**Abstract:** Cervical cancer is one of the problems occurring world-wide also caused due to the over-expression of Oncostatin M Receptor signalling (OSMR).We have analyzed the GEO dataset (GSE27480) consisting of primary site as control and lymph node as sample. The study involved in the pathways includes epithelial cell differentiation process and the genes which were CYR61, LGALS1, MYC, EHF, KRT17 and IRF6 respectively taken from TF gene interactions which was Gene Ontology Biological Process. The GO: BP pathways include the epithelial cell differentiation, Regulation of Cell differentiation, Cell proliferation and Angiogenesis processes. These pathways can help us to analyze where the genes could serve as a promoter for apoptosis, cell cycle progression for the cervical cancer to occur.

***Index terms****:* Cervical cancer, GO: BP pathways, Oncostatin M Receptor, Epithelial cell differentiation process.

1. **INTRODUCTION**

Cervical cancer is the second most common female malignancy to stand as a global health challenge World Health Organization. Comprehensive cervical cancer control. A guide to essential practice. Geneva: WHO, 2006. Human Papilloma Virus (HPV) is known to cause cervical cancer where it develops an infection in the intraepithelial layer of the mucosa that can lead into squamous cell carcinoma (SCC) (IARC 1995; WHO 2009). Generally, cervix contains two types of cells known as squamous cells and glandular cells collectively known as the transition zone where most cervical pre-cancers and cancers begin to grow. The studies suggest that 90% of cervical cancer arise from squamous cell carcinomas (Small et al., 2017). The OSMR overexpression due to SCC shows greater response to the major ligand Oncostatin-M (OSM), which induced increased cell migration, invasion and angiogenesis ([Winder et al, 2011](https://www.nature.com/articles/bjc2016199#ref-CR45)). Cervical epithelia have numerous functions that include proliferation, differentiation, and maintenance of fluid balance where the epithelial functions must be tightly regulated during pregnancy and parturition as the cervix undergoes extensive growth and remodeling. (Brenda et al.,2007). The study of cervical cancer due to OSMR signaling involves the Data set GSE27480 where the primary site was taken as control and the lymph nodes of patients as samples. Here, the primary site acts as a tissue sample whereas the lymph node acts as a metastatic site. The significant genes were identified by adjusting the p- value of about 0.05 and log2 fold change of 1.0. The pathway enrichment gave the protein-protein interactions and Gene Regulatory Networks which provided the network pathways of KEGG, REACTOME, GO: BP, GO: MF and GO: CC. The GO: BP provided the significant genes of the Epithelial Cell Differentiation and Regulation of cell differentiation pathways and the down regulated genes were found to be CYR61 and LGALS1 whereas the up regulated genes were MYC, KRT17, IRF6 and EHF respectively. Hence, these genes may be involved in the pathways leading to cervical cancer.

1. **MATERIALS AND METHODS**

**MICROARRAY DATA ANALYSIS:**

The gene level summarisation was done using the microarray data which uses the class comparison method where it finds the genes which are differentially expressed. In view of this, the class comparison was done between the sample taken as lymph node (metastatic site) and primary site as control sample comprising 72 samples out of which 38 samples (lymph node (metastatic site) and primary site) were analysed for differential gene expression.

**DIFFERENTIAL EXPRESSION ANALYSIS:**

Differential expression analysis (DEA) is the process of taking the normalised read count data in order to perform statistical analysis to study the quantitative changes in gene expression levels between experimental groups. With the DEA analysis it was possible to obtain information about the results of up regulated and down regulated genes according to the p-value and log2 fold change. GEO2R tool was used to compare two or more groups of samples present in the GEO series showing six genes that were differentially expressed across the experimental conditions got during volcano plot normalisation.

**FUNCTIONAL ENRICHMENT ANALYSIS:**

Network analyst is an online tool which helps in potent description of molecular biomarkers and underlying biological process onto the context of protein-protein interactions, TF gene interactions and heat map clusters with support of statistical and functional evidence. The Gene Set Enrichment Analysis (GSEA) provides a complete statistic of the related genes present within the epithelial cell differentiation process alone. The enrichment analysis includes the pathways of KEGG and REACTOME to study the functional aspects of the gene sets present. Over representation analysis (ORA) gave the overall differential genes involved in the process of epithelial cell differentiation and positive regulation of epithelial cell differentiation.

