PLOS COMPUTATIONAL BIOLOGY

RESEARCH ARTICLE

Enhancing reproducibility of gene expression analysis with known protein functional relationships: The concept of well-associated protein

Joël R. Pradines☯, Victor Farutin☯*, Nicholas A. Cilfone, Abouzar Ghavami, Elma Kurtagic, Jamey Guess, Anthony M. Manning, Ishan Capila*

Momenta Pharmaceuticals, 301 Binney Street, Cambridge, Massachusetts, United States of America

☯ These authors contributed equally to this work.
* vfarutin@momentapharma.com (VF); icapila@gmail.com (IC)

Citation: Pradines JR, Farutin V, Cilfone NA, Ghavami A, Kurtagic E, Guess J, et al. (2020) Enhancing reproducibility of gene expression analysis with known protein functional relationships: The concept of well-associated protein. PLoS Comput Biol 16(2): e1007684. https://doi.org/10.1371/journal.pcbi.1007684

Editor: Ilya Ioshikhes, University of Ottawa, CANADA

Received: July 20, 2019

Accepted: January 27, 2020

Published: February 14, 2020

Data Availability Statement: The authors confirm that all data underlying the findings are fully available without restriction. All relevant data are available from the NCBI-GEO database under accession numbers (GSE9285, GSE13195, GSE13355, GSE19188, GSE19826, GSE27342, GSE30727, GSE30784, GSE30999, GSE32413, GSE34248, GSE36376, GSE41258, GSE41662, GSE43458, GSE44076, GSE44861, GSE45485, GSE51981, GSE58095, GSE63089, GSE79973) indicated in the manuscript.

Abstract

Identification of differentially expressed genes (DEGs) is well recognized to be variable across independent replications of genome-wide transcriptional studies. These are often employed to characterize disease state early in the process of discovery and prioritize novel targets aimed at addressing unmet medical need. Increasing reproducibility of biological findings from these studies could potentially positively impact the success rate of new clinical interventions. This work demonstrates that statistically sound combination of gene expression data with prior knowledge about biology in the form of large protein interaction networks can yield quantitatively more reproducible observations from studies characterizing human disease. The novel concept of Well-Associated Proteins (WAPs) introduced herein—gene products significantly associated on protein interaction networks with the differences in transcript levels between control and disease—does not require choosing a differential expression threshold and can be computed efficiently enough to enable false discovery rate estimation via permutation. Reproducibility of WAPs is shown to be on average superior to that of DEGs under easily-quantifiable conditions suggesting that they can yield a significantly more robust description of disease. Enhanced reproducibility of WAPs versus DEGs is first demonstrated with four independent data sets focused on systemic sclerosis. This finding is then validated over thousands of pairs of data sets obtained by random partitions of large studies in several other diseases. Conditions that individual data sets must satisfy to yield robust WAP scores are examined. Reproducible identification of WAPs can potentially benefit drug target selection and precision medicine studies.

Author summary

Gene expression studies are commonly used to characterize biological systems. Genes identified in such experiments as expressed at different levels between conditions (e.g. healthy vs. disease) can indicate biological functions that are important in this context.

Funding: The authors received no specific funding for this work.

Competing interests: I have read the journal's policy and the authors of this manuscript have the following competing interests: All authors are currently (or were at the time the study was conducted) employees of Momenta Pharmaceuticals and may own stock and/or stock options in Momenta Pharmaceuticals.

However, it is well-recognized that such findings can vary substantially across independent investigations. We quantified reproducibility here under a conservative control scenario that partitions a given data set in two, independently of experimental conditions, for multiple data sets characterizing several diseases in humans. Furthermore, we have shown that it is possible to obtain more reproducible findings than DEGs, which we term Well-Associated Proteins, characterizing differences in gene expression between healthy and disease states. This was accomplished by combining gene expression and prior knowledge of functional relationships between gene products accumulated over many studies and publications. Resulting Well-Associated Proteins can be computed efficiently enough to enable permutation controls and demonstrate on average higher reproducibility than differentially expressed genes, both within and across data sets. This suggests that Well-Associated Proteins may better reflect differences in biology when comparing disease and healthy states than DEGs, thus representing an important step towards identification of key disease drivers.

Introduction

Microarrays and RNA sequencing are experimental technologies convenient for generating lists of Differentially Expressed Genes (DEGs) that characterize differences between two conditions, e.g. a disease versus its absence. It is commonly recognized that DEG identification can be highly variable over independent studies [1–3]. Some of the advances aimed at improving reproducibility of the DEG discovery included removal of DEGs with small fold changes between conditions [3, 4], as well as accounting for correlation among DEGs [5, 6], and relying on ranked lists of DEGs [7] when comparing DEGs across independent experiments. It has also been argued that high variability in the identified DEGs is an inherent property of gene expression studies and that development of new metrics of reproducibility might be desirable in this context [8].

The modular vision of biology [9] inspired development of a vast array of systems-based methodologies [10] quantifying differential regulation of gene sets representing known biological functions such as Gene Ontology (GO) categories [11], curated collections of pathways (e.g. KEGG [12] or Reactome [13]) or molecular signatures observed in previous gene expression studies (e.g. MSigDB [14]). These methods can potentially improve sensitivity and interpretability of gene expression experiments and have been benchmarked in several contexts [15–18]. Networks of known interactions between gene products represent a complementary approach for organizing genes by their functional relationships and have been used for differential analysis of gene expression data by multiple methods that have been extensively reviewed [10, 19, 20] and benchmarked [21, 22] as well. Studies using pathway networks in combination with genome-wide patient profiling data include works by [23] that developed permutation tests to demonstrate significant connectedness on PPI network of genomic loci associated with several common diseases and [24] that found higher replicability of the predictions of patient response to treatment in an independent study, when patient-level gene expression data was combined with the network of causal relationships representing transcriptional regulation. These approaches utilize prior biological knowledge, are well established for the analysis of gene expression data and are routinely used within high throughput biology studies [25].

Assessing the gain in reproducibility of the systems-based analyses results due to the use of prior biological knowledge (e.g. functional categories or pathway networks) still remains

difficult to address. A survey of the studies evaluating reproducibility of the findings at the level of genes and/or gene sets [26–31] illustrates a variety of challenges associated with this task. They include: the choice of reference gene expression data and gene set knowledgebase, the selection of analyses methodologies to compare, the definition of reproducibility metrics, and, most importantly, the burden of interpreting results when comparing reproducibility metrics for different types of entities (e.g. as calculated for genes vs. those for GO categories).

This paper introduces a novel concept of a Well-Associated Protein (WAP) that quantifies the association of gene (product) on a protein interaction network (STRING [32]) with the genes that are more significantly regulated in the experiment. This development enables comparison of the reproducibility of findings across independent experiments within the same universe of genes represented in the protein interaction network that can be scored both for their individual differential expression (as DEG) or, as a WAP, for their association on the network with the most significantly regulated genes. The significance of the WAP association accounts for the total number of interactions of each protein (protein degree) on the network, and computation of this significance is fast enough to enable permutation controls. This allows for identification of gene products which have a significantly large number of known associations to the genes that are most perturbed in the experiment, without actually choosing a threshold of differential expression (consequently incurring inevitable information loss). Therefore resulting WAPs can attain statistical significance while not being themselves differentially expressed. Thus they can extend standard gene expression analysis results while leveraging the systems-level of information encoded in the protein network and the entire compendium of data obtained in a gene expression experiment.

By considering only genes which are represented both in gene expression data and on the protein interaction network, this approach enables the direct comparison of the reproducibility metrics for DEG and WAP rankings. Comparisons presented in this paper demonstrate that, under easily-quantifiable conditions, higher average reproducibility of WAP identification versus that of DEG is observed over nine types of diseases.

This paper is organized as follows. After presenting computational details, reproducibility of WAP and DEG identification is examined over four large data sets where gene expression in skin samples is compared between Systemic Sclerosis (SSc) patients and non-SSc subjects, demonstrating greater reproducibility of top WAPs as compared to that of top DEGs across these data sets. Superior average robustness of top WAPs versus top DEGs in disease versus normal comparisons is then validated over thousands of data set pairs obtained by random partitions of eighteen large gene-expression studies incorporating eight other diseases (colon cancer, gastric cancer, endometriosis, hepatocellular carcinoma, non-small-cell lung carcinoma, lung adenocarcinoma, oral squamous cell carcinoma and psoriasis). Additionally, conditions for individual data sets contributing to the higher robustness of WAPs are examined. Finally, limitations of the approach and potential for further work are discussed.

In summary, the WAP score is a robust statistic which ranks gene products by the significance of their known interactions with the genes most perturbed in the experiment without having to choose a threshold of differential expression. Because WAP scores can be efficiently computed, false discovery rates can be numerically estimated by permutation techniques that preserve correlation of gene expression [33]. Identification of genes with significant WAP scores complements standard gene expression analysis, and extends it by identifying genes that are not themselves DEGs. Enhanced reproducibility of WAP identification versus traditional selection of DEGs is demonstrated under specific conditions, and holds at least across nine types of disease. Such an increase in reproducibility of the findings from gene expression studies, driven by the prior biological knowledge encoded in a protein network, suggests that

the resulting WAPs may represent a robust description of the disease-related biology that is likely to be beneficial to drug target selection and precision medicine approaches.

Materials and methods

The concept of well-associated protein

The approach presented herein identifies proteins, termed Well-Associated Proteins (WAPs), which have a large number of known functional relationships (i.e. interactions) in a protein interaction network with the genes that are significantly perturbed in gene expression data. The method utilized to identify WAPs is illustrated with Fig 1A. Consider the n genes which are represented both in a gene expression data set and the protein interaction network. Genes are ranked from the most differentially expressed ($i = 1$) to the least ($i = n$) using results of gene-level statistical models as warranted by a given study design (e.g. t-test, linear or mixed effects model, etc.).

The association of any gene product j to the top i DEGs is a function of $x_{j,i}$, the number of known interactions in the protein network between j and this set. The value of $x_{j,i}$ depends not only on i but also on the degrees (total numbers of interactions) of all considered gene products. A possible approach to account for degrees is to compare $x_{j,i}$ to the expected number of interactions when randomly rewiring edges in the whole network, while exactly preserving the degree k of each gene product [34]. This however requires extensive numerical simulations, because explicit derivation of the expected number of interactions under this random graph model is a hard problem [35].

By considering a relaxed random graph model, which only preserves degrees on average over realizations [36], one can easily obtain the distribution of random variable $X_{j,i}$ [37]. Briefly, any possible interaction between two proteins of degrees k_u and k_v is represented by a Bernoulli random variable of parameter $b_{uv} = k_u k_v / 2M$ (b_{uv} is the probability of this interaction to exist), where M is the total number of interactions in the network, and interactions are modeled as independent variables (Section 1 in S1 Text provides additional technical details). For sparse networks ($b_{uv} \ll 1$), counting interactions is then equivalent to summing

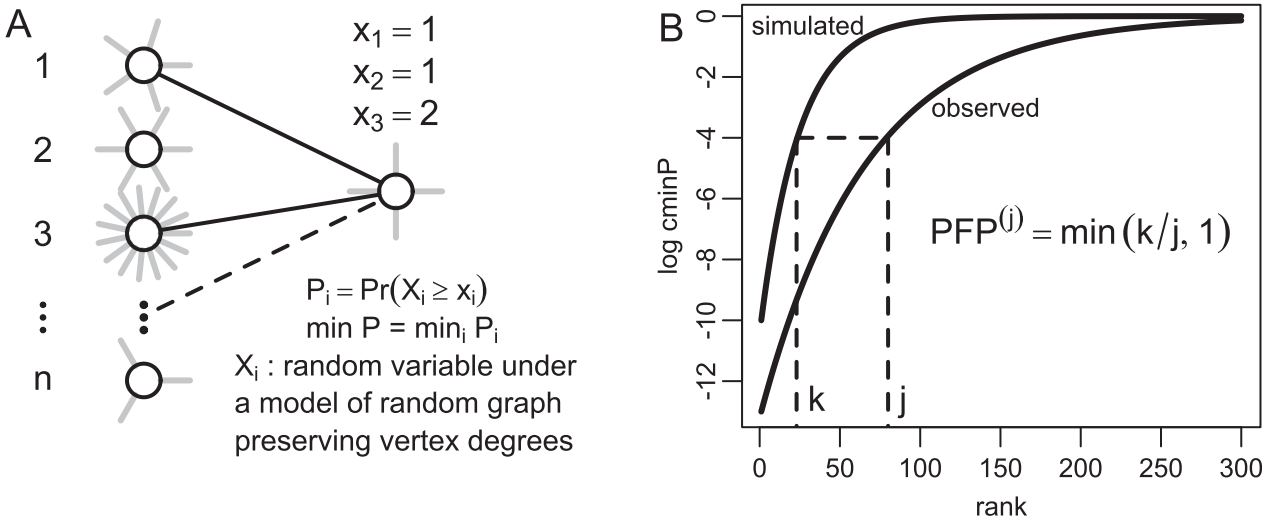

Fig 1. Illustration of WAP scores. A: Gene products are scored for their association x_i (number of interactions) to the top i DEGs, where i varies from 1 to the total number n of genes. Values of x_i are compared via edge-count probabilities P_i, and a gene product is scored with its best corrected attachment p-value: c min$_i$ P_i. B: The profile of ranked c min P values is compared to profiles estimated after randomly assigning disease and control labels to samples (10^4 permutations). This yields an estimated Proportion of False Positive (PFP) of WAPs at each rank j.

