

Results

Single-Cell RNA-Sequencing Reveals Cell Type Heterogeneity and Diet-Induced Alterations in the Intestinal Epithelium

Single-cell RNA-sequencing analysis of small intestinal tissue displayed the variation the transcriptomes of the cells in the dataset and revealed distinct cell populations across CD and HFHSD conditions. A total of 27,419 cells passed quality control criteria, with 13,363 cells from CD and 14,056 cells from HFHSD conditions. The average number of genes detected per cell was 3,566 for CD and 3,490 for HFHSD, while the mean UMI count per cell was 18,647 for CD and 20,385 for HFHSD. Unsupervised Leiden clustering and annotation based on known marker genes identified 12 distinct cell clusters (Figure 1a). These clusters represented six main cell types, ISCs, enterocytes, goblet cells, tuft cells, paneth cells, and enteroendocrine cells. Additionally, progenitor populations were identified for all main cells types excluding ISCs. Cells which did not express known epithelial cell markers were labelled as ‘Not Annotated’. UMAP visualisation of cells from CD and HFHSD conditions (Figure 1b) showed overall similarity in cluster distribution, indicating successful batch correction. Notable differences were observed in the density in the Paneth cell cluster which were more dense in the HFHSD condition. This cluster of Paneth cells were reported to have sampling issues however, and were thus removed from further analyses. Feature plots displaying the average expression of cell type-specific marker genes (Fig 1c) confirmed the identity of the annotated clusters. Each cell type exhibited a distinct pattern of marker gene expression. ISCs exhibited the expression of *Lgr5*, *Olfm4*, *Axin2*, *Ascl2*, and *Slc12a2* (max average expression: 2.22), enterocytes displayed the expression of *Fabp1*, *Alpi*, *Apoa1*, *Apoa4*, and *Lct* (max average expression: 4.27), Goblet cells displayed the expression of *Muc2*, *Tff3*, *Agr2*, *Spdef*, *Klf4*, *Ccl9* (max average expression: 4.43), and *Manf*, Paneth cells displayed the expression of *Lyz1*, *Mmp7*, *Defa17*, *Defa22*, and *Anf4* (max average expression: 6.10). Enteroendocrine cells displayed the expression of *Neurod1*, *Neurod2*, *Insm1*, *Chga*, and *Chgb* (max average expression: 4.33), and tuft cells displayed the expression of *Dclk1*, *Trpm5*, *Gfi1b*, and *Iil25* (max average expression: 2.42). Quantification of cell type proportions between CD and HFHSD conditions (Fig 1d) revealed changes in several cell populations. The ISC population was lower in the HFHSD condition, while enterocyte progenitors showed a

higher proportion, indicating increased cell turnover of ISCs in the HFHSD condition. Conversely, both tuft progenitors and tuft cells appeared in lower proportions in the HFHSD condition. These trends were consistent with previously published findings using this dataset (Aliluev et al., 2021). Enterocyte progenitors were the largest cluster with a proportion of approximately 0.3 in CD and 0.375 in HFHSD. Second to this cluster were ISCs and goblet proportions. ISCs exhibited a proportion of approximately 0.23 in the CD group and 0.13 in the HFHSD group. Goblet cells accounted for approximately a portion of 0.2. Heatmaps of marker gene expression for CD and HFHSD conditions (Figure 1e) illustrated distinct transcriptional profiles for each of the 12 identified clusters. The heatmaps revealed both shared and diet-specific gene expression patterns across cell types, with certain genes showing differential expression between CD and HFHSD conditions within the same cell type. These data exhibit the diverse cellular heterogeneity within the intestinal epithelium as well as some HFHSD induced alterations in cell type proportions and expression profiles.

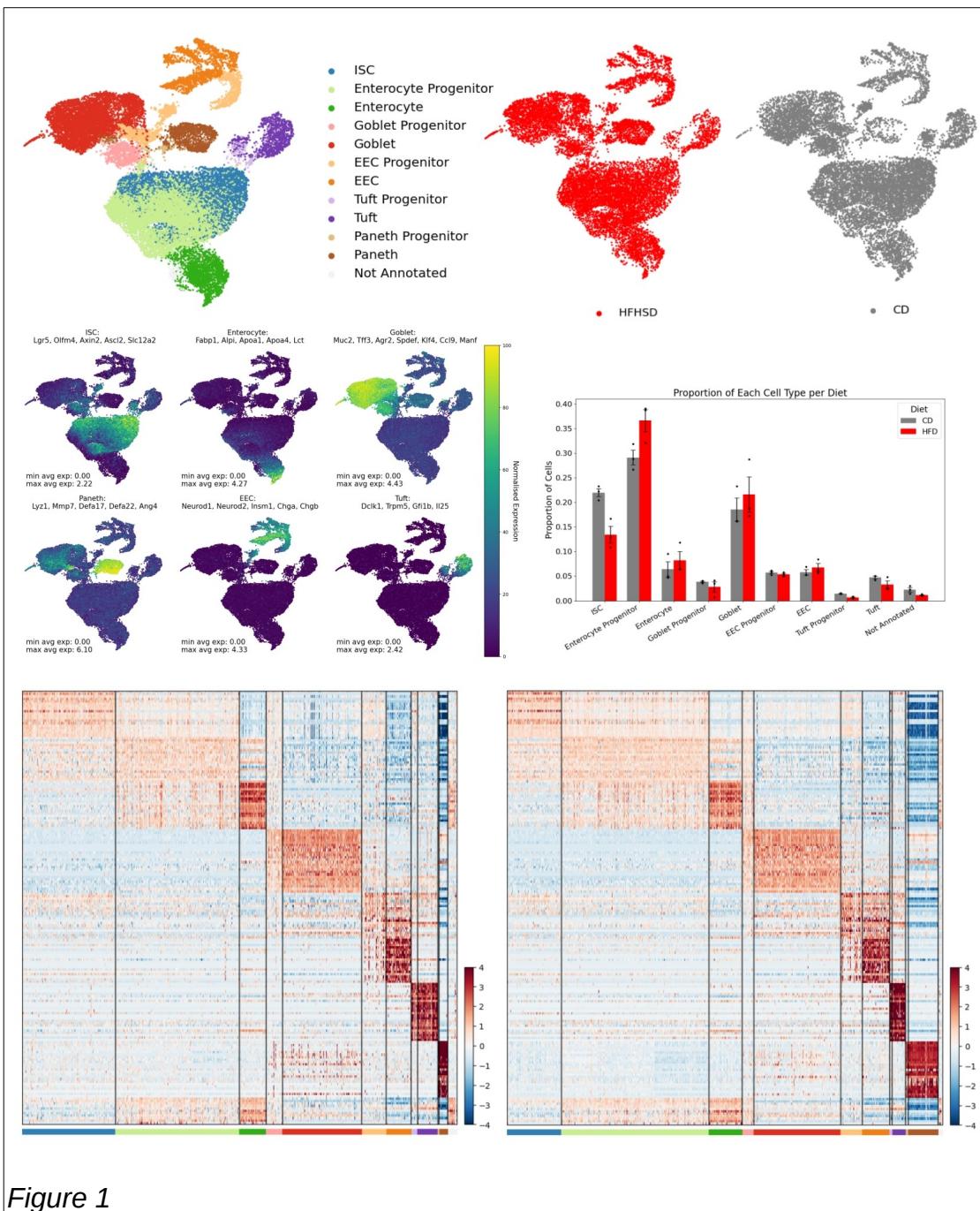


Figure 1

High-Fat High-Sugar Diet Alters ISC Function

Differential gene expression analysis of intestinal stem cells revealed a total of 337 differentially expressed genes between the CD and HFHSD conditions (Figure 2). The top 10 highest log₂ fold change significantly DEGs were *Fabp1*, *Mgll*, *H2-Aa*, *Creb3l3*, *Etv4*, *Slc16a12*, *H2-Ab1*, *H2-Eb1*, *Cst6*, and *Scd2*. The top 10 lowest log₂ fold change significantly DEGs were *Sepp1*, *Sis*, *S100g*, *Mtus2*, *Anpep*, *Emp2*, *AA467197*, *Lzts3*, *Prr15*, *Zfp467*. Trends were observed in genes relating to both fat metabolism and cell cycle regulation. Notably, genes involved in fatty acid metabolism and transport showed consistent upregulation. Among these, *Fabp1* stood out with one of the highest positive log₂ fold change (log₂FC) of approximately 2.7, while *Fabp2*, a key intracellular protein involved in the uptake and transport of long-chain fatty acids, also displayed significantly increased expression albeit with a lower log₂FC of approximately 0.6. Similarly, *Hmgcs2*, a mitochondrial enzyme essential for the first steps of ketogenesis, exhibited higher expression levels under the HFHSD with a log₂FC of around 1. Upregulation was also seen in *Acot1*, an enzyme that facilitates the conversion of acyl-CoAs into fatty acids and CoA, and *Scd2*, which plays a role in fatty acid biosynthesis both highly significant with log₂FC values around 1. The enzyme encoded by *Mgll*, responsible for breaking down monoacylglycerols into free fatty acids, also showed increased expression with a log₂FC over 2.5.

