Finding disease related gene ontology subnets using DNet

XX, *Member, IEEE,* XX, *Fellow, OSA,* XX, *Life Fellow, IEEE*

Abstract—The abstract goes here.

Index Terms—Computer Society, IEEE, IEEEtran, journal, LaTeX, paper, template.

1 Introduction

It is still popular to find differentially expressed genes by microarray analysis. Independent of the platform and the analysis methods used, the result of a microarray experiment is, in most cases, a list of differentially expressed genes [1]. There are many techniques used to identify differentially expressed genes. These techniques can be divided into three categories: individual genes, gene pathways and gene classes approaches [2]. In the past, peoples focus is in the function of one gene. However, the expression level of this gene in the cells (including tumor cells) is not so high. What really works in cell activity is not a single gene, but a group of genes that participate in a function. So it is necessary to introduce the systematic biology perspective to slightly macroscopically observe and evaluate the function of the cells. Recently, more and more research is devoted to gene sets instead of individual genes. Gene sets are group of differentially expressed genes, these genes are usually relating to diseases. Here, we call the differentially expressed genes between patients and normal for the disease genes. The gene sets blur the powerful function of the individual gene, and is more concerned with the role of the whole functional group, thus closer to the normal condition of the cell, which is also its advantage.

Studying gene sets is more conducive to understanding the cause of disease and find out which part of biological functions is affected in the disease. However, the increasing complexity of gene expression data presents several challenges for researchers [2]. A big challenge faced by the researchers is how to choose disease genes according to patients gene expression value and how to translate disease genes into a better understanding of the underlying biological process. There are many microarray analysis methods proposed to solve the problem. Such as GSEA, SNet, PFSNet and so on. Those methods are useful for diagnosis of a disease. However, there are also some shortcomings for traditional methods: GSEA is a method based on pathways. Its enrichment analysis was carried out to find out relevant pathways. The essence of this method is the whole pathways are considered and each pathway is scored by a variation of Kolmogorov-Smirnov statistic which determines its importance. But pathways are often large and the statistical score is easily affected by the genes in each pathway. When a pathway contains too few useful genes, the score of it will be low and finally this pathway will be missed, hence causing a big impact on the results. As for SNet, it is a network-based method. In this method, differentially expressed genes is selected by setting a threshold on the gene expression levels. SNet use gene rank value rather than the absolute gene expression values to select genes. This approach makes the results more consistent in two independent datasets. The shortcoming of this method is the threshold is hard to choose and genes around threshold is easily missed leading the results are not accurate. Based on SNet, there is another method called PFSNet. In PFSNet, the genes around threshold are considered by setting two threshold to choose significant genes. Another meaningful improvement of PFSNet is PFSNet do not use the absolute gene expression values but use a fuzzy value to substitute gene expression level according to gene ranks. The fuzzy value is a number between 0 and 1 [3]. This method scores each subnet of pathways using a paired t-statistic based on the fuzzy score of two phenotypes. Finally the p-value of every single subnet is estimated within the subnets list and keep those which are significant. This principle of network-based or pathway-based association [4] is now being applied to effectively map the genetics underlying complex phenotypes, including cancer and other common diseases [5]. However, the scoring process of these methods do not consider the structure of each subnet. The significance of each biological process is measured by gene sets, which only contain differentially expressed genes. So the question is: Is it possible that the causative gene is not highly expressed but affects the expression of genes that are closely linked surrounding it?

This paper looks at a problem of finding biological process according to gene sets based on GO. Different from the previous method, the gene sets we study here are composed of differential expressed genes and some new genes. These new genes have potential to play the same function as differential expression genes. Directly, the gene sets are first mapped onto knowledge of GO terms; affected subnets are then statistically associated with the disease phenotype. We identify some shortcomings of the previous methods in finding consistent disease subnets. Thinking they did not consider the biological process structural information and just focus on the absolute differentially expressed genes. In this article we present our technique, DNet, to identify significant biological process within a phenotype of microarray experiments. The method drives its power by focusing on gene sets as well as GO term network structure information. This method can greatly increase our power to identify relevant associations between phenotype and biological process [5].

