Current Trends in Biotechnology and Pharmacy
Vol. 20(1) 2763-2773 January 2026, ISSN 0973-8916 (Print), 2230-7303 (Online)
DOI: 10.5530/ctbp.2026.1.5

Toward Precision Surgery in Rectal Cancer: Integrating Machine Learning for Molecular Subtype Identification and Surgical Planning

Alfred Augustin[1], Supriya Pathi[2], Usha Adiga[2]*,Vasishta Sampara[2]

[1]Department of Surgery, Apollo Institute of Medical Sciences & Research Chittoor, India
[2]Department of Biochemistry, Apollo Institute of Medical Sciences & Research Chittoor, India

* Corresponding author:ushachidu@yahoo.com

Abstract

Because rectal cancer is a diverse disease, improving patient outcomes requires precision medicine and individualized surgical techniques. To identify important molecular subtypes that could affect therapeutic decision-making and precision surgery, this study combined unsupervised machine learning techniques with high-throughput microarray analysis. Using R software (v4.0.1), microarray-based gene expression data from the GSE253106 dataset were extracted from the Gene Expression Omnibus (GEO). The molecular subtypes of rectal cancer were identified using unsupervised machine learning techniques. Markov Clustering (MCL) mapped molecular networks involved in DNA repair and microenvironment interactions, while K-Means clustering grouped genes according to similarities in their expression. Density-Based Spatial Clustering of Applications with Noise, or DBSCAN, identified uncommon tumor subtypes linked to aggressive characteristics. The findings showed that rectal cancer had 411 DEGs, 348 of which were upregulated and 63 were downregulated. Genes such as TTTY15, RPS4Y1, and KDM5D were upregulated, whereas IGF2 and INS-IGF2 were downregulated. K-means clustering highlighted immune and metabolic regulation by grouping genes, such as C3AR1, TREM2, IGF2, and APOE. Networks involving DNA repair (FANCA, DAPK1) and the tumor microenvironment (AIF1, C1QA) were mapped using Markov Clustering (MCL). Rare aggressive subtypes were identified using DBSCAN, which also identified PLTP, ISG15, and TYROBP as indicators of immune evasion. TTTY15, KDM5D, and IGF2 were identified by outlier detection, indicating their involvement in tumor progression and treatment response. By highlighting the important molecular subtypes of rectal cancer, this study demonstrates how machine learning can be used to improve precision oncology and surgical techniques. Biomarker-driven treatment strategies may benefit from additional functional validation, which would improve therapeutic results and patient stratification.

Keywords: Precision surgery, Genomics, Machine learning, Rectal cancer

Introduction

Globally, colorectal cancer (CRC) is the third most common cancer and the third most common cause of cancer-related death (1,2). Nearly one-third of new CRC diagnoses are for rectal cancer, which poses particular difficulties because of its intricate anatomy and high risk of local recurrence. Interestingly, it is becoming more common among younger people (3). A stepwise genetic model that explains the progression of colorectal cancer from

adenoma to carcinoma has been established by the research of Fearon and Vogelstein (4,5). Although this framework is generally applicable to the large intestine, new developments in sequencing technologies have shed more light on the distinct genetic makeup of rectal cancer. Despite advancements in systemic therapies, radiotherapy, and surgery, overall survival and recurrence rates remain below ideal levels.

Tumor behavior, response to treatment, and disease course are all greatly impacted by the molecular heterogeneity of rectal cancer(6). Although clinical staging is a major component of traditional treatment approaches, the management of rectal cancer may be completely transformed by integrating molecular and metabolic profiling into clinical decision making(7). To improve precision surgery and customized treatment plans, this study aimed to identify important molecular signatures, examine transcriptomic and miRNA interactions, examine protein-protein networks, and investigate metabolic changes.

Erratic responses to conventional treatments are a significant obstacle in the treatment of rectal cancer. Although patient responses vary, the current standard treatment for locally advanced rectal cancer is neoadjuvant chemoradiotherapy followed by total mesorectal excision (TME). Some patients show resistance, resulting in recurrence and metastasis, while others attain a complete pathological response. Overtreatment or undertreatment is caused by the absence of trustworthy molecular markers to predict therapy outcomes, highlighting the critical need for biomarkers to direct patient stratification and facilitate tailored therapy. Treatment resistance, immune evasion, and tumor progression are all significantly influenced by differentially expressed genes (DEGs). Genes with altered expression patterns can be identified using high-throughput microarray data. APC, TP53, and KRAS are the most commonly mutated genes in colorectal cancer (CRC) (8,9), and are also frequently altered in rectal cancer. Although KRAS mutations are less common in rectal cancer (39% vs. 65%), APC (78% vs. 70%), and TP53 (81% vs. 65%) mutation rates are higher in rectal tumors than in proximal colon tumors (9). PI3K, EGFR, Wnt/β-catenin, IGF, TGFβ, p53, DNA mismatch repair (MMR), extracellular matrix (ECM) remodelling, and epithelial-mesenchymal transition (EMT) are among the oncogenic pathways that are impacted by genomic, transcriptomic, and epigenetic changes that cause rectal cancer (10-13). PTEN, APC, TP53, and SMAD4 inactivation, KRAS mutations, and MYC overexpression are examples of common genetic changes (14).

