## Results (1343 words without figures/tables + their descriptions)

From the pre-processing of the subject datafile, 60 subjects remained of which 32 were IR and 28 were IS. The corresponding subject IDs were used to filter the proteome and metabolome data for the DAs, combined pathway analysis and network analysis. This resulted in 686 samples (316 IR and 370 IS) being included in the proteome data and 726 samples (425 IR and 302 IS) in the metabolome data. Moreover, the proteome data included 302 proteins annotated with HGNC symbols while the metabolome data was reduced from 724 metabolites to 323 metabolites after excluding those with no HMDB identifier specified.

After making the metagenomic and metabolomic datafiles consistent in order to run the integrative analysis, both datasets contained 49 subjects (26 IR and 23 IS) and 441 samples. This number of samples was reduced to 402 (200 IR and 202 IS) after excluding those with a PCA Axis1 value of less than -2.8. Despite excluding metabolites that had an abundance of 0 across many samples, the number of metabolites used in this integrative analysis remained the same (323 metabolites). Finally, filtering of the metagenomic data to only include sufficiently classified and abundant taxa, resulted in the phyloseq object containing 362 taxa.

### *Identification of differences in the gut microbiome of IR and IS subjects*

To determine whether differences in microbiome composition exist between the 2 groups, Bray-Curtis dissimilarities were illustrated using a PCoA. This PCoA demonstrated some clustering of the 2 groups, with 2 clusters being able to be distinguished per group (Figure 2). The separation seen in this plot was deemed to be significant by the PERMANOVA (p = 0.010) and was found to likely not be due to a significant difference in the variance of the groups by the multivariate analysis of the group dispersion homogeneity (p = 0.932). Nevertheless, there did exist a large variation in the absolute microbial abundances and absolute phyla abundances across both the IR and IS samples (Figure 3).



*Figure 2: Principal Coordinate Analysis (PCoA) based on Bray-Curtis dissimilarity illustrating the separation in the microbial composition of the insulin resistant (IR) and insulin sensitive (IS) samples. Variation explained by the corresponding principal coordinates are given in %.*



*Figure 3: bar plots depicting the variation in the absolute microbial abundance and in the absolute phyla abundances across the A) insulin resistant (IR) samples and B) the insulin sensitive (IS) samples.*

The mean abundance of each phyla in the different groups was investigated to determine which phyla were most likely to contribute to the distinction between the IR and IS metagenomes. Small differences in mean abundance between the 2 conditions were observed for the Actinobacteria, Firmicutes and Proteobacteria phyla, with the Actinobacteria (Figure 4.A) and Firmicutes (Figure 4.C) both being slightly more abundant and the Proteobacteria (Figure 4.D) slightly less abundant in the IR condition. However, only the differences in the Firmicutes and Proteobacteria phyla were significant (p = 0.009 and p < 0.0001, respectively).

More notable differences were seen for the remaining phyla. The Bacteroidetes were less abundant in the IR samples compared to the IS samples (Figure 4.B) while the Verrucomicrobia were more abundant in the IR samples (Figure 4.E). However, only the difference in Bacteroidetes was significant (p < 0.0001).

****

*Figure 4: box plots representing the differences in mean phylum abundance between the insulin resistant (IR) and insulin sensitive (IS) group. Significant p-values (p < 0.05) are indicated by \*\*\* = p < 0.0001, \*\* = p < 0.001 or \* = p < 0.01.*

To further investigate the taxa separating the groups, the top 20 taxa generated by the PERMANOVA were identified in order to try recognise the microbes contributing the most to the metagenomic differences between the 2 groups. The majority (70%) of these microbes belonged to the Firmicutes phylum. Consistent with the mean phyla abundance boxplots (Figure 4), the microbe most abundant in the IR condition compared to the IS condition was of the Bacteroidetes phylum while the microbe least abundant in the IR condition was of the Firmicutes phylum. 3 of the 10 microbes more abundant in IR and none of those less abundant in IR were Ruminoccocus. On the other hand, 2 out of the 10 microbes less abundant in IR and none of those more abundant in IR were Lachnospira. (Figure 5)



*Figure 56: bar plot of the taxa generated by the multivariate ANOVA with permutations (PERMANOVA) deemed to contribute the most to the separation between the insulin resistant and insulin sensitive groups. Taxa with negative values are less abundant in the IR condition and taxa with positive values are more abundant in the IS condition. Blue bars represent taxa belonging to the Bacteroidetes phylum, orange bars represent taxa belonging to the Firmicutes phylum and green bars represent taxa belonging to the Proteobacteria phylum.*

### *Identification of differences in the proteome and metabolome of IR and IS subjects*

From the DA using limma, 23 proteins were deemed to be significantly altered in the IR condition compared to the IS condition based on p-value. 13 of these proteins were more abundant in the IR subjects (logFC > 0) and 10, less abundant (logFC < 0) (Appendix A4). When a sufficient change was deemed to be a logFC > 0.50 or < -0.50, 3 of the significantly altered proteins remained: LPA and SHBG were both less abundant in the IR condition (logFC = -1.074 and -0.522, respectively) while APOC4 was more abundant in the IR condition (logFC = 0.624). Despite these proteins having significant p-values, it should be noted that non possessed a significant adjusted p-value.

