# AN *INSILICO* ANALYSIS OF DIFFERENTIAL GENE EXPRESSION IN TUBERCULOSIS MICROARRAY DATASET

[1]Jeyabaskar Suganya, [*1]Mahendran Radha, [2]Sharanya M, [3]Hemamalini
[1]Assistant Professor, [*1]Professor, [3]Assistant Professor, [4]Student
[1]Department of Bioinformatics,
[1]School of Life Sciences, VISTAS, Pallavaram, Chennai-600117, Tamil Nadu, India.

**ABSTRACT:**

Tuberculosis (TB) is a chronic infectious disease that is transmitted by cough-propelled droplets that carry the etiologic bacterium, *Mycobacterium tuberculosis*. The capreomycin is a polypeptide antibiotic used as the second-line standard drug for tuberculosis. In recent years, prescription of the drug increased in clinical sides but the therapeutic efficiency of the drug is decreased gradually. The drug-resistant can occurs when genes of *Mycobacterium tuberculosis* become resistant to the prescribed drugs which are used to treat TB and now these drugs will longer inhibit the gene function of the *Mycobacterium tuberculosis*. To understand the reason behind the TB resistance towards drug capreomycin, an *in silico* research were carried out. The GEO (Gene Expression Omnibus) is a public repository that archives the high-throughput gene expression data originated from scientific experiments. The GEO2R, a Bioinformatics tool to compare two or more set of sample in a GEO Datasets, in order the identify genes which exhibit the drug resistance activity towards drug. In the current study the tuberculosis micro-array datasets with the standard drug Capreomycin were analyzed using *in silico* analysis. The database predicted the genes which showed the down-regulations towards the standard drug. The result of the current study revealed that the genes which is responsible for the drug resistance activity towards the drug capreomycin. This study might provide novel clues in designing of anti-TBdrug which inhibits the function of the above genes.

**Keywords:** Tuberculosis, Capreomycin, GEO Datasets, GEO2R.

## 1. INTRODUCTION:

Tuberculosis can affect any age people in all parts of the world especially to the women in the age of 15 – 44 [1]. The World Health Organization (WHO) evaluated that, in a year 9 million people were affected by TB and in that 3 million TB peoples dies due to the health systems [2]. TB is caused by airborne pathogen that can spread infection through the air from one person to another. The WHO estimated that about 1/3 of the world population is supposed to have latent (an inactive state) TB and about 10% of latent TB can be transformed into active TB. The 2/3 of TB person dies due to the improper treatment and drug resistance developed by the pathogen [3-5].

Tuberculosis is triggered mainly by group of closely related bacterial species of *Mycobacterium tuberculosis* complex. The complex of *Mycobacterium tuberculosis* consists of 7 species: M. africanum, M. bovis, M. canettii, M.caprae, M. microti, M. pinnipedii, M. tuberculosis [6]. Among the seven bacteria species: TB infection caused by M. africanum, M. canettii and M.caprae is very rare, M. bovis mainly affect the animals and it affect only 6% of humans, M. microti cause TB mainly to the pet animals (dogs and cats), M. pinnipedii infects TB to seals and M. tuberculosis was one of the main TB causative bacterium in Human. On analyzing further it was exposed that 90% of human TB was caused by M. tuberculosis, which spread TB infection via aerosolized nuclei from persons of TB disease [7, 8]. The WHO reported that the TB infected person could infect at least 10-15 other person in a year through close contacts [9].

The recent report on tuberculosis revealed that the during last ten decades the death rate of TB was increased drastically than HIV infected patience [10]. On analyzing the reason behind the cause of death, it was found that the stains of TB bacterium developed drug resistance towards all types of TB drugs [11]. When the drugs are mismanaged, there was possibility for development of drug resistances (ability of survival towards the drug) by the stains of Tb bacterium [12]. Drug-resistant tuberculosis is a form of TB infection caused by *Mycobacterium tuberculosis*, that is the genes present in the bacterium developed resistance capacity towards first line and second line tuberculosis drugs [13, 14]. The TBdrug capreomycin was one of the important second line drugs and now a day the bacterium developed the resistance towards drug capreomycin [15]. If the micro-organism developed resistance towards one drug within a same group of drug, there was the possibility of development of resistance towards the drug which belongs to the same class [16]. So there was an urgent need for the development of the novel TB drug towards *Mycobacterium tuberculosis* [17].