**RESULTS AND DISCUSSION:**

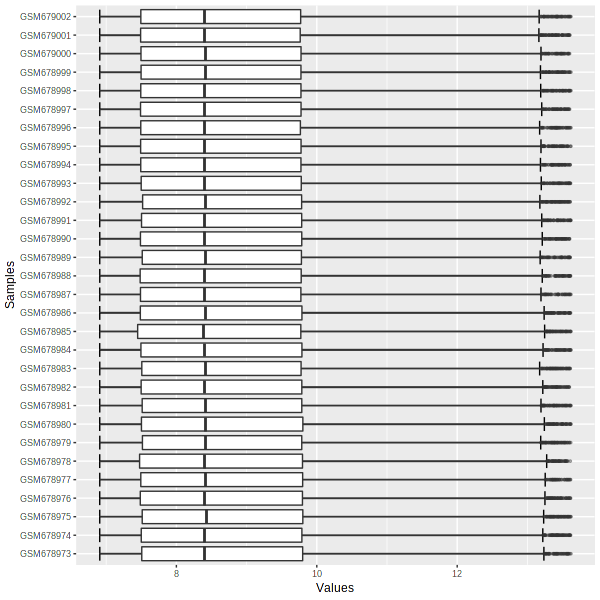
**DATA FILTERING AND NORMALIZATION:**

Data filtering increases the statistical power by removing unresponsive genes prior to differential expression analysis where the lower expression analysis of genes can be removed by variance filter adjustments.

Normalisation is the process of reorganising in a data base so that there is no redundancy in data and to become logical. Proper normalisation is essential to draw conclusions from DEA.

**BOX PLOT:**

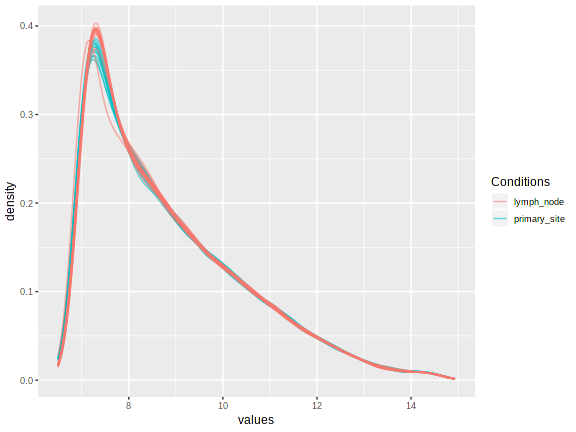
The Box Plot is a graphical data where relies on the concentration of data and to show the extreme values that are far related from most of the data. Box plots uses the quantile normalisation to make each quantile same across the sample and shows the outliers between the samples present in the data according to the number of experiments.



**Figure 1: Box Plot.**

**DENSITY PLOT:**

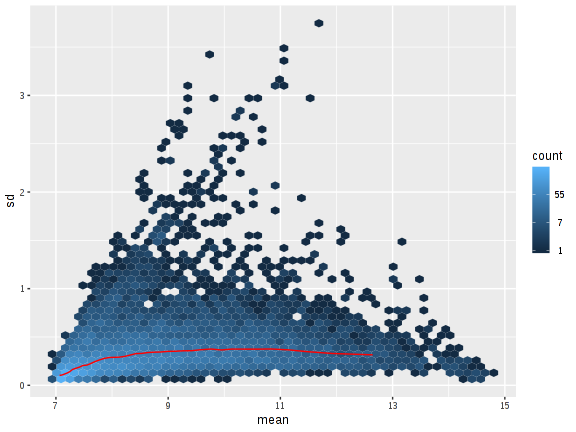
The density plot is an important parameter where it shows the plot between the density values and the log2 transformed values. This indicates the occurrence where each of the samples are concentrated and associated with the data set.



**Figure 2: Density plot.**

**MSD PLOT:**

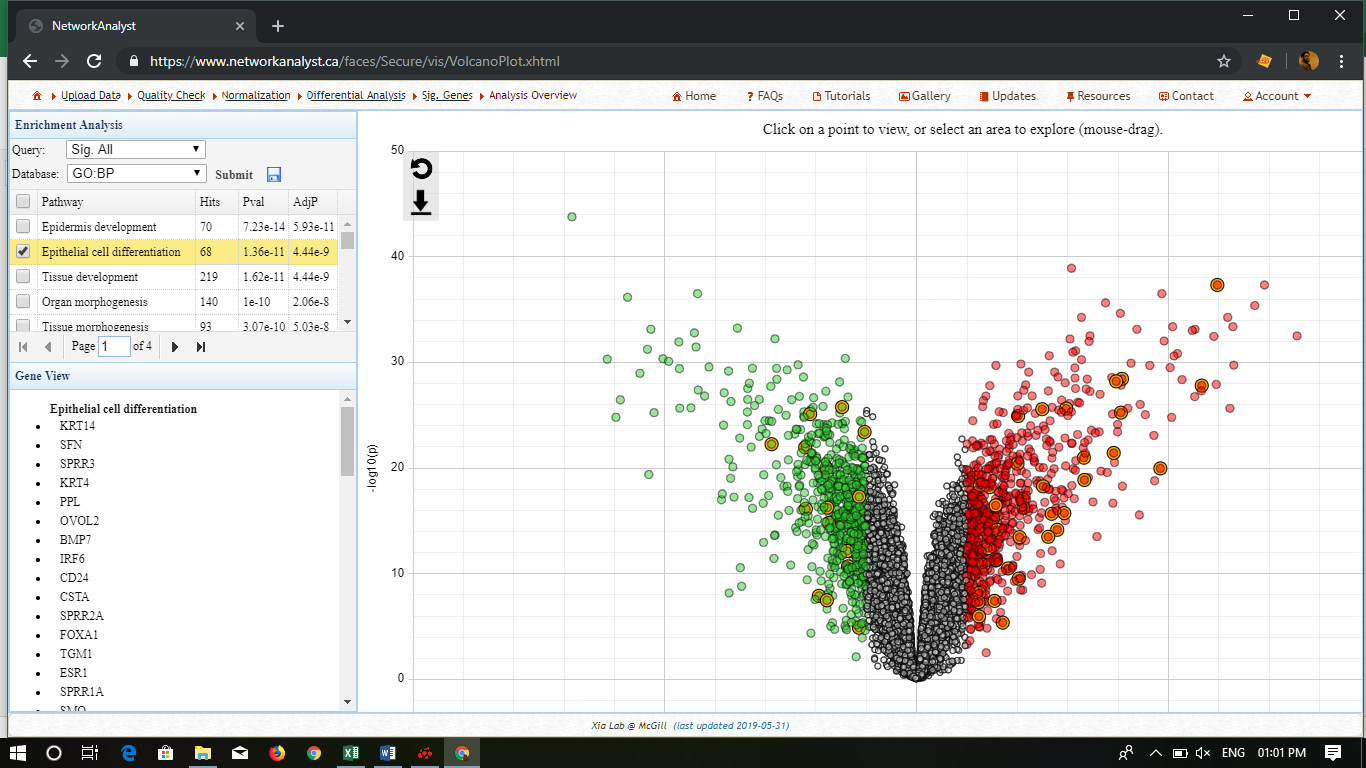
It is a diagnostic plot in normalisation which gives the information on the deviation of the genes from a mean point with respect to its position. The red line gives a check on the whether there is a dependence between the counts and the variance and the blue dots represent the cluster of genes.



