

The difference between Treg cells and naive T cells analyzed by cluster analysis

Yuchen Wang^{1, a}

¹School of Public Health, Southern Medical University, Guangzhou 510515, China

^a907622192@qq.com

Abstract: This article hopes to find differences in the function of Treg cells and naive T cells by high-throughput analysis. With the increasing demand for low-cost sequencing, the development of high-throughput sequencing has been promoted, which multiplexes the sequencing process and produces thousands or millions of sequences. RNA-seq is a transcriptional sequencing technique in which mRNA, small RNA, and non-coding RNA, or some of them, are identified by high-throughput sequencing techniques. Reflecting their level of expression. It is a powerful tool for deep research of the complexity of transcription groups.

Key words: RNA sequence, immunologic tolerance

1. Introduction

In recent years, it has been found that regulatory T cells (Treg cells) can be used to treat certain autoimmune diseases, transplant rejection, and play an important role in mediating tumor immune surveillance. Therefore, regulatory T cells provide a new approach to immunotherapy for clinical disease.

So far, people have different understanding of the characteristics and mechanisms of action of different regulatory T cell subsets, but there are still many problems that are not solved, such as how to activate in vitro and in vivo amplification of regulatory T cells to achieve the purpose of clinical treatment. For the more advanced understanding of the body's regulatory T cell population, this article to the inherent CD4 + CD25 + regulatory cells, through high-throughput sequencing, the regulation of several T cells and the mechanism of action to explore a mechanism.

The high-throughput analysis technique can be used in study on the structure of transcription such as gene fusion, coding region SNP study, non-coding region functional studies, transcriptional variation studies (gene boundary identification, variable shear studies, etc.), transcript variation studies Gene expression levels and new transcripts.

2. Conclusions: Since a growing number of Treg and naive T cells RNA sequence have been doing presently, I would like to continue to explore related content.

The following analysis of the results of the data source is Ann-Jay Tong's article^[1]. We used the published list of genes that had significant differences in T-reg and naive T cells published by the authors for statistical and cluster analysis. It is desirable to explain the reasons for the differences in function between the two cells in some ways.

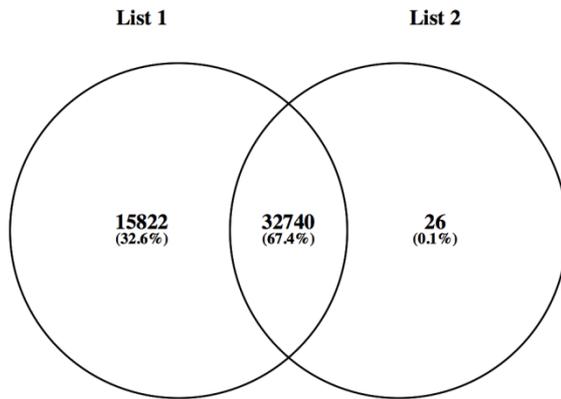


figure1: This figure is done with venny (<http://bioinfo.cnb.csic.es/tools/venny/>). Enter List1 and List2 First in Venny firstly. List1 stands for t-reg RNA sequence, and list2 stands for Naive T cells sequence. The two overlapping parts are the two groups of co-expressed genes, a total of 32740, and the non-overlapping part is the number of genes present in only one experimental group, there are 15822 in Treg cells and 26 in naive T cells.

