Content available at Bio-Research

Journal of Biological Research and Biotechnology

Journal Home page: https://bio-research.com.ng/index.php/home

ORIGINAL ARTICLE

Integrative Transcriptomics Reveals Diabetes-Associated Immune–Metabolic and Signalling Programs in Colorectal Cancer

[1*]**Henry Okuchukwu Ebili**

[1]Morbid Anatomy and Histopathology Department, Olabisi Onabanjo University, Sagamu Campus, Sagamu, Ogun State.

ARTICLE INFO

Received: 29th May 2026
Revised: 24th August 2026
Accepted: 26th August 2026
Published online: 31st August 2026
DOI:https://dx.doi.org/10.4314br.v24i2.22

e-ISSN: 2705-3822
print-ISSN: 1596-7409

Official. Faculty of Biological Sciences, University of Nigeria.

ABSTRACT

Diabetes mellitus (DM) and colorectal cancer (CRC) share metabolic and signalling alterations, suggesting that CRC transcriptomes may provide a systems-level framework for investigating DM-associated molecular programmes. Transcriptomic data from 988 CRC tumours across TCGA (n=534), CPTAC-2 (n=106), and Sidra-LUMC (n=348) cohorts were integrated. Five DM-related transcriptional signatures were quantified, and hyperglycaemia, insulin resistance, and mitochondrial dysfunction scores were integrated into a core composite DM-like score. Gene set enrichment analysis (GSEA), differential expression analysis, gene ontology enrichment analysis (GOEA), pathway ontology enrichment analysis (POEA), and disease ontology enrichment analysis (DOEA) were used to characterize the DM-like phenotype. CRC tumours exhibited heterogeneous DM-related transcriptional programmes. GSEA independently identified diabetes-associated pathways, including insulin resistance, β-cell destruction, type 1 diabetes, and immune-regulatory programmes. DM-like-high tumours showed enrichment of glucose and insulin signalling, mitochondrial dysfunction, and diabetes-associated clinical features and complications, including atherosclerosis, nephropathy, fatty liver disease, retinopathy, and neuropathy. CRC transcriptomes recapitulate multiple molecular features of DM and provide a systems-level framework for investigating diabetes-associated molecular alterations, biomarkers, and potential therapeutic targets.

Keywords: Colorectal cancer; Diabetes mellitus; Diabetes-like phenotype; Metabolic signature; Immune–metabolic interactions; Functional enrichment analyses; Microsatellite instability.

Cite this article: Henry Okuchukwu Ebili. (2026). Integrative Transcriptomics Reveals Diabetes-Associated Immune–Metabolic and Signalling Programs in Colorectal Cancer. (2026). *Journal of Biological Research and Biotechnology*, 24(2), 682-691

Corresponding author: Henry Okuchukwu Ebili. Email. ebili.henry@oouagoiwoye.edu.ng. **DOI**: https://dx.doi.org/10.4314br.v24i2.22

1.0 INTRODUCTION

Diabetes mellitus (DM) is a multifactorial metabolic disorder characterized by chronic hyperglycaemia resulting from defects in insulin secretion, insulin action, or both. The disease encompasses a spectrum of metabolic phenotypes, including insulin resistance, impaired glucose tolerance, and metabolic syndrome, all of which arise from complex interactions between genetic, environmental, and molecular factors (DeFronzo et al., 2015; Holt et al., 2021; Taylor, 2013). Despite advances in understanding DM pathogenesis, elucidating the systems-level molecular architecture underlying these phenotypes remains a major challenge. Cancer biology, particularly colorectal cancer (CRC), provides a unique lens through which metabolic dysregulation can be studied. CRC is characterized by profound metabolic reprogramming, including enhanced aerobic glycolysis (the Warburg effect), altered mitochondrial function, and dysregulated nutrient-sensing pathways (Vander Heiden et al., 2009; Pavlova & Thompson, 2016). These metabolic alterations are driven by oncogenic mutations and are essential for tumour growth and survival. Importantly, many of the pathways implicated in CRC metabolism--including insulin/IGF signalling, PI3K/AKT/mTOR signalling, and AMP-activated protein kinase (AMPK) pathways--are also central to the development and progression of DM (Hopkins et al., 2016; Caturano et al., 2025; Saxton & Sabatini, 2017). This overlap suggests that CRC may serve as a biologically relevant human model for studying metabolic phenotypes associated with diabetes. Unlike traditional experimental systems, clinical CRC datasets capture the complexity of human genetic variation, epigenetic regulation, and environmental exposure. Furthermore, epidemiological evidence supports a strong link between CRC and metabolic disorders. Individuals with type 2 diabetes have an increased risk of developing CRC, while obesity, hyperinsulinemia, and chronic inflammation--hallmarks of metabolic syndrome--are established risk factors for colorectal tumorigenesis (Giovannucci et al., 2010; Sun et al., 2022). Conversely, CRC tumours frequently exhibit transcriptional and metabolic signatures that are linked to systemic metabolic dysregulation (Gutiérrez-Salmerón, et al., 2021). At the molecular level, insulin resistance and cancer share common signalling disruptions, including aberrant activation of the PI3K/AKT pathway, altered glucose transporter expression, and dysregulation of key metabolic enzymes (Tsilidis et al., 2015). These shared mechanisms reinforce the concept that oncogenic processes may recapitulate or amplify metabolic states observed in diabetes. Therefore, analysing CRC transcriptomes may reveal gene expression signatures reflective of underlying metabolic phenotypes such as glucose intolerance and insulin resistance. The availability of large-scale, high-throughput datasets has enabled integrative analyses across diverse patient cohorts. Resources such as The Cancer Genome Atlas (TCGA), the Sidra-Leiden University Medical Center (Sidra-LUMC) cohort, and the Clinical Proteomic Tumor Analysis Consortium (CPTAC2) provide comprehensive molecular profiles of CRC, including transcriptomic and proteomic data (Cancer Genome Atlas Network, 2012; Vasaikar et al., 2019; Roelands et al., 2023). These datasets offer an unprecedented opportunity to investigate the convergence between cancer-associated and diabetes-associated molecular pathways in human tissues. Gene set-based analytical frameworks are particularly well-suited for this purpose. Gene Set Enrichment Analysis (GSEA) enables the identification of coordinated changes in predefined gene sets, while Gene Ontology

Enrichment Analysis (GOEA) provides functional annotation of differentially expressed genes (Subramanian et al., 2005; Ashburner et al., 2000). Additionally, single-sample GSEA (ssGSEA) allows for the quantification of pathway activity at the level of individual samples, facilitating the characterization of metabolic heterogeneity within and across cohorts (Barbie et al., 2009).

The molecular overlap between CRC and DM provides the biological rationale for investigating whether CRC transcriptomic heterogeneity can be exploited to model DM-associated molecular phenotypes. However, the presence of shared metabolic and signalling pathways does not, by itself, establish that CRC recapitulates DM. We therefore hypothesized that coordinated transcriptional programmes representing major features of DM can be identified and quantified within CRC transcriptomes and used to interrogate molecular alterations associated with DM biology. To test this hypothesis, we integrated transcriptomic data from three independent CRC cohorts and quantified transcriptional signatures representing hyperglycaemia, insulin resistance, lipid metabolism, mitochondrial dysfunction, and metabolic syndrome. These signatures were subsequently integrated and interrogated using complementary transcriptomic and enrichment analyses to determine whether CRC can provide a systems-level framework for studying DM-associated molecular alterations.

2.0 MATERIALS AND METHODS

2.1 Study design

This study employed an integrative transcriptomic framework to determine whether CRC can serve as a systems-level model for metabolic phenotypes associated with diabetes mellitus. Multiple independent primary CRC cohorts (Cancer Genome Atlas Network, 2012; Vasaikar et al., 2019; Roelands et al., 2023) were analysed, and gene set-based approaches were used to construct, quantify, and validate diabetes-related transcriptional signatures. The complementary enrichment approaches were applied sequentially to address distinct analytical objectives. Sample-level gene-set scoring was first used to quantify the activity of curated DM-related transcriptional programmes in individual CRC tumours and to derive the composite DM-like phenotype. GSEA was subsequently used as an independent gene-ranking approach to determine whether genes associated with the derived DM-like phenotype were enriched for established DM-related gene sets. Differentially expressed genes between DM-like-high and DM-like-low tumours were then subjected to GOEA, POEA, and DOEA to characterize the biological processes, signalling pathways, and disease/clinical phenotypes associated with the DM-like transcriptional state. Thus, these analyses were complementary rather than redundant, progressing from phenotype quantification, through independent enrichment-based validation, to functional and disease-level characterization.

