

Mining gene sets for measuring similarities

CHRISTINE NARDINI¹, DANIELE MASOTTI², SUNGROH YOON³, ENRICO MACII², MICHAEL D. KUO⁴, GIOVANNI DE MICHELI⁵, AND LUCA BENINI¹

¹University of Bologna, Italy, cnardini@deis.unibo.it, lbenini@deis.unibo.it

²Politecnico di Torino, Italy, daniele.masotti@polito.it, enrico.macii@polito.it

³Stanford University, CA, USA, sryoon@stanford.edu

⁴UCSD Medical Center HillCrest, CA, USA, mkuo@ucsd.edu

⁵EPFL, Switzerland, giovanni.demicheli@epfl.ch

Abstract

In recent years, the development of high throughput devices for the massive parallel analyses of genomic data has led to the generation of large amount of new biological evidences and has triggered the proliferation of data mining algorithms for the extraction of meaningful information. Microarrays for gene expression analyses are part of this revolution and provide important insight in molecular biology often in the form of coherent sets of genes representing previously uncharacterized processes. Large amount of data are continuously produced in this form, and computational approaches can significantly improve the efficient use of these results, since comparison among numbers of genes sets can give new meaningful information at no cost from the experimental biology point of view. To address this opportunity we designed and implemented FIT, a scalable, unsupervised algorithm that quantitatively compares different populations of gene sets using two distinct measures of similarity between any two gene sets. These measures are then used to obtain a summary statistic that describes the tightness of fit between sets belonging to two distinct populations of gene sets. We present the results of FIT on two data sets for the study of Lymphoma and Acute Lymphoblastic Leukemia. In both cases FIT was able to recapitulate the previous analyses on these datasets, to extend the results and to extract information likely to offer potential insights into the underlying biology.

1 Introduction

Since the decoding of the human genome, an important change in life science has occurred that involves the strong contribution of information technology to biological research. In fact, the availability of such an amount of information requires automated approaches for the efficient management and data mining of these new, large types of data. The proliferation of devices able to process in parallel thousands of experiments and almost capable to perform genome-wide screens is a consequence of this new paradigm. Expression microarrays are such one tool and allow the evaluation of the activation of number of genes under different environmental conditions. Data mining algorithms demanded to the analysis of expression data are used for both extracting and defining clusters of genes (sometimes called *signatures*) whose pattern of expression is concurrently shared by and that helps to characterize a particular class of samples and/or genes to identify for example different sub-populations within a disease and to predict survival and disease-free status (i.e.[2]). A complete review on classical and recent algorithms is presented in [12, 10], for algorithm specifically designed for the integration of clinical and expression information see also [17, 18]. What all these approaches share in common is the generation and definition of new gene sets which continually expand and populate our genomic knowledge base. Gene sets are defined as lists of genes IDs and through the process of *annotation* it is possible to characterize the global meaning of a gene set in terms of its role in some molecular process, or disease evolution and more. Against this background, when exploring new hypotheses it is often useful to interrogate this broad

population of gene sets. In fact, computational approaches can help the mining of new information at no experimental cost from a biological point of view. For example, the possibility to compare newly found or already characterized gene sets against an appropriate subset of annotated gene sets, can give meaningful insight about the new sets or enrich the characterization to mine new relevant information. Moreover, on the computational side, ongoing research in and application of data mining techniques to genomics often lead to comparisons of different algorithms. This comparison is necessary to measure the differences in computational efficiency or complexity of the algorithm, but also to check and confirm that the results found make biological sense. In such cases there is an obvious need for comparison with known reference sets. The necessity to have a standard approach to this problem was a strong practical motivation to the implementation of FIT to evaluate a measure of similarity between any two sets of genes from two different populations of gene sets, where a population is defined as a set of gene sets.