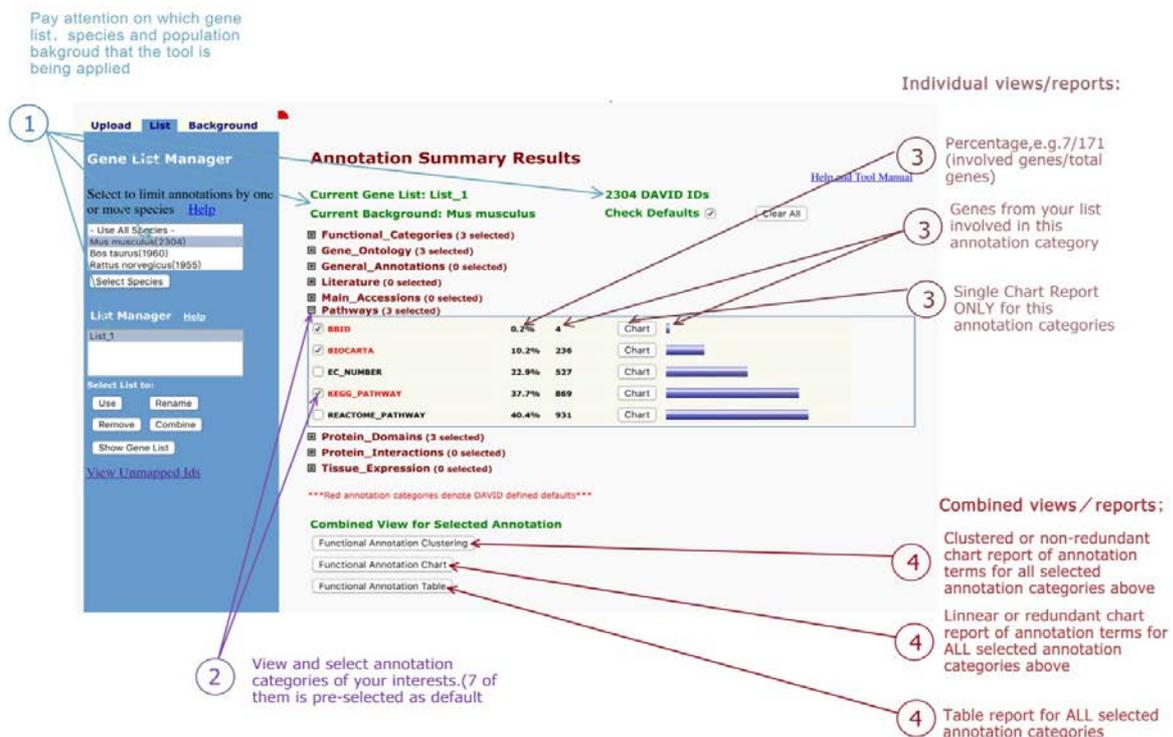


Figure2: Functional Annotation Tool

Figure 2, the figure is done with DAVID Bioinformatics Resources 6.8(<https://david.ncifcrf.gov/home.jsp>). The input gene is the overlapping gene from figure 1,

First click on the left side button > **Functional Annotation** , Then enter the overlapping part of the gene in Figure 1 , Step 2: Select Identifier , select > OFFICIAL GENE SYMBOL , Step 3: List Type, select Gene List, Step 4: Submit List. And I get Annotation Summary Results. Then select the species Mus musculus (2304 genes). One of the results is as above.

Pathway analysis returns a p-value for each pathway that differentiates the gene, and the small p value indicates that the differential gene appears in the pathway. Enrichment, pathway analysis has a role in presenting the results. Pathway analysis of the differential gene can be used to find the Pathway entry that enriches the differential gene, and the difference gene of different samples may be related to the change of the cell signaling pathway. Unlike the GO analysis, the results of the pathway analysis are more indirect, because the pathway is the interaction between the proteins, and the changes in the pathway can be caused by the expression of the protein involved in this pathway or the activity of the protein.

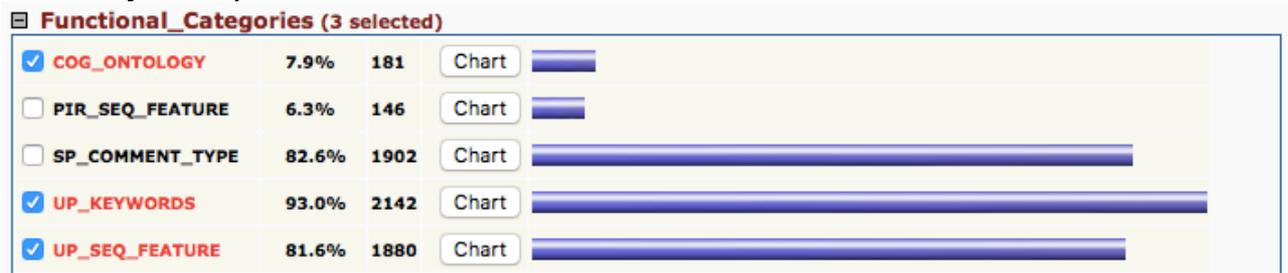


Figure 3:Functional Annotation Tool

Figure 3 is done with DAVID Bioinformatics Resource6.8(<https://david.ncifcrf.gov/home.jsp>).

