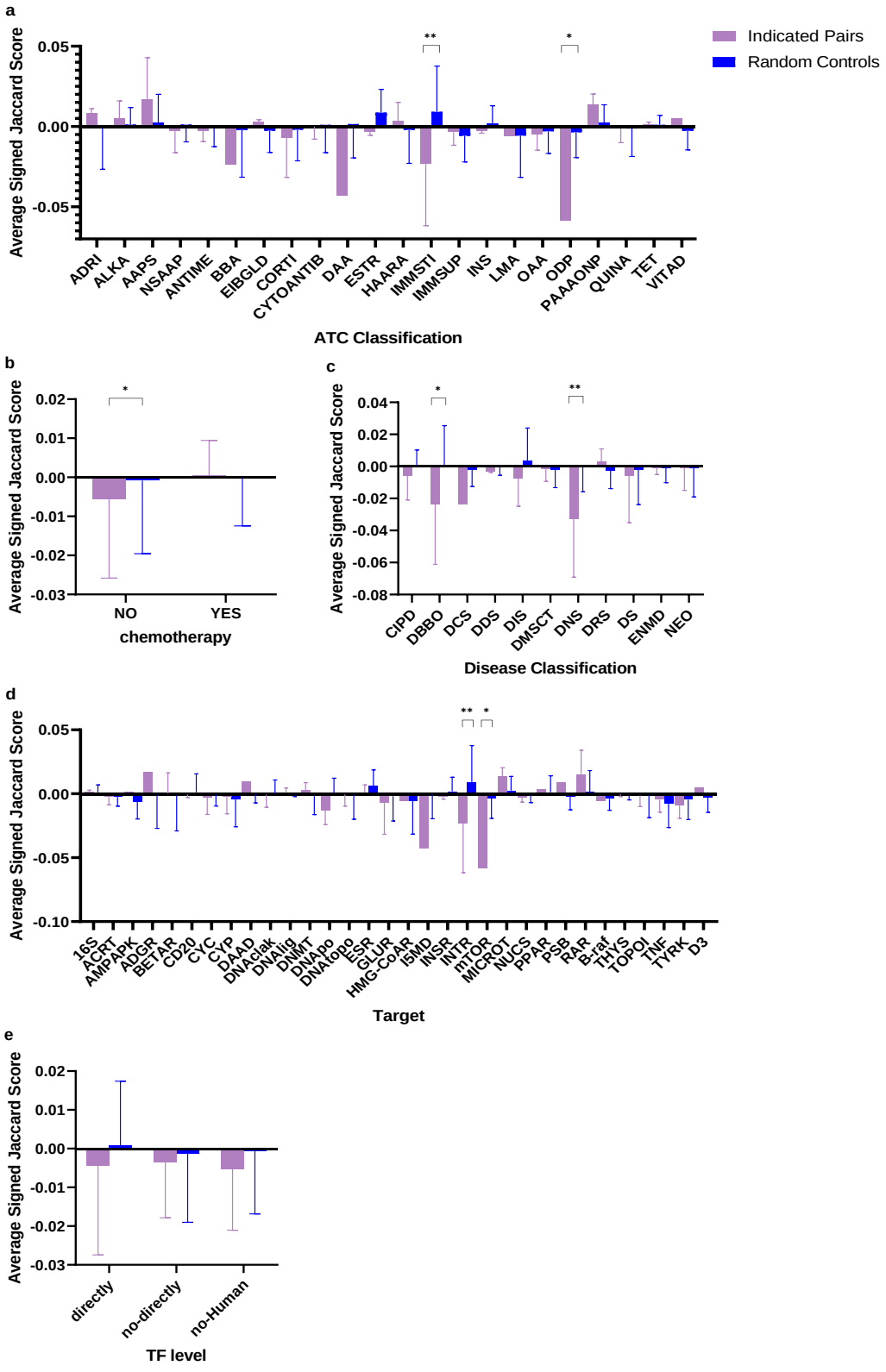

Figure 2 The average SJI score of drug-disease pairs split by different categories of subgroups.

The average SJI Score of unique drug-disease pairs split by different Categories of subgroups. (A) ATC classification. ADRI: Adrenergics, inhalants, ALKA: Alkylating agents, AAPS: Anti-acne preparations for systemic use, NSAAP: Anti-inflammatory and antirheumatic products, non-steroids, ANTIME: Antimetabolites, BBA: Beta blocking agents, EIBGLD: Blood glucose lowering drugs, excl. insulins, CORTI: Corticosteroids for systemic use, plain, CYTOANTIB: Cytotoxic antibiotics and related substances, DAA: Direct acting antivirals, ESTR: Estrogens, HAARA: Hormone antagonists and related agents, IMMSTI: Immunostimulants, IMMSUP: Immunosuppressants, INS: Insulins and analogues, LMA: Lipid modifying agents, plain, OAA: Other antineoplastic agents, ODP: Other dermatological preparations, PAAAONP: Plant alkaloids and other natural products, QUINA: Quinolone antibacterials, TET: Tetracyclines, VITAD: Vitamin a and d, incl. combinations of the two. **indicates FDR Q < 0.01, * Q < 0.05 (B) Chemotherapy. “YES” or “NO” indicates the drug is a chemotherapy drug or not. *indicates FDR Q < 0.05 (C) Disease classification. CIPD: Certain infectious or parasitic diseases, DBBO: Diseases of the blood or blood-forming organs, DCS: Diseases of the circulatory system, DDS: Diseases of the digestive system, DIS: Diseases of the immune system, DMSCT: Diseases of the musculoskeletal system or connective tissue, DNS: Diseases of the nervous system, DRS: Diseases of the respiratory system, DS: Diseases of the skin, ENMD: Endocrine, nutritional or metabolic diseases, NEO: Neoplasms; **indicates FDR Q < 0.01, * Q < 0.05 (D) Target. 16S: 16S ribosomal RNA, ACRT: Aminoimidazole caboxamide ribonucleotide transformylase, AMPAPK: AMP-activated protein kinase, ADGR: Androgen receptor, BETAR: Beta adrenergic receptor, CD20: CD20 antigen, CYC: Cyclooxygenase, CYP: Cytochromes P450, DAAD: Delta-aminolevulinic acid dehydratase, DNAclak: DNA cross-linking/alkylation, DNAlig: DNA/ligase, DNMT: DNA/methyltransferase, DNApo: DNA/polymerase, DNAtopo: DNA/topoisomerase-human, ESR: Estrogen receptor, GLUR: Glucocorticoid receptor, HMG-CoAR: HMG-CoA reductase, I5MD: Inosine-5'-monophosphate dehydrogenase, INSR: Insulin receptor, INTR: Interferon receptor, mTOR: Kinase mTOR, MICROT: Microtubules, NUCS: Nucleotide synthesis, PPAR: Peroxisome proliferator-activated receptors, PSB: Proteasome subunit beta, RAR: Retinoic acid receptor, B-raf: Serine/threonine-protein kinase B-raf, THYS: Thymidylate synthase, TOPOI: topoisomerase-no-human, TNF: Tumor necrosis factor, TYRK: Tyrosine kinase, D3: Vitamin D3 receptor; **indicates FDR Q < 0.01, * Q < 0.05 (E) TF level. “directly” stands for drugs with main therapeutic targets that can directly interact with at least one TF. “not-directly” indicates drugs with main therapeutic targets which are human DNA structures or human proteins but not TFs. “non-Human” represent drugs interacting with proteins or structures of non-human (for example, from virus or bacterial) as main therapeutic targets.

Table 5 The results of SJI score of indicated drug-disease pairs' subgroups

Category	Subgroups	Size	SD	Mean
Target	16S ribosomal RNA	2	0.00144	0.00134
Target	Aminoimidazole caboxamide ribonucleotide transformylase	4	0.00656	-0.00218
Target	AMP-activated protein kinase	1	---------	0.00198
Target	Androgen receptor	1	---------	0.01733
Target	Beta adrenergic receptor	4	0.01611	0.00012
Target	CD20 antigen	3	0.00233	-0.00057
Target	Cyclooxygenase	8	0.01331	-0.00293
Target	Cytochromes P450	3	0.01320	-0.00253
Target	Delta-aminolevulinic acid dehydratase	1	---------	0.00974
Target	Estrogen receptor	4	0.00589	0.00107
Target	Glucocorticoid receptor	48	0.02451	-0.00715
Target	HMG-CoA reductase	1	---------	-0.00594
Target	Inosine-5'-monophosphate dehydrogenase	1	---------	-0.04285
Target	Insulin receptor	2	0.00157	-0.00259
Target	Interferon receptor	5	0.03866	-0.02314
Target	kinase mTOR	1	---------	-0.05846
Target	Peroxisome proliferator-activated receptors	1	---------	0.00390
Target	Proteasome subunit beta	1	---------	0.00923
Target	Retinoic acid receptor	3	0.01861	0.01548
Target	Serine/threonine-protein kinase B-raf	1	---------	-0.00551
Target	Topoisomerase-non-Human	6	0.00879	-0.00129
Target	Tumor necrosis factor	5	0.00965	-0.00473
Target	Tyrosine kinase	13	0.01001	-0.00917
Target	Vitamin D3 receptor	1	---------	0.00537
Target	DNA cross-linking/alkylation	10	0.00989	-0.00067
Target	DNA/ligase	6	0.00368	0.00084
Target	DNA/methyltransferase	2	0.00561	0.00305
Target	DNA/polymerase	2	0.01131	-0.01287
Target	DNA/topoisomerase-human	13	0.00915	-0.00053
Target	Microtubules	5	0.00665	0.01363
Target	Thymidylate synthase	4	0.00182	-0.00033
Target	Nucleotide synthesis	5	0.00334	-0.00323
ATC Classification	Tetracyclines	2	0.00144	0.00134
ATC Classification	Immunosuppressants	9	0.00803	-0.00360
ATC Classification	Blood glucose lowering drugs, excl. insulins	2	0.00136	0.00294
ATC Classification	Hormone antagonists and related agents	6	0.01166	0.00343
ATC Classification	Adrenergics, inhalants	3	0.00292	0.00816
ATC Classification	Beta blocking agents	1	---------	-0.02369
ATC Classification	Other antineoplastic agents	26	0.01008	-0.00467

Table 5 (continued)

Category	Subgroups	Size	SD	Mean
ATC Classification	Anti-inflammatory and antirheumatic products, non-steroids	8	0.01331	-0.00293
ATC Classification	Estrogens	2	0.00221	-0.00327
ATC Classification	Corticosteroids for systemic use, plain	48	0.02451	-0.00715
ATC Classification	Lipid modifying agents, plain	1	---------	-0.00594
ATC Classification	Direct acting antivirals	1	---------	-0.04285
ATC Classification	Insulins and analogues	2	0.00157	-0.00259
ATC Classification	Immunostimulants	5	0.03866	-0.02314
ATC Classification	Other dermatological preparations	1	---------	-0.05846
ATC Classification	Anti-acne preparations for systemic use	2	0.02613	0.01675
ATC Classification	Quinolone antibacterials	6	0.00879	-0.00129
ATC Classification	Vitamin a and d, incl. combinations of the two	1	---------	0.00537
ATC Classification	Alkylating agents	4	0.01093	0.00502
ATC Classification	Cytotoxic antibiotics and related substances	19	0.00775	-0.00010
ATC Classification	Antimetabolites	13	0.00652	-0.00286
ATC Classification	Plant alkaloids and other natural products	5	0.00665	0.01363
TF Level	non-Human	9	0.01574	-0.00533
TF Level	not-directly	98	0.01443	-0.00344
TF Level	Directly	60	0.02310	-0.00433
Disease Classification	Certain infectious or parasitic diseases	10	0.01500	-0.00600
Disease Classification	Diseases of the blood or blood-forming organs	6	0.03746	-0.02368
Disease Classification	Diseases of the circulatory system	1	---------	-0.02369
Disease Classification	Diseases of the digestive system	2	0.00078	-0.00297
Disease Classification	Diseases of the immune system	10	0.01758	-0.00723
Disease Classification	Diseases of the musculoskeletal system or connective tissue	26	0.00768	-0.00170
Disease Classification	Diseases of the nervous system	4	0.03648	-0.03264
Disease Classification	Diseases of the respiratory system	5	0.00792	0.00300
Disease Classification	Diseases of the skin	16	0.02951	-0.00569
Disease Classification	Endocrine, nutritional or metabolic diseases	5	0.00403	-0.00105
Disease Classification	Neoplasms	82	0.01394	-0.00103
chemotherapy	NO	120	0.02026	-0.00556
chemotherapy	YES	47	0.00894	0.00048

"---------" indicates this subgroup's sample size is one and standard deviation cannot be calculated.

