Journal of
Theoretical & Computational Science

In Silico Based Analysis of CKD Expressions Data in Correlation with Diabetes Mellitus Unveils Biomarker Gene

Mohd. Murshad Ahmed, Safia Tazyeen, Aftab Alam, Anam Farooqui, Shahnawaz Ali, Md. Zubbair Malik and Romana Ishrat*

Centre for Interdisciplinary Research in Basic Sciences, Jamia Millia Islamia, New Delhi, India

Abstract

Chronic kidney disease (CKD) is becoming an extensive public health problem worldwide. The current anxiety of disease might be due to the change of the underlying pathogenicity. The aim of our study was to provide a detailed analysis of microarray gene expression data of CKD in correlation with diabetes and identification of biomarker genes. Here, Affymetrix expression arrays were used to identify differentially expressed genes in 22 and 69 samples of CKD and diabetes respectively. It further outlines few of the principal biological alterations observed in the CKD state and depicts specific procedures for conducting quality assessment of Affymetrix Gene chip using GEO datasets (GSE70528, GSE11045) and also illustrates quality control packages to remark the visualization for detailed analysis. We identified 912 differentially expressed genes in CKD and 629 in diabetes. From extensive comparison of CKD with diabetes, we found 80 common genes, of which 29 were found up regulated and 51 down. Further, analysis with NCG of these 80 genes, 10 common genes were found involved in various types of cancer. Thus, the results emphasize the importance of these 10 common differentially expressed genes in considering them as biomarkers for three conditions diabetes, CKD and cancer. Our studies have cataloged differentially expressed genes that may play important role in the pathogenesis of CKD and could serve as biomarkers.

Introduction

Chronic kidney disease (CKD) is chronic, non-communicable, age related renal function decline disease that affects people worldwide [1]. The prolonged CKD is a major risk factor of cardiovascular diseases and death [2]. The disease prevalence increased with age and gender specific affecting more females than males. The healthcare burden is highest in early stages due to increased prevalence, affecting around 35% of those over 70 years [3]. The unhealthy lifestyle, diabetes, hypertension, BMI, obesity and smoking are some major factors that accelerate CKD. Since, high blood pressure and hypertension are the major causes of kidney failure as these provide stress on blood vessels throughout the body, including nephrons [4]. Further, increasing cases of diabetes complicates the CKD disease thus making it healthcare priority in both developed and developing countries. CKD symptoms include abnormalities in kidney structure and function. It can be manifested with or without decrease in GFR that is considered to be as the markers of kidney damage [5]. Late appearance of symptoms makes it hazardous, often causing irreparable damages to the kidney requiring expensive treatments and renal transplant [6]. According to Global Burden of Disease report 2015 of ISN (International Society of Nephrology's) [7], it is reported that the CKD was ranked 18th, which is second only to that for HIV and AIDS [8]. The overall prevalence of CKD is 13% to 25% as calculated for the world population [9].

Diabetes accounts for approximately 2% death of patients starting treatment for kidney failure each year, including dialysis and the renal transplantation for both type 1 and 2 [10]. Nearly 44% of dialysis patients have Type 2 diabetes, making it the fastest growing risk factor for CKD. Type 2 diabetes is the number one cause of kidney failure, responsible for more than one of every three new cases [11]. CKD emerged as a global burden requiring attention in understanding the disease dynamics rendering prevention and early detection as a health care priority in both developed and developing countries. Novel and affordable treatment strategies, identification of prognostic biomarkers for progression and treatment response are urgently required for this endemic disorder [12].

High throughput gene expression profiling has emerged as a dynamic research tool that uncovers transcriptional patterns in numerous biological systems to acquire insight into disease dynamics [13]. The development of high-throughput microarray technology, permitting concurrent measurement of changes in expression of multiple genes within the human genome, equips the opportunity for novel insight into disease processes and molecular pathways of biological dysfunction [14]. A vital, yet challenging arena is quality control and visualization of microarray gene expression data. One of the most prominent platforms for microarray analysis is Bioconductor, an open source and open development software project for the analysis and discernment of genomic data [15].

The aim of our study was to provide an unbiased comprehensive catalog of change in gene expression for the human CKD samples. We tried analyzing and visualizing the differentially expressed genes in CKD together with the effect and correlation to the Diabetes type 2. Here, we presented a gene-expression analysis of CKD (22 samples each) from healthy persons and patients. The heuristic comparison of three conditions CKD, Diabetes and Cancer has revealed the occurrence of 10 prominent genes. These genes are required for the basic functioning of the cell like in the rearrangement of the cytoskeleton, activation of immunity and development of tissues and kidney. So, the Up regulated genes (4 in number) directly confer the increase in the conditions like leukemia (increase in WBC) and alter the immunity by hampering

*Corresponding author: Romana Ishrat, Centre for Interdisciplinary Research in Basic Sciences, Jamia Millia Islamia, New Delhi, India; Tel: 9891813458; E-mail: romana05@gmail.com

the T-cell maturation. Similarly Down regulated genes (6 out of 10) disturbs the development of kidney i.e., the formation of new tissues and also disturbs the GFR (effect of the Osmotic pressure) that directly effects normal functioning of the kidney. Thus, this study emphasize the importance of these 10 genes (both Up and Down regulated) as biomarkers for three conditions CKD, diabetes and cancer. We also expect that this work will benefit the researchers working in *in-vitro* to collect evidences against CKD.