**Figure 3: MSD Plot**

**VOLCANO PLOT:**

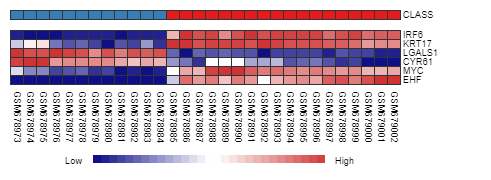
A volcano plot is a visualisation tool to visualise the differentially expressed genes that are statistically significant. The volcano plot emphasised the upregulated genes and downregulated genes which were coloured in red and green respectively whereas the non- significant genes were shaded in grey colour. The distribution of genes into the upregulated and downregulated portions of the plot represented as circles were the genes involved in the narrowed down pathways.



**Figure 4: Volcano plot**

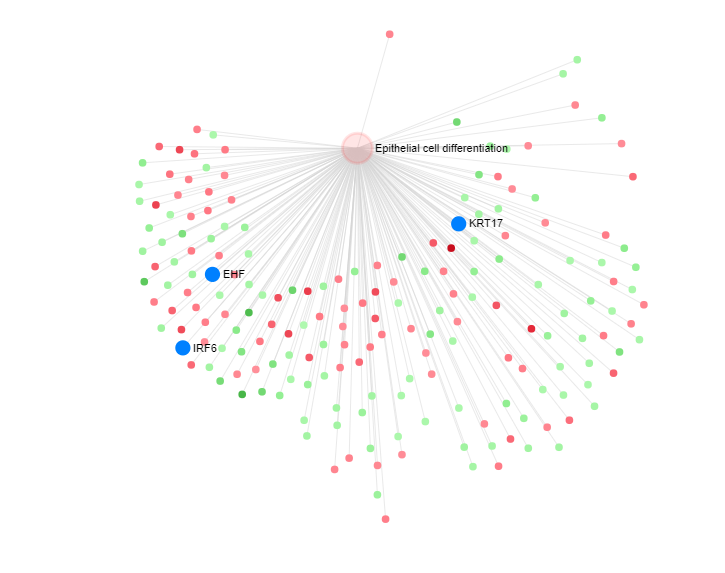
**HEAT MAP CLUSTERING:**

A Heat map is a two-way display of a data matrix in which the individual cells are displayed as coloured rectangles. It helps in identifying the intensity of expression pattern in genes. The blue coloured rectangles displays’ the low intensity whereas the red coloured rectangles display the high intensity of genes.