Table 6 The results of SJI score of random drug-disease pairs' subgroups

Category	Subgroups	Size	SD	Mean
Target	16S ribosomal RNA	46	0.00587	0.00106
Target	Aminoimidazole caboxamide ribonucleotide transformylase	92	0.00741	-0.00223
Target	AMP-activated protein kinase	23	0.01326	-0.00637
Target	Androgen receptor	23	0.02626	-0.00085
Target	Beta adrenergic receptor	92	0.02759	-0.00149
Target	CD20 antigen	69	0.01436	0.00118
Target	Cyclooxygenase	184	0.00887	-0.00064
Target	Cytochromes P450	69	0.02139	-0.00440
Target	Delta-aminolevulinic acid dehydratase	23	0.00706	-0.00008
Target	Estrogen receptor	92	0.01238	0.00637
Target	Glucocorticoid receptor	1104	0.01943	-0.00194
Target	HMG-CoA reductase	23	0.02600	-0.00562
Target	Inosine-5'-monophosphate dehydrogenase	23	0.01932	-0.00028
Target	Insulin receptor	46	0.01103	0.00196
Target	Interferon receptor	115	0.02849	0.00916
Target	kinase mTOR	23	0.01580	-0.00353
Target	Peroxisome proliferator-activated receptors	23	0.01289	0.00118
Target	Proteasome subunit beta	23	0.01031	-0.00219
Target	Retinoic acid receptor	69	0.01689	0.00131
Target	Serine/threonine-protein kinase B-Raf	23	0.00903	-0.00396
Target	Topoisomerase-non-Human	138	0.01749	-0.00121
Target	Tumor necrosis factor	115	0.01891	-0.00753
Target	Tyrosine kinase	299	0.01544	-0.00471
Target	Vitamin D3 receptor	23	0.01184	-0.00267
Target	DNA cross-linking/alkylation	230	0.01040	0.00047
Target	DNA/ligase	137	0.00165	-0.00056
Target	DNA/methyltransferase	44	0.01455	-0.00180
Target	DNA/polymerase	44	0.01139	0.00073
Target	DNA/topoisomerase-human	286	0.01900	-0.00090
Target	Microtubules	110	0.01101	0.00258
Target	Thymidylate synthase	88	0.00329	-0.00152
Target	Nucleotide synthesis	110	0.00595	-0.00096
ATC Classification	Tetracyclines	46	0.00587	0.00106
ATC Classification	Immunosuppressants	207	0.01623	-0.00580
ATC Classification	Blood glucose lowering drugs, excl. insulins	46	0.01357	-0.00260
ATC Classification	Hormone antagonists and related agents	138	0.02098	-0.00202
ATC Classification	Adrenergics, inhalants	69	0.02579	-0.00083
ATC Classification	Beta blocking agents	23	0.02941	-0.00213
ATC Classification	Other antineoplastic agents	598	0.01379	-0.00298

Table 6 (continued)

Category	Subgroups	Size	SD	Mean
ATC Classification	Anti-inflammatory and antirheumatic products, non-steroids	184	0.00887	-0.00064
ATC Classification	Estrogens	46	0.01443	0.00869
ATC Classification	Corticosteroids for systemic use, plain	1104	0.01943	-0.00194
ATC Classification	Lipid modifying agents, plain	23	0.02600	-0.00562
ATC Classification	Direct acting antivirals	23	0.01932	-0.00028
ATC Classification	Insulins and analogues	46	0.01103	0.00196
ATC Classification	Immunostimulants	115	0.02849	0.00916
ATC Classification	Other dermatological preparations	23	0.01580	-0.00353
ATC Classification	Anti-acne preparations for systemic use	46	0.01748	0.00257
ATC Classification	Quinolone antibacterials	138	0.01749	-0.00121
ATC Classification	Vitamin a and d, incl. combinations of the two	23	0.01184	-0.00267
ATC Classification	Alkylating agents	92	0.01068	0.00120
ATC Classification	Cytotoxic antibiotics and related substances	423	0.01550	-0.00079
ATC Classification	Antimetabolites	286	0.01150	-0.00108
ATC Classification	Plant alkaloids and other natural products	110	0.01101	0.00258
TF Level	non-Human	207	0.01627	-0.00057
TF Level	not-directly	2224	0.01785	-0.00116
TF Level	Directly	1378	0.01671	0.00070
Disease Classification	Certain infectious or parasitic diseases	230	0.00999	0.00031
Disease Classification	Diseases of the blood or blood-forming organs	138	0.02470	0.00075
Disease Classification	Diseases of the circulatory system	23	0.01064	-0.00186
Disease Classification	Diseases of the digestive system	46	0.00555	-0.00001
Disease Classification	Diseases of the immune system	230	0.02018	0.00376
Disease Classification	Diseases of the musculoskeletal system or connective tissue	598	0.01129	-0.00199
Disease Classification	Diseases of the nervous system	92	0.01528	-0.00054
Disease Classification	Diseases of the respiratory system	115	0.01108	-0.00278
Disease Classification	Diseases of the skin	367	0.02185	-0.00204
Disease Classification	Endocrine, nutritional or metabolic diseases	115	0.00930	-0.00084
Disease Classification	Neoplasms	1855	0.01789	-0.00117
chemotherapy	NO	2760	0.01872	-0.00086
chemotherapy	YES	1049	0.01221	-0.00022

Table 7 Important results of SJI Score of drug-disease pairs' subgroups

Classification Category	Subgroups	Average SJI of Indicated Pairs ± SD	N	Average SJI of Control Pairs ± SD	N	Q value
Disease classification	Diseases of the blood or blood-forming organs	-0.02368±0.03746	6	0.00075±0.02470	138	0.01322
	Diseases of the nervous system	-0.03264±0.03648	4	-0.00054±0.01528	92	0.00704
Drug target classification	Interferon receptor	-0.02314±0.03866	5	0.00916±0.02849	115	0.00110
	kinase mTOR	-0.05846± ---------	1	0.00353±0.01580	23	0.01755
Chemotherapy classification	Chemotherapy drugs	0.00048±0.00894	47	-0.00022±0.01221	1049	0.99509
	Non-chemotherapy drugs	-0.00556±0.02026	120	-0.00086±0.01872	2760	0.03937
ATC classification	Immunostimulants	-0.02314±0.03866	5	0.00916±0.02849	115	0.00110
	Other dermatological preparations	-0.05846± ---------	1	-0.00353±0.01580	23	0.01755
Transcription factor level	Directly	-0.00433±0.02310	60	0.00070±0.01671	1378	0.22309
	Not-directly	-0.00344±0.01443	98	-0.00116±0.01785	2224	0.99509
	Non-Human	-0.00533±0.01574	9	-0.00057±0.01627	207	0.79080

"---------" indicates the subgroup's sample size is one and standard deviation cannot be calculated.

3.5 Gene and pathway analysis on an example drug-disease GES pair

"Interferon receptor" (Same drug-disease pair content as the "Immunostimulants" subgroup), the subgroup with the lowest Q value, is chosen as a case report of pathway analysis. The top 5% (93/1898) genes with relatively reversed expression probability according to $G^{I\text{-}R\%}$ scores are showed in **Table 8**. Top 10 significant biological pathways identified by IPA are showed in **Table 9**.

Table 8 Top Top 5% genes with relatively expression probability $G^{I\text{-}R\%}$

Gene	$G^{I\text{-}R\%}$	Gene	$G^{I\text{-}R\%}$	Gene	$G^{I\text{-}R\%}$	Gene	$G^{I\text{-}R\%}$
MX1	-46.87%	FTL	-25.22%	USP18	-19.56%	DUSP6	-16.90%
IFIT3	-41.45%	RPL24	-25.18%	CERS2	-19.38%	TPT1	-16.66%
NME1	-40.50%	ERP29	-23.86%	RPLP0	-19.36%	RSAD2	-16.59%
RPL3	-39.19%	RSL24D1	-23.86%	KLRB1	-19.28%	ADAR	-16.48%
RPS5	-37.61%	PTMA	-23.65%	ADM	-19.23%	DDX58	-16.44%
RPL6	-36.57%	HLA-DRA	-22.88%	PLSCR1	-19.23%	APOBEC3A	-16.40%
MT1HL1	-35.52%	IFIT1	-22.22%	RPLP0P6	-19.14%	PPIB	-16.17%
MT2A	-34.80%	MX2	-22.22%	RPS3A	-19.07%	RGS2	-16.09%
RPSA	-33.55%	LDHB	-22.12%	TRIM22	-19.00%	IRF7	-16.08%
TGFBI	-33.47%	DYNLT1	-21.90%	DDX21	-18.66%	PSMA6	-16.00%
MT1X	-32.30%	ALDH1A1	-21.64%	GCH1	-18.64%	RPL9	-15.94%
HERC5	-32.15%	HSPA1A	-21.53%	GAPDH	-18.55%	OAS1	-15.91%
FAU	-31.82%	SLC25A5	-21.53%	OAS3	-18.48%	RPL31	-15.74%
PLS3	-29.66%	IFIT2	-21.38%	RPS25	-18.40%	PTTG1IP	-15.74%
HLA-A	-29.15%	RPS4X	-21.28%	NDUFB11	-18.40%	BIRC2	-15.74%
RPL22	-28.88%	EIF3E	-20.88%	SNHG6	-18.15%	MYD88	-15.67%
FBL	-28.52%	HMGN2	-20.88%	PSAT1	-18.06%	RPS14P3	-15.64%
RPS8	-27.57%	FTH1P5	-20.80%	IER2	-18.02%	FTH1	-15.62%
ISG15	-26.91%	YWHAZ	-20.72%	UXT	-17.65%	C4orf46	-15.45%
EEF1B2	-26.88%	PFDN5	-20.57%	PARP12	-17.58%	PPT1	-15.42%
PHB2	-26.48%	TMA7	-20.20%	MAFB	-17.40%	YBX1	-15.33%
MT1H	-26.29%	CCT7	-20.12%	LYZ	-17.25%		
RPL8	-26.11%	OASL	-19.89%	NARS	-17.15%		
ATF4	-25.36%	SNHG5	-19.64%	AKR1B1	-17.02%		

Table 9 Top 10 significant biological pathways according to high relatively expression probability genes

Ingenuity Canonical Pathways	-log(p-value)	Ratio	Genes overlapped with datasets
EIF2 Signaling	16.50	8.02% (17/212)	ATF4, EIF3E, FAU, RPL22, RPL24, RPL3, RPL31, RPL6, RPL8, RPL9, RPLP0, RPS25, RPS3A, RPS4X, RPS5, RPS8, RPSA
Activation of IRF by Cytosolic Pattern Recognition Receptors	6.60	9.84% (6/61)	ADAR, DDX58, IFIT2, IRF7, ISG15, PPIB
Regulation of eIF4 and p70S6K Signaling	6.48	5.23% (8/153)	EIF3E, FAU, RPS25, RPS3A, RPS4X, RPS5, RPS8, RPSA
Interferon Signaling	6.34	13.90% (5/36)	IFIT1, IFIT3, ISG15, MX1, OAS1
mTOR Signaling	5.57	3.96% (8/202)	EIF3E, FAU, RPS25, RPS3A, RPS4X, RPS5, RPS8, RPSA
NRF2-mediated Oxidative Stress Response	3.80	3.23% (6/186)	ATF4, CCT7, ERP29, FTH1, FTL, PPIB
Role of Pattern Recognition Receptors in Recognition of Bacteria and Viruses	3.39	3.47% (5/144)	DDX58, IRF7, MYD88, OAS1, OAS3
Neuroinflammation Signaling Pathway	2.78	2.06% (6/291)	ATF4, BIRC2, HLA-A, HLA-DRA, IRF7, MYD88
SPINK1 General Cancer Pathway	2.63	4.92% (3/61)	MT1H, MT1X, MT2A
Systemic Lupus Erythematosus in B Cell Signaling Pathway	2.23	1.89% (5/265)	IFIT2, IFIT3, IRF7, ISG15, MYD88

All of the top 10 pathways showed in the table are reported to be related with interferon regulation[67-79]. These pathways are mostly functioning with interferon in inflammatory and immune (See **Table 10**).

