ISSN: 0974-2115
www.jchps.com Journal of Chemical and Pharmaceutical Sciences

Molecular Understanding of Zellweger Spectrum Disorder (Zsd): Microarray Gene Expression Analysis

Ruchi Yadav*
AMITY Institute of Biotechnology, AMITY University, Uttar Pradesh Lucknow, UP, INDIA
***Corresponding author: E-Mail: ryadav@lko.amity.edu,**

ABSTRACT

Zellweger spectrum disorder (ZSD) is a Human peroxisome biogenesis disorders that are lethal disease. ZSD is a continuum caused due to mutation in peroxins gene (PEX) .Peroxins is essential protein for proper peroxisome assembly and normal functioning. Peroxisome is crucial for proper development and function of central nervous system, liver, adrenal, kidney and other organs. Defects inperoxisome assembly leads to abnormal brain development and causes severe damage to nervous system, dysfunction of many organs for eg. Kidney, liver, adrenal, also loss of vision and hearing. Clinical studies have shown the importance of peroxisome assembly and function but still the molecular basis of disease gene is not fully identified and is under study. In present work we have used microarray gene expression data from GEO database to identify genes that are expressed in disease condition .In this paper we have used t-test, SAM analysis, clustering analysis to identify genes that are involved in Zellweger spectrum disorder (PBD-ZSD).This gene can be used as potential marker for disease diagnosis and also as potential drug target for drug development.

KEYWORDS: Zellweger spectrum disorder, Peroxisome Disorders, Microarray, Gene expression analysis, PEX gene, statistical analysis of microarray

1. INTRODUCTION

Peroxisome Biogenesis Disorders: Peroxisome are small organelles include almost 50 types of different enzymes that are crucial for different metabolic reactions and are involved in different biochemical reactions in different cell type. Peroxisome is a multiprotein complex that are synthesized on ribosome and assembled to make complex peroxisome structure. Proteins that are associated with biogenesis of peroxisome are peroxins proteins that are encoded by PEX genes**.** Different proteins are targeted to core of peroxisome for degradation or proper folding of protein. Protein with signal sequences at carboxy terminal Ser-Lys-Leu or PTS1(Peroxisome targeting sequence-1) or other signal sequence that is combination of nine amino acids PTS2 are targeted into interior of by different receptors that traslocates protein to peroxisome .any mutation in pts1 or pts2 leads to severe childhood diseases. Assembly of peroxisome. Proteins destined for peroxisome are synthesized on free ribosome and imported into preexisting peroxisome as completed polypeptide chains .any mutation in Peroxisome protein leads to abnormal function of peroxisome or loss of function in multiple enzyme mutation it causes human diseases .typical example of peroxisomal disorder is Zellweger syndrome which very fatal in initial years of new born child.

Zellweger Spectrum Disorder: Zellweger spectrum is a peroxisome biogenesis disorders (PBD-ZSD) that is caused due to mutation in different peroxisomal genes or receptor of PTS1 and PTS2. Peroxisome biogenesis disorders are of three phenotypes - Zellweger syndrome (the most severe form), neonatal adrenoleukodystrophy (NALD), and infantile Refsum disease (the least severe). Zellweger spectrum is most severe form of peroxisomal disorders PBDs are caused by mutations of PEX genes that code peroxins protein responsible or peroxisome assembly and functioning.

Zellweger syndrome is type of neurological disorder that involves demyelination, retinal dystrophy, hearing loss and seizures, abnormal neuronal migration, neuronal positioning, and brain development. The most commonly mutated genes are PEX1, PEX2, PEX3, PEX5, PEX6, PEX10, PEX12, PEX13, PEX14, PEX16, PEX19, or PEX26. Due to the mutations the peroxisome machinery of the individual is highly affected. Peroxisomes are required to break down branched fatty acids, plyamines, very long chain fatty acids and D-amino acids. They are also involved in the biosynthesis of plasmanogens and are involved in the reduction of Hydrogen peroxide a reactive oxygen species.

