Analyze omics from SLIMM-T2D trial

jd

2020-10-02

The installation of required R packages takes approximately 15 minutes. The RMD file can be knit to Word in approximately 1 minute.

## Install & load R packages

Install related R packages from GitHub, if needed. This is also described in the ezlimmaplot README at <https://github.com/jdreyf/ezlimmaplot>. Other packages can be installed with install.packages().

Load and attach the above and other R packages.

# Read data and take the difference from baseline per person

Clinical metadata at [clin\_metadata.csv](./data/clin_metadata.csv). This file also used for uploads to GEO. Here, change column names in workflow, so can use them more easily in ggplot2. After taking differences of phenotypes, we remove individual points from R object *diff.pheno*, who have NA in every measure, such as Topaz\_3.

The original SOMAscan dataset was parsed to write [soma\_annot.csv](./data/soma_annot.csv) and log2-transformed to write [soma\_mat.csv](./data/soma_mat.csv).

The original metabolomics dataset was parsed, filtered, and log2-transformed to write [met\_mat\_norm.csv](./data/met_mat_norm.csv). Annotations for metabolites did not originally have CHEBI annotations, so we added these automatically using CTSgetR. However, CTSgetR wasn’t given the chirality of many compounds, so we found that 3-hydroxyisobutyrate had CHEBI:11805, whereas SMPDB had (S)-3-Hydroxyisobutyric acid (CHEBI:37373), so we manually fixed 3-hydroxyisobutyrate to have CHEBI:37373, and wrote [met\_annot\_with\_chebi.csv](./data/met_annot_with_chebi.csv).

The table of numbers in each arm that had HbA1c measurements:

|  |  |  |
| --- | --- | --- |
|  | DWM | RYGB |
| 0 | 19 | 19 |
| 3 | 19 | 19 |
| 12 | 18 | 19 |
| 18 | 11 | 17 |
| 24 | 10 | 16 |
| 36 | 10 | 15 |

The table of numbers in each arm that were assayed by SomaLogic:

|  |  |  |
| --- | --- | --- |
|  | DWM | RYGB |
| 0 | 19 | 19 |
| 3 | 19 | 19 |
| 12 | 16 | 19 |
| 24 | 10 | 15 |
| 36 | 9 | 14 |

The able of numbers in each arm that were assayed by metabolomics:

|  |  |  |
| --- | --- | --- |
|  | DWM | RYGB |
| 0 | 19 | 19 |
| 3 | 18 | 18 |
| 12 | 16 | 19 |
| 36 | 9 | 13 |

Samples in 12mo metabolomics and proteomics are identical, so create diff\_mats at 1 year. Also, create a combined annotation.

Calculate metadata differences from baseline and remove samples with too many NAs, who shouldn’t be used in Hitman. Write [clin\_metadata\_diff.csv](./results/clin_metadata_diff.csv) for Table S1.

Test association of metabolite abundance to missingness

# Differential abundance between groups

Analyze differences between groups at each time point, and these differences after taking each person’s difference from baseline, with and without adjusting for each person’s change in BMI.

## Proteome

The Soma supp table with means per group per time point and statistics on differences between groups on differences from baseline is at [soma\_supp\_table.csv](./results/soma_supp_table.csv).

The Venn table using FDR but not logFC is soma\_FDR15\_venn.csv.

The heatmap w/ 19 analytes is at [soma\_FDR15\_fc50pct\_heat.pdf](./results/soma_FDR15_fc50pct_heat.pdf).

### Text

We summarize that the number of proteins significant at FDR 0.15:  
\* At baseline: 0  
\* At 3 mo in unadj: 14; whereas in adj: 8. In both, not inc. lfc, we have

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | RYGBvsDWMat3mo.sig | RYGBvsDWMat3mo\_bmi.sig | EntrezGeneSymbol | TargetFullName |
| SL004660 | 1 | 1 | IBSP | Bone sialoprotein 2 |
| SL000466 | 1 | 1 | IGFBP2 | Insulin-like growth factor-binding protein 2 |
| SL010458 | 1 | 1 | ESM1 | Endothelial cell-specific molecule 1 |
| SL000522 | 1 | 1 | MMP12 | Macrophage metalloelastase |
| SL006694 | -1 | -1 | CNDP1 | Beta-Ala-His dipeptidase |
| SL018548 | 1 | 1 | SERPINA3 | Alpha-1-antichymotrypsin complex |
| SL003187 | 1 | 1 | CCL22 | C-C motif chemokine 22 |
| SL006777 | -1 | -1 | FETUB | Fetuin-B |

* After baseline correction, there are 56 proteins that are significant at any time point without BMI adjustment. Of these, 19 have sufficient FC to be in the heatmap.