Table 10. Top 10 Pathways and their functions

Ingenuity Canonical Pathways	Function	Reference
EIF2 Signaling	Immune Responses	[80]
Activation of IRF by Cytosolic Pattern Recognition Receptors	Regulate Interferon	[69]
Regulation of eIF4 and p70S6K Signaling	Inflammatory	[70, 81]
Interferon Signaling	Immune Responses	[82, 83]
mTOR Signaling	Immune Responses	[71]
NRF2-mediated Oxidative Stress Response	Antioxidant Response	[73]
Role of Pattern Recognition Receptors in Recognition of Bacteria and Viruses	Regulate Interferon	[74]
Neuroinflammation Signaling Pathway	Inflammatory	[75]
SPINK1 General Cancer Pathway	Cancer Diagnose	[84]
Systemic Lupus Erythematosus in B Cell Signaling Pathway	Inflammatory	[85]

4.0 Discussion and Conclusion

It's well recognized that similar gene expression patterns are supposed to reflect a similar function[86]. From the overall mean score, we can see that those FDA approved drugs listed in CREEDS database and diseases they indicated generally have inverse GES patterns compared to random controls. This may imply that even a simple GES score-guided drug repositioning study may have the chance to find new potentially therapeutic use of existing drugs. However, the absolute difference between the indicated group and random control group is not obvious. As is reported in a recent study[87], the authors investigated the relationship between drug-disease GES similarity and drug therapeutic effect using Connectivity Map[88], and they also showed a result with significant statistical value($p=0.03$) with a relatively low overall Area Under Curve (AUC) of 0.57, which indicated that this inverse relationship may be real but relatively weak. This phenomenon may be attributed to that the effectors of drug treatment are likely to be the protein products of the genes, and there is only a moderate correlation between gene expression and levels of the corresponding proteins[89]. Therefore, a study or analysis aiming to find out which kind of drugs or diseases have a better association between gene expression and pharmacological effect or symptom is necessary. In our analysis, some subgroups of drugs-diseases pairs with indication associations have positive similarity mean scores (which means, this drug may exacerbate the disease according to the assumption on similarity of gene expression signatures) or mean scores higher than random drug-disease pairs, but are not statistically significant. Additionally, 7 of 70 subgroups (10.0%) showed significantly lower similarity mean scores when drug-disease association is indicated. This study may provide some hints for the future studies on utilizing GES-based method for drug repositioning. That is, some certain types of drugs may have a stronger

ability to reverse the GES of the diseases they intent to treat. Also, the type of diseases may influence this ability, too. As such, in specific kinds of subgroups, the drug-disease pairs with higher opposite similarities probably have higher chances of potential therapeutic relationships, which means focusing on certain kinds of diseases or drugs can increase the true positive rate of the GES-guided drug repositioning method.

Such as, over a half (4/7) of the significant subgroups (Immunostimulants, Interferon receptor, other dermatological preparations, and diseases of the blood or blood-forming organs) are related to diseases associated with immune system (The disease include in "other dermatological preparations" is atopic dermatitis). This indicates a drug with drug-disease pairs that associated with immune system tends to perform lower similarity scores when compared diseases it indicated with random diseases. This means in a GES-guided drug repositioning analysis, an immune-associated drug is more likely to have a potential therapeutic effect on diseases that have a higher inverse similarity with it. Also, chemotherapy drugs may not be a good area for GES-guided drug repositioning method as the mechanism of these kinds of drugs is not selective and its similarity scores show no significance. On the other hand, non-chemotherapy drugs show a significant Q-value (0.03937).

To our surprise, the TF-level of the drugs is not a significant factor that affects by the indication relationship (FDR Q-value 0.22309 in GLM least squares mean partitions F tests). This indicates that a drug's ability to disturb the regulation of gene expression may not reflect its ability to reverse the GES of its indicated diseases.

For the case of "interferon receptor" subgroup biological pathway analysis, as the genes involved to construct these pathways are genes with the lowest $G^{I\text{-}R\%}$ scores in the subgroup with most significant indicated-random drug-disease pairs' SJI difference, it is reasonable that GES-

guided drug repositioning methods are more sensitive (have a higher true positive rate) to drugs or diseases which these pathways involved in. Also, the significance of "mTOR Signaling" accordance with the result which subgroup "kinase mTOR" has a significant indicated-random drug-disease pairs' SJI difference. This result confirmed the high sensitivity of GES-guided drug repositioning method to this pathway on the other side.

We can notice that the standard deviation of the method is relatively high, which indicated the similarity scores are not very stable. To exclude the possibility that this unstable result is due to the scoring method (SJI), we calculated the similarity score with another commonly used method called connectivity score[88] as a reference. The results are shown in **Table 11**.

Table 11 The defference of results between SJI and connectivity score

Groups	Mean of Signed Jaccard Index	SD	Mean of Connectivity Score	SD
Indicated group	-0.00386	0.01794	-0.03582	0.18435
Random controls	-0.00072	0.01750	-0.01152	0.21524

We can see from the table that fold of the scale between the score mean and the standard deviation is quite similar among these two scoring methods. The reference score result (connectivity score) indicated that the high standard deviation of the similarity scores may not be caused by a certain scoring method. As other study reported[55, 90], this high variation may due to gene expression measurement method itself with a low reproducibility.

Unavoidably, there are some limitations in this study. First, the tissues used for testing the drug effects may not match with the body parts/organs where the diseases affected. Second, in real biological process, the weights of each gene apparently are not the same. However, it is not practical to estimate the weight of every gene in all therapeutic relationships appeared in this study. Also, is it unreasonable to measure the weight for random control pair as the therapeutic relationships do not even exist. Third, some bias may be caused by limited number of CREEDS

bio-assay collection which may not have the ability to fully present the patterns of all kinds of drugs and diseases. What's more, there are different types of "treatment effect". Some drugs may actually cure the diseases and others may just provide symptomatic relief. As such, drugs may also induce a different pattern of GES compared with diseases. Also, some indicated subgroups ("kinase mTOR" and "other dermatological preparations") have so few unique drug-disease pairs (n=1) that may impair the power of the analysis. This may result from the strict indication criterion. It is hard to say that all the random control pairs do not have strong therapeutic relationship comparable to the indicated pairs. It is easy to find reports of drugs with their off-label usage. However, due to the large quantity and varying quality of the drug experiment reports, to increase the sample size is not possible until we could find a feasible way to figure out all off-label usage with the same criterion of evidence power.

In this study, we systematically analyzed the similarity of gene expression profiles from known drug-disease associations and we found that indicated pairs do show a more inversed overall similarly score. Also, we found 7 subgroups of which drugs or diseases may have a more reversed pattern when there is a clear therapeutic effect. That means a GES-guided drug repositioning method should be used with more caution based on drug or disease type differences. That is, drugs or diseases associated with immune system or non-chemotherapy drugs may have a higher true positive rate. And, as the case of our biological pathway enrichment analysis, some certain pathways may be also more sensitive to this method, such as "mTOR Signaling" pathway.

Appendix A CREEDS signatures

Information of 230 manual disease signatures and 244 manual drug perturbation signatures

Appendix Table 1. Information of manual huaman drug perturbation signatures

DRID	Cell type	Drug name	GEO ID
drug:3474	PC-3 cells	Gefitinib	GSE53180
drug:3233	NB4 APL 9 acute promyelocytic leukemia cells - 72 Hours	Tretinoin	GSE23702
drug:3232	NB4 APL 9 acute promyelocytic leukemia cells - (TG2-knockdown) 72 Hours	Tretinoin	GSE23702
drug:3231	NB4 APL 9 - acute promyelocytic leukemia cells - (TG2-knockdown) 48 Hours	Tretinoin	GSE23702
drug:3237	MCF-7 breast cancer cells	Tretinoin	GSE32161
drug:3300	No Data	Abiraterone	GSE49244
drug:3160	OVCAR-8 ovarian carcinoma cell line	Cisplatin	GSE47856
drug:2786	NA	Metoprolol	GSE3356
drug:2542	H1299	Azacitidine	GSE29077
drug:2543	HT29	Methotrexate	GSE11440
drug:2540	A549	Decitabine	GSE29077
drug:2546	H1299	Decitabine	GSE29077
drug:2547	H1299	Decitabine	GSE29077
drug:2544	H1299	Azacitidine	GSE29077
drug:2545	H1299	Azacitidine	GSE29077
drug:2548	H1299	Decitabine	GSE29077
drug:2763	NA	Imatinib	GSE23743
drug:2723	NA	Doxorubicin	GSE763
drug:2481	uninfected hepatoma Huh7.5.1 cells	Ribavirin	GSE23031
drug:3587	Lymphoblastoid cell line (Bleomycin-insensitive)	Bleomycin	GSE3598
drug:3058	HL60 promyelocytic leukemia cell line	Diclofenac	GSE28185
drug:2591	plaque on epidermis	Etanercept	GSE47751
drug:2593	plaque on epidermis	Etanercept	GSE47751
drug:3055	SLM2 melanoma cells	Sorafenib	GSE39192
drug:3252	SKBR3 parental and resistant (SKBR3-R) cell lines	Lapatinib	GSE38376
drug:3588	Lymphoblastoid cell line (Bleomycin-sensitive)	Bleomycin	GSE3598
drug:3158	C13 ovarian carcinoma cell line	Cisplatin	GSE47856
drug:3159	IGROV-1 ovarian carcinoma cell line	Cisplatin	GSE47856
drug:2541	PBMC	Methotrexate	GSE23687
drug:2577	CD19+ selected B cells	Rituximab	GSE15490
drug:2576	left internal mammary artery	Paclitaxel	GSE19136

Appendix Table 1 (continued)

DRID	Cell type	Drug name	GEO ID
drug:2575	NCI-H69	Aminolevulinic acid	GSE8920
drug:2574	Skin	Pimecrolimus	GSE32473
drug:2573	C33KD2 cells	Doxycycline	GSE11422
drug:2570	M238_R1 melanoma resistant sub-line	Plx4032	GSE24862
drug:3129	Mtb H37Rv-infected THP-1 macrophages	Calcitriol	GSE52819
drug:2498	skin cells	Isotretinoin	GSE10433
drug:2681	NA	Bexarotene	GSE12791
drug:3155	OVCA433 ovarian carcinoma cell line	Cisplatin	GSE47856
drug:3063	MCF7/BUS human breast cancer cells - 60 pM	Estradiol	GSE4668
drug:2594	plaque on epidermis	Etanercept	GSE47751
drug:2597	whole-blood leukocytes	Etanercept	GSE36177
drug:3148	PA-1 ovarian carcinoma cell line	Cisplatin	GSE47856
drug:2596	whole-blood leukocytes	Etanercept	GSE36177
drug:3062	MCF7/BUS human breast cancer cells - 30 pM	Estradiol	GSE4668
drug:2652	MCF-7 breast cancer (BC) cells	Estradiol	GSE53394
drug:3368	Skeletal muscle - Vastus lateralis muscle biopsies - from insulin-sensitive subjects	Insulin	GSE22309
drug:2560	SW1736 thyroid cancer cell line (vemurafenib-refractory cell line) - 6 Hours	Vemurafenib	GSE37441
drug:2561	SW1736 thyroid cancer cell line (vemurafenib-refractory cell line) - 48 Hours	Vemurafenib	GSE37441
drug:2562	SK-MEL-28 melanoma cell line (vemurafenib sensitive cell line) - 1 Hour	Vemurafenib	GSE37441
drug:2563	SK-MEL-28 melanoma cell line (vemurafenib sensitive cell line) - 6 Hours	Vemurafenib	GSE37441
drug:2564	SK-MEL-28 melanoma cell line (vemurafenib sensitive cell line) - 48 Hours	Vemurafenib	GSE37441
drug:2566	M249_R4 melanoma resistant sub-line	Plx4032	GSE24862
drug:2567	M249 melanoma cell line	Plx4032	GSE24862
drug:3140	ME180 squamous cell carcinoma cell line - 3 hours 10Î¼M	Tretinoin	GSE54464
drug:2740	hme-cc	Fluorouracil	GSE1647
drug:2741	mcf-7	Fluorouracil	GSE1647
drug:2742	me16c	Fluorouracil	GSE1647
drug:2487	Breast adenocarcinoma MCF-7 cell line	Estradiol	GSE46924
drug:2743	zr-75-1	Fluorouracil	GSE1647
drug:2484	Primary umbilical vein endothelial cell	Atorvastatin	GSE2450
drug:2483	Human primary fibroblasts (IMR90) stably expressing H-RasV12	Metformin	GSE33612
drug:3072	HT29 colon adenocarcinoma cell line	Methotrexate	GSE11440
drug:3075	HeLa cells - 1 Hour	Doxycycline	GSE2624
drug:2691	NA	Chlorambucil	GSE8832
drug:3074	HeLa cells - 0 Hour	Doxycycline	GSE2624