In addition to changes in metabolism-related genes, several genes associated with cell cycle regulation and proliferation were upregulated. *Hsp90aa1*, a molecular chaperone known to be involved in cell cycle control, showed a marked increase in expression. *Nme1*, involved in nucleoside triphosphate synthesis. In contrast, *Zfas1*, a non-coding RNA linked to cellular differentiation, was downregulated, as was *Sepp1*, a protein that plays a role in redox balance and epithelial cell proliferation. Additionally, the tumor suppressor *Pdc4* showed significant downregulation, along with *Ypel3*, a gene associated with the regulation of cellular senescence.

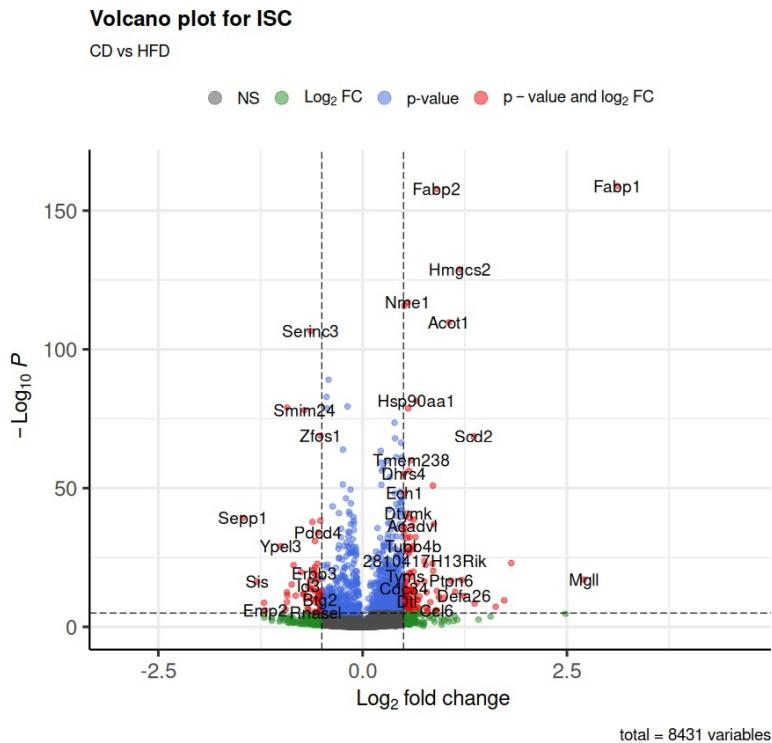


Figure 2

Gene Ontology Enrichment Analysis

GO enrichment analysis of DEGs in ISCs exposed to a HFHSD revealed significant alterations across CC, MF, and BP categories (Figure 3). The analysis revealed 16 distinct clusters of enriched GO terms and demonstrated a trend in the upregulation of cell cycle-related processes consistent with findings described from volcano plot analysis. As the top 50 GO terms from each class were selected by adjusted p-value, all terms are highly significant.

Cluster 3 in the BP category displayed enrichment for terms associated with chromosome separation, mitotic cell cycle, and cell cycle regulation. This pattern was mirrored in the CC section, where the enrichment of terms in the same cluster were related to chromosomes and mitotic spindles. Complementing these findings, cluster 1 showed enrichment of DNA replication GO terms in the BP category, corresponding to replication fork-related terms in the CC group. The MF category further supported this trend, with upregulation of terms associated with helicase activity, ATP hydrolysis acting on DNA, and single-strand DNA binding. Each of these GO terms were positively enriched with NESs between 1.5 and 2.5.

Mitochondrial function also appeared significantly enriched, as evidenced by cluster 6. This cluster showed upregulation of mitochondrial translation and gene expression in the BP category, while the CC section indicated increased expression of mitochondrial ribosome-related terms.

RNA processing mechanisms were positively. Cluster 5 revealed upregulation of spliceosome and ribonucleoprotein complex-related terms, particularly in protein-RNA complex assembly. Similarly, cluster 2 showed increased ribonucleoprotein complex biogenesis, with corresponding upregulation of snoRNA binding in the MF category.

Cluster 12 showed decreased DNA-binding transcription factor activity specific to RNA polymerase II. Cluster 4 indicated downregulation in genes related to adherens junctions in the CC category, with associated decreases in transmembrane protein kinase activity and cell adhesion molecule binding in the MF category. All related GO terms were found in the -1.5 to -2 NES range.

Overall this GO enrichment analysis of ISCs under prediabetic HFHSD conditions, revealed trends in the broad upregulation of processes related to cell cycle progression, DNA replication, mitochondrial function, and RNA processing. Moreover, it showed downregulation in specific transcription factor activities and cell adhesion processes.