Materials and Methods

Microarray data

The framework for microarray analysis is shown in the Figure 1. The gene expression data bearing GEOID: GSE70528, GSE11045 and GSE30122 (contain 69 cel files with diabetes) were retrieved from Gene Expression Omnibus (GEO) (http://www.ncbi.nlm.nih.gov/geo/) [16]. The GSE70528 RNA microarray analysis was performed on 19 patients without diabetes. These patients were categorized into three groups: HT (8 hypertension patients), CKD (7 patient's not on dialysis) and HD group (4 patients Haemodialysis). We extracted three normal CEL files from GSE11045, thus, total 22 CEL files. In addition, we also retrieved gene expression data for capturing the scenario in diabetic patients.

Microarray analysis

The data fetched above was then subjected to R packages given in Table 1. According to the corresponding correlation between the probe and gene from the GPL570 platform HGU 133 Plus 2 (Affymetrix Human Genome U 133 plus 2.0 Array, Affymetrix, Santa Clara, CA, USA) [17]. The probe numbers of the expression profile were converted into the corresponding gene symbols. One gene corresponds with multiple probes, thus each gene has more than one expression value. Therefore, the average value was calculated and selected as the only representative value. A total of 54675 genes were involved in pre-processed data. Background correction and quartile data normalization were performed by the robust multi-array (RMA) average in R affy package (Affymetrix). In pre-processing, convolution background was corrected, missing values were estimated, expression values were log2 transformed and then normalized. Both data cel files pre-process under the MAS5 method [18]. The expression value after summarization is further use for DEG's.
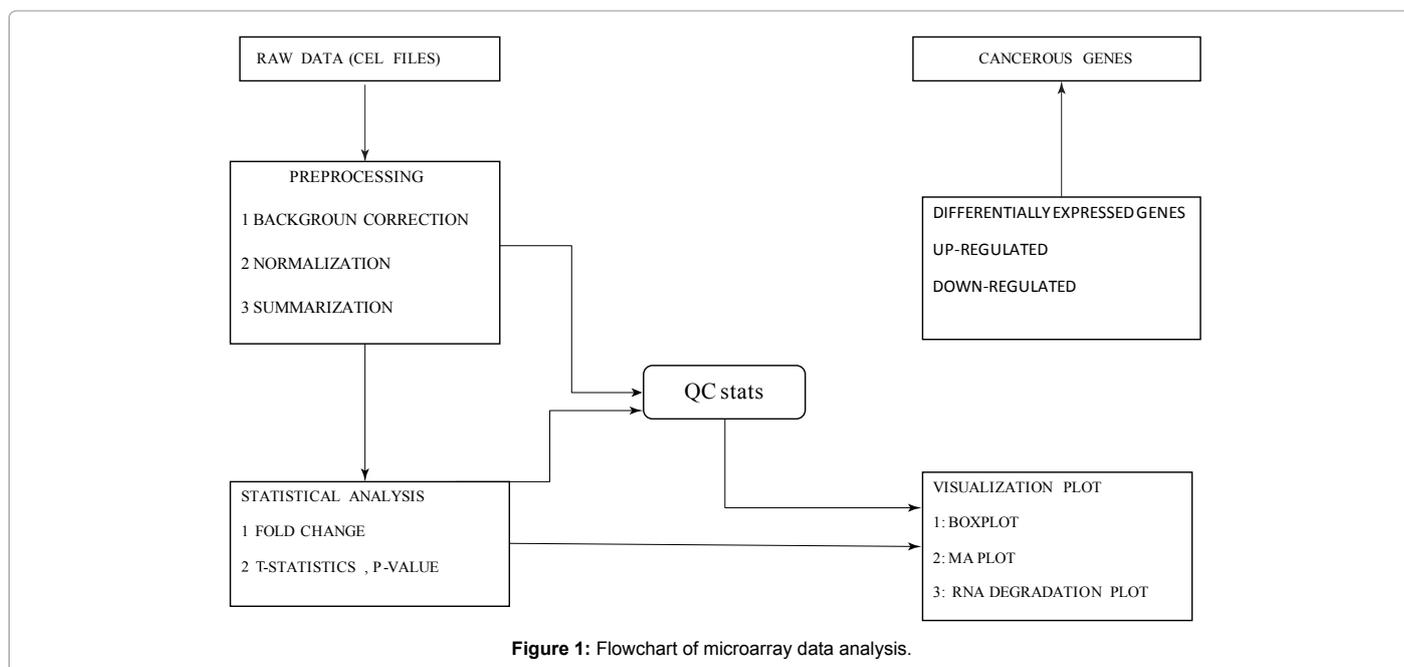

Figure 1: Flowchart of microarray data analysis.