https://doi.org/10.1371/journal.pcbi.1007684.g001

independent Bernoulli variables of small parameters and the resulting distribution can be approximated by a Poisson distribution [38]. The distribution of random variable $X_{j,i}$ is therefore Poisson with parameter

$$\lambda_{j,i} = \frac{k_j}{2M} \sum_{h \le i, h \neq j} k_h. \tag{1}$$

After some normalization [37], the probability of observing at least $x_{j,i}$ interactions is given by

$$\Pr\left(X_{j,i} \geq x_{j,i}\right) = \frac{e^{-\lambda_{j,i}}}{\alpha_{j,i}} \sum_{h=x_{j,i}}^{i} \frac{\lambda_{j,i}^h}{h!}; \quad \alpha_{j,i} = \sum_{h=0}^{i} \frac{\lambda_{j,i}^h}{h!}. \tag{2}$$

Notice that, because $\lambda_{j,i}$ increases with i, the p-value is conditional not only to the observed number x_i of interactions, but also to the considered number i of DEGs. This implies that there is no need for choosing a threshold (bound on i) which defines differential expression. The above p-value is referred to as $P_{j,i}$ for short. This type of p-value enables comparison of diverse protein sets for their connectedness in a network, while taking into account protein degrees [37, 39–41]. One can compare values of $P_{j,1}$, $P_{j,2}$,. . ., $P_{j,n}$ and the best association of gene product j to DEGs corresponds to $\min_i P_{j,i}$. Again, this quantity does not require choosing a threshold of differential expression. To compare best association scores across proteins one must further correct by their degree, because the operation of taking the minimum can yield bias towards proteins with large degrees and due to lower values of $P_{j,i}$ attainable for them (Section 3 in S1 Text). The corrected best association score for gene product j of degree k_j in the protein interaction network used herein is given by

$$\begin{aligned} &\text{c min } P_j = (\rho(k_j)/\rho(1)) \min_i P_{j,i}, \\ &\text{with } \rho(k) = k^{-\alpha} e^{-\beta}, \ \alpha = 0.1799, \ \text{and } \beta = 1.056. \end{aligned} \tag{3}$$

When the total number n of proteins in the network is at least a few thousand, the correction depends only on k_j. Gene products are next sorted by ascending values of corrected best association score c min P:

$$\text{c min } P^{(1)} \leq \ldots \leq \text{c min } P^{(j)} \leq \ldots \leq \text{c min } P^{(n)}. \tag{4}$$

Computation of all c min P values can be optimally performed in $\mathcal{O}(M)$, where M is the total number of interactions in the network (Section 2 in S1 Text).

Fast computation makes it possible to estimate a false discovery rate (FDR; the expected proportion of false positives at a given cutoff) for the observed values of c min $P^{(j)}$ via permutation techniques in a reasonable time. Permutation, randomizing sample labels (healthy vs. disease) as described below, was employed to estimate proportion of false positives under the null hypothesis of interchangeability of observations in these two groups. The null hypothesis of randomizing sample labels is that of the primary interest for identifying WAPs representing differences between these two groups, unlike the null hypothesis of random rewiring of the pathway network, yielding actual values of minP, that was used to account for the wide ranging disparity of vertex degrees when scoring them for connectedness to more differentially regulated genes. For the sake of clarity, different acronyms will be used to emphasize the distinction between permutation-based estimates of FDR for WAPs and Benjamini-Hochberg [42] estimates of FDR for DEGs. Throughout this paper the former will be referred to as a Proportion of False Positives (PFP) and use of FDR will be reserved to represent the latter—Benjamini-Hochberg FDR for DEGs. To estimate PFP values for WAPs disease states are randomly

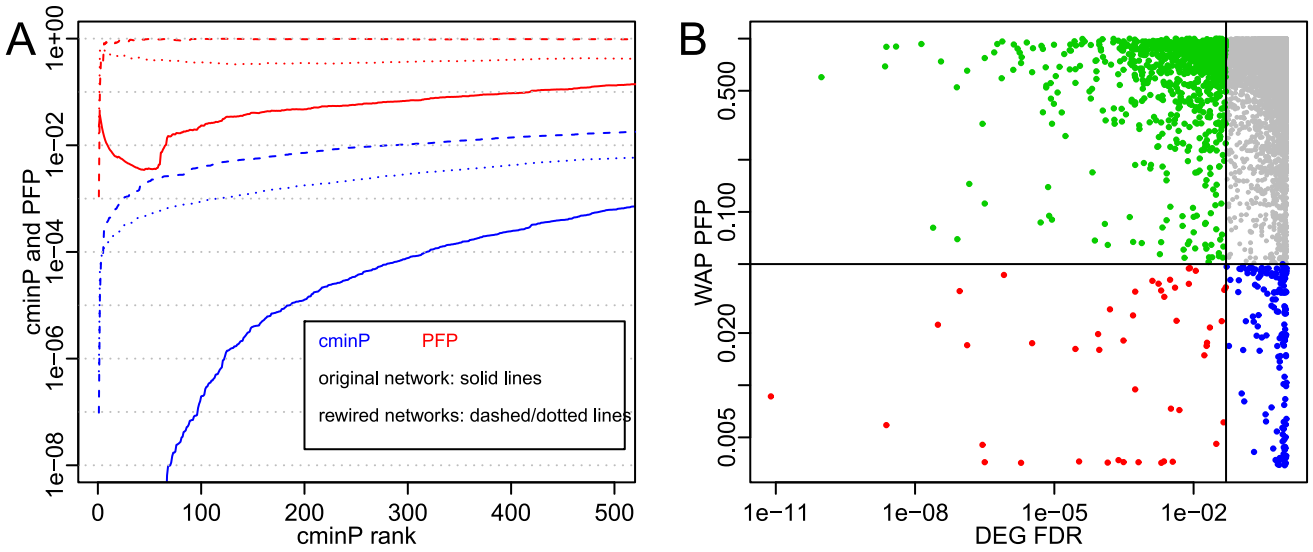

Fig 2. WAP and DEG scores for data set GSE58095. A: c min $P(j)$ values (blue solid line) and PFP values (red solid line, 10^4 permutations) as a function of their rank j. Dashed lines correspond to values obtained with a randomly rewired protein network and dotted lines to a randomly rewired network preserving the number of triangles. B: Scatterplot of WAP PFP and DEG FDR values. Thresholding DEG FDR and WAP PFP at 5% yields four categories of genes based on their significance: DEG and not WAP (green), DEG and WAP (red), WAP and not DEG (blue) and neither DEG nor WAP (gray).

https://doi.org/10.1371/journal.pcbi.1007684.g002

shuffled between samples (thus preserving gene co-expression), genes are ranked for differential expression and values of c min $P^{(j)}$ are estimated again. Observed values are compared to simulated ones: for the observed WAP score of rank j, the simulated profile yields k more significant scores, hence a first estimation of $\mathrm{PFP}^{(j)} = \min(k/j, 1)$. The estimation is refined by averaging values of $\mathrm{PFP}^{(j)}$ over at least 10^3 simulations. This process is illustrated with Fig 1B. WAPs with small PFP values are said to be significantly associated to the most differentially expressed genes.

Fig 2 provides an example of a c min P profile obtained with data set GSE58095 [43], where skin samples of Systemic Sclerosis (SSc) patients are compared to skin samples of healthy subjects. Differential expression is assessed by a two-sided t-test. The solid blue line in Fig 2A displays observed values of c min $P^{(j)}$ as a function of rank j. For the sake of display, the vertical scale has been restricted to values between 10^{-8} and 1, and ranks $j \leq 70$ yield smaller values down to c min $P^{(1)} \simeq 10^{-38}$. The red solid line shows PFP values for each observed c min $P^{(j)}$ estimated with 10^4 permutations. There are 215 genes with PFP value less than 0.05, suggesting that DEGs in SSc versus healthy are specifically organized within the network of known protein functional relationships. To illustrate the importance of biological knowledge encoded in the protein network on WAP scores, results are also displayed for a randomly rewired network, as described in [34], which preserves vertex degrees (dashed lines) and for a similarly rewired network but containing the same number of triangles as the original network (dotted lines). The utilized method to create random triangles while preserving vertex degrees is described in [44]. Randomly rewired networks do not yield small PFP values, except for one gene: Ubiquitin C. This is because it connects to nearly half of the vertices on the network, so that the probability of the removal by random rewiring of all of the existing ubiquitin C interactions is negligibly small. All other proteins yield PFP values greater than 0.1 after rewiring, showing that true protein functional relationships are required to obtain significant WAP scores.

In Fig 2B, genes are partitioned into four categories by choosing a threshold of 0.05 on WAP PFP and/or DEG FDR, so as to yield an expected false-positive rate of only 5% for both DEGs and WAPs that are deemed significant. Gray dots correspond to genes which are neither significant DEGs (FDR > 0.05) nor significant WAPs (PFP > 0.05). Significant DEGs are represented by the union of green and red dots, for a total of $\approx$1100 genes. Among those, $\approx$50 are significant WAPs as well (red dots). These genes are not only differentially expressed, but also significantly connected to other DEGs via known interactions and this can be seen as additional evidence for their potential involvement in SSc. Lastly, significant WAPs which are not significant DEGs are represented by blue dots ($\approx$160 genes). These genes can only be identified by combining gene expression and protein network knowledge: even though they are not significantly perturbed at the mRNA level, their significant connectedness on the network to the most perturbed genes in the SSc-normal comparison suggests their potential involvement in SSc.

Quantification of gene scores reproducibility across data sets

In order to compare rankings of WAP and DEG scores for their robustness across data sets, it is required to define a measure of reproducibility. Reproducibility is quantified with the Jaccard index, which is a well-accepted measure of overlap [45]. Call $\mathcal{A}_i$ and $\mathcal{B}_i$ the top i genes, based on WAP or DEG scores, in two data sets $\mathcal{A}$ and $\mathcal{B}$. The two data sets have been reduced to their common n genes, which are also represented in the protein network. The relative overlap of sets $\mathcal{A}_i$ and $\mathcal{B}_i$ is the Jaccard index:

$$R_i = \frac{I_i}{U_i}, \quad \text{with } I_i = |\mathcal{A}_i \cap \mathcal{B}_i|, \quad U_i = |\mathcal{A}_i \cup \mathcal{B}_i|, \tag{5}$$

and where $|\mathcal{S}|$ stands for the number of genes in set $\mathcal{S}$. Plotting R_i as a function of U_i provides a display of overlap between two data sets. The two curves obtained with WAP and DEG scores can then be visually compared. For $i \to n$ one trivially has $R_i \to 1$. But local maxima of R_i for values of U_i less than n are indicative of remarkable overlap.

To illustrate the quantification of reproducibility, two Systemic Sclerosis (SSc) data sets are considered: GSE58095 [43] and GSE32413 [46]. Gene expression was measured in skin samples of both SSc patients and non-SSc subjects. Differential expression is assessed by an absolute t-statistic between the two groups. Fig 3A displays overlap profiles across the two data sets, i.e. graphs of R_i as a function of U_i, for WAP (blue) and DEG (green) scores. Relative overlap R_i tends to be higher with WAP scores than with DEG scores.

