Metabolite changes in blood predict the onset of tuberculosis

Supplementary data

The GC6 Consortium

2018-03-29

# Supplementary tables

**Supplementary Table 1.** Institutional Review Boards that provided ethics approvals for the different GC6-74 sites

|  |  |  |  |
| --- | --- | --- | --- |
| **Institution** | **Country** | **Ethics Review Committee** | **Protocol no.** |
| SUN | South Africa | Stellenbosch University Institutional Review Board | N05/11/187 |
| MRC | The Gambia | Joint Medical Research Council and Gambian Government | SCC.1141vs2 |
| MAK | Uganda | Uganda National Council for Science and Technology | MV 715 |
| MAK | Uganda | University Hospitals Case Medical Centre | 12-95-08 |
| AHRI | Ethiopia | Armauer Hansen Research Institute (AHRI)/All Africa Leprosy, TB and Rehabilitation Training Center (ALERT) | P015/10 |
| UCT | South Africa | University of Cape Town Human Research Ethics Committee HREC | 013/2013 |

#### 

**Supplementary Table 2.** GC6-74 recruitment and follow up dates.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Country** | **Institution** | **First date of enrolment** | **Last date of enrolment** | **Last date of follow-up** | **Number of Index cases** | **Number of Household Contacts** |
| South Africa | SUN | February 27th, 2006 | December 14th, 2010 | November 28th, 2012 | 209 | 1,197 |
| The Gambia | MRC | March 5th, 2007 | October 21st, 2010 | October 22nd, 2012 | 402 | 1,948 |
| Uganda | MAK | June 1st, 2006 | June 8th, 2010 | May 4th, 2012 | 181 | 499 |
| Ethiopia | AHRI | February 12th, 2007 | August 3rd, 2011 | August 16th, 2012 | 154 | 818 |

#### 

**Supplementary Table 3.** Criteria for tuberculosis diagnosis in GC6-74 progressors.

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Category | Culture 1 | Culture 2† | AFB 1 | AFB 2† | CXR | Symptoms | TB Treatment Response | Failed AB | New Class |
| A | + | + |  |  |  |  |  |  | Definite |
| B | + |  |  |  | + |  |  |  | Definite |
| C | + |  | + |  |  |  |  |  | Definite |
| D |  |  | + | + | + |  |  |  | Probable |
| E |  |  | + | + |  | + |  |  | Probable |
| F |  |  | + |  | + |  |  |  | Probable |
| G | + |  |  |  |  | + | + |  | Probable |
| H | + |  |  |  |  | + |  |  | Possible |
| I |  |  | + | + |  |  | + |  | Possible |
| J |  |  | + |  |  | + |  |  | Possible |
| K |  |  |  |  | + | + | + | + | Possible |
| L | + |  |  |  |  |  |  |  | Questionable |
| M |  |  | + | + |  |  |  |  | Questionable |
| N |  |  | + |  |  |  |  |  | Questionable |
| O |  |  |  |  | + | + |  |  | Questionable |
| P |  |  |  |  | + |  |  |  | Questionable |
| Q |  |  |  |  |  | + |  |  | Questionable |
| R | Neg/ND | Neg/ND | Neg/ND | Neg/ND | Neg/ND | + | Rx not started |  | Non-TB case |
| S | Neg/ND | Neg/ND | Neg/ND | Neg/ND | Neg/ND | Neg/ND | Rx not started |  | Non-TB case |

Culture

Either liquid or solid agar positive with confirmed speciation for M.tuberculosis complex

AFB

Acid-fast bacilli: sputum smear ≥scanty (≥1-9 acid-fast organisms per 100x oil field)

CXR

Chest X-ray compatible with active TB

Failed AB

Failed antibiotics treatment; no response of symptoms to a 7-day, broad spectrum oral or IV antibiotics

†

Positivity on a separate sample of culture or AFB (not an aliquot of the same sample) provided by the participant preferrably but not necessarily on separate days

Note: Questionable classification is excluded in both progressors and non-progressors (controls)

#### 

**Supplementary Table 4.** Sample types by cohort.

|  |  |  |  |
| --- | --- | --- | --- |
|  | Plasma | Plasma/RPMI | Serum |
| **AHRI** | 79 (P=20, C=59) |  |  |
| **MAK** |  | 45 (P=10, C=35) | 40 (P=9, C=31) |
| **MRC** | 251 (P=61, C=190) |  |  |
| **SUN** | 36 (P=11, C=25) |  | 300 (P=70, C=230) |

#### 

**Supplementary Table 5.** Performance of the different machine learning models in internal k-fold cross-validation within the training data sets. The models were trained and evaluated on the training data set. *Site*, the name of the site from which the samples were used; *Timepoints*, the selected timepoints, either all samples, or samples close (proximate) to the diagnoses, or samples further away (distal) from the diagnosis (see column 2). *AUC*, Area Under Curve; CI, 95% confidence interval; *p-value*, p-value associated with the AUC; *q-value*, p-value adjusted with the Benjamini-Hochberg procedure.

