Annotation Enrichment Analysis: An Alternative Method for Evaluating the Functional Properties of Gene Sets

Kimberly Glass [1–3,*], and Michelle Girvan [3–5]
[1] *Department of Biostatistics and Computational Biology, Dana-Farber Cancer Institute, Boston, MA, USA*
[2] *Department of Biostatistics, Harvard School of Public Health, Boston, MA, USA*
[3] *Department of Physics, University of Maryland, College Park, MD, USA*
[4] *Institute for Physical Science and Technology, University of Maryland, College Park, MD, USA*
[5] *Santa Fe Institute, Santa Fe, NM*

Gene annotation databases (compendiums maintained by the scientific community that describe the biological functions performed by individual genes) are commonly used to evaluate the functional properties of experimentally derived gene sets. Overlap statistics, such as Fishers Exact test (FET), are often employed to assess these associations, but don't account for non-uniformity in the number of genes annotated to individual functions or the number of functions associated with individual genes. We find FET is strongly biased toward over-estimating overlap significance if a gene set has an unusually high number of annotations. To correct for these biases, we develop Annotation Enrichment Analysis (AEA), which properly accounts for the non-uniformity of annotations. We show that AEA is able to identify biologically meaningful functional enrichments that are obscured by numerous false-positive enrichment scores in FET, and we therefore suggest it be used to more accurately assess the biological properties of gene sets.

1. INTRODUCTION

Evaluating the functional properties of gene sets is a routine step in understanding high-throughput biological data [1, 2] and is commonly used both to verify that the genes implicated in a biological experiment are functionally relevant [1] and to discover unexpected shared functions between those genes [3, 4]. One of the most widely used databases for functional annotations is the Gene Ontology (GO) [5, 6]. This database is highly regarded both for its comprehensiveness and its unified approach for annotating genes in different species to the same basic set of underlying functions [5]. In order to evaluate the strength of connection between an experimentally derived gene set and the set of genes that are annotated to a given biological function in this database, most functional enrichment analysis tools rely on set-overlap statistics [7]. Because these approaches are subject to an increase in type I errors associated with multiple hypothesis testing [], corrections such as the Benjamini, Bonferroni, and FDR are often also applied [8] (discussed in more detail below).

Young et. al. recently pointed out that these standard statistical approaches are sometimes inadequate, specifically when evaluating the functional properties of gene-sets derived from RNA-seq data since these experiments are prone to selection-bias due to variability in gene length [9]. Despite the wide use of functional analysis tools, however, little attention has been paid to whether or not the underlying properties of the functional databases themselves may contribute to spurious statistical results. For example, functional terms in GO are related through a directed acyclic graph (DAG) whose structure contributes to the heavy-tailed distribution seen in the number of gene annotations to individual terms [10]. In this work we investigate whether such annotation properties lead to a bias in the results of traditional functional analysis techniques. We find that traditional methods spuriously identify significant enrichment between function terms and *random* gene sets, if the number of annotations made to genes in the gene set is high.

We also investigate the properties of experimentally-derived gene signatures, i.e. sets of genes whose combined expressed patterns are associated with specific biological conditions. We study signatures from the Gene Signatures Database [11] and find that most contain a disproportionate number of highly annotated genes. Furthermore, traditional overlap statistics report significant associations between these signatures and randomly constructed annotation sets. Consequently, we propose a scheme, called Annotation Enrichment Analysis (AEA), that evaluates the overlap in *annotations* between a set of genes and the set of terms belonging to a branch of the GO hierarchy, using a randomization protocol to build a null model. By looking at annotation overlap instead of gene overlap, our approach takes into account the annotation properties of the Gene Ontology. It effectively eliminates biases due to database construction and highlights relevant biological functions in experimentally-defined gene signatures. We also provide a simple analytic approximation to AEA (which we call AEA-A, for

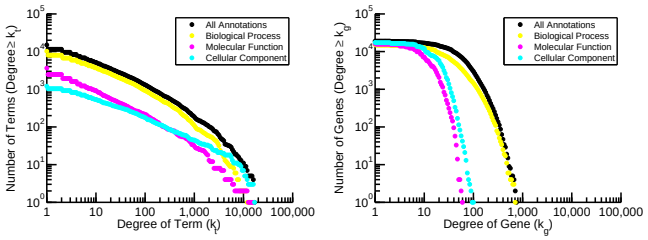
*contact: kglass@jimmy.harvard.edu

Annotation Enrichment Analysis Approximation) that is able to partially compensate for the biases we find using traditional approaches. Implementations of both approaches are provided at http://www.networks.umd.edu.

There are many functional annotation databases that have been developed in order to classify genes according their various roles in the cell [12–16]. Here, we focus our analysis on functional annotations made to the Gene Ontology because of its wide use by many functional enrichment tools (for example [1, 2, 17–19]). It has been observed that the annotation properties of the Gene Ontology are often shared by other databases [10]. For example, several categorize biological information in a hierarchical fashion, including the Keywords published by the Universal Protein Resource consortium [20] and the *E.coli* multi-functional classification scheme [16]. Furthermore many classification databases also appear to exhibit a heavy-tailed distribution in the number of genes annotated to individual categories (see Figure 1, discussed in more detail below, and Supplemental Figure **??**). Therefore, we believe that the methods we develop here, although tested using the Gene Ontology, could be applied to functional enrichment analysis using other classification databases.

The Gene Ontology [5] takes the form of a directed acyclic graph (DAG) in which "child" functional categories ("terms") are subclassified under one or more other, more general categories, called "parent" terms. "Branches" in the Gene Ontology can therefore be defined as sets of terms that contain a parent term and all of its progeny. Note that these branches will contain overlapping sets of terms since each term can be a descendant of multiple ancestors at each level of the DAG. Using this structure, individual genes are annotated to various functional categories. These annotations are transitive up the hierarchy such that a parent term will take on all the genes annotations associated with any of its progeny [21]. Consequently, terms with many progeny often contain many gene annotations whereas terms with few progeny generally have fewer associated genes. "Biological Process," "Molecular Function," and "Cellular Component" are the three most general terms in GO, defining three independent branches such that every other term can only belong to one of these three categories. As a consequence all genes in GO are annotated to at least one, and often all three, of these categories.

Since we want to determine the influence of annotation database properties on functional enrichment analysis, especially in the context of experimental gene signatures in commonly studied diseases such as cancer, we focus our study on GO annotations that are associated with human genes. With this in mind, we downloaded information regarding gene-term annotations for human genes from the Gene Ontology website (geneontology.org) and used this data to construct a gene-term bipartite graph, represented as an $n_G \times n_T$ adjacency matrix, where n_G is the total number of genes and n_T is the total number of terms listed in the annotation file. In this matrix a value