Every nucleus contains complete genome i.e DNA'S of all differentiated cells are identical [18]. The introns present in the genes do not code for the protein and they retain the potential function of the genes [19]. The small percentage of the genome is encoded into protein; during that process RNA synthesis is unique for every cell [20]. As the result, even the same genes can produce proteins with different functions. The special components of a cell provides a special characteristics, these components are either synthesized by protein or are by themselves [21]. The expressions of different subsets of genes produce different subsets of gene product [22]. Differential gene expressions (DEGs) refer to the function of different genes within a cell without the loss genetic informations.

For the advance research, the numerous computational approaches can be employed to evaluate level of gene expression in the sample [23]. *In silico* databases, tools and softwares are used to identify DEGs based on the statistical analysis of Pvalue, adj.p.value, log FC value and FDR (false discovery rate) value [24]. Some of the *in silico* tools like GEO2R, IDEG6, Netgestalt, BRBarray were commonly used tools to predict the DEGs from the microarray datasets [25].

The main objective and focus of the current study is to identify the genes which enhances the tuberculosis even after the treatment of the standard drug Capreomycin using Bioinformatics GEO2R analyzer tool.

## 2. MATERIALS AND METHODS:

**Geo profiles**:

The GEO Profiles database stores gene expression profiles derived from curate GEO Data Sets. Each Profile is presented as a chart that displays the expression level of one gene across all samples within a data set. Experimental context is provided in the bars along the bottom of the charts making it possible to understand whether a gene is differentially expressed across different experimental conditions. Profiles have various types of links including internal links that connect genes that exhibit similar behavior and external links to relevant records in other NCBI databases.GEO Profiles can be searched using many different attributes including keywords, gene symbols, gene names, Gen Bank accession numbers, or Profiles flagged as being differentially expressed.