On the other hand, the DA using MetaboDiff generated 40 metabolites whose abundances were significantly altered (p < 0.05) in the IR group compared to the IS group: 21 were less abundant (logFC < 0) and 19 more abundant in the IR condition (logFC > 0) (Appendix A5). Based on a logFC > 0.05 or < -0.05, 9 out of the 21 less abundant metabolites and 3 out of the 19 more abundant metabolites were sufficiently changed (Table 1). Unlike the results of the proteomic DA, 7 metabolites possessed a significant adjusted p-value with 5 also having a logFC suggesting a sufficient change in abundance (Table 1). The remaining 2 metabolites with a logFC > 0.05 or < -0.05 were HMDB02759 (logFC = -0.327) and HMDB00705 (logFC = 0.212).

*Table 1: table specifying the sufficiently less abundant (logFC < -0.50, blue cells) and sufficiently more abundant (logFC > 0.05, red cells) metabolites as determined by the differential analysis (DA) using MetaboDiff. The logFC and adjusted p-value are given for each metabolite. All metabolites listed were significantly changed in the IR condition compared to the IS condition based on p value (p < 0.05).*

|  |  |  |
| --- | --- | --- |
| Metabolite | logFC | Adjusted p-value |
| LysoPE(0:0/22:6(4Z,7Z,10Z,13Z,16Z,19Z)) (Lysophospholipid) | -1.336 | 0.005 |
| LysoPE(0:0/16:0) (Lysophospholipid) | -1.010 | 0.281 |
| LysoPE(P-16:0/0:0) (Phospho-ether lipid) | -1.089 | 0.042 |
| Ethylmalonic acid | -0.795 | 0.040 |
| 2-arachidonoylglycerol | -0.758 | 0.005 |
| 5-dodecenoic acid | -0.720 | 0.162 |
| MG(0:0/14:1(9Z)/0:0) (Monoacylglycerol) | -0.683 | 0.026 |
| LysoPE(0:0/20:3(11Z,14Z,17Z)) (Lysophospholipid) | -0.591 | 0.158 |
| LysoPE(0:0/22:0) (Lysophospholipid) | -0.553 | 0.326 |
| Glucaric acid | 0.556 | 0.326 |
| 2-trans,4-cis-Decadienoylcarnitine (Fatty acid ester) | 0.604 | 0.106 |
| LysoPC(20:0/0:0) (Lysophospholipid) | 0.867 | 0.326 |

### *Identification of altered biological pathways involved these metabolomic and proteomic changes*

Combined pathway analysis of the proteome and metabolome data was done using PathVisio and MetaboAnalyst.

Out of all the 323 metabolites and 302 proteins run through PathVisio, only 15 metabolites and 25 proteins met the expression criterion (p < 0.05). The overrepresentation analysis of these compounds found 18 pathways to contain significantly more changed compounds than expected (Z-score > 1.96). All these pathways possessed 1-2 compounds that had significantly different levels in the IR subjects compared to in the IS subjects. The majority of the significantly altered pathways involved metabolites that met the expression criteria but none of these pathways contained both proteins and metabolites with a p < 0.05 (Table 2).

*Table 2: Significantly altered pathways (1.96 < Z-score < -1.96) and the associated significantly altered compounds (p < 0.05) as determined by combined pathway analysis using PathVisio. The common name for each metabolite is given while proteins are annotated using HGNC symbols. Arrows indicate* the compound being more abundant in insulin resistance (**↑)** or less abundant in insulin resistance (**↓**), in comparison to the insulin sensitive condition.