(a)Term Degree Distribution (b)Gene Degree Distribution

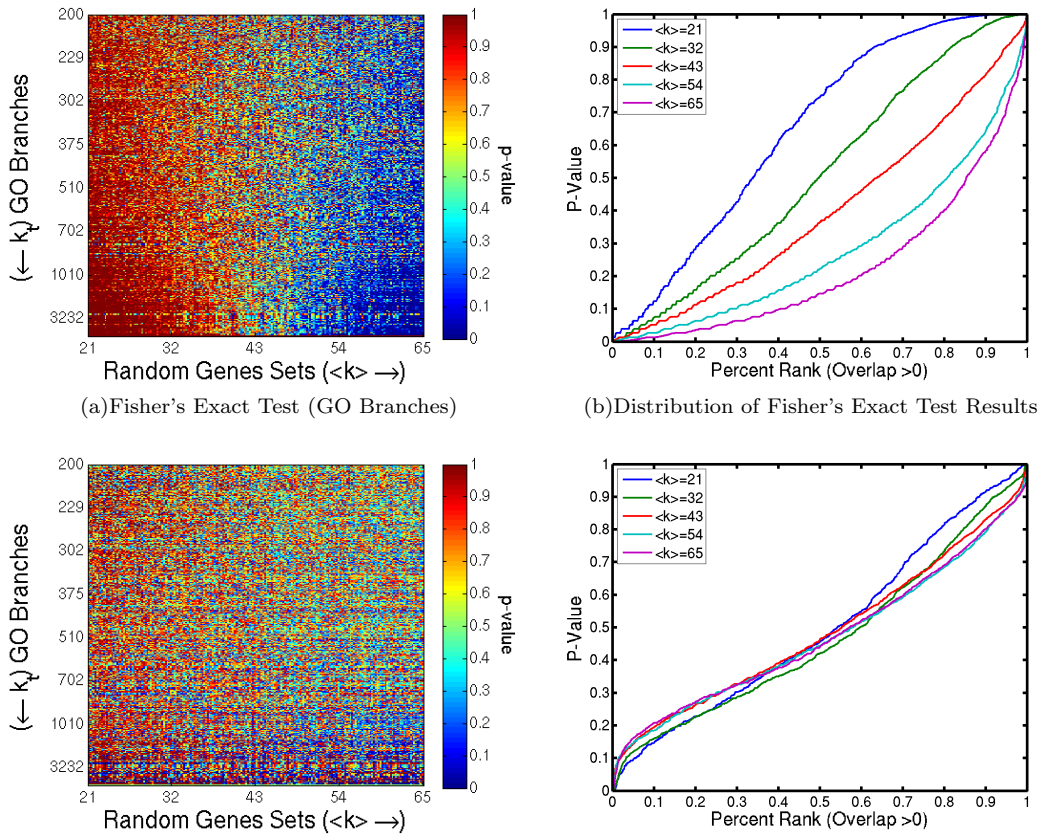
FIG. 1: The cumulative degree distributions of (a) terms and (b) genes in human GO annotations. "Biological Process" terms make up the majority of terms and annotations. The average number of "Biological Process" terms to which an individual gene is annotated is 43.2 while the average number of annotations made to an individual term is 64.4.

of one indicates a known connection between the corresponding gene and term, and a value of zero indicates that the gene is not associated with that term.

In this bipartite graph many terms are only associated with a small handful of genes, while some terms are associated with many genes. A histogram of the "degree", k_t, of terms (the number of genes annotated to individual terms) reveals a heavy-tailed relationship (Figure 1(a)). In contrast, a histogram of the "degree", k_g, of genes (the number of terms to which individual genes are annotated) shows that although some genes have many more annotations than others, the distribution is not as heavy-tailed as the term degree distribution. (Figure 1(b)).

The "Biological Process" ontology contains a significant fraction of the total annotations. Although all three ontologies are used in functional enrichment analysis, many studies using GO focus on this ontology, both for its size and because its members describe dynamical processes performed by the cell. We do the same in the following analysis. The total number of annotations made to the "Biological Process" ontology is 656783, originating from 15213 genes to 10192 terms. Consequently, the average number of annotations made by an individual gene to this ontology is 43.2 and the average number of annotations made to an individual term is 64.4. These values will be useful to keep in mind, especially as we investigate the annotation properties of gene signatures and of the terms for which they are enriched.

The most widely used statistics for evaluating which functional categories are enriched in a set of genes are based on gene counts and include Fisher's Exact Test, the binomial test, and the chi-squared test [7]. Although these statistics vary in exact implementation, they all rely on the same basic underlying assumption that all genes have an equal probability of being selected under the null hypothesis. Of these tests, Fisher's Exact Test (FET) is the most common statistic and is used by many of the most popular functional enrichment tools (see Table 2 in [8]), and therefore we choose it to represent a "typical" evaluation of gene set functional enrichment. Although it is recognized that this statistic makes as-



(a)Fisher's Exact Test (GO Branches)

(b)Distribution of Fisher's Exact Test Results

(c)Annotation Enrichment Analysis (GO Branches)

(d)Distribution of Annotation Enrichment Analysis Results

FIG. 2: (a) The enrichment (measured by p-value) of 200 randomly generated gene sets in annotations made to GO branches. The branches are ordered based on how many genes are annotated to the parent term (k_t) and the gene sets are ordered based on total the number of annotations (M_g) made by the 200 genes in that set. Although we tested enrichment for all branches, for simplicity we only visualize subset of branches with 200 or more unique gene annotations. There is an obvious bias toward significant enrichment between high degree gene-set/term pairs using Fisher's Exact Test (FET). (b) A plot of the p-values predicted by FET as a function of rank for five of the random gene sets shows that FET is both overly conservative for low degree gene sets and anti-conservative for high degree gene sets. (c)-(d) the same plots illustrating the results using AEA. The bias can be correctly mitigated by using Annotation Enrichment Analysis (AEA) (c,d).

sumptions in its null hypothesis that fail to reflect the complex properties of the Gene Ontology, it is widely regarded as a good guide in determining which functions are represented by a given set of genes.

Since most functional enrichment analysis compares a gene set to all the terms in GO, multiple-hypothesis testing corrections are often applied to these p-values [8]. These corrections decrease the value at which a comparison between a gene set and a GO term should be considered significant. Commonly used multiple-hypothesis corrections include the Bonferroni, Benjamini and the False Discovery Rate. Of these, the Bonferroni is the most conservative and adjusts the value at which a test is considered "significant" by the number of tests made []. The False Discovery Rate (FDR) adjusts the value at which a test is considered "significant" based on the rank of the predicted level of significance []. It provides approximately the same correction as the Bonferroni for the most significantly-ranked p-values but will not adjust tests that are the least-significant. It is important to note that although these corrections will change the critical value of individual tests, they do not affect the rank ordering of these tests.

2. RESULTS

2.1. Annotation Properties Influence the Results of Functional Enrichment Analysis

One of our goals is to determine the effect of annotation database properties on functional enrichment analysis. To do this, we first created random gene sets with N_g members each, but in which we controlled the total number of annotations (M_g) made by the genes belonging to each gene set, such that the average degree of the genes

in the set ($k_{avg} = M_g/N_g$) varies from approximately 21 to 65, or from around half to 1.5 times the expected average degree of 43 (see Introduction and Figure 1). In practice experimentally derived sets of genes can range from only a handful (≈ 10) to a few thousand genes. In this analysis we chose $N_g = 200$ since this represents a "typical" gene set size.

As an initial test, we used our random gene sets to evaluate how annotation bias might effect the enrichment significance predicted by traditional methods to asses functional enrichment. To do this, we used FET to determine the enrichment of our randomly constructed gene sets in the 10192 GO terms from the "Biological Process" ontology. Figure 2(a) shows the results for the subset of those terms that have 200 or more unique gene annotations, ordered based on their total number of unique gene annotations. The trend is striking. Even though they have the same number of members, gene sets with a higher number of *annotations* are more enriched in GO terms compared to gene sets with a lower number of annotations. For high degree gene sets, we see that high degree GO branches tend to be more significantly enriched (i.e., have lower p-values) than low degree branches. However, note that the trend is reversed for low degree gene sets. For each random gene set, since we tested for significant overlap for all 10192 "Biological Process" terms, the expected minimum p-value across all the terms should be approximately 10^{-4}. However, instead we observe that random gene sets with the fewest annotations have a minimum p-value around 10^{-3}, while random gene sets with the highest annotation levels have a minimum p-value across all tests close to 10^{-6} (Supplemental Figure S2(a)). We point out that although multiple-hypothesis corrections will sufficiently raise a p-value such that either very few or no false positives occur using these random gene sets, the biases themselves cannot be overcome in this manner (see Supplemental Figure S2(b)).