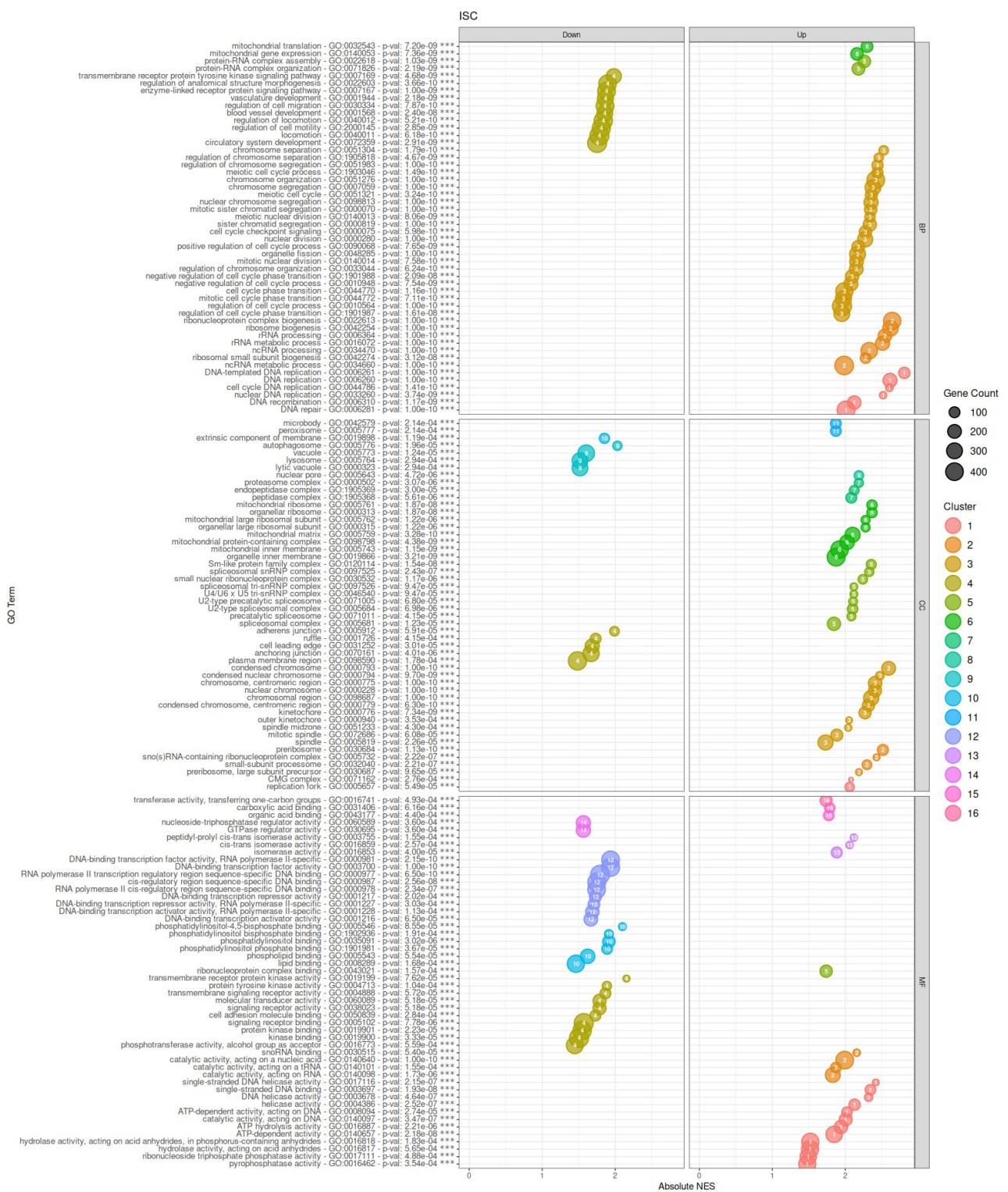


Figure 3

The Gene Ontology enrichment network analysis (Figure 4) further corroborated and expanded upon the findings observed in the dotplot visualisation. This network

representation integrated Cellular Component, Biological Process, and Molecular Function categories, revealing 33 distinct clusters of functionally related GO terms.

A prominent cluster centered around the "regulation of mitotic sister chromatid segregation" GO term emerged, exhibiting positive Normalised Enrichment Scores (NES). This cluster encompassed various cell cycle-related processes, reinforcing the earlier observation of cell cycle upregulation in ISCs under HFHSD. Similarly, a separate cluster with most edges leading to the "double-strand break repair via homologous recombination" GO term also displayed positive NES values, further emphasising the enrichment of DNA replication and repair processes.

Conversely, a cluster associated with "brush border" and "regulation of developmental process" GO terms showed negative NES values. This cluster is consistent with the previous findings relating to the downregulation of adherens junction-related GO terms, providing additional evidence for altered cell adhesion and developmental processes in the ISCs.

The enrichment map also highlighted several other functional clusters. Notably, terms related to "mitochondrial translation" and "ribosomal large subunit biogenesis" formed distinct clusters with positive NES values, aligning with the earlier findings of upregulated mitochondrial and protein synthesis processes. Additionally, clusters associated with "fatty acid binding" and "carboxylic acid metabolic process" were observed, indicating significant alterations in lipid metabolism pathways.

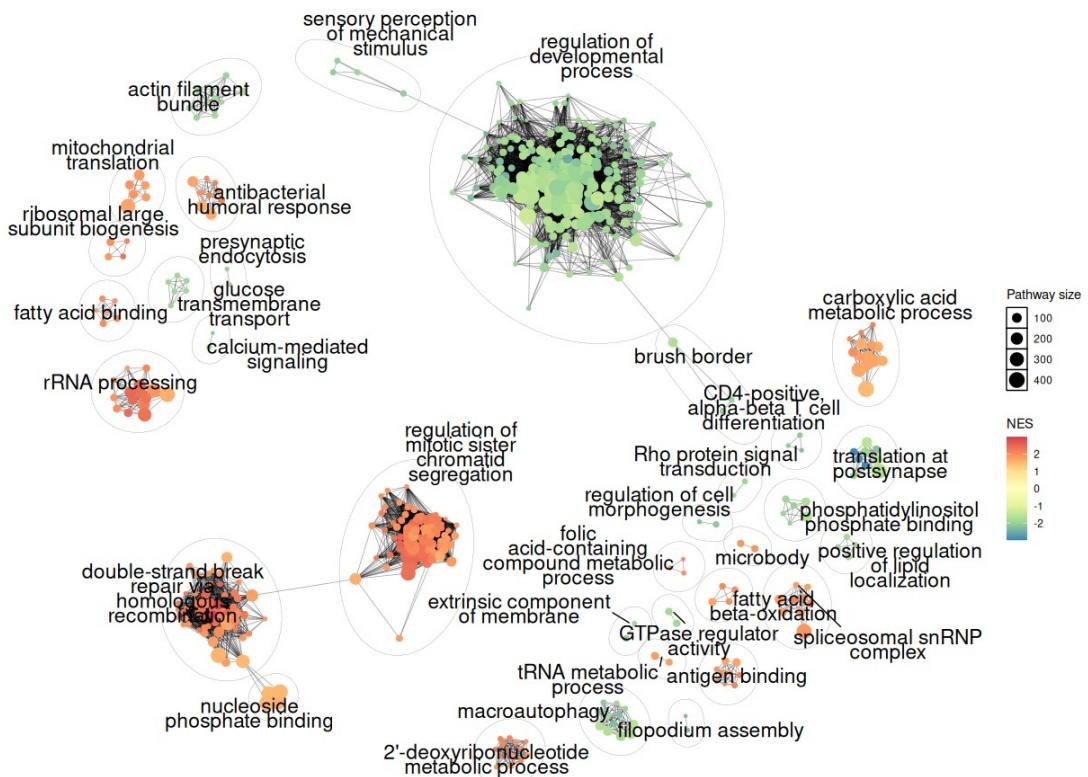


Figure 4

Building upon the GO enrichment analysis, KEGG pathway enrichment analysis of ISCs further revealed the functional alterations associated with the prediabetic state (Figure 5). The analysis found 23 significant enrichment of pathways. 5 of these pathways were related to genetic information processing, each with a spread of DEGs in the upregulated region, with DNA replication showing high significance ($p = 3.57e-04$) and a high gene ratio of approximately 0.9. This was complemented by enrichment of related pathways including ribosome biogenesis, mismatch repair, and spliceosome. 3 pathways were enriched in the cellular pathways category with the cell cycle pathway most significantly enriched ($p = 6.33e-03$), and with most significant DEGs upregulated (gene ratio approximately 0.5), reinforcing the earlier observations of cell cycle upregulation. Notably, the cell adhesion molecules pathway was enriched ($p = 1.29e-02$) with a predominance of downregulated genes. The analysis also highlighted alterations in the endocrine system, with significant enrichment of the PPAR signaling pathway ($p = 1.24e-06$), with most of the significant DEGs in the positive log₂ fold change region. Furthermore, the fatty acid metabolism pathway was significantly enriched ($p = 2.08e-02$) with many of its genes

upregulated, coinciding with the alterations revealed in the DEG and GO analyses. Interestingly, several immune-related pathways showed enrichment, including the *Staphylococcus aureus* infection pathway ($p = 1.23e-07$) and the antigen processing and presentation pathway ($p = 8.69e-03$), with many of their respective significant DEGs in displayed in the positive og2 fold change region.