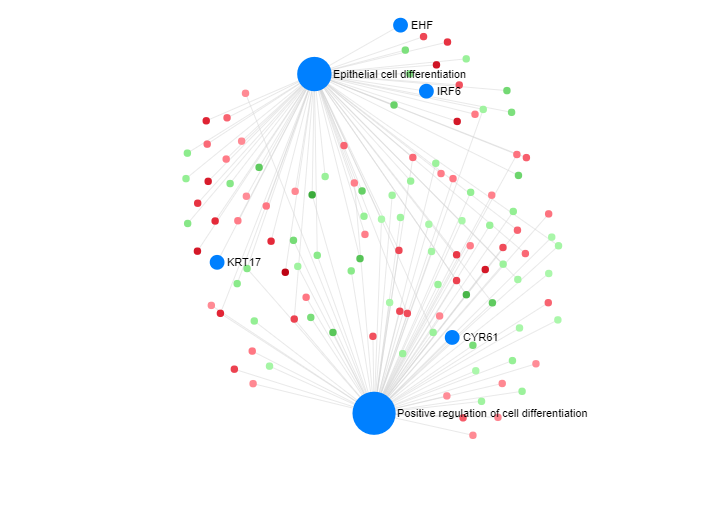
**Figure 5: ORA heat- map clustering**

**PATHWAY ENRICHMENT ANALYSIS:**

The Pathway Enrichment Analysis is used to interpret the gene set data for better understanding of gene function annotation. The enrichment of this analysis was done using KEGG, REACTOME and gene ontology pathways. The Pathway Enrichment Analysis included Over-Representation Analysis Enrichment and Gene Set Enrichment Analysis. In the Over- Representation Analysis Network, the association of biological genes was found out whereas for the Gene Set Enrichment Analysis network the differential expression of genes associated with a certain biological process or not was found.



A)

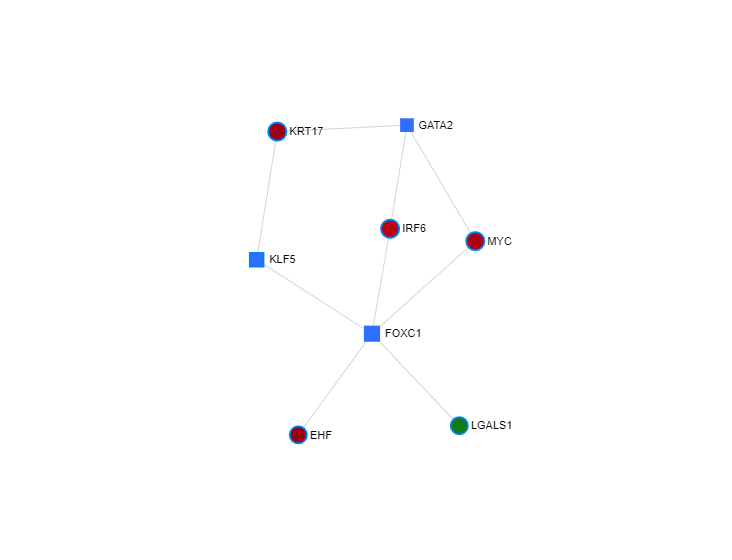


**B)**

**Figure 6: Network Analysis using GSEA and ORA**

In fig. 1: **A)** The Gene set enrichment analysis of biological differential genes KRT17, IRF6 and EHF (highlighted in blue colour) whereas the red and green dots represent the up and down regulated genes involved in the epithelial cell differentiation process. **B)** The Over- Representation analysis of genes involved in epithelial cell differentiation and positive regulation of cell differentiation process involving the biological differential genes of EHF, IRF6, KRT17 and CYR61 (highlighted in blue colour) whereas the red and green dots represent the up and down regulated genes involved between these processes.