**GEO Datasets:**

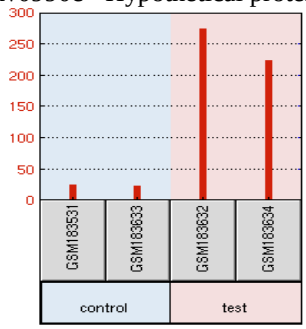
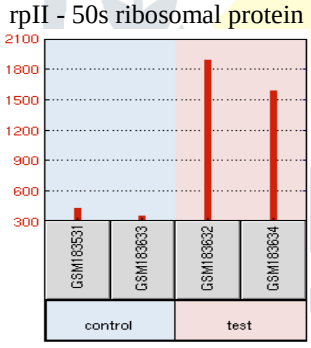
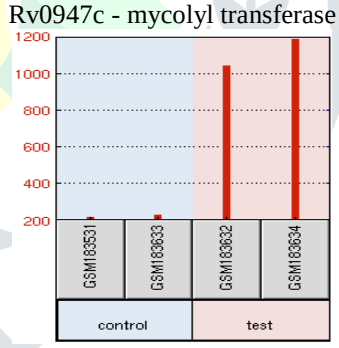
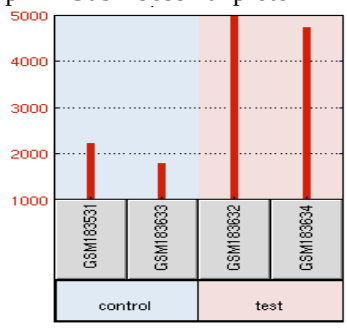
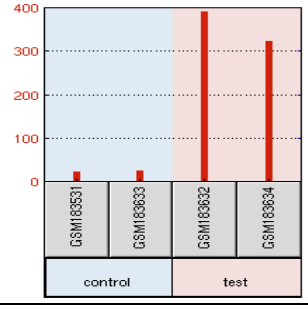
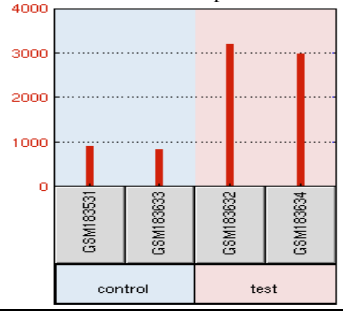
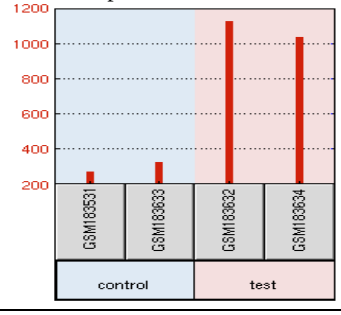
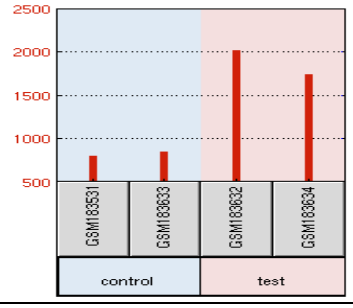
The Gene Expression Omnibus (GEO) datasets contain microarray samples, next-generation sequencing and other forms of high-throughput functional genomic data. Approximately 90% of the data in GEO are from gene expression studies that investigate a broad range of biological properties including disease, development, ecology, evolution, immunity, toxicology and metabolism. The database stores curate gene expression dataset and original series & platform records in the Gene Expression Omnibus (GEO) repository. GEO dataset records contain additional database including cluster tools and differential gene expression (DEG) queries.

**GEO2R:**

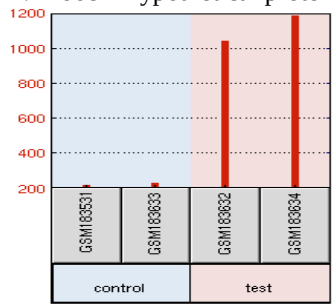
GEO2R database compare two or more groups of microarray samples in a GEO Series in order to identify genes that are differentially expressed across experimental conditions. Results predicted by the database are presented in form of table and genes order in the table based its significant properties. In GEO2R database, there are 5 steps to be followed for analyzing the data [26]. They are as follows:

1. Selection of experiment from the GEO Profile (Capreomycin effect on *Mycobacterium tuberculosis*).
2. Define sample groups (GSE7588)
3. Assign samples to groups (Test and Control)
4. Perform the GEO2R tests
5. Interpret the results table

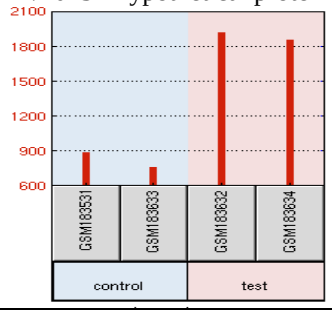
## 3. RESULTS AND DISSCUSSION:

| Gene Symbol , Description of gene and its Profile Graph | Gene Symbol , Description of gene and its Profile Graph | Gene Symbol , Description of gene and its Profile Graph | Gene Symbol , Description of gene and its Profile Graph |
|---|---|---|---|
| Rv0990c - Hypothetical protein   | rpII - 50s ribosomal protein   | Rv0947c - mycolyl transferase   | rplM - 50S ribosomal protein L13   |
| Rv2824c - CRISPR-associated endoribonuclease Cas6   | rpsR2/rpsR1 - 30S ribosomal protein S18/30S ribosomal protein S18   | rim - 16S rRNA processing protein rimM   | snzP - pyridoxine biosynthesis prote   |