## Metabolome

Analyze differences between groups at each time point, and these differences after taking each person’s difference from baseline. The met supp table with means per group per time point and statistics on differences between groups on differences from baseline is at [met\_supp\_table.csv](./results/met_supp_table.csv).

We find that the results from not imputing missing values are very similar to those from imputing. For example, if we consider the top 20 metabolites between groups at 3 months, we find that

The Venn table using FDR but not logFC is met\_FDR15\_venn.csv.

The heatmap w/ 85 analytes is at [met\_FDR15\_fc50pct\_heat.pdf](./results/met_FDR15_fc50pct_heat.pdf).

### Text

We summarize that the number of proteins significant at FDR 0.15:  
\* At baseline: 0  
\* At 3 mo in unadj: 96; whereas in adj: 74. In both, not inc. log fold-change threshold, we have 45 metabolites.  
\* 3-hydroxyisobutyrate at 3mo:

|  |  |  |
| --- | --- | --- |
|  | RYGBvsDWMat3mo\_bmi.sig | RYGBvsDWMat3mo.sig |
| C1549 | -1 | -1 |

* After baseline correction, there are 85 proteins that are significant at any time point without BMI adjustment. Of these, 85 have sufficient FC to be in the heatmap.

# Correlate soma vs. met

We correlate top somalogic proteins with top metabolites. There are duplicate probes for one protein, so we remove the duplicated probe and only have one row for that protein.

Cluster proteins and metabolites.

Build corr matrix

Find strong correlations to comment upon

# Hitman

We apply Hitman to 12mo outcome change from baseline using change at 3mo from baseline as mediators and arm as exposure. The results of these mediation tests are written as CSVs to [results](./results/) folder using the naming convention outcome-timepoint\_vs\_mediator-timepoint\_hitman, such as “HbA1c12\_vs\_soma3\_hitman.csv,” whose *EMY* columns hold the overall p-values and FDRs, and corresponding chi-squared on 1 degree of freedom, where larger is more significant.

## HbA1c

We apply Hitman to 12mo HbA1c change using change at 3mo from baseline in analytes and clinical parameters as mediators.

We also test protein changes at 3 months vs. HbA1c changes at 24 months, and protein & metabolite changes at 36 months.

### Compare clinical vs analytes

The protein analytes more significant in p-value & FDR than any clinical mediator are:

|  |  |  |  |
| --- | --- | --- | --- |
| probe | ratio3 | nm | EntrezGeneSymbol |
| SL005168 | -0.163 | Growth hormone receptor | GHR |

The metabolite mediators are pro-hydroxy-pro

### Compare Hitman to joint significance

Top analytes: GHR has FDR 0.26 & pro-hydroxy-pro has FDR 0.73.

### Plot inconsistent mediation fig

We combine several mediation plots into the Hitman figure using ggplot2 facet\_wrap. We modify this figure by copy-pasting pieces into PPT, then adding text boxes for Inconsistent and Consistent mediators.

## HOMA-IR

We apply Hitman to 12mo HOMA-IR change using 3mo analyte change as mediators.

## dins030

We apply Hitman to dins030 [change in insulin from 0 to 30 min in mixed-meal tolerance test (mcU/ml)] change at 12mo from baseline using 3mo analyte change from baseline as mediators.

# Pathways

For pathway input data, differences from baseline are combined for people who have both somascan & metabolomics. Analytes are subset and summarized (by averaging over analytes with the same ID) to match network, for network plotting.

We plot the top pathways. The t-statistics from ezlimma have sufficiently many degrees of freedom to be approximately z-scores, which are better known, so we annotate them as z-scores. For a network plotting of differences, we use Pathway Commons PC9 and SMPDB files downloaded around 2016.