Machine learning is a potent tool for evaluating intricate genomic data and categorizing tumors into molecular subtypes (15). This study used unsupervised machine learning techniques to stratify patients with rectal cancer according to their gene expression patterns. This could help guide personalized treatment decisions by offering insights into prognosis, therapeutic response, and recurrence risk.

Rectal cancer remains a major clinical challenge because of its heterogeneity and inconsistent response to treatment (16). Promising approaches to enhance patient outcomes include identifying important molecular signatures and incorporating machine learning for molecular subtyping. This study aims to improve personalized medicine and precision surgery for rectal cancer by bridging the gap between molecular research and clinical practice. The knowledge acquired will not only improve our comprehension of the biology of rectal cancer but also set the stage for innovative patient-specific diagnostic, prognostic, and treatment strategies.
The study's goals
1. To use high-throughput microarray data to examine differentially expressed genes (DEGs) in patients with rectal cancer.
2. Examine gene expression data for rectal cancer and use patterns in gene expression to determine molecular subtypes.

Materials and Methods

Analysis of gene expression and microarray data

Understanding the molecular landscape of rectal cancer requires knowledge of gene expression data derived from microarray analysis. High-throughput microarray and sequencing datasets were stored in the Gene Expression Omnibus (GEO), a public repository for functional genomic data. The GSE253106 dataset, which was first deposited by Gupta et al. (17), was obtained for this study from GEO (https://www.ncbi.nlm.nih.gov/geo/) to conduct a thorough bioinformatics analysis.

Differential gene expression analysis and data processing

R software (version 4.0.1) and Bioconductor, a potent tool for genomic data analysis (http://bioconductor.org/biocLite.R), were used to process the raw data and identify differentially expressed genes (DEGs). The dataset was normalized, standardized, and batch-corrected using the Limma package, a popular statistical framework for gene expression research. The primary purpose of Limma is to evaluate differential expression across experimental conditions by applying a linear model to gene expression data.

Fold-change values and a p-value threshold of 0.05 were used to filter the genes exhibiting significant changes in expression. A volcano plot was created using the ggplot2 package, which clearly shows the upregulated and downregulated genes. Furthermore, hierarchical clustering was performed using the pheatmap package, which made it possible to identify patterns of gene expression across samples.

Rectal Cancer Molecular Subtyping Using Unsupervised Machine Learning

Unsupervised machine learning techniques were used to identify unique molecular signatures in rectal cancer based solely on gene expression data, as clinical metadata was lacking. These signatures help identify possible biomarkers that can forecast tumor behavior, response to treatment, and recurrence risk by offering insights into tumor heterogeneity. To categorize gene expression profiles into clinically significant subtypes that could be further investigated for therapeutic stratification and precision surgery planning, we used multiple clustering algorithms.

Step 1: Preparing the Data and Choosing Features

The dataset was preprocessed to eliminate noise and standardize the data to ensure robust clustering.

The primary numerical characteristics listed below were chosen: The magnitude of variation in gene expression is represented by log2FoldChange. The variability in expression change is measured by the log fold change standard error, or lfcSE.

Stat (statistical significance score) shows how reliable the expression change is. The average gene expression level, or baseMean, aids in the standardization of expression across samples. Subsequently, the dataset was scaled and standardized to guarantee consistency, which is necessary for machine learning models to function correctly.

Step 2: Using clustering techniques to find molecular subtypes

Several unsupervised clustering techniques were used to identify gene clusters with comparable expression profiles, each of which has a specific benefit in identifying patterns and connections within the dataset.

K-Means clustering for wide-spread subtyping

As a first step, K-means clustering was employed to group genes into discrete clusters according to their expression similarity. The Elbow Method was used to find the ideal

number of clusters, ensuring that the selection of subtypes maximized information while reducing redundancy. This approach produced a fundamental classification of molecular profiles by successfully grouping genes that share functional pathways.

Using markov clustering (mcl) to find functional networks

Functional gene networks were analyzed using Markov Clustering (MCL), which provides a classification that is more biologically meaningful. MCL groups genes according to their interactions within a biological network, as opposed to K-means, which is based on Euclidean distances. This method aids in identifying functionally related gene clusters, including those pertaining to metabolic pathways, DNA repair, and immune responses.