In order to better interpret these results, for five of our random gene sets ($\langle k \rangle \approx \{21, 32, 43, 54, 65\}$), we directly compared the distributions of the p-values predicted by FET to the expected distribution (evenly distributed values from zero to one). Figure 2(b) plots, in rank order, the p-values calculated for these random gene sets for the set of terms that contain at least one gene annotation from a member of the given random gene set. The deep dip below the diagonal for the the more highly-annotated gene sets demonstrates that FET is anti-conservative for these gene sets; in addition, it also appears to be overly conservative for gene sets with a lower overall annotation level. Plots using all terms are shown in Supplemental Figure S3(a)).

2.2. Annotation Enrichment Analysis Corrects for Annotation Bias

Clearly annotation properties of both gene sets and functional categories can influence the results of functional enrichment analysis. In order to mitigate these effects, we suggest that instead of evaluating the overlap between genes, as is traditionally done in functional enrichment analysis, one instead considers the overlap between *annotations* made to a gene set and a branch of terms in the Gene Ontology. To accurately capture the significance of annotation overlap we develop a randomization scheme that preserves the transitive annotation features of the GO DAG while calculating the probability of obtaining a certain number of annotations between a gene set and a GO branch. We call this approach Annotation Enrichment Analysis (AEA) and illustrate it in Figure 3.

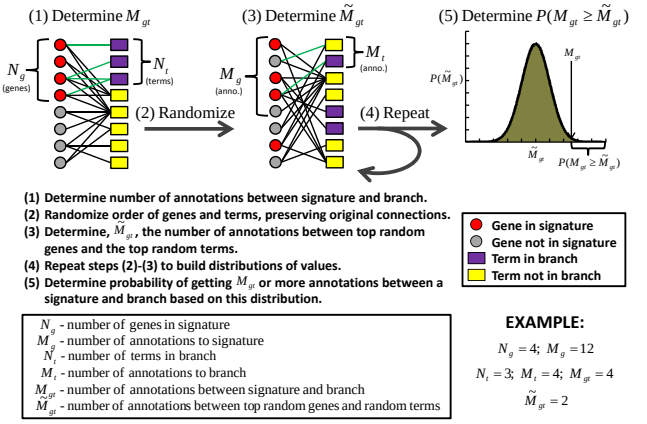
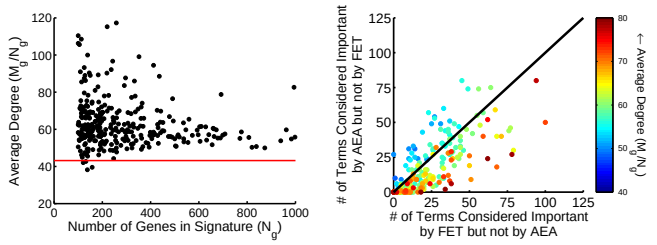


FIG. 3: An outline of how Annotation Enrichment Analysis (AEA) calculates the significance of association between a given gene signature and the collection of terms that belong to a branch in the GO hierarchy.

In this randomization is it useful to think of the Gene Ontology as a bipartite graph (see Introduction). We begin by determining M_g, the total number of annotations to a gene set, M_t, the total number of annotations to the terms in a GO branch, and M_{gt}, the number of annotations stretching between this gene set and branch. We then determine a distribution for the expected number of co-annotations. To do this we, simultaneously, randomly permute the order of genes and terms while still preserving the original connections from the GO bipartite graph. By preserving the original connections, we retain the transitive annotation properties of the GO DAG. We then take annotations connected to the top randomly shuffled genes until we've selected M_g annotations, and annotations connected to the top randomly shuffled terms until we've selected M_t annotations, and determine $\tilde{M}_{gt}$, the number of edges in the bipartite graph that extend between these top randomly shuffled genes and top randomly shuffled terms. In the (fairly common) case where selecting the top M_g/M_t annotations does not correspond to selecting a whole number of genes/terms, we take the top number of genes/terms whose total annotations is closest to M_g/M_t, respectively. We repeat the randomization process many times in order to determine a distribution of values for $\tilde{M}_{gt}$. We define a new p-value,



(a)Signature Properties (b)FET vs. AEA on Signatures

FIG. 4: Annotation properties of experimental gene signatures. (a) The number of genes versus the average number of annotations made to the genes in each signature. Genes from signatures generally contain many more GO annotations than one would expect if selecting genes randomly (red line). (b) The number of terms that are considered important (top 10% by rank) by one of the measures (either AEA or FET), but not important (bottom 80% by rank) by the other, plotted for each gene signature. The signatures are colored according to the average level of annotation ($k_{avg} = M_g/N_g$).

$p_A(M_{gt})$ which reflects the probability that $\tilde{M}_{gt} \geq M_{gt}$:

$$p_A(M_{gt}) = P(\tilde{M}_{gt} \geq M_{gt}). \quad (1)$$

We determined the significance of all GO branches in our randomly generated gene sets with AEA (using 10^4 randomizations), and created a heat map of these values analogous to the one produced using standard set-overlap statistics (Figure 2(c)). The results of AEA are close to uniform across varying gene set degree (Figure 2(d)), demonstrating that AEA works well at eliminating annotation bias.

2.3. Experimental Gene Signatures are Often Highly-Annotated

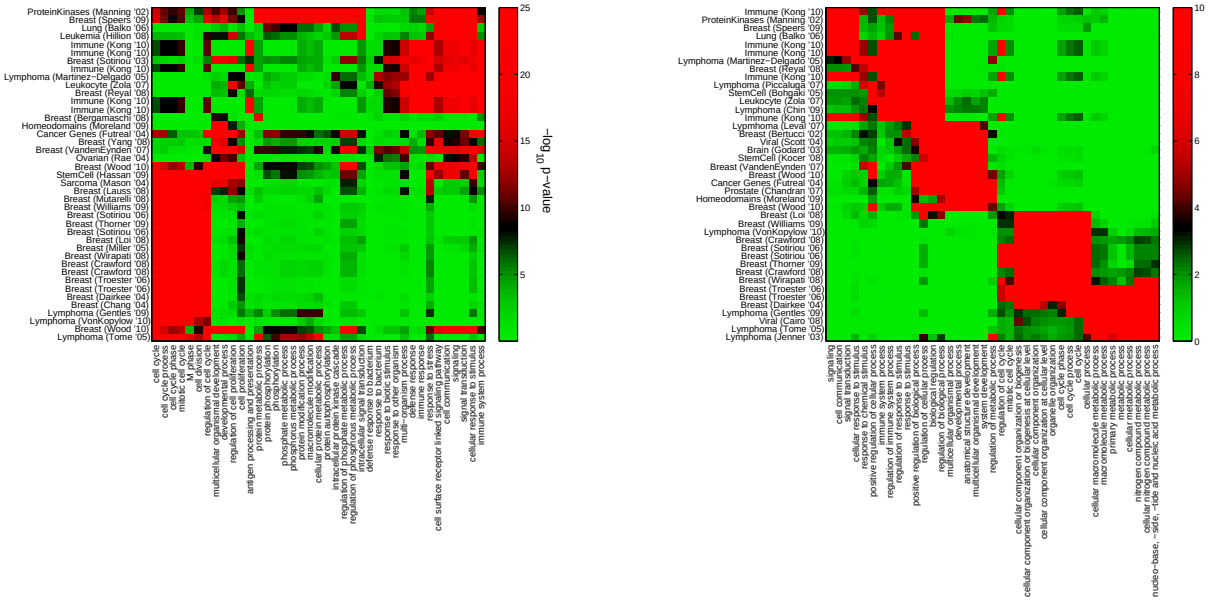
One of the most common applications of enrichment analysis is to ascertain the functional properties of a gene "signature" (an experimentally determined set of genes). Although we have demonstrated that AEA corrects for annotation bias with randomly generated gene sets, we also want to know how well this analysis can recapitulate biologically-relevant results. With this in mind, we downloaded signatures as recorded in the Gene Signatures Database (GeneSigDB) [11]. This database is a manual curation of previously published gene expression signatures, focusing primarily on cancer and stem cell signatures [22]. In the following analysis we will use all 309 human signatures from this database that contain at least 100 and less than 1000 genes that also are annotated to a term in the "Biological Process" ontology.