2.2 Data acquisition and processing

Transcriptomic datasets for three independent primary CRC transcriptomic cohorts, including TCGA COAD/READ (n=534) (Cancer Genome Atlas Network, 2012), CPTAC-2 CRC (n=106) (Vasaikar et al., 2019), and Sidra-LUMC AC-ICAM (n=348) (Roelands et al., 2023) were obtained from cBioPortal (Cerami et al., 2012; Gao et al., 2013). These datasets were integrated to increase statistical power and capture molecular heterogeneity across independently generated CRC datasets. Gene expression data were provided as RSEM-normalized values (Li & Dewey, 2011). Gene identifiers were harmonised, and analyses were restricted to genes shared across all three datasets. Because the datasets originated from independent cohorts and platforms, principal component analysis (PCA) was performed on the combined expression matrix before batch correction to assess cohort-related variation. The matrix was subsequently corrected for cohort effects using ComBat (Johnson et al., 2007) implemented in the sva R package (Leek et al., 2012), and principal component analysis (PCA) was repeated after correction to evaluate the effectiveness of harmonisation. Downstream analyses were performed on the ComBat-corrected expression matrix.

2.3 Construction of diabetes-related metabolic signatures

Transcriptional signatures were constructed for Hyperglycaemia, Insulin resistance, Mitochondrial dysfunction, Lipid metabolism, and Metabolic syndrome. Diabetes-related gene sets were curated from Gene Ontology Biological Process (GO:BP) (Gene Ontology Consortium, 2026), Kyoto Encyclopedia of Genes and Genomes (KEGG) (Kanehisa et al., 2025), and Reactome pathway (Milacic et al., 2024) databases, supplemented with literature-derived markers (Subramanian et al., 2005; Ashburner et al., 2000; Pavlova & Thompson, 2016). Each phenotype comprised paired upregulated and downregulated modules. Gene sets were curated to minimize redundancy, and candidate genes were filtered against the expression matrix and uniquely assigned to individual phenotypes to prevent score inflation from shared genes. Signature genes are listed in Supplementary Table 1.

2.4 ssGSEA-based phenotype scoring

Diabetes-related transcriptional modules were scored independently in each tumour using a custom weighted rank-based running-sum algorithm adapted from the single-sample gene-set enrichment framework (Barbie et al., 2009). For each sample, genes were ranked in decreasing order of expression. To permit non-negative weighting while preserving the within-sample expression ranking, expression values were shifted by subtracting the minimum expression value and adding 10^{-12}. For each gene set, genes belonging to the set were assigned hit weights proportional to the shifted expression magnitude raised to α=0.25, normalized by the sum of hit weights, whereas genes outside the set received equal miss weights of $1/N_m$. A running enrichment statistic was calculated as the cumulative difference between hit and miss probabilities across the ranked gene list. The module enrichment score was defined as the signed maximum absolute deviation of this running statistic and normalized by the square root of the number of detected genes in the module. Module scores were subsequently Z-standardized across samples. For each DM-related phenotype, the final score was calculated as the difference between the standardized upregulated and downregulated module scores:

Phenotype score = Z(upregulated module) − Z(downregulated module)

Thus, positive scores indicate relative predominance of the transcriptional programme expected to increase with the corresponding DM phenotype, whereas negative scores indicate relative predominance of the opposing module. Signature genes were protected from low-expression filtering. Module coverage was assessed before scoring, and modules were considered usable when at least three genes and ≥50% of the original module were represented in the expression matrix. The complete R implementation is provided as Supplementary Script 1.

2.5 Composite DM-like Signature

A composite DM-like score was constructed as the mean of Hyperglycaemia, Insulin resistance, and Mitochondrial dysfunction scores:

DM-like score = mean(Z(Hyperglycaemia), Z(InsulinResistance), Z(MitoDysfunction))

Lipid metabolism and metabolic syndrome were excluded to allow independent evaluation of downstream processes.

2.6 Correlation Analysis

Pairwise phenotype score relationships were assessed using Spearman correlation with Benjamini-Hochberg FDR-adjusted p correction for multiple testing.

2.7 Differential Expression Analysis

Differential expression between phenotype-defined subsets (DM-like high vs low) was performed using limma with empirical Bayes moderation (Smyth, 2004; Ritchie et al., 2015). Genes with adjusted $p < 0.05$ were considered significant.

2.8 Gene Set Enrichment Analysis (GSEA)

GSEA was conducted using two complementary approaches: median dichotomisation of the core DM-like score and continuous pre-ranking of genes according to their Spearman correlation with the DM-like score. Enrichment analysis was performed using MSigDB GSEA v4.3.3 (Subramanian et al., 2005). Diabetes-related gene sets were curated from the Enrichr gene-set resource (https://maayanlab.cloud/Enrichr/) (Chen et al., 2013; Kuleshov et al., 2016; Xie et al., 2021). Candidate gene sets were filtered to remove redundancy and those with substantial overlap with signature-construction genes, yielding 57 non-redundant diabetes-related gene sets for enrichment analysis (Supplementary Table 1b). Significance was defined as nominal $p < 0.05$ and FDR q < 0.25 (Subramanian et al., 2005).

2.9 Functional enrichment analyses

Gene ontology enrichment analysis (GOEA) was performed on upregulated genes using GO:BP on Enrichr (Chen et al., 2013; Kuleshov

et al., 2016; Xie et al., 2021). Complementary Pathway ontology enrichment analysis (POEA) and Disease ontology enrichment analysis (DOEA) were conducted using DisGeNET and Elsevier Pathway Collection 2024 libraries to evaluate disease-associated and complication-related pathways, including vascular, inflammatory, and metabolic processes.

2.10 Integration and comparative analysis

Signature scores and enrichment results were compared across cohorts to identify consistent metabolic patterns, focusing on phenotype coherence, composite DM-like states, and enrichment of diabetes-related pathways and complications.

2.11 Statistical analysis and data visualisation

Clinicopathological association and survival analyses were performed using SPSS v29, while transcriptomic and enrichment analyses were conducted in R. Where applicable, multiple-testing correction was performed using the Benjamini–Hochberg procedure, with FDR-adjusted $p < 0.05$ considered significant. For GSEA, significance was defined as nominal $p < 0.05$ and FDR q < 0.25, using the conventional GSEA significance threshold (Subramanian et al., 2005). Visualisations were generated using SRPlot (https://www.bioinformatics.com.cn/en) (Tang et al., 2023).

3.0 RESULTS

3.1 ComBat correction attenuates cohort-related transcriptional variation

PCA before ComBat correction demonstrated pronounced cohort-dependent separation, with PC1 accounting for 53.4% of total variance. Following ComBat correction, the three cohorts showed substantial overlap, while the variance explained by PC1 decreased to 12.0%, indicating marked attenuation of the dominant cohort-related structure while retaining inter-sample transcriptional heterogeneity (Figure 1).

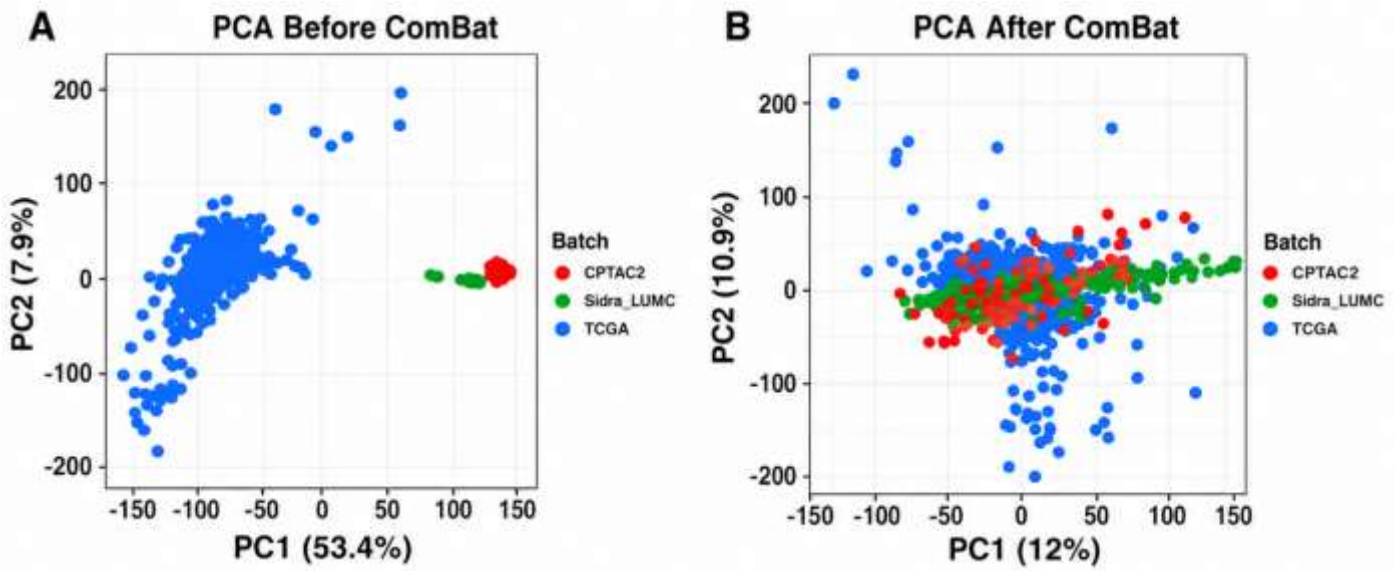

Figure 1: Principal component analysis before and after multicohort batch correction. PCA of the integrated TCGA, CPTAC2, and Sidra-LUMC CRC transcriptomes before ComBat correction showed pronounced cohort-dependent separation, with PC1 explaining 53.4% of the variance. Following ComBat correction, cohort separation was markedly reduced, with substantial overlap among the three cohorts and PC1 explaining 12.0% of the variance, demonstrating attenuation of dominant cohort-related effects.