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Site | Timepoints | AUC | CI | p-value | q-value | NA | NA | NA | NA | NA | NA |
| **SUN** | **all** | **0.63** | **0.55 – 0.71** | **0.0032** | **0.029** | **0.31** | **0.25** | **0.38** | **0.97** | **0.87** | **1** |
|  | **distal** | **0.65** | **0.57 – 0.72** | **0.0013** | **0.013** | **0.34** | **0.25** | **0.44** | **0.85** | **0.78** | **0.91** |
|  | proximate | 0.76 | 0.55 – 0.98 | 0.018 | 0.12 | 0.17 | 0.056 | 0.35 | 0.99 | 0.95 | 1 |
| **MRC** | **all** | **0.71** | **0.61 – 0.81** | **5.5e-05** | **0.00066** | **0.68** | **0.49** | **0.83** | **0.85** | **0.78** | **0.9** |
|  | distal | 0.61 | 0.49 – 0.73 | 0.061 | 0.37 | 0.32 | 0.18 | 0.5 | 0.86 | 0.78 | 0.91 |
|  | **proximate** | **0.83** | **0.70 – 0.95** | **0.00012** | **0.0013** | **0.36** | **0.18** | **0.57** | **0.97** | **0.91** | **0.99** |
| AHRI | all | 0.57 | 0.38 – 0.76 | 0.46 | 1 | 0.39 | 0.17 | 0.64 | 0.82 | 0.65 | 0.93 |
|  | distal | 0.57 | 0.34 – 0.80 | 0.53 | 1 | 0.33 | 0.12 | 0.62 | 0.88 | 0.71 | 0.96 |
|  | proximate | 0.66 | 0.27 – 1.00 | 0.31 | 1 | 0.4 | 0.053 | 0.85 | 0.95 | 0.82 | 0.99 |
| MAK | all | 0.56 | 0.40 – 0.73 | 0.51 | 1 | 0.32 | 0.17 | 0.51 | 0.91 | 0.71 | 0.99 |
|  | distal | 0.55 | 0.39 – 0.71 | 0.61 | 1 | 0.3 | 0.16 | 0.49 | 0.95 | 0.74 | 1 |
| **TOT** | **all** | **0.64** | **0.59 – 0.70** | **1.6e-06** | **2.2e-05** | **0.45** | **0.35** | **0.56** | **0.79** | **0.75** | **0.83** |
|  | distal | 0.59 | 0.53 – 0.65 | 0.0063 | 0.05 | 0.26 | 0.21 | 0.31 | 0.86 | 0.8 | 0.9 |
|  | **proximate** | **0.77** | **0.67 – 0.87** | **6.8e-06** | **8.8e-05** | **0.27** | **0.15** | **0.42** | **0.97** | **0.94** | **0.98** |

#### 

**Supplementary Table 6.** Performance of the machine learning tested cross-wise between cohorts. The models were trained and evaluated only on the training data set. *Train*, the name of the site from which the samples were used for testing; *Test*, name of the site on which the given model was tested. *AUC*, Area Under Curve; CI, 95% confidence interval; *p-value*, p-value associated with the AUC; *q-value*, p-value adjusted with the Benjamini-Hochberg procedure.

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Train | Test | AUC | CI | p-value | q-value | NA | NA | NA | NA | NA | NA |
| **SUN** | **MRC** | **0.7** | **0.61 – 0.79** | **0.00012** | **0.0014** | **0.46** | **0.31** | **0.61** | **0.83** | **0.76** | **0.9** |
|  | AHRI | 0.57 | 0.40 – 0.75 | 0.44 | 1 | 0.12 | 0.016 | 0.38 | 0.69 | 0.51 | 0.83 |
|  | MAK | 0.6 | 0.42 – 0.77 | 0.32 | 1 | 0.29 | 0.15 | 0.46 | 0.93 | 0.68 | 1 |
| **MRC** | **SUN** | **0.64** | **0.56 – 0.73** | **0.00084** | **0.0093** | **0.4** | **0.29** | **0.52** | **0.81** | **0.73** | **0.86** |
|  | AHRI | 0.56 | 0.38 – 0.75 | 0.51 | 1 | 0.18 | 0.061 | 0.37 | 0.65 | 0.43 | 0.84 |
|  | MAK | 0.55 | 0.35 – 0.75 | 0.62 | 1 | 0.4 | 0.12 | 0.74 | 0.81 | 0.67 | 0.92 |
| AHRI | SUN | 0.52 | 0.43 – 0.60 | 0.69 | 1 | 0.24 | 0.17 | 0.31 | 0.68 | 0.55 | 0.79 |
|  | MRC | 0.56 | 0.46 – 0.65 | 0.26 | 1 | 0.31 | 0.22 | 0.41 | 0.83 | 0.73 | 0.91 |
|  | MAK | 0.66 | 0.50 – 0.82 | 0.097 | 0.88 | 0.35 | 0.19 | 0.55 | 0.95 | 0.77 | 1 |
| MAK | SUN | 0.6 | 0.52 – 0.69 | 0.019 | 0.19 | 0.37 | 0.27 | 0.47 | 0.82 | 0.74 | 0.88 |
|  | MRC | 0.53 | 0.43 – 0.64 | 0.5 | 1 | 0.35 | 0.2 | 0.53 | 0.78 | 0.7 | 0.85 |
|  | AHRI | 0.61 | 0.43 – 0.79 | 0.25 | 1 | 0.34 | 0.19 | 0.53 | 0.89 | 0.67 | 0.99 |