2 METHOD

We blend information on GO and genes from GO annotations. The GO network we study here is a directed graph (DAG) of terms and hierarchical relations. It consists of three branches: biological processes, cellular components and molecular functions. Here, we focus on researching biological process. The graph is made up of vertices and nodes. It is defined as G= (V, E). V is the vertex set whose elements are the nodes of the graph. This set is often denoted V(G) or just V. E is the edge set whose elements are edges of the graph. This set is often denoted E(G) or just E [6]. Each vertices represents a GO term and each edges represent the relationship between two terms.

The subnets we found here is an undirected graph, defined as S = (V′, E′). Each vertices represents a gene and each edges represent the relationship between the two genes. Every subnet belong to a term and represent a small biological process. The size of significant subnets we choose here is more than five.

The density of the graph directly reflects the sparseness of the graph. For undirected simple graphs, the graph density is defined as:

$$D = \frac{2 * |E|}{|V| * (|V| - 1)} \tag{1}$$

where E is the number of edges and V is the number of vertices in the graph.

The maximum number of edges for an undirected graph is $|V| * (|V| - 1)/2$, so the maximal density is 1 (for complete graphs) (Fig. 1a) and the minimal density is 0 (Fig. 1c) [7]. Also, for a undirected graph with n nodes, the condition of the connected graph is that there is at least $n - 1$ edges (Fig. 1b).

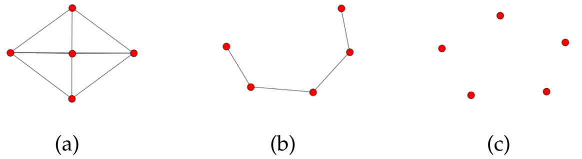


Fig. 1. (a) The maximum density of undirected graph is 1. (b) To make the undirected graph which has n nodes connected, there is at least $(n - 1)$ edges. (c) The minimal density of a undirected graph is 0.

2.1 Subnet genereation

For each disease we study here, we consider test patients and control patients gene expression level together. The phenotype of test patients is defined as D and the phenotype of normal patients is defined as $\neg D$. For the patients respect to D and $\neg D$, we both rank each gene for each patients based on gene expression value. After ranking, we give each gene a weight value $w(g_i, p_j)$. Genes whose rank is up 95% get a weight value 1 and genes whose rank is below 85% get a weight value 0. Genes between 85% and 95% are given a weight value between 0 and 1 based on their rank. These weights reflect the expression levels of different genes in different patients. Then we calculate each genes average weight value using each genes weight value and choose genes whose average weight value greater than 0.5 into the gene list L. These genes are considered to be differentially expression genes in the disease because their weight value are higher than the average in most patients.

When we complete the construction of the gene list of test patients and control patients, the next step is to generate subnets using the gene sets. In this step, we first generate the subnets S according to L. Then we add genes which are tightly connected to subnets into S. Since the differentially expressed genes may be caused by genes that are not themselves differentially expressed. For example, a gene has a mutation which impact on the shape of its protein product but its expression level are not affected. In this case, it can affect the surrounding genes which closely linked to it causing other genes high expressed. It has a great impact on many differentially genes but its own expression level has no change. Thus it is easily to be ignored since the traditional method only pick out genes that have great changes in expression. So it is more likely that the genes which have more connection to gene sets L plays a similar function. The function of these genes just like a catalyst and we should pay attention to these extra genes as well. E.g. for subnet S with n nodes, when a node that dont belong to S connect to more than $2 * n/3$ nodes in the subnet S, we add this node into S. (Fig. 2)

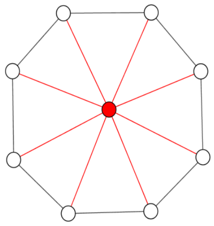


Fig. 2. We add nodes that closed to subnets into subnets S

2.2 Subnets scoring

The goal in this step is to score every subnet and find out the significant subnet relate to disease. In this step, we consider the difference between the D and $\neg D$. For every subnet S, each patients of phenotype D and $\neg D$ can be scored as follows:

$$scores_1^{p_k}(S) = \sum_{g_i \in S} w(g_i, p_k) * \sum_{p_k \in D} \frac{w(g_i, p_k)}{|D|} \tag{2}$$

$$scores_2^{p_k}(S) = \sum_{g_i \in S} w(g_i, p_k) * \sum_{p_k \in \neg D} \frac{w(g_i, p_k)}{|\neg D|} \tag{3}$$

Where D is the phenotype for test patients and $\neg D$ is the phenotype for normal patients. p_k ranges over the patients of phenotype D and $w(g_i, p_j)$ is the weight value for each patient of each gene.

