**Gene expression analysis in type 2 diabetes patients from Abu Dhabi**

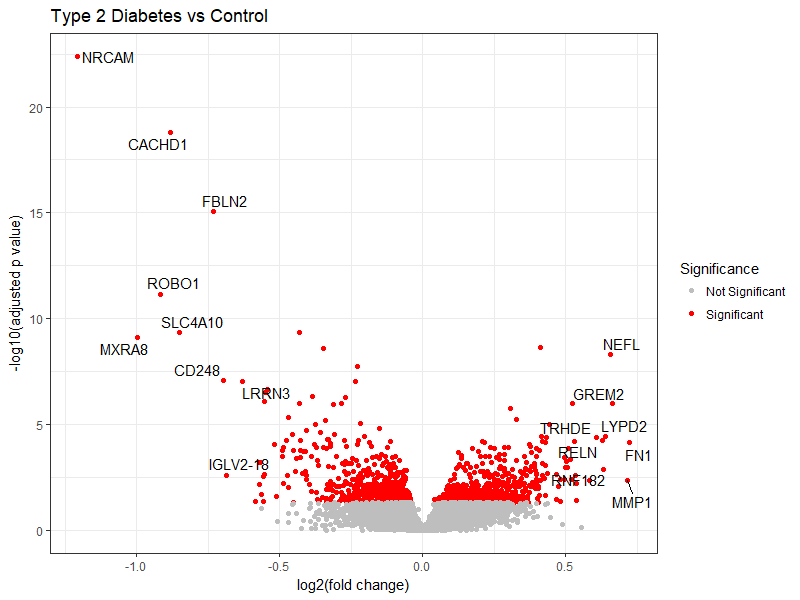
**Background**

Gene expression is under precise spatial and temporal control, and disruption of normal regulatory networks can have severe consequences on phenotype. In a disease context, differential gene expression analysis highlights genes and biological functions that are particularly relevant to pathogenesis, enhancing understanding of disease mechanisms and informing drug development. Additionally, the majority of genetic risk factors in complex disease are thought to act through regulation of gene expression. Exploring the relationship between genetic variation and gene expression in a disease context can therefore help interpret such associations.

**Data Generation and Preliminary Findings**

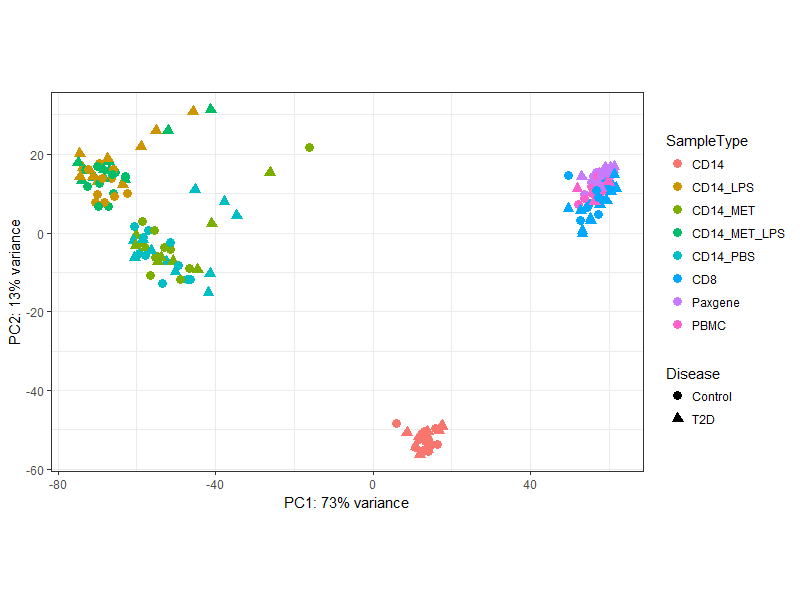
RNA sequencing is a powerful approach to quantify expression of all transcripts in an unbiased manner, allowing discovery of novel genes, comparison of different isoforms, and identification of regulatory elements. We have now carried out RNA-seq for a total of 806 samples from the Abu Dhabi type 2 diabetes cohort, divided into two complementary data sets.

The PaxGene data set comprises gene expression measurements in whole blood from 620 case and control individuals, including a small number of family trios. Initial analysis comparing type 2 diabetes patients to non-diabetic individuals identifies a large number of differentially expressed genes (Figure 1), including *TRHDE* and *SLC4A10*, which have been previously associated with diabetes and BMI respectively through GWAS.



**Figure 1:** Volcano plot summarising differential gene expression between cases and controls, with –log10(adjusted p value) for each gene tested plotted against its log2 fold change (positive fold change indicates higher expression in type 2 diabetes). Significance is determined using FDR<0.05.

Given the complex nature of signalling networks in blood cells and the mixed cell populations involved, analysis of sorted cells is helpful in identifying cell-specific effects. The cell stimulation data set comprises 186 samples from 19 individuals, allowing comparison of gene expression between type 2 diabetes and controls in CD14+, CD8+, and PBMCs as well as whole blood. This demonstrates that gene expression is highly cell specific, and cell type has an impact on the specific genes that are differentially regulated in disease (Figure 2). Additionally, CD14+ cells were treated with LPS and metformin to test for differential responses according to disease status. Whilst global changes were observed with treatment, this response did not differ significantly between cases and controls. However, power calculations suggest that additional samples are required for confidence in this finding.



**Figure 2:** Principal component analysis of RNA-seq count data from cell stimulation experiment shows the relationship between sample types on the basis of global gene expression.

**Future Plans**

We will continue to analyse the gene expression data, leveraging the rich clinical information available to better understand the differences observed. Results from the cell stimulation data set will be used to inform selection of additional samples for further RNA-sequencing. We will also integrate genotyping information from patients and controls to identify regulatory variants through association with gene expression. We hypothesise that we will find examples of population-specific gene regulation, and of disease risk variants that are associated with the expression of biologically relevant genes in the context of diabetes.