Plasmanogens are the most abundant phospholipids in the myelin. Low or no production of it leads to myelin anomalies. In the same way loss of the capacity of beta-oxidation of VLCFAs leads to their accumulation in myelin and other parts of the CNS leading to severe dysfunctions. ZS patients have a very low life expectancy with only a few surviving the 6 month period. Important clinical findings associated to it are large fontanels, flat supraorbital ridges, Hyperteloirism, Prominent forehead.

There is no proper cure of this disorder and the diagnosis is also dependent on hectic biochemical analysis mainly centered on the level of VLCFAs in the body. This work is aimed to furnish new information in this scenario and identification of non peroxisomal genes which can be used as targets for drugs as well as for diagnosis of disease.

Microarray Study in Zellweger Spectrum: The Microarray technique has lately become one of the most important and significant tool in Biology and particularly in the field of Genomics, Transcriptomics and proteomics. It used to study and analyze differential expression of gene, within an organism as well as to compare the features of two

different organisms even at their genomic level. The most important uses of DNA Microarray data are Gene Expression profiling that is comparative study of expression profile of a diseased individual with a control. The data generated tells about which genes are differentially expressed down regulated in one disease condition or up regulated in a another disease condition.

In GEO database 7 microarray experiment detail on Zellweger spectrum disorder has been deposited that signifies the importance of peroxisomes in the growth and functions of the central nervous system, liver, and other organs. But no experiment has worked on differential gene expression that shows the regulation of genes in control and in patients with disease. Till date the mechanism and understanding of genes that specify the disease is not completely elucidated.

Microarray data is taken in which study has been done on Zellweger spectrum using microarray experiment to identify gene expression biomarkers of hepatocyte-like cells derived from PBD-ZSD patients and healthy controls and data is submitted in GEO database but other analysis like differential gene expression analysis ,co expression analysis, clustering analysis has been not done these analysis and careful study can put more light into genes involved in Zellweger spectrum and can further studied to identify target protein or biomarker.

Differential Gene Expression Analysis: It was discovered that not all the genes of a genome are expressed in all the cells. This leads to different protein production in different cells. Initially it was thought the two conditions would exist. Either the gene is not express or the gene is expressed. This was because the experiment to proof the differential expression of gene was done through simple blotting. Later on with the advent of techniques like Microarrays and SAGE it was observed that even the extent of expression of a gene in two different cells is differential in nature, this observation as possible because of the presence of millions of probes at each spot. It was gradually observed that certain disease conditions prevail due to up regulation or down regulation of certain genes which are involved in the processes that can be attributed to the diseases symptoms. Certain other nonrelated genes were also seen to be differentially regulated in accordance to a disease condition. These differentially expressed genes can be helpful in many ways. The proteins they encode can be potential targets for new drugs. In case of non-related genes the proteins can act as markers and even the gene expression data can also be used in diagnosis.

2. MATERIALS AND METHOD

Data Set: Microarray data for differential gene expression analysis is retrieved from Gene Express Omnibus database GEO ID: GSE69066 (http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=gse69066)

This data is released on May 21, 2015 with Title: Gene expression profiles of hepatocyte-like cells derived from induced pluripotent stem cells (iPSCs) from donors with the Zellweger spectrum of peroxisome biogenesis disorders (PBD-ZSD) and healthy controls (Wang XM, Yik WY, Zhang P, Lu W. Induced pluripotent stem cell models of Zellweger spectrum disorder show impaired peroxisome assembly and cell type-specific lipid abnormalities. *Stem Cell Res Ther* 2015 Aug 29;6:158)

In this microarray experiment total [HG-U133A_2] Affymetrix Human Genome U133A 2.0 Array chip is used and 79 samples were used for gene expression analysis out of which 23 control samples and other PBD-ZSD patients samples. These samples consisted of information of controls, induced pluripotent cells of control, cells with defects of various peroxins coding genes including PEX1, PEX 10, PEX 12, PEX 26. In this study 21 samples were used and grouped according to mutation in gene along with its control samples. Statistical analysis is done to identify differentially expressed gene

Generation of Gene Expression File: RMA express software (http://rmaexpress.bmbolstad.com/) is used to generate gene expression file .rma uses Robust Multichip Average expression method to calculate gene expression file. RMA Express is a standalone GUI program for Windows, OS X and Linux to compute gene expression files for Affymetrix Genechip Following steps were used in RMA software to generate

- The required CDF and CEL files were downloaded from GEO Database
- These files were loaded in RMA express as unprocessed files
- RMA Measures were computed by allowing Background adjust, Quantile Normalization and Median Polish Summarization Method.
- The raw data was visualized and analyzed.
- Gene expression file is written in a log file.