In this step, we get two scores for each subnet. When we get two scores that both describe the network S, we expect the scores calculated by D and $\neg D$ is really diffident since they come from different phenotype and use the different patients datasets. So a paired t-test was done under the null hypothesis that the difference in scores gives us a distribution with mean=0. While doing a t-test, we considered the structural information of the subnet as well. It is easy to think that sparse networks and dense networks are really different. A dense network has more edges and is more likely to have informations. The nodes in dense network are closely connected and have the tendency to play a same function. Thus, the density is also listed as one of the factors that influence the subnet score. So a variant t-test is used to score each subnet. The final score of subnets not only use the gene average weight value but also the density of the subnet. For every subnet S, the score is defined as follows:

$$DF^{p_k}(S) = scores_1^{p_k}(S) - scores_2^{p_k}(S) \tag{4}$$

$$\overline{x} = \frac{\sum_{p_k \in D} DF^{p_k}(S)}{n} \tag{5}$$

$$Score(S) = -\frac{1}{\log DS(S)} * \frac{n * \overline{x}}{\sqrt{\sum_{p_k \in D} (DF^{p_k}(S) - \overline{x})^2}} \tag{6}$$

Where DS(S) is the density of subnet S.

3 RESULTS

We use GO from Gene Ontology Consortium, a project that address the need for consistent description of gene products across databases. The GO project has develop three structured ontologies that describe gene products in term of their associated biological processes, cellular components and molecular functions in a species-independent manner. A biological process is a recognized series of events or molecular functions. Here, we study the biological process across GO database. The biological process of GO related to more than 20000 terms and more than 60000 relationships between terms. It describe the knowledge on the molecular interaction and reaction. We tested DNet on independent datasets of two diseases. For each of the two disease types studied herełLeukemia [8] [9]and Duchenne Muscular Dystrophy (DMD) [10] [11]łwe obtain two independent datasets which are produced using different microarray platforms (HG-U85Av2, HGU133A and HU6800). For each disease type, we run DNet and PFSNet on the two datasets independently and obtain a corresponding outputs about the significant biological process. We compared the results form DNet and PFSNet. When comparing DNet and PFSNet, we set θ_1= 5% and θ_2 = 15% for PFSNet. In this way, genes above 95th percentile are given a total vote and genes below 85th are given no vote at all. This allows same genes to be considered in DNet and PFSNet. In the results, we compare top 10, top 20, top 30 and top 40 significant subnets and significant genes from two datasets using Jaccard similarity coefficient method. It is defined as follows:

$$J(A, B) = \frac{|A \cap B|}{|A \cup B|} \tag{7}$$

3.1 comparing DNet and PFSnet

We run DNet and PFSNet on two datasets and analysis the results of top 10, top 20, top 30 and top 40 subnets. In DNet, we get even higher subnet-level agreement and gene-level agreement than PFSNet in the two datasets (Fig.3). In the top 10 subnets of the results, DNet achieves the maximum subnet agreement of 25% in Leukemia datasets and 11.1% in DMD datasets whereas PFSNet achieves the maximum subnet agreement of 17.6% in Leukemia datasets and 5.3% in DMD datasets. This shows the subnet structure plays an important role in consistent disease subnet. Besides, as we allow more genes which closely related to the subnets to be considered, we also measure the gene-level agreement from significant subnets between two datasets to see whether these genes are similar. The result shows that adding these genes into subnets makes the gene-agreement of DNet even higher. In the top 10 subnets of results, DNet achieves the maximum gene agreement of 58% in the Leukemia dataset whereas PFSNet achieves maximum gene agreement of 53.3%. In the DMD datasets, DNet achieves the maximum gene agreement of 91.5% whereas PFSNet achieves maximum gene agreement of 50%. The good results provide that our methods has the ability to find more consistent disease genes in the independent datasets for the same disease. Since the extra genes we find is similar to the differentially expressed genes in both datasets.

