

Article

A Protocol for Identifying Potentially Repurposable Drugs Using Online Tools and Databases

Aditi Chattopadhyay¹ and Madhavi K. Ganapathiraju^{2,3,*}

¹ Grade 12, Upper St. Clair High School, Pittsburgh, PA 15241, USA; aditic@verizon.net

² Department of Biomedical Informatics, University of Pittsburgh, Pittsburgh, PA 15260, USA

³ Intelligent Systems Program, University of Pittsburgh, Pittsburgh, PA 15260, USA

* Correspondence: madhavi@pitt.edu

Abstract: Traditional methods for discovery and development of new drugs can be a very time-consuming and expensive process because it includes several stages such as compound identification, pre-clinical and clinical trials before the drug is approved by the US Food and Drug Administration (FDA). Therefore, drug repurposing, namely using currently FDA-approved drugs as therapeutics for other diseases than what they are originally prescribed for, is emerging to be a faster and more cost-effective alternative to current drug discovery methods. In this paper, we have described a three-step *in silico* protocol for analyzing transcriptomics data using online databases and bioinformatics tools for identifying potentially repurposable drugs. The efficacy of this protocol was evaluated by comparing its predictions with the findings of two case studies of recently reported repurposed drugs: HIV treating drug Zidovudine for the treatment of Dry Age-Related Macular Degeneration and the antidepressant Imipramine for Small-Cell Lung Carcinoma. The proposed protocol successfully identified the published findings, thus demonstrating the efficacy of this method. In addition, it also yielded several novel predictions that have not yet been published, including the finding that Imipramine could potentially treat Severe Acute Respiratory Syndrome (SARS), a disease that currently does not have any treatment or vaccine. Since this *in-silico* protocol is simple to use and does not require advanced computer skills, we believe any motivated participant with access to these databases and tools would be able to apply it to large datasets to identify other potentially repurposable drugs in the future.

Keywords: drug repurposing; translational bioinformatics; transcriptomics; transcriptome analysis; drug discovery; protocol; gene expression

Introduction

De novo methods for drug discovery can be an incredibly expensive and time-consuming. A target molecule – such as a gene that is causally linked to a disease is first identified to initiate the drug discovery process. Scientists then systematically screen for small molecules that modulate the target protein-product and carry out a series of optimization steps to improve the efficacy of the lead drug. Next, periods of animal studies are followed by clinical trials before a drug can be approved for the market by the USA Food and Drug Administration (FDA) (or its counterpart in other countries). Usually, this procedure costs 4 to 12 billion dollars and takes an average of 12 to 17 years to complete (1). Unfortunately, in some cases, this money could also go waste, without resulting in any successful drug that is safe for human use. For example, if at a very late stage of the discovery process a promising drug starts to reveal severe side effects during clinical trials, the drug is discarded, and the resources put into it is subsequently will be wasted. Due to these limitations, only 50 new drugs could be approved by the U.S. FDA during a period of ten years from 1999 to 2008 (2).

In order to improve the lengthy and costly nature of current drug discovery practices, a new approach based on repurposing – the application of existing FDA approved drugs on new diseases – has recently gained traction (1). Drug repurposing, which is typically based on systems biology and data analytics, could drastically reduce both the cost and time associated with finding suitable drugs for diseases. There are several drugs that were successfully repurposed to treat new diseases (3), of which some well-known examples include: anti-angina drug *Sildenafil Citrate* repurposed to treat erectile dysfunction (4), the diabetes medicine *Metformin* that has potential to use for cancer treatment (5), HIV drug *Nucleoside Reverse Transcriptase Inhibitors* that could treat “dry” Age-Related Macular Degeneration (6), and *Imipramine*, a tricyclic antidepressant, with potential to treat small-cell lung carcinoma (7).

A disease is typically caused by the irregular expression of certain genes; hence, each disease reveals a signature gene expression pattern (8). When a suitable drug is administered to treat a disease, it tends to correct the aberration in the body by bringing the gene expression pattern back to its normal state (8,9). A schematic diagram representing opposing gene expression patterns under a disease state and under its drug treatment is presented in Fig. 1.

Advancements in high-throughput technologies such as microarrays and next-generation sequencing based RNA-Sequencing methods allow scientists to measure gene expression patterns under different experimental conditions. Scientists routinely deposit gene expression data derived from such experiments into online repositories, oftentimes making the data freely available (10).