We map ChEBI IDs in PC9 to chemical names, so that we can show the names in our network plots.

## Roast between-group

We write the z-score for the between-arm comparison of each analyte’s 3 month difference from baseline from ezlimma::limma\_contrasts to use in results.

Run roast.

Roast barplot.

Plot Roast pwys.

# Pathway mediation

We apply CAMERA to Hitman scores to test pathways. We need to know inter-gene correlation for method cameraPR.

## HbA1c

We find no significant pathways.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | NGenes | EMY.chisq.Direction | EMY.chisq.p | EMY.chisq.FDR |
| Beta-Alanine Metabolism | 10 | 1 | 0.005 | 0.279 |
| Histidine Metabolism | 10 | 1 | 0.008 | 0.279 |
| Glycine and Serine Metabolism | 22 | 1 | 0.042 | 0.849 |

## HOMA-IR

We find 15 significant pathways.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | NGenes | EMY.chisq.Direction | EMY.chisq.p | EMY.chisq.FDR |
| Fatty Acid Biosynthesis | 9 | 1 | 0 | 0 |
| Valine, Leucine and Isoleucine Degradation | 15 | 1 | 0 | 0 |
| Phenylalanine and Tyrosine Metabolism | 10 | 1 | 0 | 0 |

## dins030

We find caffeine metabolism significant.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | NGenes | EMY.chisq.Direction | EMY.chisq.p | EMY.chisq.FDR |
| Caffeine Metabolism | 9 | 1 | 0.000 | 0.000 |
| Sphingolipid Metabolism | 10 | 1 | 0.051 | 0.988 |
| Phospholipid Biosynthesis | 4 | 1 | 0.090 | 0.988 |

