**4. RESULTS**

**4.1 Sample collection**

The datasets for this paper were retrieved from the NCBI-GEO (Gene Expression Omnibus) a free public database of microarray or gene profile, and we obtained the gene expression profile of GSE6731-platform [HG\_U95Av2] Affymetrix Human Genome U95 Version 2 Array for healthy, UC and CD samples of the dataset. High throughput sequencing data profiling of gene expression is included in the data. The present data for GSE6731 includes out of these 36 cases, 4 are controlled cases, while 32 are samples with adult inflammatory bowel disease (IBD). Majority of cases have Crohn’s disease i.e., 19 while 9 have ulcerative colitis. Although the age group has not been considered for the analysis for this paper, the number of cases for each dataset is summarised below in Table 1. The research data was collected and conducted at the Johns Hopkins University School of Medicine,specifically in the Department of Medicine, Baltimore, USA. The study was submitted on January 12, 2007, and last updated on December 13, 2018.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Data set Accession ID** | **Type of the data**  **(file format)** | **Technology type (if any)** | **Number of cases** | **Reference** |
| GSE6731 | Microarray( CEL files) | in situ oligonucleotide | 36 samples (34 considered) | <https://pubmed.ncbi.nlm.nih.gov/17262812/> |

**Table 1:** The following table consists of dataset used for a genome wide sequencing analysis, each containing control/non-inflammatory and inflammatory samples, where GSE6731 is taken from Affymetrix Human Genome U95 Version 2 Array, and dataset GSE9452 and GSE16879 taken from U133 Plus 2.0 Array

**4.1.1 GSE6731**

The following dataset has been procured from NCBI GEO website where it matched the desired criteria of IBD conditions dataset. In order to perform Genome-wide sequencing of genes, cel files from the Affymetrix array are required. The expression data shows the variations in Crohn's and ulcerative colitis's genome-wide gene expression from endoscopic pinch biopsies. The research study compared active and inactive areas of UC and CD with infectious colitis and healthy controls. A total of 36 samples underwent unsupervised classification, revealing distinctive gene expression patterns among affected and unaffected IBD tissues, non-IBD colitis, and normal controls.

|  |  |  |  |
| --- | --- | --- | --- |
| **RAW FILE FORMAT** | **SAMPLES NUMBERS** | **CONDITIONS** | **SHORTLISTED CONDITIONS** |
| celfiles | 36 | Ulcerative colitis-9 | Ulcerative colitis-9 |
|  |  | Crohn’s disease-19 | Crohn’s disease-19 |
|  |  | Normal-4 | Normal-4 |
|  |  | Indeterminate colitis -2 |  |
|  |  | INF(bacterial infectious colitis)-2 |  |

**Table 2:** the GEO dataset GSE6731 has .CEL format taken from Affymetrix Human Genome U95 Version 2 Array where the shortlisted conditions are normal, UC and CD which are in total 34 out of the 36 samples present in the original dataset.

**4.2 Pre-processing of the extracted sample files**

Data preparation, processing and normalisation are among the first steps into the analysis of the first dataset GSE6731. In order to retrieve the expression values of the samples in IBD, the pre-processing of data was performed for the collection and the retrieval of DEGs afterwards. Then to retrieve the RAW file either we can use a set of R CODE or directly download and unzip the compressed version of the raw data file.

Once the RAW files have been decompressed the .*cel* files also have to be retrieved, i.e. unzipped in order to be read by the R studio console. *Library affy* has been used to perform the extraction of *cel* files from the system and further perform the basic normalisation and extract the expression values of the 32 samples of UC, CD and normal control are considered from the dataset.

**4.3** **Normalisation of expressed values**

*Readaffy* function reads the cel files into the R environment and prepares it for normalisation, *exprs* and *rma*are both functions used for normalising the data. Successfully extracting the expression values of the samples, now QC must be performed on the normalised data to anomalies and outliers. To detect the accurate normalisation of the dataset, Figure 3 shows a *boxplot* plotted to visualise the normalisation of the expression values from the samples.

A graph with numbers and lines

AI-generated content may be incorrect.

(a)

A rainbow colored bars on a black background

AI-generated content may be incorrect.