Transcriptomics data from experiments comparing cells under two different conditions (for e.g., disease vs. normal state, or drug treatment vs. non-treatment) provides a list of genes that are differentially expressed under the two conditions. Large scale analysis of transcriptomics data encompassing differentially expressed genes (DEG) has demonstrated that it is possible to identify potentially repurposable drugs; for a disease and a drug that is suitable for that disease, the DEG pattern from a “disease vs. normal” experiment generally reveals a strong negative correlation with the DEG profile linked with a “drug treatment vs. non-treatment” experiment (8,9). While prior work presented results of large scale data analysis, here we sought to develop a protocol that employs readily available online tools and databases for continued application of this approach by motivated individuals to identify repurposable drugs for diseases of interest.

We have developed an *in silico* protocol to identify repurposable drugs. It starts with an FDA approved drug and then searches for diseases that show negatively correlated gene expression in relation to that drug. This protocol is based on transcriptomics data analysis and employs online databases, bioinformatics tools, and literature search engine. The efficacy of this protocol was assessed by comparing the protocol prediction with the results of two recently published articles of repurposed drugs: Zidovudine (6) and Imipramine (7).

Methods

BaseSpace Correlation Engine

BaseSpace Correlation Engine, formerly NextBio, is an searchable online database maintained and delivered by Illumina Inc., San Diego, CA, USA¹ (11). This database collects raw experimental data from high-throughput gene expression experiments submitted to global repositories such as the Gene Expression Omnibus (GEO)² and Array Express³.

BaseSpace Correlation Engine utilizes proprietary statistical algorithms to convert raw experimental data into a list of genes that are differentially expressed in certain conditions along with their corresponding fold change and p-value calculations. The fold change value indicates how a given gene is differentially expressed in a test condition compared to the control condition of an experiment. Examples are experiments with drug treatment vs. non-treatment or disease state vs. normal state. Rank-based enrichment statistics are then used to compute the pairwise correlation scores between all gene expression signatures present in the database. The most correlated gene expression study present for each query was assigned a numerical score of 100 and scores for the rest of the results were normalized to the top-ranked study. This resource enables users to find gene expression experiments with given drug-treated versus untreated conditions. Hence, diseases showing strong negative correlation with a given drug can be identified to predict potentially repurposable drugs.

PubMed

An online database⁴ comprised of over 26 million citations published in 2600 life sciences Journals. Many of these citations provide links to the abstract and full text of the article. This database is freely accessible and maintained by the National Library of Medicine, a division of the US National Institute of Health.

In silico protocol to identify repurposed drugs

A protocol encompassing the systematic search of the online databases mentioned above enables users to quickly and efficiently identify potential target diseases for a drug (Figure 2). The steps are described as follows:

1. Enter a drug name in the search box of the BaseSpace Correlation Engine and then click on the icon named "Curated Studies" displayed at the top of the web page. Once the search result is returned, click on the "Filter By" option and select "Data Types" available at the top of the page and select "RNA Expression."

Browse the search results returned that will display numerous independent gene expression studies. Identify the studies that compare the gene expressed data for drug treatment vs. control (untreated). Select the appropriate study by clicking on the hyperlink of that study. Each study may have multiple experiments measured under different conditions such as a different dosage of the drug or different treatment time points. Clicking on the study will bring up a page with a detailed description of the study as well as links to the gene expression data associated with each experiment. When the experiment was selected by clicking the hyperlinked title and a table consisting of the differential gene expression data will appear on the screen.

¹ <https://www.nextbio.com/>

² <http://www.ncbi.nlm.nih.gov/geo/>

³ <https://www.ebi.ac.uk/arrayexpress/>

⁴ <https://www.ncbi.nlm.nih.gov/pubmed>

2. Select the icon "Disease Atlas" available at the top of the page. A web page displaying a table of various disease names with their corresponding correlation scores will appear. Sort through the table by selecting "Rank" from the drop-down menu available under the "View By" option displayed at the top left corner of the page. Re-sort the results by selecting the "-ve Correlation" option present in the drop-down menu under the "Correlation with Query" Column heading at the top of the right-most column. Finally, select the diseases that have the largest number of studies. The queried drug would have the potential to treat the selected diseases.
3. Search literature database to validate disease predictions.

