Shilpa’s Olink Data Analysis Report

Approach

There were quite a lot of different conditions for these experiments so I decided to mainly make use of ANOVAs to help disentangle these conditions and find the true drivers for differential protein expression. The main aims going in were:

|  |  |  |
| --- | --- | --- |
| **Aim** | **Variable Name** | **Conditions** |
| Determine the effect of media fat and sugar concentration on inflammatory protein release in each cell type | Media | MO – Media Only  LFLS – Low Fat, Low Sugar  HFHS – High Fat, High Sugar |
| Determine the differences in inflammatory protein release across the cell types | Cell Origin | HO – Hepatocyte Only  H – Hepatocytes in co-culture  L – Endothelial cells in co-culture |
| To determine if there is any effect of fatty acid composition on inflammatory protein release | Fat | O – OPLA  P – POLA |

Early analysis showed that Fat composition had very minimal effect and so it was separated from Media variable. A final aim was also to examine any Media : Cell Origin interactions where cell types responded differently to changes in Media composition.

Methods

Data was kindly provided by Shilpa 😊, manipulated on Excel and analysed using the “OlinkAnalyze” package on R version 4.2.3. Firstly, to account for proteins which were present in the serum that was added to the cell media, an average was taken from samples of stock media and subtracted from our sample data for all proteins. All samples passed the Olink QC however some samples were below the limits of detection for the assay. Where >50% of the samples in an assay were below this limit the entire assay was removed, otherwise all other samples and assays were included.

For analysis of variables with multiple levels or analysis of multiple variables at once, one-way or two-way ANOVAs were performed (respectively) across all assays and used the Benjamini & Hochberg (1995) False Discovery Rate (FDR) method to adjust for multiple comparisons. Significant assays then underwent post-hoc pairwise T tests using the Tukey correction for multiple comparisons. For comparison of variables with two levels, Welch 2-sample t-tests were performed using the same FDR method to adjust for multiple comparisons. Significance was set to p < 0.05.

Results

**Principle Component Analysis (PCA)**

PCA plots were generated using the entire dataset based on relative protein concentrations. Principle component 1 (PC1) clearly identified as corresponding to Cell Origin variable and accounted for 43.38% of the total data variance (*fig.1*). Principle component 2 (PC2) was identified as Media variable and accounted for 20.96% of total data variance. (*fig. 2*).

**Cell Origin**

This variable had the biggest effect on the dataset and was investigated first. A two-way ANOVA of variables Media and Cell Origin showed that 96% of proteins were significantly effected by Cell Origin. Of these, the majority had the same pattern of protein expression with hepatocyte monoculture being highest, followed by hepatocyte co-culture and then endothelial cell co-culture (*fig.3*). However there were a few exceptions were the co-cultured cells had higher protein expression than the monoculture or the endothelial cells had significantly different protein expression to the hepatocytes.