3.2 Construction and distribution of diabetes-related metabolic signatures

Transcriptional signatures representing key diabetes-related metabolic phenotypes - hyperglycaemia, insulin resistance, mitochondrial dysfunction, lipid metabolism, and metabolic syndrome - were successfully derived across the combined CRC cohort. Each phenotype was quantified using a paired module ssGSEA-based approach, yielding continuous, standardized scores for all tumours. All metabolic signatures were centred around zero following Z-standardization but demonstrated substantial variability and wide dynamic ranges across samples. Several signatures exhibited skewed distributions and high kurtosis, indicating the presence of extreme values and supporting marked inter-tumoural heterogeneity in metabolic phenotypes (Figure 2, Supplementary Table 2).

3.3 Correlation structure reveals a metabolic syndrome-centred axis

Pairwise correlation analysis of phenotype scores revealed a structured network of metabolic relationships. Metabolic syndrome exhibited significant positive correlations with hyperglycaemia ($\rho \approx 0.25$, FDR-adjusted $p < 10^{-14}$), mitochondrial dysfunction ($\rho \approx 0.39$, FDR-adjusted $p < 10^{-35}$), insulin resistance ($\rho \approx 0.17$, FDR-adjusted $p < 10^{-7}$). Hyperglycaemia also correlated strongly with mitochondrial dysfunction ($\rho \approx 0.41$, FDR-adjusted $p < 10^{-39}$), supporting the biological coupling between glucose dysregulation and mitochondrial stress. In contrast, lipid metabolism showed weak or negative correlations with other phenotypes, indicating relative independence from the core metabolic axis (Figure 3). Collectively, these results identify a metabolic syndrome-centred transcriptional axis, linking hyperglycaemia, mitochondrial dysfunction, and insulin resistance, and suggest that this axis may represent the dominant diabetes-related program in CRC.

3.4 Derivation of a Composite DM-like transcriptional signature

Based on the observed correlation structure and biological relevance, a composite DM-like score was constructed from hyperglycaemia, insulin resistance and mitochondrial dysfunction signature. The resulting DM-like signature demonstrated a continuous distribution across tumours, enabling stratification into DM-like high and DM-like low subsets (Figure 4). Importantly, lipid metabolism and metabolic syndrome were not included in the composite score, allowing independent evaluation of downstream metabolic processes.

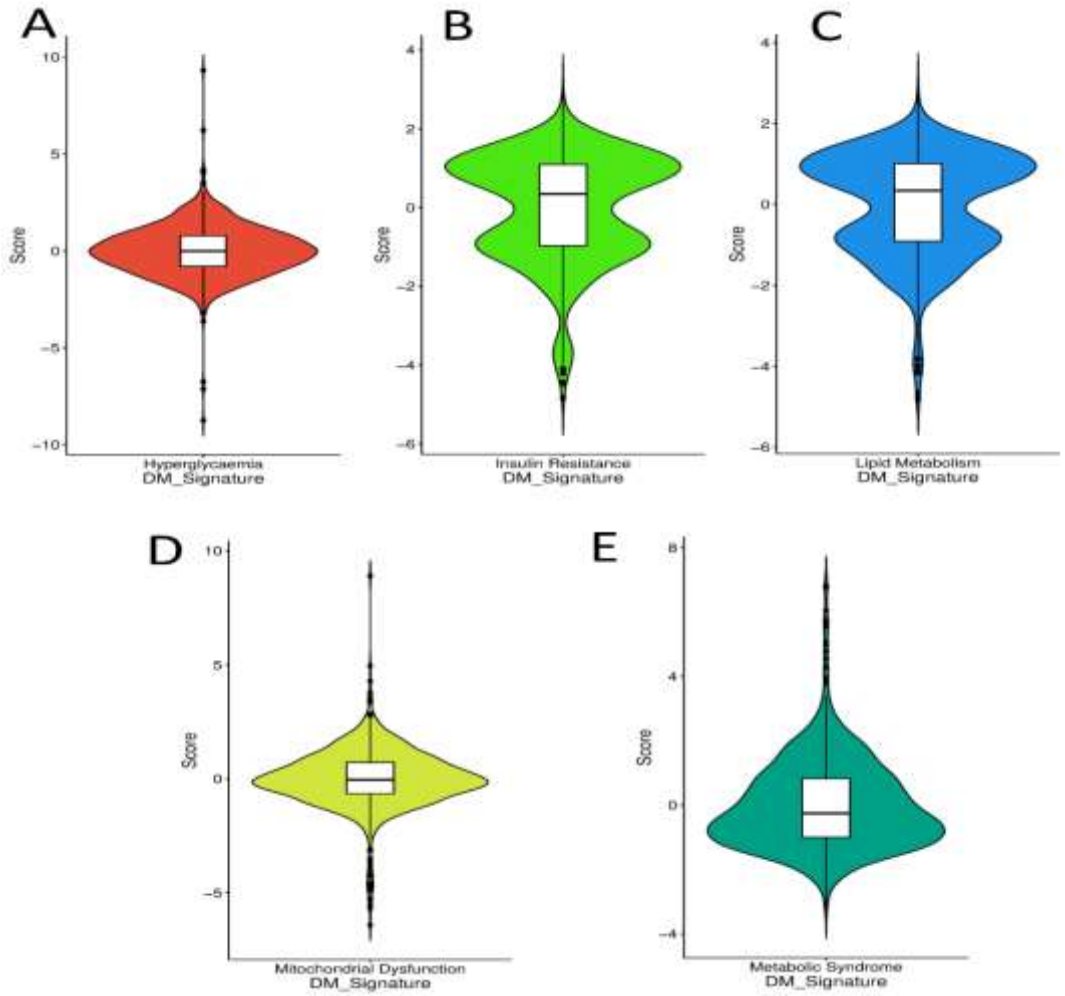

Figure 2. Distribution of diabetes-related metabolic signatures in CRC. Violin plots showing sample-level scores for (A) hyperglycaemia, (B) insulin resistance, (C) lipid metabolism, (D) mitochondrial dysfunction, and (E) metabolic syndrome. Embedded boxplots indicate the median and interquartile range. The distributions demonstrate substantial inter-tumoural heterogeneity in diabetes-related transcriptional programmes.

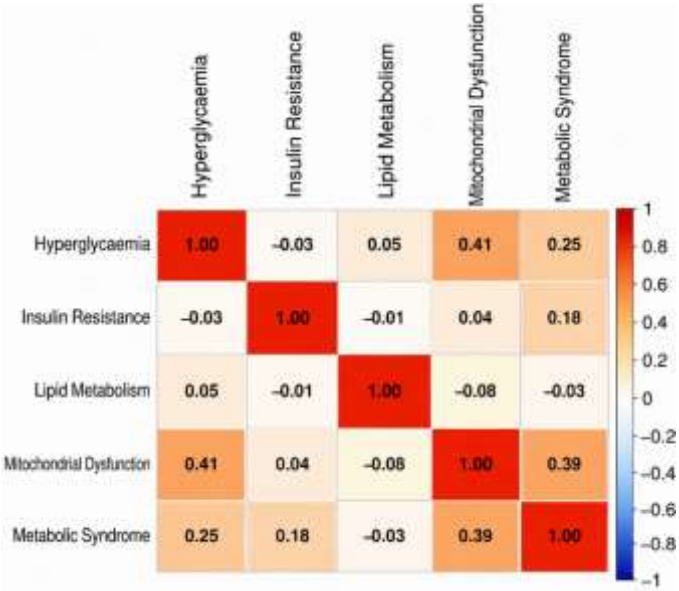

Figure 3. Correlation matrix of diabetes-related metabolic signatures in CRC. Heatmap showing pairwise Spearman correlations (ρ) among hyperglycaemia, insulin resistance, lipid metabolism, mitochondrial dysfunction, and metabolic syndrome signatures. The strongest positive correlations were observed between hyperglycaemia and mitochondrial dysfunction (ρ = 0.41) and between mitochondrial dysfunction and metabolic syndrome (ρ = 0.39), whereas the remaining associations were weak to modest.