Go to PubMed and enter "the drug name AND the predicted disease name" in the search box and look for citations.

Results

This protocol was applied to two previously published repurposed drugs and the protocol predictions were compared with the reported findings.

Case study 1: Zidovudine

Through experiments performed in mouse models, Fowler et al. reported that the drug Zidovudine, a Nucleoside Reverse Transcriptase Inhibitor (NRTI) usually used to treat HIV patients, showed strong potential to be used against the untreatable Dry Form of Age-related Macular Degeneration (AMD) (6).

Study Selection from BaseSpace Correlation Engine:

A search for curated studies using the query Zidovudine in the BaseSpace Correlation Engine retrieved 10 RNA-Expression studies: 6 studies based on rat experiments and 4 from experiments done on human data. A study titled "Drug Matrix *In Vitro* Toxicogenomic Study – Rat Hepatocytes [Affymetrix]" (12,13) was selected for further analysis. Although drug-treatment studies performed on normal human cell lines would be preferred for repurposed disease prediction, for this query, all human studies were only carried out of cancer cell lines or virus- infected cells. Hence they were not suitable for study selection. On the other hand, the studies based on rat experiments were performed under normal conditions. Since the liver is a principal site of drug metabolism, a study measuring the gene expression patterns of rat hepatocyte cells was selected.

In this study, rat hepatocyte cells isolated from male Sprague-Dawley rats were co-cultured *in vitro* with varying doses of the drug during different time points. Microarray experiments were then performed on an Affymetrix platform to measure and compare the gene expression patterns of treatment versus control conditions. Typically, a gene expression study is comprised of multiple experiments. Among the three experiments listed under this study, the experiment titled "Primary rat hepatocytes + ZIDOVUDINE at 14800uM in DMSO 1D_vs_vehicle" covered the largest number of genes (8000). Therefore, the DEG profile of the said experiment was used as a query to seek out negatively-correlated (NC) diseases for Zidovudine. Fig. 3 shows a screenshot taken from the BaseSpace Correlation Engine result page displaying top ten differentially expressed genes.

Negatively Correlated Diseases:

The *BaseSpace Correlation Engine* ranks the negatively correlated diseases based on its assigned correlation score. The most correlated gene expression study, with the lowest p-value, present for each query is assigned a numerical score of 100 and scores for the rest of the results were normalized to the top-ranked study. However, for this protocol, the ranking was manually changed to reflect the number of supporting, independent studies. This step was taken to bolster the prediction efficiency with the notion that if more independent experiments found a negative correlation between the drug and the disease, then its ranking should be stronger (while still maintaining an acceptable score of at least 50/100 to keep correlation as a factor). Table 1 shows top 10 Negatively-Correlated Diseases for HIV-drug Zidovudine ordered by the number of supported gene expression studies.

Among the diseases listed in the table, Human Immunodeficiency Virus Infection, the intended target of Zidovudine, is present (no. 5). The gene expression profiles derived from twenty-one studies revealed a strong statistically significant negative correlation (score: 62) with the expression pattern of Zidovudine. Figure 4 displays a comparison of gene expressions between the query experiment and the experiment "CD8+ T cells from chronic HIV infection patients _vs_ negative control" listed under the study "HIV-infected individuals with various clinical stages of HIV infection" performed by Hyrcza M *et al.*(14). Since the protocol correctly predicted that Zidovudine could treat HIV – its intended target – this finding serves as a valuable result to prove the efficacy of this protocol.

Age-Related Macular Degeneration (AMD), the published finding (6) that we sought to find falls under the broader term "Retinol Disorder," which is present in the list the NC-diseases (no. 9). The DEG profile of the experiment "Macular retina – GA Age-related macular degeneration *vs.* normal tissue" as part of the study "Age-related macular degeneration subtype expression analysis" revealed a strong negative correlation with the gene expression profile of Zidovudine (15).

Apart from HIV, the intended target of Zidovudine, all other NC diseases are novel findings as they are not listed as a therapeutic target of Zidovudine in well-known drug information resources such as DrugBank (16) and National Library of Medicine Drug Information Portal (17). While the potential of this drug to treat AMD has been recently published, a literature search was conducted to collect evidence on whether the other NC diseases found have been connected to Zidovudine. A literature search found Marcais A *et al.* reported that the treatment of Zidovudine is highly effective in the treatment of a leukemic subtype of Adult T-Cell Lymphoma (18). Furthermore, Beck-Engeser GB *et al.* published a paper on the efficacy of Zidovudine treatment against Lupus erythematosus (19). However, no literature evidence was found for Cardiovascular Disease, Neuropathy, Mycobacteriosis, Rheumatoid Arthritis, Myopathy, or Dermatitis.