(b)

**Figure 3**: a Boxplot illustration to showcase data variance and normalisation in the samples of the dataset GSE6731. Samples are listed on the x-axis where 32 patient data is listed and the expression values are represented on the y-axis (a) A boxplot showing expression values of the data with unnormalized configuration, (b) Boxplot depicting normalised expression values after quality control of the data. The colour palette can be changed according to the user’s choice(note: here the palate has been set as “rainbow” to give a visually appealing aesthetic touch to the box plot).

Normalised values often show the variance of the samples showcasing the regulated genes through their expression values, this can be again visualised using a heatmap plot Figure 4 where the upregulated and downregulated genetic variance can be noticed using different colour themes often present in the *Bioconductor* package. *Library (ComplexHeatmap)*for this purpose has been utilised to illustrate the expression values of the normalised data and further the *colorRampPalette*command to make a combination of different colour schemes that are used to construct the heatmap.

A screenshot of a test

AI-generated content may be incorrect.

**Figure 4:** A heatmap constructed with the help of ComplexHeatmap package in R version 4.3.2, in the above graph, columns consist of the 32 samples and the associated regulated genes are somewhere represented in the rows. Due to the large number of expressed values here the map shows significant boxes overlapping with mild to high variance in the values. The **Violet** colour represents the Upregulated genes whereas the downregulated genes are shown in the **green** colour. The value ranges from +6 to -6 in the plot.

**4.4 Principal Component Analysis**

In the following dataset, a PCA is done using the expression values of Inflammatory Bowel Disease dataset to demonstrate any possible sample clustering within and variation between cohorts. It shows the correlation or variability of the samples from each other and gives a visual characterization to differentiate the normal vs diseased samples. In order to generate the plot, *ggfortify library* has been used in the R pipeline, further the pc data has to be generated where each sample contains the individual pc score using the expression values. The function *autopilot* has been used to generate a pc plot for and has been shown in the figure 5. The Principal Component 1 (PC1) is the first principal component extracted from the data which captures the direction of maximum variance in the data whereas Principal Component 2 (PC2) is the second principal component, orthogonal (perpendicular) to PC1, capturing the next highest amount of variance that is uncorrelated with PC1. The percentage numbers for each primary component reflect how much of the total variance is explained by that component. For example, if PC1 explains 55.59% of the variance, this means that when the data points are projected onto PC1, this single component accounts for 55.59% of the total variability in the data. In a similar manner if PC2 represents 11% of the variation, it means that PC2 accounts for an additional 11% of overall variability in the data set that PC1 does not explain. It is interpreted that "55.59%" for PC1 and "11.18%" for PC2 indicate that PC1 is the dominating component, accounting for the majority of the variation in the data. PC2 with a percentage that is lower (11.18%), detects less variation, but it provides additional information not described by PC1. When reading a PCA graphic, the axes (PC1 and PC2) do not correspond to the original variables in the dataset.

**A graph with red and blue dots

AI-generated content may be incorrect.**

**Figure 5:** Principal component analysis performed on the IBD dataset GSE6731. The points here represent the samples and are coloured according to the subject cohort. Results are plotted according to the PC1 and PC2 scores from the pc data matrix generated using ggfortify library, with the percent variation explained by the respective axis where the PCA plot is demonstrating variation between IBD and non-IBD samples.The PC1 score on x axis demonstrates a 55.59% showing the maximum variance and the orthogonal PC2 score is 11.18% capturing the next highest amount of variance.

**4.5 Differential Genes Expression analysis**

Differential Gene Expression (DGE) searches for the genes with expression that varies in response to therapy or between groups. After successfully normalising our data the next step is to perform differentiation of the expressed genes or retrieve differentially expressed genes which will be further used for biological interpretation. We initiate by accessing the *limma library*, limma is an *R/Bioconductor* software package that provides an integrated solution for analysing data from gene expression experiments. It is a very popular package for analysing microarray and RNA-seq data. We construct a design matrix for the healthy control samples, Crohn's disease and Ulcerative colitis samples, and further make a *contrast matrix* of IBD vs Non IBD samples which consist of CD and UC. Depending on the type of genes you wish to retrieve you may construct different contrast matrices. For the analysis, CD - normal, UC - normal and Combined - normal conditions were taken into account.