Overlap profiles provide visual representation of reproducibility across two data sets. To facilitate assessment over many pairs, two summary statistics are defined and illustrated with Fig 3B. The first statistic is called maxR and represents the height of the peak in the overlap profile:

$$\text{maxR} = \max_{a \le U_i \le b} R_i. \tag{6}$$

Lower bound $a = 50$ is chosen to avoid large values due solely to the discrete nature of R_i at small i values. Upper bound $b = 1,000$ is chosen, so that R_i does not trivially increase with i ($R_n = 1$ and $n \simeq 14,000$). The second statistic is the area under the curve (overlap profile) between $i = 1$ and $i = b$:

$$\text{AUC} = \frac{w_1 R'_1 + w_b R'_b}{2} + \sum_{i=2}^{b-1} w_i R'_i \quad \text{with} \quad w_i = U'_{i+1} - U'_i. \tag{7}$$

Notice that lower bound $a = 50$ on i is not utilized, because contribution of $i \le a$ to the sum is

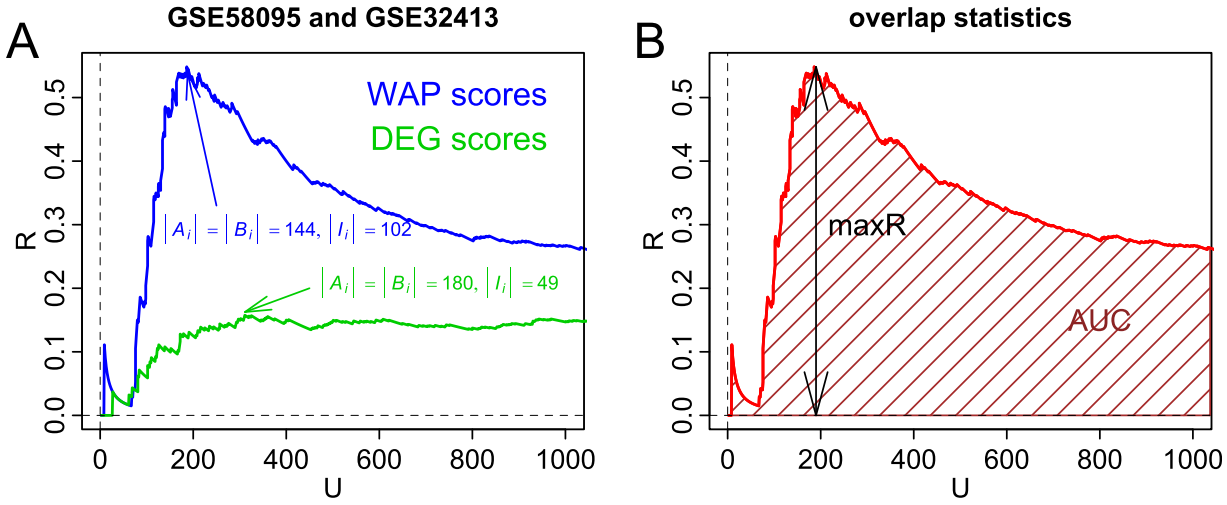

Fig 3. Illustrations of overlap profiles. A: Overlap profiles of WAP (blue) and DEG (green) scores across two gene expression data sets, where skin samples of SSc patients are compared to samples of non-SSc subjects. Values of $|\mathcal{A}_i| = |\mathcal{B}_i|$ and $|\mathcal{I}_i|$ indicate sizes of gene sets and intersection between them corresponding to maxR values for WAP and DEG score profiles. B: The two summary statistics of an overlap profile.

https://doi.org/10.1371/journal.pcbi.1007684.g003

small ($a/b = 0.5\%$). The prime notation is used with U and R to reflect that these variables need to be modified when two or more successive values of U_i are equal: R'_i is then the average of the R_i values and U'_i is the smallest U value. As empirically demonstrated below using random partitioning of gene expression datasets, maxR and AUC statistics, although strongly correlated, are far from being entirely collinear, therefore representing complementary metrics for quantifying reproducibility of ranked lists of genes.

Random partitions of a data set in two

The number of available pairs of independent gene expression studies which describe the same disease is limited. To enable larger sampling, an approach consists in partitioning a large data set in two. Namely, a data set having n disease samples and m control samples is randomly split in two data sets of sizes $\lfloor n/2 \rfloor + \lfloor m/2 \rfloor$ and $\lceil n/2 \rceil + \lceil m/2 \rceil$. Even with $n = m = 10$ there are more than 3×10^4 possible partitions. This enables reasonable sampling and thus accurate estimation of overlap statistics distributions for DEG and WAP scores. In addition to being partitioned, data sets can also be perturbed for their disease/control composition. This enables sampling of data set pairs in which individual data sets have tunable characteristics (Section 4 in S1 Text).

A multivariate statistic based on sample dissimilarities

A statistic to summarize the entire differential expression (i.e. over all genes) between two groups of samples is estimated as follows. Call d_{ij} the Pearson dissimilarity [47] between two samples i and j. Values of d_{ij} close to 0 mean that the two samples have similar expression values across all genes, and values close to 1 indicate dissimilarity. Such dissimilarities are commonly used, for instance in cluster analysis [48]. Call $\mathcal{D}$ and $\mathcal{C}$ disease and control groups of samples. A statistic called MVT, for Multi-Variate T [49, 50], is defined by

$$\text{mvt}\,(\mathcal{D},\mathcal{C}) = \frac{d(\mathcal{D},\mathcal{C})}{s(\mathcal{D}) + s(\mathcal{C})} \quad \text{with} \quad d(\mathcal{D},\mathcal{C}) = \frac{\sum_{i\in\mathcal{D}, j\in\mathcal{C}} d_{ij}}{|\mathcal{D}||\mathcal{C}|}$$
$$\text{and} \quad s(\mathcal{D}) = \frac{2\sum_{i<j\in\mathcal{D}} d_{ij}}{|\mathcal{D}|(|\mathcal{D}|-1)}. \tag{8}$$

Much like a t-statistic, mvt is the ratio of difference between groups (d) to their spread (s). Large values of mvt therefore indicate separation of the two groups. To define what large is, permutation testing is utilized. The observed value (mvt) is compared to values (MVT) obtained when randomly shuffling samples between groups $\mathcal{D}$ and $\mathcal{C}$, and the numerically estimated p-value

$$p = \Pr(\mathrm{MVT} \geq \mathrm{mvt}) \tag{9}$$

is small compared to 1 if the two groups are markedly different.

Protein networks and gene expression data sets

The network of protein functional relationship utilized throughout this study is based on STRING V10 [32]. Interactions were restricted to those having a confidence score of at least 0.7 [51]. Results presented next for Systemic Sclerosis data sets are shown to be robust to changing this threshold (Section 4.2 in S1 Text). Additionally, PPI networks evaluated and/or derived in [52] and deposited to the Network Data Exchange (NDEx) repository [53–55] have been retrieved from NDEx using their identifiers provided by [52] ("Deposited Data" in "Key Resources Table" therein) using the Bioconductor package "ndexr" [56]. All utilized gene expression data sets were downloaded from the Gene Expression Omnibus database [57].

Results

Robustness of WAP scores across Systemic Sclerosis studies

Greater robustness of WAP scores as compared to that of traditional DEG scores is first illustrated with four gene expression Systemic Sclerosis (SSc) data sets (GSE58095 [43], GSE32413 [46], GSE9285 [58] and GSE45485 [59]). Skin samples of SSc patients are compared to skin samples of healthy subjects.

To first assess reproducibility in a visual way, overlap profiles over the entire set of the top 1000 DEGs and WAPs are utilized. Panels A and B of Fig 4 display overlap profiles obtained with two pairs of data sets. Solid green lines show results obtained when ordering genes by their differential expression (SSc vs. healthy) based on a two-sided t-test. Solid blue lines display overlap profiles over the two data sets for WAP scores and green lines correspond to DEG scores. Ranks of WAP scores tend to be more reproducible than those of DEG scores, as attested by larger values of R over most of the overlap profile. WAP scores obtained with randomly rewired networks (brown) are less reproducible than DEG scores, illustrating that superior robustness of WAP scores over DEG scores relies on the existence of true biological interactions between gene products and is not a trivial artifact of different scoring between WAPs and DEGs.

Dashed lines in Fig 4A correspond to the results obtained by combining two-sided t-test and average fold change to account for the observation that "ranking and selecting differentially expressed genes solely by the t-test statistic predestine a poor concordance in results" made by one of the early community-wide cross-institutional assessments of the quality of microarray data [4]. Namely, genes are ordered for differential expression by the average of their absolute values of t-statistic and fold-change ranks. Green dashed lines tend to be higher than solid green lines. This is in agreement with the fact that combining fold change and t-test tends to increase reproducibility between data sets [4]. It is also informative to notice that blue dashed lines tend to be higher than solid blue lines. That is, an increase in DEG score reproducibility tends to yield a larger WAP score reproducibility, and it remains above that of DEG scores.

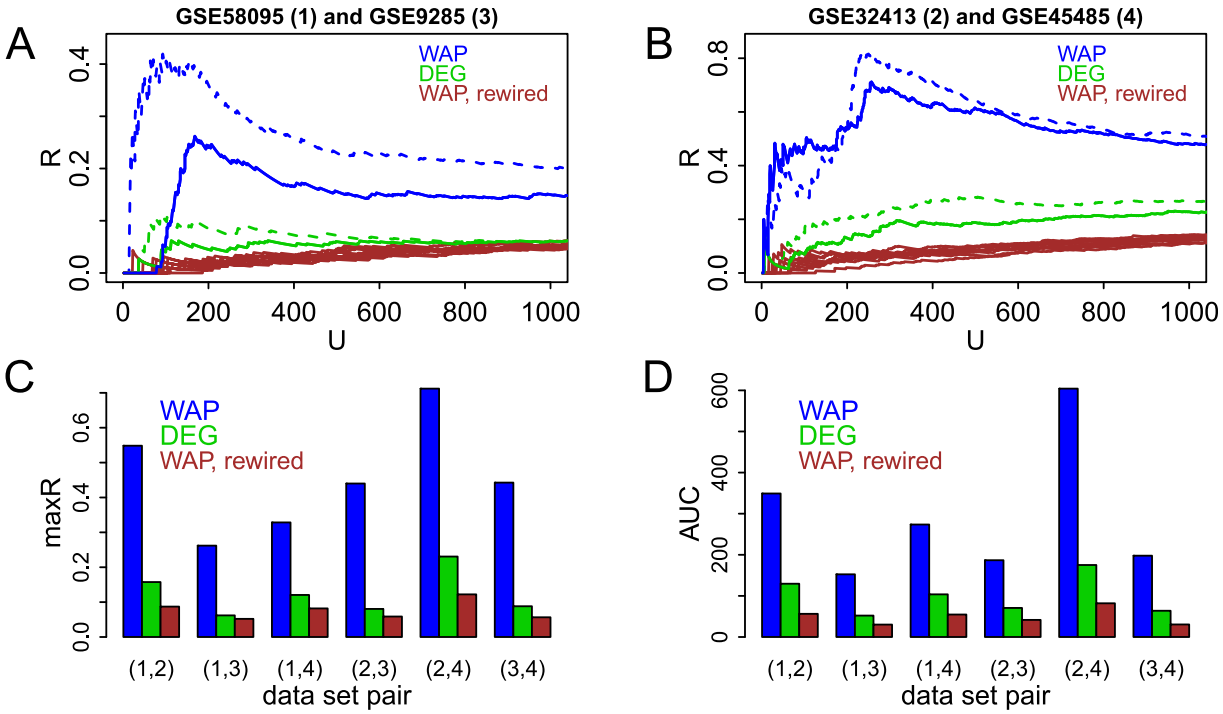

Fig 4. Reproducibility of WAPs and DEGs among four SSc data sets. Four gene expression data sets comparing skin samples between SSc patients and non-SSc subjects—GSE58095 (1), GS32413 (2), GSE9285 (3) and GSE45485 (4)—are used to obtain WAP and DEG scores. Blue and green colors represent overlaps between top scoring WAPs and DEGs respectively. A-B: Overlap profiles for two pairs of data sets. Solid profiles correspond to genes ordered by a two-sided t-test and dashed profiles—to ordering which further accounts for fold change (a gene rank is the average of its t-statistic and fold-change ranks). Brown profiles display results for ten randomly rewired protein networks. C-D: Overlap statistics maxR and AUC for all six data set pairs, when genes are ordered by a two-sided t-test. For randomly rewired networks, displayed values are averages.

https://doi.org/10.1371/journal.pcbi.1007684.g004

Panels C and D of Fig 4 summarize score reproducibility over the six possible pairs of SSc data sets with overlap statistics maxR and AUC (as explained in "Materials and methods"). Both statistics are larger with WAP scores than DEG scores, hence demonstrating the larger reproducibility of WAP scores' rankings over the six data set pairs. WAP scores obtained with randomly rewired networks on average tend to yield lower reproducibility than DEG scores, showing again that true biological knowledge encoded in the protein network is a likely driver of WAP score robustness. Additionally, results shown in Sections 4.3 and 4.4 of S1 Text illustrate that such higher reproducibility of the ordering of WAP scores is relatively robust with respect to: a) removal of interactions between co-expressed genes in the network, b) partial random rewiring of the network (as approximation for simultaneous introduction of both false negative and false positive interactions to the graph), and c) reproducibility of WAP findings is less sensitive to false positives as compared to false negatives randomly added to the pathway networks.