Case study 2: Imipramine

In 2013, Jahchan and colleagues reported that the tricyclic antidepressant (TCA) Imipramine could be efficiently repurposed to treat small cell lung carcinoma (7). In a protocol of their own, this study sought small molecules with the ability to treat the apparently recalcitrant form of lung cancer first by analyzing the disease derived transcriptomic data and then by running experiments in an animal model. One of the small molecules identified was Imipramine.

Study/Experiment Selection from *BaseSpace Correlation Engine*:

A search for Imipramine (performed on June 20th, 2016) in the *BaseSpace Correlation Engine* results in eight studies. The DEG profile derived from the experiment titled "Hepatocytes of female donors treated 24hr with 15uM imipramine _vs_ 0uM" done in human as a part of the study "Genomics Assisted Toxicity

Evaluation system study - Human Hepatocytes" was selected for target disease prediction through this protocol (20).

Negatively Correlated Diseases:

A list of NC Diseases based on the DEG profile expressed by the selected query experiment is presented in Table 2. The list shows a preponderance of cancer subtypes. Note that Small-Cell Lung Carcinoma, the reported therapeutic target for Imipramine, is present in the NC-disease list under Lung Cancer (no. 3). Hence, the findings reported by Jahchen et al., (7) was successfully replicated by this *in silico* protocol. A literature search on Imipramine retrieved previously published articles that studied the treatment of Imipramine on two of the predicted target diseases: Breast Cancer (21) and Brain Cancer (22). However, no supporting evidence from the literature was found for the treatment of Liver Cancer, Kidney Cancer, Inflammatory Bowel Disease, or Severe Acute Respiratory Syndrome (SARS) using Imipramine. The prediction of SARS is notable because it revealed the highest correlation score of 100. However, the lack of reported evidence on the repositioning of Imipramine against SARS further emphasizes the novelty of this prediction.

Limitations

This online repurposed drug prediction protocol solely depends on the availability of the *BaseSpace Correlation Engine* hosted transcriptomics data. If the data for a given drug is not available, the protocol cannot be applied to that drug.

Furthermore, the DEG data gathered from experiments that compare the effect of the drug treatment on a normal (non-diseased) human cell line or tissue to the untreated or vehicle treated condition are considered to be ideal datasets for this method. However, for many drugs such datasets are not available. Instead, the available DEG data come from experiments on a human cancer cell lines or from experiments performed on animal models such as mouse and rat. Consequently, the protocol utilizing the non-ideal dataset may infer erroneous predictions and require additional validations.

The *BaseSpace Correlation engine* was freely available to the research community, and the author received free access upon request to the vendor. However, currently, the tool requires paid subscription with a fifteen-day free trial option.

Discussion

The case studies mentioned in this study substantiate the target disease prediction accuracy for a drug by this protocol. The finding that Imipramine could be repurposed to treat Small-Cell Lung Cancer was an expected finding as transcriptomics data analysis was employed by both the literature described methods as well as by this *in silico* protocol. However, the correct prediction that the Dry Form of AMD may be a target disease is noteworthy because the underlying discovery methods were different. The literature (6) used a small molecule screening technique while this *in silico* protocol applied transcriptomics data analysis. This finding further adds credence to the strategy adopted by this protocol to predict diseases a given drug can be repurposed toward.

The prediction that Imipramine could be repurposed against SARS is a significant observation. SARS is a deadly viral disease, and during the period of 2002 to 2003, an outbreak caused over 8000 cases with 772 deaths reported in 37 countries (23). As of today, there is no treatment or vaccines available for SARS. This makes the predictions made by this protocol particularly advantageous for such deadly diseases by giving patients a possible treatment (24). Because this is a novel finding, further clinical investigation measuring the potency of Imipramine against SARS in human would be needed.