For Analysis of microarray data gene expression files were created using RMA express with combination of PBD-ZSD patient and control samples.

Table.1. Dataset used for gene expression file A and B

S.No	FILE	CDF File	CEL Files
1	Gene Expression file A	HG-U133A_2.cdf	Cont1_iPS1.CEL Cont1_iPS2.CEL Cont1_iPS3.CEL Cont2_iPS1.CEL Cont2_iPS2.CEL Cont2_iPS3.CEL PEX1fs1_iPS1.CEL PEX1fs1_iPS2.CEL PEX1fs2_iPS1.CEL PEX1fs2_iPS2.CEL PEX1ms1_iPS1.CEL PEX1ms1_iPS3.CEL PEX10_iPS1.CEL PEX10_iPS2.CEL PEX12_iPS2.CEL PEX12_iPS3.CEL PEX26_iPS1.CEL PEX26_iPS2.CEL
2	Gene Expression file B	HG-U133A_2.cdf	GSM1691854_Control_1_HLC.CEL Cont1_iPS1.CEL Cont1_iPS2.CEL, GSM1691855_PBD_PEX1ms1_HLC.CEL GSM1691856_PBD_PEX1fs1_HLC.CEL PEX1fs1_iPS1.CEL PEX1fs1_iPS2.CEL PEX1fs2_iPS1.CEL PEX1fs2_iPS2.CEL PEX10_iPS1.CEL PEX10_iPS2.CEL PEX26_iPS1.CEL PEX26_iPS2.CEL

Statistical Analysis of Deferentially Expressed Gene: Microarray studies often aim to identify genes that are differentially regulated across different classes of samples; examples are: finding the genes affected by a treatment, or finding marker genes that discriminate diseased from healthy subjects. Statistical tests, rather than cluster analysis, are the right tool for this purpose. In this study Multi Experiment Viewer (MeV) software is used for t-test, significant analysis of microarray (SAM) and clustering analysis

3. RESULT

T-Test Analysis: The t-test is probably the most commonly used Statistical Data Analysis procedure for hypothesis testing. We have used student t –test between control and patient with disease sample and list of significant genes table is saved .This significant genes are used for further analysis to identify differentially expressed genes. Volcano plot is plotted between significant p values of genes on y-axis that shows the statistical significance of gene and fold change value on x –axis that shows the biological significance of genes .Figure 1 shows the volcano plot of gene expression file A shown in table 1.Microarray dataset containing 22277 genes out of which 207 genes are found significant for further study

T-test on gene expression file B was done between different set of control and patient samples, figure 2 shows the volcano plot of significant genes out of 22277 gene 1390 are found significant and further studied

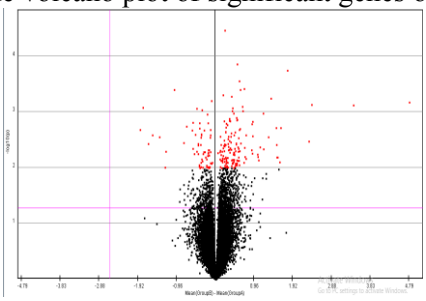

Figure .1.Volcano plot of gene expression file A, showing significant genes under study in red

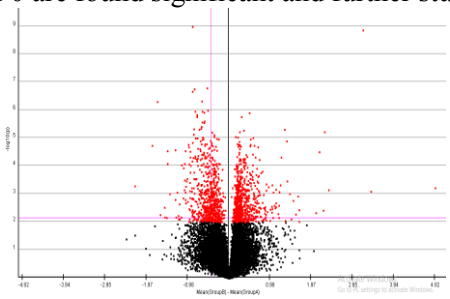

Figure.1. Volcano plot of gene expression file B, showing significant genes under study in red

Significant Analysis of Microarray (SAM): Several statistical methods have been described to interpret gene expression data. Non-parametric methods such as Significance Analysis of Microarrays (SAM) (Tusher, 2001) are widely applicable because they do not rely on explicit assumptions about the error structure. SAM can be used to predict significant genes on the basis of differential gene expression between sets of samples. Two class unpaired test is used to identify positively and negatively significant genes. Figure 3 shows the SAM graph for gene expression file A.