While higher reproducibility of WAP findings as compared to that of DEGs is observed with the considered SSc data sets, it is obviously not a feature universal to all possible data set pairs. Robustness actually requires that a data set contains differential expression for disease versus control which is specifically organized within the protein network. This topic will be explored later. First, higher robustness of top WAPs than that of top DEGs is shown to hold over diseases others than SSc.

Validation of WAP score robustness in other diseases

To further evaluate robustness of the rankings of WAP scores versus those of DEG scores, a large number of data set pairs is required. There however exists only a limited number of available independent gene expression studies which focus on the same disease and the same tissue. This limitation can be alleviated by partitioning individual data sets with a sufficiently large number of samples.

Namely, a data set can be randomly split in two data sets with the goal of comparing disease and control samples. If n is the smallest number of disease or control samples, then a lower bound for the possible number of random partitions is $n!^2/(n/2)!^4/2$. With n as small as 10 this gives over 3×10^4 possible splits, and thus enables reasonable statistical estimation. Note that comparing robustness of WAPs' rankings against that of DEGs in partitions of individual data sets is actually a quite conservative approach, because reproducibility of gene expression findings is expected to be the highest within individual studies.

In the following random data-set partitions, DEGs are defined by comparing disease and control samples via a two-sided t-test. Included diseases are colon cancer (GSE41258 [60], GSE44076 [61], GSE44861 [62]), endometriosis (GSE51981 [63]), gastric cancer (GSE13195 [64], GSE19826 [65], GSE27342 [66], GSE30727 [67], GSE63089 [68], GSE79973 [69]), hepatocellular carcinoma (GSE36376 [70]), non-small cell lung carcinoma (NSCLC; GSE19188 [71]), lung adenocarcinoma (GSE43458 [72]), oral squamous cell carcinoma (OSCC; GSE30784 [73]) and psoriasis (GSE13355 [74], GSE30999 [75], GSE34248 [76], GSE41662 [76]). Distributions of overlap statistic maxR which are estimated over one thousand random partitions of each data set are displayed in Fig 5. Panels A to H detail results over eight individual data sets and diseases. Vertical bars near the top display 99% confidence intervals on median values of maxR. Reproducibility of the top WAPs (blue) tends to be on average significantly higher than that of the top DEGs (green) and approximately comparable to the reproducibility of the WAPs ordered by their PFP values (pink). Similar plots for the entire collection of the data sets cited above are shown in Section 5 of S1 Text that also includes using AUC to quantify reproducibility of top WAPs and DEGs (median Spearman correlation of maxR and AUC values for WAP score profiles over 1,000 random partitions of each dataset in two across all 22 NCBI-GEO datasets analyzed in this study is $\rho = 0.86$, interquartile range 0.27). For reference, plots shown therein also include results obtained on the network rewired to preserve on average 50% of the original edges.

As was already observed with SSc data sets, rankings of WAP scores with randomly rewired protein networks (brown) tend to have lower reproducibility than those of DEG scores. This, along with demonstrated lower informativeness of recently introduced differential expression (DE) prior [77] for predicting significant WAPs (as compared to that for DEGs—Section 7 of S1 Text), demonstrates again that true biological knowledge encoded in the protein network is required for robustness of WAP scores. Furthermore, comparison of the reproducibility of gene ranking by their WAP and DEG scores for a selected subset of these datasets that represent the same disease (and therefore are also impacted by inter-study variability) across 23 different PPI networks made available through NDEx [53–55] by Huang et al. [52] (see Section 6 in S1 Text) illustrates higher robustness of WAP score ranks as compared to that of DEGs for multiple PPI networks, especially for those with higher information content.

Fig 5I shows distributions of maxR differences (WAP maxR minus DEG maxR, values being paired in each partition) that are obtained with all eighteen data sets. Based on the raw WAP score, but not its PFP value, all data sets but one yield distributions which are shifted towards positive values, indicating higher reproducibility of WAP scores versus DEG scores. The only exception is colon-cancer data set GSE44861. Values of maxR in GSE44861 are small

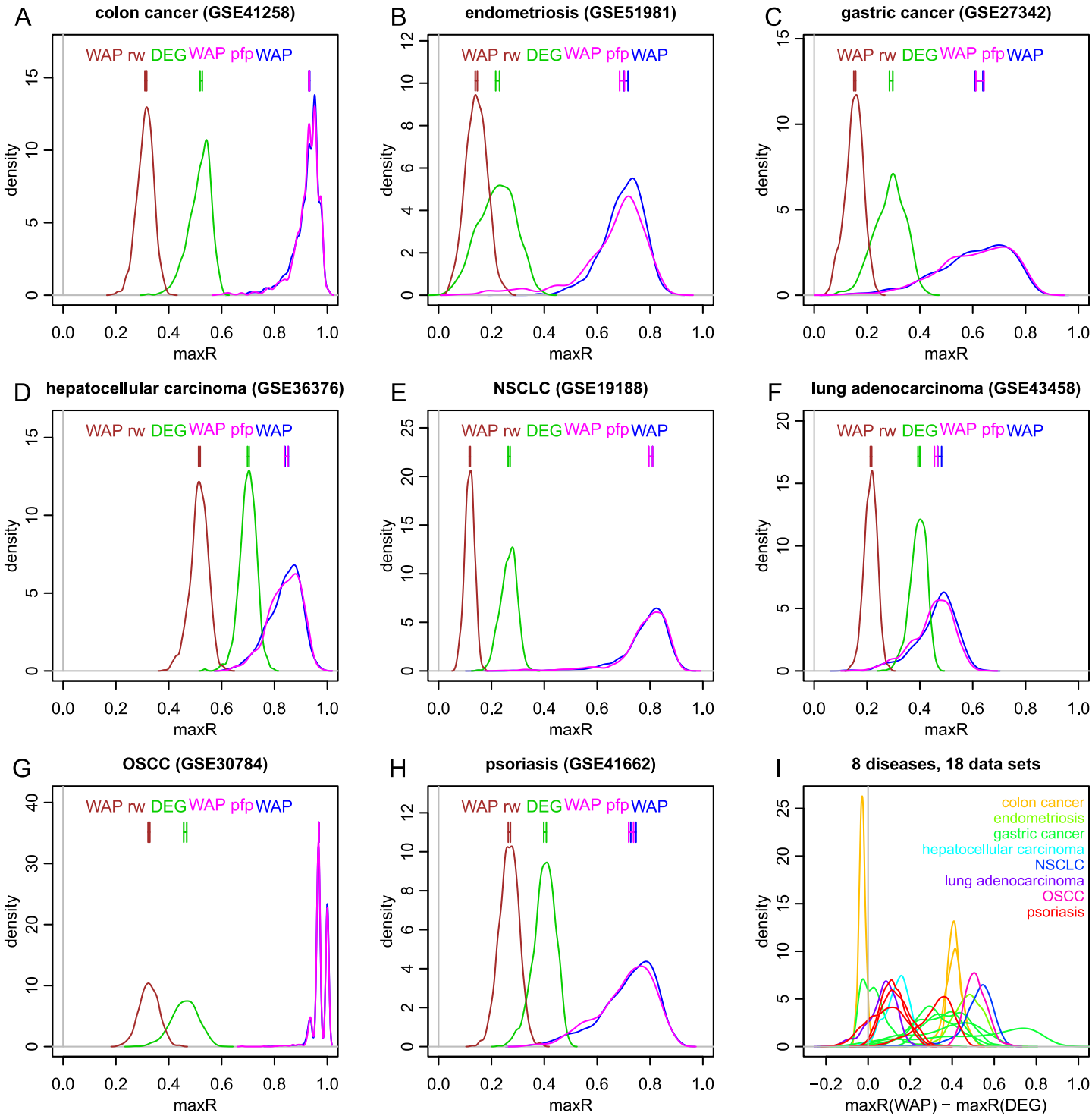

Fig 5. Comparison of WAP and DEG scores reproducibility in eighteen large gene expression data sets representing eight diseases. A-H: Each data set is randomly split in two sets one thousand times, and the resulting distributions of overlap statistic maxR are estimated for WAP scores (blue), PFP values of WAPs (pink), DEG scores (green) and WAP scores with randomly rewired protein network (brown). Vertical bars display 99% confidence intervals on median values estimated by bootstrapping with 10^6 samples. I: summary of results over 18 data sets for distributions of maxR differences between DEG and WAP scores (paired values of maxR in each partition (WAP versus DEG scores)).

https://doi.org/10.1371/journal.pcbi.1007684.g005

(less than 0.1) for both WAP scores and DEG scores. Moreover, for this data set overlap statistics on WAP scores have similar distributions with original and randomly rewired protein networks and even lower for WAP scores ordered by their PFP values (Figure 18, bottom row, in S1 Text). Differential expression in this data set is therefore not specifically organized within the protein network and the smallest PFP value is quite large (0.88). The other seventeen data sets, representing eight diseases, show that reproducibility of WAP scores tends to be on average higher than that of DEG scores. In summary, even though the random partition of a data set in two is a rather conservative control that disregards between studies variability, WAP scores tend to be on average significantly more robust than DEG scores over such partitions in seventeen data sets which represent eight diseases.

Utility of WAP scores for small data sets

The potential value of the WAP score when a data set has a limited number of samples is illustrated next with Psoriasis data set GSE30999 [75], which has a rather large number of samples (83 lesions and 81 no-lesion samples). Pairs of data sets of varying sizes $2m$ (m lesions and m no-lesion) are drawn one thousand times for each value of m, and average values of maxR are estimated for both DEG and WAP scores. Bootstrapping with 10^6 trials is then utilized to estimate 95% confidence intervals on these averages. Results are presented in Fig 6. Panel A shows that for large values of m ($m \geq 30$) average values of maxR are quite large (e.g. at least 0.5) for DEG scores, with even larger values for WAP scores. Large DEG reproducibility is less likely to happen with small sample sizes m. This can be seen in Fig 6A: values of maxR become smaller for both DEG and WAP scores when m decreases.