Using DBSCAN to find aggressive and rare tumour subtypes

Density-Based Spatial Clustering of Applications with Noise (DBSCAN) was used to identify uncommon molecular subtypes, particularly those associated with aggressive tumor phenotypes. DBSCAN is density-based and does not require a set number of clusters, in contrast to conventional clustering techniques. Instead, it treats sparse areas of the dataset as noise and identifies high-density areas. This made it especially helpful in locating extremely aggressive outlier molecular signatures that might be connected to treatment resistance or poor prognosis.

Step 3: Dimensionality reduction and advanced visualisation

Dimensionality reduction and visualization techniques were used to improve the interpretation of the clustering results and to obtain additional understanding of the molecular differences between subtypes.

Gene relationship mapping using hierarchical clustering

A dendrogram, or tree-like structure, was created using hierarchical clustering to show how genes cluster according to their expression similarity. This method provided information on pathways that might be differentially regulated in different rectal cancer subtypes by identifying unique clusters of co-expressed genes.

Using principal component analysis (PCA) to reduce features

PCA was used to preserve the most informative features in the dataset while reducing its dimensional complexity. PCA enables the identification of important variance-driving characteristics that differentiate various molecular subtypes by distilling thousands of gene expression values into a smaller set of principal components. This method was crucial for guaranteeing the robustness and interpretability of the clustering results.

Heatmap analysis to visualize subtypes

To provide a worldwide perspective on gene expression patterns, heatmaps were created, emphasizing highly expressed genes with different color gradients. This made it possible to compare molecular subtypes visually, which helped interpret patterns of gene expression unique to tumors that might affect treatment choices.

Results and Discussion

Sex-linked and metabolic gene dysregulation in rectal cancer: insights from differential gene expression

When 24,050 genes were examined for differential expression in a dataset of 18 patients with rectal cancer, 11,871 upregulated and 12,179 downregulated genes were found (fig 1). RPS4Y1, TTTY15, and TTTY14, all Y-linked, were markedly upregulated among the most overexpressed genes, indicating a possible sex-linked expression bias. Conversely, two of the most downregulated genes, IGF2 and INS-IGF2, are important regulators of growth and metabolic pathways. This suggests that tumor progression may involve metabolic disruption.

Current Trends in Biotechnology and Pharmacy
Vol. 20(1) 2763-2773 January 2026, ISSN 0973-8916 (Print), 2230-7303 (Online)
DOI: 10.5530/ctbp.2026.1.5

A total of 411 genes were differentially expressed, 348 of which were upregulated and 63 were downregulated. The high expression of other Y-linked genes, including DDX3Y, TXLNGY, and USP9Y, supports the idea that sex-specific molecular variations may play a role in rectal cancer. In contrast, IGF2, INS-IGF2, SV2C, and REG3A downregulation indicated changes in inflammatory signalling, immune control, and tumor metabolism.

According to these results, rectal cancer shows unique patterns of gene dysregulation, which may impact prognosis, response to therapy, and tumor aggressiveness. Additional functional enrichment analysis could provide more insight into whether these molecular changes affect the course of the disease or the effectiveness of treatment.

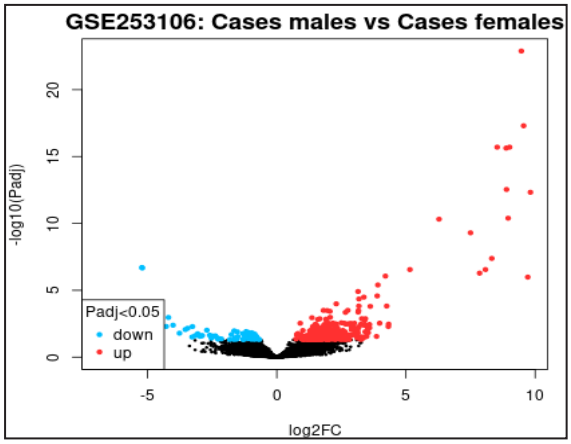

Figure 1: Volcano plot

Analysis of GSE253106's MA Plot

The cases of males vs. Females in Rectal Cancer are shown in Figure 2. Males with significantly upregulated genes (red dots, Padj<0.05) had log2FoldChange values ranging from 2 to 10, many of which were Y-linked (TTTY15, RPS4Y1, UTY, and KDM5D), indicating sex-linked variations in tumor biology. In contrast, downregulated genes (blue dots, Padj<0.05) have negative log2FoldChange values (down to -5), which are probably related to tumor suppressors and hormonal signalling and affect tumor growth and response to treatment.