S. No.	Package	Description
1	Affy	Exploratory oligonucleotide array analysis
2	AffyExpress	Quality assessment and to identify differentially expressed genes in the Affymetrix affymetrix chip
3	AffyQCReport	Quality of a set of affymetrix arrays
4	AffyRNADegradation	Assessment and correction of RNA degradation effects in Affymetrix 3' expression arrays
5	ArrayTools	Quality assessment and to detect differentially expressed genes for the Affymetrix GeneChips
6	DEGseq	To identify differential expressed genes from rna-seq data.
7	DEGreport	Report of deg analysis
8	Simpleaffy	High level analysis of Affymetrix data
9	AffyPLM	Quality assessment tools for affymetrix chip
10	Affy comp	Compare expression measures for Affymetrix Oligonucleotide Arrays
11	Derfinderplot	To find the plot for derfind.
12	Affyio	Parsing Affymetrix data files
13	ArrayQualityMetrics	Microarray quality metrics reports
14	AffyPLM	Quality assessment tools for affymetrix chip

Table 1: Analyzing and Visualizing of CKD Data Using Quality Control Packages of Bio-Conductor.

Visualization plots

For visualizing of chips affyqcreport package was downloaded from Bioconductor [19]. Starting from library affy package and readaffy function read the 22 samples from directories celfiles names save as in the same directories having all plot described below. There are number of plots built in quality control packages for visualization and analysis of chips. These plots were also used for comparing chips before and after normalization [20]. Here we describe some plots and their analysis.

Box plot: Boxplot also called as box-and-whisker plot, is statistical plot that graphically plots numerical values for comparison of chips. Five values were plotted in box plot that are lower min value, lower quartile, mean, upper quartile, max value of chip files in parallel for comparison of chip intensities.

Intensity plot: Intensity plot is similar to Boxplot but it gives more detailed view. Intensity plot x-axis represents probe density and y-axis probe intensity.

MA plot: MA plot is used to determine if there any biasness in intensities of red and green dye and other systemic errors or instrumental errors in experiment.

RNA Degradation Plot: RNA degradation plot is used to assess the quality of RNA molecule used as probe in chips. This plot is used for quality analysis of probes spotted on affymetrix chip.

NUSE plot: Normalized Unscaled Standard Errors (NUSE) is used for standard error for each probe on all chips. All probes are normalized to scale one across all arrays.

QC stats plot: QC plot is recommended by affymetrix. Any array that is shown in red is indication of error and in blue indicates the array within limits of scale factors **e** within 3-fold.

Median Absolute Deviation (MAD): The low or high median of the absolute deviations was computed from the median and adjusts by a factor for asymptotically normal consistency.

RLE plot Relative Log Expression: RLE plot is a plot of RLE values that are calculated by assuming that all probe set across arrays against median expression value is constant.

Results and Discussion

This study provides information on the structure of correlation based tuning between genes in multiple microarray data sets. Our main goal was to establish the set of Fundamental genes by comparing analyses across data sets that is relevant in understanding gene function. Thus it was found that many genes showed patterns of correlated expression that are reproducible across data sets, and there is a clear relationship between confirmation of correlated expression and related gene function. From GSE70528 data, total **19 cel** files are selected out of which 8 hypertension data, 7 CKD data and **4 hemodialysis** data chip used in this experiment is **HGU 133 Plus 2**. On the other **GSE11045** from **GEODATASETS3** normal kidney cel files downloaded and taken into cel files folder. Gene expression file is created using affy package of bioconductor and visualization packages are used for quality control analysis and summarizing the output generated by these packages [21].

Visualization plot

Geneexpressionfile created usingaffypackage.

>*library(affy)*

>*cirbs=ReadAffy(changedir ≤ CELFILES)*

>*eset.mas5=mas5(cirbs)*

>*exprSet=exprs(eset.mas5)*

>*exprSet=log(exprSet,2)*

>*write.table(exprSet,file="myab", quote=F, sep="\t")*

We used rmaan another method which create expression file for visualization plots

>*eset=rma(cirbs)*

>*write.exprs(eset,file="cirbs.txt")*

Using these above commands in R, the file containing the expression data can be created and will be used for further analysis using statistics. The object **'exprs'** created in R for expression file using the simple command. This object **exprs** can be used directly in R to visualize expression file and analyze the intensity [22]. Four Affymetrix-specific metrics were evaluated if the input object is an AffyBatch. The RNA degradation plot from the *affy* package, the relative log expression (RLE) boxplots and the normalized unscaled standard error (NUSE) boxplots from the affyPLM package and the QC stat plot from the simple affy package are represented [23]. For visualizing data given by affyQCReport an R package the following set of command are required.

>*Library (QC Report)*

>*QCReport(data, file="cirbs.pdf")*

Boxplot: Each box shows the distribution of expression values of one array. The shape of the box plot for all the arrays looked similar, thus we concluded that there were less systematic biases in the data. The box plot of raw intensities of the data across the 22 chips is depicted in see Figure 2A. The raw intensities differed between HT 1 and HT 2, CKD 6, CKD 7 and HD 4.

Intensity plot: Density curves illustrating the distribution of signal intensities in different arrays shown in Figure 2B. Note the peak to the left which corresponds to the background signal: genes expressed below this level will show up within this area. Above this are the probes that are notably expressed. The number of probes decreases exponentially as the intensity increases, not how the curves all have the same, linear slope. However, the upper tail is shifted to the right or left depending on whether the signal is strong or weak.