The input gene is the overlapping gene from figure 1, First click on the left side button > **Functional Annotation** , Then enter the overlapping part of the gene in Figure 1 , Step 2: Select Identifier , select > OFFICIAL GENE SYMBOL , Step 3: List Type, select Gene List, Step 4: Submit List. Then I get Annotation Summary Results. Then select the species Mus musculus (2304 genes). One

of the results is as above

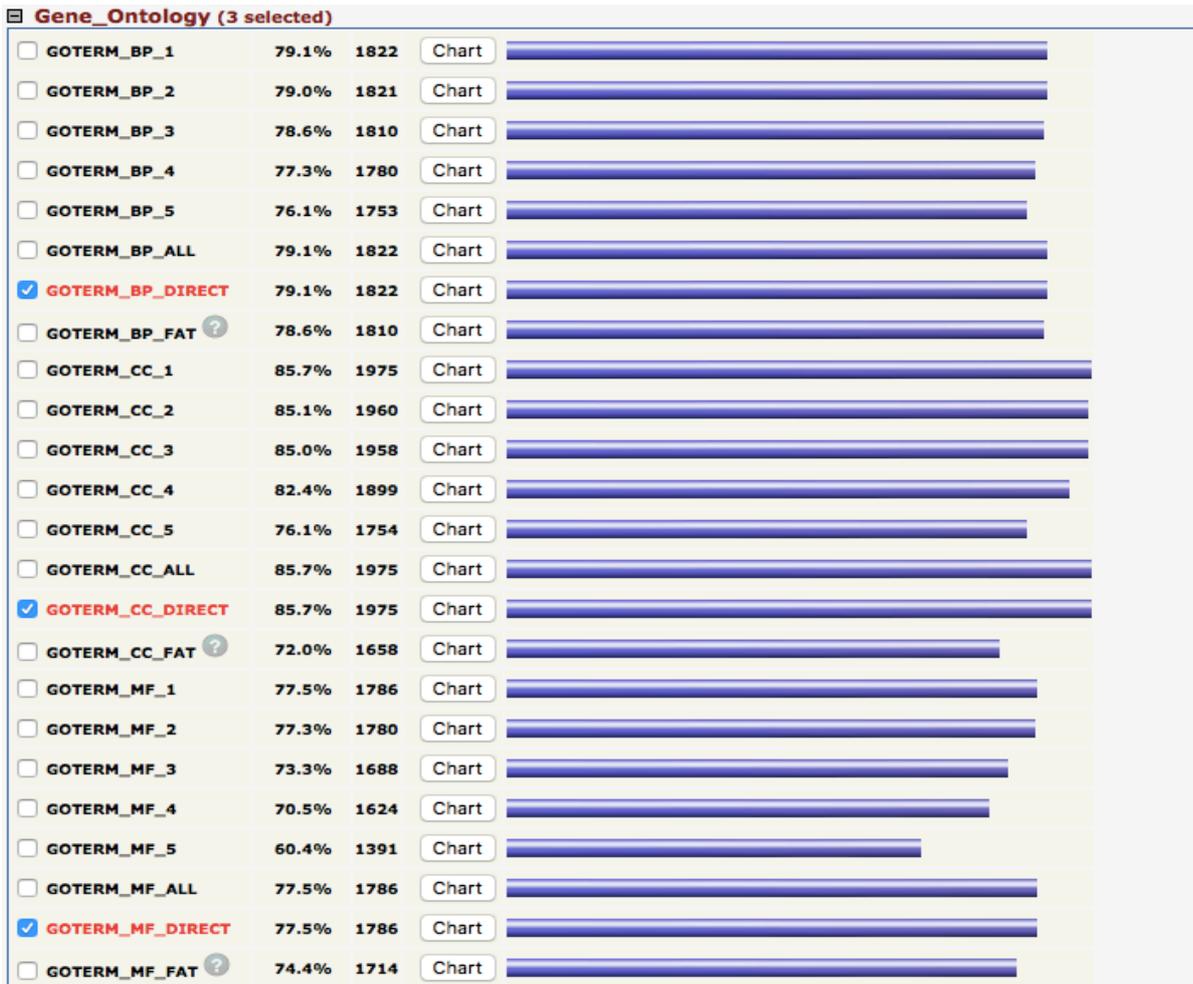


Figure 4:Functional Annotation Tool

Figure 4 is done with DAVID Bioinformatics Resources 6.8(<https://david.ncifcrf.gov/home.jsp>). The input gene is the overlapping gene from figure 1, First click on the left side button- **Functional Annotation** , Then enter the overlapping part of the gene in Figure 1 , Step 2: Select Identifier , select > OFFICIAL GENE SYMBOL , Step 3: List Type, select Gene List, Step 4: Submit List. Then I get Annotation Summary Results. Then select the species Mus musculus (2304 genes). One of the results is as above.