The main contribution of this work is the definition of a complete and flexible measure of the relationship among two populations of genes sets in terms of their similarity. This approach overcomes two main limitations of previously published techniques based on statistical similarity measures, namely using enrichment and its significance. In fact, previous approaches allowed the selection of only one population of gene sets (test population) to measure its similarity against the categories of Gene Ontology [16], a hierarchically structured vocabulary for the characterization of gene sets in terms of defined classes (reference population) [1, 14]). In contrast, we designed an approach to gene cluster characterization allowing the choice of both user-defined test and reference populations. This has two main advantages. In fact, it allows to approach two different problems related to gene cluster similarity evaluation. First, it is possible to compare the test population issued from a new data mining algorithm against a known benchmark, a common issue that is so far addressed by ad hoc approaches (i.e. [5, 8]). Second, it allows contrasting two populations of genes sets, related by some common biology and often issued from different studies, to inquire about possible commonality, likely to give more insight into the biological problem under study. This evaluation performed *in silico* can save part of costly and time consuming biological experiments. Moreover, the fact that both populations are specified by the user often requires a more subtle evaluation of the measure of similarity. For this reason we also improved the score used so far (enrichment) adding information about the specificity of the enrichment. In fact, in case of ties in the value of enrichment, the specificity and its significance are a unique property of the considered set. Fit is freely available at: <http://www-micrel.deis.unibo.it/~fit/>.

2 Related Work

To date, the automated characterization of genes sets produced by microarray experiments has often been performed using the definitions stored in the controlled hierarchical vocabulary curated by the Gene Ontology Consortium. Gene Ontology (GO) is as a structured vocabulary used to classify molecules, cellular products and activities of genes from three main points of view (Molecular Function, Biological Process, Cellular Component) called ontologies. However, a large number of genes is still not characterized, due to the lack of information about their role. Thus, the main application of the GO vocabulary consists in labeling the genes in a set of interest with the appropriate GO term, when available, (*annotation* process) and to extend the annotation to uncharacterized genes, provided that the number of genes in the set related to the specific GO term is bigger than the number expected by chance (*enrichment* analysis). Because of its free availability, its plasticity and its continuous growth, GO has become the gold standard for automated gene annotation, triggering interest around the notions of similarity on which it is based, and motivating the design of several tools to automate the process of relating newly found clusters to it (only as a sample [1, 14]). However, besides the undeniable importance of having such a structured and controlled classifier, some limitations arise, notably related to the necessity of a manual curation. For this we elaborated this new approach to annotate genes in a more flexible way. In fact, very often the design of a microarray experiment aims at studying in deeper detail the mechanisms peculiar to a disease or environment stress. For this reason, gene clusters are not only annotated with GO terms, but also labeled in terms of the effects of a disease or of environment stress. This labeling is sometimes transversal to GO and carries information of diverse origin.

Finally, it is a common and necessary issue to compare the results of new data mining algorithms to known benchmarks. There is no standard method to approach this typically computational problem, conversely to what happens in biologically oriented research. Because of this lack, various approaches have been implemented. Cheng et al. ([5]) designed a well known biclustering algorithm able to cluster genes and samples (generically called items or conditions) in parallel. To validate their approach on microarray experiments they defined two scores. The first score (*mix*) was meant to measure the affinity of their bicluster to a benchmark made of only 2 clusters called primary clusters. It measures the percentage of items in the bicluster that belong to the less represented primary cluster. The second score (*row variance*) was defined to help the selection of the most interesting bicluster, discarding trivial, non-informative biclusters made of items (genes or samples) whose value was

constant through the conditions (samples or genes). It gives a measure of the variance of an item through some conditions. A combination of low mix score and high row variance is likely to select biologically interesting biclusters. A limitation of this approach is that it involves the contemporary observation of two measures and the interpretation of all possible combination of values. Moreover, it is limited to the benchmark chosen, that consists of only two clusters. Finally, especially for the gene biclusters validation, this approach does not actually compare to clusters with known biological meaning, since primary clusters are poorly characterized. Thus, this implies a subsequent manual curation of the most interesting biclusters, difficult to be performed and biologically validated with information technology facilities alone. Getz et al. ([8]) also defined a biclustering algorithm: the Coupled Two Way Clustering (CTWC). In their validation approach they gave more attention to the biological meaning of their results, however this consisted of a careful but time consuming manual curation.