#### 

**Supplementary Table 7.** Performance of the different machine learning models on the blinded data sets. All ML models were trained on the training data set. *Model*, the name of the model applied either to all samples, or samples close (proximate) to the diagnoses, or samples further away (distal) from the diagnosis (see column 2). *AUC*, Area Under Curve; CI, 95% confidence interval; *p-value*, p-value associated with the AUC; *q-value*, p-value adjusted with the Benjamini-Hochberg procedure.

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Model | Timepoints | AUC | CI | p-value | q-value | NA | NA | NA | NA | NA | NA |
| Distal serum | all | 0.61 | 0.52 – 0.71 | 0.011 | 0.2 | 0.32 | 0.23 | 0.41 | 0.86 | 0.79 | 0.92 |
|  | proximate | 0.68 | 0.53 – 0.83 | 0.029 | 0.41 | 0.35 | 0.2 | 0.53 | 0.91 | 0.76 | 0.98 |
|  | distal | 0.59 | 0.48 – 0.70 | 0.1 | 0.96 | 0.3 | 0.2 | 0.42 | 0.85 | 0.76 | 0.91 |
|  | BL | 0.56 | 0.45 – 0.68 | 0.27 | 1 | 0.3 | 0.2 | 0.43 | 0.85 | 0.76 | 0.92 |
| BL serum | all | 0.59 | 0.50 – 0.68 | 0.045 | 0.54 | 0.36 | 0.23 | 0.5 | 0.82 | 0.76 | 0.88 |
|  | proximate | 0.58 | 0.41 – 0.75 | 0.32 | 1 | 0.3 | 0.17 | 0.46 | 0.89 | 0.72 | 0.98 |
|  | distal | 0.59 | 0.48 – 0.70 | 0.096 | 0.96 | 0.4 | 0.24 | 0.58 | 0.83 | 0.76 | 0.89 |
|  | BL | 0.58 | 0.46 – 0.69 | 0.19 | 1 | 0.35 | 0.2 | 0.54 | 0.82 | 0.75 | 0.89 |
| Distal plasma | all | 0.6 | 0.51 – 0.70 | 0.021 | 0.34 | 0.36 | 0.24 | 0.5 | 0.83 | 0.77 | 0.88 |
|  | proximate | 0.69 | 0.53 – 0.85 | 0.022 | 0.34 | 0.57 | 0.29 | 0.82 | 0.86 | 0.74 | 0.94 |
|  | distal | 0.57 | 0.46 – 0.68 | 0.22 | 1 | 0.42 | 0.22 | 0.63 | 0.82 | 0.75 | 0.88 |
|  | BL | 0.54 | 0.42 – 0.65 | 0.52 | 1 | 0.25 | 0.16 | 0.36 | 0.83 | 0.72 | 0.91 |
| BL plasma | all | 0.64 | 0.55 – 0.72 | 0.0026 | 0.054 | 0.45 | 0.3 | 0.6 | 0.84 | 0.78 | 0.89 |
|  | proximate | 0.45 | 0.27 – 0.62 | 0.51 | 1 | 0.2 | 0.091 | 0.36 | 0.74 | 0.55 | 0.88 |
|  | **distal** | **0.68** | **0.58 – 0.77** | **0.0011** | **0.023** | **0.31** | **0.22** | **0.42** | **0.88** | **0.79** | **0.94** |
|  | BL | 0.65 | 0.54 – 0.75 | 0.011 | 0.2 | 0.41 | 0.25 | 0.59 | 0.84 | 0.76 | 0.9 |
| **TotalF** | **all** | **0.71** | **0.62 – 0.79** | **4.2e-06** | **0.00011** | **0.81** | **0.61** | **0.93** | **0.85** | **0.8** | **0.9** |
|  | **proximate** | **0.78** | **0.62 – 0.94** | **0.00071** | **0.017** | **1** | **0.66** | **1** | **0.89** | **0.78** | **0.95** |
|  | **distal** | **0.68** | **0.58 – 0.79** | **0.00071** | **0.017** | **0.8** | **0.52** | **0.96** | **0.84** | **0.78** | **0.89** |
|  | BL | 0.64 | 0.52 – 0.76 | 0.017 | 0.28 | 0.71 | 0.42 | 0.92 | 0.84 | 0.76 | 0.89 |
| **Total BL** | **all** | **0.73** | **0.65 – 0.81** | **2.7e-07** | **7.6e-06** | **0.59** | **0.42** | **0.74** | **0.85** | **0.8** | **0.9** |
|  | proximate | 0.75 | 0.59 – 0.91 | 0.0026 | 0.054 | 0.64 | 0.35 | 0.87 | 0.88 | 0.76 | 0.95 |
|  | **distal** | **0.72** | **0.63 – 0.82** | **3.2e-05** | **0.00084** | **0.33** | **0.23** | **0.44** | **0.89** | **0.81** | **0.95** |
|  | **BL** | **0.71** | **0.61 – 0.81** | **0.00021** | **0.0052** | **0.34** | **0.23** | **0.47** | **0.89** | **0.81** | **0.95** |
| GGC | all | 0.59 | 0.50 – 0.67 | 0.041 | 0.54 | 0.27 | 0.2 | 0.36 | 0.86 | 0.78 | 0.92 |
|  | proximate | 0.59 | 0.44 – 0.74 | 0.28 | 1 | 0.31 | 0.17 | 0.48 | 0.88 | 0.71 | 0.96 |
|  | distal | 0.59 | 0.49 – 0.70 | 0.078 | 0.86 | 0.32 | 0.19 | 0.47 | 0.83 | 0.75 | 0.89 |
|  | BL | 0.53 | 0.43 – 0.64 | 0.54 | 1 | 0.26 | 0.17 | 0.37 | 0.84 | 0.74 | 0.92 |