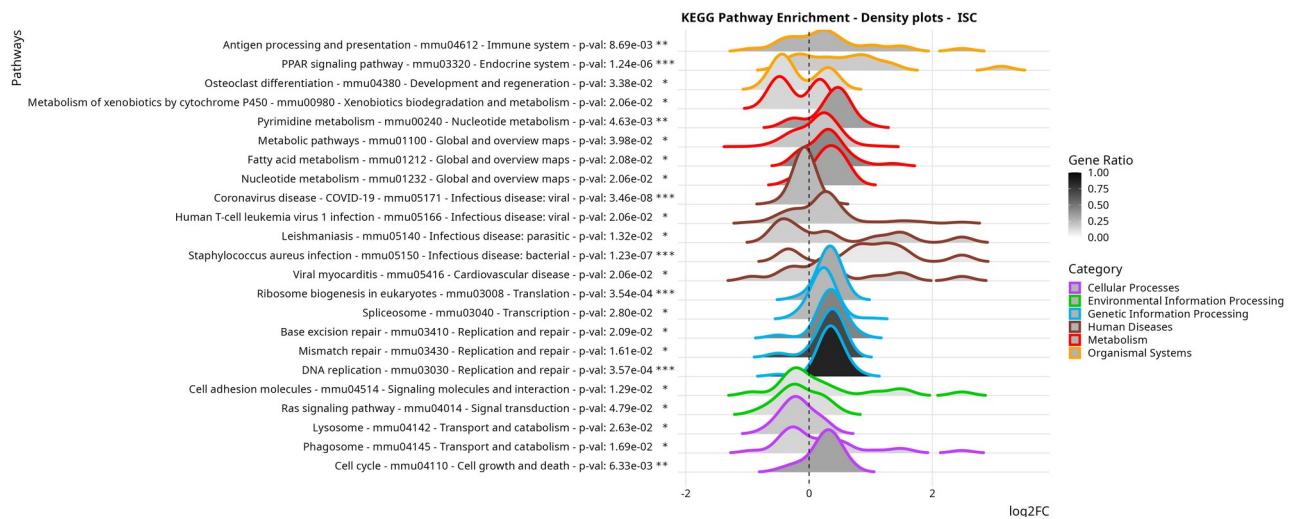


Figure 5

Further examination of the cell cycle pathway through KEGG graphs and heatmaps of log 2 fold change values of DEGs provided additional insights into the mechanisms of cell cycle alterations in the ISCs under prediabetic conditions (Figure 6). The heatmap analysis revealed that approximately 52% of genes involved in the cell cycle were reported significantly differentially expressed in the epithelium with ISCs and enterocyte progenitors being the most affected cell types within the pathway. ISCs displayed widespread upregulation of cell cycle-related genes, with a slightly attenuated signal observed in enterocyte progenitors. This trend builds on the current evidence of the rapid proliferation of ISCs followed by differentiation into enterocyte cell types, consistent with the cell type proportions seen in Figure 1d. The KEGG graph visualisation of the cell cycle pathway highlighted significant upregulation of key cyclins, including Cyclin D, E, A, and B, as well as Cyclin-dependent kinase 1 (CDK1). The graph also indicated increased expression of CDK45, with the regulatory signals predominantly tending towards DNA replication upregulation. These observations at the pathway level corroborate and extend the earlier findings of cell cycle upregulation in ISCs, providing a more detailed view of the specific cell cycle components and mechanisms affected in the prediabetic intestinal epithelium.

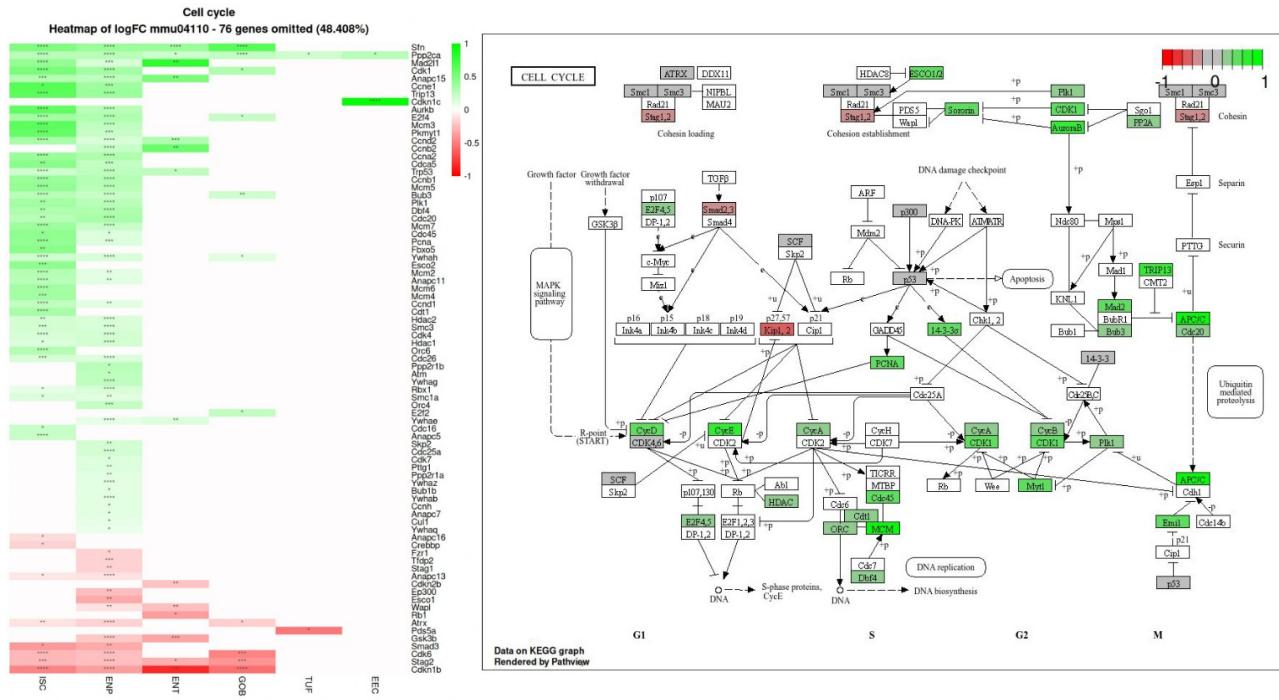


Figure 6

High-Fat High-Sugar Diet Alters Enterocyte Progenitor Function

Differential gene expression analysis of enterocyte progenitors in the intestinal epithelium revealed a total of 339 significantly differentially expressed genes between the CD and HFHSD conditions (Figure 7). The top 10 highest log₂ fold change significantly DEGs were *Cyp4a10*, *Fabp1*, *Mgll*, *Creb3l3*, *Acot1*, *Me1*, *Hmgcs2*, *Akr1b8*, *Acaa1b*, *Acacb*. The top 10 lowest log₂ fold change significantly DEGs were *Sepp1*, *Sis*, *Gngt2*, *Sh3bgr*, *Mfsd4*, *S100g*, *Sectm1b*, *AA467197*, *Abi3*, *Tnfsf10*. Trends were observed in genes relating to nutrient and energy utilisation pathways between CD and prediabetic HFHSD conditions, which were closely related to lipid metabolism processes. Genes involved in fatty acid uptake and intracellular transport showed marked upregulation. *Fabp1* and *Fabp2*, encoding fatty acid binding proteins crucial for fatty acid absorption in enterocytes, exhibited log₂ fold changes approximately 3.5 and 0.7 respectively, similar to what was seen in the ISC population. *Hmgcs2*, a key enzyme in ketogenesis, also showed significant upregulation with a log₂ fold change value of approximately 2. The prediabetic model further induced upregulation of genes associated with mitochondrial fatty acid metabolism and beta-oxidation. *Ech1* (enoyl-CoA hydratase), *Acaa2* (acetyl-CoA

acyltransferase 2), *Etfb* (electron transfer flavoprotein beta subunit), and *Acad1* (acyl-CoA dehydrogenase) all showed increased expression, all related in their capacity for fatty acid breakdown. *Me1*, encoding malic enzyme 1, which generates NADPH for fatty acid biosynthesis, was also upregulated. Additionally, *Cyp4a10*, involved in fatty acid omega-oxidation, showed increased expression, suggesting activation of alternative lipid metabolism pathways.

Conversely, several genes related to carbohydrate metabolism and cellular stress responses were downregulated in the HFHSD condition. *Aldob*, encoding fructose-bisphosphate aldolase B, a key enzyme in glycolysis and fructose metabolism, showed reduced expression. *Oat*, encoding ornithine aminotransferase, a mitochondrial enzyme involved in amino acid metabolism, was also downregulated.