GO annotation analysis (gene ontology analysis):

categorize each gene according to three large criteria (the function of the gene, the process of metabolism, and the localization of the gene product)

According to the selected differential gene, the super-geometric distribution of these differential genes is calculated for GO (GO), and the GO analysis returns a p-value for each GO that exists in the GO classification. The value indicates that the differential gene is enriched in the GO. GO analysis of the experimental results suggest that the difference between the GO analysis of genes can be found in the enrichment of differentiated gene GO classification entries, to find different samples of different genes may be related to what gene function changes.

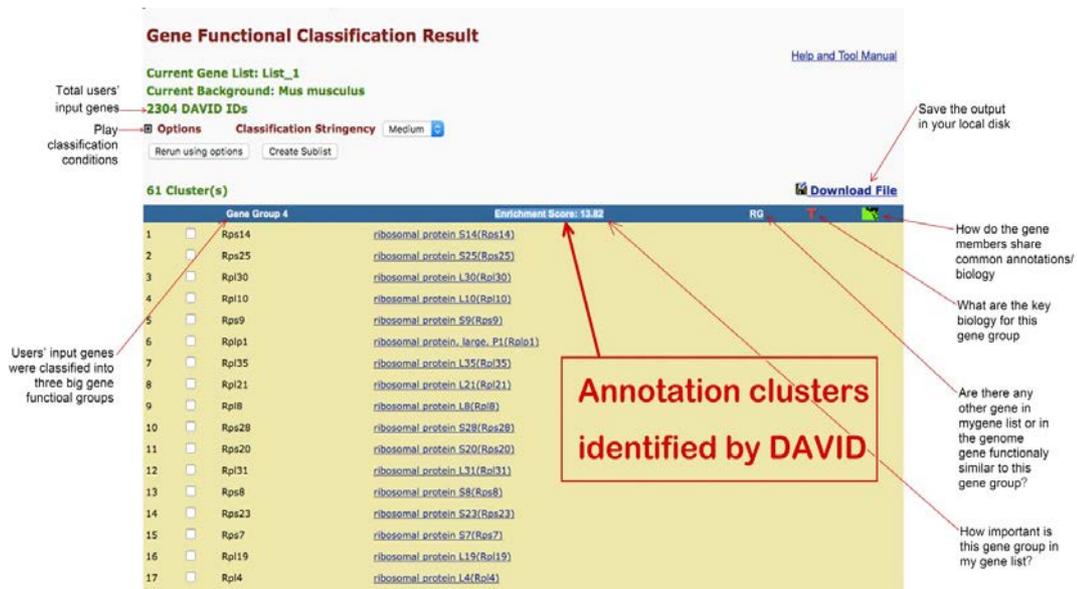


Figure 5: gene functional classification

Figure 5 is done with DAVID Bioinformatics Resource6.8(<https://david.ncifcrf.gov/home.jsp>). The input gene is the overlapping gene from figure 1. First click on the left side button > **Gene Functional Classification**, Then enter the overlapping part of the gene in Figure 1 , Step 2: Select Identifier , select > OFFICIAL GENE SYMBOL , Step 3: List Type, select Gene List, Step 4: Submit List. Then select the species Mus musculus (2304 genes).Terms are divided into many groups and given the scores of the clusters. The score of this term is13.82. The higher the score, the more important the gene representing the group is in the gene list. The gene ID in the list is used as a clustering object, and the function-related gene is displayed