Out of 207 significant genes 104 genes found to positively regulated. Red dots shoes the significant over expression of genes and green dots denote the significant under expression of gene when control samples are compared with patients sample, whereas black points shows genes with no alteration in expression or genes that are not differentially expressed.

These genes were further studied for annotations and biological meaning. Figure 4 shows the SAM graph for gene expression file B, in this analysis out of 1390 significant genes 3 genes are positively regulated while others are negatively regulated.

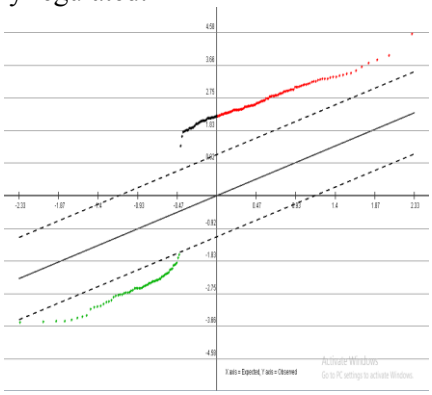

Figure.2.SAM analysis for gene expression file A

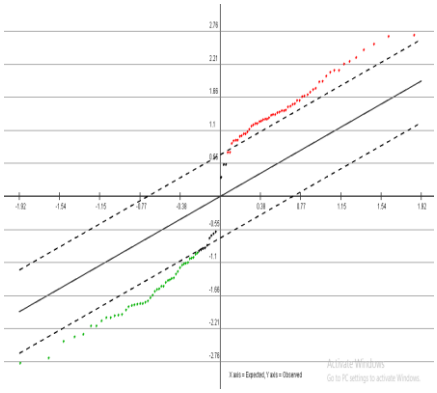

Figure.3. SAM analysis for gene expression file B

Clustering Analysis: Clustering genes on the basis of gene expression profile is helpful to identify genes that are co-expressed or to identify gene functions that are poorly annotated. Clustering is one of the unsupervised approaches to classify data into groups of genes or samples with similar patterns that are characteristic to the group. Clustering of microarray gene expression data is done to identify genes that have differential expression in different samples. Careful investigations of cluster can be done for meaningful biological inference about the set of genes or samples.

Figure.4. Cluster 1 showing expression of genes in different samples.

Figure 5 shows the clustering result of gene expression file, three genes are identified that are differentially expressed in patient sample with mutation in PEX1, PEX26 gene. Affymetrix id 205000_at and 205001_s_at is DDX3Y gene and affymetrix id 201909_at is RPS4Y1 gene molecular function of these genes are later defined.

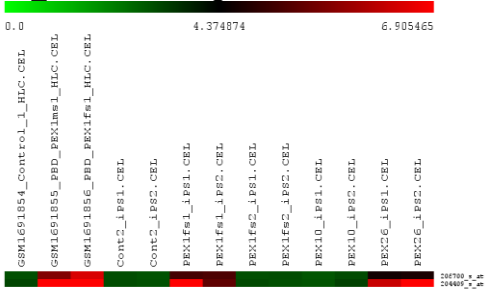

Figure.5. Cluster 2 showing expression of genes in different samples.

Figure 6 shows result of second cluster in which 2 genes are identified that are differentially expressed in patient with mutation in PEX1 and PEX26 gene. Affymetrix id 206700_s_at is KDM5D (lysine (K)-specific demethylase 5D) and 204409_s_at is EIF1AY (eukaryotic translation initiation factor 1A, Y-linked) gene that are upregulated in Zellweger spectrum disorder (ZSD).