With our approach, the number of clusters in the benchmark is not a limit, and it is straightforward to compare to biologically meaningful clusters experimentally validated. Besides the more informative validation obtained, this also simplifies the difficult and time consuming validation performed on a cluster by cluster basis.

3 Computing Similarity

When comparing two populations of clusters there is generally a clear distinction in the quality of the two populations in terms of their semantic characterization. In fact, the comparison is often performed between a population whose sets' meaning is known, here called *reference*, and one whose sets' meaning is unknown, here called *test* population. The attempt is to describe the test sets in terms of the reference sets.

Enrichment. To obtain a measure of the similarity between the information contained in a test set with respect to that contained in a reference set we calculated the enrichment score (η) ([19]). With this approach, we assume that the reference sets are the base of knowledge for this comparison and we represent every set as a distribution of genes, where the categories are the reference sets. The enrichment score evaluates the proportion of the relative frequency observed in test set i for category j ($f_j^{test_i}$) with respect to the relative frequency expected in the reference population (i.e. the relative size of reference set j , $|ref_j|$) and is calculated as follows: $\eta_{i,j} = \frac{f_j^{observed}}{f_j^{expected}} = \frac{f_j^{test_i}/|test_i|}{|ref_j|/|RefPop|}$ where $|set_z|$ represents the size of set z and $|RefPop|$ the size of the ref population. To evaluate the significance of the enrichment we used the hypergeometric distribution [13], to calculate the probability α of being wrong when assum-

ing that the enrichment is not due to chance alone ([3, 4]). Namely, we evaluated the probability that a specific amount of items (r genes in the test set) of a given type (n_1 genes in the reference set) can be selected without replacement in a sample (of n genes in the test set) from a broader population (of N genes of test set population), where N is the size of the background population. The complement to 1 of the cumulated probability of all the events that are less likely than having r items belonging to the sample, gives the statistical significance of the enrichment.

$$p_{i,j}^\eta = 1 - \sum_{i=1}^{r-1} \frac{\binom{n_1}{i} \cdot \binom{N-n_1}{n-i}}{\binom{N}{n}}$$

Specificity of the Enrichment. In order to express as correctly as possible the test sets with the *vocabulary* represented by the reference population, the use of the enrichment alone was not informative enough. For this reason we designed an extra measure for the evaluation of the similarity, here called the *specificity of the enrichment* to evaluate how exclusive the enrichment of the test set i for reference set j is, with respect to the other reference sets (for an example see Figure 1). To this aim we compare the profiles of the enrichments through all the reference categories of both the test and the reference set. The more similar the profiles are, the more tight is the similarity. We used Kendall's coefficient of correlation τ [13], a non parametric measure for the measure of association between variables based on rankings. Assuming the null hypothesis in our test is $H_0 : \tau(test_i, ref_j) = 0$, we also defined the significance of the specificity p^τ . $\tau_{i,j} = \tau(test_i, ref_j)$, $p_{i,j}^\tau = \alpha$, for $H_0 : \tau_{i,j} = 0$

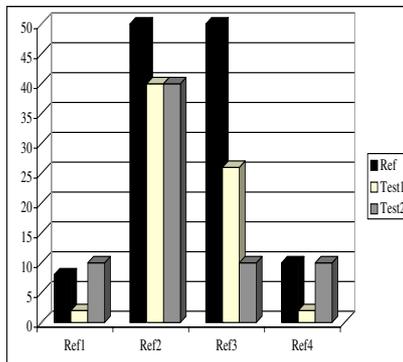
Fit. Finally, we designed a global measure to synthesize the information previously obtained. Namely, to quantify the refinement of the enrichment with the supplementary information of specificity, we multiplied the values of η and τ , and defined the measure of fitness ϕ : $\phi_{i,j} = \eta_{i,j} \times \tau_{i,j}$. For the evaluation of ϕ 's significance, we use Fisher's method for meta-analysis [9], that allows to combine the results of several tests, merging their significance values computing Fisher's statistic $S = -2 \sum_{i=1}^{\#tests} \ln(p_i)$. In this case $\#tests = 2$, $p_1 = p^\eta$, $p_2 = p^\tau$. When the tests are independent and the scores have a uniform distribution in $[0, 1]$, S has a $\chi^2_{(\alpha, 2 \cdot \#tests)}$ distribution, where α summarizes the results of the two tests and $2 \cdot \#tests$ are the degrees of freedom. $p_{i,j}^\phi = \chi^2(-2(\ln p_{i,j}^\eta + \ln p_{i,j}^\tau), 4)$. A summary of the scores' characteristics is given in Table 1.