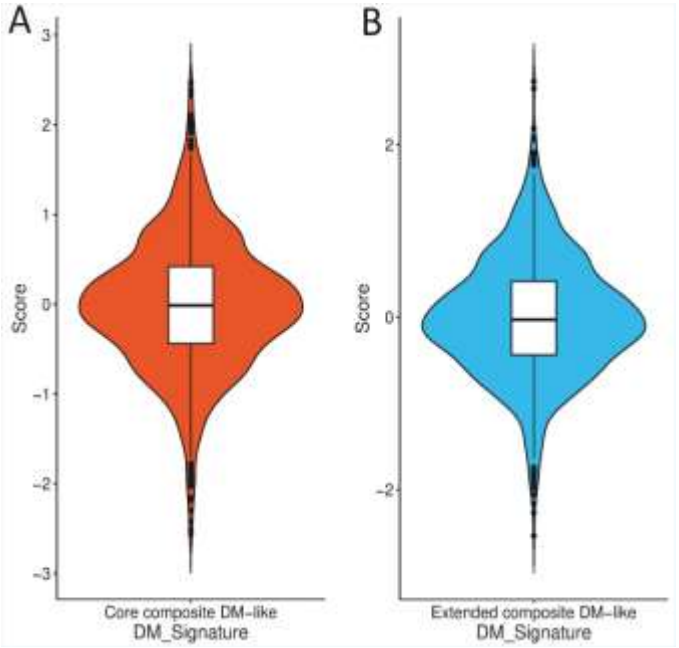

Figure 4. Distribution of composite diabetes-like (DM-like) transcriptional scores in CRC. Violin plots showing (A) core and (B) extended composite DM-like scores. Embedded boxplots indicate the median and interquartile range. Both scores show broad continuous distributions, demonstrating inter-tumoural heterogeneity in DM-like transcriptional states.

3.5 Clinicopathological and Molecular/Genomic Correlates of Diabetes-Related Phenotypes in Colorectal Cancer

a. Associations with clinicopathological features

Nonparametric analyses revealed that diabetes-related metabolic phenotypes demonstrated selective but biologically meaningful associations with clinicopathological characteristics (Supplementary Table 3A). Hyperglycaemia showed broad clinicopathological and molecular associations. Clinically, hyperglycaemia scores were significantly higher in right-sided tumours than in other tumour locations ($p < 0.001$), in mucinous adenocarcinomas relative to non-mucinous tumours ($p < 0.001$), in node-positive (N+) tumours (p = 0.001), and in stage III/IV tumours ($p < 0.001$). In contrast, insulin resistance showed no significant associations with clinicopathological variables. Mitochondrial dysfunction showed the broadest and most consistent clinicopathological associations among the individual phenotypes. Scores were significantly higher in older patients ($p = 0.017$), in right-sided tumours ($p = 0.044$), in mucinous adenocarcinomas ($p = 0.002$), in pT3/pT4 tumours ($p = 0.027$), in node-positive tumours ($p = 0.011$), and in stage III/IV disease ($p < 0.001$). Metabolic syndrome also showed a strong and coherent pattern of associations. Clinically, metabolic syndrome scores were significantly higher in right-sided tumours ($p < 0.001$), in mucinous adenocarcinomas ($p < 0.001$), in pT3/pT4 tumours ($p < 0.001$), and in stage III/IV disease ($p = 0.020$). Composite scores demonstrated enhanced clinical relevance. Clinically, core DM-like scores were significantly higher in right-sided tumours ($p < 0.001$), mucinous tumours ($p = 0.001$), pT3/pT4 tumours ($p = 0.048$), and stage III/IV disease ($p = 0.024$), while the extended DM-like signature displayed the strongest overall clinicopathological profile. It was significantly higher in right-sided tumours ($p < 0.001$), mucinous adenocarcinomas ($p < 0.001$), pT3/pT4 tumours ($p = 0.010$), node-positive tumours (p = 0.040), and stage III/IV tumours ($p < 0.001$). Lipid metabolism showed limited clinicopathological associations. Among clinicopathological variables, lipid metabolism was significantly lower in metastatic (M1) tumours ($p = 0.027$), while no significant associations were observed with age, sex, tumour location, histology, tumour stage, nodal stage, or overall stage. Collectively, these findings indicate that diabetes-related metabolic phenotypes are differentially linked to tumour progression and anatomical characteristics, with mitochondrial dysfunction and metabolic syndrome showing the strongest clinicopathological associations.

b. Associations with Molecular and Genomic Features

Analysis of molecular and genomic features demonstrated robust and widespread associations across multiple diabetes-related phenotypes (Supplementary Table 3B). Hyperglycaemia was significantly higher in *TP53*-mutant tumours ($p = 0.001$), *BRAF*-mutant tumours ($p < 0.001$), MSI-H tumours ($p < 0.001$), and in the Hypermutated/MSI molecular subtype compared with other subtypes ($p < 0.001$). In contrast to these positive associations, hyperglycaemia was significantly lower in tumours with high aneuploidy ($p < 0.001$) and lower in tumours with high fraction genome altered (FGA) ($p = 0.001$), but significantly higher in tumours with high tumour mutational burden (TMB) ($p < 0.001$). Mitochondrial dysfunction was significantly higher in BRAF-mutant tumours ($p < 0.001$), MSI-H tumours ($p < 0.001$), and the Hypermutated/MSI subtype ($p < 0.001$). In contrast, mitochondrial dysfunction was significantly lower in tumours with high aneuploidy ($p = 0.004$) and lower in tumours with high FGA ($p = 0.045$). No significant association with TMB was observed. Metabolic syndrome was significantly higher in BRAF-mutant tumours ($p < 0.001$), MSI-H tumours ($p < 0.001$), and the Hypermutated/MSI subtype ($p < 0.001$). At the same time, metabolic syndrome was significantly lower in tumours with high aneuploidy ($p < 0.001$) and lower in tumours with high FGA (p = 0.001), but *BRAF* significantly higher in tumours with high TMB ($p < 0.001$). Lipid metabolism significantly higher in *TP53*-mutant tumours (p = 0.044), -mutant tumours ($p < 0.001$), MSI-H tumours ($p < 0.001$), and the Hypermutated/MSI subtype relative to other molecular subtypes ($p < 0.001$). Lipid metabolism was significantly lower in tumours with high aneuploidy ($p = 0.011$) and lower in tumours with high FGA ($p < 0.001$), but significantly higher in tumours with high TMB ($p < 0.001$). Insulin resistance was significantly higher in MSI-H tumours than in MSS tumours ($p = 0.001$), and it also varied significantly by molecular subtype, with the Hypermutated/MSI subtype showing higher scores than other subtypes ($p = 0.001$). No significant associations were observed with *TP53* mutation, *BRAF* mutation, aneuploidy, FGA, or TMB. The core DM-like signature was significantly higher in *TP53*-mutant tumours ($p < 0.001$), *BRAF*-mutant tumours ($p < 0.001$), MSI-H tumours ($p < 0.001$), and the Hypermutated/MSI subtype ($p < 0.001$). Like several individual phenotypes, the core signature was significantly lower in tumours with high aneuploidy ($p < 0.001$) and lower in tumours with high FGA ($p < 0.001$), but significantly higher in tumours with high TMB ($p < 0.001$). The extended DM-like signature was significantly higher in *TP53*-mutant tumours ($p = 0.013$), *BRAF*-mutant tumours ($p < 0.001$), MSI-H tumours ($p < 0.001$), and the Hypermutated/MSI subtype ($p < 0.001$). It was also significantly lower in tumours with high aneuploidy ($p < 0.001$) and lower in tumours with high FGA ($p < 0.001$), while being significantly higher in tumours with high TMB ($p < 0.001$).

c. Survival Analysis

In univariable Cox regression analysis for disease-free survival (DFS) (Supplementary Table 3C), hyperglycaemia was associated with poorer outcomes (HR = 1.11, 95% CI: 1.01-1.23, $p = 0.033$), while lipid metabolism was associated with improved DFS (HR = 0.89, 95% CI: 0.80-0.98, $p = 0.025$). Mitochondrial dysfunction showed a borderline association with worse DFS ($p = 0.057$). In multivariable analysis, the overall model was significant ($\chi^2 = 11.52$, $p = 0.042$), with lipid metabolism remaining an independent predictor of improved DFS (HR = 0.89, 95% CI: 0.81-0.99, $p = 0.027$). No significant associations were observed for overall survival across any of the metabolic phenotypes.

d. Integrated Interpretation

Taken together, diabetes-related metabolic phenotypes showed associations with both clinicopathological features and molecular/genomic characteristics in CRC. Composite DM-like signatures exhibited the most consistent associations across these domains, integrating transcriptional, clinicopathological, and molecular features of the CRC cohort.