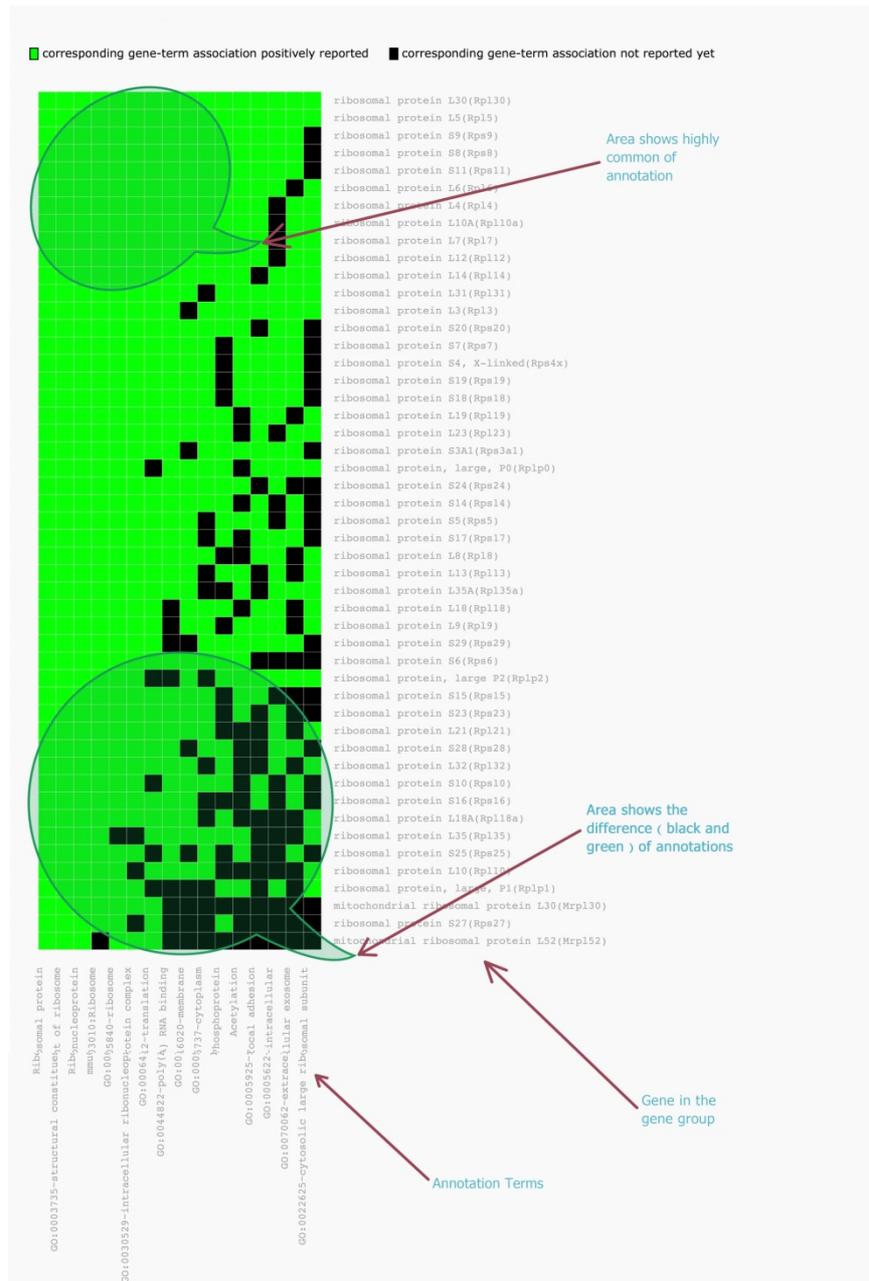


Figure 6: 2D view shows the gene-term relationship

Figure 6 is done with DAVID Bioinformatics Resources 6.8(<https://david.ncifcrf.gov/home.jsp>). The input gene is the overlapping gene from figure 1. First click on the left side button > **Gene Functional Classification**, Then enter the overlapping part of the gene in Figure 1 , Step 2: Select Identifier , select > OFFICIAL GENE SYMBOL , Step 3: List Type, select Gene List, Step 4: Submit List. Then select the species Mus musculus(2304 genes)Then I get 2D View Results.Gene Group 4(Enrichment Score: 13.82) is as above, which shows the relationship between the genes clustered into the same group and the relationship between the gene-term in the form of heat map.

the green part stands for corresponding gene-term association positively reported
the black part stands for corresponding gene-term association not reported yet

The abscissa of the graph represents the specific division of the function, the vertical expression of the specific gene, the green part of the gene that has the following functions, such as Acetylation (acetylation) function of the gene "ribosomal protein, large, p1 (Rp1p1) "

3. Conclusion

The site Venny (<http://bioinfogp.cnb.csic.es/tools/venny/>) can list the gene expression, shows which are the co-expression genes, which are differential gene express clearly. The advantage is given some statistical parameters, the disadvantage is that you can not use the file upload data, the amount of data will be a big problem. DAVID is a bioinformatics database , Integrating biological data and analysis tools, Providing comprehensive biometric annotation information for large-scale gene or protein lists (list of thousands of genes or protein IDs) to help users extract biological information from them. David is doing clustering analysis, can be used to compare the correlation between the two genes list. Can achieve the function that gathering function-related genes as a unit to study these genomes in the biological network level. Score the clustering results, The higher the score, the more important the genes on the gene list in the group. As with other similar analytical tools, the genes in the input list are linked to biological annotations, and from the statistical level, the most significant enrichment of biological annotations is found in thousands of associated genes. The most important feature is the annotation and information link.