4 Algorithm Implementation

The tool core is developed under R, a statistical analysis package, compatible with S (and Splus), freely available on the web under the Free Software Foundation's GNU general public license. The user interface is developed using php. Users requests are processed in the following pipeline:

Name	Enr.	Enr.Sign.	Spec.	Spec.Sign.	Fit	Fit Sign.
Symbol	η	p^η	τ	p^τ	ϕ	p^ϕ
Origin	f_{obs} / f_{exp}	Hyp.Dist.	Kend.Corr.	Sign.Kend.Corr.	Product	Fisher's met.
Range	$[0, +\infty)$	$[0, 1]$	$[-1, 1]$	$[0, 1]$	$[0, +\infty)$	$[0, 1]$

Table 1. Summarized characteristics and mathematical function or formula allowing the calculation and range of variability of the scores calculated in FIT. η gives a relative measure of the increased presence of items of a specific reference type in the studied test set. τ allows to discriminate test sets with equal enrichment for a specific reference category, but different distribution of the genes among other reference categories. Finally, ϕ synthesizes the results of the previously defined scores. p^η , p^τ and p^ϕ respectively give the statistical significance of each score.



(a) Sets Distributions

VS Ref2	$\eta(p_\eta)$	$\tau(p_\tau)$	$\phi(p_\phi)$
Test1	2.96 (0.00)	0.77(1.14 e-01)	2.29(0.00)
Test2	2.96 (0.00)	-0.26(5.99 e-01)	-0.76(0.00)

(b) Similarity Scores

Figure 1. Reference and Test population (Figure 1(a)) are build in such a way that the enrichment of the two test sets for Ref2 appears to be identical. However, since the distributions are not the same, the evaluation of the *specificity* allows to discriminate between the two test sets and identify the one that is best suited to represent the information contained in Ref2. Namely, Test2 has a distribution that is uniform across all the other reference sets, and thus Test1 is more suited to describe Ref2, as the final values of ϕ and p_ϕ show in Table 1(b).

they are first stored in a database (MySQL), where a batch scheduled task retrieves and processes them using various R scripts, while the user interface is waiting. Then, once the results are stored, the web site can access and display them on the user browser. Each request has an assigned code, which allows the users to retrieve their results later. Input of

the algorithms are plain text files, where each column represents a gene set. All scores and their significance are stored in tables where every row represents a test set, every column a reference set and every cell represents the score. Output is represented by a text and graphical version of these output tables. Graphs represent score tables where cells are filled in darker or lighter color proportional to the score value. Rows and columns are then clustered to obtain the so called *heat-maps* where sets of genes that are more similar are also spatially closer. We designed this graphical output with in the specific intent to recall the one of TreeView (freely downloadable at: <http://rana.lbl.gov/EisenSoftware.htm>), a familiar tool for microarray users, designed to visualize the results of the data mining performed on microarray data using hierarchical clustering. This is a cluster analysis algorithm, based on the iterative pairwise agglomeration of items. Agglomeration is defined based on a similarity criterion, evaluated by some metric (i.e. euclidean distance, correlation). Pairwise agglomerations are organized in a hierarchical way, starting in this case from genes and moving to sets (*nodes*) of increasing size, until all the items are comprised in a single set (*root* of the hierarchy). Generation of clusters is performed selecting nodes of the tree corresponding to a threshold value of the distance metric, or a level of agglomeration (*depth*) and defining the subtrees having as root these nodes as clusters. Despite the fact that genes organized this way can only fall in one cluster at a time, which is not biologically realistic, this method is broadly used, due to its simple and intuitive visualization.