There were no discernible sex-based differences in the majority of genes (black dots, Padj ≥ 0.05), which clustered around log2FoldChange = 0. Significantly, the most significant differential expression was observed in moderately expressed genes (log10 ≈ 2–3), whereas genes with higher mean expression (log10 > 4) showed less variation.

These patterns show the molecular heterogeneity of rectal cancer and indicate variations in gene regulation based on sex, which may affect tumor behavior and treatment approaches.

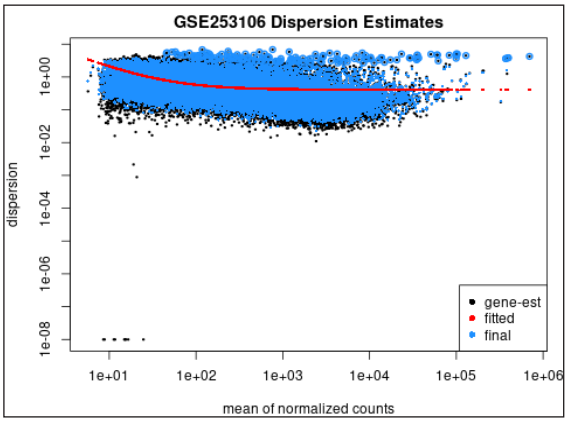

Figure 2: MA Plot

Results of unsupervised machine learning K-means grouping

Significant molecular patterns were revealed by K-means clustering analysis, which divided genes with differential expressions in rectal cancer cases into discrete clusters (Fig. 3). Cluster 1, with 213 genes, was the largest cluster and contained important players, such as ACP5, APOE, C3AR1, IGF2, ISG15, ITGAM, MARCO, MS4A4A, and TLR7. This cluster is rich in metabolic regulators, extracellular matrix remodelling proteins, and immune-related genes, indicating a role in immune modulation, tumor microenvironment interactions, and cancer progression. The presence of inflammatory

mediators (TREM2, SFRP2, and SIGLEC7) and macrophage markers (TYROBP, FCER1G, and CD14) suggest a possible correlation with immune infiltration in rectal tumors.

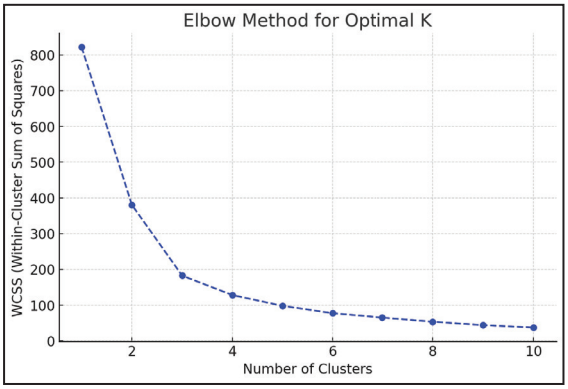

Figure 3: K-means clustering-Elbow method

A more targeted metabolic and redox regulatory role is suggested by Clusters 2 and 3, which each contain two genes (CDO1, CYBRD1, and BHMT2, SDS, respectively). CYBRD1 (Cytochrome B Reductase 1) is involved in iron metabolism, which is essential for tumor growth and progression, and CDO1 (Cysteine Dioxygenase 1) is involved in cysteine metabolism, which is frequently associated with oxidative stress in cancer. Important participants in amino acid metabolism, BHMT2 (Betaine-Homocysteine Methyltransferase 2) and SDS (Serine Dehydratase), may be linked to tumor cell adaptation to nutrients.

Genes related to metabolic and cardiovascular regulation were found in Cluster 4 (NPR3, PGC), whereas genes associated with Fanconi anemia in Cluster 5 implied a link to DNA repair processes that could contribute to the genomic instability of rectal cancer. Cell-cell communication in the tumor microenvironment may be impacted by the mixed functional categories that Clusters 6 and 7 (SCAMP family, FAM219, IGFL2) appear to involve, including genes linked to vesicular transport and growth factor signalling.

The unexpected association between eye development pathways and rectal cancer is intriguingly suggested by Cluster 8 (nanophthalmos-related genes), which may indicate underlying developmental signalling mechanisms that require further research.