While transcriptomics data is freely available, running sophisticated correlation statistics on large data sets to predict drug-target diseases requires proficiency in the fields of computer science, bioinformatics and statistics, along with access to high powered computers. The *in silico* protocol described in this study leverages pre-processed transcriptomics data available through the BaseSpace Correlation Engine database. This method is straight forward, simple to use, and does not require advanced computer skills. Hence, it may be beneficial to users including researchers, clinicians, patients, and even high school students who can apply this method and identify the existing drugs that can be repurposed to serve as new therapeutic targets for various diseases in future.

Author Contributions

AC is a High School student (12th grade) and carried out this work and wrote the manuscript with supervision from MKG. Manuscript has been read and approved by both authors.

Conflict of Interest

Authors declare that they do not have any conflict of interest.

References

1. Hurlle MR, Yang L, Xie Q, Rajpal DK, Sanseau P, Agarwal P. Computational drug repositioning: from data to therapeutics. *Clin Pharmacol Ther.* 2013 Apr;93(4):335–341.
2. Swinney DC, Anthony J. How were new medicines discovered? *Nat Rev Drug Discov.* 2011 Jul;10(7):507–519.
3. Chong CR, Sullivan DJ. New uses for old drugs. *Nature.* 2007 Aug 9;448(7154):645–646.
4. Ghofrani HA, Osterloh IH, Grimminger F. Sildenafil: from angina to erectile dysfunction to pulmonary hypertension and beyond. *Nat Rev Drug Discov.* 2006 Aug;5(8):689–702.
5. Morales DR, Morris AD. Metformin in cancer treatment and prevention. *Annu Rev Med.* 2015;66:17–29.
6. Fowler BJ, Gelfand BD, Kim Y, Kerur N, Tarallo V, Hirano Y, et al. Nucleoside reverse transcriptase inhibitors possess intrinsic anti-inflammatory activity. *Science.* 2014 Nov 21;346(6212):1000–1003.
7. Jahchan NS, Dudley JT, Mazur PK, Flores N, Yang D, Palmerton A, et al. A drug repositioning approach identifies tricyclic antidepressants as inhibitors of small cell lung cancer and other neuroendocrine tumors. *Cancer Discov.* 2013 Dec;3(12):1364–1377.
8. Sirota M, Dudley JT, Kim J, Chiang AP, Morgan AA, Sweet-Cordero A, et al. Discovery and preclinical validation of drug indications using compendia of public gene expression data. *Sci Transl Med.* 2011 Aug 17;3(96):96ra77.
9. Iorio F, Bosotti R, Scacheri E, Belcastro V, Mithbaekar P, Ferriero R, et al. Discovery of drug mode of action and drug repositioning from transcriptional responses. *Proc Natl Acad Sci U S A.* 2010 Aug 17;107(33):14621–14626.
10. Barrett T, Wilhite SE, Ledoux P, Evangelista C, Kim IF, Tomashevsky M, et al. NCBI GEO: archive for functional genomics data sets--update. *Nucleic Acids Res.* 2013 Jan;41(Database issue):D991–5.
11. Kupershmidt I, Su QJ, Grewal A, Sundaresh S, Halperin I, Flynn J, et al. Ontology-based meta-analysis of global collections of high-throughput public data. *PLoS ONE.* 2010 Sep 29;5(9).