Appendix Table 1 (continued)

DRID	Cell type	Drug name	GEO ID
drug:3076	HeLa cells - 3 Hours	Doxycycline	GSE2624
drug:3374	A673	Cytarabine	GSE6930
drug:3370	Skeletal muscle - Vastus lateralis muscle biopsies - from diabetic patients	Insulin	GSE22309
drug:2519	K562 leukemia cell line (II) - 24 Hours	Imatinib	GSE1922
drug:2517	K562 leukemia cell line (III) - 24 Hours	Imatinib	GSE1922
drug:2516	K562 leukemia cell line (IV) - 24 Hours	Imatinib	GSE1922
drug:2510	Chronic myelogenous leukemia CD34+CD38- cells	Imatinib	GSE20876
drug:2513	K562 leukemia cell line (VII) - 24 Hours	Imatinib	GSE1922
drug:2512	K562 leukemia cell line (VIII)	Imatinib	GSE1922
drug:2764	NA	Imatinib	GSE24493
drug:2734	Eralpha	Estradiol	GSE1153
drug:2731	NA	Estradiol	GSE8597
drug:2641	MCF-7 BREAST CANCER cell line	Doxorubicin	GSE13477
drug:3424	A673	Cytarabine	GSE6930
drug:2714	NA	Dexamethasone	GSE34313
drug:2624	primary human hepatocytes (PHH)	Ciprofloxacin	GSE9166
drug:3346	MCF7 cells depleted of ERK2 - 4 Hours after treatment	Estradiol	GSE24592
drug:3306	Pancreatic Cancer Cell Lines	Dasatinib	GSE59357
drug:3296	Colon	Celecoxib	GSE11237
drug:3343	Peripheral mononuclear blood cells (NAB+) - 3 Months of treatment	Interferon beta-1b	GSE26104
drug:3347	MCF7 cells depleted of ERK2 - 24 Hours after treatment	Estradiol	GSE24592
drug:3169	Airway smooth muscle cells	Dexamethasone	GSE34313
drug:3344	MCF7 cells depleted of ERK1 - 4 Hours after treatment	Estradiol	GSE24592
drug:2508	Primary leukemic cells (purified from peripheral blood) - 48 Hours	Vemurafenib	GSE63790
drug:2509	Primary leukemic cells (purified from peripheral blood) - 72 Hours	Vemurafenib	GSE63790
drug:2501	AGS cells	Celecoxib	GSE54657
drug:2720	NA	Doxorubicin	GSE12972
drug:2721	NA	Doxorubicin	GSE763
drug:3123	SKOV3 ovarian cancer xenograft tumor	Trastuzumab	GSE31432
drug:3126	Macrophages	Calcitriol	GSE52819
drug:3124	SKOV3 ovarian cancer xenograft tumor	Pertuzumab	GSE31432
drug:2603	lesioned skin	Etanercept	GSE41663
drug:2601	PBMC (peripheral blood mononuclear cells)	Methotrexate	GSE41831
drug:3422	A673	Cytarabine	GSE6930
drug:2496	LY2	Tamoxifen	GSE28645
drug:2533	MCF-7/ADR	Doxorubicin	GSE24460
drug:2532	A549	Cisplatin	GSE6410

Appendix Table 1 (continued)

DRID	Cell type	Drug name	GEO ID
drug:2536	A549	Azacitidine	GSE29077
drug:2535	A549	Azacitidine	GSE29077
drug:2534	A549	Azacitidine	GSE29077
drug:2539	A549	Decitabine	GSE29077
drug:3184	Peripheral mononuclear blood cells (NAB+) - 12 months	Interferon beta-1b	GSE26104
drug:3039	RT112 cancer cell line (FGFR3 knocked down with shRNA 2-4)	Doxycycline	GSE41035
drug:3141	ME180 squamous cell carcinoma cell line - 6 hours 10Î¼M	Tretinoin	GSE54464
drug:2538	A549	Decitabine	GSE29077
drug:2526	GIST882 cells	Imatinib	GSE22433
drug:2527	K562 cells	Imatinib	GSE19567
drug:2520	K562 leukemia cell line (I) - 24 Hours	Imatinib	GSE1922
drug:2521	Chronic myelogenous leukemia CD34+ cells	Imatinib	GSE1418
drug:2528	K562 cells	Nilotinib	GSE19567
drug:3149	TYK-nu ovarian carcinoma cell line	Cisplatin	GSE47856
drug:2810	NA	Rosiglitazone	GSE7035
drug:3146	OVCA420 ovarian carcinoma cell line	Cisplatin	GSE47856
drug:2730	NA	Estradiol	GSE11506
drug:2474	lung epithelial A549 cells, 24h treatment	Natural alpha interferon	GSE5542
drug:3139	ME180 squamous cell carcinoma cell line - 1 hour 10Î¼M	Tretinoin	GSE54464
drug:3394	Endothelial cells (cultured umbilical vein endothelial cells) - treated with 1 nmol/L estradiol for 24 hours	Estradiol	GSE16683
drug:2461	HL60 cells	Tretinoin	GSE5007
drug:2606	skin lesion	Etanercept	GSE11903
drug:3375	A673	Cytarabine	GSE6930
drug:3373	A673	Cytarabine	GSE6930
drug:2584	lapatinib-sensitive ErbB2-positive cells treated with 1 uM lapatinib	Lapatinib	GSE38376
drug:2585	lapatinib-resistant ErbB2-positive cells treated with 0.1 uM lapatinib	Lapatinib	GSE38376
drug:2583	lapatinib-sensitive ErbB2-positive cells treated with 0.1 uM lapatinib	Lapatinib	GSE38376
drug:2580	BT474 (lapatinib-sensitive) HER2+ breast cancer cells	Lapatinib	GSE16179
drug:2581	BT474-J4 (acquired lapatinib-resistance) HER2+ breast cancer cells	Lapatinib	GSE16179
drug:2733	NA	Estradiol	GSE26834
drug:3238	paclitaxel-resistant MDA-MB-231 cancer cells	Paclitaxel	GSE12791
drug:2518	Philadelphia chromosome positive CML CD34+ cells	Imatinib (glivec)	GSE12211
drug:3093	OE-E6/7 cells	Dexamethasone	GSE54608
drug:3097	breast cancer biopsies	Letrozole	GSE5462

Appendix Table 1 (continued)

DRID	Cell type	Drug name	GEO ID
drug:3096	NCI-H460 human lung large cell carcinoma cell line - 24 Hours	Cisplatin	GSE42172
drug:3095	NCI-H460 human lung large cell carcinoma cell line - 2 Hours	Cisplatin	GSE42172
drug:3098	A549 lung cancer cells	Actinomycin d	GSE6400
drug:2514	synovial knee tissue	Rituximab	GSE24742
drug:2477	lung epithelial A549 cells, 24h treatment	Interferon gamma-1b	GSE5542
drug:3423	A673	Cytarabine	GSE6930
drug:3185	Peripheral mononuclear blood cells (NAB+) - 24 months	Interferon beta-1b	GSE26104
drug:2735	NA	Estradiol	GSE1153
drug:2732	NA	Estradiol	GSE24592
drug:3187	Peripheral mononuclear blood cells (NAB-) 12 months	Interferon beta-1a	GSE26104
drug:3170	Primary smooth muscle cells	Calcitriol	GSE5145
drug:3180	KJD SV40 virus transformed epidermal keratinocyte	Doxorubicin	GSE58074
drug:3144	HeyC2 ovarian carcinoma cell line	Cisplatin	GSE47856
drug:3087	Breast Cancer tumor core biopsies	Anastrozole	GSE33658
drug:3088	DHT-stimulated LNCaP prostate cells	Bicalutamide	GSE7708
drug:3089	epithelial ovarian cancer cell line	Carboplatin	GSE13525
drug:3241	MCF-7 breast cancer cells - 24 Hours	Estradiol	GSE26834
drug:3060	NSCLC adenocarcinoma cell line	Carboplatin	GSE7035
drug:3061	MCF7/BUS human breast cancer cells - 10 pM	Estradiol	GSE4668
drug:2599	diffuse large B cell lymphoma SKI-DLCL cells	Cytarabine	GSE5681
drug:2628	primary human hepatocytes (PHH)	Levofloxacin	GSE9166
drug:2626	primary human hepatocytes (PHH)	Gatifloxacin	GSE9166
drug:2820	NA	Tamoxifen	GSE4025
drug:2828	NA	Tretinoin	GSE23702
drug:2476	lung epithelial A549 cells, 6h treatment	Interferon gamma-1b	GSE5542
drug:2686	NA	Bortezomib	GSE30931
drug:2473	lung epithelial A549 cells, 6h treatment	Natural alpha interferon	GSE5542
drug:3174	OV1002 ovarian cancer cell line - Day 2	Carboplatin	GSE49577
drug:3176	OV1002 ovarian cancer cell line - Day 1	Carboplatin	GSE49577
drug:2822	NA	Tamoxifen	GSE28645
drug:3077	HeLa cells - 6 Hours	Doxycycline	GSE2624
drug:2586	lapatinib-resistant ErbB2-positive cells treated with 1 uM lapatinib	Lapatinib	GSE38376
drug:3057	THP-1 acute monocytic leukemia cells	Diclofenac	GSE28185
drug:3053	Liver	Diclofenac	GSE54255
drug:2457	K562 leukemia cell line	Imatinib	GSE1922
drug:3188	Peripheral mononuclear blood cells (NAB-) 24 months	Interferon beta-1a	GSE26104

Appendix Table 1 (continued)

DRID	Cell type	Drug name	GEO ID
drug:3186	Peripheral mononuclear blood cells (NAB-) 3 months	Interferon beta-1a	GSE26104
drug:2458	Primary Colorectal Adenocarcinoma	Celecoxib	GSE11237
drug:2631	primary lung fibroblasts	Formoterol	GSE30242
drug:3245	lapatinib sensitive and lapatinib resistant ErbB2-positive breast cancer cells	Lapatinib	GSE38376
drug:3179	SCC25 tongue epidermal keratinocyte	Doxorubicin	GSE58074
drug:2680	NA	Bexarotene	GSE6914
drug:3260	A673	Doxorubicin	GSE6930
drug:3261	A673	Doxorubicin	GSE6930
drug:3262	A673	Doxorubicin	GSE6930
drug:3263	A673	Doxorubicin	GSE6930
drug:3264	A673	Doxorubicin	GSE6930
drug:3265	A673	Doxorubicin	GSE6930
drug:3040	RT112 cancer cell line (FGFR3 knocked down with shRNA 4-1)	Doxycycline	GSE41035
drug:3041	RT112 cancer cell line (FGFR3 knocked down with shRNA 6-16)	Doxycycline	GSE41035
drug:2515	K562 leukemia cell line (V) - 24 Hours	Imatinib	GSE1922
drug:2488	glucocorticoid (GC)-resistant lymphoblastic leukemia CEM-C7H2 T-ALL cell line	Dexamethasone	GSE22152
drug:2608	skin lesion	Etanercept	GSE11903
drug:2751	normal scar	Hydrocortisone	GSE7890
drug:2592	plaque on epidermis	Etanercept	GSE47751
drug:2568	M229_R5 melanoma resistant sub-line	Plx4032	GSE24862
drug:2569	M238 melanoma cell line	Plx4032	GSE24862
drug:2722	NA	Doxorubicin	GSE763
drug:2610	skin lesion	Etanercept	GSE11903
drug:3031	36M2 epithelial ovarian cancer cells	Carboplatin (30 h)	GSE13525
drug:3030	36M2 epithelial ovarian cancer cells	Carboplatin (24 h)	GSE13525
drug:3032	36M2 epithelial ovarian cancer cells	Carboplatin (36 h)	GSE13525
drug:2493	RWPE1 cells	1,25 dihydroxyvitamin d	GSE15947
drug:2495	BRAFV600E A375 human melanoma cells	Vemurafenib	GSE42872
drug:2497	HepG2	Decitabine	GSE5230
drug:3143	HeyA8 ovarian carcinoma cell line	Cisplatin	GSE47856
drug:3145	A2780 ovarian carcinoma cell line	Cisplatin	GSE47856
drug:3147	OVCA429 ovarian carcinoma cell line	Cisplatin	GSE47856
drug:2727	NA	Estradiol	GSE4668
drug:2565	M229 melanoma cell line	Plx4032	GSE24862
drug:2724	NA	Doxorubicin	GSE763
drug:2559	SW1736 thyroid cancer cell line (vemurafenib-refractory cell line) - 1 Hour	Vemurafenib	GSE37441