# Session info

## - Session info ---------------------------------------------------------------  
## setting value   
## version R version 4.0.2 (2020-06-22)  
## os Windows 7 x64 SP 1   
## system x86\_64, mingw32   
## ui RTerm   
## language (EN)   
## collate English\_United States.1252   
## ctype English\_United States.1252   
## tz America/New\_York   
## date 2020-10-02   
##   
## - Packages -------------------------------------------------------------------  
## package \* version date lib source   
## assertthat 0.2.1 2019-03-21 [1] CRAN (R 4.0.2)   
## backports 1.1.7 2020-05-13 [1] CRAN (R 4.0.0)   
## callr 3.4.3 2020-03-28 [1] CRAN (R 4.0.2)   
## cli 2.0.2 2020-02-28 [1] CRAN (R 4.0.2)   
## colorspace 1.4-1 2019-03-18 [1] CRAN (R 4.0.2)   
## crayon 1.3.4 2017-09-16 [1] CRAN (R 4.0.2)   
## desc 1.2.0 2018-05-01 [1] CRAN (R 4.0.2)   
## devtools 2.3.0 2020-04-10 [1] CRAN (R 4.0.2)   
## digest 0.6.25 2020-02-23 [1] CRAN (R 4.0.2)   
## dplyr \* 1.0.0 2020-05-29 [1] CRAN (R 4.0.2)   
## ellipsis 0.3.1 2020-05-15 [1] CRAN (R 4.0.2)   
## evaluate 0.14 2019-05-28 [1] CRAN (R 4.0.2)   
## ezlimma \* 0.2.3.9036 2020-08-29 [1] local   
## ezlimmaplot \* 0.0.1.9023 2020-07-10 [1] Github (jdreyf/ezlimmaplot@ee8b01c)  
## fansi 0.4.1 2020-01-08 [1] CRAN (R 4.0.2)   
## farver 2.0.3 2020-01-16 [1] CRAN (R 4.0.2)   
## fs 1.4.1 2020-04-04 [1] CRAN (R 4.0.2)   
## generics 0.0.2 2018-11-29 [1] CRAN (R 4.0.2)   
## ggplot2 \* 3.3.2 2020-06-19 [1] CRAN (R 4.0.2)   
## glue 1.4.1 2020-05-13 [1] CRAN (R 4.0.2)   
## gtable 0.3.0 2019-03-25 [1] CRAN (R 4.0.2)   
## highr 0.8 2019-03-20 [1] CRAN (R 4.0.2)   
## Hitman \* 0.0.1.9001 2020-08-30 [1] local   
## htmltools 0.5.0 2020-06-16 [1] CRAN (R 4.0.2)   
## igraph 1.2.5 2020-03-19 [1] CRAN (R 4.0.2)   
## knitr \* 1.29 2020-06-23 [1] CRAN (R 4.0.2)   
## labeling 0.3 2014-08-23 [1] CRAN (R 4.0.0)   
## lattice 0.20-41 2020-04-02 [1] CRAN (R 4.0.2)   
## lifecycle 0.2.0 2020-03-06 [1] CRAN (R 4.0.2)   
## limma \* 3.44.3 2020-06-12 [1] Bioconductor   
## magrittr 1.5 2014-11-22 [1] CRAN (R 4.0.2)   
## Matrix \* 1.2-18 2019-11-27 [1] CRAN (R 4.0.2)   
## memoise 1.1.0 2017-04-21 [1] CRAN (R 4.0.2)   
## mgcv 1.8-31 2019-11-09 [1] CRAN (R 4.0.2)   
## munsell 0.5.0 2018-06-12 [1] CRAN (R 4.0.2)   
## nlme 3.1-148 2020-05-24 [1] CRAN (R 4.0.2)   
## pheatmap 1.0.12 2019-01-04 [1] CRAN (R 4.0.2)   
## pillar 1.4.4 2020-05-05 [1] CRAN (R 4.0.2)   
## pkgbuild 1.0.8 2020-05-07 [1] CRAN (R 4.0.2)   
## pkgconfig 2.0.3 2019-09-22 [1] CRAN (R 4.0.2)   
## pkgload 1.1.0 2020-05-29 [1] CRAN (R 4.0.2)   
## prettyunits 1.1.1 2020-01-24 [1] CRAN (R 4.0.2)   
## processx 3.4.2 2020-02-09 [1] CRAN (R 4.0.2)   
## ps 1.3.3 2020-05-08 [1] CRAN (R 4.0.2)   
## purrr 0.3.4 2020-04-17 [1] CRAN (R 4.0.2)   
## R6 2.4.1 2019-11-12 [1] CRAN (R 4.0.2)   
## RColorBrewer 1.1-2 2014-12-07 [1] CRAN (R 4.0.0)   
## remotes 2.1.1 2020-02-15 [1] CRAN (R 4.0.2)   
## rlang 0.4.6 2020-05-02 [1] CRAN (R 4.0.2)   
## rmarkdown 2.3 2020-06-18 [1] CRAN (R 4.0.2)   
## rprojroot 1.3-2 2018-01-03 [1] CRAN (R 4.0.2)   
## scales 1.1.1 2020-05-11 [1] CRAN (R 4.0.2)   
## sessioninfo 1.1.1 2018-11-05 [1] CRAN (R 4.0.2)   
## stringi 1.4.6 2020-02-17 [1] CRAN (R 4.0.0)   
## stringr 1.4.0 2019-02-10 [1] CRAN (R 4.0.2)   
## testthat 2.3.2 2020-03-02 [1] CRAN (R 4.0.2)   
## tibble 3.0.1 2020-04-20 [1] CRAN (R 4.0.2)   
## tidyselect 1.1.0 2020-05-11 [1] CRAN (R 4.0.2)   
## usethis \* 1.6.1 2020-04-29 [1] CRAN (R 4.0.2)   
## vctrs 0.3.1 2020-06-05 [1] CRAN (R 4.0.2)   
## withr 2.2.0 2020-04-20 [1] CRAN (R 4.0.2)   
## writexl \* 1.3 2020-05-05 [1] CRAN (R 4.0.2)   
## xfun 0.17 2020-09-09 [1] CRAN (R 4.0.2)   
## yaml 2.2.1 2020-02-01 [1] CRAN (R 4.0.0)   
## zeallot \* 0.1.0 2018-01-28 [1] CRAN (R 4.0.2)   
##   
## [1] C:/Program Files/R/R-4.0.2/library

Explore missingness