Furthermore, it is instructive to examine the difference of maxR between WAP and DEG scores. Results are displayed in Fig 6B. Because WAP and DEG scores are not independent, confidence intervals on maxR differences are estimated via bootstrapping on paired values for WAPs and DEGs. One can see that even though maxR values of WAP and DEG scores decrease as m becomes smaller (A), the maxR difference (WAP minus DEG) instead increases until m becomes less than 5 (B). This clearly shows that WAP scores are less sensitive than DEG scores to reduction of sample size m. When DEG score reproducibility is small due to a

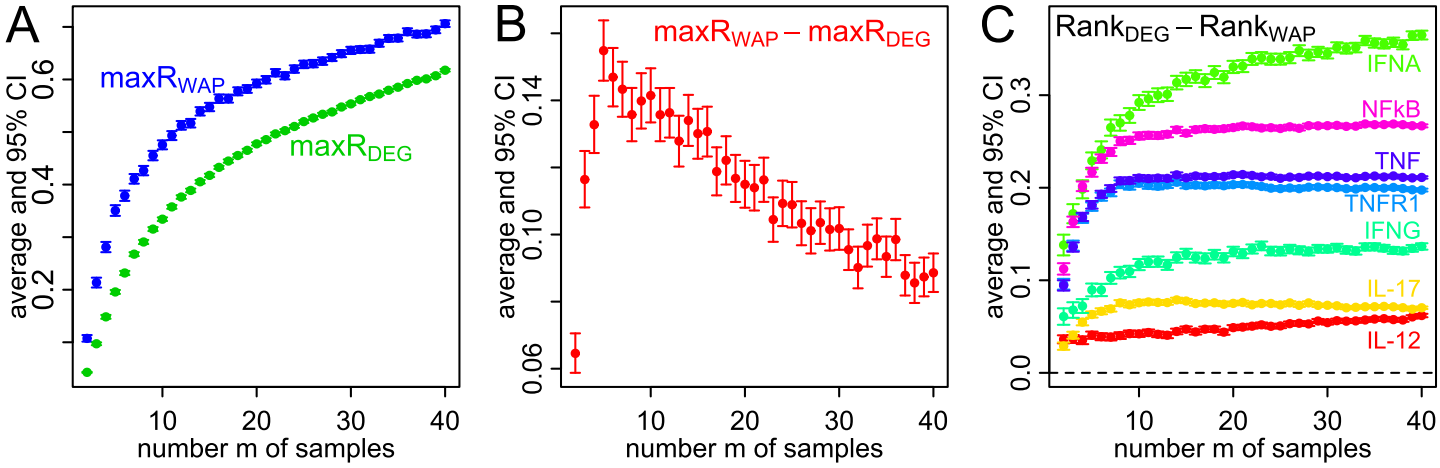

Fig 6. Comparison of gene expression in m skin samples with psoriasislesions to m healthy skin samples. Statistic maxR is estimated with 10^3 draws of two data sets of size $2m$ from GSE30999. Differences between average ranks of pathway members by WAP and DEG are estimated for 10^3 data sets of size $2m$ randomly drawn from GSE30999. Confidence Intervals (CI, 95%, bootstrapping with 10^6 trials, vertical lines) are estimated for average maxR values (dots) based on WAP scores, DEG scores (A), their differences (B) and average difference between pathway member ranks based on DEG and WAP scores (C). Color and labels represent the following Reactome pathways—IFNA (green): Regulation of IFNA signaling, NFkB (magenta): TNFR1-induced NFkappaB signaling pathway, TNF (dark blue): TNF signaling, TNFR1 (light blue): Regulation of TNFR1 signaling, IFNG (teal): Regulation of IFNG signaling, IL-17 (yellow): Interleukin-17 signaling, IL-12 (red): Interleukin-12 family signaling.

https://doi.org/10.1371/journal.pcbi.1007684.g006

small number of samples, WAP scores can in some cases yield a significantly more robust gene ranking than DEG scores.

Finally, it is interesting to note that for this psoriasis dataset besides achieving higher reproducibility as quantified by maxR statistic, WAP scores also result in more prominent, as compared to that by DEG, ranking of gene sets that are representative of pathways known to be involved in the psoriasis pathogenesis. Fig 6C depicts average differences between the ranks of the significance of DEG and WAP scores (where lower rank represents greater significance, i.e. smaller p-value) for the genes included in Reactome [13] pathways selected for their importance in, and history of clinical development for, this disease (e.g. signaling by TNF, IFN-gamma, IL-12, etc.) [78].

Across all sample sizes evaluated here, the average ranks of the members of these pathways are consistently lower (representing smaller, more significant, p-values) by WAP than by DEG scores. Here the rank of zero corresponds to the most significant DEG or WAP score, and the rank of one represents the least significant WAP or DEG score across the entire set of genes included in the analysis. On average, in case of this psoriasis dataset, WAP scores of the members of Reactome pathways depicted in Fig 6C rank more significantly than DEG scores by about 5% to 35% of the size of the entire gene set. Such more prominent ranking by WAP scores of pathways well recognized for their involvement for psoriasis pathogenesis further emphasizes potential merits of the WAP score for identification of new molecular targets for therapeutic intervention.

Necessary conditions for WAP score robustness

Robustness of the gene ranking by their WAP scores is obviously not guaranteed for an arbitrary data set. Conditions which are required for robustness of WAP scores are now examined. Robustness relies on meaningful signal contained in differential expression with respect to biological information encoded in the protein network. This was first demonstrated by showing that WAP scores obtained with randomly rewired networks yield lower reproducibility than with the original network and on average lower than reproducibility of DEG scores (Figs 4 and 5). An already-introduced method for quantifying how specifically differential expression is organized within the protein network is estimation of a PFP profile (Fig 2). Intuitively, small PFP values of the top ranking WAPs suggest that the top WAPs might be robust to small changes of DEG ranks.

To test this conjecture, simulations based on SSc data set GSE58095 [43] are utilized. GSE58095 is chosen because it yields large differences in overlap statistics (maxR and AUC) between genes ordered by their WAP and DEG scores (skin samples of SSc patients versus non-SSc subjects), when this large data set is randomly partitioned in two (Section 4.4 in S1 Text). Partitions tend to yield two data sets which have small WAP PFP values, as measured for instance by the average PFP value over the top 200 WAPs. In order to explore a wider range of average PFP values, the two data sets of a partition are randomly perturbed for their composition in disease/control samples. Perturbations are controlled by a parameter $0 \leq \rho \leq 1$, which can be seen as the probability of swapping disease/control state between two samples. Setting $\rho = 0.41$ yields a close to uniform distribution of PFP values averaged over the 200 top WAPs (Section 8.1 in S1 Text). With this value of ρ, 9×10^4 pairs of data sets are sampled, so as to approximately draw 100 pairs in each cell of a 30×30 uniform grid. An additional 9×10^4 partitions are generated with $\rho = 1$ for reasons which will be explained next, and 2×10^4 partitions are also sampled with $\rho = 0$.

Fig 7 summarizes results obtained with the 2×10^5 sampled partitions of GSE58095. On average, both overlap statistics (maxR and AUC) tend to be higher for WAP scores as

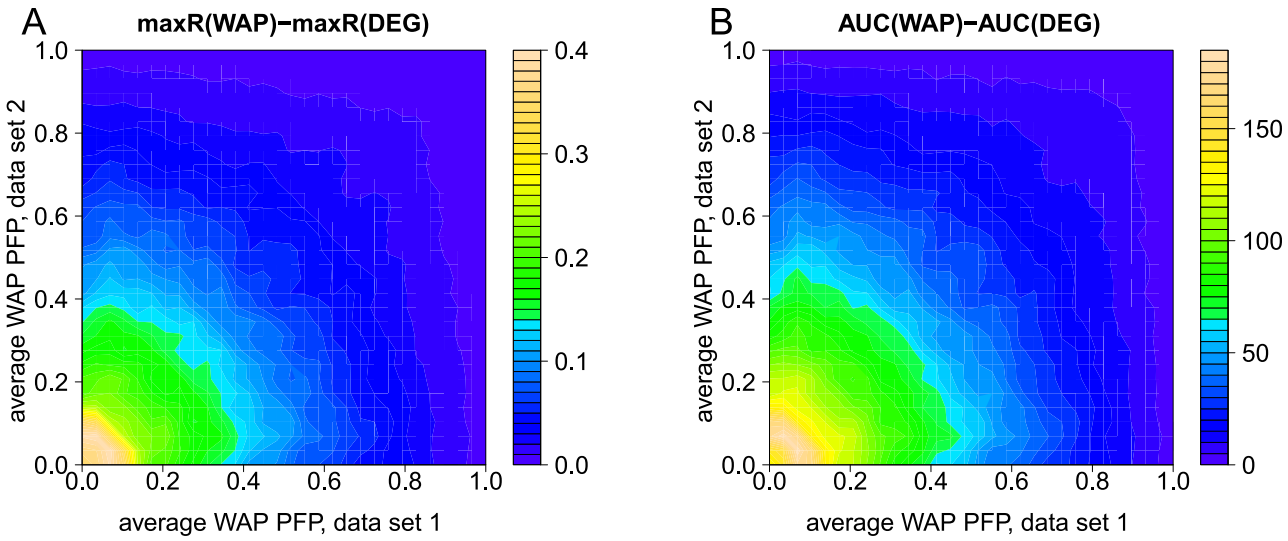

Fig 7. Differences in overlap statistics between WAP and DEG scores as a function of the average PFP value. A: Differences between maxR values for WAP and DEG scores. B: Differences between AUC values for WAP and DEG scores. Top 200 WAPs in each data set of a simulated pair are used to calculate maxR, AUC and PFP. Plots summarize results obtained with 2×10^5 random partitions of GSE58095.

https://doi.org/10.1371/journal.pcbi.1007684.g007

compared to DEG scores, when both data sets are rich in WAPs having small PFP values (lower left corners). Absence of small PFP values in both data sets yields similar reproducibility of WAP and DEG scores (upper right corners). Notice that if one of the data sets has a small average PFP value (e.g. 0.05), then it might still yield better reproducibility of WAP scores versus DEG scores even if the second data set has a larger average PFP value (e.g. 0.3). One can therefore state that a necessary condition for larger reproducibility of WAP scores versus DEG scores between two data sets is that at least one of the data sets yields WAP scores having small PFP values. A large proportion of small PFP values indicates that observed differential expression is specifically organized within the protein network, as compared to differential expression obtained when randomly reassigning disease states to samples.

One can also characterize differential expression just for its magnitude, independently of the protein network. This can be done, for instance, with the average FDR value [42] over the top 200 DEG scores. This would however be too computationally expensive given the very large number of sampled data set pairs. Instead, a more efficient approach is based on a statistic of Pearson dissimilarities between entire samples, i.e. dissimilarities based on all genes. The statistic is called Multi-Variate T (MVT) and significance of its value is assessed via permutation testing (as explained in "Materials and methods"). Small MVT p-values correlate well with small average FDR values over the top 200 DEGs (Section 8.3 in S1 Text). The advantage of the MVT p-value is that the control distribution can be estimated once and then rapidly utilized with all simulated data set pairs (a similar approach was utilized to rapidly estimate average PFP values over the top 200 WAPs). The disease/control mixing parameter is now set to $\rho = 1$, and 9×10^4 pairs of data sets are sampled. The rationale is to sample approximately 100 data set pairs in each cell of a 30×30 uniform grid for MTV p-values between 0 and 1 (Section 8.1 in S1 Text).

Average overlap statistics maxR obtained with all 2×10^5 random partitions of GSE58095 are displayed in Fig 8 as a function of MVT p-values. Panel A indicates that average maxR of DEG scores never reaches values higher than 0.2. Even when the two data sets of a pair both have small MVT p-values (lower left corner), DEG reproducibility tends to remain low, and