MA plot: In MA plot, genes with similar expression levels in two experiments will appear around the horizontal line y=0. In our array every chip indicates outliers see Figure 3A1-3A3. An MA-plot is a plot of log-fold change (*M*-values, that is, the log of the ratio of level counts for each gene between two samples) against the log-average (*A*-values, that is, the average level counts for each gene across the two samples).

NUSE plot: Each box shows NUSE values of the array of Expression files (CEL) that is, $NUSE_i$. In general, value of $NUSE_{i=1}$. In our case array *i=1,6; i=5,7 and i=4* corresponding to HT, CKD and HD respectively are found deviated from the normal. Thus from the NUSE value one can easily identify lower quality arrays. This is much easier than the Affymetrix quality standards to distinguish problematic arrays (Figure 2C).

RNA degradation plot: The degradation plot shows the (shifted and scaled) mean intensity $< RNA_i^s >$ for each position within probe sets. High slopes indicate degradation. At minimal degradation, we can see an upward trend in expression levels as probe number increase. Therefore,
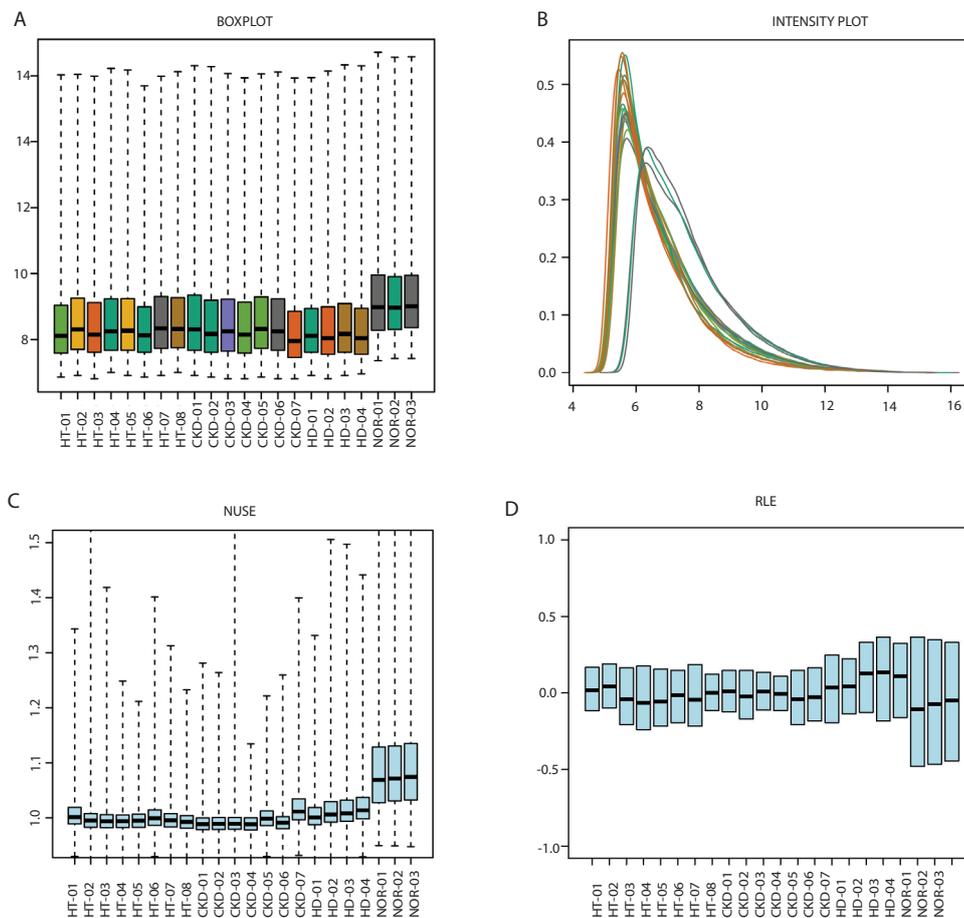
Figure 2: A. Boxplots for arrays GSM1808377.HT-01, GSM1808378.HT-02, GSM1808379.HT-03, GSM1808380.HT-04, GSM1808381.HT-05, GSM1808382. HT-06, GSM1808383.HT-07, SM1808384.HT-08, GSM1808385.CKD-01, GSM1808386.CKD-02, GSM1808387.CKD-03, GSM1808388.CKD-04, GSM1808389. CKD-05, GSM1808390.CKD-06, GSM1808391.CKD07, GSM1808392.HD-01, GSM1808393.HD-02, GSM1808394.HD-03, GSM1808395.HD-04, GSM279060, GSM279061, GSM279062.; **B.** Density estimates (histograms) for arrays GSM1808377.HT-01, GSM1808378.HT-02, GSM1808379.HT-03, GSM1808380.HT-04, SM1808381.HT-05, GSM1808382.HT-06, SM1808383.HT-07, GSM1808384.HT-08, GSM1808385.CKD-01, SM1808386.CKD-02, GSM1808387.CKD-03, SM1808388.CKD-04, GSM1808389.CKD-05, GSM1808390.CKD-06, GSM1808391.CKD-07, SM1808392.HD-01, GSM1808393.HD-02, GSM1808394.HD-03, GSM1808395.HD-04, GSM279060, GSM279061, GSM279062; **C.** NUSE plot.; **D.** RLE plot.

$$RNA_i^s = Avg \text{ intensity}$$

Where S represents index for cell files and i the index for probe set ID see in Figure 4B.