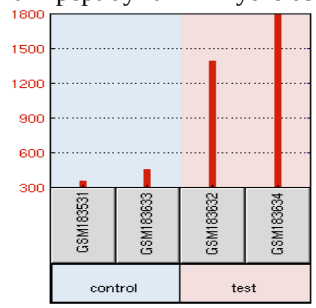
Rv2466c - Hypothetical protein



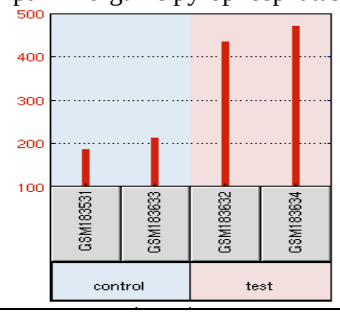

Rv1073 - hypothetical protein



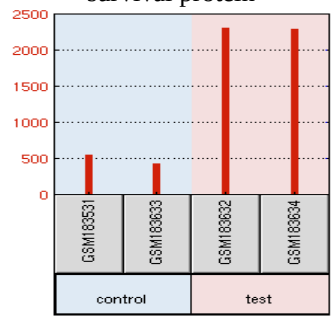

Pth - peptidyl-tRNA hydrolase



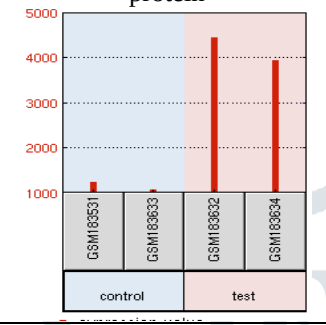

Ppa – Inorganic pyrophosphatase



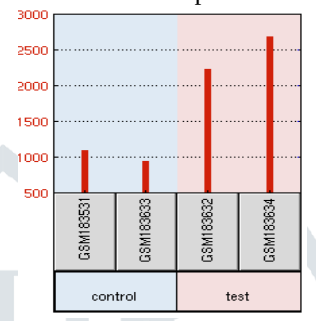

Eis - enhanced intracellular survival protein



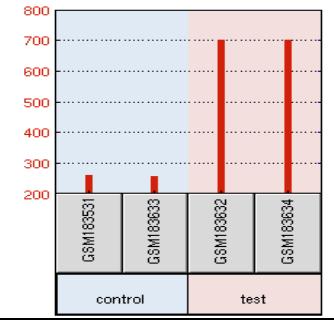

Ssb - single-strand DNA-binding protein



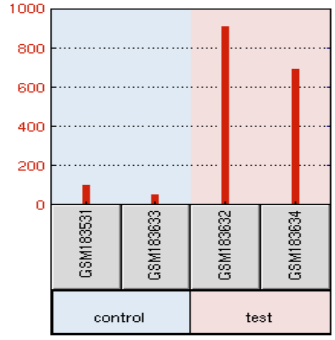

sucD - succinyl-CoA ligase subunit alpha



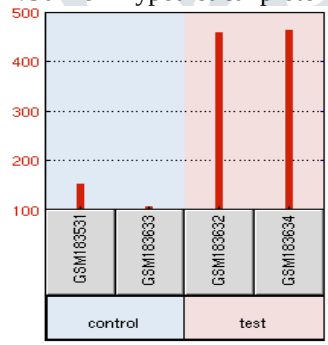

rplL - 50S ribosomal protein L7/L12



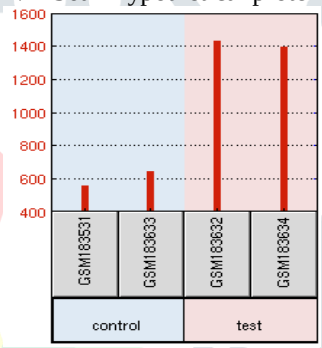

Rv0057 - Hypothetical protein



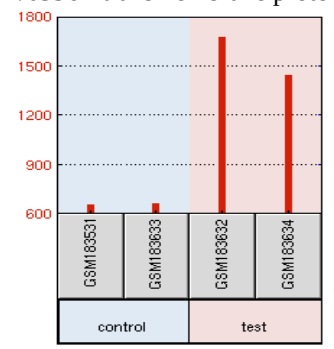

Rv3047c - Hypothetical protein



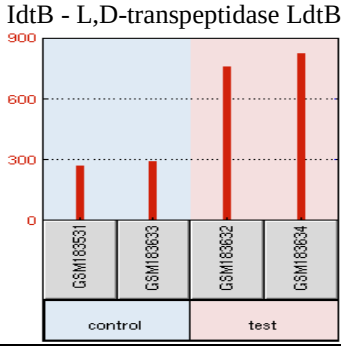

Rv2256a - hypothetical protein



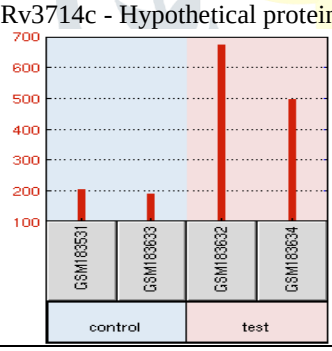

Rv0996 - transmembrane protein



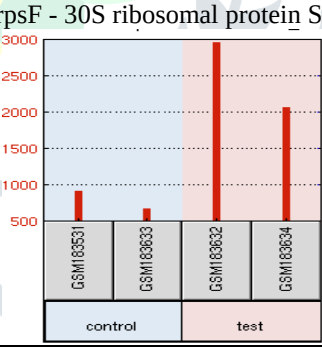

IdtB - L,D-transpeptidase LdtB



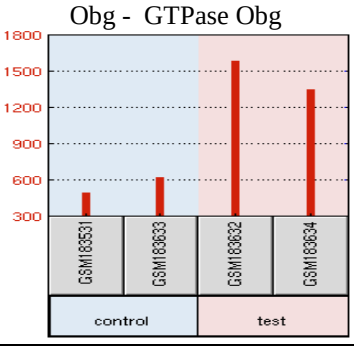

Rv3714c - Hypothetical protein



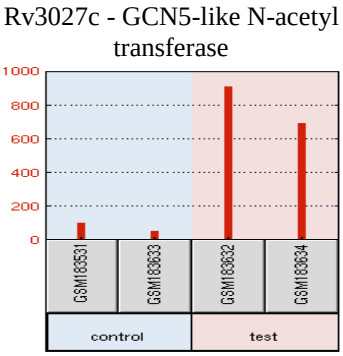

rpsF - 30S ribosomal protein S6



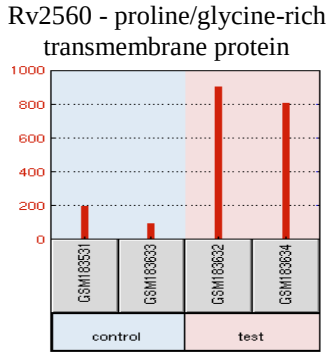

Obg - GTPase Obg



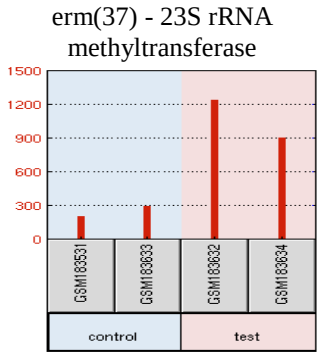

Rv3027c - GCN5-like N-acetyl transferase



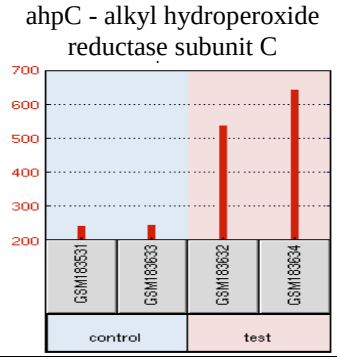

Rv2560 - proline/glycine-rich transmembrane protein




erm(37) - 23S rRNA methyltransferase




ahpC - alkyl hydroperoxide reductase subunit C