#### 

**Supplementary Table 8.** Performance of the machine learning models "Total" and "Total.BL" on the blinded data sets, stratified by site. All ML models were trained on the training data set. *Model*, the name of the model applied all samples from one of the four cohorts; *Site*, the subset of samples on which the model was evaluated. *AUC*, Area Under Curve; CI, 95% confidence interval; *p-value*, p-value associated with the AUC; *q-value*, p-value adjusted with the Benjamini-Hochberg procedure.

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Model | Site | AUC | CI | p-value | q-value | NA | NA | NA | NA | NA | NA |
| **TotalF** | **SUN** | **0.8** | **0.69 – 0.91** | **4e-05** | **0.00032** | **0.41** | **0.25** | **0.58** | **0.92** | **0.83** | **0.97** |
|  | MRC | 0.68 | 0.51 – 0.84 | 0.02 | 0.099 | 1 | 0.63 | 1 | 0.85 | 0.75 | 0.92 |
|  | AHRI | 0.59 | 0.28 – 0.91 | 0.5 | 0.63 | 1 | 0.29 | 1 | 0.84 | 0.64 | 0.95 |
|  | MAK | 0.63 | 0.33 – 0.93 | 0.32 | 0.63 | 1 | 0.29 | 1 | 0.86 | 0.68 | 0.96 |
| **Total BL** | **SUN** | **0.75** | **0.62 – 0.88** | **0.00053** | **0.0037** | **0.75** | **0.43** | **0.95** | **0.88** | **0.79** | **0.94** |
|  | **MRC** | **0.73** | **0.59 – 0.87** | **0.0024** | **0.014** | **0.73** | **0.39** | **0.94** | **0.85** | **0.74** | **0.92** |
|  | AHRI | 0.67 | 0.39 – 0.95 | 0.2 | 0.61 | 0.5 | 0.16 | 0.84 | 0.85 | 0.62 | 0.97 |
|  | MAK | 0.72 | 0.50 – 0.94 | 0.087 | 0.35 | 0.43 | 0.18 | 0.71 | 0.94 | 0.73 | 1 |