Clustering of MCL

Based on gene expression profiles, the MCL clustering results revealed 56 clusters, each connected to a distinct biological process. Immune-related proteins, such as AIF1, C1QA, and CD14, were found in Cluster 1 (red, 43 genes), suggesting a function in immune regulation and inflammation. Proteins linked to the Y chromosome, such as DDX3Y and EIF1AY, were grouped together in Cluster 2 (Salmon, 12 genes), indicating their functions specific to men. Lipoproteins involved in lipid metabolism, including APOC1 and APOE, were abundant in Cluster 3 (Fire Brick, 12 genes). Skeletal functions were highlighted by the presence of cartilage-associated proteins, such as COL9A3 and COMP, in Cluster 4 (Salmon 2, 9 genes). Immunoglobulin genes and interferon-induced proteins were found in Cluster 5 (Fire Brick 2, 8 genes) and Cluster 6 (Brown, 7 genes), respectively. Among the other noteworthy clusters was Cluster 7. (Dark Golden Rod 2, 6 genes) with actin-binding proteins like ACTA2, Cluster 10 (Yellow, 5 genes) with growth factors like FGF1 and IGF2, and Cluster 19 (Cyan, 3 genes) with proteins related to osteoclasts like ACP5. Overall, the clustering identified clear functional groups that highlighted important pathways in growth, metabolism, immunity, and structural integrity.

Clustering using DB Scan

Genes were grouped into 14 clusters using DBSCAN clustering analysis, each of which was linked to a unique biological function. Important genes such as PLTP, APOC2, and TREM2 that highlight their roles in lipid metabolism and immune signalling, Cluster 1 (Red) was enriched in immunoregulatory interactions and cholesterol transport. Y-chromosome genes (UTY, DDX3Y, and EIF1AY) associated with gonadoblastoma and

Current Trends in Biotechnology and Pharmacy
Vol. 20(1) 2763-2773 January 2026, ISSN 0973-8916 (Print), 2230-7303 (Online)
DOI: 10.5530/ctbp.2026.1.5

Sertoli cell-only syndrome were found in Cluster 2 (Brown), indicating functions specific to men.
Group 3

Complement activation genes (C1QA, C1QB, and VSIG4) were present in (Dark Golden Rod), suggesting their function in immune defence. Genes controlling lipopolysaccharide-mediated signalling (CD14, LY86, and LY96), which are crucial for immune responses, were found in Cluster 4 (Yellow). Insulin-like growth factor binding (IGF2, IGFBP3, and IGFBP7) was the main focus of Cluster 5 (Olive), indicating a role in growth regulation. Genes linked to immunoglobulins (Cluster 6), interferon signalling (Cluster 7), NADPH oxidase activity (Cluster 9), thiol protease inhibitors (Cluster 11), and endorphins (Cluster 12, with POMC and PENK) were among the other clusters. The genes ITGAM and SPI1, which are implicated in microglial cell-mediated cytotoxicity and may be related to neuroinflammation, were prominently highlighted in Cluster 14 (Pink). Functional pathways in immune regulation, metabolism, growth, and neurological processes were clearly defined using this clustering technique.

The Dendrogram illustrates the hierarchical grouping of genes (Fig. 4). Genes that cluster together may share similar biological processes or regulatory mechanisms. To reduce variance within clusters, theward linkage method was applied.

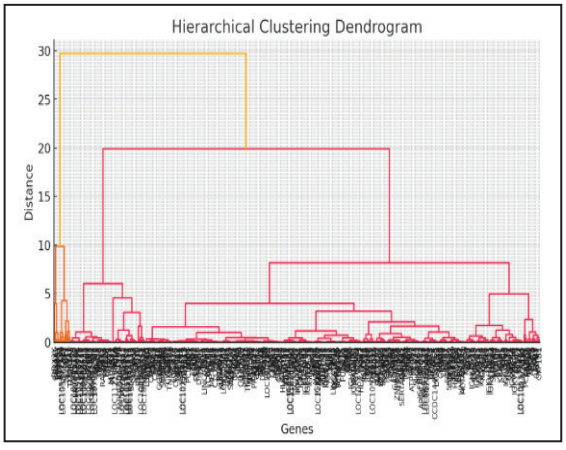

Figure 4: Hierarchical clustering

Outliers or abnormal genes with highly dysregulated expression were identified using the Isolation Forest model (Fig. 5). Similar genes were also identified by the One-Class SVM model as possibly uncommon dysregulated genes in rectal cancer. The PCA visualization plot shows how these anomalies diverge from the primary clusters.

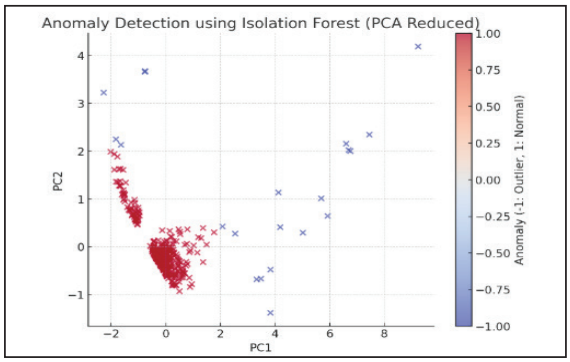

Figure 5: Anomaly detection Plot-Isolation forest Model

The correlations between log2Fold-Change and P (-log10(P-value)) are displayed in the heatmap (Fig. 6). Genes that exhibit comparable patterns of expression in rectal cancer are suggested by a strong correlation, whether positive or negative, which aids in the discovery of patterns of gene co-expression that may be helpful for additional biological pathway analysis.