The GEO2R tool predicted DEGs (differentially expressed genes) profile graphs for the dataset GS7588 which consist tuberculosis patient samples treated by the drug Capreomycin. On further analyzing the profile graphs, it was identified that the 206 tuberculosis genes showed down-regulation towards drug and other 44 genes (1.Rv0990c, 2.rpII, 3.Rv0947c, 4.rplM, 5.Rv2824c, 6.RpsR2/rpsR1, 7.rim, 8.snzP, 9.Rv2466c 10.Rv1073, 11.Pth, 12.Ppa, 13.Eis, 14.Ssb, 15.sucD, 16.rplL, 17.Rv0057, 18.Rv3047c, 19.Rv2256a, 20.Rv0996, 21.IdtB, 22.Rv3714c, 23.rpsF, 24.Obg, 25. Rv3027c, 26.Rv2560, 27.Erm (37), 28. ahpC, 29. LppK, 30.Rv0147, 31. Rv2908c, 32. LppU, 33. recR, 34.Rv0654, 35.cysA2, 36.hadA, 37.vapC25, 38.vapB33, 39. rplJ, 40. rpIN, 41. rpIY, 42. galE1, 43. rpsP and 44. htpG) showed up-regulation towards drug Capreomycin. From the profile graph, it was clearly revealed that the drug Capreomycin does not suppress the functions of above 44 genes which were responsible for tuberculosis. The results of current study could be useful for the identification and designing of novel anti-tuberculosis drug which could suppress the function of the 44 genes.

**4. CONCLUSION**:

Tuberculosis dataset of GSE7588 were analyzed by GEO2R in order to identify genes differentially expressed across studied samples. Results were ordered by the significance of the gene towards tuberculosis. Further comparison study on test and control, showed that the 44 genes responsible for tuberculosis showed up-regulation even after the ingestion of standard drug Capreomycin which confirms that the standard drug does not showed any inhibitory activity towards those genes and rather enhances the genes activity. The current *insilco* research would lead to the development of powerful novel anti-tuberculosis drugs.