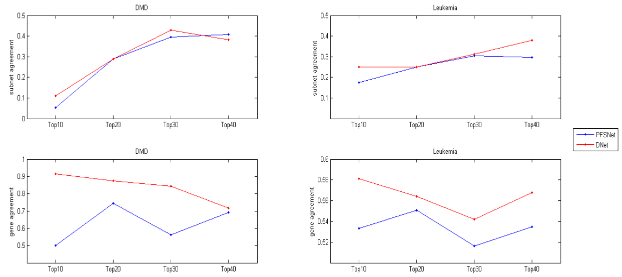


Fig. 3. Consistency of subnets and their genes in Leukemia and DMD dataseet.

3.2 comparing GO and subnets

GO is a big directed graph. It contains more than 40000 terms and 80000 relationships. The GO data provides a very effective way of linking biological knowledge with the analysis of the large datasets of post-genomics research [12]. As for GO terms, each GO term can be seen as a network made up by genes. We chose terms whose annotation genes between 50 and 100 from the biological process of GO in our methods. These terms total contain more than 7000 genes and almost 100000 relationships. After that, we break each terms network into small subnets to see whether it is associated with diseases in two independent datasets. As there are two independent datasets for each disease, we can get two results which describe the significant subnets and genes found from independent datasets.

We first analysis the significant subnets to see how many subnets are appear in the results generated by different datasets. For the subnets both appear in two datasets, we also compare our methods with PFSNet about the intersection genes for each disease. (Fig.4). Comparison results shows that whether it is DMD or Leukemia, the number of intersection subnets and genes in DNet is more than PFSNet. The overlap we obtained from the PFSNet is low. For example, in the DMD datasets, the number of intersection subnets in DNet is 2 whereas the number in FSNet is 1. We compare the number of intersection genes in the intersection subnets as well. The result is the number of intersection genes in DNet is 86 while FSNet can only find 4 genes. It is obvious that the structure of each subnet affects its ranking. When we consider the density of each subnet, we can get a higher agreement among independent datasets. Thus it is more probable that the subnets selected by our method can provide more informations about disease. Also, the genes that not differentially expressed but closely connect to gene sets is similar to differentially expressed genes, consider these genes leading the gene agreement become higher.

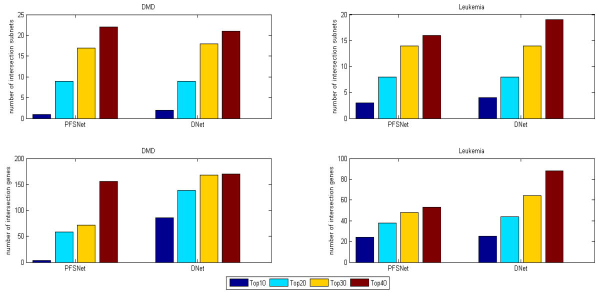


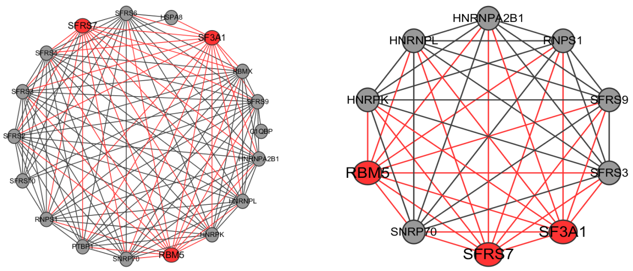
Fig. 4. Number of intersection subnets and genes

3.3 biologically significant subnets

For the subnets and extra genes we find here, we check them for biologically significance. We discover that the genes we add to subnets participate in many important process of the disease.