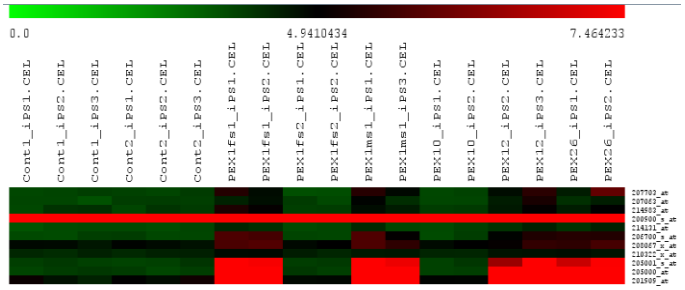

Figure.6 . Cluster 3 showing expression of genes in different samples.

Cluster 3 shows other genes that are differentially expressed in control and patient samples figure 7 shows the cluster 3 result. Genes with affymetrix id 205001_s_at, 205000_at 201909_at shows upregulation in samples with mutation in PEX1, PEX12, PEX26.

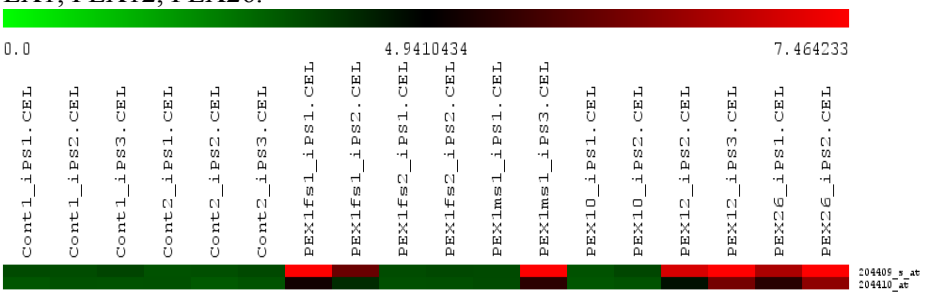

Figure.7. Cluster 4 showing expression of genes in different samples.

Cluster 4 shows the genes that are upregulated in samples with mutation in PEX1, PEX12, PEX26 with affymetrix ids 204409_s_at and 204410_at details of genes that are differentially expressed is defined in Table.2

Table.2.Significant upregulated genes in Zellweger spectrum disorder (ZSD)

Probesets	Gene Symbol	Gene Title	Associated samples
205000_at 205001_s_at	DDX3Y	DEAD (Asp-Glu-Ala-Asp) box helicase 3, Y-linked	PEX1,PEX26
201909_at	RPS4Y1	Ribosomal protein S4, Y-linked 1	PEX1,PEX26
206700_s_at	KDM5D	Lysine (K)-specific demethylase 5D	PEX1,PEX12 ,PEX26
204409_s_at 204410_at	EIF1AY	Eukaryotic translation initiation factor 1A, Y-linked	PEX1,PEX12, PEX26

Genes that are significantly upregulated in patients with Zellweger spectrum disorder (ZSD) having mutation in PEX1, PEX12, and PEX26 are:-

DDX3Y , DEAD-box helicase 3, Y-linked (205000_at, 205001_s_at): This gene is responsible for encoding a protein which is a member of the DEAD-box RNA helicase family, characterized by nine conserved motifs, included the conserved Asp-Glu-Ala-Asp (DEAD) motif. These motifs are thought to be involved in the formation of intermolecular interactions hydrolysis ATP binding, and RNA binding. This protein shares high similarity to DDX3X, on the X chromosome, but a deletion of this gene is not complemented by DDX3X. Mutations in this gene result in male infertility, a reduction in germ cell numbers, and can result in Sertoli-cell only syndrome. Pseudogenes sharing similarity to both this gene and the DDX3X paralogs are found on chromosome 4 and the X chromosome. Alternative splicing results in multiple transcript variants encoding different isoforms

RPS4Y1, ribosomal protein S4 Y-linked 1(201909_at): Cytoplasmic ribosomes, organelles that catalyze protein synthesis, consist of a small 40S subunit and a large 60S subunit. Together these subunits are composed of 4 RNA species and approximately 80 structurally distinct proteins. This gene encodes ribosomal protein S4, a component of the 40S subunit. Ribosomal protein S4 is the only ribosomal protein known to be encoded by more than one gene, namely this gene and ribosomal protein S4, X-linked (RPS4X). The 2 isoforms encoded by these genes are not identical, but are functionally equivalent. Ribosomal protein S4 belongs to the S4E family of ribosomal proteins. It has been suggested that haploinsufficiency of the ribosomal protein S4 genes plays a role in Turner syndrome.