12. Natsoulis G, Pearson CI, Gollub J, P Eynon B, Ferng J, Nair R, et al. The liver pharmacological and xenobiotic gene response repertoire. *Mol Syst Biol*. 2008 Mar 25;4:175.
13. Waters M, Stasiewicz S, Merrick BA, Tomer K, Bushel P, Paules R, et al. CEBS--Chemical Effects in Biological Systems: a public data repository integrating study design and toxicity data with microarray and proteomics data. *Nucleic Acids Res*. 2008 Jan;36(Database issue):D892–900.
14. Hyrcza MD, Kovacs C, Loutfy M, Halpenny R, Heisler L, Yang S, et al. Distinct transcriptional profiles in ex vivo CD4+ and CD8+ T cells are established early in human immunodeficiency virus type 1 infection and are characterized by a chronic interferon response as well as extensive transcriptional changes in CD8+ T cells. *J Virol*. 2007 Apr;81(7):3477–3486.
15. Newman AM, Gallo NB, Hancox LS, Miller NJ, Radeke CM, Maloney MA, et al. Systems-level analysis of age-related macular degeneration reveals global biomarkers and phenotype-specific functional networks. *Genome Med*. 2012 Feb 24;4(2):16.
16. Wishart DS, Knox C, Guo AC, Shrivastava S, Hassanali M, Stothard P, et al. DrugBank: a comprehensive resource for in silico drug discovery and exploration. *Nucleic Acids Res*. 2006 Jan 1;34(Database issue):D668–72.
17. Drug Information Portal - U.S. National Library of Medicine - Quick Access to Quality Drug Information [Internet]. [cited 2016 Sep 18]. Available from: <https://druginfo.nlm.nih.gov/drugportal/drugportal.jsp>
18. Marçais A, Suarez F, Sibon D, Frenzel L, Hermine O, Bazarbachi A. Therapeutic options for adult T-cell leukemia/lymphoma. *Curr Oncol Rep*. 2013 Oct;15(5):457–464.
19. Beck-Engeser GB, Eilat D, Wabl M. An autoimmune disease prevented by anti-retroviral drugs. *Retrovirology*. 2011 Nov 8;8:91.
20. Uehara T, Ono A, Maruyama T, Kato I, Yamada H, Ohno Y, et al. The Japanese toxicogenomics project: application of toxicogenomics. *Mol Nutr Food Res*. 2010 Feb;54(2):218–227.
21. Rajamanickam S, Panneerdoss S, Gorthi A, Timilsina S, Onyeagucha B, Kovalsky D, et al. Inhibition of FoxM1-Mediated DNA Repair by Imipramine Blue Suppresses Breast Cancer Growth and Metastasis. *Clin Cancer Res*. 2016 Feb 29;
22. Shipman L. Glioma: Repurposed drugs combined to amplify autophagy. *Nat Rev Cancer*. 2015 Nov;15(11):636.
23. WHO | Summary of probable SARS cases with onset of illness from 1 November 2002 to 31 July 2003 [Internet]. [cited 2016 Sep 10]. Available from: http://www.who.int/csr/sars/country/table2004_04_21/en/
24. Jiang S, Lu L, Du L. Development of SARS vaccines and therapeutics is still needed. *Future Virol*. 2013 Jan;8(1):1–2.

Figure 1: A contrasting gene expression pattern between studies – “Disease vs Normal” and “Drug treatment vs Untreated.” Red colored bars indicate over-expression of genes while green-colored bars represents under-expressed genes. A. Differential expression of genes in disease. B. Opposing differential expression due to administering a drug (relative to ‘normal’ expression).

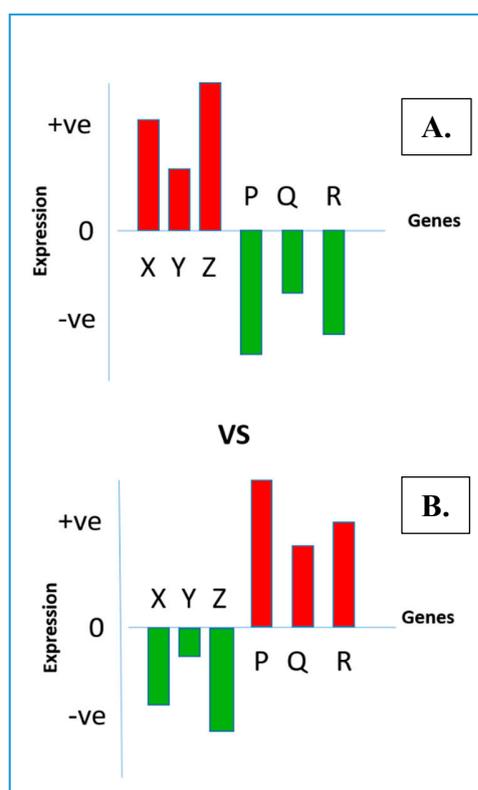


Figure 2: A three-step online protocol for repurposed drug prediction.