After the formation of the matrix, we further proceed with ranking genes in order of evidence for differential expression using the function *eBayes*. This step is critical for stabilising variance estimates, especially when there are a limited number of samples to analyse or genes with low expression. *"Fit"* refers to outcomes of fitting a linear model to gene expression data. This linear model have been fitted with a function like *lmFit()* from the *limma* package for microarrays. To compute the contrasts indicated in the *contrastMatrix*, we used the *fit()* function from the *limma* package (or a comparable function from another package). It takes the fitted model (*fit*) and the contrast matrix as inputs and computes the desired contrasts. After this procedure, we need to produce a *top table* of the conditions used in the contrast matrix we constructed. This table now contains the differentially expressed genes from the conditions we set from the samples of the dataset. The top table resulted in expression of **12,625** genes along with seven variables.

Proceeding with the expression, we use *dplyr* to filter the null values from the DEGs that have been found in the *toptables* of 3 conditions. After filtration, around we are left with *prob ids*, which we have to convert into gene symbols/ids for which the *library(hgu133plus2.db)* is used as it contains mappings between a manufacturer's identifiers and manufacturers accessions. Either you can download the annotation data manually or through a set of R code, which is then read into the R studio environment for mapping purposes. After merging, filtering the required columns and reordering columns, we have successfully produced a genelist for a specific condition i.e., CD vs normal, UC vs normal or combined vs normal. We retrieved around 2,450 entries for condition Ulcerative colitis vs Normal, 5,157 entries for Crohn’s disease vs Normal and 5,072 entries for Combined vs Normal condition.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **S.no** | **gene.symbol** | **logfc** | **AvgExpr** | **t** | **P.value** | **adj.P.val** | **B** |
| 1 | RABGGTA | 0.112717432 | 7.300086 | 1.04211036 | 3.051403e-01 | 0.522784076 | -5.7643117 |
| 2 | MAPK3 | 0.011684912 | 8.176038 | 0.06287469 | 9.502559e-01 | 0.976666137 | -6.2943555 |
| 3 | TIE1 | 0.316040658 | 3.985100 | 2.71566402 | 1.055796e-02 | 0.082077746 | -2.9996039 |
| 4 | CYP2C19 | 0.083767482 | 4.166605 | 0.54176689 | 5.917174e-01 | 0.759267407 | -6.1507259 |
| 5 | CXCR5 | 0.200358888 | 5.404175 | 2.21942411 | 3.364000e-02 | 0.150200659 | -4.0179155 |

**Table 3:** A demonstration of the example of how a retrieved set of resulted genes with important variable columns such as *logFC* and P value looks like, the genelist contains upregulated genes responsible for a particular condition in this case Combined UC and CD vs normal.

After downloading the gene lists from the R studio environment into the system, visualisation of the genes can be performed using a volcano plot from the function of *EnhancedVolcano*, which is a useful way to envision differential expressed genes from any specific condition. Depending on the variance of the values of *Fold change* and *P values*, we can set a limit which can draw a separation path to investigate the up and down regulated genes.

**4.5.1 Differential genes identified for condition Ulcerative Colitis vs control:**

After filtration, a total of 9156 genes were retrieved for this condition, with fold change values ranging from -4 to +4. The statistical significance of p-values varied from 0 to 1. The number of upregulated genes for the condition ulcerative colitis vs control are 286 and 1,548 downregulated genes. Among the top significantly upregulated differentially expressed genes are REG1A, REG3A, LCN2, OLFM4, and MIR8071-2. The GSE6731 dataset demonstrated that REG1A was apparently higher in the inflamed tissues of active IBD patients relative to healthy tissues. REG1A, being the most significant gene is a protein in the regenerating islet-derived (REG) gene family, is notably highly expressed in inflamed tissues of patients with inflammatory bowel disease (IBD) (Mao; Jia et al., 2021). Within intestinal epithelial cells (IEC), REG1A plays a pivotal role by inhibiting inflammatory responses, promoting cell proliferation, and suppressing epithelial apoptosis. Additionally, other members of the REG gene family, such as REG1B, REG3A, and REG4, which belong to the c-type lectin superfamily, also show significant expression.