QC stats: Each array is represented by a separate line in the Figure 4A. The central vertical line corresponds to fold change, **FC=0**, the dotted lines on either side correspond to up and down regulation. The blue bar represents the region in which all arrays have scale factors within, by default, three-fold of each other. Its position is found by calculating the mean scale factor for all chips and placing the center of the region such that the borders are **-1.5** fold up or down from the mean value. Each array is plotted as a line from the 0-fold line to the point that corresponds to its scale factor. If the ends of all of the lines are in the blue region, their scale-factors are compatible. The lines are colored blue if OK, red if not (Figure 4A).

RLE plots: The Relative Log Expression (RLE) values were computed by calculating for each probe-set, the ratio between the expression of a probe-set and the median expression of this probe-set across all arrays of the experiment. It is assumed that most probe-sets were not changed across the arrays, so it is expected that these ratios

$RLE \cong 0$ on a log scale. The boxplots presenting the distribution of these log-ratios should then be centered near 0 and have similar spread. Other behavior would be a sign of low quality (see in Figure 2D)

Median Absolute Deviation (MAD): Figure 3B shows a false color display of between arrays distances, computed as the MAD of *M-values* of each pair of arrays.

$$d_{ij} = c \times median\left[X_{mi} - X_{mj} \right]$$

Here, x_{mi} is the normalized intensity value of the mth probe on the ith array, on the original data scale. **c=1:4826** is a constant. Pairwise differences between arrays, computed as the MAD of the differences of the *M-values*.

Differentially Expressed Genes (DEGs) analysis

Total 54675 genes from $GSE70528$ and 22278 genes from $GSE30122$ were filtered at FC=6 and 2 respectively. We got 21 Up-regulated genes on comparing the expression of set 1 (HCH and $GLOMERULI$ diabetes genes), while 8 genes were found for the $set2$ containing HCH and TUBULI diabetes genes. In addition, some down regulated genes were also been extracted and 36 and 32 were found respectively from
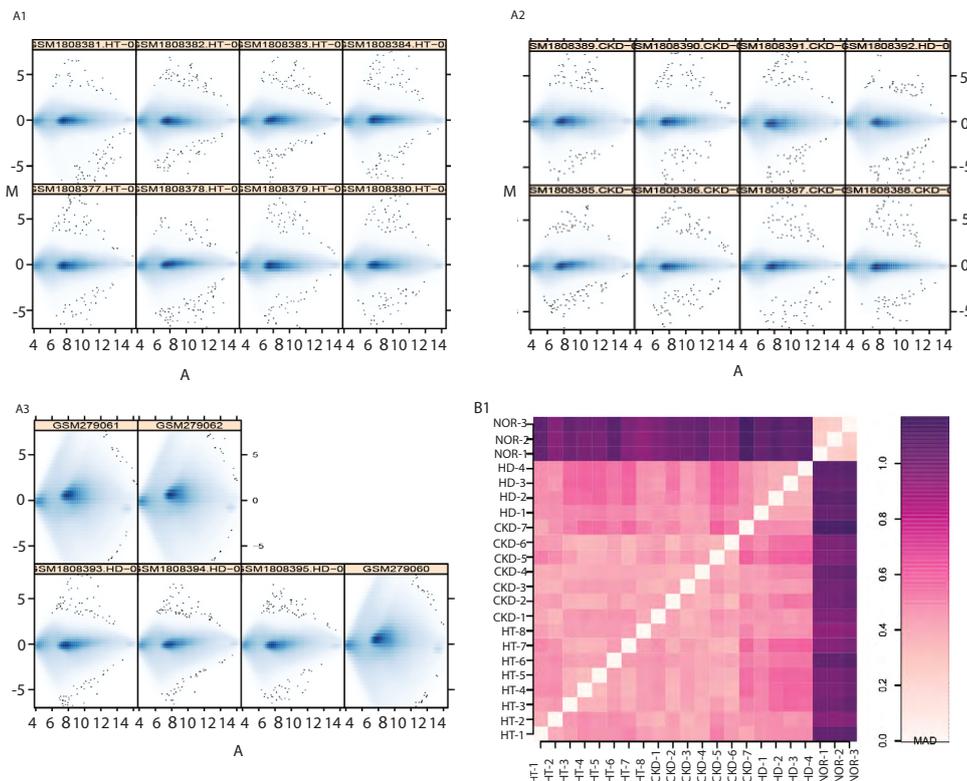
Figure 3: A1, A2 and A3 MA plots: A reference array is calculated from the median across arrays and for each array M and A values are calculated for the comparison to that reference.