For Leukemia dataset, one of the significant subnet we find is term GO: 0048024. The name of this term is regulation of mRNA splicing, via spliceosome. This subnet is associate with Leukemia. We compare the output of this subnets nodes between DNet and PFSNet and get two gene list. Comparing the two gene lists we find 3 genes that appear in DNet while PFSNet do not contain. They are RBM5, SFRS7 and SF3A1 (Fig. 5). Through observation the structure of the subnet, we found that this three genes are closely linked to the subnet and they are inseparable.

Serine/arginine-rich splicing factor 7 (SFRS7) is related to Leukemia. The protein encoded by it is a member of the serine/arginine (SR)-rich family of pre-mRNA-splicing factors, which constitute part of the spliceosome (Table 1) [13] [14] [15]. The SR family plays an important role in the alternative splicing and it is closely related to the occurrence and development of tumor. In the past years, the study of DNA and RNA sequencing (RNAseq) has been very mature. Comparative DNA and RNA sequencing studies have revealed that humanspecific distal regulatory elements, RNA editing, and alternative splicing play key roles in human embryonic stem cell (hESC) self-renewal and cell fate determination. Several of the phosphoproteins regulated during differentiation are components of the posttranscriptional RNA modification machinery, including double-stranded RNA-specific adenosine deaminase (ADAR) and serine/arginine-rich splicing factor 7 (SFRS7) [16]. In many studies of Leukemia, various spliceosome gene mutations were detected and SFRS7 is one of the members [17]. Recent studies have shown that some spliceosome genes involved in the early steps of U2-dependent splice site recognition are commonly mutated in hematologic malignancies and solid cancers. For example, exome-sequencing studies found that SRSF7 was mutated in the patients with chronic lymphocytic leukemia.



(a) (b)

Fig. 5. The closely linked genes we find from term GO:0048024 in different datasets.

TABLE 1
SR-rich familye

Gene name	SR protein	Chromosomal	location UniProt
SFRS1	SF2/ASF/SRp30a	17q21.3-q22	Q07955
SFRS2	SC35/SRp30b	17q25.1	Q01130
SFRS3	SRp20	6p21.31	P84103
FRS4	SRp75	1p35.3	Q08170
SFRS5	SRp40	14q24.2	Q13243
SFRS6	SRp55	20q13.11	Q13247
SFRS7	9G8	2p22.1	Q16629
SFRS9	SRp30c	12q24.23	Q13242
SFRS11	SRp54	1p31.1	Q05519

The RNA maturation is an important and complex biological process. It requires several small nuclear ribonucleoproteins (snRNPs) that comprise the two forms of spliceosomes. The major form of spliceosome (U2-type) is composed of U1, U2, U4/6 and U5 snRNPs, and catalyzes most splicing events in metazoans.Mutations of genes, such as SF3B1, SRSF2, U2AF1, ZRSR2, and to a lesser extent SF1, SF3A1, U2AF2 or PRPF40B, encoding spliceosome compounds have been found to occur at high frequencies in myelodysplastic syndromes (MDS) and chronic lymphocytic leukemia (CLL). Subsequently, SF3B1 mutations were also found in solid tumors such as endometrial, lung, bladder, pancreatic and breast carcinomas and cutaneous melanomas [18]. In myelodysplastic syndrome, spliceosome genes were reported to be mutated in 45C85% of patients; mutations were found in SF3A1, PRPF40B and so on. These findings illustrate that RNA splicing-related genes appear to be associated with cancer [19].

The study about RBM5 suggest that RBM6-RBM5 transcription-induced chimerism might be a process that is linked to the tumour-associated increased transcriptional activity of the RBM6 gene. It appears that none of the transcription-induced chimeras generates a protein product; however, the novel alternative splicing, which affects putative functional domains within exons 3, 6 and 11 of RBM6, does suggest that the generation of these chimeric transcripts has functional relevance. Finally, the association of chimeric expression with diseases suggests that RBM6-RBM5 chimeric expression may be a potential tumour differentiation marker [20]. Thus, the disruption of these genes results in splicing of key genes and pathways in myelodysplastic syndrome hematopoietic stem and progenitor cells.