On the other side, the most significantly downregulated differentially expressed genes are PLOD2, PKN2, ABCB1, DYRK2, GUCA2A, PRKACB, NR1D2, PCK1, SLC26A2, and CLDN8. CLDN8, or claudin 8, is a critical tight junction (TJ) molecule that is underexpressed in IBD. CLDN8 in the colon prevents absorbed Na+ from leaking into the lumen by sealing the lateral paracellular gap, establishing a protective barrier. It is interesting to note that CLDN8 is the most dramatically downregulated gene in the setting of IBD, indicating a potential role in disease aetiology and intestinal barrier disruption.

PLOD2 is also one of the downregulated genes in UC and according to studies, it is an enzyme that performs post-translational modification on collagen fibrils. Machine learning discovered PLOD2 as a significant gene for identifying Crohn's disease (CD)recurrence after surgery. PLOD2 is a member of the Procollagen-Lysine, 2-Oxoglutarate 5-Dioxygenase Family (PLOD). When PLOD family members are dysregulated, they can cause cancer to develop and metastasis (van Haaften et al., 2020). The differentially expressed genes of IBD for condition Ulcerative colitis are also visualised using the volcano plot by filtering the list according to the log fold change and P values of the respective genes (Figure 6 (a)).

**4.5.2 Differential genes identified for condition Crohn’s disease vs control**

A total of 5,157 genes were retrieved for this condition, with fold change values ranging from 1 to 14. The statistical significance of p-values varied from 0 to 0.045. This condition contains a maximum number of upregulated genes. The number of upregulated genes for the condition Crohn’s disease vs control are of number 5,100. Among the top significantly upregulated differentially expressed genes are IGHM, GAPDH, IGHA1, REG1A, RPL37A, HLA-C, HLA-B, EEF1A1, MIR8071-2, and HLA-A, along with RPL41. Of particular interest, the upregulation of IGHM (Immunoglobulin heavy constant mu) expression has been observed in the context of mucosal inflammation and immune responses associated with inflammatory bowel disease (IBD). Acute exacerbations of chronic IBD, including ulcerative colitis and Crohn's disease, are characterised by an increase in immunoglobulin G (IgG) positive cells in the mucosa (Rüthlein et al., 1992). IGHM - Immunoglobulins are immune system components that can have a role in inflammation. This suggests that IGHM may play a significant role in the immune response and inflammatory processes underlying IBD.

Apart from this, one of the most important genes are HLA-C, HLA-B, HLA-A (the family of Human leukocyte antigen genes) which are most likely the most significant genes related with Crohn's disease. These are major histocompatibility complex (MHC) genes, with certain alleles highly associated with Crohn's disease development and susceptibility. They have an important role in immune response and control, which are critical in the pathophysiology of Crohn's disease and other autoimmune illnesses. The HLA class II genes are thought to play a role in the development of IBD because their products are important in the immune response. Numerous studies have found links between the HLA-DR or -DQ phenotypes and either ulcerative colitis or Crohn's disease (Ahmad et al., 2006). According to the dataset and volcano plot of CD vs Normal ( figure 6(b)), it shows the maximum number of genes are upregulated as compared to the previous gene list.

The downregulation of GCFC2, GABRG2, APOM, CHEK2, and KSR1 in IBD, particularly Crohn's disease, indicates that these proteins may play a role in disease aetiology. These genes play roles in a variety of physiological activities, including transcriptional control (GCFC2), neurotransmission (GABRG2), lipid metabolism (APOM), DNA damage response (CHEK2), and signalling pathways (KSR1).