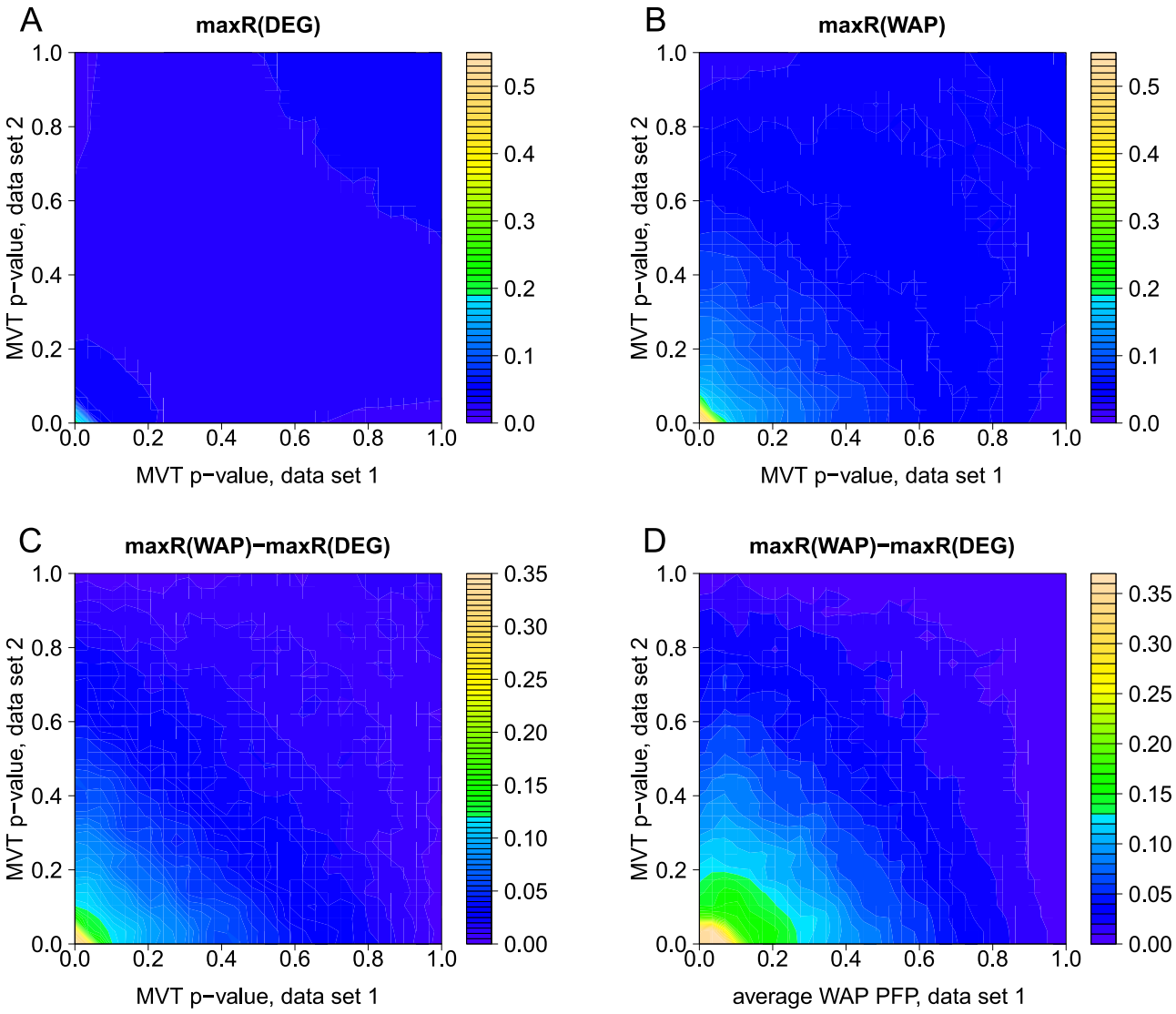

Fig 8. Overlap statistic maxR for DEG and WAP scores as a function of MVT p-values. A-B: Overlap statistic maxR for DEG and WAP scores as a function of MVT p-values in each data set of a simulated pair. C: Difference of maxR statistic (WAP minus DEG) as a function of MVT p-values in each data set of a simulated pair. D: Difference of maxR statistic (WAP minus DEG) as a function of MVT p-value in one data set and average WAP p-value in the other. Plots summarize results obtained with 2×10^5 random partitions of GSE58095.

https://doi.org/10.1371/journal.pcbi.1007684.g008

this occurs with data set pairs extracted from the same study. Fig 8B reveals a stronger effect of small MVT p-values on reproducibility of WAP findings. Average values of maxR go up to 0.55 (bottom left corner). Increase of average maxR values of WAP scores with small MVT p-values is larger than that of maxR of DEG scores, as can be seen in Fig 8C. Finally, Fig 8D illustrates that small PFP values of WAP scores are better predictors of superior robustness of top WAP ranks versus top DEG ranks than small MVT p-values.

In summary, robustness of WAP scores across data sets tends to increase when data sets exhibit significant degree of differential gene expression, as measured for instance by a small MVT p-value or a large proportion of small FDR values on DEG scores. Most importantly, higher reproducibility of top WAPs versus top DEGs is more pronounced when differential

expression appears to be specifically organized within the protein network, i.e. when WAP scores yield small PFP values.

Discussion

The importance of interactome annotation for identification and prioritization of targets for therapeutic strategies [79–81] and disease characterization [82–84] is well established with growing evidence for distinct properties of network connectivity associated with drug targets [85, 86] and disease genes [87]. The need for improving reproducibility of the findings from genome-wide studies characterizing differences between healthy state and disease remains an important analytical challenge for the field [88].

Results presented in this paper have rigorously demonstrated that utilizing prior biological knowledge in the form of known protein functional relationships can significantly enhance reproducibility of the findings from gene expression analysis, at least in the context of characterizing a disease, therefore potentially providing more robust description of the phenomenon under study. Demonstration was performed with thousands of data set pairs, which were generated with twenty-two large gene expression data sets representing nine different diseases. By scoring the same universe of genes with both WAP and DEG scores, comparison of the reproducibility of the findings by these two approaches could be made in a rigorous way. This focus on making the reproducibility of the findings made with and without prior biological knowledge in the form of the PPI network directly comparable sets it apart from other methodologies that enable analysis of genome-wide molecular characterization data with pathway networks (e.g. [23, 24] and others reviewed in [10, 19, 20]).

The method to score gene products for their association to differentially expressed genes, i.e. the WAP score, is well-conditioned for the total number of interactions of each gene. This was achieved by combining published methods [37] and novel correction factors (Section 3 in S1 Text). Besides not requiring to choose a threshold which defines differential expression, another advantage of the WAP score is that its estimation can be implemented efficiently, i.e. in $\mathcal{O}(M)$ where M is the total number of interactions in the protein network. This means that false discovery rates can be numerically estimated via permutation techniques in reasonable time. It also implies that several hundred thousand pairs of data sets can be rapidly scored, so as to provide statistical evidence for higher reproducibility of top WAPs as compared to that of top DEGs. Throughout this study, for illustration purposes, the DEGs were identified by a two-sided t-test (disease vs. healthy) for the reasons of clarity and simplicity. However, ranking of genes in protein interaction network by their WAP scores can be readily obtained for genome-wide ranking of genes by their statistical significance as estimated by more sophisticated approaches [89–91], when necessary for the studies with more complex designs. Evaluating the impact of WAP methodology on the reproducibility of the findings from such more advanced models represents one of the exciting possibilities for follow-up investigations.

Higher reproducibility of WAPs rankings as compared to those of DEGs is obviously not a given fact. Such robustness relies upon sound biological knowledge in the protein network, as attested by the fact that WAPs computed with randomly rewired networks are less reproducible than DEGs. Significant coverage of the interactome is potentially an important factor, even though reducing interactions down to a few tens of thousands having the highest confidence levels [41] can still yield robust WAP findings in the case of systemic sclerosis data sets (Section 4.2 in S1 Text). Evaluation of the impact of edge confidence score on the robustness of WAP score rankings did not yield compelling evidence for choosing a specific threshold for using a subset of STRING network for the analysis of these datasets. This in combination with the observed higher sensitivity of WAP procedure to false negatives than false positives and its

tendency to result in greater gain of reproducibility for larger networks among those evaluated in [52], calls for a generic recommendation of using larger compendiums of PPI data when possible. However, when the size of gene expression data enables evaluation of its intra-study reproducibility by random partitioning, it could be worthwhile verifying the lack of drastic sensitivity of the reproducibility of the findings on the edge confidence threshold if/when available by following a procedure demonstrated in Section 4.4 in S1 Text.

Similar to the recent report [52] that surveyed available genome-wide interaction networks for their ability to recover known disease gene sets, and concluded that the larger PPI networks perform better in this context, greater gain in the reproducibility of WAP findings for larger networks was also observed here as well (Section 6 in S1 Text). Additionally, it could be instructive to evaluate relative contributions of different types and/or source of interactome annotation to WAP score robustness in the future. Preliminary work suggests that, while including co-expression knowledge in interactions is beneficial to the robustness of WAPs versus DEGs, it is unlikely to be the critical element (Section 4.3 in S1 Text). Furthermore, comparison of the rankings of the significant WAPs and DEGs by a recently introduced DE prior [77] across NCBI-GEO datasets evaluated in this study did not detect increased enrichment of WAPs for the genes that are more likely to be differentially expressed across large compendium of transcriptional profiling studies (Section 7 in S1 Text). Finally, the observed tendency of WAP scores to yield more reproducible ranking of genes as compared to that by their DEG scores has been shown to be robust to moderate amounts of random noise introduced in the protein interaction network, especially to false positives (Sections 4.4 and 5 in S1 Text) and to hold for multiple PPI networks evaluated and/or derived by [52] (Section 6 in S1 Text).

Besides being dependent on quality of protein interactions, the robustness of WAP findings also relies upon the signal encoded in gene expression data. Top WAPs are more robust when differential expression is significantly high, e.g. when it yields a large proportion of DEGs with small FDR values. More importantly, it was shown that robustness requires differential expression to be specifically organized within the protein network, i.e. it must yield small PFP values on WAP scores. When this easily-testable condition is satisfied, identified WAPs with small PFP values have the potential to be more robust than top DEGs across data sets, and this can be valuable in the case of small studies, as was demonstrated with a psoriasis data set.

Such an increase in the reproducibility of the findings from gene expression studies with smaller sample sizes by utilizing PPI information with the WAP framework may provide an appealing and cost-effective alternative to increasing reproducibility of DEGs by increasing the number of samples characterized by gene expression. The positive impact of the increase of sample size on the reproducibility of the DEGs observed in gene expression studies is well-recognized and has been extensively studied in the context of microarray and RNA-seq technology (e.g. [92, 93] and references therein). More detailed evaluation of the gains in reproducibility of the findings from gene expression findings due to the use of PPI data and WAP methodology and comparing it to that solely due to the increase in the sample size of gene expression datasets across broad number of biological phenomena and experimental designs represents another promising area of future research.

Limiting gene expression data used by WAP methodology to that for the genes represented in PPI network is an inherent source of information loss for this approach. In the light of observed lower impact of false positive interactions on the robustness of WAP findings and its tendency to yield more robust findings (as compared to those of DEGs) for larger networks, this shortcoming could be partly alleviated by using larger compendiums of PPI data for WAP analyses. Additionally, the analyses reported herein were purposely limited to the overlap of genes represented both in the gene expression data and PPI network for the reasons of clarity and direct comparability of the robustness of WAP and DEG rankings that was the main focus

of this study. It is straightforward to extend WAP calculation to score nodes in pathway network that are not in gene expression data themselves and/or are not reliably detected due to low levels of expression, cross-hybridization, etc., but their network neighbors are. Such an extension, that is technically trivial, would further advance the potential of WAP methodology to reveal important biological aspects of phenomena studied in a manner complementary to the conventional differential expression analysis.

Although the comparison of disease vs. lack thereof was employed in the examples used in this study, the same methodology can be readily applied to characterization of a broader variety of biological systems for which genome-wide measurements of gene expression data and corresponding interactome information are available. For instance, it would be interesting to apply this approach to the data sets characterizing chemosensitivity of cancer cell lines [94] as recent publication by [95] suggests relevance of interactome information in this context. Findings from this type of analysis might be particularly amenable for experimental follow-up to directly test the hypothesis of high pertinence of top WAPs for the phenomenon studied. Extending WAP methodology to the generation of the predictions about individual patient outcomes, similarly to the approach presented in [24], represents another promising direction of future research.

In conclusion, this paper has rigorously demonstrated that utilizing systems-level knowledge about protein functional relationships can significantly enhance reproducibility of disease description via gene expression analysis. Such enhanced reproducibility, which soundly makes use of accumulated prior biological knowledge of diverse types, is likely to be beneficial to devise targeted therapeutic interventions, drug repurposing and potentially to benefit precision medicine investigations.

Supporting information

S1 Text. Supplementary material. Algorithmic details and numerical results further characterizing comparative reproducibility of WAPs and DEGs.
(PDF)

Acknowledgments

The authors are grateful to Vlado Dančík and James J. Collins for their comments.

Author Contributions

Conceptualization: Joël R. Pradines, Victor Farutin, Elma Kurtagic, Anthony M. Manning, Ishan Capila.

Data curation: Nicholas A. Cilfone, Elma Kurtagic.

Formal analysis: Joël R. Pradines, Victor Farutin, Nicholas A. Cilfone, Abouzar Ghavami.

Methodology: Joël R. Pradines, Victor Farutin, Abouzar Ghavami, Jamey Guess.

Software: Joël R. Pradines, Victor Farutin, Nicholas A. Cilfone, Abouzar Ghavami.

Supervision: Anthony M. Manning, Ishan Capila.

Writing – original draft: Joël R. Pradines, Victor Farutin, Elma Kurtagic, Ishan Capila.

Writing – review & editing: Joël R. Pradines, Victor Farutin, Nicholas A. Cilfone, Abouzar Ghavami, Elma Kurtagic, Jamey Guess, Anthony M. Manning, Ishan Capila.