5 Experimental Results

We present here two different applications of our method. The first example shows how it is possible to compare the results of a new clustering algorithm against a population of annotated gene sets. The goal is to assess if the new clustering algorithm is able to capture the biological information obtained and assessed with other techniques. The second example explores the possibility to relate gene sets annotated in two different studies, to infer new information

not apparent in the two studies alone.

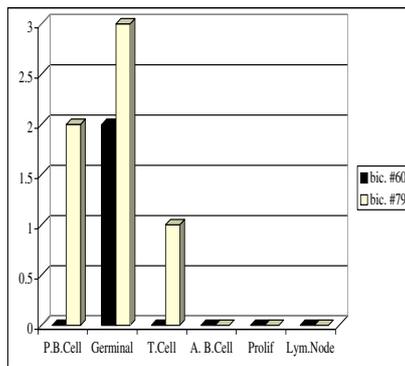
Lymphoma Dataset. This dataset ([2]) was processed using a hierarchical clustering approach. From this clustering analysis, two reference populations were extracted: (*R1*) the primary clusters (generated by the root division in the hierarchical tree) and (*R2*) a population of six annotated clusters (Pan B Cell, Germinal Centre Cell, T Cell, Activated B Cell, Proliferation, Lymph Node) which were selected from the nodes of the hierarchical tree at different levels of depth. For the test population we chose the gene bicluster population (*T*) obtained through the use of the biclustering algorithm on the same dataset (curated by Cheng et al. [5]). We then applied FIT in order to measure how well the biclustering algorithm of Cheng et al. was able to capture in *T* the information obtained with hierarchical clustering in (*R1*) and (*R2*). We first compared *T* with *R1*. We found that 10 of the 100 gene biclusters were exclusively related to one of the primary gene clusters, thus confirming the results obtained with the *mix* measure by Cheng et al. (see Section 2). We then explored the comparison between *T* and *R2* and ranked for every set in *R2* the best fitting sets in *T*. Interestingly, the set of best biclusters in *T* extracted by the combined statistic ϕ was not the same extracted by the enrichment statistic alone. Namely, bicluster #60 (best fitness score) had only 2 genes in common with *R2* and both of them were related to Germinal Centre Cell cluster. Conversely, bicluster #79 (best enrichment score), had 3 genes related to Germinal Centre Cell cluster, but with overall 6 genes in common with *R2*, spread over 2 more categories (Pan B Cell, T Cell). For a graphical visualization of these facts see Figure 2(a). Thus, since bicluster #79 had a more uniform distribution of the genes across the reference sets, the final fitness score was worse than the one evaluated for bicluster #60 (6.3 vs 3.6). In conclusion FIT was able to recapitulate the finding of previous work and also to give a new meaning to the genes biclusters in term of the biochemical information contained in sets of genes annotated from the same dataset.

Acute Lymphoblastic Leukemia Datasets. We compare here the results of two studies conducted on Acute Lymphoblastic Leukemia (ALL). The first study ([7]) aims at finding the gene signatures shared by both cell lines and clinical samples, defining three subtypes of the ALL disease corresponding to three chromosomal abnormalities: TEL/AML1, BCR/ABL and MLL. We used these three corresponding clusters, extracted with SAM [17], as reference population. The second study [6] focuses on the sensitivity of ALL samples to L-Asparaginase (L-Asp), an important component of most treatment regimens for ALL. Resistance to this enzyme-derived drug is correlated to poor prognosis ([11]). Since the mechanism of activity of this compound are still largely unknown, this study aimed at finding the set of genes that are differentially expressed in