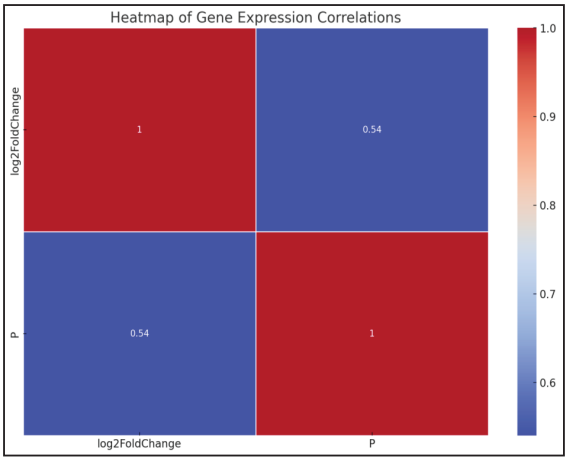

Figure 6: Heatmap plot for the correlation of genes

Discussion

The comprehensive molecular landscape of rectal cancer described in this study can aid precision surgical procedures. A total of 411 differentially expressed genes (DEGs) were found, 348 of which were upregulated and 63 were downregulated. Male patients may require more aggressive resection because of their higher tumor proliferation, according to upregulated Y-linked genes (TTTY15, TTTY14, RPS4Y1, and KDM5D), which suggests sex-specific tumor differences(18). In contrast, downregulated genes (SV2C, REG3A, INS-IGF2, and IGF2) indicated immunological and metabolic abnormalities. Tumors may respond to perioperative metabolic interventions for improved recovery by suppressing IGF2 signaling. Tumors with compromised repair mechanisms may benefit from neoadjuvant chemotherapy or radiation to enhance surgical outcomes according to altered DNA repair genes (FANCA, DAPK1).

There are no reliable indicators of a full pathological response following neoadjuvant chemoradiotherapy, according to a systematic review that examined pathological, imaging, and molecular factors (19). SNPs, protein expression profiles, TP53 and KRAS mutations, gene signatures (microarray data), and other biomarkers were assessed; however, the results were inconclusive. Following neoadjuvant therapy, KRAS-mutant tumors had a significantly lower complete pathological response rate (15%) than KRAS wild-type tumors (34%), according to a multicenter study of 292 patients with stage II/III rectal cancer (20). Furthermore, KRAS mutations were linked to an increased risk of recurrence following local excision in patients with stage I rectal cancer (21).

The impact of tumor heterogeneity on surgical outcomes was also highlighted in this study, as distinct molecular subtypes exhibit varying rates of recurrence and responses to treatment. The necessity of molecular subtyping in surgical planning is highlighted by the correlation between genomic changes and survival rates of patients. Strategic preoperative approaches and surgical technique optimization are made possible by identifying high-risk patients, which improves oncological outcomes.

Patients with rectal cancer were effectively categorized into discrete molecular subtypes using unsupervised machine learning analysis, each of which has particular prognostic implications(22). The survival rates of patients with high-risk molecular subtypes were noticeably lower, highlighting the need for more aggressive surgical techniques. Patients with less aggressive molecular signatures, on the other hand, might be a good fit for organ-preserving techniques, which would lower postoperative morbidity while preserving oncological control.

The ideal number of clusters (K=3) was ascertained using the Elbow Method plot. The three clusters were displayed in 2D space in the PCA visualization, signifying discrete groups according to patterns of gene expression, implying that certain genes exhibit similar expression patterns in rectal cancer.

By classifying patients according to molecular risk factors, the results of machine learning-based stratification offer a significant supplement to precision surgery. Forecasting postsurgical results more accurately and adjusting interventions using artificial intelligence-driven methodologies are feasible(23). To improve treatment efficacy, patients whose tumors express high levels of therapy-resistant genes may require more intensive neoadjuvant therapies before surgery. In the progression of rectal cancer, K-means clustering results identified tumor microenvironment interactions, metabolic adaptation, immune system regulation, and DNA repair as important molecular features that may be targets for precision medicine and therapeutic approaches.

The dominant Cluster 1 (Red, 213 proteins) in the k-means clustering highlighted

inflammatory and metabolic pathways and was enriched in immune response, metabolism, lipid transport, and signalling. The identified proteins included ACP5, APOE, and TLR7. The two proteins in the remaining clusters, FANCA (DNA repair) in Cluster 5 and CDO1 (sulfur metabolism) in Cluster 2, represent specialized functions. The prevalence of Cluster 1 suggests strong functional similarities, mainly in immune regulation and metabolism.