First, to assess whether annotation bias might play a role in evaluating the functional properties of these gene signatures, we determined the average number of annotations made to the genes occurring in each signature. Figure 4(a) shows the number of genes in a signature plotted against the average level of annotation for each signature. The expectation for a random selection of genes (the average number of annotations made to all genes – see Introduction) is shown as a red line. The plot suggests that many genes belonging to these signatures are also more highly annotated in GO. Almost a third (99) of the signatures have an average level of annotation that is greater than any of our randomly generated gene sets and all but four signatures have an average level of annotation greater than expected by chance. Since we have shown that random gene signatures with these annotation levels encounter a bias in traditional functional enrichment analysis, we believe these experimental signatures are an appropriate biological set with which to evaluate how AEA compares to FET when investigating and discovering the functions of genes derived from experimental biological data.

It is common practice when evaluating the functional properties of a gene signature to focus on a set of "top" categories based on p-value rank. We investigate how different the results of AEA or FET might appear in this context. To this end, we selected the top 10% of terms based on their enrichment score in FET and AEA to designate as "important" according each to these measures. We compared this list of terms to the list of terms that are "not important" (in the bottom 80% of terms by rank) according to each measure. The number of terms considered important in AEA but not by FET versus the number of terms considered important by FET but not AEA for each signature is plotted in Figure 4(b). Complete agreement between FET and AEA on this plot is represented by a point at $(0, 0)$, and complete disagreement is represented by a point at $(1019, 1019)$. In order to see how annotation properties influenced any differences, we colored signatures based on the average level of annotation to their member genes.

Figure 4(b) shows overall agreement between AEA and FET, as many points fall fairly close to the origin and, at most, reflect only a 10% difference in identified "important" terms. However, annotation bias is evident. In signatures containing the highest levels of annotation, i.e. those represented by reddish marks, the terms deemed most "important" by FET are more likely to be considered "unimportant" according to AEA, and vice versus. These results are consistent with the previous analysis in random gene sets that showed a bias by FET to place more significance between gene sets and terms with a higher number of annotations (see Figure 2). It also demonstrates that annotation bias is present when evaluating experimentally-derived gene signatures and it not an artifact of how we constructed our random gene sets.

In the supplement we also directly compare FET and AEA p-values and observe that, in these experimental signatures, a high annotation level is correlated with increased significance of FET compared to AEA and vice versus (Supplemental Figure S3(c)). These results are consistent with the results shown in Figure 4(b).



(a)Fisher's Exact Test

(b)Annotation Enrichment Analysis

FIG. 5: Clustergrams representing enriched term-signature pairs. (a) A clustering of signatures and terms selected based on their enrichment-score according to FET. These signatures include those reported in [23–55]. (b) A clustering of signatures and terms selected based on their enrichment-score according to AEA. The signatures and terms break into several, biologically distinct units. One is associated with immune-response, and includes signatures published in [23–25, 27, 29–31, 56–58]. A second includes signatures related to cellular-differentiation published in [33, 34, 36, 38, 59–64]. Another cluster includes breast cancer signatures published in [43–46, 48–51, 54]. Finally three lymphoma [53, 55, 65] and a viral signature [66] associated with proliferation are also included. The colorscales for the p-values were chosen to give approximately the same red/green balance in each clustergram.

2.4. Annotation Enrichment Analysis Uncovers Meaningful Biological Associations

Next, we investigated the specific biology that is highlighted using AEA and FET. For each measure, we chose approximately forty signatures having the most significant enrichment scores across all terms. Similarly, for each measure, we chose forty terms having the most significant enrichment scores across all signatures. Using these sets of terms and signatures and the p-values associating all pairs in the set, we then performed a standard hierarchical clustering analysis (using the "clustergram" function in Matlab with default settings) on the terms and signatures selected for each measure. The results are shown in Figure 5. For AEA a small number (981 out of 3149328 possible) term-signature pairs have an estimated p-value of $p < 10^{-6}$ after one million randomizations, therefore, when necessary, we broke ties by the number of signatures/terms enriched in the terms/signatures at this level.

Clustering the FET results gives rise to a weak visual segregation of terms and signatures into groups (Figure 5(a)). These groups highlight the relationship between the gene signatures and several important biological processes. For example, the FET clustering shows an enrichment of cell-cycle related processes in breast cancer signatures [67] and includes immune-related terms enriched in immune gene signatures. These two groups, however, account for only about half of the selected terms; the clustergram also includes a number of functional categories related to "proteins" and "phosphorylation" that are only enriched in a small number of signatures. From this analysis we suggest that the results of FET might be muddled by a signal driven by annotation bias, highlighting either highly-annotated signatures or more general biological processes.

In contrast, using AEA, distinct clusters of signatures and terms emerge (Figure 5(b)). The first includes signatures from immune-systems, lymphoma and leucocytes, and is logically also enriched in terms such as "immune system" and "response to stimulus" as well as terms related to "biological regulation". Interestingly, one of the breast signatures associated with this cluster [31] represents a list of genes defined based on immune response in breast cancer and the stem cell signature [57] is from a study on patients with systemic sclerosis, a type of autoimmune disorder. In addition, the inclusion of a protein-kinase signature [23] is interesting as MAP kinases have been shown to play an important role in immune response [68].

Another cluster is enriched in categories such as "system development" and "developmental process" and includes several signatures associated with stem cells or identified based on their role in cellular differentiation. It also includes a signature of oncogenes [34], as well as a signature of homeodomain proteins, known to initiate cascades of genes that in turn will induce cellular differentiation into tissues and organs (e.g. [69, 70]). The next cluster, associated primarily with breast cancer signatures, shows a strong enrichment for terms related to the cell cycle and cellular component organization, processes known to be differentially regulated in breast cancer [67]. Finally, two lymphoma and one viral signature that were identified based on cell proliferation (for example, by association with Myc targeting [53, 66]) are enriched for terms such as "cellular metabolic process." This is consistent with expectation since there is evidence that a connection exists between proliferation and metabolic pathways in cancer cells [71, 72].

2.5. Some Predictions Made by FET are Likely a Consequence of Annotation Bias in Experimental Gene Signatures

Next, we also created random term sets by randomly selecting GO terms, constructing each individual random term set such that it has approximately the same number of unique genes annotated to its members as a real GO branch (for more details see Methods). We used these random term sets to study the application of functional enrichment analysis methods to experimental gene signatures and to systematically determine if annotation database properties might be a source of false positives. Specifically, using both the traditional FET and our proposed AEA, we investigated the enrichment of experimental gene signatures in randomly constructed term sets as well as real GO branches. We determined the number of term-signature comparisons considered significant at several different thresholds and present the results in Figure 6.

Surprisingly, using FET, there is almost no difference between the number of significant comparisons made using real GO branches and using the randomly generated term sets. This striking similarity can be understood as follows. When calculating the significance between two gene sets, FET assumes all genes in those sets have an equal probability of being chosen. This is a false assumption as some genes are actually more likely to be annotated to any given term in GO. Just as high degree genes are more likely to be annotated to a randomly chosen GO branch, so too are they more likely to be annotated to a randomly set of GO terms. As noted previously, experimental gene signatures include an abundance of genes with higher levels of annotations. Combined together, this bias means that these genes are likely to be enriched in random sets of functional categories, just because their members have more annotations overall. We believe this illustrates a fundamental flaw of using FET in this manner, as it will predict significant functional associations, not because of biological signal, but as a result of a bias in signature annotation properties.