**GO: BP PATHWAYS:**



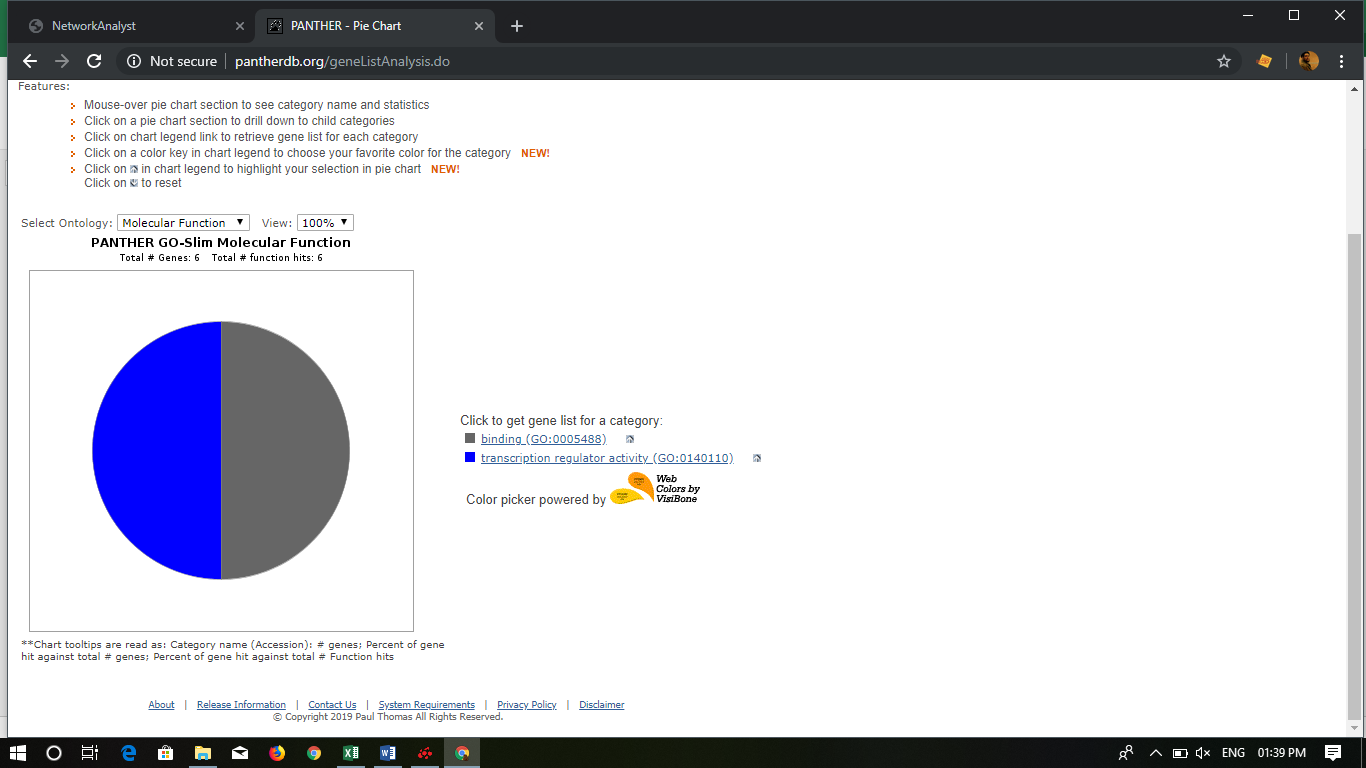
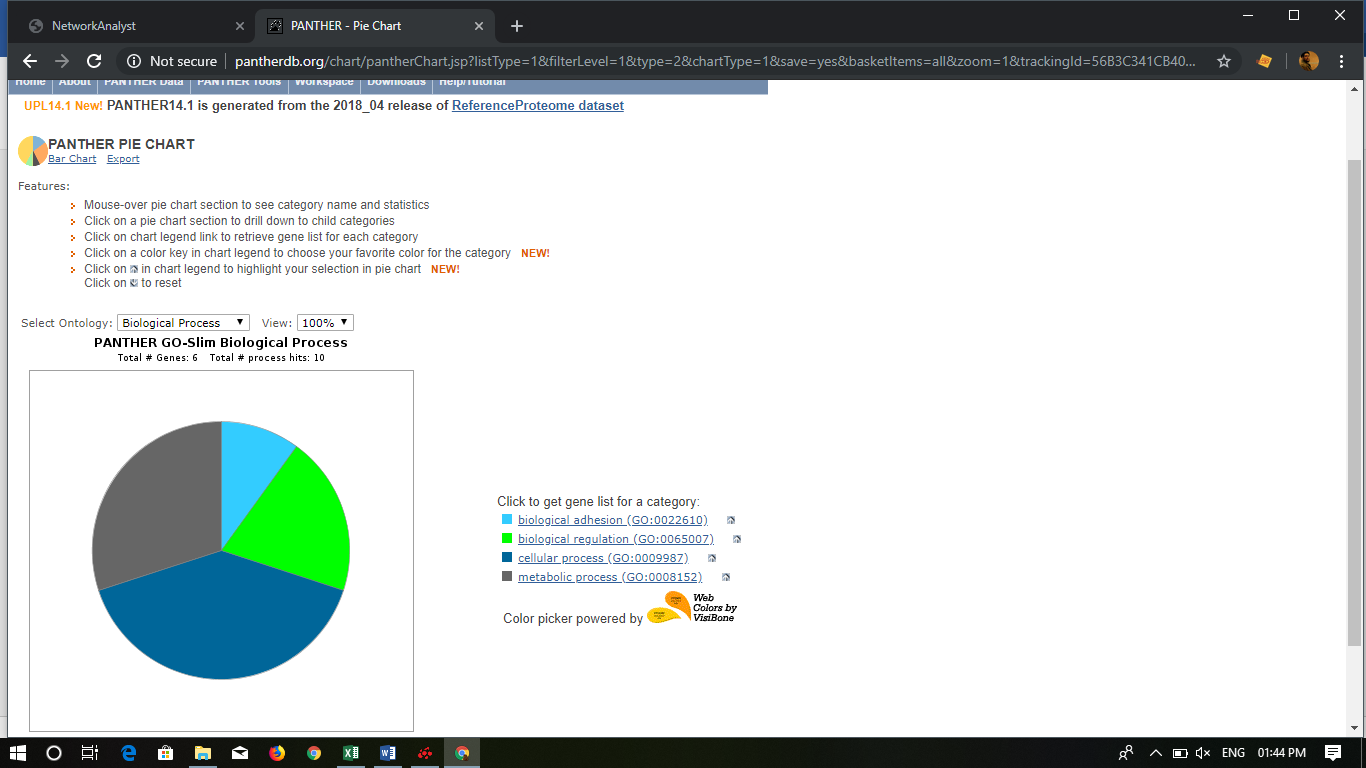
**Figure 7: Differential gene interactions in GO: BP pathways**