Appendix Table 1 (continued)

DRID	Cell type	Drug name	GEO ID
drug:2558	COLO829 human melanoma cell line-MITF Knockdown (BRAFV600E mutated)	Plx4032	GSE50649
drug:2605	lesioned skin	Etanercept	GSE41663
drug:3175	OV1002 ovarian cancer cell line - Day 4	Carboplatin	GSE49577
drug:2607	skin lesion	Etanercept	GSE11903
drug:3177	OV1002 ovarian cancer cell line - Day 7	Carboplatin	GSE49577
drug:3178	OV1002 ovarian cancer cell line - Day 14	Carboplatin	GSE49577
drug:2600	Granulocyte	Methotrexate	GSE41831
drug:3203	Endometrium (postmenopausal)	Estradiol	GSE12446
drug:2809	NA	Rosiglitazone	GSE5679
drug:2602	lesioned skin	Etanercept	GSE41663
drug:3152	FU-OV-1 ovarian carcinoma cell line	Cisplatin	GSE47856
drug:3153	A2008 ovarian carcinoma cell line	Cisplatin	GSE47856
drug:3150	CH1 ovarian carcinoma cell line	Cisplatin	GSE47856
drug:3151	OV90 ovarian carcinoma cell line	Cisplatin	GSE47856
drug:3156	OVCAR-10 ovarian carcinoma cell line	Cisplatin	GSE47856
drug:3154	DOV13 ovarian carcinoma cell line	Cisplatin	GSE47856
drug:2555	U87	Temozolomide	GSE43452
drug:2557	COLO829 human melanoma cell line (BRAFV600E mutated)	Plx4032	GSE50649
drug:2556	WM164 BRAF mutant Melanoma cells	Plx4032	GSE54711
drug:2772	NA	Isotretinoin	GSE10432

Appendix Table 2 information of manual huaman disease perturbation signatures

DZID	Cell type	Disease name	GEO ID
dz:595	Colon mucosa - colorectal adenocarcinomas with microsatellite instability (MSI CRCs)	colorectal adenocarcinoma	GSE24514
dz:733	Primary lung tumor and Normal lung	large cell neuroendocrine carcinoma	GSE51852
dz:749	CD34+ hematopoietic stem cells (refractory anemia with excess blasts 2)	myelodysplastic syndrome	GSE19429
dz:748	CD34+ hematopoietic stem cells (refractory anemia with excess blasts 1)	myelodysplastic syndrome	GSE19429
dz:747	CD34+ hematopoietic stem cells (refractory anemia)	myelodysplastic syndrome	GSE19429
dz:750	CD34+ hematopoietic stem cells (refractory anemia with ringed sideroblasts)	myelodysplastic syndrome	GSE19429
dz:812	HSC cells	myelodysplastic syndrome	GSE19429
dz:755	low-grade serous ovarian carcinomas (tumors from patients with matched border control)	ovarian serous carcinoma	GSE56443
dz:605	Mammary gland	ductal carcinoma in situ	GSE21422
dz:761	Colon	Colitis	GSE6731
dz:130	Muscle - Striated (Skeletal) (MMHCC)	dermatomyositis	GSE1551
dz:705	Pathological skeletal muscle fibers	dermatomyositis	GSE48280
dz:334	Muscle - Striated (Skeletal) (MMHCC)	dermatomyositis	GSE5370
dz:135	Prostate	prostate cancer	GSE3868
dz:485	LNCaP	prostate cancer	GSE39452
dz:835	Radical prostatectomy tissue samples (TMPRSS2:ERG gene fusion NEGATIVE) and Benign prostate tissue	prostate cancer	GSE55945
dz:834	Radical prostatectomy tissue samples (TMPRSS2:ERG gene fusion POSITIVE) and Benign prostate tissue	prostate cancer	GSE55945
dz:639	Normal prostate epithelial cells and Primary epithelial cell culture (Gleason 7 score tumor)	prostate cancer	GSE3868
dz:638	Normal prostate epithelial cells and Primary epithelial cell culture (Gleason 6 score tumor)	prostate cancer	GSE3868
dz:990	Prostate tissue (primary prostate tumor)	prostate cancer	GSE3325
dz:991	Prostate tissue (metastatic prostate tumor)	prostate cancer	GSE3325
dz:866	Peripheral white blood cells	prostate cancer	GSE30174
dz:603	Stromal cells	prostate cancer	GSE26910
dz:322	Peripheral blood mononuclear cell	bacterial infectious disease	GSE3026
dz:414	Dendritic cell	bacterial infectious disease	GSE4748
dz:786	B lymphocytes	chronic lymphocytic leukemia	GSE6691
dz:194	Peripheral blood mononuclear cell	chronic lymphocytic leukemia	GSE8835
dz:708	Peripheral blood mononuclear cells	sarcoidosis	GSE19314
dz:12	T lymphocyte	sarcoidosis	GSE2657
dz:163	Blood monocyte	osteoporosis	GSE2208
dz:424	Bone Marrow	aplastic anemia	GSE3807
dz:472	Fibroblasts	Diamond-Blackfan anemia	GSE14335

Appendix Table 2 (continued)

DZID	Cell type	Disease name	GEO ID
dz:916	Lungs	pulmonary sarcoidosis	GSE16538
dz:773	Peripheral blood monocytes	familial combined hyperlipidemia	GSE11393
dz:137	Lymphoblast	familial combined hyperlipidemia	GSE1010
dz:267	Blood monocyte	familial combined hyperlipidemia	GSE11393
dz:909	Monocytes - Heterozygous FH	familial hypercholesterolemia	GSE6054
dz:908	T lymphocytes - Homozygous FH	familial hypercholesterolemia	GSE6088
dz:907	T lymphocytes - Heterozygous FH	familial hypercholesterolemia	GSE6088
dz:910	Monocytes - Homozygous FH	familial hypercholesterolemia	GSE6054
dz:771	Skeletal muscle biopsies - long duration	childhood type dermatomyositis	GSE11971
dz:772	Skeletal muscle biopsies - short duration	childhood type dermatomyositis	GSE11971
dz:815	Skin	Urticaria	GSE57178
dz:392	Mammary Gland Tissue	breast cancer	GSE1379
dz:978	Sections from non-Basal Like Cancer specimens	breast cancer	GSE3744
dz:448	Breast Epithelium	breast cancer	GSE9574
dz:148	Mammary Gland Tissue	breast cancer	GSE2429
dz:602	Stromal cells	breast cancer	GSE26910
dz:504	MDA-MB-231 breast cancer cells	breast cancer	GSE14943
dz:11	Mammary Epithelium	breast cancer	GSE53
dz:478	MDA-MB231 cells	breast cancer	GSE34925
dz:24	Mammary Gland Tissue	breast cancer	GSE3744
dz:39	Epithelial Cell	breast cancer	GSE2155
dz:52	Mammary Gland Tissue	breast cancer	GSE1378
dz:483	Pancreas	pancreatic cancer	GSE18670
dz:555	Peripheral blood mononuclear cells	pancreatic cancer	GSE49515
dz:475	Pancreatic tissue	pancreatic cancer	GSE16515
dz:798	PANC-1 PANCREATIC ADENOCARCINOMA CELL LINE	pancreatic cancer	GSE23952
dz:597	Huh7 hepatoma cells (infected with JFH-1 HCV) - 6 Hours post-infection	hepatitis C	GSE20948
dz:598	Huh7 hepatoma cells (infected with JFH-1 HCV) - 12 Hours post-infection	hepatitis C	GSE20948
dz:599	Huh7 hepatoma cells (infected with JFH-1 HCV) - 18 Hours post-infection	hepatitis C	GSE20948
dz:600	Huh7 hepatoma cells (infected with JFH-1 HCV) - 24 Hours post-infection	hepatitis C	GSE20948
dz:601	Huh7 hepatoma cells (infected with JFH-1 HCV) - 48 Hours post-infection	hepatitis C	GSE20948
dz:309	Hepatocyte	hepatitis C	GSE2067
dz:326	Epidermis	Melanoma	GSE4587

Appendix Table 2 (continued)

DZID	Cell type	Disease name	GEO ID
dz:117	Epidermis	Melanoma	GSE3189
dz:950	CD8T cells (sorted peripheral blood lymphocytes)	Melanoma	GSE6887
dz:951	Natural Killer (NK) cells (sorted peripheral blood lymphocytes)	Melanoma	GSE6887
dz:949	CD4T cells (sorted peripheral blood lymphocytes)	Melanoma	GSE6887
dz:948	B cells (sorted peripheral blood lymphocytes)	Melanoma	GSE6887
dz:418	Renal Tissue	nephroblastoma	GSE2712
dz:502	HT29 Colo205	colon cancer	GSE34299
dz:245	Intestine - Large Intestine - Colon (MMHCC)	colon cancer	GSE4107
dz:377	B Cell Lymphocyte	multiple sclerosis	GSE10064
dz:746	Peripheral blood mononuclear cells	multiple sclerosis	GSE23832
dz:743	peripheral blood mononuclear cells	multiple sclerosis	GSE21942
dz:738	brain lesion (MS after inflammation - late stage)	multiple sclerosis	GSE38010
dz:737	brain lesion (MS after demyelination - active inflammation)	multiple sclerosis	GSE38010
dz:879	CD4+ T cells Lymphocytes (sorted from PBMCs from monozygotic twin pairs discordant for RRMS)	relapsing-remitting multiple sclerosis	GSE16461
dz:611	ovarian epithelial cells	ovarian cancer	GSE14407
dz:827	Bronchial epithelial cells - Female - Mild asthma	Asthma	GSE43696
dz:753	Airway epithelial cells	Asthma	GSE18965
dz:712	Circulating CD4+ T-cells (severe asthma)	Asthma	GSE31773
dz:565	white blood cells - Male Mild Asthma	Asthma	GSE27011
dz:567	white blood cells - Female Severe Asthma	Asthma	GSE27011
dz:566	white blood cells - Female Mild Asthma	Asthma	GSE27011
dz:568	white blood cells - Male Severe Asthma	Asthma	GSE27011
dz:635	Male bronchial epithelial cell	Asthma	GSE43696
dz:828	Bronchial epithelial cells - Female - Severe asthma	Asthma	GSE43696
dz:829	Bronchial epithelial cells - Male - Mild asthma	Asthma	GSE43696
dz:830	Bronchial epithelial cells - Male - Severe asthma	Asthma	GSE43696
dz:713	circulating CD4+ T-cells (non-severe asthma)	Asthma	GSE31773
dz:729	PBMC	Asthma	GSE16032
dz:634	Female bronchial epithelial cell	Asthma	GSE43696
dz:467	bronchial epithelial cells	Asthma	GSE43696
dz:318	Epithelial Cell	Asthma	GSE4302
dz:714	Circulating CD8+ T-cells (severe asthma)	Asthma	GSE31773
dz:715	Circulating CD8+ T-cells (non-severe asthma)	Asthma	GSE31773
dz:246	Testis	testicular cancer	GSE1818
dz:925	nickel-allergic patients whose SKINS were not exposed to nickel	allergic contact dermatitis	GSE6281
dz:927	nickel-allergic patients whose SKINS were exposed to nickel - 48 hours	allergic contact dermatitis	GSE6281

Appendix Table 2 (continued)