**5. CONFLICT OF INTEREST:**

The authors declare they have no competing interests.

**6. ACKNOWLEDGEMENT:**

We acknowledge Vels Institute of Science, Technology and Advanced Studies (VISTAS) for providing us with required infrastructure and support system needed.

**REFERENCE:**

**[1]** T. Yoshikawa Thomas Shobita Rajagopalan. 2001. Tuberculosis and Aging: A Global Health Problem. Clinical Infectious Diseases, 33 (7): 1034–1039.

**[2]** Siroka A, Ponce NA, Lönnroth K. 2015. Association between spending on social protection and tuberculosis burden: a global analysis. Lancet Infect Dis, 16(4): 473-9.

**[3]** Ai JW, Ruan QL, Liu QH, Zhang WH. 2016. Updates on the risk factors for latent tuberculosis reactivation and their managements. Emerg Microbes Infect, 5(2): e10.

**[4]** Ahmad S. 2010. New approaches in the diagnosis and treatment of latent tuberculosis infection. Respir Res, 11(1): 169.

**[5]** Esmail H, Barry CE, Young DB, Wilkinson RJ. 2014. The ongoing challenge of latent tuberculosis. Philos Trans R Soc Lond B Biol Sci, 369(1645): 20130437.

**[6]** Riojas MA, McGough KJ, Rider-Riojas CJ, Rastogi N, Hazbón MH. 2018. Phylogenomic analysis of the species of the Mycobacterium tuberculosis complex demonstrates that Mycobacterium africanum, Mycobacterium bovis, Mycobacterium caprae, Mycobacterium microti and Mycobacterium pinnipedii are later heterotypic synonyms of Mycobacterium tuberculosis. Int J Syst Evol Microbiol, 68(1): 324-332.