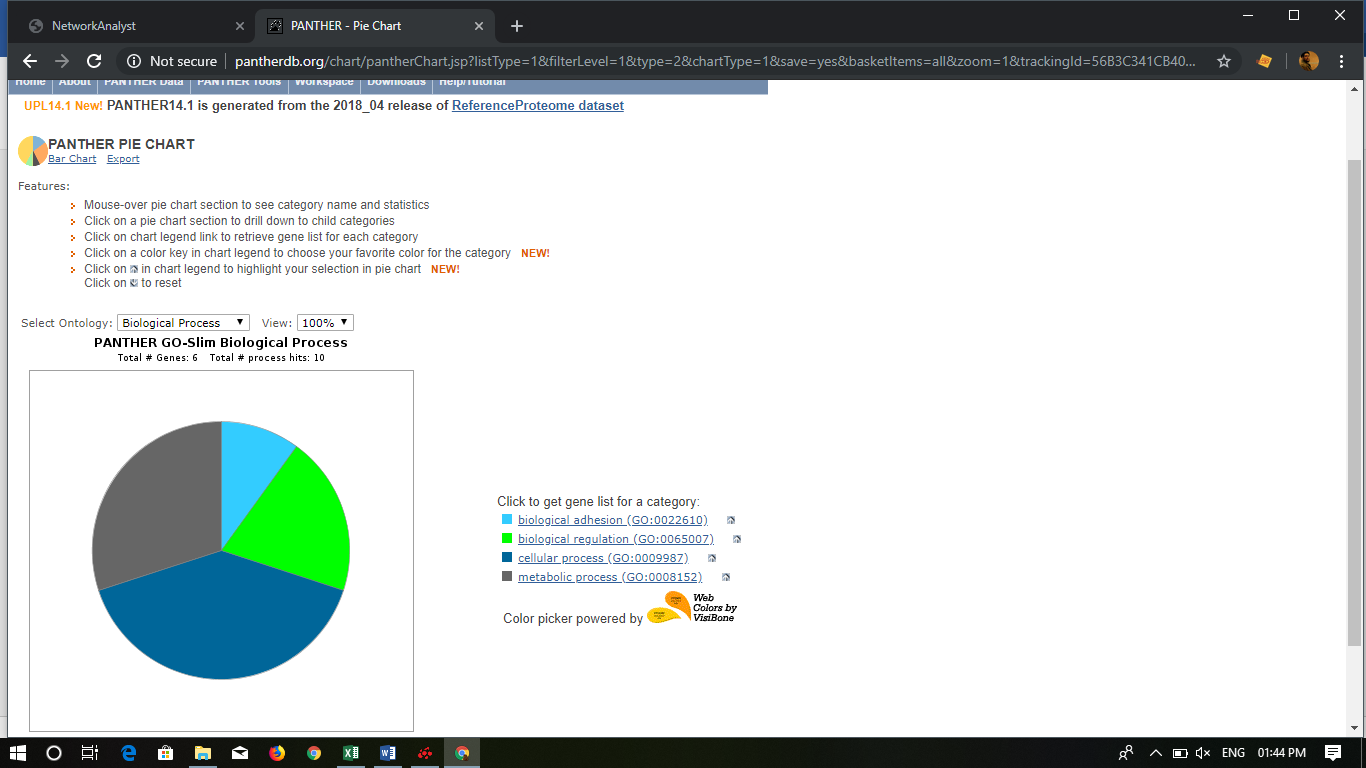
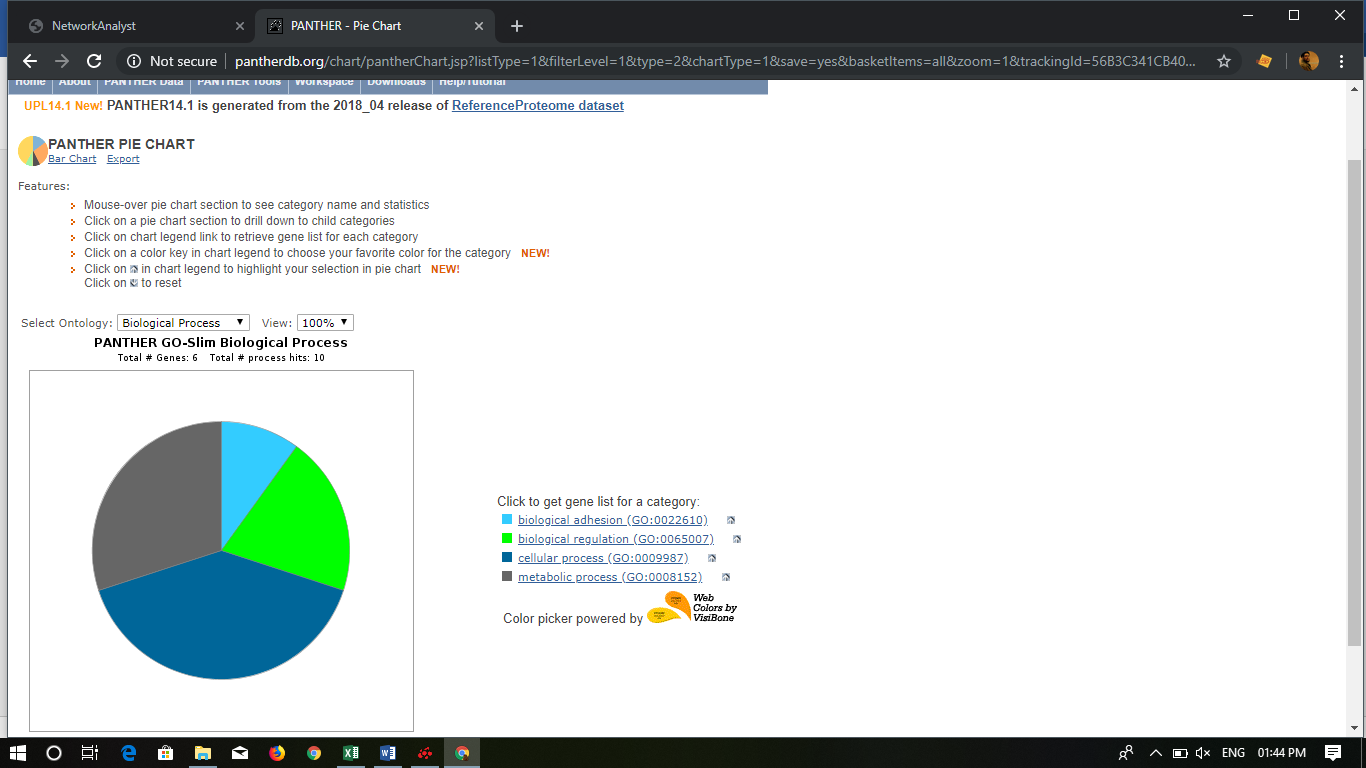
The differential genes were involved in the processes of epithelial cell differentiation, regulation of cell differentiation, Angiogenesis and Cell proliferation where its main role is Transcriptional binding activity and in epithelial cell proliferation and differentiation. The differential genes were highlighted in blue circles respectively. The LGALS1 gene is downregulated indicated in green whereas the upregulated genes (KRT17, IRF6, MYC, EHF) were indicated in red colour.