#### 

**Supplementary Table 9.** Performance of the TB-HEALTHY model when applied to GC6 data. AUC, area under curve; in parentheses, 95% confidence intervals; p-value for the AUC; q-value: p-value corrected using the Benjamini-Hochberg method; PPV, positive predictive value; NPV, negative predictive value.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Site | Model | AUC | p-value | q-value | PPV | NPV |
| **SUN** | **all** | **0.71 (0.64-0.78)** | **2e-08** | **2.5e-07** | **0.42 (0.33-0.51)** | **0.85 (0.80-0.90)** |
|  | **distal** | **0.71 (0.64-0.78)** | **5.9e-08** | **6.4e-07** | **0.49 (0.37-0.62)** | **0.86 (0.81-0.90)** |
|  | **proximate** | **0.81 (0.69-0.93)** | **0.00098** | **0.0059** | **0.09 (0.04-0.16)** | **0.99 (0.97-1.00)** |
| **MRC** | **all** | **0.74 (0.66-0.81)** | **3e-08** | **3.6e-07** | **0.58 (0.44-0.72)** | **0.85 (0.79-0.90)** |
|  | **distal** | **0.68 (0.58-0.77)** | **3e-04** | **0.0024** | **0.47 (0.30-0.65)** | **0.87 (0.81-0.91)** |
|  | **proximate** | **0.86 (0.75-0.96)** | **3.1e-07** | **2.8e-06** | **0.42 (0.26-0.59)** | **0.98 (0.94-0.99)** |
| AHRI | all | 0.55 (0.39-0.72) | 0.48 | 1 | 1.00 (0.54-1.00) | 0.81 (0.70-0.89) |
|  | distal | 0.45 (0.25-0.64) | 0.54 | 1 | 0.50 (0.07-0.93) | 0.84 (0.73-0.92) |
|  | **proximate** | **0.89 (0.75-1.00)** | **0.00077** | **0.0054** | **1.00 (0.48-1.00)** | **0.97 (0.89-1.00)** |
| MAK | all | 0.57 (0.43-0.72) | 0.33 | 1 | 0.29 (0.17-0.44) | 0.86 (0.71-0.95) |
|  | distal | 0.50 (0.34-0.67) | 0.96 | 1 | 0.21 (0.11-0.34) | 0.91 (0.72-0.99) |
|  | proximate | 0.75 (0.54-0.97) | 0.061 | 0.3 | 0.14 (0.04-0.33) | 0.98 (0.88-1.00) |
| **TOT** | **all** | **0.68 (0.64-0.73)** | **1.1e-13** | **1.6e-12** | **0.35 (0.30-0.40)** | **0.86 (0.82-0.90)** |
|  | **distal** | **0.65 (0.59-0.70)** | **8e-08** | **8e-07** | **0.30 (0.25-0.37)** | **0.86 (0.82-0.89)** |
|  | **proximate** | **0.82 (0.75-0.89)** | **7.7e-12** | **1.1e-10** | **0.24 (0.16-0.33)** | **0.97 (0.95-0.98)** |

#### 

**Supplementary Table 10.** Performance of the TB-ORD model when applied to GC6 data. AUC, area under curve; CI, 95% confidence interval for the AUC; p-value for the AUC; q-value: p-value corrected using the Benjamini-Hochberg method.

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Site | Model | AUC | CI | p-value | q-value | NA | NA | NA | NA | NA | NA |
| **SUN** | **all** | **0.67** | **0.60 – 0.74** | **4.2e-06** | **5.5e-05** | **0.33** | **0.26** | **0.4** | **0.86** | **0.79** | **0.91** |
|  | **distal** | **0.64** | **0.57 – 0.71** | **0.00031** | **0.0036** | **0.27** | **0.21** | **0.33** | **0.9** | **0.82** | **0.95** |
|  | **proximate** | **0.76** | **0.57 – 0.94** | **0.0063** | **0.044** | **0.38** | **0.14** | **0.68** | **0.98** | **0.95** | **0.99** |
| **MRC** | **all** | **0.63** | **0.55 – 0.72** | **0.002** | **0.018** | **0.44** | **0.3** | **0.59** | **0.81** | **0.74** | **0.86** |
|  | distal | 0.57 | 0.47 – 0.66 | 0.18 | 0.54 | 0.21 | 0.15 | 0.29 | 0.88 | 0.79 | 0.94 |
|  | **proximate** | **0.73** | **0.59 – 0.88** | **0.00075** | **0.0075** | **0.19** | **0.11** | **0.3** | **0.96** | **0.92** | **0.99** |
| AHRI | all | 0.56 | 0.38 – 0.74 | 0.42 | 0.85 | 0.5 | 0.25 | 0.75 | 0.81 | 0.69 | 0.9 |
|  | distal | 0.64 | 0.44 – 0.83 | 0.13 | 0.52 | 0.13 | 0.058 | 0.24 | 0.55 | 0.23 | 0.83 |
|  | **proximate** | **0.92** | **0.79 – 1.00** | **3e-04** | **0.0036** | **0.75** | **0.35** | **0.97** | **0.98** | **0.91** | **1** |
| MAK | all | 0.62 | 0.48 – 0.77 | 0.1 | 0.5 | 0.67 | 0.3 | 0.93 | 0.83 | 0.73 | 0.91 |
|  | distal | 0.48 | 0.33 – 0.64 | 0.85 | 0.85 | 0.14 | 0.053 | 0.28 | 0.78 | 0.62 | 0.9 |
|  | **proximate** | **0.89** | **0.74 – 1.00** | **0.0042** | **0.033** | **0.57** | **0.18** | **0.9** | **0.98** | **0.92** | **1** |
| **TOT** | **all** | **0.63** | **0.58 – 0.67** | **2.9e-07** | **4e-06** | **0.68** | **0.55** | **0.8** | **0.8** | **0.76** | **0.82** |
|  | **distal** | **0.57** | **0.52 – 0.63** | **0.0069** | **0.044** | **0.29** | **0.22** | **0.38** | **0.83** | **0.79** | **0.86** |
|  | **proximate** | **0.83** | **0.75 – 0.91** | **1.1e-12** | **1.7e-11** | **0.21** | **0.15** | **0.29** | **0.98** | **0.96** | **0.99** |