responders versus non-responders to L-Asp treatment, using SAM. We used this cluster as our unique test set. We used our method to perform an exploratory comparison to see if and how the signature of the genes related to L-Asp might be related to the three subtypes of disease. Results in Table 2(b) show that the signature related to TEL/AML1 chromosomal aberration is strongly related to the signature for L-Asp sensitivity. This appears to be in accordance with literature findings since TEL/AML1 mutation is known to be associated with good prognosis and sensitivity to L-Asp ([15]). Moreover, the highly significant values of ϕ and p^ϕ suggest, at no extra experimental cost, a specific, significant set of genes (the overlap between the signatures) on which further study could be conducted to possibly understand more about the molecular mechanisms underlying the sensitivity to L-Asp and better prognosis. FIT has thus been used to extract meaningful and better selected information embedded in sets already annotated. Because of the large amount of signatures available, such an approach can lead to the discovery of new information based on data already studied, accelerating the process of discovery of interesting sets of genes to be investigated. This can shorten the expensive and time consuming processes that allow the experimental identification of meaningful candidate genes.

6 Conclusions

Because of the high density of the data treated and produced, microarrays are a continuous source of new biological information. A necessary resulting effect of this property however, is the continuous challenge with respect to the management, evaluation and interpretation of the large scale data they produce. Here we show a new way of navigating through the sea of complexity produced by microarray experiments, devising a method to compare and contrast gene sets, discovering or uncovering potentially hidden associations and mining new similarities at no new biological experimental cost. Namely, we showed how it is possible to deal with large populations of sets, confirm and enhance the results of comparison between a benchmark and a newly found population of gene sets. We were also able to extract a significant set of genes to obtain new insights for the important functioning of L-Asp in the treatment of ALL patients, hopefully shortening and easing the process of discovery of meaningful candidate genes. Our approach also addresses the issue of the set population comparison when dealing with validation of new data mining algorithms. By combining two measures for similarity, we obtain the global score ϕ and p^ϕ , able to give a statistical measure of the similarity among sets. Obviously, more statistical tests could be added to enhance the robustness of the final score, and different statistical approaches could be devised to describe in as many ways as possible the similarity among sets. How-



(a) Sets Distributions

	MLL	BCR/ABL	TEL/AML1
L-Asp η (p^{η})	1.680 (0.028)	1.994 (0.142)	3.544 (5.687e-09)
L-Asp τ (p^{τ})	-0.816 (0.201)	0 (1)	0.816 (0.201)
L-Asp ϕ (p^{ϕ})	-1.371 (0.034)	0 (0.419)	2.894 (2.466e-08)

(b) Lymphoma Dataset

Figure 2. Figure 2(a) plots the distributions of two biclusters (#79 and #60) from the Lymphoma data set. Both biclusters are enriched for Germinal Center, however bicluster #60 is more specific. Table 2(b) shows the scores for comparison between signatures of 3 subtypes of ALL due to chromosomal abnormalities and the L-Asp cluster. The best fitting is for the TEL/AML1 signature, consistently with literature findings.

ever, this method provides interesting results, and can be used for broad general purposes applications, as the input required are very simple. In the complex landscape of the deciphering of genes interactions, FIT can give a meaningful contribution in allowing quantitative comparisons of entire sets of populations for computational validation, and biological insight.