For DMD, one of the significant subnet is term GO: 0019886. The name of this term is antigen processing and presentation of exogenous peptide antigen via MHC class II. It is a subsystem of the immune system process. In this subnet, we find gene HLA-DMB is associate with subnet. We also check its biological significance. DMD is a severe type of muscular dystrophy. Typically muscle loss occurs first in the upper legs and pelvis followed by those of the upper arms. This can result in trouble standing up [21]. Studies of DMD have shown that these diseases are closely related to HLA class II region encode products [22]. Genes of the HLA encode products plays a center role in immune response. HLA-DMB is a part of HLA family. The variants of HLA-DMB were found in most autoimmune disease patients. This disease is same to DMD. It is a chronic systemic autoimmune disease consisting mainly of joint disease. The main clinical manifestations of joint swelling and pain, joint stiffness, deformity, dysfunction.

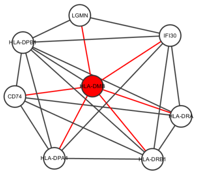


Fig. 6. The closely linked genes in term GO:0019886

4 Conclusion

Analyzing gene sets rather than individual genes has many advantages. The most fundamental thing is that the disease is not caused by a single gene but a group of interacting genes. A variety of methods have been developed for discover biological process about disease based on the differential expression genes. Pathways and GO is always used to analysis. They are all represent the biological process in a form of network. Here, we are committed to finding GO terms related to differentially expression genes. But GO is very large and each term is associated to a biological process. The network of GO describe the whole process of human life. Thus, the total network is not conductive to analysis because they will cause false positives. In order to solve this problem, we choose the differentially genes composing gene sets, then we map the gene sets to the subnets. Methods that analysis subnets such as SNet, PFSNet have achieved by dividing the biological process into small parts. But there are still some shortcomings in these methods. First, the gene sets in these methods are composed of differentially expression genes only. Genes that do not change expression value is filtered. Second, when scoring the subnets, the process of scoring is only related to the expression value of each gene. As a graph, the structure is as important as the nodes value.

In this article, we improved the above two shortcomings by consider the closely connected genes and the density of each subnets. We compare our method with PFSNet in two independent datasets of DMD and Leukemia and analysis the result of two methods. We have found that the results turns to be more consistent between two independent datasets and genes we find are meaningful and biologically significant, which prove that our approach is right.

References

[1] Purvesh Khatri and Sorin Drăghici. Ontological analysis of gene expression data: current tools, limitations, and open problems. *Bioinformatics*, 21(18):3587–3595, 2005.

[2] Donny Soh, Difeng Dong, Yike Guo, and Limsoon Wong. Finding consistent disease subnetworks across microarray datasets. *BMC bioinformatics*, 12(13):S15, 2011.

[3] K Lim and L Wong. Finding consistent disease subnetworks using pfsnet. *Bioinformatics*, 30(2):189–196, 2014.

[4] Andrea Califano, Atul J Butte, Stephen Friend, Trey Ideker, and Eric Schadt. Leveraging models of cell regulation and gwas data in integrative network-based association studies. *Nature Genetics*, 44(44):841–847, 2011.

[5] Michael Ku Yu, Michael Kramer, Janusz Dutkowski, Rohith Srivas, Katherine Licon, Jason F Kreisberg, Cherie T Ng, Nevan Krogan, Roded Sharan, and Trey Ideker. Translation of genotype to phenotype by a hierarchy of cell subsystems. *Cell systems*, 2(2):77–88, 2016.

[6] Thomas F Coleman and Jorge J Moré. Estimation of sparse jacobian matrices and graph coloring blems. *SIAM journal on Numerical Analysis*, 20(1):187–209, 1983.

[7] Norman Biggs, E Keith Lloyd, and Robin J Wilson. *Graph Theory, 1736-1936*. Oxford University Press, 1976.

[8] Scott A Armstrong, Jane Staunton, Lewis B Silverman, Rob Pieters, Monique L Den Boer, Mark D Minden, Stephen E Sallan, Eric S Lander, Todd R Golub, and Stanley J Korsmeyer. Mll translocations specify a distinct gene expression profile that distinguishes a unique leukemia. *Nature Genetics*, 30(1):41–47, 2002.