#### 

**Supplementary Table 12.** Significant differences in metabolic profiles between cases and compounds for the last collection time point (model 1).

|  |  |  |
| --- | --- | --- |
|  | Compound | q value |
| **M.59** | histidine | 0.00019 |
| **M.46695** | X - 18913 | 0.0021 |
| **M.27447** | 1-linoleoylglycerol (1-monolinolein) | 0.0021 |
| **M.47640** | androsterone glucuronide | 0.0023 |
| **M.46500** | 3-CMPFP\*\* | 0.0023 |
| **M.3155** | 3-ureidopropionate | 0.0029 |
| **M.54** | tryptophan | 0.0029 |
| **M.46458** | N-acetyl-2-aminoctanoate | 0.0049 |
| **M.53** | glutamine | 0.0049 |
| **M.1302** | methionine | 0.0049 |
| **M.42489** | 2-hydroxydecanoate | 0.0049 |
| **M.32675** | C-glycosyltryptophan\* | 0.0076 |
| **M.46908** | 6-bromotryptophan | 0.0076 |
| **M.46670** | 3,4-methylene heptanoyl carnitine | 0.0082 |
| **M.35136** | 5-methyluridine (ribothymidine) | 0.0093 |
| **M.46598** | X - 11438 | 0.0093 |
| **M.1712** | cortisol | 0.0093 |
| **M.43802** | guanidinoacetate | 0.01 |
| **M.32198** | acetylcarnitine | 0.011 |
| **M.46689** | X - 17145 | 0.011 |
| **M.1126** | alanine | 0.016 |
| **M.46608** | X - 11880 | 0.016 |
| **M.44872** | gamma-glutamylmethionine | 0.022 |
| **M.46997** | X - 12822 | 0.025 |
| **M.1564** | citrate | 0.025 |
| **M.15336** | tartarate | 0.025 |
| **M.43488** | N-acetylcarnosine | 0.025 |
| **M.46366** | X - 18249 | 0.025 |
| **M.37253** | 2-hydroxyglutarate | 0.031 |
| **M.43343** | 2-aminooctanoate | 0.031 |
| **M.46928** | valine ion | 0.031 |
| **M.41377** | phenylalanyltryptophan | 0.031 |
| **M.44876** | gamma-CEHC | 0.033 |
| **M.1573** | guanosine | 0.033 |
| **M.32599** | glycocholenate sulfate\* | 0.033 |
| **M.1589** | N-acetylmethionine | 0.033 |
| **M.31787** | 3-carboxy-4-methyl-5-propyl-2-furanpropanoate (CMPF) | 0.035 |
| **M.22036** | 2-hydroxyoctanoate | 0.035 |
| **M.1649** | valine | 0.04 |
| **M.42374** | 2-aminobutyrate | 0.04 |
| **M.46638** | glycodeoxycholate sulfate | 0.041 |
| **M.33935** | piperine | 0.041 |
| **M.46681** | X - 16935 | 0.042 |
| **M.1284** | threonine | 0.042 |
| **M.47648** | X - 17138 | 0.042 |
| **M.46738** | 3-formylindole | 0.042 |
| **M.33950** | N-acetylphenylalanine | 0.049 |
| **M.42582** | pyruvate | 0.05 |
| **M.37063** | gamma-glutamylalanine | 0.05 |

#### 

**Supplementary Table 13.** Significant differences in metabolic profiles between cases and compounds for all collection time points (model 2).