Gene chip is a traditional method for detecting differences in gene expression. There are many advantages to digital gene expression profiles and chip ratios based on high throughput sequencing. First the signal is digitized, and the chip signal is an analog signal

In addition, DGE no need to repeat the test, the results are very stable, the chip will need to repeat three times.

DGE can also detect. These are the chip can not do.

I have introduce differential expression analysis based on RNA-seq data in this paper,. We divide this workflow into three parts: (1) differential analysis, (2) visualization and (3) functional annotation. We not only describe the theoretical meaning of each step, but also a few useful tools in order to follow our guidance readers can cross the gap between theories and practice.

Data analysis is based on three major analysis steps^[2-3] that require specific health Material information tools to achieve:

The tools used in this article, Venny , DAVID, are well known and credible. They are used to manipulate and analyze RNA-seq data and have stable performance in many different circumstances. There are many other tools that all have different characteristics , which can Carding data we input to get convincible Well organized point.

Finally, differential expression analysis is only one of the most important applications of RNA-seq data. RNA-seq data, due to the high quality and quantity of information it carries, can be used in many different research directions. Many instructions and tools are available to biologists using their RNA-seq data. RNA-seq more fully reveals the biological individuals at specific times and specific tissue global gene expression^[5-9], such as the discovery of new transcripts.^[10]

Gene chip is a traditional method for detecting differences in gene expression. There are many advantages to digital gene expression profiles and chip ratios based on high throughput sequencing. First the signal is digitized, and the chip signal is an analog signal.

4. References

- [1]Ann-Jay Tong, Xin Liu, Brandon J. Thomas, Michelle M. Lissner, Mairead R. Baker, Madhavi D. Senagolage, Amanda L. Allred, Grant D. Barish, and Stephen T. Smale A Stringent Systems Approach [1]Uncovers Gene-Specific Mechanisms Regulating Inflammation(2015) Cell Biology 3/187 Q1
- [2] OliverGR, HartSN, Klee EW. Bioinformatics for clinical next generation sequencing[J]. Clin Chem, 2015, 61(1):124—135.
- [3] Gargis AS, Kalman L, Bick DP, et al. Good laboratory practice for clinical next-generation sequencing informatics pipeline[J]. Nat Biotechnol, 2015, 33(7):689-693

- [4] Li Xiaobai, Xiang Lin, Luo Jie, Hu Biaolin, Tian Shengping, Xie Ming, Sun Chongbo The Strategy of RNA-seq, Application and Development of Molecular Marker Derived from RNA-seq (2013) Chinese Journal of Cell Biology ,35(5)
- [5] Barakat A, DiLoreto DS, Zhang Y, Smith C, Baier K, Powell WA, et al. Comparison of the transcriptomes of American chestnut (*Castanea dentata*) and Chinese chestnut (*Castanea mollissima*) in response to the chestnut blight infection. *BMC Plant Biol* 2009; 9: 51.
- [6] Dassanayake M, Haas JS, Bohnert HJ, Cheeseman JM. Shedding light on an extremophile lifestyle through transcriptomics. *New Phytol* 2009; 183(3): 764-75.
- [7] Wang W, Wang Y, Zhang Q, Qi Y, Guo D. Global characterization of *Artemisia annua* glandular trichome transcriptome using 454 pyrosequencing. *BMC Genomics* 2009; 10(1): 465.
- [8] Novaes E, Drost DR, Farmerie WG, Pappas GJ, Grattapaglia D, Sederoff RR, et al. High-throughput gene and SNP discovery in *Eucalyptus grandis*, an uncharacterized genome. 2008 30;9:312.
- [9] Alagna F, D'Agostino N, Torchia L, Servili M, Rao R, Pietrella M, et al. Comparative 454 pyrosequencing of transcripts from two olive genotypes during fruit development. *BMC Genomics* 2009; 10: 399.
- [10] Denoeud F, Aury JM, Da Silva C, Noel B, Rogier O, Delledonne M, et al. Annotating genomes with massive-scale RNA sequencing. *Genome Biol* 2008; 9(12): R175.
- [11] Mortazavi, A., Williams, B.A., McCue, K., Schaeffer, L. & Wold, B. Mapping and quantifying mammalian transcriptomes by RNA-seq. *Nat. Methods* 5, 621–628 (2008).
- [12] Cloonan, N. et al. Stem cell transcriptome profiling via massive-scale mRNA sequencing. *Nat. Methods* 5, 613–619 (2008).