Compared to FET, AEA finds fewer enriched pairs at each threshold, but, unlike FET, finds no signatures enriched in the random term sets, demonstrating its ability to correct for annotation biases introduced from the hierarchical relationships between those terms in the ontology. These results give us confidence that AEA is highlighting the connections between gene sets and branches that are most likely to be truly biologically relevant and is robust against biases introduced by annotation properties.

For more analysis comparing the effects of term set properties on FET and AEA see the supplementary material.

2.6. A Quantitative Approximation to Annotation Enrichment Analysis Partially Corrects for Annotation Bias

One significant strength of AEA is that it makes no assumptions regarding the structure of gene-term annotations; however, because it uses a randomization scheme to estimate the null hypothesis, the precision of the estimated p-values is dependent upon the number of randomizations, and each run of the algorithm will give slightly different results. Therefore, we sought an analytic approximation of AEA in order to overcome these limitations.

Given that we want to estimate the significance of annotation overlap, one logical approach is to simply count the number of annotations made to a gene set, the number of annotations made to a branch in GO, and the number of annotations extending between that gene set and branch, and use the hypergeometric probability to determine the significance of this overlap. We point out that this approach makes the false assumption that annotations are independent, implying that a gene could be annotated to the same term multiple times. Another more limiting problem is that this approach, unlike AEA, erases the hierarchical organization of annotations encoded in the GO DAG. Because of these assumptions, the approximation predictions do not have the reliability of the randomization protocol specified by AEA; however, it can be computed quickly and without the need for any randomization to generate a null hypothesis.

Acknowledging that we are making some false assumptions regarding the structure of gene-term annotations, we propose an analytic framework for evaluating functional enrichment which we call the Annotation Enrichment Analysis Approximation (AEA-A). This approximation makes use of the hypergeometric probability to calculate the significance (or p-value approximating AEA, $p_a(M_{gt})$) of overlap between annotations made to a given gene set and branch in the GO hierarchy. Given

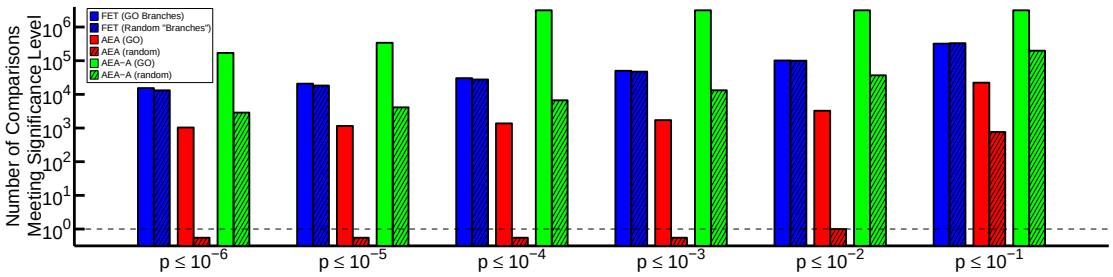


FIG. 6: A plot of the number of term-signature comparisons deemed "significant" at various p-value thresholds. A dotted line indicates only one comparison with a p-value less than or equal to the indicated threshold, and cases where no significant comparisons were found for the corresponding p-value are indicated by a bar not exceeding this line. Evaluations using gene annotations to GO branches are shown as solid colors, whereas evaluations using genes annotated to random term sets are striped. AEA-A refers to the results when using a quantitative approximation to AEA.

M_g annotations to a gene set, M_t annotations to terms belonging to a GO branch, and M_{tot} annotations made in the GO ontology, the probability of finding M_{gt} or more annotations in common between these two sets can be written as:

$$\begin{aligned} p_a(M_{gt}) &= P(M \geq M_{gt} | M_g, M_t, M_{tot}) \\ &= \sum_{i=M_{gt}}^{min[M_g,M_t]} \frac{\binom{M_t}{i}\binom{M_{tot}-M_t}{M_g-i}}{\binom{M_{tot}}{M_g}}. \end{aligned} \quad (2)$$

We tested the performance of this approximation by determining the functional enrichment of GO terms in our randomly generated gene sets. The results of AEA-A are uniform across varying gene set degree (see Supplemental Figure S5), demonstrating that AEA-A works well at eliminating annotation bias. However, the predicted p-values are often misleadingly low due to the independence assumption. This limitation is evident in analysis performed on the experimental signatures (Figure 6) – many more comparisons are deemed "significant" at each threshold using AEA-A than either AEA or FET. Furthermore, compared to AEA, the approximation is only partially able to discern between real GO branches and random term sets. However, we note that it does significantly outperform FET in this regard. Therefore, we conclude that although this analytic approximation is conceptually appealing and has some advantages over FET, it does not provide results that are as discerning as AEA.

3. DISCUSSION

We have demonstrated that evaluating the functional enrichment of gene sets using traditional set-overlap statistics, such as FET, is susceptible to producing false positives as a result of certain annotation database properties. We offer a solution, Annotation Enrichment Analysis, or AEA, that fully considers these properties, eliminating potential annotation bias in the predicted enrichment scores. The importance of using this approach is highlighted by the fact that many published gene-signatures include a large number of highly-annotated genes. This is likely in part due to a non-independence between identified signatures and functional annotations, since genes that are involved in a well-studied phenomena such as cancer are also more likely to be frequently annotated in these databases. Although it is possible that newly-derived gene signatures may not exhibit the same level of annotation-bias as these previously-published signatures, it is also very probable that highly annotated genes are important in a wide variety of well-studied systems and will continue to show up and influence the results of functional enrichment analysis on newly generated gene sets.

The annotation-bias associated with FET results and the bias for higher annotation-levels among experimentally-derived gene signatures is largely unrecognized. Although significant p-values for functional enrichment in experimental signatures may initially seem compelling for the bioinformatician, we suggest that these results do not always reflect biological properties but instead have a high potential to be a result of statistical bias. In light of our analysis we suggest using the AEA approach either alongside or in place of other traditional measures, especially for gene signatures that are known to contain significantly more or less annotations than one would expect by chance. Furthermore, we urge the bioinformatics community to consider annotation properties of gene signatures and annotation databases before utilizing results from the wide variety of available gene set enrichment tools. We believe that considering annotation enrichment will allow biologists to better interpret the functional roles of genes identified as important in their experimental system.

4. METHODS

4.1. Calculating Functional Enrichment using Set-Overlap Statistic

In this analysis we use Fisher's Exact Test (FET) to evaluate biases in the performance of "traditional" functional enrichment analysis. FET is related to the hypergeometric probability and can be used to calculate the significance, or p-value estimated using FET ($p_F(N_{gt})$) of an overlap between two independent sets. For example, given a gene set containing N_g genes, a GO term with annotations to k_t different genes, and N_{tot} total genes annotated in GO, the probability that N_{gt} or more genes belong both to this gene set and are annotated to the GO term can be calculated as:

$$\begin{aligned} p_F(N_{gt}) &= P(N \geq N_{gt}|N_g, k_t, N_{tot}) \\ &= \sum_{i=N_{gt}}^{min[N_g,k_t]} \frac{\binom{k_t}{i}\binom{N_{tot}-k_t}{N_g-i}}{\binom{N_{tot}}{N_g}}. \end{aligned} \quad (3)$$

Together with the FET, we also sometimes determine the False Discovery Rate (FDR) of the tests in order to account for any potential Type I error ([73, 74]). We calculate the FDR using the matlab function "mafdr" and report the associated q-values.