References

1. Tan P, Downey T, Spitznagel E, Xu P, Fu D, Dimitrov D, et al. Evaluation of gene expression measurements from commercial microarray platforms. Nucleic Acids Res. 2003; 31(19):5676–84. https://doi.org/10.1093/nar/gkg763 PMID: 14500831
2. Fortunel N, Otu H, Ng H, Chen J, Mu X, Chevassut T, et al. Comment on "'Stemness': transcriptional profiling of embryonic and adult stem cells" and "a stem cell molecular signature". Science. 2003; 302 (5644):393. https://doi.org/10.1126/science.1086384 PMID: 14563990
3. Shi L, Jones W, Jensen R, Harris S, Perkins R, Goodsaid F, et al. The balance of reproducibility, sensitivity, and specificity of lists of differentially expressed genes in microarray studies. BMC Bioinformatics. 2008; 9 Suppl 9:S10. https://doi.org/10.1186/1471-2105-9-S9-S10 PMID: 18793455
4. Shi L, Reid LH, Jones WD, Shippy R, Warrington JA, Baker SC, et al. The MicroArray Quality Control (MAQC) project shows inter- and intraplatform reproducibility of gene expression measurements. Nature biotechnology. 2006; 24(9):1151–61 PMID: 16964229
5. Zhang M, Zhang L, Zou J, Yao C, Xiao H, Liu Q, et al. Evaluating reproducibility of differential expression discoveries in microarray studies by considering correlated molecular changes. Bioinformatics. 2009; 25(13):1662–8. https://doi.org/10.1093/bioinformatics/btp295 PMID: 19417058
6. Li R, Lin X, Geng H, Li Z, Li J, Lu T, et al. A network-based method to evaluate quality of reproducibility of differential expression in cancer genomics studies. Oncotarget. 2015; 6(42):44714–27. https://doi.org/10.18632/oncotarget.5987 PMID: 26556852
7. Ni S, Vingron M. R2KS: a novel measure for comparing gene expression based on ranked gene lists. J Comput Biol. 2012; 19(6):766–75. https://doi.org/10.1089/cmb.2012.0026 PMID: 22697246
8. Zhang M, Yao C, Guo Z, Zou J, Zhang L, Xiao H, et al. Apparently low reproducibility of true differential expression discoveries in microarray studies. Bioinformatics. 2008; 24(18):2057–63. https://doi.org/10.1093/bioinformatics/btn365 PMID: 18632747
9. Hartwell LH, Hopfield JJ, Leibler S, Murray AW. From molecular to modular cell biology. Nature. 1999; 402(6761supp):C47. https://doi.org/10.1038/35011540 PMID: 10591225
10. Khatri P, Sirota M, Butte AJ. Ten years of pathway analysis: current approaches and outstanding challenges. PLoS computational biology. 2012; 8(2):e1002375. https://doi.org/10.1371/journal.pcbi.1002375 PMID: 22383865
11. Consortium TGO. Gene Ontology Consortium: going forward. Nucleic Acids Res. 2015; 43(Database issue):D1049–56. https://doi.org/10.1093/nar/gku1179
12. Kanehisa M, Goto S, Sato Y, Furumichi M, Tanabe M. KEGG for integration and interpretation of large-scale molecular data sets. Nucleic acids research. 2011; 40(D1):D109–D114. https://doi.org/10.1093/nar/gkr988 PMID: 22080510
13. Fabregat A, Jupe S, Matthews L, Sidiropoulos K, Gillespie M, Garapati P, et al. The reactome pathway knowledgebase. Nucleic acids research. 2017; 46(D1):D649–D655. https://doi.org/10.1093/nar/gkx1132
14. Liberzon A, Subramanian A, Pinchback R, Thorvaldsdóttir H, Tamayo P, Mesirov JP. Molecular signatures database (MSigDB) 3.0. Bioinformatics. 2011; 27(12):1739–1740. https://doi.org/10.1093/bioinformatics/btr260 PMID: 21546393
15. Glazko GV, Emmert-Streib F. Unite and conquer: univariate and multivariate approaches for finding differentially expressed gene sets. Bioinformatics. 2009; 25(18):2348–2354. https://doi.org/10.1093/bioinformatics/btp406 PMID: 19574285
16. Ackermann M, Strimmer K. A general modular framework for gene set enrichment analysis. BMC bioinformatics. 2009; 10(1):47. https://doi.org/10.1186/1471-2105-10-47 PMID: 19192285
17. Tarca AL, Bhatti G, Romero R. A comparison of gene set analysis methods in terms of sensitivity, prioritization and specificity. PloS one. 2013; 8(11):e79217. https://doi.org/10.1371/journal.pone.0079217 PMID: 24260172
18. Mathur R, Rotroff D, Ma J, Shojaie A, Motsinger-Reif A. Gene set analysis methods: a systematic comparison. BioData mining. 2018; 11(1):8. https://doi.org/10.1186/s13040-018-0166-8 PMID: 29881462
19. Mitra K, Carvunis AR, Ramesh SK, Ideker T. Integrative approaches for finding modular structure in biological networks. Nature Reviews Genetics. 2013; 14(10):719. https://doi.org/10.1038/nrg3552 PMID: 24045689
20. Nguyen H, Shrestha S, Tran D, Shafi A, Draghici S, Nguyen T. A comprehensive survey of tools and software for active subnetwork identification. Frontiers in genetics. 2019; 10. https://doi.org/10.3389/fgene.2019.00155
21. Jaakkola MK, Elo LL. Empirical comparison of structure-based pathway methods. Briefings in bioinformatics. 2015; 17(2):336–345. https://doi.org/10.1093/bib/bbv049 PMID: 26197809

22. He H, Lin D, Zhang J, Wang Yp, Deng Hw. Comparison of statistical methods for subnetwork detection in the integration of gene expression and protein interaction network. BMC bioinformatics. 2017; 18 (1):149. https://doi.org/10.1186/s12859-017-1567-2 PMID: 28253853

23. Rossin EJ, Lage K, Raychaudhuri S, Xavier RJ, Tatar D, Benita Y, et al. Proteins encoded in genomic regions associated with immune-mediated disease physically interact and suggest underlying biology. PLoS genetics. 2011; 7(1):e1001273. https://doi.org/10.1371/journal.pgen.1001273 PMID: 21249183

24. Zarringhalam K, Enayetallah A, Reddy P, Ziemek D. Robust clinical outcome prediction based on Bayesian analysis of transcriptional profiles and prior causal networks. Bioinformatics. 2014; 30(12): i69–i77. https://doi.org/10.1093/bioinformatics/btu272 PMID: 24932007

25. Reimand J, Isserlin R, Voisin V, Kucera M, Tannus-Lopes C, Rostamianfar A, et al. Pathway enrichment analysis and visualization of omics data using g:Profiler, GSEA, Cytoscape and EnrichmentMap. Nature Protocols. 2019; 14(2):482–517. https://doi.org/10.1038/s41596-018-0103-9 PMID: 30664679

26. Manoli T, Gretz N, Gröne H, Kenzelmann M, Eils R, Brors B. Group testing for pathway analysis improves comparability of different microarray datasets. Bioinformatics. 2006; 22(20):2500–6. https://doi.org/10.1093/bioinformatics/btl424 PMID: 16895928

27. Kadota K, Nakai Y, Shimizu K. Ranking differentially expressed genes from Affymetrix gene expression data: methods with reproducibility, sensitivity, and specificity. Algorithms Mol Biol. 2009; 4:7. https://doi.org/10.1186/1748-7188-4-7 PMID: 19386098

28. Maglietta R, Distaso A, Piepoli A, Palumbo O, Carella M, D'Addabbo A, et al. On the reproducibility of results of pathway analysis in genome-wide expression studies of colorectal cancers. J Biomed Inform. 2010; 43(3):397–406. https://doi.org/10.1016/j.jbi.2009.09.005 PMID: 19796710

29. Zhang L, Zhang J, Yang G, Wu D, Jiang L, Wen Z, et al. Investigating the concordance of Gene Ontology terms reveals the intra- and inter-platform reproducibility of enrichment analysis. BMC Bioinformatics. 2013; 14:143. https://doi.org/10.1186/1471-2105-14-143 PMID: 23627640

30. da Silva S, Perrone G, Dinis J, de Almeida R. Reproducibility enhancement and differential expression of non predefined functional gene sets in human genome. BMC Genomics. 2014; 15:1181. https://doi.org/10.1186/1471-2164-15-1181 PMID: 25539829

31. Lim K, Li Z, Choi K, Wong L. A quantum leap in the reproducibility, precision, and sensitivity of gene expression profile analysis even when sample size is extremely small. J Bioinform Comput Biol. 2015; 13(4):1550018. https://doi.org/10.1142/S0219720015500183 PMID: 26166210

32. Szklarczyk D, Franceschini A, Wyder S, Forslund K, Heller D, Huerta-Cepas J, et al. STRING v10: protein-protein interaction networks, integrated over the tree of life. Nucleic Acids Res. 2015; 43(Database issue):D447–52. https://doi.org/10.1093/nar/gku1003 PMID: 25352553

33. Tian L, Greenberg SA, Kong SW, Altschuler J, Kohane IS, Park PJ. Discovering statistically significant pathways in expression profiling studies. Proceedings of the National Academy of Sciences. 2005; 102 (38):13544–13549. https://doi.org/10.1073/pnas.0506577102

34. Milo R, Shen-Orr S, Itzkovitz S, Kashtan N, Chklovskii D, Alon U. Network motifs: simple building blocks of complex networks. Science. 2002; 298(5594):824–7. https://doi.org/10.1126/science.298.5594.824 PMID: 12399590

35. Itzkovitz S, Milo R, Kashtan N, Ziv G, Alon U. Subgraphs in random networks. Phys Rev E Stat Nonlin Soft Matter Phys. 2003; 68(2 Pt 2):026127. https://doi.org/10.1103/PhysRevE.68.026127 PMID: 14525069

36. Chung F, Lu L. The average distances in random graphs with given expected degrees. Proc Natl Acad Sci U S A. 2002; 99(25):15879–82. https://doi.org/10.1073/pnas.252631999 PMID: 12466502

37. Pradines J, Farutin V, Rowley S, Dancik V. Analyzing protein lists with large networks: edge-count probabilities in random graphs with given expected degrees. J Comput Biol. 2005; 12(2):113–28. https://doi.org/10.1089/cmb.2005.12.113 PMID: 15767772

38. Le Cam L. An approximation theorem for the poisson binomial distribution. Pacif J Math. 1960; 10:1181–97. https://doi.org/10.2140/pjm.1960.10.1181

39. Farutin V, Robison K, Lightcap E, Dancik V, Ruttenberg A, Letovsky S, et al. Edge-count probabilities for the identification of local protein communities and their organization. Proteins. 2006; 62(3):800–18. https://doi.org/10.1002/prot.20799 PMID: 16372355

40. Pradines J, Dancik V, Ruttenberg A, Farutin V. Connectedness profiles in protein networks for the analysis of gene expression data. Lecture Notes in Bioinformatics. 2007; 4453(RECOMB2007):296–310.

41. Franceschini A, Szklarczyk D, Frankild S, Kuhn M, Simonovic M, Roth A, et al. STRING v9.1: protein-protein interaction networks, with increased coverage and integration. Nucleic Acids Res. 2013; 41 (Database issue):D808–15. https://doi.org/10.1093/nar/gks1094 PMID: 23203871

42. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J Roy Statist Soc Ser B. 1995; 57:289–300.

43. Assassi S, Swindell W, Wu M, Tan F, Khanna D, Furst D, et al. Dissecting the heterogeneity of skin gene expression patterns in systemic sclerosis. Arthritis Rheumatol. 2015; 67(11):3016–26. https://doi.org/10.1002/art.39289 PMID: 26238292

44. Bansal S, Khandelwal S, Meyers L. Exploring biological network structure with clustered random networks. BMC Bioinformatics. 2009; 10(405). https://doi.org/10.1186/1471-2105-10-405 PMID: 20003212

45. Levandowski M, Winter D. Distance between Sets. Nature. 1971; 234:34–35. https://doi.org/10.1038/234034a0

46. Pendergrass S, Lemaire R, Francis I, Mahoney J, Lafyatis R, Whitfield M. Intrinsic gene expression subsets of diffuse cutaneous systemic sclerosis are stable in serial skin biopsies. J Invest Dermatol. 2012; 132(5):1363–73. https://doi.org/10.1038/jid.2011.472 PMID: 22318389

47. Goshtasby AA. Similarity and dissimilarity measures. In: Image registration. Springer; 2012. p. 7–66.

48. Kaufman L, Rousseeuw PJ. Finding groups in data: an introduction to cluster analysis. vol. 344. John Wiley & Sons; 2009.