|  |  |  |
| --- | --- | --- |
|  | Compound | q value |
| **M.47391** | 2'-O-methyluridine | 3.9e-07 |
| **M.1712** | cortisol | 6.1e-06 |
| **M.1126** | alanine | 9.9e-05 |
| **M.584** | mannose | 9.9e-05 |
| **M.46608** | X - 11880 | 9.9e-05 |
| **M.553** | cotinine | 0.00018 |
| **M.43488** | N-acetylcarnosine | 0.00022 |
| **M.46515** | X - 21470 | 3e-04 |
| **M.12129** | beta-hydroxyisovalerate | 0.00051 |
| **M.46681** | X - 16935 | 0.0018 |
| **M.46626** | X - 12739 | 0.0022 |
| **M.46294** | X - 21285 | 0.0024 |
| **M.59** | histidine | 0.0024 |
| **M.47715** | argininate | 0.0028 |
| **M.33950** | N-acetylphenylalanine | 0.0031 |
| **M.53** | glutamine | 0.0048 |
| **M.46595** | X - 11378 | 0.0048 |
| **M.47709** | 2-oxoarginine\* | 0.0079 |
| **M.46366** | X - 18249 | 0.0086 |
| **M.37202** | 4-androsten-3beta,17beta-diol disulfate (1) | 0.0086 |
| **M.37063** | gamma-glutamylalanine | 0.0099 |
| **M.542** | 3-hydroxybutyrate (BHBA) | 0.011 |
| **M.31453** | cysteine | 0.012 |
| **M.46368** | X - 18914 | 0.012 |
| **M.54** | tryptophan | 0.012 |
| **M.1769** | cortisone | 0.013 |
| **M.32599** | glycocholenate sulfate\* | 0.014 |
| **M.18349** | indolelactate | 0.014 |
| **M.1564** | citrate | 0.014 |
| **M.46652** | X - 14658 | 0.016 |
| **M.33935** | piperine | 0.016 |
| **M.20693** | tartronate (hydroxymalonate) | 0.016 |
| **M.46390** | X - 11308 | 0.016 |
| **M.41377** | phenylalanyltryptophan | 0.016 |
| **M.2730** | gamma-glutamylglutamine | 0.016 |
| **M.46622** | thioproline | 0.017 |
| **M.46660** | myristoleoylcarnitine\* | 0.017 |
| **M.32857** | 5-dodecenoylcarnitine (C12:1) | 0.017 |
| **M.37207** | 4-androsten-3alpha,17alpha-diol monosulfate (2) | 0.017 |
| **M.27738** | threonate | 0.018 |
| **M.44630** | 1-eicosatrienoylglycerophosphoethanolamine\* | 0.02 |
| **M.47640** | androsterone glucuronide | 0.02 |
| **M.46471** | hexanoylglutamine | 0.022 |
| **M.46663** | 1-ribosyl-imidazoleacetate\* | 0.023 |
| **M.46928** | valine ion | 0.023 |
| **M.32562** | pregnen-diol disulfate\* | 0.023 |
| **M.32786** | hydroxy-CMPF\* | 0.025 |
| **M.46905** | X - 21736 | 0.031 |
| **M.27710** | N-acetylglycine | 0.031 |
| **M.34406** | valerylcarnitine | 0.036 |
| **M.35884** | 2-eicosatrienoylglycerophosphocholine\* | 0.036 |
| **M.46497** | hydroxy-CMPF\* | 0.037 |
| **M.57** | glutamate | 0.037 |
| **M.38661** | hydroxycotinine | 0.037 |
| **M.46751** | 3-hydroxybutyrylcarnitine (2) | 0.037 |
| **M.27447** | 1-linoleoylglycerol (1-monolinolein) | 0.037 |
| **M.37190** | 5alpha-androstan-3beta,17beta-diol disulfate | 0.04 |
| **M.46739** | laurylcarnitine\* | 0.05 |
| **M.33952** | myristoylcarnitine | 0.05 |

#### 

**Supplementary Table 14.** Significant differences in metabolic profiles between cases and compounds for all collection time points for serum metabolites.