Genes involved in cellular stress responses and redox regulation also showed decreased expression. *Sepp1*, which encodes selenoprotein P, an extracellular antioxidant, and *Txnip*, a regulator of cellular redox state and glucose metabolism, both exhibited downregulation.

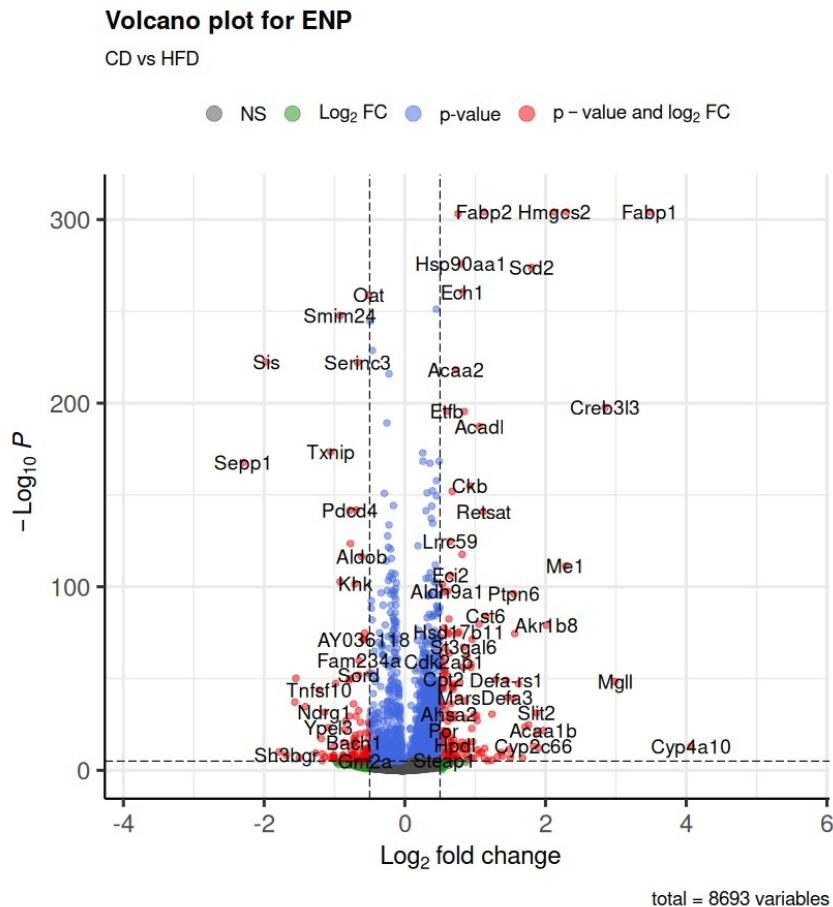


Figure 7

GO enrichment analysis of DEGs in enterocyte progenitors revealed extensive alterations across BP, CC, and MF categories in response to the prediabetic HFHSD model (Figure 8). The analysis uncovered 22 distinct clusters of enriched GO terms with trends in x y and z consistent with findings of the volcano plot. As seen in the previous GO enrichment analysis of ISCs, in this analysis of enterocyte progenitors, the top 50 GO terms from each class were selected by adjusted p-value, and thus all terms are highly significant.

Cluster 3 emerged as a focal point of upregulation, predominantly in lipid metabolism processes. This cluster showed enrichment in fatty acid beta-oxidation, lipid catabolism, fatty acid metabolism, and monocarboxylic acid transport. Notably, long-chain fatty acid metabolic process and fatty acid oxidation terms displayed high enrichment scores. These processes were primarily associated with the peroxisome cellular component. The cluster also revealed upregulation of molecular functions crucial for lipid processing, including

carboxylic ester hydrolase activity, fatty acyl-CoA hydrolase activity, NADP binding, and various hydrolase and oxidoreductase activities.

In contrast, cluster 5 exhibited consistent downregulation across all three GO categories, suggesting a coordinated change in cell adhesion and signaling. Biological processes affected included transmembrane receptor protein tyrosine kinase signaling pathway, positive regulation of cell migration, cell junction organisation, and actin filament-based processes. The analysis revealed downregulation of cellular components including adherens junctions, focal adhesions, and the actin cytoskeleton. Corresponding molecular functions, such as protein kinase activity, transmembrane signaling receptor activity, and cell adhesion molecule binding, were also downregulated.

Several other smaller clusters also displayed HFHSD induced alterations in enterocyte progenitor functions. Cluster 22 displayed a downregulation of calcium-dependent protein binding molecular functions, affected calcium-mediated cellular processes. Cluster 21 indicated reduced cysteine-type endopeptidase regulator activity involved in apoptotic processes. Cluster 20 revealed downregulation of beta-catenin binding molecular function. Additionally, Cluster 16 demonstrated a set of downregulated GO terms associated with DNA-binding transcription repressor activity, including specific RNA polymerase II-related transcription factor activities.

GO Term

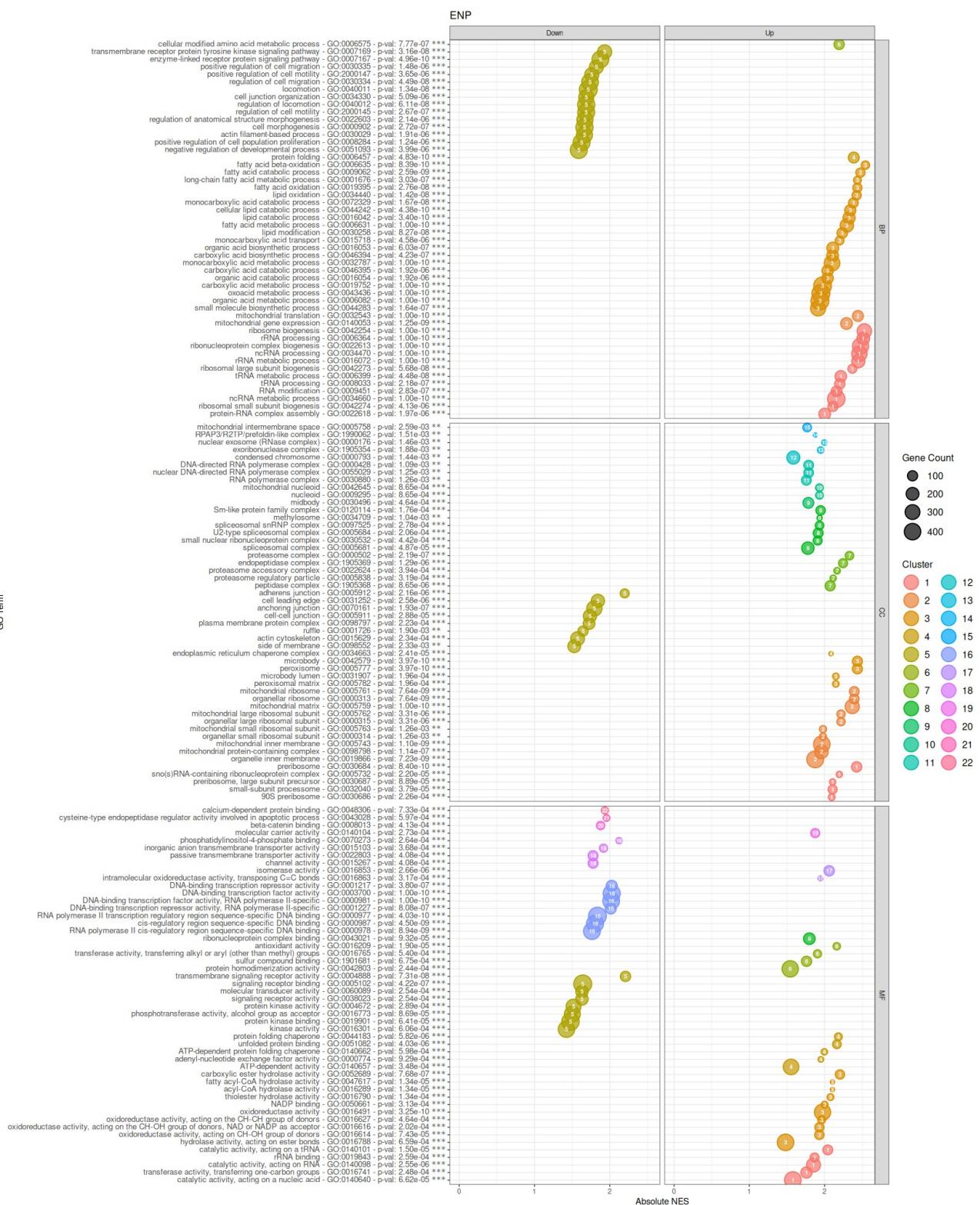


Figure 8

KEGG pathway enrichment analysis of enterocyte progenitors revealed 32 significantly enriched pathways across various categories (Figure 9). The figure displays 32 significantly enriched KEGG pathways from various categories such as cellular processes, Environmental information processing, genetic information processing, human diseases, metabolism, and organismal systems. Notably, the peroxisome pathway from the cellular processes category is enriched, showing a wide distribution of upregulated DEGs with a gene ratio of approximately 0.5. The Notch signaling pathway is also significantly enriched, exhibiting the highest gene ratio in its category and a p-value of 2.78e-02.