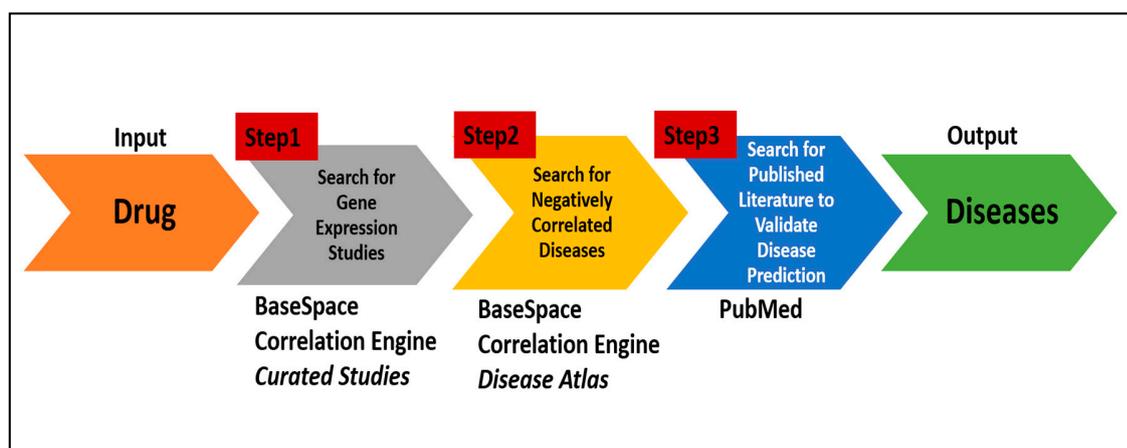


Figure 3: A screenshot taken from the *BaseSpace Correlation Engine* generated result page displaying top ten differentially expressed genes of the selected query study – *Zidovudine treatment on rat hepatocyte cells vs untreated cells*.

Primary rat hepatocytes + ZIDOVUDINE at 14800uM in DMSO 1d_vs_vehicle

BioSet from study: Drug Matrix In Vitro Toxicogenomic Study - Rat Hepatocytes [Affymetrix]

Rattus norvegicus | RE: RNA Expression | 8,000 features (mapped to 7,026 genes)

View BioSet Details

Selected Study

Selected Experiment

Search by genes or keywords show all Data Filter NEW Pathway Viewer NEW Pathway Studio NEW

8,000 features

Top Ten Differentially Expressed Genes

Symbol	EntrezGene ID	Imported ID	Control Expression	Fold Change	P-Value	Rank ▲	Test Expression
Tagln	25123	1367570_at	38241.5	-50.1	4.3E-5	1	764.1
Cyp2c12	25011	1368155_at	5401.4	-22.5	2.8E-14	2	240.5
Gna14	309242	1381557_at	4696.4	-21.1	6.6E-21	3	222.3
1385005_at	NA	1385005_at	4431.2	-21	0.0008	4	211.4
Cav1	25404	1393281_at	5585.8	-20.4	0.0004	5	273.7
Sult1a1	83783	1370019_at	14915.7	-18.5	4.5E-10	6	806.9
Cav1	25404	1372111_at	6620.5	-16.5	0.0007	7	401.5
Lox	24914	1368171_at	12932.9	-13.1	0.0037	8	986.3
Cd55	64036	1387951_at	190.4	12.5	0.0003	9	2384.2
Pck1	362282	1372264_at	3483.2	12.3	2.3E-5	10	42798.1

Table 1. Top 10 Negatively-Correlated Diseases for HIV-drug Zidovudine ordered by the number of supported gene expression studies

Negatively Correlated Disease	Number of Supporting Studies in <i>BaseSpace</i> Correlation Engine	Correlation Score Determined by <i>BaseSpace</i> Correlation Engine	Published Supporting Literature in <i>PubMed</i>
1. Lupus Erythematosus	45	65	Yes
2. Cardiovascular Diseases	26	54	No
3. Lymphoid Leukemia	25	58	Yes
4. Neuropathy	24	65	No
5. Human Immunodeficiency Virus Infection	21	62	Yes
6. Mycobacteriosis	19	67	No
7. Rheumatoid Arthritis	19	58	No
8. Myopathy	17	64	No
9. Retinal Disorder	17	60	Yes
10. Dermatitis	17	50	No

Figure 4: A screenshot was taken from the *BaseSpace Correlation Engine* showing a statistically significant negative correlation among differentially expressed genes from two experiments – “CD8+ T cells from acute HIV-infected patients’ vs negative control” and the query study “Zidovudine treatment on rat hepatocyte vs negative control (vehicle treated cells).”