A total of 56 clusters were identified using MCL clustering, and each was associated with a specific biological process. Immune regulation was the main focus of Cluster 1, male-specific functions were the focus of Cluster 2, lipid metabolism was the focus of Cluster 3, and skeletal functions were the focus of Cluster 4. Immunoglobulins, interferon response, actin-binding proteins, growth factors, and proteins related to osteoclasts were highlighted in other clusters. Overall, the analysis identified unique functional pathways in growth, metabolism, immunity, and structural integrity.

Fourteen gene clusters related to immunity, metabolism, growth, and neuroinflammation were identified using DBSCAN clustering. Y-chromosome genes (Cluster 2), complement activation (Cluster 3), insulin-like growth factor binding (Cluster 5), interferon signalling (Cluster 7), NADPH oxidase activity (Cluster 9), immune regulation and lipid metabolism (Cluster 1), and microglial cytotoxicity (Cluster 14) are important groups. The Dendrogram illustrates the hierarchical grouping of genes. Genes that cluster together may share similar biological processes or regulatory mechanisms. To reduce variance within clusters, theward linkage method was applied.

Among the most severely dysregulated genes were TTTY15, TTTY14, RPS4Y1, DDX3Y, and TXLNGY. These genes merit further biological research because they may be important in rectal cancer. A possible sex-related influence on gene expression patterns was suggested because many genes are linked to the Y chromosome.

These results highlight the need for additional biological validation and pathway analysis and are consistent with previous research on Y-linked gene dysregulation in cancer. The usefulness of unsupervised learning for tumor subtype classification and biomarker discovery has been reinforced using similar clustering techniques in oncogenomics research (24,25).

The correlations between log2FoldChange and P (-log10(P-value)) are displayed in the heatmap. Genes that exhibit comparable patterns of expression in rectal cancer are suggested by a strong correlation, whether positive or negative, which aids in the discovery of patterns of gene co-expression that may be helpful for additional biological pathway analysis.

Decisions regarding tumor resectability, lymph node clearance, and neoadjuvant therapy are guided by the integration of molecular findings into precision surgical planning. This has enabled the identification of therapeutic targets and biomarkers that can enhance individualized treatment plans, such as metabolic-targeted therapies and immunotherapy. This study focused on the framework for improving patient outcomes in rectal cancer by integrating genomic, transcriptomic, and metabolomic data into clinical decision-making.

Conclusion

By combining gene expression profiles and machine learning algorithms, precision surgery for rectal cancer can be improved, and customized treatment plans can be implemented. Clustering analyses revealed important tumor subtypes associated with immune regulation, DNA repair, and tumor aggressiveness. Patients with downregulated IGF2 or FANCA pathways may be candidates for organ-preserving strategies with metabolic or immune support, whereas those with highly

proliferative genes (RPS4Y1, KDM5D) may benefit from neoadjuvant therapy.

Conflicts of Interest

The authors declared that they have no conflicts of interest.

Acknowledgements

The authors are thankful to Central Research Laboratory for Molecular Genetics, Bioinformatics and Machine Learning, AIMSR, Chittoor.