4.2. Constructing Biased Random Gene Sets and Random Term Sets

In order to investigate potential bias due to the annotation properties of gene sets, we constructed random gene sets with the same number of members, but with varying amounts of annotations made by those members. Each set with a desired total number of annotations, M_g, was created by first randomly selecting N_g genes. We then randomly selected one gene in this gene set (gene i) and one gene not in the gene set (gene j). If replacing gene i with gene j caused the total number of annotations made by genes in the gene set to approach M_g, we replaced gene i with gene j with a high probability ($p = 0.95$), but if the replacement caused the average degree of the gene set to move farther away from M_g we replaced gene i with gene j with a low probability ($p = 0.05$). This swapping continued until the total number of annotations made by the gene set was within 0.1% of M_g. In this way we created 200 gene sets with $N_g = 200$ genes each, but whose average degree ($k_{avg} = M_g/N_g$) varies from approximately 21 to 65.

We also constructed sets of random GO terms to investigate the how the significance levels determined between experimental gene signatures differed for these random sets as compared to real GO branches. Specifically, for comparison with a branch in the GO DAG that has a parent term with k_t gene annotations, we randomly ordered all the terms in GO and selected the top N_t terms until the number of unique genes annotated to those N_t random terms (k'_t) is within a small percentage of k_t ($|k_t - k'_t|/k_t < 0.01$). In the case where selecting both N_t and N_{t+1} terms were within this limit we chose N_t to minimize the absolute difference between k_t and k'_t. If selecting the top N_t terms did not lead to a situation within this limit, we reshuffled the terms and selected the top N_t terms in this new list, repeating until a suitable random collection of terms could be chosen. In this way we created 10192 random term sets with approximately the same *number* of unique genes annotated to each as to real GO branches.

Acknowledgements

We would like to thank Emanuele Mazzola for helpful discussions regarding this work.

[1] D. W. Huang, B. T. Sherman, Q. Tan, J. Kir, D. Liu, D. Bryant, Y. Guo, R. Stephens, M. W. Baseler, H. C. Lane, et al., Nucl. Acids Res. **35**, gkm415+ (2007), ISSN 1362-4962, URL http://dx.doi.org/10.1093/nar/gkm415.

[2] A. Subramanian, P. Tamayo, V. K. Mootha, S. Mukherjee, B. L. Ebert, M. A. Gillette, A. Paulovich, S. L. Pomeroy, T. R. Golub, E. S. Lander, et al., Proceedings of the National Academy of Sciences of the United States of America **102**, 15545 (2005), ISSN 0027-8424, URL http://dx.doi.org/10.1073/pnas.0506580102.

[3] O. D. King, R. E. Foulger, S. S. Dwight, J. V. White, and F. P. Roth, Genome research **13**, 896 (2003), ISSN 1088-9051, URL http://dx.doi.org/10.1101/gr.440803.

[4] S. Mostafavi and Q. Morris, Bioinformatics (Oxford, England) **26**, 1759 (2010), ISSN 1367-4811, URL http://dx.doi.org/10.1093/bioinformatics/btq262.

[5] M. Ashburner, C. A. Ball, J. A. Blake, D. Botstein, H. Butler, J. M. Cherry, A. P. Davis, K. Dolinski, S. S. Dwight, J. T. Eppig, et al., Nature genetics **25**, 25 (2000), ISSN 1061-4036, URL http://dx.doi.org/10.1038/75556.

[6] T. G. O. Consortium, Nucleic Acids Research **38**, D331 (2010), ISSN 1362-4962, URL http://dx.doi.org/10.1093/nar/gkp1018.

[7] I. Rivals, L. Personnaz, L. Taing, and M.-C. Potier, Bioinformatics **23**, 401 (2007), ISSN 1460-2059, URL http://dx.doi.org/10.1093/bioinformatics/btl633.

[8] P. Khatri and S. Drăghici, Bioinformatics **21**, 3587 (2005), ISSN 1460-2059, URL http://dx.doi.org/10.1093/bioinformatics/bti565.

[9] M. D. Young, M. J. Wakefield, G. K. Smyth, and A. Oshlack, Genome biology **11**, R14+ (2010), ISSN 1465-6914, URL http://dx.doi.org/10.1186/gb-2010-11-2-r14.

[10] K. Glass, E. Ott, W. Losert, and M. Girvan, Journal of the Royal Society, Interface / the Royal Society (2012), ISSN 1742-5662, URL http://dx.doi.org/10.1098/rsif.2011.0585.

[11] A. C. Culhane, M. S. Schröder, R. Sultana, S. C. Picard, E. N. Martinelli, C. Kelly, B. Haibe-Kains, M. Kapushesky, A.-A. St Pierre, W. Flahive, et al., Nucleic Acids Research **40**, D1060 (2012), ISSN 1362-4962, URL http://dx.doi.org/10.1093/nar/gkr901.

[12] R. L. Tatusov, N. D. Fedorova, J. D. Jackson, A. R. Jacobs, B. Kiryutin, E. V. Koonin, D. M. Krylov, R. Mazumder, S. L. Mekhedov, A. N. Nikolskaya, et al., BMC bioinformatics **4**, 41+ (2003), ISSN 1471-2105, URL http://dx.doi.org/10.1186/1471-2105-4-41.

[13] M. Kanehisa and S. Goto, Nucleic Acids Research **28**, 27 (2000), ISSN 0305-1048, URL http://dx.doi.org/10.1093/nar/28.1.27.

[14] M. H. Serres, S. Goswami, and M. Riley, Nucleic Acids Research **32**, D300 (2004), URL http://view.ncbi.nlm.nih.gov/pubmed/11471834.

[15] R. Caspi, T. Altman, J. M. Dale, K. Dreher, C. A. Fulcher, F. Gilham, P. Kaipa, A. S. Karthikeyan, A. Kothari, M. Krummenacker, et al., Nucleic acids research **38**, D473 (2010), ISSN 1362-4962, URL http://dx.doi.org/10.1093/nar/gkp875.

[16] M. H. Serres and M. Riley, Microb Comp Genomics **5**, 205 (2000), ISSN 1090-6592, URL http://view.ncbi.nlm.nih.gov/pubmed/11471834.

[17] T. Beissbarth and T. P. Speed, Bioinformatics (Oxford, England) **20**, 1464 (2004), ISSN 1367-4803, URL http://dx.doi.org/10.1093/bioinformatics/bth088.

[18] D. Martin, C. Brun, E. Remy, P. Mouren, D. Thieffry, and B. Jacq, Genome Biol **5** (2004), ISSN 1465-6914, URL http://dx.doi.org/10.1186/gb-2004-5-12-r101.

[19] A. Alexa, J. Rahnenführer, and T. Lengauer, Bioinformatics **22**, 1600 (2006), ISSN 1367-4803, URL http://dx.doi.org/10.1093/bioinformatics/btl140.

[20] E. Boutet, D. Lieberherr, M. Tognolli, M. Schneider, and A. Bairoch, Methods in molecular biology (Clifton, N.J.) **406**, 89 (2007), ISSN 1064-3745, URL http://view.ncbi.nlm.nih.gov/pubmed/18287689.

[21] The_gene_ontology_consortium, Genome Res. **11**, 1425 (2001), ISSN 1088-9051, URL http://dx.doi.org/10.1101/gr.180801.

[22] A. C. Culhane, T. Schwarzl, R. Sultana, K. C. Picard, S. C. Picard, T. H. Lu, K. R. Franklin, S. J. French, G. Papenhausen, M. Correll, et al., Nucleic acids research **38**, D716 (2010), ISSN 1362-4962, URL http://dx.doi.org/10.1093/nar/gkp1015.

[23] G. Manning, D. B. Whyte, R. Martinez, T. Hunter, and S. Sudarsanam, Science **298**, 1912 (2002), ISSN 1095-9203, URL http://dx.doi.org/10.1126/science.1075762.

[24] C. Speers, A. Tsimelzon, K. Sexton, A. M. Herrick, C. Gutierrez, A. Culhane, J. Quackenbush, S. Hilsenbeck, J. Chang, and P. Brown, Clinical Cancer Research **15**, 6327 (2009), ISSN 1557-3265, URL http://dx.doi.org/10.1158/1078-0432.CCR-09-1107.