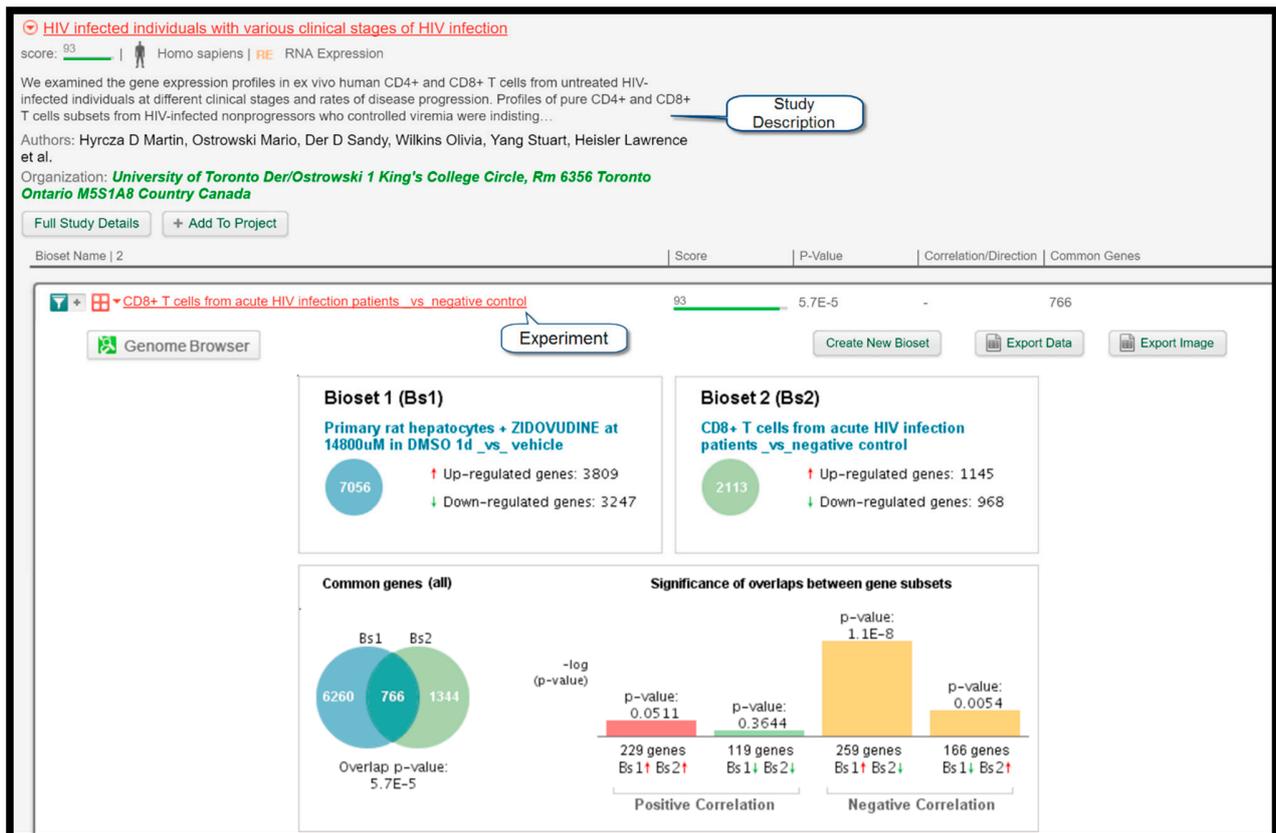


Table 2. Top 10 Negatively-Correlated Diseases ordered by the number of supported gene expression studies for antidepressant Imipramine

Negatively Correlated Disease	Number of Supporting Studies in <i>BaseSpace Correlation Engine</i>	Correlation Score Determined by <i>BaseSpace Correlation Engine</i>	Published Supporting Literature in PubMed
1. Breast Cancer	57	65	Yes
2. Allergic Disorder	50	65	Yes
3. Lung Cancer	49	65	Yes
4. Liver Cancer	47	65	No
5. Brain Cancer	44	66	Yes
6. Inflammatory Bowel Disease	32	66	No
7. Cardiomyopathy	28	68	No
8. Kidney Cancer	26	68	No
9. Nerve Injury	23	79	Yes
10. Severe Acute Respiratory Syndrome	14	100	No



© 2016 by the authors; licensee *Preprints*, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons by Attribution (CC-BY) license (<http://creativecommons.org/licenses/by/4.0/>).