49. D'Alessandro J, Duffner J, Pradines J, Capila I, Garofalo K, Kaundinya G, et al. Equivalent Gene Expression Profiles between Glatopa™and Copaxone©. PLoS One. 2015; 10(10):e0140299. https://doi.org/10.1371/journal.pone.0140299 PMID: 26473741

50. Sobolev O, Binda E, O'Farrell S, Lorenc A, Pradines J, Huang H, et al. Adjuvanted influenza-H1N1 vaccination reveals lymphoid signatures of age-dependent early responses and of clinical adverse events. Nature Immunology. 2016; 17(2):204–13. https://doi.org/10.1038/ni.3328 PMID: 26726811

51. von Mering C, Jensen L, Snel B, Hooper S, Krupp M, Foglierini M, et al. STRING: known and predicted protein-protein associations, integrated and transferred across organisms. Nucleic Acids Res. 2005; 33 (Database issue):D433–7. https://doi.org/10.1093/nar/gki005 PMID: 15608232

52. Huang JK, Carlin DE, Yu MK, Zhang W, Kreisberg JF, Tamayo P, et al. Systematic evaluation of molecular networks for discovery of disease genes. Cell systems. 2018; 6(4):484–495. https://doi.org/10.1016/j.cels.2018.03.001 PMID: 29605183

53. Pratt D, Chen J, Welker D, Rivas R, Pillich R, Rynkov V, et al. NDEx, the network data exchange. Cell systems. 2015; 1(4):302–305. https://doi.org/10.1016/j.cels.2015.10.001 PMID: 26594663

54. Pillich RT, Chen J, Rynkov V, Welker D, Pratt D. NDEx: a community resource for sharing and publishing of biological networks. In: Protein Bioinformatics. Springer; 2017. p. 271–301.

55. Pratt D, Chen J, Pillich R, Rynkov V, Gary A, Demchak B, et al. NDEx 2.0: a clearinghouse for research on cancer pathways. Cancer research. 2017; 77(21):e58–e61. https://doi.org/10.1158/0008-5472.CAN-17-0606 PMID: 29092941

56. Auer F, Kramer F, Ishkin A, Pratt D. ndexr: NDEx R client library; 2019. Available from: https://github.com/frankkramer-lab/ndexr.

57. Barrett T, Wilhite S, Ledoux P, Evangelista C, Kim I, Tomashevsky M, et al. NCBI GEO: archive for functional genomics data sets—update. Nucleic Acids Res. 2013; 41(Database issue). PMID: 23193258

58. Milano A, Pendergrass S, Sargent J, George L, McCalmont T, Connolly M, et al. Molecular subsets in the gene expression signatures of scleroderma skin. PLoS One. 2008; 3(7):e2696. https://doi.org/10.1371/journal.pone.0002696 PMID: 18648520

59. Hinchcliff M, Huang C, Wood T, Matthew M, Martyanov V, Bhattacharyya S, et al. Molecular signatures in skin associated with clinical improvement during mycophenolate treatment in systemic sclerosis. J Invest Dermatol. 2013; 133(8):1979–89. https://doi.org/10.1038/jid.2013.130 PMID: 23677167

60. Sheffer M, Bacolod M, Zuk O, Giardina S, Pincas H, Barany F, et al. Association of survival and disease progression with chromosomal instability: a genomic exploration of colorectal cancer. Proc Natl Acad Sci U S A. 2009; 106(17):7131–6. https://doi.org/10.1073/pnas.0902232106 PMID: 19359472

61. Cordero D, Sole X, Crous-Bou M, Sanz-Pamplona R, Pare-Brunet L, Guino E, et al. Large differences in global transcriptional regulatory programs of normal and tumor colon cells. BMC Cancer. 2014; p. 14–708.

62. Ryan B, Zanetti K, Robles A, Schetter A, Goodman J, Hayes R, et al. Germline variation in NCF4, an innate immunity gene, is associated with an increased risk of colorectal cancer. Int J Cancer. 2014; 134(6):1399–407. https://doi.org/10.1002/ijc.28457 PMID: 23982929

63. Tamaresis J, Irwin J, Goldfien G, Rabban J, Burney R, Nezhat C, et al. Molecular classification of endometriosis and disease stage using high-dimensional genomic data. Endocrinology. 2014; 155 (12):4986–99. https://doi.org/10.1210/en.2014-1490 PMID: 25243856

64. Yang Y, Zhang W, Gao H, Zhang Q. Gene expression and alternative splicing in human gastric cancer. Gene Expression Omnibus. 2009;GSE13195.

65. Wang Q, Wen Y, Li D, Xia J, Zhou C, Yan D, et al. Upregulated INHBA expression is associated with poor survival in gastric cancer. Med Oncol. 2012; 29(1):77–83. https://doi.org/10.1007/s12032-010-9766-y PMID: 21132402

66. Cui J, Li F, Wang G, Fang X, Puett J, Xu Y. Gene-expression signatures can distinguish gastric cancer grades and stages. PLos One. 2011; 6(3):e17819. https://doi.org/10.1371/journal.pone.0017819 PMID: 21445269

67. Goh S, Choi I, Kim N. Comparison of exon-wise expression profiling between normal and cancer tissues of human stomach. Gene Expression Omnibus. 2014;GSE30727.

68. Zhang X, Ni Z, Duan Z, Xin Zea. Overexpression of E2F mRNAs associated with gastric cancer progression identified by the transcription factor and miRNA co-regulatory network analysis. PLoS One. 2015; 10(2):e0116979. https://doi.org/10.1371/journal.pone.0116979 PMID: 25646628

69. Shao Q, Yao H, He J, Jin Y. Expression data from gastric cancer and paired normal tissues. Gene Expression Omnibus. 2016;GSE79973.

70. Lim H, Sohn I, Deng S, Lee J, Jung S, Mao M, et al. Prediction of disease-free survival in hepatocellular carcinoma by gene expression profiling. Ann Surg Oncol. 2013; 20(12):3747–53. https://doi.org/10.1245/s10434-013-3070-y PMID: 23800896

71. Hou J, Aerts J, den Hamer B, van Ijcken W, den Bakker M, Riegman P, et al. Gene expression-based classification of non-small cell lung carcinomas and survival prediction. PLos One. 2010; 5(4):e10312. https://doi.org/10.1371/journal.pone.0010312 PMID: 20421987

72. Kabbout M, Garcia M, Fujimoto J, Liu D, Woods D, Chow C, et al. ETS2 mediated tumor suppressive function and MET oncogene inhibition in human non-small cell lung cancer. Clin Canc Res. 2013; 19 (13):3383–95. https://doi.org/10.1158/1078-0432.CCR-13-0341

73. Chen C, Mendez E, Houck J, Fan W, Lohavanichbutr P, Doody D, et al. Gene expression profiling identifies genes predictive of oral squamous cell carcinoma. Cancer Epidemiol Biomarkers Prev. 2008; 17 (8):2152–62. https://doi.org/10.1158/1055-9965.EPI-07-2893 PMID: 18669583

74. Swindell W, Johnston A, Carbajal S, Han G, Wohn C, Lu J, et al. Genome-wide expression profiling of five mouse models identifies similarities and differences with human psoriasis. PLoS One. 2011; 6(4): e18266. https://doi.org/10.1371/journal.pone.0018266 PMID: 21483750

75. Suarez-Farinas M, Li K, Fuentes-Duculan J, Hayden K, Brodmerkel C, Krueger J. Expanding the psoriasis disease profile: interrogation of the skin and serum of patients with moderate-to-severe psoriasis. J Invest Dermatol. 2012; 132(11):2552–64. https://doi.org/10.1038/jid.2012.184 PMID: 22763790

76. Bigler J, Rand H, Kerkof K, Timour M, Russell C. Cross-study homogeneity of psoriasis gene expression in skin across a large expression range. PLoS One. 2013; 8(1):e52242. https://doi.org/10.1371/journal.pone.0052242 PMID: 23308107

77. Crow M, Lim N, Ballouz S, Pavlidis P, Gillis J. Predictability of human differential gene expression. Proceedings of the National Academy of Sciences. 2019; 116(13):6491–6500. https://doi.org/10.1073/pnas.1802973116

78. Lowes MA, Suarez-Farinas M, Krueger JG. Immunology of psoriasis. Annual review of immunology. 2014; 32:227–255. https://doi.org/10.1146/annurev-immunol-032713-120225 PMID: 24655295

79. Yildirim MA, Goh KI, Cusick ME, Barabasi AL, Vidal M. Drug-target network. Nature biotechnology. 2007; 25(10):1119–1126. https://doi.org/10.1038/nbt1338 PMID: 17921997

80. Hopkins AL. Network pharmacology: the next paradigm in drug discovery. Nature chemical biology. 2008; 4(11):682. https://doi.org/10.1038/nchembio.118 PMID: 18936753

81. Cheng F, Liu C, Jiang J, Lu W, Li W, Liu G, et al. Prediction of drug-target interactions and drug repositioning via network-based inference. PLoS computational biology. 2012; 8(5):e1002503. https://doi.org/10.1371/journal.pcbi.1002503 PMID: 22589709

82. Creixell P, Reimand J, Haider S, Wu G, Shibata T, Vazquez M, et al. Pathway and network analysis of cancer genomes. Nature methods. 2015; 12(7):615. https://doi.org/10.1038/nmeth.3440 PMID: 26125594

83. Parikshak NN, Gandal MJ, Geschwind DH. Systems biology and gene networks in neurodevelopmental and neurodegenerative disorders. Nature Reviews Genetics. 2015; 16(8):441. https://doi.org/10.1038/nrg3934 PMID: 26149713

84. Hu JX, Thomas CE, Brunak S. Network biology concepts in complex disease comorbidities. Nature Reviews Genetics. 2016; 17(10):615. https://doi.org/10.1038/nrg.2016.87 PMID: 27498692

85. Dancik V, Seiler KP, Young DW, Schreiber SL, Clemons PA. Distinct biological network properties between the targets of natural products and disease genes. Journal of the American Chemical Society. 2010; 132(27):9259–9261. https://doi.org/10.1021/ja102798t PMID: 20565092

86. Guney E, Menche J, Vidal M, Barábasi AL. Network-based in silico drug efficacy screening. Nature communications. 2016; 7:10331. https://doi.org/10.1038/ncomms10331 PMID: 26831545

87. Kim SS, Dai C, Hormozdiari F, van de Geijn B, Gazal S, Park Y, et al. Genes with high network connectivity are enriched for disease heritability. The American Journal of Human Genetics. 2019; 104(5):896–913. https://doi.org/10.1016/j.ajhg.2019.03.020 PMID: 31051114

88. Karczewski KJ, Snyder MP. Integrative omics for health and disease. Nature Reviews Genetics. 2018; 19(5):299. https://doi.org/10.1038/nrg.2018.4 PMID: 29479082

89. Leng N, Dawson JA, Thomson JA, Ruotti V, Rissman AI, Smits BM, et al. EBSeq: an empirical Bayes hierarchical model for inference in RNA-seq experiments. Bioinformatics. 2013; 29(8):1035–1043. https://doi.org/10.1093/bioinformatics/btt087 PMID: 23428641

90. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome biology. 2014; 15(12):550. https://doi.org/10.1186/s13059-014-0550-8 PMID: 25516281

91. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic acids research. 2015; 43(7):e47–e47. https://doi.org/10.1093/nar/gkv007 PMID: 25605792

92. Stretch C, Khan S, Asgarian N, Eisner R, Vaisipour S, Damaraju S, et al. Effects of sample size on differential gene expression, rank order and prediction accuracy of a gene signature. PloS one. 2013; 8 (6):e65380. https://doi.org/10.1371/journal.pone.0065380 PMID: 23755224

93. Schurch NJ, Schofield P, Gierliński M, Cole C, Sherstnev A, Singh V, et al. How many biological replicates are needed in an RNA-seq experiment and which differential expression tool should you use? Rna. 2016; 22(6):839–851. https://doi.org/10.1261/rna.053959.115 PMID: 27022035

94. Ghandi M, Huang FW, Jané-Valbuena J, Kryukov GV, Lo CC, McDonald ER, et al. Next-generation characterization of the Cancer Cell Line Encyclopedia. Nature. 2019; p. 1.

95. Wang S, Huang E, Cairns J, Peng J, Wang L, Sinha S. Identification of pathways associated with chemosensitivity through network embedding. PLoS computational biology. 2019; 15(3):e1006864. https://doi.org/10.1371/journal.pcbi.1006864 PMID: 30893303