DZID	Cell type	Disease name	GEO ID
dz:926	nickel-allergic patients whose SKINS were exposed to nickel - 7 hours	allergic contact dermatitis	GSE6281
dz:928	nickel-allergic patients whose SKINS were exposed to nickel - 96 hours	allergic contact dermatitis	GSE6281
dz:862	Brain tumor tissue samples of human gliomas and normal brain - Grade:IV - Primary tumor	glioblastoma multiforme	GSE15824
dz:863	Brain tumor tissue samples of human gliomas and normal brain - Grade:IV - Secundary tumor	glioblastoma multiforme	GSE15824
dz:860	Brain tumor tissue samples of human gliomas and normal brain - Grade:AII - Primary tumor	astrocytoma	GSE15824
dz:861	Brain tumor tissue samples of human gliomas and normal brain - Grade:AIII - Primary tumor	astrocytoma	GSE15824
dz:36	Lung Tissue	chronic obstructive pulmonary disease	GSE1650
dz:196	Alveolar Macrophage	chronic obstructive pulmonary disease	GSE3212
dz:255	Bronchial epithelium	chronic obstructive pulmonary disease	GSE3320
dz:627	Normal skin and squamous cell carcinoma (SCC) tumor biopsies	skin squamous cell carcinoma	GSE2503
dz:657	Skin from SCCs and Normal human epidermis	skin squamous cell carcinoma	GSE45164
dz:371	Blood neutrophil	chronic granulomatous disease	GSE935
dz:857	Skin (LESIONAL atopic dermatitis (AL) skin lesions)	atopic dermatitis	GSE32924
dz:981	nonlesional epithelium	atopic dermatitis	GSE26952
dz:509	derma (skin)	atopic dermatitis	GSE32924
dz:1071	Breast epithelium	breast adenocarcinoma	GSE61304
dz:604	Pancreatic tissue	pancreatic ductal adenocarcinoma	GSE15471
dz:607	Esophageal normal adjacent tissue	esophagus squamous cell carcinoma	GSE20347
dz:658	KYSE human esophageal squamous cell carcinoma cell line and Normal esophageal tissue	esophagus squamous cell carcinoma	GSE63941
dz:734	Primary lung tumor and Normal lung	lung squamous cell carcinoma	GSE51852
dz:441	Lung Tissue	lung squamous cell carcinoma	GSE3268
dz:431	Lung Tissue	lung adenocarcinoma	GSE1987
dz:397	Thyroid Gland (MMHCC)	papillary thyroid carcinoma	GSE3467
dz:306	Thyroid Gland (MMHCC)	papillary thyroid carcinoma	GSE3678
dz:653	Papillary thyroid carcinoma tumors (with a BRAF mutation) and Normal thyroid specimens	papillary thyroid carcinoma	GSE54958
dz:652	Papillary thyroid carcinoma tumors (without a BRAF mutation) and Normal thyroid specimens	papillary thyroid carcinoma	GSE54958
dz:664	Blood myelomonocytic cells from RCC and Healthy blood myelomonocytic cells	renal cell carcinoma	GSE38424
dz:732	Primary lung tumor and Normal lung	lung large cell carcinoma	GSE51852
dz:731	Primary lung tumor and Normal lung	adenosquamous cell lung carcinoma	GSE51852
dz:164	Esophageal Tissue	esophagus adenocarcinoma	GSE1420

Appendix Table 2 (continued)

DZID	Cell type	Disease name	GEO ID
dz:644	Esophagus epithelium	esophagus adenocarcinoma	GSE1420
dz:362	Lung Tissue	lung small cell carcinoma	GSE1037
dz:458	Head and Neck Squamous Cell (Normal vs. Tumor)	head and neck squamous cell carcinoma	GSE6631
dz:623	Testicular seminoma tumors (at pT2) and normal testicular tissue	testis seminoma	GSE8607
dz:624	Testicular seminoma tumors (at pT3) and normal testicular tissue	testis seminoma	GSE8607
dz:622	Testicular seminoma tumors (at pT1) and normal testicular tissue	testis seminoma	GSE8607
dz:459	Skin	acne	GSE6475
dz:430	Peripheral blood mononuclear cell	juvenile rheumatoid arthritis	GSE1402
dz:493	Huh-7	hepatocellular carcinoma	GSE10393
dz:554	Peripheral blood mononuclear cells	hepatocellular carcinoma	GSE49515
dz:663	HCC tumor and matched non-tumor surrounding tissues	hepatocellular carcinoma	GSE39791
dz:735	Peripheral blood mononuclear cell	hepatocellular carcinoma	GSE58208
dz:660	Tumor tissues and Adjacent non-tumorous tissues of Hepatocellular Carcinoma	hepatocellular carcinoma	GSE57957
dz:407	Liver	hepatocellular carcinoma	GSE6764
dz:656	Hepatocellular carcinoma and Adjacent non-tumorous liver tissues	hepatocellular carcinoma	GSE60502
dz:803	AT2 E/R positive BCP ALL cell line (following E/R -ETV6/RUNX1- knockdown)	precursor B lymphoblastic lymphoma/leukemia	GSE29639
dz:202	Macrophage	ankylosing spondylitis	GSE11886
dz:305	Synovial Membrane	rheumatoid arthritis	GSE2053
dz:904	Peripheral blood mononuclear cells	rheumatoid arthritis	GSE15573
dz:110	Synovial Membrane	rheumatoid arthritis	GSE1919
dz:779	Synovial fluid macrophages	rheumatoid arthritis	GSE10500
dz:609	Epithelial cells (pancreatic duct))	pancreatic non-invasive intraductal papillary-mucinous carcinoma	GSE19650
dz:979	Sections from sporadic, primary Basal Like Cancer specimens	sporadic breast cancer	GSE3744
dz:610	Epithelial cells (pancreatic duct)	pancreatic invasive intraductal papillary-mucinous carcinoma	GSE19650
dz:752	chondrocytes monolayer culture	osteoarthritis	GSE16464
dz:751	Hyaff-11 scaffold culture	osteoarthritis	GSE16464
dz:576	PBMC (peripheral blood mononuclear cells)	arthritis	GSE21521
dz:574	PBMC (peripheral blood mononuclear cells)	arthritis	GSE21521
dz:575	PBMC (peripheral blood mononuclear cells)	arthritis	GSE21521
dz:578	PBMC (peripheral blood mononuclear cells)	arthritis	GSE21521
dz:579	PBMC (peripheral blood mononuclear cells)	arthritis	GSE21521
dz:577	PBMC (peripheral blood mononuclear cells)	arthritis	GSE21521

Appendix Table 2 (continued)

DZID	Cell type	Disease name	GEO ID
dz:580	PBMC (peripheral blood mononuclear cells)	arthritis	GSE21521
dz:4	Haematopoietic stem cell	chronic myeloid leukemia	GSE4170
dz:403	Haematopoietic stem cell	chronic myeloid leukemia	GSE11889
dz:456	CD34+ hematopoietic stem and progenitor cells from the bone marrow of untreated patients	chronic myeloid leukemia	GSE5550
dz:924	Colonic mucosa - Non-Inflamed	ulcerative colitis	GSE9452
dz:759	Colon	ulcerative colitis	GSE6731
dz:594	colon mucosa	ulcerative colitis	GSE37283
dz:590	ascending colon	ulcerative colitis	GSE11223
dz:591	ascending colon	ulcerative colitis	GSE11223
dz:593	sigmoid colon	ulcerative colitis	GSE11223
dz:589	descending colon	ulcerative colitis	GSE11223
dz:588	descending colon	ulcerative colitis	GSE11223
dz:587	terminal ileum	ulcerative colitis	GSE11223
dz:586	sigmoid colon	ulcerative colitis	GSE11223
dz:585	sigmoid colon	ulcerative colitis	GSE11223
dz:762	PBMC (peripheral blood mononuclear cells)	ulcerative colitis	GSE3365
dz:454	Colon	ulcerative colitis	GSE9452
dz:710	Intestinal biopsies	ulcerative colitis	GSE22619
dz:249	Intestine - Large Intestine - Colon (MMHCC)	ulcerative colitis	GSE6731
dz:760	Colon	ulcerative colitis	GSE6731
dz:810	Intestinal mucosa	ulcerative colitis	GSE38713
dz:993	sigmoid colons (mucosal biopsy)	ulcerative colitis	GSE1710
dz:188	Peripheral blood mononuclear cell	ulcerative colitis	GSE3365
dz:264	Sigmoid colon	ulcerative colitis	GSE1710
dz:923	Colonic mucosa – Inflamed	ulcerative colitis	GSE9452
dz:739	Peripheral blood B cells (inactive lupus)	lupus erythematosus	GSE30153
dz:350	Skin tissue	actinic keratosis	GSE2503
dz:628	Normal skin and actinic keratotic (AK) lesion	actinic keratosis	GSE2503
dz:754	Non-lesional skin	psoriasis	GSE14905
dz:837	Non-lesional (NL) Skin - (Uninvolved samples)	psoriasis	GSE13355
dz:836	Non-lesional (NL) Skin	psoriasis	GSE32407
dz:813	Epidermis	psoriasis	GSE53431
dz:982	nonlesional epithelium	psoriasis	GSE26952
dz:697	Whole blood (Cutaneous Psoriasis)	psoriasis	GSE61281
dz:690	Skin	psoriasis	GSE52471
dz:358	T lymphocyte	autoimmune thrombocytopenic purpura	GSE574
dz:696	Whole blood	psoriatic arthritis	GSE61281
dz:693	CD19+ B cells	systemic lupus erythematosus	GSE10325

Appendix Table 2 (continued)

DZID	Cell type	Disease name	GEO ID
dz:512	synovial biopsies from affected knees	systemic lupus erythematosus	GSE36700
dz:816	Connective Tissue in Joints	systemic lupus erythematosus	GSE61635
dz:1075	CD8+ T cells - European-American SLE patients	systemic lupus erythematosus	GSE55447
dz:1074	CD8+ T cells - African-American SLE patients	systemic lupus erythematosus	GSE55447
dz:1073	CD4+ T cells - European-American SLE patients	systemic lupus erythematosus	GSE55447
dz:1079	CD20+ B cells - European-American SLE patients	systemic lupus erythematosus	GSE55447
dz:692	CD4+ T cells	systemic lupus erythematosus	GSE10325
dz:691	CD33+ myeloid cells	systemic lupus erythematosus	GSE10325
dz:450	CD3+ T cells	systemic lupus erythematosus	GSE13887
dz:112	Mononuclear Leukocyte	acute myeloid leukemia	GSE2191
dz:782	Bone marrow	acute myeloid leukemia	GSE9476
dz:783	Peripheral blood	acute myeloid leukemia	GSE9476
dz:76	Gastric Tissue	gastrointestinal stromal tumor	GSE15966
dz:270	Connective Tissue	gastrointestinal stromal tumor	GSE2719
dz:552	Colon	colorectal cancer	GSE32323
dz:581	skeletal muscle	type 2 diabetes mellitus	GSE36297
dz:882	Mammary arterial tissue	type 2 diabetes mellitus	GSE13760
dz:895	Percutaneous needle LIVER biopsies - (hepatokines) – Female	type 2 diabetes mellitus	GSE23343
dz:893	Percutaneous needle LIVER biopsies - (hepatokines) – Male	type 2 diabetes mellitus	GSE23343
dz:274	Muscle tissue	type 2 diabetes mellitus	GSE12643
dz:510	Peripheral blood	acute myocardial infarction	GSE48060
dz:360	Bronchial epithelium	allergic asthma	GSE3004
dz:716	Bronchial biopsies	allergic asthma	GSE41649
dz:561	Bone marrow plasma cells	multiple myeloma	GSE47552
dz:707	Bone marrow mesenchymal stromal cells	multiple myeloma	GSE36474
dz:787	Plasma cells	multiple myeloma	GSE6691
dz:9	T lymphocyte	type 1 diabetes mellitus	GSE10586
dz:654	Anaplastic thyroid carcinoma tissue and Normal thyroid tissue	anaplastic thyroid carcinoma	GSE65144
dz:190	Colon	Carcinoma in situ of large intestine	GSE4183
dz:745	Peripheral blood	Chronic Lymphocytic Leukemia (Chronic B-lymphocytic leukemia)	GSE26725
dz:689	Skin	discoid lupus erythematosus	GSE52471
dz:176	Peripheral blood mononuclear cell	JRA - Juvenile rheumatoid arthritis	GSE7753
dz:74	T lymphocyte	Leukemia, Chronic T-Cell	GSE5788
dz:47	Peripheral blood mononuclear cell	LGLL - Large granular lymphocytic leukemia	GSE10631

Appendix Table 2 (continued)

DZID	Cell type	Disease name	GEO ID
dz:206	Bone marrow stem cell	MDS - Myelodysplastic syndrome	GSE4619
dz:980	Biopsy	melanoma in situ	GSE4587
dz:199	Chondrocyte	Osteoarthritis	GSE16464
dz:227	Muscle tissue	Polymyositis	GSE3112
dz:35	Skin tissue	Psoriasis vulgaris	GSE13355
dz:93	Skin tissue	Psoriasis vulgaris	GSE14905
dz:444	Skin tissue	Psoriasis vulgaris	GSE6710
dz:307	Whole blood	Septic Shock	GSE9692
dz:295	Oropharynx Epithelium	Squamous cell carcinoma of mouth	GSE3524

Bibliography

[1] H. Xue, J. Li, H. Xie, Y.J.I.j.o.b.s. Wang, Review of drug repositioning approaches and resources, 14 (2018) 1232.