In the genetic information processing category, the proteasome pathway and protein processing in the endoplasmic reticulum show significant enrichment, with high gene ratios (approximately 0.9 for the proteasome and approximately 0.6 for protein processing in the ER). The proteasome pathway has a p-value of 5.47e-03, while the ER protein processing pathway has a p-value of 2.20e-02.

In the metabolism category, 10 pathways are enriched, all showing a predominance of upregulated DEGs. Of these, six belong to the lipid metabolism subcategory, each displaying high gene ratios. Retinol metabolism is particularly significant, with a p-value of 1.64e-04. In the amino acid metabolism subcategory, glutathione metabolism is enriched with a p-value of 3.52e-02, alongside the valine, leucine, and isoleucine degradation pathway, which has a p-value of 1.14e-02.

Additionally, 10 enriched pathways fall under the organismal systems category, with two in the digestive system, six in the endocrine system, and the remaining two in the sensory and immune system subcategories. The fat digestion and absorption pathway is highly significant ($p=7.00e-04$) and displays a majority of genes in the upregulated region, while the carbohydrate digestion and absorption pathway ($p=1.18e-02$) predominantly shows downregulated genes. Within the endocrine system, the PPAR signaling pathway is the most significant, with an extremely low p-value of 1.48e-08.

These trends display vast alterations in metabolism in the prediabetic enterocyte progenitor population with widespread upregulation of lipid metabolism and digestion pathways. Furthermore, upregulatory trends are apparent in the endocrine system pathways, and protein folding, sorting and degradation pathways. Moreover, the prediabetic enterocyte population display downregulatory trends in notch signalling, and carbohydrate digestion.

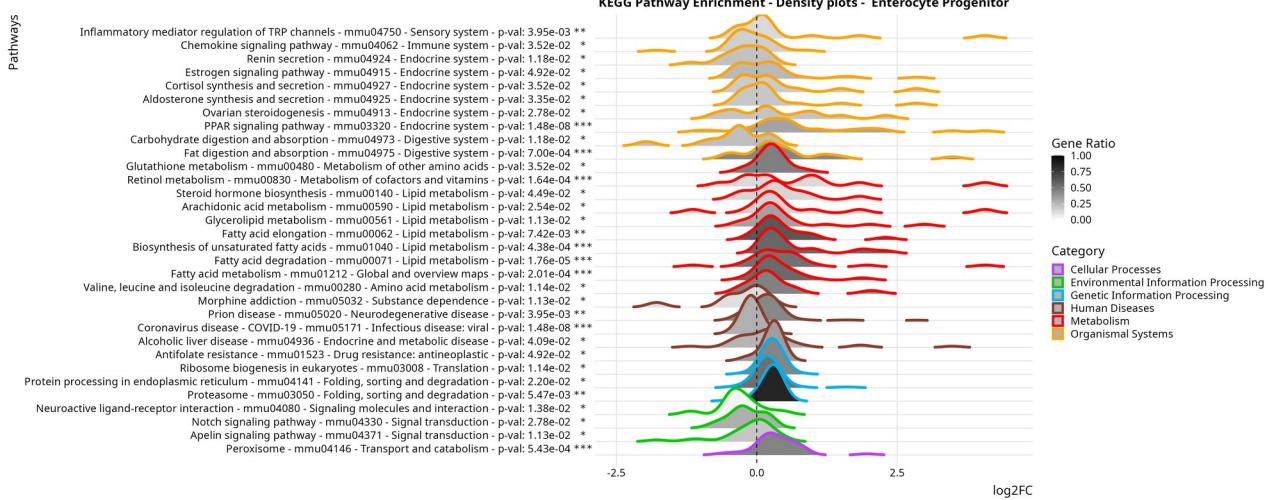


Figure 9

Further examination of the tight junction pathway through KEGG graphs and heatmaps of log 2 fold change values of DEGs provided additional insights into the mechanisms and components altered under the prediabetic condition (Figure 10).

In enterocyte progenitors, several claudin genes showed notable downregulation, particularly *Cldn15*, *Cldn4*, and *Cldn3*. The mature enterocyte population exhibited downregulation of *Cldn23*. Genes related to actin cytoskeleton also displayed significant downregulation in both enterocyte and enterocyte progenitor populations, with *Actb*, *Actg1*, and *Actn4* showing the most pronounced decreases.

Conversely, genes encoding alpha-tubulin proteins showed upregulatory trends across multiple cell types. *Tuba4a* and *Tuba1c* were significantly upregulated in most cell populations examined. *Tuba1b* showed specific upregulation in the ISC and enterocyte progenitor populations.

The KEGG graph visualisation of the tight junction pathway (Figure 10) provided further insights into the altered gene expression patterns. It highlighted the upregulation of PP2A, known to interact with occludin. The graph also illustrated downregulation of claudins in the tight junctional space. Furthermore, the visualisation indicated decreased expression in genes associated with cell polarity and paracellular permeability mechanisms within the tight junction complex. Collectively, these results demonstrate substantial changes in the expression of tight junction-related genes in the intestinal epithelium under HFHSD

conditions, with distinct patterns observed across different cell types and specific pathway components.