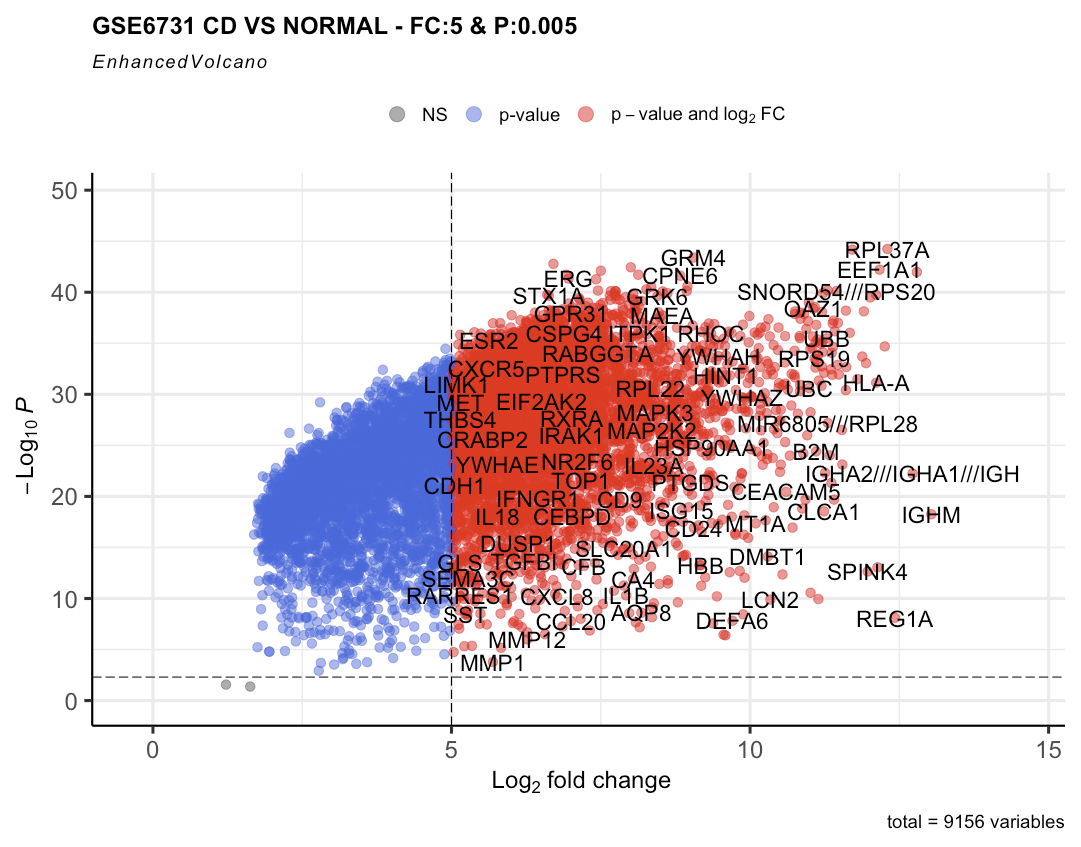
**4.5.3 Differential genes identified for condition IBD vs control samples:**

This condition yielded 9156 genes with fold change values ranging from -1 to 16. The statistical significance of p-values ranged from zero to 0.05. This condition contains more upregulated genes. The number of upregulated genes for the condition combined IBD samples vs control are of 5,071 upregulated genes whereas downregulated genes were around 3,080 in number. A visual illustration (figure 6 (c)) shows that most of the genes were upregulated for this case as well. One of the top significantly elevated differentially expressed genes in this dataset is REG1A, which is strongly expressed in inflamed tissues of individuals with inflammatory bowel disease (IBD) and plays a role in suppressing inflammatory responses (Mao; Jia et al., 2021). The group MIR8071-2 and MIR8071-1 contain microRNAs and immunoglobulin heavy chain genes, indicating an aggressive immune response in IBD. OLFM4, which is known for its role in tissue repair and mucosal healing, is also elevated. DEFA5, an antimicrobial peptide vital for host defence, is another notable gene in this dataset.

LCN2, which is involved in innate immunity and increases in response to inflammation, is likewise among the top elevated genes. Furthermore, GAPDH, a popular housekeeping gene frequently employed as a reference in gene expression investigations, had increased expression. The collection of genes IGHA2, IGHA1, IGH, like the previously stated immunoglobulin genes, is elevated, indicating that the immune system plays a role in IBD pathogenesis (Rüthlein et al., 1992). These findings indicate a complex interplay of immune response, inflammation, tissue healing, and antimicrobial defence mechanisms in the context of IBD, emphasising the disease's varied molecular profile (Stallhofer J et al., 2015).