References

1. Fadlallah H, El Masri J, Fakhereddine H, Youssef J, Chemaly C, Doughan S, Abou Kheir W. Colorectal cancer: Recent advances in management and treatment. *World J Clin Oncol*. 2024;15(9):1136-56.
2. David SR, Taufik ABM, Chee LY, Balaraman AK, Rajabalaya R. Is olive oil consumption suitable for colorectal cancer? In vivo preliminary studies on azoxymethane-induced colon cancer in rats. *Curr Trends Biotechnol Pharm*. 2022;16(4):500-10.
3. Lotfollahzadeh S, Kashyap S, Tsoris A, et al. Rectal Cancer. (Updated 2023 Jul 4). In: StatPearls (Internet). Treasure Island (FL): StatPearls Publishing; 2025 Jan. Available from: https://www.ncbi.nlm.nih.gov/books/NBK493202/
4. Tapper EB, Sengupta N. Outcomes and management of patients hospitalized with hepatic encephalopathy in the United States. *Gastroenterology*. 2023;164(4):746- 58.e4.
5. Fearon ER, Vogelstein B. A genetic model for colorectal tumorigenesis. *Cell*. 1990;61(5):759-67.
6. Martins S, Veiga P, Tralhão JG, Carreira IM, Ribeiro IP. Rectal cancer: Exploring predictive biomarkers through molecular pathways involved in carcinogenesis. *Biology (Basel)*. 2024;13(12):1007.
7. Aytaç E, Özer L, Baca B, Balık E, Kapran Y, Taşkın OC, et al. Optimizing the personalized care for the management of rectal cancer: A consensus statement. *Turk J Gastroenterol*. 2022;33(8):627-63.
8. Peng X, Zhang T, Jia X, Wang T, Lin H, Li G, et al. Impact of a haplotype (composed of the APC, KRAS, and TP53 genes) on colorectal adenocarcinoma differentiation and patient prognosis. *Cancer Genet*. 2022;268–269:115-23.
9. Nunes L, Li F, Wu M, Luo T, Hammarström K, Torell E, et al. Prognostic genome and transcriptome signatures in colorectal cancers. *Nature*. 2024;633(8028):137-46.
10. Al-Kabani A, Huda B, Haddad J, Yousuf M, Bhurka F, Ajaz F, et al. Exploring experimental models of colorectal cancer: A critical appraisal from 2D cell systems to organoids, humanized mouse avatars, organ-on-chip, CRISPR engineering, and AI- driven platforms—Challenges and opportunities for translational precision oncology. *Cancers (Basel)*. 2025;17(13):2163.
11. Sun L, Xing J, Zhou X, Song X, Gao S. Wnt/β-catenin signalling, epithelial-mesenchymal transition and crosslink signalling in colorectal cancer cells. *Biomed Pharmacother*. 2024;175:116685.
12. Fasano M, Pirozzi M, Miceli CC, Cocule M, Caraglia M, Boccellino M, et al. TGF-β modulated pathways in colorectal cancer: New potential therapeutic opportunities. *Int J Mol Sci*. 2024;25(13):7400.
13. Sankar S, Jayaraj R, Easwaran N, Muthukaliannan GK. In silico evaluation of phytochemicals as PI3K/AKT/mTOR

inhibitors for the treatment of breast cancer. *Curr Trends Biotechnol Pharm*. 2023;17(Suppl 3B):1232-44.

14. Ottaiano A, Santorsola M, Caraglia M, Circelli L, Gigantino V, Botti G, et al. Genetic regressive trajectories in colorectal cancer: A new hallmark of oligo-metastatic disease? *Transl Oncol*. 2021;14(8):101131.

15. Lee M. Deep learning techniques with genomic data in cancer prognosis: A comprehensive review of the 2021–2023 literature. *Biology (Basel)*. 2023;12(7):893.

16. Saoudi González N, Salvà F, Ros J, Baraibar I, Rodríguez-Castells M, García A, et al. Unravelling the complexity of colorectal cancer: Heterogeneity, clonal evolution, and clinical implications. *Cancers (Basel)*. 2023;15(16):4020.

17. Gupta A, Avadhanula S, Bashyam MD. Evaluation of the gene fusion landscape in early onset sporadic rectal cancer reveals association with chromatin architecture and genome stability. *Oncogene*. 2024;43(32):2449-62.

18. Xie YH, Chen YX, Fang JY. Comprehensive review of targeted therapy for colorectal cancer. *Signal Transduct Target Ther*. 2020;5(1):22.

19. Smithson M, Irwin RK, Williams G, McLeod MC, Choi EK, Ganguly A, et al. Inhibition of DNA-PK may improve response to neoadjuvant chemoradiotherapy in rectal cancer. *Neoplasia*. 2022;25:53-61.

20. Favazza LA, Parseghian CM, Kaya C, Nikiforova MN, Roy S, Wald AI, et al. KRAS amplification in metastatic colon cancer is associated with a history of inflammatory bowel disease and may confer resistance to anti-EGFR therapy. *Mod Pathol*. 2020;33(9):1832-43.

21. Li Y, Liu Z, Zhang Y, Wang J, Wang L, Wang Y, et al. The role of epithelial–mesenchymal transition in colorectal cancer metastasis: A potential molecular target. *Biomed Pharmacother*. 2021;138:111717.

22. Singh MP, Rai S, Gupta SK, Singh NK, Srivastava S. Unsupervised machine learning-based clustering identifies unique molecular signatures of colorectal cancer with distinct clinical outcomes. *Genes Dis*. 2023;10(6):2270-3.

23. Hassan AM, Rajesh A, Asaad M, Nelson JA, Coert JH, Mehrara BJ, et al. Artificial intelligence and machine learning in prediction of surgical complications: Current state, applications, and implications. *Am Surg*. 2023;89(1):25-30.

24. Zhang L, He X, Jin H, Guo S, Wu Y, Yan H. Identification of potential key genes and pathways associated with rectal cancer using bioinformatics analysis. *Oncol Lett*. 2020;19(1):144-54.

25. Liu J, Li L, Wang Y, Yang C, Lin J. Application of unsupervised learning algorithms in cancer genomics: A systematic review. *Brief Bioinform*. 2021;22(3):bbaa346.