KDM5D, lysine (K)-specific demethylase 5D (206700_s_at): This gene encodes a protein containing zinc finger domains. A short peptide derived from this protein is a minor histocompatibility antigen which can lead to graft rejection of male donor cells in a female recipient. It plays a role in histone code and in spermatogenesis Diseases associated with KDM5D include y chromosome infertility.

EIF1AY, eukaryotic translation initiation factor 1A, Y-linked (204409_s_at, 204410_at): This gene is essential for maximal rate of protein biosynthesis. Enhances ribosome dissociation into subunits and stabilizes the binding of

the initiator Met-tRNA(I) to 40 S ribosomal subunits. This gene is located on the non-recombining region of the Y chromosome. It encodes a protein related to eukaryotic translation initiation factor 1A (EIF1A), which may function in stabilizing the binding of the initiator Met-tRNA to 40S ribosomal subunits. Alternative splicing results in multiple transcript variants

4. CONCLUSION

The proteins encoded by these genes can be used as potential molecular markers for the diagnosis of Zellwegers Syndrome. These proteins can also serve as a marker to know the exact peroxins gene which is mutated. Once the mutated gene is diagnosed the Peroxins which are lacked can be administered artificially and prognosis of disease can be halted. These proteins can also serve as markers to calculate the probability of the mutation of PEX genes to be dominant in the child once these proteins are studied in the parents' blood samples. The greatest difficulty faced by doctors in combating Zellwegers Syndrome is that even though it has an early onset and accelerated and aggravated action, it is already late by the time it is diagnosed. These molecular markers can save a lot of time and can provide very early diagnosis and hence can help in treating the patient with the drugs that would be developed. The other important application of these markers is that even the fetus can be diagnosed with the disease and if time permits such fetuses could be removed.

5. ACKNOWLEDGEMENT

I acknowledge to all those who are directly or indirectly helped me in this work. Further I wish to acknowledge Bioinformatics analysis tools and data bases for conducting this study.

REFERENCES

Bolstad B.M, Irizarry R.A, Astrand M and Speed T.P, A Comparison of Normalization Methods for High Density Oligonucleotide Array Data Based on Bias and Variance. Bioinformatics, 19, 2, 185-193

Brown M.P, Grundy W.N, Lin N, Cristianini, Sugnet C.W, Furey T.S, Ares M and Haussler D, Knowledge-based analysis of microarray gene expression data by using support vector machines. Proceedings of the National Academy of Sciences USA, 97, 2000, 262-267

Burkard T.R, Planyavsky M, Kaupe I, Breitwieser F.P, Buerckstuemmer T, Bennett K.L, Superti-Furga G, Colinge J, Initial characterization of the human central proteome , BMC Syst. Biol, 5, 2011,17-17.

Chu, Gil, Significance Analysis of Microarrays Users Guide and Technical Document, 2001.

Churchill G, Fundamentals of experimental design for cDNA microarrays. Nature Genetics, 2002, 490–495.

Dudoit S, Yang Y.H, Callow M.J and Speed T, Statistical methods for identifying differentially expressed genes in replicated cDNA microarray experiments. Technical report, Statistics Department, University of California, Berkeley, 2000.

Eisen M.B, Spellman P.T, Brown P.O and Botstein D, Cluster analysis and display of genome-wide expression patterns. Proc. Natl. Acad. Sci. USA, 95, 1998, 14863-14868.

Eissenberg J.C, Lee M.G, Schneider J, Ilvarsonn A, Shiekhattar R, Shilatifard A, The trithorax-group gene in Drosophila little imaginal discs encodes a trimethylated histone H3 Lys4 demethylase, Nat. Struct. Mol. Biol. 14, 2007, 344-346

Gueler B, AZFa protein DDX3Y is differentially expressed in human male germ cells during development and in testicular tumours, new evidence for phenotypic plasticity of germ cells, Hum Reprod, 2012.