3.6 DM-like Signature Captures Diabetes-Associated Gene Sets

To evaluate the biological validity of the composite metabolic phenotype, GSEA was performed using two complementary approaches: (i) median dichotomization of the core DM-like score and (ii) continuous ranking of genes based on their Spearman correlation with the DM-like score. Enrichment significance was defined as nominal p-value < 0.05 and FDR q-value < 0.25.

a. Enrichment in the DM-like High State and Positive DM-like Axis

Across both dichotomized and continuous analyses, the DM-like phenotype was consistently associated with strong enrichment of diabetes-related transcriptional programs, particularly those linked to type 1 diabetes, immune regulation, and insulin resistance (Supplementary Table 4 and Supplementary Table 5). The most prominent enrichments were observed for type 1 diabetes-associated pathways, including type 1 diabetes mellitus (BioPlanet 2019) (NES = 2.27, FDR $q < 0.001$), type I diabetes mellitus (KEGG 2026) (NES = 2.10, FDR $q < 0.001$), and beta-cell destruction in type 1 diabetes mellitus (NES ≈ 2.09-2.14, FDR $q \leq 0.002$). Additional significant enrichment was observed for type 1A diabetes mellitus (Jensen_DISEASES_Curated 2025) (NES = 1.94, FDR $q = 0.002$) and type 1B diabetes mellitus (Jensen_DISEASES_Experimental 2025) (NES = 1.64, FDR $q = 0.028$), demonstrating a consistent and dominant type 1 diabetes-like transcriptional signature within the DM-like high state. Immune-regulatory pathways were also strongly enriched, including Treg-cell activation in diabetes mellitus (NES ≈ 1.8-1.9, FDR $q \leq 0.006$) and pathways related to altered susceptibility to autoimmune diabetes (decreased susceptibility: NES ≈ 1.85, FDR q = 0.005; increased susceptibility: NES ≈ 1.62, FDR $q = 0.033$). These findings indicate that the DM-like phenotype incorporates a substantial immune-mediated component, consistent with autoimmune diabetes biology. Metabolic pathways were concurrently enriched. Significant enrichment was observed for insulin resistance in myocytes induced by obesity (NES ≈ 1.69-1.81, FDR $q \leq 0.016$) and proteins involved in insulin resistance (NES ≈ 1.61-1.73, FDR $q \leq 0.025$), indicating activation of insulin resistance-related processes. In addition, advanced glycation end product (AGE)-related pathways were significantly enriched (NES = 1.70, nominal p = 0.007, FDR $q = 0.017$), linking the DM-like phenotype to downstream metabolic complications of diabetes. Beyond these core pathways, broader metabolic categories--including proteins involved in type 2 diabetes

mellitus (NES ≈ 1.31-1.44, FDR $q = 0.113$-0.191) and metabolic syndrome (NES ≈ 1.27-1.39, FDR $q = 0.145$-0.226) - also met the predefined significance threshold, albeit with lower enrichment magnitude. This indicates that the DM-like phenotype encompasses elements of type 2 diabetes and systemic metabolic dysfunction, but these are less dominant than the type 1 diabetes-associated and immune pathways. Importantly, these enrichment patterns were highly concordant between dichotomized and continuous analyses, demonstrating that diabetes-related pathways are progressively activated along the DM-like continuum rather than being restricted to arbitrarily defined high/low groups.

b. Enrichment in the DM-like Low State and Negative DM-like Axis

In contrast, the negative direction of the DM-like axis was characterized by significant enrichment of monogenic diabetes pathways, specifically those related to maturity-onset diabetes of the young (MODY) (Supplementary Table 5). Genes negatively correlated with the DM-like score were significantly enriched for MODY (Jensen_DISEASES_Curated 2025) (NES = −1.77, nominal p = 0.002, FDR $q = 0.086$) and MODY (BioPlanet 2019) (NES = −1.68, nominal p = 0.007, FDR $q = 0.097$), both of which met the predefined significance thresholds. These pathways reflect β-cell intrinsic dysfunction and genetically driven diabetes mechanisms, in contrast to the immune and metabolic processes enriched in the positive DM-like axis. Other negatively enriched pathways, including monogenic diabetes, type II diabetes mellitus, and insulin resistance-related gene sets, showed weaker enrichment (NES range: −1.22 to −0.81) and did not meet statistical significance (FDR $q > 0.25$), indicating that the negative axis is specifically driven by MODY-related biology rather than generalized depletion of diabetes pathways.

c. Integrated Interpretation

Taken together, these analyses demonstrate that the core DM-like phenotype captures a continuous, bidirectional diabetes-related transcriptional axis within colorectal cancer. The positive DM-like axis is characterized by coordinated activation of pathways associated with type 1 diabetes and autoimmune β-cell destruction, immune regulation (including Treg-associated processes), insulin resistance and metabolic dysregulation and glycation and downstream metabolic complications. In contrast, the negative DM-like axis is defined by enrichment of MODY-related pathways, reflecting β-cell intrinsic and monogenic diabetes mechanisms. The concordance between the dichotomized and continuous GSEA approaches indicates that these enrichment patterns are detectable using both categorical and continuous representations of the DM-like phenotype.

3.7 Differential expression analysis and GOEA reveals DM-associated mechanistic processes, clinical features and complication signatures

Differential expression analysis identified a large set of genes significantly dysregulated between DM-like-high and DM-like-low CRC cases (Supplementary Table 6). To focus on the most biologically relevant signals, genes were ranked based on the DE metric, and subsets representing the most strongly upregulated (ranking metric ≥ 0.1) and downregulated (ranking metric ≤ −0.1) genes, with FDR-adjusted $p < 0.05$, were selected. These filtered gene sets were subsequently used as input for gene ontology enrichment analysis (GOEA). The analysis focused on extracting the ontology terms relevant to DM-related pathogenetic features, i.e., mechanistic processes, clinical features and DM-related complications, that were significantly enriched by the GOEA (Supplementary Table 7).

a. Gene Ontology and Pathway Enrichment Analysis of the Core DM-like High Subset

Gene ontology and pathway enrichment analysis of genes upregulated in the core DM-like high subset showed that the diabetes-like program was dominated by immune-inflammatory and vascular-complication biology, with more limited support for direct canonical insulin/glucose signalling terms (Figure 5, Supplementary Table 8). Within the GO Biological Process library, the strongest enrichments were inflammatory response (adjusted $p = 1.80 \times 10^{-16}$; odds ratio [OR] = 3.46), cytokine-mediated signalling pathway (*adjusted* $p = 3.08 \times 10^{-14}$; OR = 3.25), and regulation of inflammatory response (adjusted $p = 4.29 \times 10^{-14}$; OR = 3.27). T-cell-associated processes were also prominent, including positive regulation of T-cell activation (adjusted p = 7.04×10^{-12}; OR = 4.71), regulation of T-cell activation (adjusted $p = 4.02 \times 10^{-10}$; OR = 5.99), and T-cell activation itself (adjusted $p = 3.32 \times 10^{-7}$; OR = 3.35). In keeping with the immune-regulatory component seen in GSEA, regulation of regulatory T-cell differentiation was also enriched (adjusted $p = 0.00125$; OR = 5.77). Reactive oxygen species biology was represented by positive regulation of reactive oxygen species metabolic process (adjusted $p = 0.00375$; OR = 3.05). Together, these results indicate that the core DM-like high subset is characterized primarily by cytokine-driven, NF-κB-linked, T-cell-regulated inflammatory signalling, rather than by a purely metabolic transcriptomic program. The Elsevier Pathway Collection supported this interpretation while also adding a more explicit diabetes-mechanistic component. Significant pathways included NF-kB canonical signalling (adjusted $p = 0.00120$; OR = 3.05), STAT3 facilitates the function of Treg cells and cancer progression (adjusted $p = 0.00122$; OR = 8.46), and adiponectin in insulin resistance prevention (adjusted $p = 0.0265$; OR = 5.83), indicating that the DM-like high state integrates immune activation, Treg-linked regulation, and adipocytokine/insulin-resistance biology. Disease-oriented enrichment further showed that the DM-like high subset preferentially mapped to diabetes complications and vascular pathology. In the DisGeNET library, the most significant disease term was atherosclerosis (adjusted $p = 5.74 \times 10^{-33}$; OR = 2.43), followed by diabetic nephropathy (adjusted $p = 5.31 \times 10^{-16}$; OR = 2.35) and carotid atherosclerosis (adjusted $p = 9.20 \times 10^{-11}$; OR = 4.32). Consistent with this complication-oriented profile, the Elsevier database also identified proteins involved in atherosclerosis (adjusted p = 2.02×10^{-25}), proteins involved in diabetic nephropathy (adjusted $p = 6.35 \times 10^{-12}$), endothelial cell dysfunction in progressive diabetic nephropathy (adjusted p = 3.43×10^{-4}), mesangial cell dysfunction in diabetic nephropathy (adjusted $p = 1.83 \times 10^{-4}$), proteins involved in pulmonary hypertension (adjusted $p = 8.26 \times 10^{-11}$), and vascular endothelial cell activation by cytokines (adjusted $p = 5.79 \times 10^{-9}$). These findings indicate that the DM-like high subset is not only diabetes-related at the signalling level, but is also enriched for pathways linked to macrovascular and microvascular diabetic complications. Notably, several focused mechanistic terms classically associated with diabetes were significantly enriched at the predefined threshold. In particular, direct GO terms such as insulin receptor signalling pathway and cellular response to insulin stimulus remained significant after multiple-testing correction in the GOBP analysis, confirming activation of core insulin signalling processes within the DM-like-high subset. This suggests that, in this dataset, the core DM-like high phenotype is captured more strongly by immune-inflammatory regulation, adipocytokine/insulin-resistance signalling, and complication-related vascular programs than by isolated canonical insulin/glucose homeostasis terms. Overall, the GOEA results support the interpretation that the core DM-like high subset in CRC represents a systems-level diabetes-like state dominated by inflammation, T-cell and Treg regulation, ROS-associated stress, adiponectin-linked insulin-resistance biology, and enrichment of nephropathic and atherosclerotic complication pathways. This pattern is more consistent with a complex immune-metabolic-vascular diabetes phenotype than with a narrowly defined glucose-handling program alone.