|  |  |  |
| --- | --- | --- |
| Pathway | Z-score | Proteins and Metabolites |
| Gastric acid production | 3.05 | **↑** Acetylcholine |
| Heroin metabolism | 3.05 | **↑** Pseudo-ChE |
| IL-1 signaling pathway | 3.05 | **↓** IL1RAP (CHECK) |
| Irinotecan pathway | 3.05 | **↑** BCHE |
| Monoamine GPCRs | 3.05 | **↑** Acetylcholine |
| Oligodendrocyte Specification and differentiation(including remyelination), leading to Myelin Components for CNS | 3.05 | **↓** Sphingomyelin (d18:0/18:1(11Z))  HMDB12088  (CHECK!!!) |
| Phosphodiesterases in neuronal function | 3.05 | **↑** L-Glutamate |
| Secretion of Hydrochloric Acid in Parietal Cells | 3.05 | ↑ Acetylcholine |
| Signal transduction through IL1R | 3.05 | ↓ IL1RAP |
| Sphingolipid pathway | 3.05 | ↓ Sphinganine |
| Structural Pathway of Interleukin 1 (IL-1) | 3.05 | **↓** IL1RAP |
| Thermogenesis | 3.05 | **↓** 2-Arachidonoylglycerol |
| Vitamin D-sensitive calcium signaling in depression | 3.05 | **↑** Acetylcholine |
| Ebola Virus Pathway on Host | 2.3 | **↓** GSN |
| Phosphatidylcholine catabolism | 2.3 | **↑** LysoPC(20:0/0:0)  **↓** Sphingomyelin (d18:0/18:1(11Z))  HMDB12088  (CHECK!!!) |
| PPAR Alpha Pathway | 2.3 | **↓** APOA1  **↓** PLTP |
| Senescence and Autophagy in Cancer | 2.3 | **↑** VTN  **↓** GSN |
| Glutathione metabolism | 1.97 | **↑** L-Glutamate  5-Oxoproline |

MetaboAnalyst was unable to recognise one metabolite out of the 40 significantly altered metabolites (HMDB61112), hence, meaning 39 metabolites and 23 proteins were used in the corresponding joint pathway analysis. 6 pathways were found to be significantly changed (p < 0.05) in the IR subjects compared to the IS subjects. Like PathVisio, all these pathways concerned 1-2 significantly altered compounds, however, all were metabolites. Furthermore, the glutathione metabolism pathway as well as pathways associated with sphingolipids were present in the results of both programmes (Table 3).

*Table 3: List of significantly altered pathways ( p < 0.05) from the combined pathway analysis using MetaboAnalyst. The compounds whose abundance is significantly different between the insulin resistant (IR) and insulin sensitive (IS) subjects are given for each corresponding pathway. Arrows indicate compounds more abundant (*↑) *and less abundant (*↓) *in the IR condition.*

|  |  |  |
| --- | --- | --- |
| Pathway | p-value | Proteins and Metabolites |
| Porphyrin and chlorophyll metabolism | 0.019 | **↓** Biliverdin (isomer 1)  **↑** Biliverdin (isomer 2)  **↑** L-Glutamate |
|  |
| Glutathione metabolism | 0.021 | **↑** L-Glutamate  **↑** 5-Oxoproline |  |
|  |
| Sphingolipid metabolism | 0.022 | **↓** Sphinganine  **↓** Sphingomyelin (d18:0/18:1(11Z)) |  |
|  |
| Nitrogen metabolism | 0.040 | **↑** L-Glutamate |  |
| D-Glutamine and D-glutamate metabolism | 0.040 | **↑** L-Glutamate |  |
| Glycerophospholipid metabolism | 0.046 | **↓** 1-Acyl-sn-glycero-3-phosphocholine  **↑** Acetylcholine |  |
|  |

Since overrepresentation analysis can bury information concerning pathways that it deems to not be significantly altered, an extended network analysis of the significantly altered proteins was performed to uncover alterations in biological pathways that may have been lost by this analysis. The results revealed the human complement system and the complement and coagulation cascades pathways to possess the highest out-degree values: the first of these pathways having a value of 5 and the second a value of 4. Both of these pathways involved CFH and F10, with the 3 remaining proteins involved in the human complement system being C4A, APOA1 and VTN and the 2 remaining proteins in the complement and coagulation cascades pathway being SERPIND1 and C1QB.

### *Identification of associated microbes and metabolites*

Analysis of which features explain the covariation between the metabolomic and metagenomic data generated 13 microbes and 11 metabolites. 8 of these metabolites were monoacylglycerols with the remaining metabolites being arachidonic acid and 2-arachidonoylglycerol (2-AG). One of these monoacylglycerols (MG(0:0/14:1(9Z)/0:0)) as well as 2-AG were also shown to have significantly different abundances between the 2 groups by the DA.

All 13 of the covariance-explaining microbes were Firmicutes. This coincides with the majority of the taxa contributing most to the metagenomic separation of the IR and IS group (Figure 5). 6 of these 13 Firmicutes were Faecalibacterium prausnitzii, 3 were Oscillospira and 2 Coprococcus with the remaining 2 taxa being Dorea and Blautia producta.