References

- [1] F. Al-Shahrour, R. Díaz-Uriarte, and J. Dopazo. FatiGO: a web tool for finding significant associations of Gene Ontology terms with groups of genes. *Bioinformatics*, 20(4):578–580, 2004.
- [2] A. Alizadeh, M. Eisen, R. Davis, I. L. C. Ma, A. Rosenwald, J. Boldrick, H. Sabet, T. Tran, X. Yu, J. Powell, L. Yang, G. Marti, T. Moore, J. H. Jr, L. Lu, D. Lewis, R. Tibshirani, G. Sherlock, W. Chan, T. Greiner, D. Weisenburger, J. Armitage, R. Warnke, R. Levy, W. Wilson, M. Grever, J. Byrd, D. Botstein, P. Brown, , and L. Staudt. Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. *Nature*, 403(6769):503–511, 2000.
- [3] E. I. Boyle, S. Weng, J. Gollub, H. Jin, D. Botstein, J. M. Cherry, and G. Sherlock. GO::TermFinder-open source software for accessing Gene Ontology information and finding significantly enriched Gene Ontology terms associated with a list of genes. *Bioinformatics*, 20(18):3710–3715, 2004.
- [4] C. I. Castillo-Davis and D. L. Hartl. GeneMerge-post-genomic analysis, data mining, and hypothesis testing. *Bioinformatics*, 19(7):891–892, 2003.
- [5] Y. Cheng and G. M. Church. Biclustering of expression data. *Proceedings of ISMB*, 2000.
- [6] B. M. Fine, G. J. L. Kaspers, M. Ho, A. H. Loonen, and L. M. Boxer. A genome-wide view of the *in vitro* response to L-Asparaginase in acute lymphoblastic leukemia. *Cancer Res.*, 65(1):291–299, 2005.
- [7] B. M. Fine, M. Stanulla, M. Shrappe, M. Ho, S. Viehmann, J. Harbott, and L. M. Boxer. Gene expression patterns associated with recurrent chromosomal translocations in acute lymphoblastic leukemia. *Blood*, 103(3):1043–1049, 2004.
- [8] G. Getz, E. Levine, and E. Domany. Coupled Two-Way Clustering analysis of gene microarray data. *PNAS*, 97(22):12079–12084, 2000.
- [9] L. B. Hedges and I. Olkin. *Statistical Methods in Meta-Analysis*. Academic Press, New York, (1985).
- [10] S. C. Madeira and A. L. Oliveira. Biclustering algorithms for biological data analysis: A survey. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 1(1):24–45, 2004.
- [11] T. McLean, S. Ringold, D. Neuberg, K. Stegmaier, R. Tantravahi, J. Ritz, H. Koefler, S. Takeuchi, J. Janssen, T. Seriu, C. Bartram, S. Sallan, D. Gilliland, and T. Golub. TEL/AML-1 dimerizes and is associated with a favorable outcome in childhood acute lymphoblastic leukemia. *Blood*, 8(11):1252–4258, 1996.
- [12] J. Quackenbush. Computational analysis of microarray data. *Nat Rev Genet*, 2(6):418–427, 2001.
- [13] R. R. Sokal and F. J. Rohlf. *Biometry*. Freeman, New York, 2003.
- [14] J. L. Sevilla, V. Segura, A. Podhorski, E. Guruceaga, J. M. Mato, L. A. Martínez-Cruz, F. J. Corrales, and A. Rubio. Correlation between gene expression and GO semantic similarity. *IEE/ACM TCCB*, 2(4):330–338, 2005.
- [15] W. A. G. Stams, H. B. B. M. L. den Boer, J. P. P. Meijerink, R. L. Stigter, E. R. van Wering, G. E. Janka-Schaub, R. Slater, and R. Pieters. Sensitivity to L-asparaginase is not associated with expression levels of asparagine synthetase in $t(12;21)^+$ pediatric ALL. *Blood*, 101(7):2743–2747, 2003.
- [16] The Gene Ontology Consortium. Creating the Gene Ontology resource: Design and implementation. *Genome Res.*, 11(8):1425–1433, 2001.
- [17] V. G. Tusher, R. Tibshirani, and G. Chu. Significance analysis of microarrays applied to the ionizing radiation response. *PNAS*, 98(9):5116–5121, 2001.
- [18] S. Yoon, L. Benini, and G. D. Micheli. Finding co-clusters of genes and clinical parameters. In *Proceedings of IEEE 27th Annual International Conference of IEEE Engineering in Medicine and Biology*, number 2109, 2005.
- [19] B. Zhang, S. Kirov, and J. Snoddy. Webgestalt: an integrated system for exploring gene sets in various biological contexts. *Nucleic Acids Res.*, 33(Web Server issue)(doi:10.1093/nar/gki475):W741–W748, 2005.