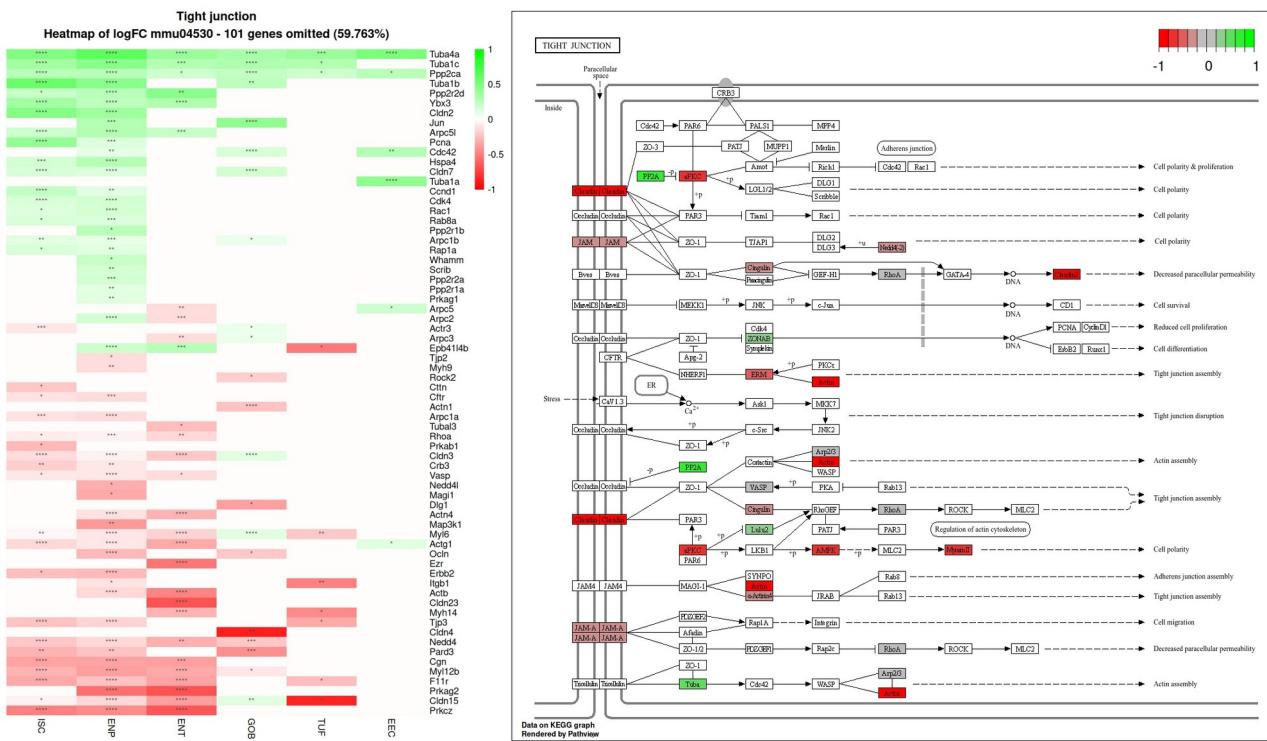


Figure 10

Analysis of the insulin signaling pathway, carbohydrate digestion and absorption, and fatty acid degradation pathways revealed significant alterations in gene expression patterns across different cell types in response to the HFHSD condition (Figure 11).

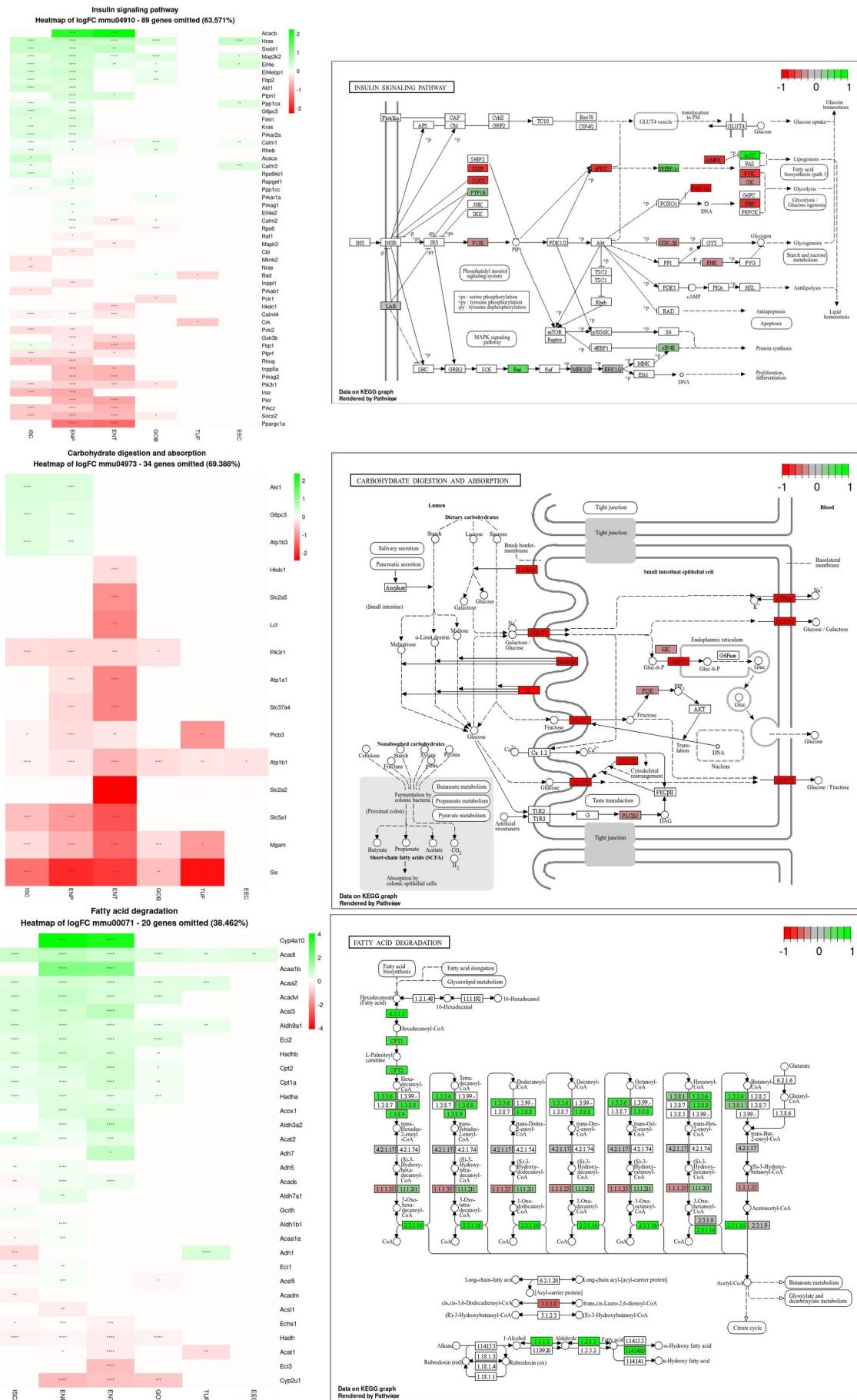
In the insulin signaling pathway, notable changes were observed in genes related to fatty acid biosynthesis and glycolysis/gluconeogenesis. The AMPK protein showed downregulation, while the Acetyl-CoA carboxylase (ACC) protein exhibited upregulation. Sreb-1c gene expression was increased in ISCs, enterocyte progenitors, and enterocytes, coinciding with decreased expression of the pyruvate kinase (PYK) protein. The *Hras* gene, involved in proliferation and differentiation pathways, showed upregulation in these cell types.

The carbohydrate digestion and absorption pathway analysis revealed significant changes, particularly in enterocytes. The *Sis* gene was markedly downregulated across all cell types except enteroendocrine cells (EECs). The *Slc2a2* gene, encoding the GLUT2 protein, a

glucose/fructose transporter, showed substantial downregulation in the heatmap and KEGG graph visualisations.

Examination of the fatty acid degradation pathway in enterocytes revealed widespread changes in enzyme expression. Upregulated enzymes included long-chain fatty acid-CoA ligase (EC:6.2.1.3), carnitine O-palmitoyltransferase 1 and 2 (CPT1, CPT2), fatty acyl-CoA oxidase (EC:1.3.3.6), long-chain acyl-CoA dehydrogenase (EC:1.3.8.8), very-long-chain acyl-CoA dehydrogenase (EC:1.3.8.9), short-chain acyl-CoA dehydrogenase (EC:1.3.8.1), acetyl-CoA C-acyltransferase (EC:2.3.1.16), alcohol dehydrogenase (EC:1.1.1.1), aldehyde dehydrogenase (NAD⁺) (EC:1.2.1.3), and long-chain fatty acid omega-monooxygenase (EC:1.14.14.80). Conversely, Delta3-Delta2-enoyl-CoA isomerase (EC:5.3.3.8) and alkane 1-monooxygenase (EC:1.1.1.35) showed downregulation.

These results collectively demonstrate substantial alterations in gene expression patterns related to insulin signaling, carbohydrate metabolism, and fatty acid degradation in the intestinal epithelium under HFHSD conditions, with distinct changes observed across different cell types and specific pathway components.