[25] J. M. Balko, A. Potti, C. Saunders, A. Stromberg, E. B. Haura, and E. P. Black, BMC Genomics **7**, 289+ (2006), ISSN 1471-2164, URL http://dx.doi.org/10.1186/1471-2164-7-289.

[26] J. Hillion, S. Dhara, T. F. F. Sumter, M. Mukherjee, F. Di Cello, A. Belton, J. Turkson, S. Jaganathan, L. Cheng, Z. Ye, et al., Cancer research **68**, 10121 (2008), ISSN 1538-7445, URL http://dx.doi.org/10.1158/0008-5472.CAN-08-2121.

[27] Y. Megan Kong, C. Dahlke, Q. Xiang, Y. Qian, D. Karp, and R. H. Scheuermann, Journal of biomedical informatics (2010), ISSN 1532-0480, URL http://dx.doi.org/10.1016/j.jbi.2010.05.001.

[28] C. Sotiriou, S.-Y. Y. Neo, L. M. McShane, E. L. Korn, P. M. Long, A. Jazaeri, P. Martiat, S. B. Fox, A. L. Harris, and E. T. Liu, Proceedings of the National Academy of Sciences of the United States of America **100**, 10393 (2003), ISSN 0027-8424, URL http://dx.doi.org/10.1073/pnas.1732912100.

[29] B. Martínez-Delgado, M. Cuadros, E. Honrado, A. Ruiz de la Parte, G. Roncador, J. Alves, J. M. Castrillo, C. Rivas, and J. Benítez, Leukemia **19**, 2254 (2005), ISSN 0887-6924, URL http://dx.doi.org/10.1038/sj.leu.2403960.

[30] H. Zola, B. Swart, A. Banham, S. Barry, A. Beare, A. Bensussan, L. Boumsell, C. D Buckley, H.-J. J. Bühring, G. Clark, et al., Journal of immunological methods **319**, 1 (2007), ISSN 0022-1759, URL http://dx.doi.org/10.1016/j.jim.2006.11.001.

[31] F. Reyal, M. H. van Vliet, N. J. Armstrong, H. M. Horlings, K. E. de Visser, M. Kok, A. E. Teschendorff, S. Mook, L. van 't Veer, C. Caldas, et al., Breast cancer research : BCR **10**, R93+ (2008), ISSN 1465-542X, URL http://dx.doi.org/10.1186/bcr2192.

[32] A. Bergamaschi, E. Tagliabue, T. Sørlie, B. Naume, T. Triulzi, R. Orlandi, H. G. Russnes, J. M. Nesland, R. Tammi, P. Auvinen, et al., The Journal of pathology **214**, 357 (2008), ISSN 0022-3417, URL http://dx.doi.org/10.1002/path.2278.

[33] R. T. Moreland, J. F. Ryan, C. Pan, and A. D. Baxevanis, Database : the journal of biological databases and curation **2009** (2009), ISSN 1758-0463, URL http://dx.doi.org/10.1093/database/bap004.

[34] P. A. Futreal, L. Coin, M. Marshall, T. Down, T. Hubbard, R. Wooster, N. Rahman, and M. R. Stratton, Nature reviews. Cancer **4**, 177 (2004), ISSN 1474-175X, URL http://dx.doi.org/10.1038/nrc1299.

[35] S. X. Yang, S. M. Steinberg, D. Nguyen, T. D. Wu, Z. Modrusan, and S. M. Swain, Clinical cancer research : an official journal of the American Association for Cancer Research **14**, 5893 (2008), ISSN 1078-0432, URL http://dx.doi.org/10.1158/1078-0432.CCR-07-4762.

[36] G. G. Van den Eynden, S. J. Van Laere, I. Van der Auwera, L. Gilles, J. L. Burn, C. Colpaert, P. van Dam, E. A. Van Marck, L. Y. Dirix, and P. B. Vermeulen, Clinical & experimental metastasis **24**, 13 (2007), ISSN 0262-0898, URL http://dx.doi.org/10.1007/s10585-006-9049-3.

[37] M. T. Rae, D. Niven, A. Ross, T. Forster, R. Lathe, H. O. Critchley, P. Ghazal, and S. G. Hillier, The Journal of endocrinology **183**, 19 (2004), ISSN 0022-0795, URL http://dx.doi.org/10.1677/joe.1.05754.

[38] C. E. Wood, J. R. Kaplan, M. B. Fontenot, J. K. Williams, and J. M. Cline, Clinical cancer research : an official journal of the American Association for Cancer Research **16**, 946 (2010), ISSN 1078-0432, URL http://dx.doi.org/10.1158/1078-0432.CCR-09-1541.

[39] K. A. Hassan, G. Chen, G. P. Kalemkerian, M. S. Wicha, and D. G. Beer, Clinical Cancer Research **15**, 6386

(2009), URL http://dx.doi.org/10.1158/1078-0432.CCR-09-1105.

[40] D. X. Mason, T. J. Jackson, and A. W. Lin, Oncogene **23**, 9238 (2004), ISSN 0950-9232, URL http://dx.doi.org/10.1038/sj.onc.1208172.

[41] M. Lauss, A. Kriegner, K. Vierlinger, I. Visne, A. Yildiz, E. Dilaveroglu, and C. Noehammer, Breast cancer research and treatment **110**, 235 (2008), ISSN 1573-7217, URL http://dx.doi.org/10.1007/s10549-007-9716-3.

[42] M. Mutarelli, L. Cicatiello, L. Ferraro, O. M. M. Grober, M. Ravo, A. M. Facchiano, C. Angelini, and A. Weisz, BMC bioinformatics **9 Suppl 2**, S12+ (2008), ISSN 1471-2105, URL http://dx.doi.org/10.1186/1471-2105-9-S2-S12.

[43] C. M. Williams, A. G. Scibetta, J. K. Friedrich, M. Canosa, C. Berlato, C. H. Moss, and H. C. Hurst, The EMBO journal **28**, 3591 (2009), ISSN 1460-2075, URL http://dx.doi.org/10.1038/emboj.2009.290.

[44] C. Sotiriou, P. Wirapati, S. Loi, A. Harris, S. Fox, J. Smeds, H. Nordgren, P. Farmer, V. Praz, B. Haibe-Kains, et al., Journal of the National Cancer Institute **98**, 262 (2006), ISSN 1460-2105, URL http://dx.doi.org/10.1093/jnci/djj052.

[45] A. R. Thorner, K. A. Hoadley, J. S. Parker, S. Winkel, R. C. Millikan, and C. M. Perou, Oncogene **28**, 742 (2009), ISSN 1476-5594, URL http://dx.doi.org/10.1038/onc.2008.430.

[46] S. Loi, B. Haibe-Kains, C. Desmedt, P. Wirapati, F. Lallemand, A. M. Tutt, C. Gillet, P. Ellis, K. Ryder, J. F. Reid, et al., BMC genomics **9**, 239+ (2008), ISSN 1471-2164, URL http://dx.doi.org/10.1186/1471-2164-9-239.

[47] L. D. Miller, J. Smeds, J. George, V. B. Vega, L. Vergara, A. Ploner, Y. Pawitan, P. Hall, S. Klaar, E. T. Liu, et al., Proceedings of the National Academy of Sciences of the United States of America **102**, 13550 (2005), ISSN 1091-6490, URL http://dx.doi.org/10.1073/pnas.0506230102.

[48] P. Wirapati, C. Sotiriou, S. Kunkel, P. Farmer, S. Pradervand, B. H. Kains, C. Desmedt, M. Ignatiadis, T. Sengstag, F. Schutz, et al., Breast Cancer Research **10**, R65+ (2008), ISSN 1465-5411, URL http://dx.doi.org/10.1186/bcr2124.