**[7]** Delogu G, Sali M and Fadda G. 2013. The biology of mycobacterium tuberculosis infection. Mediterranean journal of hematology and infectious diseases, 5(1): e2013070.

**[8]** Anne-Laure Banuls, Adama Sanou, Nguyen Thi Van Anh and Sylvain Godreuil. 2015. Mycobacterium tuberculosis: ecology and evolution of a human bacterium. Journal of Medical Microbiology, 64: 1261-1269

**[9]** Lin PL and Flynn JL. 2010. Understanding latent tuberculosis: a moving target. J Immunol, 185(1): 15-22.

**[10]** Lee SH. 2016. Tuberculosis Infection and Latent Tuberculosis. Tuberc Respir Dis (Seoul), 79(4): 201-206.

**[11]** Sia IG, Wieland ML. 2011. Current concepts in the management of tuberculosis. Mayo Clin Proc, 86(4): 348-61.

**[12]** Smith T, Wolff KA and Nguyen L. 2013. Molecular biology of drug resistance in Mycobacterium tuberculosis. Curr Top Microbiol Immunol, 374: 53-80.

**[13]** Palomino JC and Martin A. 2014. Drug Resistance Mechanisms in Mycobacterium tuberculosis. Antibiotics (Basel), 3(3): 317-40.

**[14]** Seung KJ, Keshavjee S and Rich ML. 2015. Multidrug-Resistant Tuberculosis and Extensively Drug-Resistant Tuberculosis. Cold Spring Harb Perspect Med, 5(9): a017863.

**[15]** Navisha Dookie Santhuri Rambaran Nesri Padayatchi Sharana Mahomed Kogieleum Naidoo. 2018. Evolution of drug resistance in Mycobacterium tuberculosis: a review on the molecular determinants of resistance and implications for personalized care. Journal of Antimicrobial Chemotherapy, 73(5): 1138-1151.

**[16]** Nikaido H. 2009. Multidrug resistance in bacteria. Annu Rev Biochem, 78:119-46.

**[17]** Jena L and Harinath BC. 2018. Anti-tuberculosis therapy: Urgency for new drugs and integrative approach. Biomed Biotechnol Res J, 2: 16 - 9.

**[18 ]** Paul M. Harrison Deyou Zheng Zhaolei Zhang Nicholas Carriero Mark Gerstein. 2005. Transcribed processed pseudogenes in the human genome: an intermediate form of expressed retrosequence lacking protein-coding ability. Nucleic Acids Research, 33 (8): 2374-2383,

**[19]** Jo BS, Choi SS. 2015. Introns: The Functional Benefits of Introns in Genomes. Genomics Inform, 13(4): 112-8.

**[20]** Arnold J. Berka. 2016. Discovery of RNA splicing and genes in pieces. PNAS, 113 (4): 801-805.

**[21]** Palazzo and Lee. 2005. Non-coding RNA: what is functional and what is junk?. Front. Genet, 6(2): 1-11.

**[22]** Thomas R. Gingeras. 2007. Origin of phenotypes: Genes and transcripts. Genome Res, 17: 682-690

**[23]** Abnizova I, Subhankulova T and Gilks W. 2007. Recent computational approaches to understand gene regulation: mining gene regulation *in silico*. Curr Genomics, 8(2): 79-91.

**[24]** Xiao Y, Hsiao TH and Suresh U. 2012. A novel significance score for gene selection and ranking. Bioinformatics, 30(6): 801-7.

**[25]** Murray D, Doran P, MacMathuna P and Moss AC. *In silico* gene expression analysis-an overview. Mol Cancer, 2007: 6:50.

**[26]** Altara R, Zouein FA, Brandão RD, Bajestani SN, Cataliotti A and Booz GW. 2018. In Silico Analysis of Differential Gene Expression in Three Common Rat Models of Diastolic Dysfunction. Front. Cardiovasc. Med. 5:11.