GENE 10	PROBE IDs	Fold Change	SD WILD	SD TREAT	T-Value	P-Value	Significance
ALB	211298_s_at	9.657745132	1.810692	0.012285	23.2458	1.82E-16	15.74058207
G6PC	206952_at	6.942372011	1.43397	0.064545	20.96892	1.45E-15	14.83764856
WT1	206067_s_at	7.379295758	1.570643	0.014634	20.47363	2.35E-15	14.6293731
THSD7A	213894_at	6.70450913	1.941436	0.113128	14.89363	1.23E-12	11.90865802
LCK	204891_s_at	-7.85954587	0.500862	0.760188	-17.3237	6.49E-14	13.18754255
IL7R	205798_at	-8.3399772	0.445559	0.982075	-14.4756	2.13E-12	11.67104792
PTPRC	207238_s_at	-6.24239763	0.640783	0.203114	-33.1956	1.24E-19	18.9060705
DOCK2	213160_at	-7.48479112	0.315255	0.777429	-16.4632	1.76E-13	12.75336841
TNFAIP6	206025_s_at	-6.70857688	1.206504	0.860337	-11.7978	9.96E-11	10.0019409
LYL1	210044_s_at	-6.52165747	0.76229	1.092371	-9.96467	2.06E-09	8.685874072

Table 2: Common 10 DEGs Showing T- Statistics, P-Value and Significance.

*set*1and *set*2. So, we found out some of the fundamental genes (that is, the genes showed differential expression in all datasets and are found in cancer too) thus, after comparing all the sets (irrespective for their involvement in cancer) we got 10 cancer genes of which 4 genes were from up-regulated and 6 from down-regulated expressions.

T- Statistics, P-value and significance

Identification of DEGs (differentially expressed gene(s)) between the normal and treated conditions can be formulated in terms of the hypothesis

$H0: \mu Ni - \mu iT = 0 \ v/s \ H1i: \mu Ni - \mu iT \neq 0$

The sampling distribution $\overline{X}_N^i - \overline{X}_T^i$ is used to test the hypothesis H_0^i, where $\overline{X}_T^i$ and $\overline{X}_T^i$ are the means of the a (array) replicates in the control group and a replicates in the treatment group, respectively, and s_{iN}^2 and s_{iT}^2 are the corresponding sample variances.

For independent control and treatment samples, the hypothesis is commonly done by computing

The two tailed t-statistic $(\overline{X}_N^i - \overline{X}_T^i) / \sqrt{\left(\frac{S_{iN}^2}{N}\right) + \left(\frac{S_{iT}^2}{N}\right)}$. As discussed, the intensities measured from the same spot are correlated. In such data one can apply the one-sample t-test for the two-group comparison. If there is no difference between the two groups, then the one-sample t-statistic has a t-distribution with $a-1$ degrees of freedom, where s_{iN}^2 and s_{iT}^2 are the variance of a arrays replicates. From these P-values were calculated for founding the significance of genes (calculated using TDIST command in Excel). The value of these statistics of 10 genes is given in Table 2.

The model for the t-tests presented in this paper performs a gene-by-gene analysis. It computes the sample variance (standard deviation) for each gene in the analysis. Thus, this approach does not require a constant variance or a constant coefficient of variation across genes.
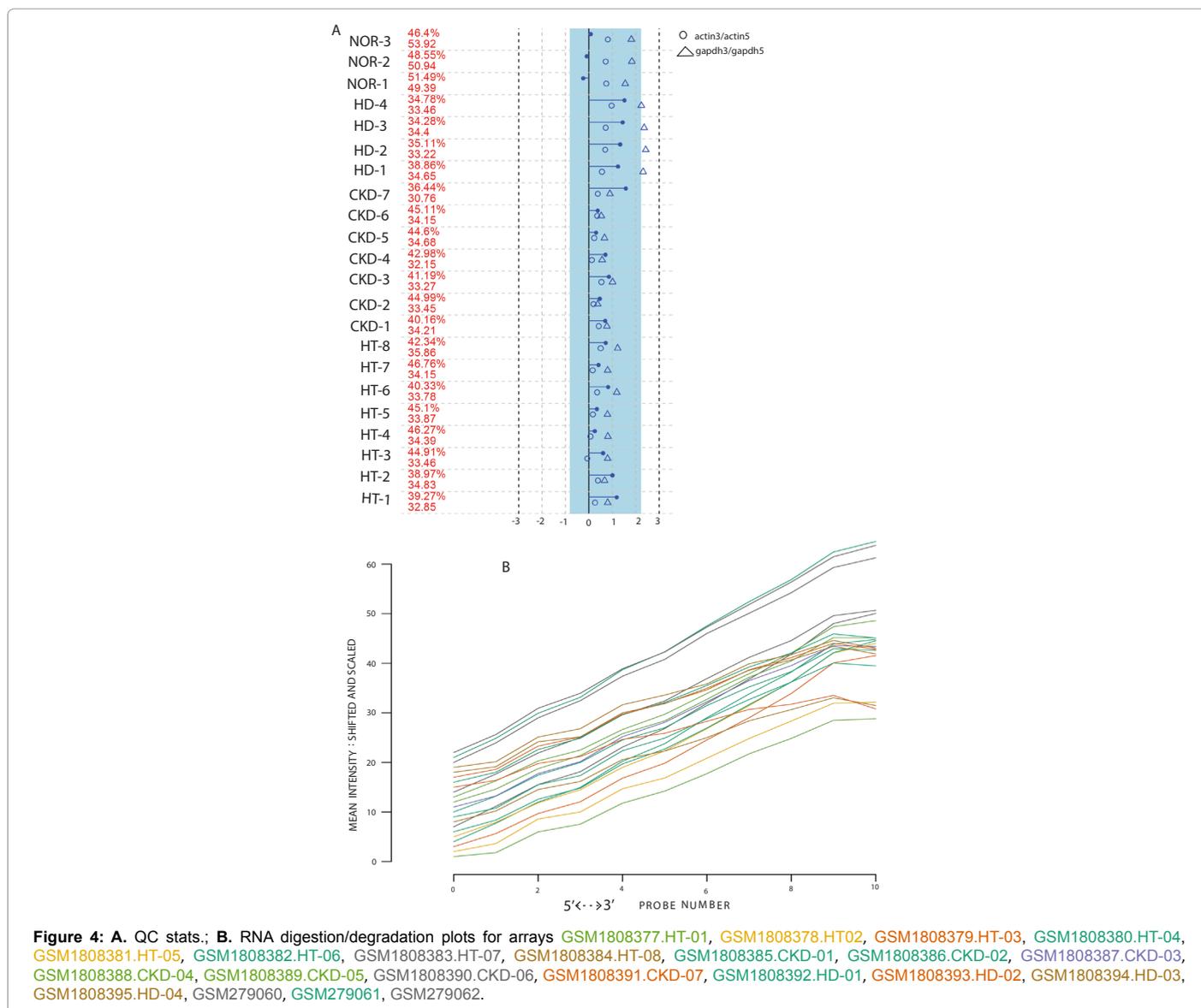
Figure 4: A. QC stats.; **B.** RNA digestion/degradation plots for arrays GSM1808377.HT-01, GSM1808378.HT02, GSM1808379.HT-03, GSM1808380.HT-04, GSM1808381.HT-05, GSM1808382.HT-06, GSM1808383.HT-07, GSM1808384.HT-08, GSM1808385.CKD-01, GSM1808386.CKD-02, GSM1808387.CKD-03, GSM1808388.CKD-04, GSM1808389.CKD-05, GSM1808390.CKD-06, GSM1808391.CKD-07, GSM1808392.HD-01, GSM1808393.HD-02, GSM1808394.HD-03, GSM1808395.HD-04, GSM279060, GSM279061, GSM279062.