[49] N. P. Crawford, J. Alsarraj, L. Lukes, R. C. Walker, J. S. Officewala, H. H. Yang, M. P. Lee, K. Ozato, and K. W. Hunter, Proceedings of the National Academy of Sciences of the United States of America **105**, 6380 (2008), ISSN 1091-6490, URL http://dx.doi.org/10.1073/pnas.0710331105.

[50] M. A. Troester, J. I. Herschkowitz, D. S. Oh, X. He, K. A. Hoadley, C. S. Barbier, and C. M. Perou, BMC cancer **6**, 276+ (2006), ISSN 1471-2407, URL http://dx.doi.org/10.1186/1471-2407-6-276.

[51] S. H. Dairkee, Y. Ji, Y. Ben, D. H. Moore, Z. Meng, and S. S. Jeffrey, BMC genomics **5** (2004), ISSN 1471-2164, URL http://dx.doi.org/10.1186/1471-2164-5-47.

[52] H. Y. Chang, J. B. Sneddon, A. A. Alizadeh, R. Sood, R. B. West, K. Montgomery, J.-T. T. Chi, M. van de Rijn, D. Botstein, and P. O. Brown, PLoS biology **2**, e7+ (2004), ISSN 1545-7885, URL http://dx.doi.org/10.1371/journal.pbio.0020007.

[53] A. J. Gentles, A. A. Alizadeh, S.-I. I. Lee, J. H. Myklebust, C. M. Shachaf, B. Shahbaba, R. Levy, D. Koller, and S. K. Plevritis, Blood **114**, 3158 (2009), ISSN 1528-0020, URL http://dx.doi.org/10.1182/blood-2009-02-202465.

[54] K. von Kopylow, C. Kirchhoff, D. Jezek, W. Schulze, C. Feig, M. Primig, V. Steinkraus, and A.-N. N. Spiess, Human reproduction (Oxford, England) **25**, 1104 (2010), ISSN 1460-2350, URL http://dx.doi.org/10.1093/humrep/deq053.

[55] M. E. Tome, D. B. Johnson, L. M. Rimsza, R. A. Roberts, T. M. Grogan, T. P. Miller, L. W. Oberley, and M. M. Briehl, Blood **106**, 3594 (2005), ISSN 0006-4971, URL http://dx.doi.org/10.1182/blood-2005-02-0487.

[56] P. P. P. Piccaluga, C. Agostinelli, A. Califano, M. Rossi, K. Basso, S. Zupo, P. Went, U. Klein, P. L. L. Zinzani, M. Baccarani, et al., The Journal of clinical investigation **117**, 823 (2007), ISSN 0021-9738, URL http://dx.doi.org/10.1172/JCI26833.

[57] T. Bohgaki, Y. Amasaki, N. Nishimura, M. Bohgaki, Y. Yamashita, M. Nishio, K.-I. I. Sawada, S. Jodo, T. Atsumi, and T. Koike, Annals of the rheumatic diseases **64**, 1165 (2005), ISSN 0003-4967, URL http://dx.doi.org/10.1136/ard.2004.030338.

[58] M. Chin, M. Herscovitch, N. Zhang, D. J. Waxman, and T. D. Gilmore, Oncogene **28**, 2100 (2009), ISSN 1476-5594, URL http://dx.doi.org/10.1038/onc.2009.74.

[59] L. de Leval, D. S. Rickman, C. Thielen, A. d. Reynies, Y.-L. L. Huang, G. Delsol, L. Lamant, K. Leroy, J. Brière, T. Molina, et al., Blood **109**, 4952 (2007), ISSN 0006-4971, URL http://dx.doi.org/10.1182/blood-2006-10-055145.

[60] F. Bertucci, R. Houlgatte, S. Granjeaud, V. Nasser, B. Loriod, E. Beaudoing, P. Hingamp, J. Jacquemier, P. Viens, D. Birnbaum, et al., Annals of the New York Academy of Sciences **975**, 217 (2002), ISSN 0077-8923, URL http://view.ncbi.nlm.nih.gov/pubmed/12538167.

[61] L. A. Scott, J. K. Vass, E. K. Parkinson, D. A. Gillespie, J. N. Winnie, and B. W. Ozanne, Molecular and cellular biology **24**, 1540 (2004), ISSN 0270-7306, URL http://www.ncbi.nlm.nih.gov/pmc/articles/PMC344183/.

[62] S. Godard, G. Getz, M. Delorenzi, P. Farmer, H. Kobayashi, I. Desbaillets, M. Nozaki, A.-C. C. Diserens, M.-F. F. Hamou, P.-Y. Y. Dietrich, et al., Cancer research **63**, 6613 (2003), ISSN 0008-5472, URL http://view.ncbi.nlm.nih.gov/pubmed/14583454.

[63] S. S. Koçer, P. M. Djurić, M. F. Bugallo, S. R. Simon, and M. Matic, BMC genomics **9**, 359+ (2008), ISSN 1471-2164, URL http://dx.doi.org/10.1186/1471-2164-9-359.

[64] U. R. Chandran, C. Ma, R. Dhir, M. Bisceglia, M. Lyons-Weiler, W. Liang, G. Michalopoulos, M. Becich, and F. A. Monzon, BMC cancer **7**, 64+ (2007), ISSN 1471-2407, URL http://dx.doi.org/10.1186/1471-2407-7-64.

[65] R. G. Jenner, K. Maillard, N. Cattini, R. A. Weiss, C. Boshoff, R. Wooster, and P. Kellam, Proceedings of the National Academy of Sciences of the United States of America **100**, 10399 (2003), ISSN 0027-8424, URL http://dx.doi.org/10.1073/pnas.1630810100.

[66] S. Cairo, C. Armengol, A. De Reyniès, Y. Wei, E. Thomas, C.-A. A. Renard, A. Goga, A. Balakrishnan, M. Semeraro, L. Gresh, et al., Cancer cell **14**, 471 (2008), ISSN 1878-3686, URL http://dx.doi.org/10.1016/j.ccr.2008.11.002.

[67] M. Loddo, S. R. Kingsbury, M. Rashid, I. Proctor, C. Holt, J. Young, S. El-Sheikh, M. Falzon, K. L. Eward, T. Prevost, et al., British journal of cancer **100**, 959 (2009), ISSN 1532-1827, URL http://dx.doi.org/10.1038/sj.bjc.6604924.

[68] C. Dong, R. J. Davis, and R. A. Flavell, Annual review of immunology **20**, 55 (2002), ISSN 0732-0582, URL http://dx.doi.org/10.1146/annurev.immunol.20.091301.131133.

[69] A. Nepveu, Gene **270**, 1 (2001), ISSN 0378-1119, URL http://view.ncbi.nlm.nih.gov/pubmed/11403998.

[70] M. C. Magli, Biotherapy (Dordrecht, Netherlands) **10**, 279 (1998), ISSN 0921-299X, URL http://view.ncbi.nlm.nih.gov/pubmed/9592016.

[71] R. J. DeBerardinis, J. J. Lum, G. Hatzivassiliou, and C. B. Thompson, Cell Metabolism **7**, 11 (2008), ISSN 15504131, URL http://dx.doi.org/10.1016/j.cmet.2007.10.002.

[72] M. G. Vander Heiden, L. C. Cantley, and C. B. Thompson, Science **324**, 1029 (2009), ISSN 1095-9203, URL http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2849637/.

[73] J. D. Storey, Journal of the Royal Statistical Society: Series B (Statistical Methodology) **64**, 479 (2002), ISSN 13697412, URL http://dx.doi.org/10.1111/1467-9868.00346.

[74] J. D. Storey and R. Tibshirani, Proceedings of the National Academy of Sciences of the United States of America **100**, 9440 (2003), ISSN 0027-8424, URL http://dx.doi.org/10.1073/pnas.1530509100.