The most significantly downregulated genes for the combined IBD genes were OPRD1, SATB1, RUNX2, SLC16A5 and RAB3GAP1. Among them, SATB1 and RUNX2 have been reportedly found to be least significant in Inflammatory bowel disease. SATB1 is a chromatin remodelling factor required for T cell lineage formation in the thymus. It controls the particular gene expression and activity of effector Th17 cells in tissue inflammation. It affects T-cell growth and immunological responses. SATB1 downregulation has the potential to alter immune cell function and regulation (Yasuda et al., 2019). Another study verified that Runx2 protein levels were shown to be lower in the intestinal epithelial cells (IECs) of CD patients, suggesting that Runx2 may protect IECs against apoptosis in CD, indicating a new biological target for CD treatment. RUNX2 is a transcription factor that plays a role in osteogenesis and bone development, and its downregulation may have an impact on bone remodelling, which has been observed in some IBD patients ( Liugen Gu a 1 et al., 2017).

A graph with red and blue dots

AI-generated content may be incorrect. 

(a) (b)

A graph with red and blue dots

AI-generated content may be incorrect.

(c)

**Figure 6:** Differentially expressed genes volcano plot on the condition demonstrating upregulated genes and downregulated genes (a)Ulcerative colitis vs normal, The following data classifies 286 upregulated genes and 1,548 downregulated genes, where fc value was >0.5 and P value set to be < 0.05 (b) Crohn’s Disease vs normal with upregulated genes of number 5,100 where fc was set to be >5 and P value < 0.005 (c) combined IBD vs normal genelist, with significant 5,071 upregulated genes where fc was set to be >5 and P value < 0.005. The red spot indicates upregulation, blue indicates downregulation, grey points are non-significant or neutral points that fall between the thresholds for upregulation and downregulation whereas green indicates no significant change.

**4.6 Biological Interpretation**

For the interpretation of genes from the different conditions of the IBD dataset, g:Profiler has been used in course of the analysis. g:Profiler is a web-based toolset for functional profiling of genes. It's used in computational biology research to analyse gene lists and perform functional enrichment analysis. Top genes have been taken from each condition and analysed using this tool for the functional profiling and molecular function analysis.

**4.6.1 Genes of ulcerative colitis vs normal:**

The profiling of the gene list shows significant functions of the particular gene mapped to a source name along with the adjusted p values arranged in a descending manner. Figure 7 shows individual characteristics of the genes plotted according to their p values and functionality. It has grounded the genes according to their activities in this significant disease. The profiling shows that UC affects mostly factor binding site motifs and binding proteins in the patients.

**A screen shot of a computer

AI-generated content may be incorrect.**

**Figure 7:** The above chart shows the description of the functional enrichment of the genes for Ulcerative colitis condition of IBD samples. Each colour signifies a group where the similar genes are grouped according to their functions, the red colour is for molecular function- protein binding, orange indicates biological process i.e., metabolic or cellular process, green represents cell components like chromosome or cytoplasm, pink, indigo and light blue for signalling pathways, grey colour shows factor binding site motifs, cyan for miRna site, purple shows HPA and maroon for the Hypothetical proteins.

**4.6.2 Condition Crohn’s disease s normal:**

For this condition a noticeable change in the number of genes associated with the functions and groups have been there. In comparison to UC, the number of genes associated with hypothetical proteins, cellular components and biological processes are significantly more in condition CD. This verifies the studies of previous research on Crohn’s disease which affects cellular components like white blood cells, red blood cell counts and C-reactive proteins (Boyapati et al., 2015).

**A screen shot of a computer

AI-generated content may be incorrect.**

**Figure 8:** This chart shows the description of the functional enrichment of the genes for Crohn’s disease condition of IBD samples. Each colour signifies a group where the similar genes are grouped according to their functions, the red colour is for molecular function- protein binding, orange indicates biological process i.e., metabolic or cellular process, green represents cell components like chromosome or cytoplasm, pink, indigo and light blue for signalling pathways, cyan for miRna site, purple shows HPA and maroon for the Hypothetical proteins.

**4.6.3 Condition combined IBD vs normal:**

In this combined condition again, there are few noticeable changes in the number of genes associated with the functions and groups. In comparison to individual UC and CD, the number of genes associated with biological processes, the mammalian cell death network includes three distinct functional types: apoptosis, autophagy, and necrosis. Each of them is distinguished by a variety of cell morphological alterations. The pathogenesis of inflammatory bowel illnesses (ulcerative colitis, Crohn's disease) entails an aberrant immune response to disrupted intestinal microbiota. One of the most critical factors in the development of chronic inflammatory disease and subsequent multiple organ dysfunction is the gut's barrier function, which regulates cellular survival. hways and hypothetical proteins, are significantly more in condition of IBD.

