SUPPLEMENTARY RESULTS

Overview of data quality

Heat map of the top 25% of most abundant proteins shows a lack of a distinguishing pattern over time (supplemental figure 2B), and the clustering of biopsy proteomes by diagnosis rather than by MS order (supplemental figure 2C) indicate the consistency of the MS performance and high reproducibility of the sample preparation. Moreover, greater than 97% of the biopsy profiles have a correlation of ≥ 0.7 (supplemental figure 2D, E), indicating that despite the macroscopically observable differences between control and IBD patients, no gross proteomic differences appeared between the samples. Notably, the low correlation was observed primarily because of the CoN and CoA biopsies from one CD patient (supplemental figure 2D). These biopsies were not identified by outlier analyses of the MS profile, suggesting that the performance of the MS was not different, but rather an indication of altered pathology or protein expression in this particular patient.

There were 3929 proteins quantified by MaxQuant analysis, 3644 of which were identified by $\geq$ 2 unique peptides and thus included for further analyses. 740 proteins were found to be significantly different by ANOVA analysis between the four groups (Control, CD CoN, CD CoA, UC CoA) after FDR adjustment with Benjamini-Hochberg[1] (supplemental table 2). This high number of significant proteins is reflective of the differences in affected and non-affected tissues originating from CD or UC patients and control subjects respectively. Student's t-test analysis of CD CoA vs UC CoA proteins identified 379 to be significantly different, though only 17 remained so following FDR adjustment, exemplifying the high degree of similarity between CD and UC. There were 225 proteins considered to be subgroup specific due to the overrepresentation in one subgroup (>70% of subgroup biopsies) when compared with at least one other subgroup (<50% of subgroup biopsies) (supplemental table 2). Based upon GOBP and KEGG categorization, approximately 10% (342) of the quantified proteome are involved in immunological processes, Of these immunological proteins, nearly 50% are involved in metabolism, 41% in cellular processes (eg cell communication), and 26% in response to stimuli (supplemental table 3). These proteins function predominantly in catalysis (as hydrolases) and in protein binding or signaling.

REFERENCES

1. Benjamini Y, Hochberg Y. On the adaptive control of the false discovery fate in multiple testing with independent statistics. Journal of Educational and Behavioral Statistics 2000;25:60-83.