50 *Figure 11*

Alterations in the Endoplasmic Reticulum and Proteasome

The KEGG graph for the enterocyte progenitors highlights widespread upregulation across the protein processing pathway in the endoplasmic reticulum (ER), with notable enhancements in the ubiquitin ligase complex, ER-associated degradation (ERAD), and protein recognition by luminal chaperones (Figure 12). Within the ubiquitin ligase complex, proteins such as Hsp40, Hsp70, and CHIP are significantly upregulated. CHIP, known for its dual role as a co-chaperone and E3 ubiquitin ligase, is crucial for the degradation of chaperone-bound misfolded proteins, indicating increased proteostasis in response to ER stress. In ERAD and chaperone recognition, key upregulated proteins include Hsp70, Hsp90, Hsp40, and nucleotide-exchange factor (NEF), encoded by *Bag1*, which facilitates the release of ADP from HSP70 proteins, thus promoting protein folding and release. Additional upregulation is seen in Derlin, encoded by *Derl2*, which plays a role in the degradation of misfolded glycoproteins, and DOA1, encoded by *Plaa*, which is involved in ubiquitin-mediated protein degradation.

The accompanying heatmap reinforces this upregulation, showing increased expression levels across key genes involved in protein processing within the ER pathway. Genes such as *Hspa1a*, *Hspa1b*, *Hsp90aa1*, *Hspf1* exhibit marked upregulation, particularly in the enterocyte progenitors and enterocytes populations.

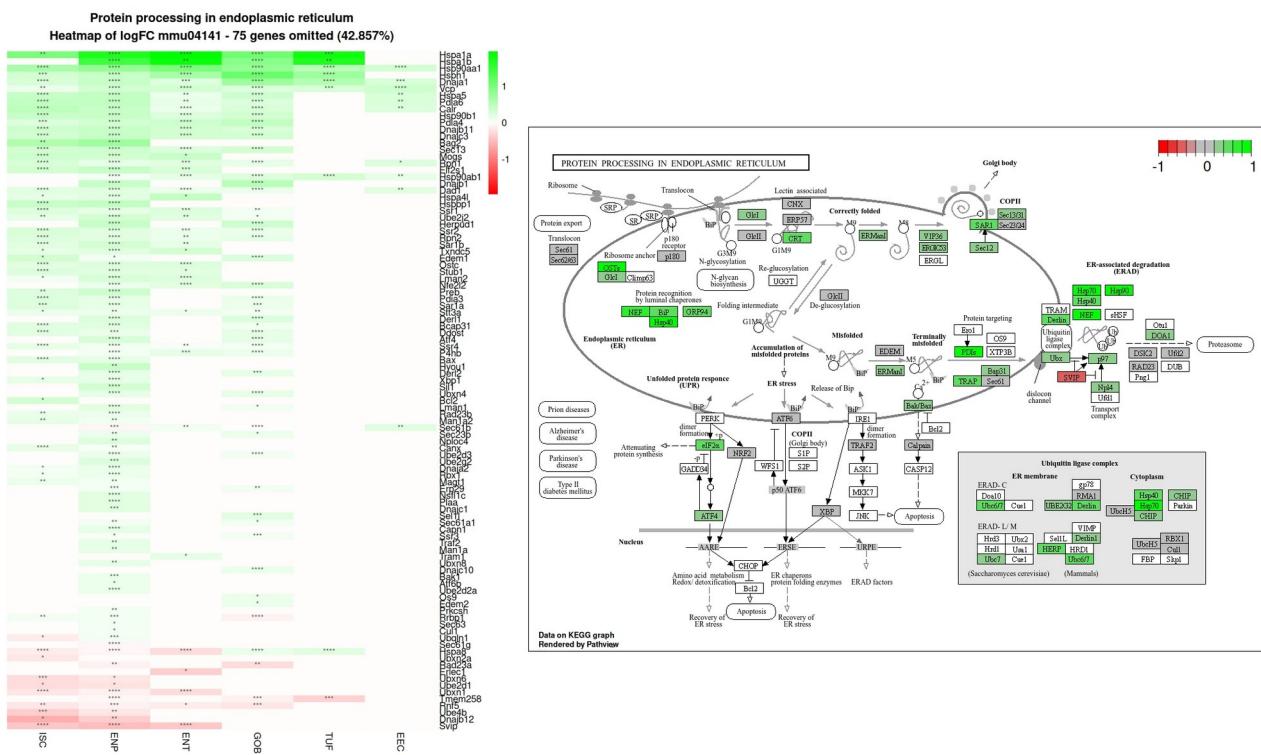


Figure 12

Following on from the ER response in protein processing, the proteasome pathway also exhibits extensive upregulation across all cell types, particularly in the 20S core particle, which involves both the alpha and beta ring subunits (Figure 13). This upregulation is mirrored in the 19S regulatory particle, especially in the PA700 lid and PA700 base subunits, indicating a broad activation of proteasome-mediated protein degradation via the HFHSD condition. Contrastingly, there is a slight downregulation observed in the immunoproteasome subunit, specifically the β 5i subunit encoded by Psmb8, within the enterocyte progenitors and enterocytes populations.

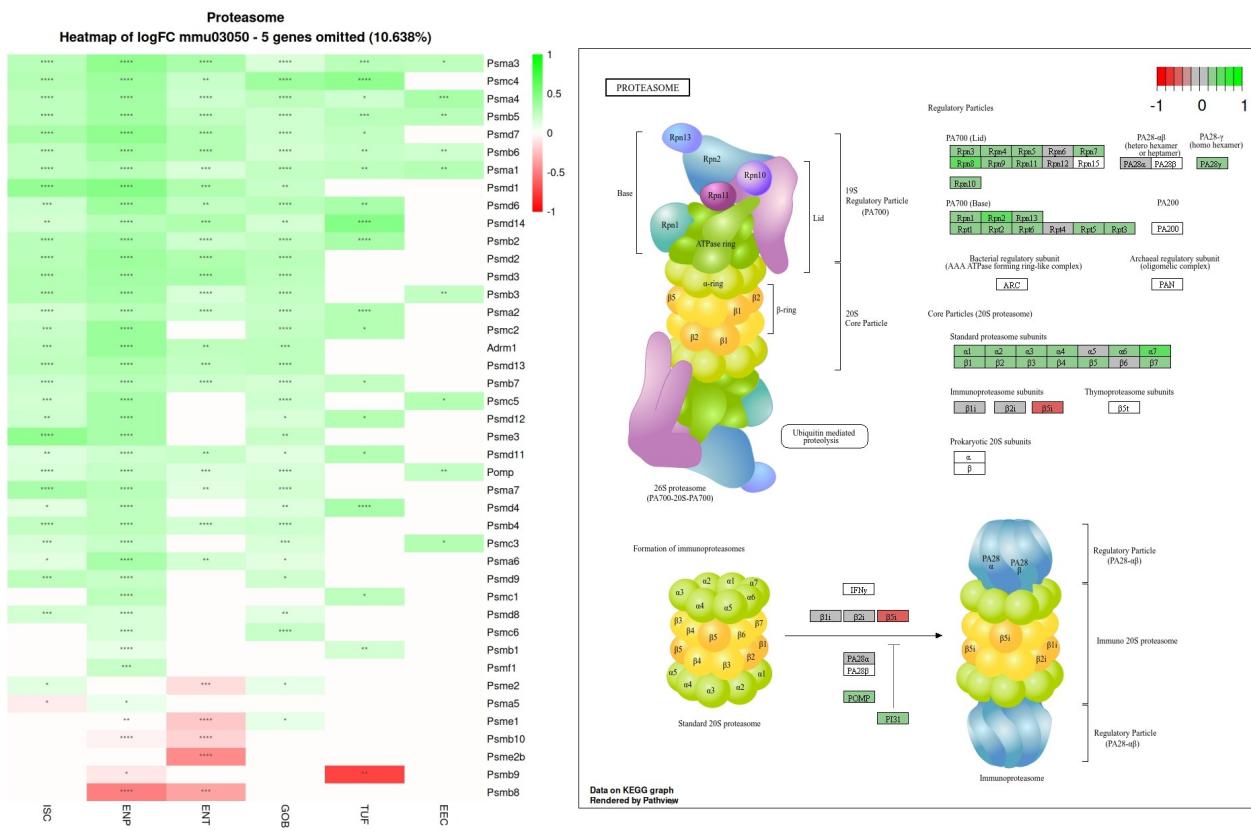


Figure 13