[2] T.T. Ashburn, K.B. Thor, Drug repositioning: identifying and developing new uses for existing drugs, Nature Reviews Drug Discovery, 3 (2004) 673-683.

[3] P.J.W.S.J. Landers, Drug industry's big push into technology falls short, 24 (2004).

[4] A. Mullard, 2014 FDA drug approvals, Nature Reviews Drug Discovery, 14 (2015) 77-81.

[5] S.-D. Zhang, T.W. Gant, sscMap: An extensible Java application for connecting small-molecule drugs using gene-expression signatures, BMC bioinformatics, 10 (2009) 236.

[6] J.M. Engreitz, R. Chen, A.A. Morgan, J.T. Dudley, R. Mallelwar, A.J. Butte, ProfileChaser: searching microarray repositories based on genome-wide patterns of differential expression, Bioinformatics, 27 (2011) 3317-3318.

[7] F. Pammolli, L. Magazzini, M. Riccaboni, The productivity crisis in pharmaceutical R&D, Nature Reviews Drug Discovery, 10 (2011) 428-438.

[8] Y. Yeu, Y. Yoon, S.J.M.B. Park, Protein localization vector propagation: a method for improving the accuracy of drug repositioning, 11 (2015) 2096-2102.

[9] P. Deotarse, A. Jain, M. Baile, N. Kolhe, A. Kulkarni, Drug repositioning: a review, Int. J. Pharma. Res Rev, 4 (2015) 51-58.

[10] T.I. Oprea, J.P.J.A. Overington, d.d. technologies, Computational and practical aspects of drug repositioning, 13 (2015) 299-306.

[11] D.S. Wishart, C. Knox, A.C. Guo, S. Shrivastava, M. Hassanali, P. Stothard, Z. Chang, J. Woolsey, DrugBank: a comprehensive resource for in silico drug discovery and exploration, Nucleic Acids Res, 34 (2006) D668-672.

[12] H. Ogata, S. Goto, K. Sato, W. Fujibuchi, H. Bono, M.J.N.a.r. Kanehisa, KEGG: Kyoto encyclopedia of genes and genomes, 27 (1999) 29-34.

[13] T. Barrett, S.E. Wilhite, P. Ledoux, C. Evangelista, I.F. Kim, M. Tomashevsky, K.A. Marshall, K.H. Phillippy, P.M. Sherman, M.J.N.a.r. Holko, NCBI GEO: archive for functional genomics data sets—update, 41 (2012) D991-D995.

[14] A.D. King, N. Pržulj, I.J.B. Jurisica, Protein complex prediction via cost-based clustering, 20 (2004) 3013-3020.

[15] K. Macropol, T. Can, A.K.J.B.b. Singh, RRW: repeated random walks on genome-scale protein networks for local cluster discovery, 10 (2009) 283.

[16] C. Wu, R.C. Gudivada, B.J. Aronow, A.G.J.B.s.b. Jegga, Computational drug repositioning through heterogeneous network clustering, 7 (2013) S6.

[17] L. Yu, J. Huang, Z. Ma, J. Zhang, Y. Zou, L.J.B.m.g. Gao, Inferring drug-disease associations based on known protein complexes, 8 (2015) S2.

[18] H. Wu, L. Gao, J. Dong, X.J.P.o. Yang, Detecting overlapping protein complexes by rough-fuzzy clustering in protein-protein interaction networks, 9 (2014).

[19] H. Luo, J. Wang, M. Li, J. Luo, X. Peng, F.-X. Wu, Y.J.B. Pan, Drug repositioning based on comprehensive similarity measures and bi-random walk algorithm, 32 (2016) 2664-2671.

[20] D. Emig, A. Ivliev, O. Pustovalova, L. Lancashire, S. Bureeva, Y. Nikolsky, M.J.P.O. Bessarabova, Drug target prediction and repositioning using an integrated network-based approach, 8 (2013).

[21] S. Köhler, S. Bauer, D. Horn, P.N.J.T.A.J.o.H.G. Robinson, Walking the interactome for prioritization of candidate disease genes, 82 (2008) 949-958.

[22] O. Vanunu, O. Magger, E. Ruppin, T. Shlomi, R.J.P.c.b. Sharan, Associating genes and protein complexes with disease via network propagation, 6 (2010).

[23] V. Martinez, C. Navarro, C. Cano, W. Fajardo, A.J.A.i.i.m. Blanco, DrugNet: Network-based drug–disease prioritization by integrating heterogeneous data, 63 (2015) 41-49.

[24] Y. Sun, Z. Sheng, C. Ma, K. Tang, R. Zhu, Z. Wu, R. Shen, J. Feng, D. Wu, D.J.N.c. Huang, Combining genomic and network characteristics for extended capability in predicting synergistic drugs for cancer, 6 (2015) 1-10.

[25] J. Lu, L. Chen, J. Yin, T. Huang, Y. Bi, X. Kong, M. Zheng, Y.-D.J.J.o.B.S. Cai, Dynamics, Identification of new candidate drugs for lung cancer using chemical–chemical interactions, chemical–protein interactions and a K-means clustering algorithm, 34 (2016) 906-917.

[26] A.A. Alizadeh, M.B. Eisen, R.E. Davis, C. Ma, I.S. Lossos, A. Rosenwald, J.C. Boldrick, H. Sabet, T. Tran, X. Yu, J.I. Powell, L. Yang, G.E. Marti, T. Moore, J. Hudson, Jr., L. Lu, D.B. Lewis, R. Tibshirani, G. Sherlock, W.C. Chan, T.C. Greiner, D.D. Weisenburger, J.O. Armitage, R. Warnke, R. Levy, W. Wilson, M.R. Grever, J.C. Byrd, D. Botstein, P.O. Brown, L.M. Staudt, Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling, Nature, 403 (2000) 503-511.

[27] S. Ramaswamy, P. Tamayo, R. Rifkin, S. Mukherjee, C.H. Yeang, M. Angelo, C. Ladd, M. Reich, E. Latulippe, J.P. Mesirov, T. Poggio, W. Gerald, M. Loda, E.S. Lander, T.R. Golub, Multiclass cancer diagnosis using tumor gene expression signatures, Proc Natl Acad Sci U S A, 98 (2001) 15149-15154.

[28] G. Wright, B. Tan, A. Rosenwald, E.H. Hurt, A. Wiestner, L.M. Staudt, A gene expression-based method to diagnose clinically distinct subgroups of diffuse large B cell lymphoma, Proc Natl Acad Sci U S A, 100 (2003) 9991-9996.

[29] Y.L. Yap, X.W. Zhang, D. Smith, R. Soong, J. Hill, Molecular gene expression signature patterns for gastric cancer diagnosis, Comput Biol Chem, 31 (2007) 275-287.

[30] A.F. Ziober, K.R. Patel, F. Alawi, P. Gimotty, R.S. Weber, M.M. Feldman, A.A. Chalian, G.S. Weinstein, J. Hunt, B.L. Ziober, Identification of a gene signature for rapid screening of oral squamous cell carcinoma, Clin Cancer Res, 12 (2006) 5960-5971.

[31] F. Chibon, Cancer gene expression signatures–the rise and fall?, European journal of cancer, 49 (2013) 2000-2009.

[32] H.Y. Chen, S.L. Yu, C.H. Chen, G.C. Chang, C.Y. Chen, A. Yuan, C.L. Cheng, C.H. Wang, H.J. Terng, S.F. Kao, W.K. Chan, H.N. Li, C.C. Liu, S. Singh, W.J. Chen, J.J. Chen, P.C. Yang, A five-gene signature and clinical outcome in non-small-cell lung cancer, N Engl J Med, 356 (2007) 11-20.

[33] B.J. Boersma, M. Reimers, M. Yi, J.A. Ludwig, B.T. Luke, R.M. Stephens, H.G. Yfantis, D.H. Lee, J.N. Weinstein, S. Ambs, A stromal gene signature associated with inflammatory breast cancer, International journal of cancer, 122 (2008) 1324-1332.

[34] J.M. Bueno-de-Mesquita, W.H. van Harten, V.P. Retel, L.J. van't Veer, F.S. van Dam, K. Karsenberg, K.F. Douma, H. van Tinteren, J.L. Peterse, J. Wesseling, Use of 70-gene signature to predict prognosis of patients with node-negative breast cancer: a prospective community-based feasibility study (RASTER), The lancet oncology, 8 (2007) 1079-1087.

[35] D.-T. Chen, A. Nasir, A. Culhane, C. Venkataramu, W. Fulp, R. Rubio, T. Wang, D. Agrawal, S.M. McCarthy, M. Gruidl, Proliferative genes dominate malignancy-risk gene signature in histologically-normal breast tissue, Breast cancer research and treatment, 119 (2010) 335.

[36] D.-J. Cheon, Y. Tong, M.-S. Sim, J. Dering, D. Berel, X. Cui, J. Lester, J.A. Beach, M. Tighiouart, A.E. Walts, A collagen-remodeling gene signature regulated by TGF-β signaling is associated with metastasis and poor survival in serous ovarian cancer, Clinical cancer research, 20 (2014) 711-723.

[37] C.W. Lee, K. Simin, Q. Liu, J. Plescia, M. Guha, A. Khan, C.-C. Hsieh, D.C. Altieri, A functional Notch–survivin gene signature in basal breast cancer, Breast Cancer Research, 10 (2008) R97.

[38] S. Loi, B. Haibe-Kains, S. Majjaj, F. Lallemand, V. Durbecq, D. Larsimont, A.M. Gonzalez-Angulo, L. Pusztai, W.F. Symmans, A. Bardelli, PIK3CA mutations associated with gene signature of low mTORC1 signaling and better outcomes in estrogen receptor–positive breast cancer, Proceedings of the National Academy of Sciences, 107 (2010) 10208-10213.

[39] J.L. Messina, D.A. Fenstermacher, S. Eschrich, X. Qu, A.E. Berglund, M.C. Lloyd, M.J. Schell, V.K. Sondak, J.S. Weber, J.J. Mulé, 12-Chemokine gene signature identifies lymph node-like structures in melanoma: potential for patient selection for immunotherapy?, Scientific reports, 2 (2012) 765.

[40] L.D. Miller, L.G. Coffman, J.W. Chou, M.A. Black, J. Bergh, R. D'Agostino, S.V. Torti, F.M. Torti, An iron regulatory gene signature predicts outcome in breast cancer, Cancer research, 71 (2011) 6728-6737.

[41] C. Sotiriou, L. Pusztai, Gene-expression signatures in breast cancer, N Engl J Med, 360 (2009) 790-800.

[42] J.K. Stratford, D.J. Bentrem, J.M. Anderson, C. Fan, K.A. Volmar, J. Marron, E.D. Routh, L.S. Caskey, J.C. Samuel, C.J. Der, A six-gene signature predicts survival of patients with localized pancreatic ductal adenocarcinoma, PLoS medicine, 7 (2010).

[43] M.J. Van De Vijver, Y.D. He, L.J. Van't Veer, H. Dai, A.A. Hart, D.W. Voskuil, G.J. Schreiber, J.L. Peterse, C. Roberts, M.J. Marton, A gene-expression signature as a predictor of survival in breast cancer, New England Journal of Medicine, 347 (2002) 1999-2009.

[44] B. Wallden, J. Storhoff, T. Nielsen, N. Dowidar, C. Schaper, S. Ferree, S. Liu, S. Leung, G. Geiss, J. Snider, Development and verification of the PAM50-based Prosigna breast cancer gene signature assay, BMC medical genomics, 8 (2015) 54.

[45] R.R. Weichselbaum, H. Ishwaran, T. Yoon, D.S. Nuyten, S.W. Baker, N. Khodarev, A.W. Su, A.Y. Shaikh, P. Roach, B. Kreike, An interferon-related gene signature for DNA damage resistance is a predictive marker for chemotherapy and radiation for breast cancer, Proceedings of the National Academy of Sciences, 105 (2008) 18490-18495.