[9] Todd R Golub, Donna K Slonim, Pablo Tamayo, C Huard, Michelle Gaasenbeek, Jill P Mesirov, Hendrik Van Coller, Mignon L Loh, James R Downing, Michael A Caligiuri, et al. Molecular classification of cancer : Class discovery and class prediction by gene expression monitoring. *Science*, 286(5439):531–537, 1999.

[10] Judith N Haslett, Despina Sanoudou, Alvin T Kho, Richard R Bennett, Steven A Greenberg, Isaac S Kohane, Alan H Beggs, and Louis M Kunkel. Gene expression comparison of biopsies from duchenne muscular dystrophy (dmd) and normal skeletal muscle. *Proceedings of the National Academy of Sciences of the United States of America*, 99(23):15000–15005, 2002.

[11] M Pescatori, A Broccolini, C Minetti, E Bertini, C Bruno, A D'Amico, C Bernardini, M Mirabella, G Silvestri, and V Giglio. Gene expression profiling in the early phases of dmd: a constant molecular signature characterizes dmd muscle from early postnatal life throughout disease progression. *Faseb Journal Official Publication of the Federation of American Societies for Experimental Biology*, 21(4):1210, 2007.

[12] Ruth C Lovering, Emily C Dimmer, and Philippa J Talmud. Improvements to cardiovascular gene ontology. *Atherosclerosis*, 205(1):9–14, 2009.

[13] Y. Zheng, J. Zhou, and Y. Tong. Gene signatures of drug resistance predict patient survival in colorectal cancer. *Pharmacogenomics Journal*, 15(2):135–143, 2014.

[14] Levi A Garraway and Eric S Lander. Lessons from the cancer genome. *Cell*, 153(1):17–37, 2013.

[15] Peter J Shepard and Klemens J Hertel. The sr protein family. *Genome Biology*, 10(10):242–242, 2009.

[16] Frida Holm, Eva Hellqvist, Cayla N Mason, Shawn A Ali, Nathaniel Delos-Santos, Christian L Barrett, Hye-Jung Chun, Mark D Minden, Richard A Moore, Marco A Marra, et al. Reversion to an embryonic alternative splicing program enhances leukemia stem cell self-renewal. *Proceedings of the National Academy of Sciences*, 112(50):15444–15449, 2015.

[17] Uri Rozovski, Michael Keating, and Zeev Estrov. The significance of spliceosome mutations in chronic lymphocytic leukemia. *Leukemia & lymphoma*, 54(7):1364–1366, 2013.

[18] D Gentien, O Kosmider, F Nguyen-Khac, B Albaud, A Rapinat, AG Dumont, F Damm, T Popova, R Marais, M Fontenay, et al. A common alternative splicing signature is associated with sf3b1 mutations in malignancies from different cell lineages. *Leukemia*, 28(6):1355, 2014.

[19] Jing Tian, Yaping Liu, Beibei Zhu, Yao Tian, Rong Zhong, Wei Chen, Xinghua Lu, Li Zou, Na Shen, Jiaming Qian, et al. Sf3a1 and pancreatic cancer: new evidence for the association of the spliceosome and cancer. *Oncotarget*, 6(35):37750, 2015.

[20] Ke Wang, Gino Ubriaco, and Leslie C Sutherland. Rbm6-rbm5 transcription-induced chimeras are differentially expressed in tumours. *BMC genomics*, 8(1):348, 2007.

[21] National Institute of Neurological Diseases and Stroke. *Muscular Dystrophy: Hope Through Research*. Information Office, National Institute of Neurological Diseases and Stroke, 1971.

[22] A Sensi, A Venturoli, S Traniello, M Lucci, C Vullo, C Conighi, PL Mattiuz, and OR Båricordi. Impaired hla capping capacity of peripheral blood lymphocytes in duchenne muscular dystrophy. *Journal of medical genetics*, 21(3):182–185, 1984.