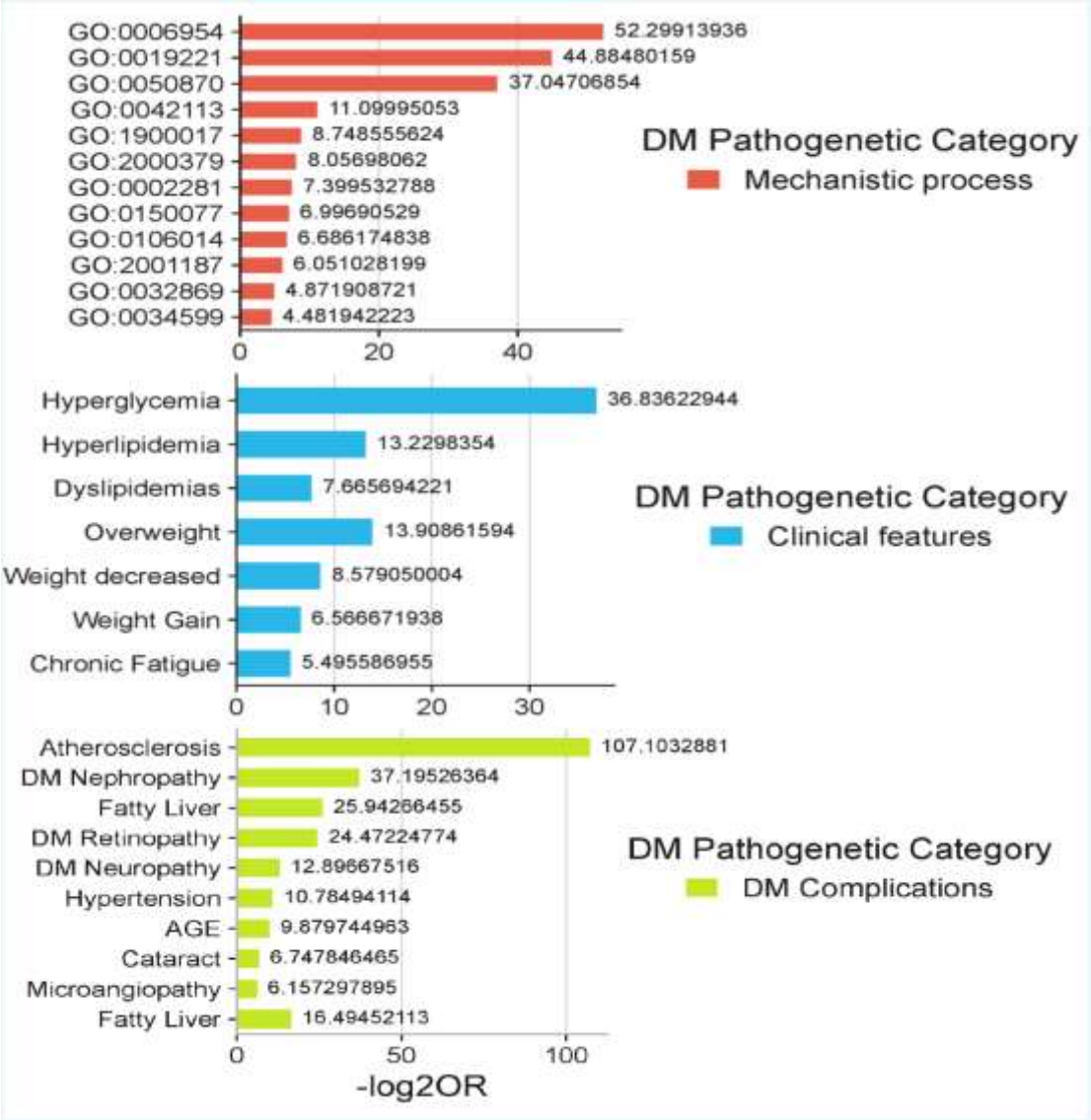

Figure **5**. Gene ontology enrichment analysis of diabetes-related pathways in DM-like-high CRC. Significantly enriched terms (FDR-adjusted $p < 0.05$) are grouped into diabetes-related (top) mechanistic processes, (middle) clinical features, and (bottom) complications. Enrichment is expressed as $-\log_2$ odds ratio ($-\log_2$OR). Full names of enriched terms are provided in Supplementary Material 8.

b. Gene Ontology Analysis of the DM-like-low Subset

In contrast to the core DM-like-high subset, the DM-like-low subset did not demonstrate enrichment of the predefined diabetes-related ontology terms across clinical features, complications, or mechanistic processes (Supplementary Table 9). Despite the presence of significant ontology terms in the global GOEA output, none mapped to the focused diabetes-related categories, including hyperglycaemia, dyslipidaemia, insulin signalling, glucose metabolism, or vascular and inflammatory processes. This indicates that the DM-like-low subset lacks transcriptional activation of diabetes-relevant biological programs. Collectively, these findings support a clear dichotomy between the subsets, whereby the DM-like-high state is characterized by coordinated activation of immune-inflammatory, metabolic, and vascular pathways consistent with diabetes biology, whereas the DM-like-low state represents a transcriptionally distinct phenotype devoid of these diabetes-associated features.

4.0 DISCUSSION

In this study, large-scale colorectal cancer (CRC) transcriptomic cohorts were leveraged to investigate whether tumours can serve as systems-level models of diabetes mellitus (DM)-associated metabolic phenotypes. The findings demonstrate that CRC recapitulates key transcriptional features of diabetes, including hyperglycaemia, insulin resistance, and mitochondrial dysfunction, which converge into a composite DM-like signature with strong biological, clinicopathological, and molecular correlates. Importantly, this DM-like state is characterized by coordinated crosstalk between metabolic, immune-inflammatory, and signalling pathways, highlighting the interconnected nature of metabolic reprogramming and cellular signalling in human disease. Collectively, these findings support the proposition that CRC may serve as a valuable human model for investigating diabetes-related metabolic dysregulation and signalling–metabolism interplay. The central premise of this study—that cancer biology can inform our understanding of metabolic disease—is grounded in the shared molecular architecture linking oncogenesis, cellular signalling, and metabolic dysfunction. CRC and diabetes mellitus share extensive dysregulation of signalling pathways involved in insulin action, nutrient sensing, inflammation, and mitochondrial homeostasis. The present findings demonstrate that CRC tumours exhibit continuous and heterogeneous expression of diabetes-related transcriptional programs, mirroring the spectrum of metabolic dysfunction observed in human diabetic populations (DeFronzo et al., 2015; Holt et al., 2021). Notably, the observed correlation structure, in which metabolic syndrome emerged as a central hub linking hyperglycaemia, insulin resistance, and mitochondrial dysfunction, closely recapitulates the clinical clustering of metabolic abnormalities characteristic of type 2 diabetes and metabolic syndrome (Tsilidis et al., 2015). These observations support the concept that CRC captures key features of signalling–metabolism crosstalk underlying diabetes-associated pathophysiology. CRC offers several advantages over traditional models of diabetes. Unlike rodent models, which often incompletely capture the complexity of human metabolic disease, clinical CRC specimens preserve natural genetic variation, epigenetic regulation,