[46] K. Yamaguchi, M. Mandai, T. Oura, N. Matsumura, J. Hamanishi, T. Baba, S. Matsui, S. Murphy, I. Konishi, Identification of an ovarian clear cell carcinoma gene signature that reflects inherent disease biology and the carcinogenic processes, Oncogene, 29 (2010) 1741-1752.

[47] F. Iorio, R. Bosotti, E. Scacheri, V. Belcastro, P. Mithbaokar, R. Ferriero, L. Murino, R. Tagliaferri, N. Brunetti-Pierri, A. Isacchi, D. di Bernardo, Discovery of drug mode of action and drug repositioning from transcriptional responses, Proc Natl Acad Sci U S A, 107 (2010) 14621-14626.

[48] M. Sirota, J.T. Dudley, J. Kim, A.P. Chiang, A.A. Morgan, A. Sweet-Cordero, J. Sage, A.J. Butte, Discovery and preclinical validation of drug indications using compendia of public gene expression data, Sci Transl Med, 3 (2011) 96ra77.

[49] G. Hu, P. Agarwal, Human disease-drug network based on genomic expression profiles, PLoS One, 4 (2009) e6536.

[50] S.D. Kunkel, M. Suneja, S.M. Ebert, K.S. Bongers, D.K. Fox, S.E. Malmberg, F. Alipour, R.K. Shields, C.M. Adams, mRNA expression signatures of human skeletal muscle

atrophy identify a natural compound that increases muscle mass, Cell metabolism, 13 (2011) 627-638.

[51] M.-H. Chen, W.-L.R. Yang, K.-T. Lin, C.-H. Liu, Y.-W. Liu, K.-W. Huang, P.M.-H. Chang, J.-M. Lai, C.-N. Hsu, K.-M. Chao, Gene expression-based chemical genomics identifies potential therapeutic drugs in hepatocellular carcinoma, PloS one, 6 (2011).

[52] Y. Ishimatsu-Tsuji, T. Soma, J. Kishimoto, Identification of novel hair-growth inducers by means of connectivity mapping, The FASEB Journal, 24 (2010) 1489-1496.

[53] M. Chang, S. Smith, A. Thorpe, M.J. Barratt, F. Karim, Evaluation of phenoxybenzamine in the CFA model of pain following gene expression studies and connectivity mapping, Molecular pain, 6 (2010) 56.

[54] S. Claerhout, J.Y. Lim, W. Choi, Y.-Y. Park, K. Kim, S.-B. Kim, J.-S. Lee, G.B. Mills, J.Y. Cho, Gene expression signature analysis identifies vorinostat as a candidate therapy for gastric cancer, PloS one, 6 (2011).

[55] Y. Donner, S. Kazmierczak, K. Fortney, Drug Repurposing Using Deep Embeddings of Gene Expression Profiles, Molecular Pharmaceutics, 15 (2018) 4314-4325.

[56] Z. Wang, C.D. Monteiro, K.M. Jagodnik, N.F. Fernandez, G.W. Gundersen, A.D. Rouillard, S.L. Jenkins, A.S. Feldmann, K.S. Hu, M.G. McDermott, Q. Duan, N.R. Clark, M.R. Jones, Y. Kou, T. Goff, H. Woodland, F.M.R. Amaral, G.L. Szeto, O. Fuchs, S.M. Schussler-Fiorenza Rose, S. Sharma, U. Schwartz, X.B. Bausela, M. Szymkiewicz, V. Maroulis, A. Salykin, C.M. Barra, C.D. Kruth, N.J. Bongio, V. Mathur, R.D. Todoric, U.E. Rubin, A. Malatras, C.T. Fulp, J.A. Galindo, R. Motiejunaite, C. Juschke, P.C. Dishuck, K. Lahl, M. Jafari, S. Aibar, A. Zaravinos, L.H. Steenhuizen, L.R. Allison, P. Gamallo, F. de Andres Segura, T. Dae Devlin, V. Perez-Garcia, A. Ma'ayan, Extraction and analysis of signatures from the Gene Expression Omnibus by the crowd, Nat Commun, 7 (2016) 12846.

[57] G.W. Gundersen, M.R. Jones, A.D. Rouillard, Y. Kou, C.D. Monteiro, A.S. Feldmann, K.S. Hu, A. Ma'ayan, GEO2Enrichr: browser extension and server app to extract gene sets from GEO and analyze them for biological functions, Bioinformatics, 31 (2015) 3060-3062.

[58] N.R. Clark, K.S. Hu, A.S. Feldmann, Y. Kou, E.Y. Chen, Q. Duan, A. Ma'ayan, The characteristic direction: a geometrical approach to identify differentially expressed genes, BMC Bioinformatics, 15 (2014) 79.

[59] P. Jaccard, Nouvelles recherches sur la distribution florale, Bull. Soc. Vaud. Sci. Nat., 44 (1908) 223-270.

[60] D.S. Wishart, C. Knox, A.C. Guo, D. Cheng, S. Shrivastava, D. Tzur, B. Gautam, M. Hassanali, DrugBank: a knowledgebase for drugs, drug actions and drug targets, Nucleic Acids Res, 36 (2008) D901-906.

[61] D.S. Wishart, Y.D. Feunang, A.C. Guo, E.J. Lo, A. Marcu, J.R. Grant, T. Sajed, D. Johnson, C. Li, Z.J.N.a.r. Sayeeda, DrugBank 5.0: a major update to the DrugBank database for 2018, 46 (2017) D1074-D1082.

[62] S.A. Lambert, A. Jolma, L.F. Campitelli, P.K. Das, Y. Yin, M. Albu, X. Chen, J. Taipale, T.R. Hughes, M.T. Weirauch, The Human Transcription Factors, Cell, 172 (2018) 650-665.

[63] Student, The probable error of a mean, Biometrika, (1908) 1-25.

[64] J.A. Nelder, R.W. Wedderburn, Generalized Linear Models, Journal of the Royal Statistical Society Series a-General, 135 (1972) 370-+.

[65] Y. Benjamini, Y.J.J.o.t.R.s.s.s.B. Hochberg, Controlling the false discovery rate: a practical and powerful approach to multiple testing, 57 (1995) 289-300.

[66] R.C. Team, R: A language and environment for statistical computing, (2013).

[67] M.F. Bustamante, R.N. Nurtdinov, J. Río, X. Montalban, M.J.P.O. Comabella, Baseline gene expression signatures in monocytes from multiple sclerosis patients treated with interferon-beta, 8 (2013).

[68] A. Bibeau-Poirier, M.J.J.C. Servant, Roles of ubiquitination in pattern-recognition receptors and type I interferon receptor signaling, 43 (2008) 359-367.

[69] K. Honda, T.J.N.R.I. Taniguchi, IRFs: master regulators of signalling by Toll-like receptors and cytosolic pattern-recognition receptors, 6 (2006) 644-658.

[70] X. Su, Y. Yu, Y. Zhong, E.G. Giannopoulou, X. Hu, H. Liu, J.R. Cross, G. Rätsch, C.M. Rice, L.B.J.N.i. Ivashkiv, Interferon-γ regulates cellular metabolism and mRNA translation to potentiate macrophage activation, 16 (2015) 838.

[71] W. Cao, S. Manicassamy, H. Tang, S.P. Kasturi, A. Pirani, N. Murthy, B.J.N.i. Pulendran, Toll-like receptor–mediated induction of type I interferon in plasmacytoid dendritic cells requires the rapamycin-sensitive PI (3) K-mTOR-p70S6K pathway, 9 (2008) 1157.

[72] S.L. Weinstein, A.J. Finn, S.H. Davé, F. Meng, C.A. Lowell, J.S. Sanghera, A.L.J.J.o.L.B. DeFranco, Phosphatidylinositol 3‐kinase and mTOR mediate lipopolysaccharide‐stimulated nitric oxide production in macrophages via interferon‐β, 67 (2000) 405-414.

[73] B.S. Staitieh, E.E. Egea, X. Fan, N. Azih, W. Neveu, D.M.J.J.o.c. Guidot, c. immunology, Activation of alveolar macrophages with interferon-γ promotes antioxidant defenses via the Nrf2-ARE pathway, 6 (2015).

[74] A.K. Perry, C. Gang, D. Zheng, T. Hong, G.J.C.r. Cheng, The host type I interferon response to viral and bacterial infections, 15 (2005) 407-422.

[75] W.V.R. Vieweg, M. Hasnain, R.H. Howland, J.M. Hettema, C. Kogut, M.A. Wood, A.K. Pandurangi, Citalopram, QTc interval prolongation, and torsade de pointes. How should we apply the recent FDA ruling?, The American journal of medicine, 125 (2012) 859-868.

[76] K. Honda, H. Yanai, H. Negishi, M. Asagiri, M. Sato, T. Mizutani, N. Shimada, Y. Ohba, A. Takaoka, N.J.N. Yoshida, IRF-7 is the master regulator of type-I interferon-dependent immune responses, 434 (2005) 772-777.

[77] J. Hilpert, J.M. Beekman, S. Schwenke, K. Kowal, D. Bauer, J. Lampe, R. Sandbrink, J.F. Heubach, S. Stürzebecher, J.J.J.o.n. Reischl, Biological response genes after single dose administration of interferon β-1b to healthy male volunteers, 199 (2008) 115-125.

[78] A.M. Becker, K.H. Dao, B.K. Han, R. Kornu, S. Lakhanpal, A.B. Mobley, Q.-Z. Li, Y. Lian, T. Wu, A.M.J.P.o. Reimold, SLE peripheral blood B cell, T cell and myeloid cell transcriptomes display unique profiles and each subset contributes to the interferon signature, 8 (2013).

[79] M.K. Crow, K.A. Kirou, J.J.A. Wohlgemuth, Microarray analysis of interferon-regulated genes in SLE, 36 (2003) 481-490.

[80] N. Shrestha, W. Bahnan, D.J. Wiley, G. Barber, K.A. Fields, K.J.J.o.B.C. Schesser, Eukaryotic initiation factor 2 (eIF2) signaling regulates proinflammatory cytokine expression and bacterial invasion, 287 (2012) 28738-28744.

[81] A. Flynn, C.G.J.C.s. Proud, The role of eIF4 in cell proliferation, 27 (1996) 293-310.

[82] A. Isaacs, Interferon, Advances in virus research, Elsevier1964, pp. 1-38.

[83] M. Sarasin-Filipowicz, E.J. Oakeley, F.H. Duong, V. Christen, L. Terracciano, W. Filipowicz, M.H.J.P.o.t.N.A.o.S. Heim, Interferon signaling and treatment outcome in chronic hepatitis C, 105 (2008) 7034-7039.

[84] R. Flavin, A. Pettersson, W.K. Hendrickson, M. Fiorentino, S. Finn, L. Kunz, G.L. Judson, R. Lis, D. Bailey, C.J.C.C.R. Fiore, SPINK1 protein expression and prostate cancer progression, 20 (2014) 4904-4911.

[85] J.W. Smoller, The genetics of stress-related disorders: PTSD, depression, and anxiety disorders, Neuropsychopharmacology, 41 (2016) 297.

[86] K.G. Le Roch, Y. Zhou, P.L. Blair, M. Grainger, J.K. Moch, J.D. Haynes, P. De la Vega, A.A. Holder, S. Batalov, D.J. Carucci, E.A. Winzeler, Discovery of Gene Function by Expression Profiling of the Malaria Parasite Life Cycle, Science, 301 (2003) 1503.

[87] J. Cheng, L. Yang, V. Kumar, P. Agarwal, Systematic evaluation of connectivity map for disease indications, Genome Med, 6 (2014) 540.

[88] J. Lamb, E.D. Crawford, D. Peck, J.W. Modell, I.C. Blat, M.J. Wrobel, J. Lerner, J.P. Brunet, A. Subramanian, K.N. Ross, M. Reich, H. Hieronymus, G. Wei, S.A. Armstrong, S.J.

Haggarty, P.A. Clemons, R. Wei, S.A. Carr, E.S. Lander, T.R. Golub, The Connectivity Map: using gene-expression signatures to connect small molecules, genes, and disease, Science, 313 (2006) 1929-1935.

[89] R. de Sousa Abreu, L.O. Penalva, E.M. Marcotte, C.J.M.B. Vogel, Global signatures of protein and mRNA expression levels, 5 (2009) 1512-1526.

[90] T.M. Filzen, P.S. Kutchukian, J.D. Hermes, J. Li, M. Tudor, Representing high throughput expression profiles via perturbation barcodes reveals compound targets, PLoS computational biology, 13 (2017).