Gene Ontology (GO) enrichment analysis

To get the idea of what could be the effect of these extracted DEGs can only be possible when we have some preliminary insight in to their individual function. As gene ontology enrichment gives us the opportunity to get the basic information about any gene. As GO provides the basic terms subdivided into three important categories namely Molecular function (MF), Biological Process (BP) and Cellular Component involving the gene of interest (CC). These associated vocabularies of biological term help us to understand the basic functioning of that gene at molecular level.

The Fundamental genes can be defined as the genes that are found to show the differential expression irrespective of the condition, state of disease and pattern of their expression. The extracted total of 10 genes obtained divided in to two set. The set of Up-Regulated namely: ALB, G6PC, WT1, and THSD7A while, another is the set of Down-Regulated genes namely: DOCK2, IL7R, LCK, PTPRC, TNFAIP6 and LYL1. These two set of genes were also found actively participating in cancer. Thus, we subjected this list of genes to DAVID, an online tool for GO enrichment and function which is described as follows.

Set of up-regulated genes: ALB it is the putative gene involved in maintaining the Colloidal Osmotic pressure of blood and binds with the elements like Zn, Ca and K. Thus, this gene product act as a transport and its down regulation will increase stress in the cell surrounding media (i.e., blood). **G6PC** is responsible for the production of Glucose thus, maintaining the level of glucose in the blood. The perturbation in its expression disturbs the homeostasis in the blood. **WT1** gene is highly expressed in developing kidney. It plays role of being an Oncogene but also work as tumor suppressor. **THSD7A** promotes migration of endothelial cells and it may be involved in cytoskeletal organization. Variations in this gene can be associated with low bone mineral density results in osteoporosis.

Set of down-regulated genes: **DCOCK2** is involved in the rearrangement of cytoskeleton and participate in the migration of lymphocyte in case of some infectious disease. It also plays important role in activating the immunity by activating Genes REC1 and REC2

that in turn activates IL2 (interleukin 2) that is required for the normal functioning of leukocytes (WBCs).

LCK is involved in the selection and maturation of T cell. It is expressed in all the stages of thymocytes maturation events. This gene is activated in response to the T-cell Receptor stimulation. **IL7R** acts as receptor for thymic stromal lymphopoetin. **PTPRC** is required for T-cell activation followed by de-phosphorylation of SKAP1 and LYN, which directly effects the positive activation of T-cell receptor by SKAP1 by enhancing the MAP kinase activity while, LYN on the other hand is a Proto-Oncogene may result in the formation of tumors in kidney under stressed conditions (infections inside kidney or increase in salt concentration). **TNFAIP6** protein is involved in cell to cell and cell to matrix interaction thus forming the base of extracellular matrix stability and cell migration. It also enhances the serine protease inhibitor activity of TNFA1 and IL1. **LYL1** is a member of helix-loop-helix family of transcriptional factors, that bind to the DNA and performs major function in controlling the process of angiogenesis.