tumour heterogeneity, and environmental exposures (Hopkins et al., 2016; Caturano et al., 2025). Furthermore, the availability of large, well-annotated datasets—including TCGA, CPTAC2, and Sidra-LUMC—enables integrative, systems-level analyses that would be logistically and ethically challenging to perform directly in patients with diabetes (Cancer Genome Atlas Network, 2012; Vasaikar et al., 2019; Roelands et al., 2023). The DM-like signature derived in this study demonstrated meaningful associations with clinicopathological progression, genomic alterations, and pathways implicated in diabetic complications, thereby supporting its biological validity. Importantly, the enrichment of immune-inflammatory, insulin-resistance, oxidative stress, and vascular-complication pathways indicates that CRC captures the complex interplay between cellular signalling and metabolic dysregulation that characterizes diabetes pathogenesis. The establishment of CRC as a DM-like systems model opens several translational applications. CRC transcriptomes provide a human in vivo system for investigating the complex interplay between insulin signalling, inflammation, oxidative stress, and mitochondrial dysfunction, all of which are central to diabetes pathogenesis but difficult to model in isolation (DeFronzo et al., 2015; Holt et al., 2021; Pavlova & Thompson, 2016). Shared signalling pathways between cancer and diabetes, including PI3K/AKT/mTOR and AMPK signalling, suggest that metabolic therapeutics such as metformin, as well as pathway-targeted agents including PI3K inhibitors, may have dual relevance in oncology and metabolic disease (Hopkins et al., 2016; Caturano et al., 2025; Saxton and Sabatini, 2017; Giovannucci et al., 2010). The DM-like transcriptional signatures identified could serve as biomarkers of metabolic vulnerability, therapeutic response, and patient risk stratification, consistent with growing interest in metabolic stratification in precision oncology (Vasaikar et al., 2019). Moreover, enrichment of vascular and nephropathic pathways indicates that CRC may provide a valuable platform for investigating molecular mechanisms underlying diabetic complications, particularly endothelial dysfunction and atherosclerosis, which are driven by chronic inflammation, oxidative stress, and metabolic dysregulation (Cole & Florez, 2020; Bornfeldt &Tabas, 2011).

An important insight is the preferential enrichment of DM-like phenotypes in MSI-high/hypermutated CRC tumours, as well as in *BRAF*-mutant tumours. These tumours consistently exhibited higher scores for hyperglycaemia, mitochondrial dysfunction, metabolic syndrome, and composite DM-like signatures. The MSI-high subtype is characterized by high tumour mutational burden, dense immune infiltration, and strong inflammatory signalling, features that closely align with the immune-mediated biology observed in the DM-like-high subset (Cancer Genome Atlas Network, 2012; Guinney et al., 2015; Llosa et al., 2015). These findings suggest that the MSI-high/hypermutated molecular subtype represents the most suitable CRC model for diabetes-related biology, particularly for studying immune–metabolic interactions. The strong enrichment of T-cell activation and regulatory T-cell pathways further supports its relevance to autoimmune diabetes mechanisms, which are driven by immune-mediated β-cell destruction (Atkinson et al., 2014; Bluestone et al., 2010). In contrast, chromosomally unstable tumours with high aneuploidy and fraction genome altered demonstrated lower DM-like scores, indicating that molecular stratification is essential when using CRC as a model for diabetes-associated processes.

The identification of a DM-like metabolic state in CRC has important implications for patient management. First, the presence of a type 1 diabetes-like immune-metabolic program may contribute to clinical features such as weight loss and cachexia. Chronic inflammation and cytokine signalling are well-established drivers of cancer-associated cachexia, promoting catabolism and energy imbalance (Argilés et al., 2014; Baracos et al., 2018). The enrichment of immune activation pathways suggests that DM-like states may exacerbate these processes through coordinated interactions between inflammatory signalling and metabolic dysregulation. Second, tumours with high DM-like scores may exhibit altered therapeutic responses. Increased insulin signalling and metabolic pathway activation can influence sensitivity to targeted therapies, while immune activation is a key determinant of response to immune checkpoint inhibitors, particularly in MSI-high tumours (Llosa et al., 2015; Le et al., 2015). Third, the enrichment of vascular complication pathways raises the possibility that CRC patients with a DM-like phenotype may be at increased risk of cardiovascular complications, thrombotic events, and metabolic comorbidities. This is consistent with evidence linking cancer, chronic inflammation, endothelial dysfunction, and prothrombotic states, as well as shared pathogenic mechanisms underlying diabetes and cardiovascular disease (Libby, 2021; Khorana, 2010). Collectively, these findings suggest that metabolic profiling of tumours could inform holistic patient management by integrating oncologic, metabolic, and cardiovascular risk assessment.

Despite the strengths of this integrative approach, several limitations should be acknowledged. First, reverse causality cannot be excluded. Although the DM-like signature is interpreted as CRC recapitulating diabetes-associated biology, patients with pre-existing diabetes may have contributed disproportionately to the DM-like-high subset. However, strong associations between the DM-like phenotype and tumour-intrinsic molecular features argue against purely patient-driven effects. Prospective studies incorporating matched metabolic phenotyping are required to clarify directionality (Giovannucci et al., 2010; Sun et al., 2022). Second, transcriptional signatures do not constitute functional assays. Although ssGSEA-based scores capture coordinated gene expression programs, they do not directly measure insulin secretion, glucose tolerance, insulin sensitivity, or metabolic flux (DeFronzo et al., 2015; Holt et al., 2021). Whether DM-like-high tumours actively modulate systemic metabolism remains to be determined using functional systems (Vander Heiden et al., 2009; Pavlova & Thompson, 2016). Third, the present model is specific to CRC and may not be generalizable to other tumour types. The unique metabolic environment of the intestine may shape the DM-like signature in ways not representative of systemic metabolism (Gurung et al., 2020; Sepich-Poore et al., 2021). Fourth, circularity is an inherent concern in enrichment-based analyses, though this was mitigated by excluding genes with substantial overlap with signature-construction gene sets and by performing sensitivity analyses. The strong concordance between dichotomized and continuous GSEA approaches supports robustness (Subramanian et al., 2005). Fifth, the absence of matched normal colonic tissues limited the ability to determine whether the DM-like signature represents a cancer-specific phenomenon (Cancer Genome Atlas Network, 2012). Finally, survival analyses were constrained by incomplete follow-up and the relatively short time horizon of cancer-specific outcomes, whereas diabetic complications typically develop over decades (Cole & Florez, 2020; Beckman et al., 2002). Despite these limitations, the present study provides evidence that CRC transcriptomes capture clinically and biologically relevant signalling–metabolism interactions associated with diabetes.

Conclusion

This study provides comprehensive evidence that colorectal cancer transcriptomes encode biologically meaningful signatures of diabetes mellitus. The DM-like composite score, derived from hyperglycaemia, insulin resistance, and mitochondrial dysfunction signatures, captures a continuous transcriptional axis enriched for type 1 diabetes and autoimmune pathways at one extreme and MODY-related monogenic diabetes pathways at the other. Importantly, the DM-like state is characterized by extensive crosstalk between metabolic programs and immune-inflammatory signalling networks, including insulin signalling, cytokine-mediated pathways, oxidative stress responses, and vascular complication-associated processes. The Hypermutated/MSI/*BRAF*-mutant molecular subtype most faithfully recapitulates this DM-like state, providing a tractable experimental system for investigating immune–metabolic interactions and signalling–metabolism interplay. These molecular associations generate testable hypotheses regarding potential relationships between the DM-like phenotype and prognosis, cachexia, cardiometabolic complications, treatment response, and immunotherapy response, which warrant investigation in appropriately annotated clinical cohorts. Although limitations remain—particularly regarding reverse causality and the need for functional validation—the findings establish CRC as a valuable human systems biology model for diabetes, with potential applications in biomarker discovery, therapeutic repurposing, precision medicine, and mechanistic investigation of diabetic complications.

Acknowledgments

The author gratefully acknowledges the contributions of The Cancer Genome Atlas and the cBioPortal for Cancer Genomics initiatives for generating, curating, and maintaining the publicly available datasets that made this study possible.

Conflicts of Interest

The author declares no conflicts of interest.

Data Availability Statement

The datasets analysed in this study were obtained from publicly accessible cancer genomics repositories, including the Genome Data Commons (GDC) Data Portal (https://portal.gdc.cancer.gov) and cBioPortal for Cancer Genomics (https://www.cbioportal.org). All data used are available from these platforms in accordance with their respective access and usage policies.

REFERENCES

Argilés, J.M., Busquets, S., Stemmler, B. and López-Soriano, F.J. (2014). Cancer cachexia. Nature Reviews Cancer, 14: 754–762.

Ashburner, M., Ball, C.A., Blake, J.A., Botstein, D., Butler, H., Cherry, J.M., et al. (2000). Gene ontology: tool for the unification of biology. Nature Genetics, 25: 25–29.

Atkinson, M.A., Eisenbarth, G.S. and Michels, A.W. (2014). Type 1 diabetes. The Lancet, 383: 69–82.

Baracos, V.E., Martin, L., Korc, M., Guttridge, D.C. and Fearon, K.C.H. (2018). Cancer-associated cachexia. Nature Reviews Disease Primers, 4: 17105.