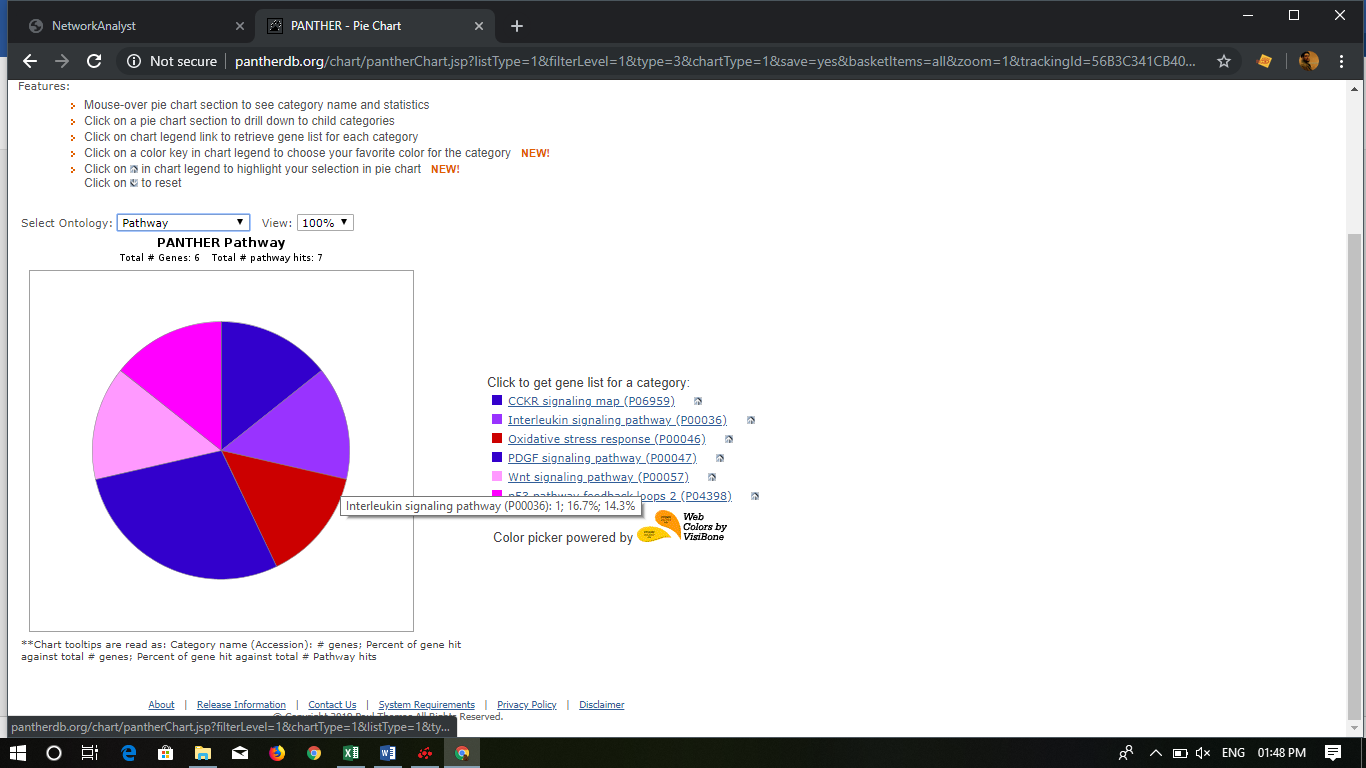
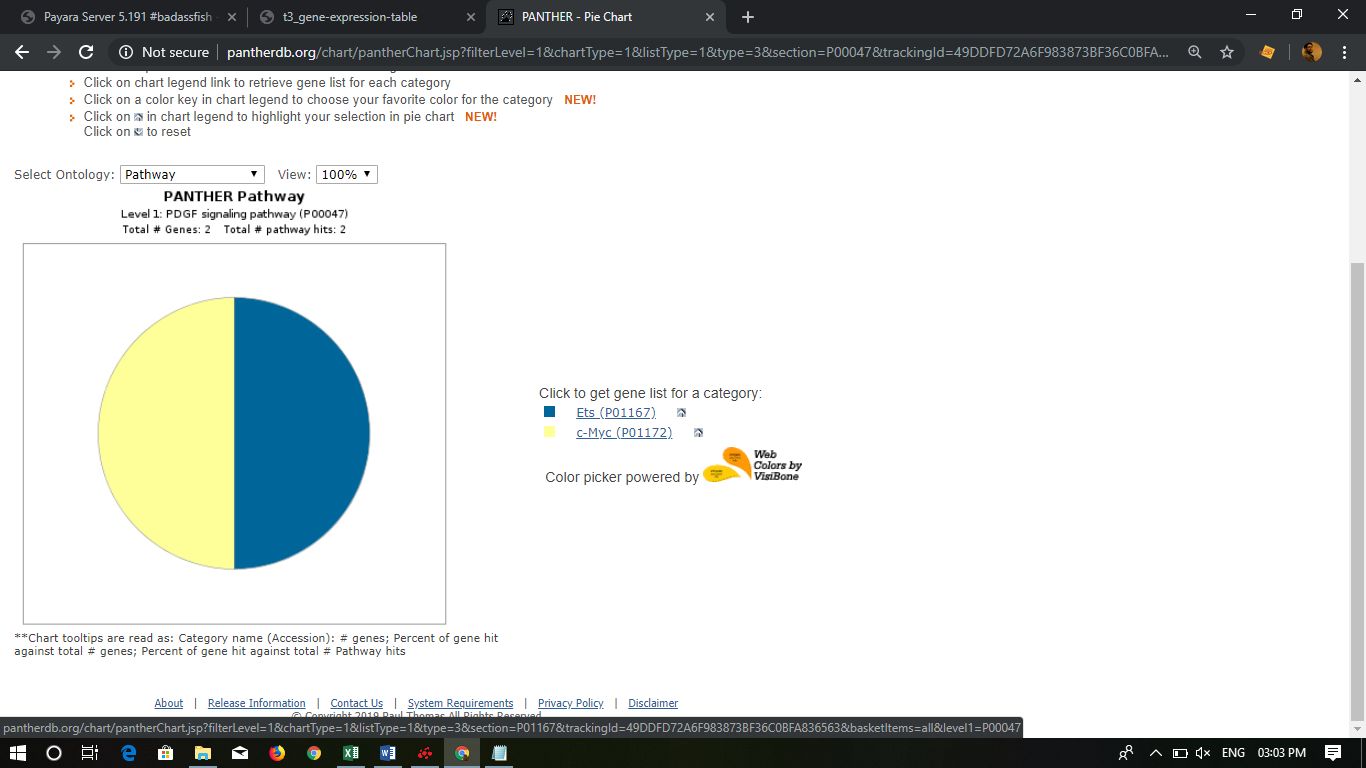
**FUNCTIONAL ENRICHMENT ANALYSIS:**

**PANTHER Analysis:**

PANTHER tool describes the Protein Analysis Through Evolutionary Relationships where it predicts the function of uncharacterised genes based on evolutionary relationships to find the differentially expressed function of genes involved in the pathway function.

Top of Form

The Molecular function involving differential expression genes were CYR61, LGALS1, MYC, IRF6 and MYC which played a role in Transcription binding activity whereas for the Biological process the genes involved were CYR61, MYC, LGALS1, EHF and IRF6 in cellular and metabolic pathways. The Pathway process involving differential gene was MYC in the process of Wnt, p53, CCKR, Oxidative stress and Interleukin signalling pathways whereas in PDGF signalling pathways, it was found that Ets played a role in Ras signalling and Myc gene played a role in cell cycle and cellular senescence.