The biological pathways are proven to be affected by the IBD genes in the patients according to this functional analysis, hence agreeing to the previously study based research on IBD. According to the study, the mammalian cell death network includes three distinct functional types: apoptosis, autophagy, and necrosis. Each of them is distinguished by a variety of cell morphological alterations. The pathogenesis of inflammatory bowel illnesses (ulcerative colitis, Crohn's disease) entails an aberrant immune response to disrupted intestinal microbiota. One of the most critical factors in the development of chronic inflammatory disease and subsequent multiple organ dysfunction is the gut's barrier function, which regulates cellular survival. (Zemljic et al., 2014).

**A screen shot of a computer

AI-generated content may be incorrect.**

**Figure 9:** This chart shows the description of the functional enrichment of the genes for IBD condition of IBD samples. Each colour signifies a group where the similar genes are grouped according to their functions, the red colour is for molecular function- protein binding, orange indicates biological process i.e., metabolic or cellular process, green represents cell components like chromosome or cytoplasm, pink, indigo and light blue for signalling pathways, cyan for miRna site, purple shows HPA and maroon for the Hypothetical proteins.

**5. DISCUSSION**

The prevalence of inflammatory bowel disease (IBD) is on the rise, prompting researchers to seek innovative treatments to alleviate its symptoms and mitigate its impact on patients' lives. IBD, a serious condition with profound consequences, significantly affects patients' overall health and quality of life. While much attention has been devoted to studying IBD in adult patients, there is growing recognition of the importance of understanding its manifestation in children to optimise treatment strategies. Despite considerable efforts in the detection, diagnosis, and management of IBD, uncertainties persist (Seyedian et al., 2019). Previous research has identified a number of genes associated with IBD, providing valuable insights into their role and potential implications for disease progression and comorbidities, including cancer. Notably, innate immunity genes such as NOD2, first identified in 2001, and autophagy genes like ATG16L1 and IRGM underscore the significance of abnormalities in intracellular bacterial processing in the pathogenesis ofInflammatory bowel disease (Noble et al., 2010). Moreover, meta-analyses and subsequent investigations into ulcerative colitis (UC) have pinpointed additional susceptibility genes within the IL-23 pathway, namely IL12B, JAK2, and STAT3, further elucidating the genetic underpinnings of IBD (Ng et al., 2017).

To assess in a conclusive way the role gene in this analysis has been used to derive the characterization of approximately 36 patients with inflamed bowel disease has generated the following conclusions: first, the expression levels of the gene products of HLA class I, REG1A, OLFM4, LCN2, IGHM and DEFA5 differentially associated with progression in patients with IBD. Secondly, the expression levels of these genes have been outlined in the previous research studies. Lastly, these genes are also closely associated with several common events in colon and rectal cancer.

In the present study, the genes REG1A, LCN2, and OLFM4, have been identified as upregulated in ulcerative colitis, each with distinct genetic mechanisms. REG1A, a member of the Reg gene family, has been associated with poor clinical outcomes in early-stage colorectal cancer (CRC) patients (Astrosini et al., 2008). Additionally, REG1A levels were found to be elevated in inflamed colorectal tissues of IBD patients, with abnormal elevations suppressing inflammatory responses, promoting cell proliferation, and reducing epithelial apoptosis in intestinal epithelial cells (IEC) (Derakhshani et al., 2022; Mao et al.). LCN2, also known as neutrophil gelatinase-associated lipocalin (NGAL) or siderocalin, is a peptide created by macrophages, neutrophil granulocytes, and other immunological and parenchymal cells. It exhibits an antibacterial effect by binding catechol-type siderophores with extraordinary affinity. Furthermore, it is associated with a variety of physiological functions, such as transporting hydrophobic ligands across cell membranes, influencing immunological responses, maintaining iron homeostasis, and stimulating epithelial cell differentiation. OLFM4 is a glycoprotein that is naturally expressed in the gastrointestinal system, neutrophils, and the prostate. It participates in innate immunity, inflammation, and cancer. Olfactomedin-4 (OLFM4) is a gene that is up-regulated in IBD tissue and colorectal cancer. It is also present in cells from colon adenocarcinomas. OLFM4 regulates intestinal inflammation and carcinogenesis and may be a therapeutic target for intestinal malignant tumours (Wang et al., 2018).