Barbie, D.A., Tamayo, P., Boehm, J.S., Kim, S.Y., Moody, S.E., Dunn, I.F., et al. (2009). Systematic RNA interference reveals that oncogenic KRAS-driven cancers require TBK1. Nature, 462: 108–112.

Beckman, J.A., Creager, M.A. and Libby, P. (2002). Diabetes and atherosclerosis. JAMA, 287: 2570–2581.

Bluestone, J.A., Herold, K. and Eisenbarth, G. (2010). Genetics, pathogenesis and clinical interventions in type 1 diabetes. Nature, 464: 1293–1300.

Bornfeldt, K.E. and Tabas, I. (2011). Insulin resistance, hyperglycemia, and atherosclerosis. Cell Metabolism, 14: 575–585.

Cancer Genome Atlas Network. (2012). Comprehensive molecular characterization of human colon and rectal cancer. Nature, 487: 330–337.

Caturano, A., D'Angelo, M., Pafundi, P.C., Galiero, R., Rinaldi, L., Salvatore, T., Vetrano, E., Cesaro, A., Rinaldi, M., Marfella, R., et al. (2025). Insulin resistance and cancer: molecular links and clinical implications. Cancers, 17: 945.

Cerami, E., Gao, J., Dogrusoz, U., Gross, B.E., Sumer, S.O., Aksoy, B.A., et al. (2012). The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. Cancer Discovery, 2: 401–404.

Chen, E.Y., Tan, C.M., Kou, Y., Duan, Q., Wang, Z., Meirelles, G.V., et al. (2013). Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. BMC Bioinformatics, 14: 128.

Cole, J.B. and Florez, J.C. (2020). Genetics of diabetes mellitus and diabetes complications. Nature Reviews Nephrology, 16: 377–390.

DeFronzo, R.A., Ferrannini, E., Groop, L., Henry, R.R., Herman, W.H., Holst, J.J., et al. (2015). Type 2 diabetes mellitus. Nature Reviews Disease Primers, 1: 15019.

Gao, J., Aksoy, B.A., Dogrusoz, U., Dresdner, G., Gross, B., Sumer, S.O., et al. (2013). Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. Science Signaling, 6: pl1.

Gene Ontology Consortium. (2026). The Gene Ontology knowledgebase in 2026. Nucleic Acids Research, 54(D1): D1779–D1792. doi: 10.1093/nar/gkaf1292.

Giovannucci, E., Harlan, D.M., Archer, M.C., Bergenstal, R.M., Gapstur, S.M., Habel, L.A., et al. (2010). Diabetes and cancer: a consensus report. CA: A Cancer Journal for Clinicians, 60: 207–221.

Guinney, J., Dienstmann, R., Wang, X., de Reyniès, A., Schlicker, A., Soneson, C., et al. (2015). The consensus molecular subtypes of colorectal cancer. Nature Medicine, 21: 1350–1356.

Gurung, M., Li, Z., You, H., Rodrigues, R., Jump, D.B., Morgun, A. and Shulzhenko, N. (2020). Role of gut microbiota in type 2 diabetes pathophysiology. EBioMedicine, 51: 102590.

Gutiérrez-Salmerón, M., Lucena, S.R., Chocarro-Calvo, A., García-Martínez, J.M., Martín Orozco, R.M. and García-Jiménez, C. (2021). Metabolic and hormonal remodeling of colorectal cancer cell signalling by diabetes. Endocrine-Related Cancer, 28(6): R191–R206. doi: 10.1530/ERC-21-0092.

Holt, R.I.G., DeVries, J.H., Hess-Fischl, A., et al. (2021). The management of type 1 diabetes in adults. A consensus report by the ADA and the EASD. Diabetologia, 64: 2609–2652.

Hopkins, B.D., Goncalves, M.D. and Cantley, L.C. (2016). Insulin-PI3K signalling: an evolutionarily insulated metabolic driver of cancer. Nature Reviews Endocrinology, 12: 548–561.

Johnson, W.E., Li, C. and Rabinovic, A. (2007). Adjusting batch effects in microarray expression data using empirical Bayes methods. Biostatistics, 8: 118–127.

Kanehisa, M., Furumichi, M., Sato, Y., Matsuura, Y. and Ishiguro-Watanabe, M. (2025). KEGG: biological systems database as a model of the real world. Nucleic Acids Research, 53(D1): D672–D677. doi: 10.1093/nar/gkae909.

Khorana, A.A. (2010). Cancer-associated thrombosis: updates and controversies. Journal of Clinical Oncology, 28: 4839–4847.

Kuleshov, M.V., Jones, M.R., Rouillard, A.D., Fernandez, N.F., Duan, Q., Wang, Z., et al. (2016). Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. Nucleic Acids Research, 44: W90–W97.

Le, D.T., Uram, J.N., Wang, H., Bartlett, B.R., Kemberling, H., Eyring, A.D., et al. (2015). PD-1 blockade in tumors with mismatch-repair deficiency. The New England Journal of Medicine, 372: 2509–2520.

Leek, J.T., Johnson, W.E., Parker, H.S., Jaffe, A.E. and Storey, J.D. (2012). The sva package for removing batch effects and other unwanted variation. Bioinformatics, 28: 882–883.

Li, B. and Dewey, C.N. (2011). RSEM: accurate transcript quantification from RNA-Seq data. BMC Bioinformatics, 12: 323.

Libby, P. (2021). The changing landscape of atherosclerosis. Nature, 592: 524–533.

Llosa, N.J., Cruise, M., Tam, A., Wicks, E.C., Hechenbleikner, E.M., Taube, J.M., et al. (2015). The vigorous immune microenvironment of microsatellite instable colon cancer. Cancer Discovery, 5: 43–51.

Milacic, M., Beavers, D., Conley, P., Gong, C., Gillespie, M., Griss, J., Haw, R., Jassal, B., Matthews, L., May, B., et al. (2024). The Reactome Pathway Knowledgebase 2024. Nucleic Acids Research, 52(D1): D672–D678. doi: 10.1093/nar/gkad1025.

Pavlova, N.N. and Thompson, C.B. (2016). The emerging hallmarks of cancer metabolism. Cell Metabolism, 23: 27–47.

Ritchie, M.E., Phipson, B., Wu, D., Hu, Y., Law, C.W., Shi, W., et al. (2015). limma powers differential expression analyses. Nucleic Acids Research, 43: e47.

Roelands, J., Kuppen, P.J.K., Ahmed, E.I., Mall, R., Masoodi, T., Singh, P., Monaco, G., Raynaud, C., de Miranda, N.F.C.C., Ferraro, L., et al. (2023). An integrated tumor, immune and microbiome atlas of colon cancer. Nature Medicine, 29: 1273–1286.

Saxton, R.A. and Sabatini, D.M. (2017). mTOR signaling in growth, metabolism, and disease. Cell, 168: 960–976.

Sepich-Poore, G.D., Zitvogel, L., Straussman, R., Hasty, J., Wargo, J.A. and Knight, R. (2021). The microbiome and human cancer. Science, 371: eabc4552.

Smyth, G.K. (2004). Linear models and empirical Bayes methods for assessing differential expression. Statistical Applications in Genetics and Molecular Biology, 3: Article 3.

Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A., Paulovich, A., Pomeroy, S.L., Golub, T.R., Lander, E.S. and Mesirov, J.P. (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proceedings of the National Academy of Sciences of the United States of America, 102: 15545–15550.

Sun, H., Saeedi, P., Karuranga, S., Pinkepank, M., Ogurtsova, K., Duncan, B.B., Stein, C., Basit, A., Chan, J.C.N., Mbanya, J.C., et al. (2022). IDF Diabetes Atlas: global, regional and country-level diabetes prevalence estimates for 2021 and projections for 2045. Diabetes Research and Clinical Practice, 183: 109119.

Tang, D., Chen, M., Huang, X., Zhang, G., Zeng, L., Zhang, G., Wu, S. and Wang, Y. (2023). SRplot: a free online platform for data visualization and graphing. PLoS ONE, 18: e0294236.

Taylor, R. (2013). Type 2 diabetes: etiology and reversibility. Diabetes Care, 36: 1047–1055.

Tsilidis, K.K., Kasimis, J.C., Lopez, D.S., Ntzani, E.E. and Ioannidis, J.P.A. (2015). Type 2 diabetes and cancer: umbrella review of meta-analyses of observational studies. BMJ, 350: g7607.

Vander Heiden, M.G., Cantley, L.C. and Thompson, C.B. (2009). Understanding the Warburg effect: the metabolic requirements of cell proliferation. Science, 324: 1029–1033.
Vasaikar, S., Huang, C., Wang, X., et al. (2019). Proteogenomic analysis of human colon cancer reveals new therapeutic opportunities. Cell, 177: 1035–1049.
Xie, Z., Bailey, A., Kuleshov, M.V., Clarke, D.J.B., Evangelista, J.E., Jenkins, S.L., et al. (2021). Gene set knowledge discovery with Enrichr. Current Protocols, 1: e90.