Conclusions

CKD is the condensed term used for the heterogeneous disorders affecting the kidney structure and function. CKD has formed as the major type of illness worldwide. Cardiovascular diseases are a very frequent implication of CKD nowadays. On the other hand High throughput gene expression recording using microarray has given the opportunity to intelligently curate the *fundamental* genes. We bolted out and report a set of ten genes that are differentially expressed in various conditions inside kidney. These extracted 10 genes were further divided in to two set. These two set of genes are found actively participating in causing cancer. We hope that the in depth expression and ontological accounts given in this article will pave a new approach to think for the novel therapeutic discoveries.

Conflict of Interest

The author declares that there is no conflict of interest.

Acknowledgements

References

1. Levey AS, Atkins R, Coresh J, Cohen EP, Collins AJ, et al. (2007) Chronic kidney disease as a global public health problem: approaches and initiatives–a position statement from Kidney Disease Improving Global Outcomes. Kidney 72: 247-259.

2. Stevens LA, Viswanathan G, Weiner DE (2010) Chronic Kidney Disease and End-Stage Renal Disease in the Elderly Population: Current Prevalence, Future Projections, and Clinical Significance. Adv. Chronic Kidney Dis 17: 293-301.

3. Hill NR, Fatoba ST, Oke JL, Hirst JA, O'Callaghan CA, et al. (2016) Global prevalence of chronic kidney disease–a systematic review and meta-analysis. PLoS One 11: e0158765.

4. Hwang SJ, Tsai JC, Chen HC (2010) Epidemiology, impact and preventive care of chronic kidney disease in Taiwan: Epidemiology of CKD in Taiwan. Nephrology 15: 3-9.

5. Levey AS, Eckardt KU, Tsukamoto Y, Levin A, Coresh J, et al. (2005) Definition and classification of chronic kidney disease: a position statement from Kidney Disease: Improving Global Outcomes (KDIGO). Kidney international 67: 2089-2100.

6. Sabanayagam C, Lim SC, Wong TY, Lee J, Shankar A, et al. (2010) Ethnic disparities in prevalence and impact of risk factors of chronic kidney disease. Nephrol. Dial. Transplant 25: 2564-2570.

7. Mills KT, Xu Y, Zhang W, Bundy JD, Chen CS, et al. (2015) A systematic analysis of worldwide population-based data on the global burden of chronic kidney disease in 2010. Kidney international 88: 950-957.

8. Perlman RL, Finkelstein FO, Liu L, Roys E, Kiser M, et al. (2005) Quality of life in chronic kidney disease (CKD): a cross-sectional analysis in the Renal Research Institute-CKD study. American journal of kidney diseases 45:658-666.

9. Jeong S, Oh JM, Oh KH, Kim IW (2017) Differentially expressed miR-3680-5p is associated with parathyroid hormone regulation in peritoneal dialysis patients. PLOS ONE 12: e0170535.

10. Woroniecka KI, Park ASD, Mohtat D, Thomas DB, Pullman JM, et al. (2011) Transcriptome Analysis of Human Diabetic Kidney Disease. Diabetes 60: 2354-2369.

11. Gu DF, Shi YL, Chen YM, Liu HM, Ding YN, et al. (2013) Prevalence of chronic kidney disease and pre-diabetes and associated risk factors: a community-based screening in Zhuhai, Southern China. Chinese medical journal 126:1213-1219.

12. Couser WG, Remuzzi G, Mendis S, Tonelli M (2011) The contribution of chronic kidney disease to the global burden of major non-communicable diseases. Kidney Int 80: 1258-1270.

13. Shannon MF (2002) TRANSCRIPTION: Of Chips and ChIPs. Science 296: 666-669.

14. He H, Mao L, Xu P, Xi Y, Xu N, et al. Ossification of the posterior longitudinal ligament related genes identification using microarray gene expression profiling and bioinformatics analysis. Gene 533: 515-519.

15. Raman T, O'Connor TP, Hackett NR, Wang W, Harvey BG, et al. (2009) Quality control in microarray assessment of gene expression in human airway epithelium. BMC genomics 10: 493.

16. Barrett T, Troup DB, Wilhite SE, Ledoux P, Evangelista C, et al. (2010) NCBI GEO: archive for functional genomics data sets—10 years on. Nucleic acids research 39: D1005-1010.

17. Gautier L, Cope L, Bolstad BM, Irizarry RA (2004) affy-analysis of Affymetrix GeneChip data at the probe level. Bioinformatics 20: 307-315.

18. Smyth GK (2004) Linear models and empirical bayes methods for assessing differential expression in microarray experiments. Stat. Appl. Genet. Mol. Biol 3: Article 3.

19. Gatto L, Breckels LM, Naake T, Gibb S (2015) Visualization of proteomics data using R and Bioconductor. Proteomics 15: 1375-1389.

20. Kauffmann A, Gentleman R, Huber W (2009) arrayQualityMetrics-a bioconductor package for quality assessment of microarray data. Bioinformatics 25: 415-416.

21. Heyer LJ, Kruglyak S, Yooseph S (1999) Exploring expression data: identification and analysis of coexpressed genes. Genome Res 9: 1106-1115.

22. Anders S, Huber W (2010) Differential expression analysis for sequence count data. Genome Biol 11: R106.

23. Bullard JH, Purdom E, Hansen KD, Dudoit S (2010) Evaluation of statistical methods for normalization and differential expression in mRNA-Seq experiment. BMC Bioinformatics 11: 94.