Iwase S, Lan F, Bayliss P, de la Torre-Ubieta L,Huarte M, Qi H.H, Whetstine J.R, Bonni A, Roberts T.M, The X-linked mental retardation gene SMCX/JARID1C defines a family of histone H3 lysine 4 demethylases, Cell 128, 2007,1077-1088

Kalnine N, Chen X, Rolfs A, Halleck A, Hines L, Eisenstein S, Koundinya M, Raphael J, Moreira D, Kelley T, LaBaer J, Lin Y, Phelan M, Farmer A, Cloning of human full-length CDSs in BD Creator(TM) system donor vector, Submitted , 2004 to the EMBL/GenBank/DDBJ databases

Kathleen Kerr M and Gary A, Churchill, Statistical design and the analysis of gene expression microarray data. Genet. Res, 77, 2001,123–128,

Lahn B.T, Page D.C, Functional coherence of the human Y chromosome, Science 278, 1997, 675-680

Lee M.G, Norman J, Shilatifard A, Shiekhattar R, Physical and functional association of a trimethyl H3K4 demethylase and Ring6a/MBLR, a polycomb-like protein, Cell, 128, 2007,877-887.

Pan W, A comparative review of statistical methods for discovering differentially expressed genes in replicated microarray experiments. Bioinformatics, 18, 2002, 546-554.

Pan W.A, comparative review of statistical methods for discovering differentially expressed genes in replicated microarray experiments. Bioinformatics, 18, 2002, 546-554.

Pepe M, Longton G, Anderson G and Schummer M, Selecting differentially expressed genes from microarray experiments, Biometrics, 2003, 59,133–142.

Saeed AI, Sharov V, White J, Li J, Liang W, Bhagabati N, Braisted J, Klapa M, Currier T, Thiagarajan M, Sturn A, Snuffin M, Rezantsev A, Popov D, Ryltsov A, Kostukovich E, Borisovsky I, Liu Z, Vinsavich A, Trush V, Quackenbush J. TM4, a free, open-source system for microarray data management and analysis. Biotechniques, 34(2), 2003, 374-378.

Sekiguchi T, Iida H, Fukumura J, Nishimoto T, Human DDX3Y, the Y-encoded isoform of RNA helicase DDX3, rescues a hamster temperature-sensitive ET24 mutant cell line with a DDX3X mutation, Exp. Cell Res. 300, 2004, 213-222

Soukas A, Cohen P, Socci N.D and Friedman J.M, Leptin-specific patterns of gene expression in white adipose tissue. Genes Dev, 14, 2000, 963-980.

Steinberg S, Dodt G, Raymond G, Braverman N, Moser A, Moser H ,Peroxisome biogenesis disorders, Biochimica et Biophysica Acta (BBA).Molecular Cell Research 1763(12), 2006,1733–48

Gerhard, Daniela S, The status, quality, and expansion of the NIH full-length cDNA project: the Mammalian Gene Collection (MGC), Genome research, 14(10B), 2004, 2121-2127.

Tusher V.G, Tibshirani R and Chu G, Significance analysis of microarrays applied to the ionizing radiation response. Proceedings of the National Academy of Sciences USA 98, 2001, 5116-5121.

Vawter MP, Gender-specific gene expression in post-mortem human brain, localization to sex chromosomes. Neuropsychopharmacology, 2004.

Wang, Xiao-Ming, Induced pluripotent stem cell models of Zellweger spectrum disorder show impaired peroxisome assembly and cell type-specific lipid abnormalities, Stem cell research & therapy, 6(1), 2015, 1-5.

Welch B.L, The generalization of ‘students’ problem when several different population variances are involved. Biometrika, 34, 1947, 28-35.

Wolfgang Huber, Anja von Heydebreck, Martin Vingron, Analysis of microarray gene expression data. The Berkeley Electronic Press, 2003

Raymond GV, Peroxisomal disorders, Handbook of Neurochemistry and Molecular Neurobiology. Springer US, 2009. 631-670.