HLA-A, HLA-B and HLA-C along with IGHM were the prominent genes of Crohn’s disease. The human leukocyte antigen (HLA) complex on chromosome 6p21.3 is the most well-studied genetic locus in inflammatory bowel disease. HLA-A and HLA-B are genes from the human leukocyte antigen (HLA) class I complex, a gene-rich area on chromosome 6p21.3. The conventional HLA genes, HLA-A, HLA-B, and HLA-C, have the strongest association with IBD development. The HLA-A and HLA-B genes may influence the aetiology of IBS by altering the gut immune system. According to one study, the frequency of HLA-B12 was significantly higher in Crohn's disease patients (52%), compared to healthy controls (21%), and ulcerative colitis patients (10%) (Michelakos et al., 2022).

Antimicrobial and cytotoxic peptides known as defensins (DEFA5) play a role in host defence. According to previous studies, aside from their role in the innate immune system and IBD, numerous studies have shown that defensins are expressed and involved in some types of cancer, such as CRC, breast cancer, and kidney cancer. The expression levels of DEFA5 and DEFA6 were considerably higher in CRC tissues than in normal tissues. α-Defensin has five members: DEFA1 (also known as DEFA2, HNP-1, and HNP-2), DEFA3 (also known as HNP-3), DEFA4 (also known as HNP-4), DEFA5, and DEFA6. KLK12, a gene involved in tumour formation, was substantially elevated in CRC among the co-expressed genes DEFA5 and DEFA6 (Arijs et al., 2009).

**6. CONCLUSION**

In conclusion, the study culminated in a thorough investigation into the gene expression profiles associated with inflammatory bowel disease (IBD), with a specific focus on Crohn's disease (CD) and ulcerative colitis (UC). Utilising data sourced from the NCBI-GEO database and employing sophisticated analytical methodologies such as pre-processing, normalisation, and differential gene expression analysis, the research team successfully pinpointed key genes implicated in the pathogenesis of IBD. Noteworthy findings included the significant upregulation of genes like REG1A, REG3A, and LCN2 in UC, whereas CD displayed upregulation of IGHM and HLA genes. Furthermore, the comprehensive analysis conducted in this study sheds light on the intricate interplay of genetic factors in the pathogenesis of IBD. Building upon previous research, the identification of upregulated genes in CD, underscores the heterogeneous nature of IBD and highlights distinct molecular signatures associated with each subtype in the gene profiling. These findings not only deepen our understanding of the disease's aetiology but also provide potential avenues for personalised therapeutic interventions. Previous research studies had identified HLA and REG3A as particularly influential in IBD development. Moreover, the discovery of downregulated genes, including PLOD2, CLDN8, GCFC2, and GABRG2, offers novel insights into the dysregulated molecular pathways underlying IBD. The differential expression of these genes hints at their crucial roles in modulating immune response, intestinal barrier integrity, and tissue homeostasis, thus implicating them as potential targets for therapeutic interventions aimed at restoring mucosal integrity and ameliorating inflammation. The functional enrichment analysis further underscores the intricate network of biological processes involved in IBD pathogenesis. By elucidating the involvement of identified genes in key pathways related to immune regulation, cytokine signalling, and epithelial barrier function, this study provides a holistic view of the molecular mechanisms driving disease progression. Such insights not only inform our understanding of IBD pathophysiology but also lay the groundwork for the development of targeted therapies tailored to address specific molecular aberrations observed in individual patients. Overall, the findings of this study represent a significant advancement in the field of IBD research, offering valuable insights into the molecular basis of the disease and paving the way for the development of precision medicine approaches aimed at improving patient outcomes and quality of life. Further research aimed at validating these findings in clinical settings and elucidating the functional significance of identified genes is warranted to translate these discoveries into actionable strategies for the diagnosis, prognosis, and treatment of IBD.