|  |  |  |
| --- | --- | --- |
|  | Compound | q value |
| **M.47391** | 2'-O-methyluridine | 2.5e-05 |
| **M.22137** | homoarginine | 0.00027 |
| **M.47964** | 3-hydroxybutyroylglycine\*\* | 0.00027 |
| **M.47642** | X - 12101 | 0.00038 |
| **M.47799** | X - 15477 | 0.00069 |
| **M.1712** | cortisol | 0.0015 |
| **M.27710** | N-acetylglycine | 0.0018 |
| **M.47495** | pyroglutamylisoleucine | 0.0018 |
| **M.46628** | X - 12472 | 0.0018 |
| **M.46906** | X - 21737 | 0.0018 |
| **M.43258** | acisoga | 0.0022 |
| **M.584** | mannose | 0.0022 |
| **M.553** | cotinine | 0.0022 |
| **M.46608** | X - 11880 | 0.0024 |
| **M.46362** | methylnaphthyl sulfate (2)\* | 0.0026 |
| **M.44876** | gamma-CEHC | 0.003 |
| **M.47718** | N-acetyl-isoputrenine\* | 0.003 |
| **M.46515** | X - 21470 | 0.003 |
| **M.38661** | hydroxycotinine | 0.003 |
| **M.32379** | scyllo-inositol | 0.003 |
| **M.43488** | N-acetylcarnosine | 0.0053 |
| **M.37202** | 4-androsten-3beta,17beta-diol disulfate (1) | 0.0055 |
| **M.46681** | X - 16935 | 0.0055 |
| **M.47698** | methylsuccinoylcarnitine (1) | 0.007 |
| **M.33950** | N-acetylphenylalanine | 0.012 |
| **M.46612** | X - 12206 | 0.012 |
| **M.46626** | X - 12739 | 0.012 |
| **M.27738** | threonate | 0.012 |
| **M.47817** | X - 17327 | 0.012 |
| **M.47715** | argininate | 0.013 |
| **M.46390** | X - 11308 | 0.016 |
| **M.47696** | 3-S-cysteinyl-2-methylpropanoate\* | 0.02 |
| **M.33954** | glycylphenylalanine | 0.022 |
| **M.46634** | X - 12846 | 0.022 |
| **M.46624** | X - 12798 | 0.022 |
| **M.46366** | X - 18249 | 0.022 |
| **M.44552** | S-(3-hydroxypropyl)mercapturic acid (HPMA) | 0.022 |
| **M.32562** | pregnen-diol disulfate\* | 0.022 |
| **M.33963** | acetoacetate | 0.028 |
| **M.37506** | palmitoyl sphingomyelin | 0.031 |
| **M.46294** | X - 21285 | 0.032 |
| **M.31453** | cysteine | 0.033 |
| **M.47845** | X - 20674 | 0.033 |
| **M.47709** | 2-oxoarginine\* | 0.033 |
| **M.41374** | phenylalanylalanine | 0.034 |
| **M.32786** | hydroxy-CMPF\* | 0.035 |
| **M.37190** | 5alpha-androstan-3beta,17beta-diol disulfate | 0.036 |
| **M.36103** | p-cresol sulfate | 0.037 |
| **M.46751** | 3-hydroxybutyrylcarnitine (2) | 0.037 |
| **M.46497** | hydroxy-CMPF\* | 0.037 |
| **M.46623** | X - 12729 | 0.038 |
| **M.38662** | cotinine N-oxide | 0.039 |
| **M.1769** | cortisone | 0.039 |
| **M.46368** | X - 18914 | 0.042 |
| **M.1444** | pipecolate | 0.043 |
| **M.47727** | 3-hydroxyhexanoylcarnitine (1) | 0.044 |
| **M.47725** | suberoylcarnitine | 0.044 |
| **M.39603** | ethyl glucuronide | 0.044 |

#### 

**Supplementary Table 15.** Significant differences in metabolic profiles between cases and compounds for all collection time points for plasma metabolites.

|  |  |  |
| --- | --- | --- |
|  | Compound | q value |
| **M.1126** | alanine | 1.3e-05 |
| **M.47656** | sulfate of piperine metabolite C16H19NO3 (3)\* | 0.016 |
| **M.54** | tryptophan | 0.02 |
| **M.46726** | glucuronide of piperine metabolite C17H21NO3 (5)\* | 0.02 |
| **M.32735** | glucuronide of piperine metabolite C17H21NO3 (4)\* | 0.02 |
| **M.35625** | 1-myristoylglycerol (1-monomyristin) | 0.02 |
| **M.46600** | sulfate of piperine metabolite C16H19NO3 (2)\* | 0.02 |
| **M.46728** | glucuronide of piperine metabolite C17H21NO3 (3)\* | 0.025 |
| **M.34393** | 1-linolenoylglycerol | 0.026 |
| **M.1592** | N-acetylneuraminate | 0.026 |
| **M.36803** | 3,7-Dihydroxy-5-cholestenoic acid | 0.028 |
| **M.53** | glutamine | 0.031 |
| **M.33935** | piperine | 0.045 |

#### 

**Supplementary Table 17.** Detailed information on the computing environment used to perform the analyses (sessionInfo() output).

## R version 3.4.3 (2017-11-30)  
## Platform: x86\_64-pc-linux-gnu (64-bit)  
## Running under: Ubuntu 16.04.3 LTS  
##   
## Matrix products: default  
## BLAS: /usr/lib/libblas/libblas.so.3.6.0  
## LAPACK: /usr/lib/lapack/liblapack.so.3.6.0  
##   
## locale:  
## [1] LC\_CTYPE=en\_US.UTF-8 LC\_NUMERIC=C   
## [3] LC\_TIME=en\_US.UTF-8 LC\_COLLATE=en\_US.UTF-8   
## [5] LC\_MONETARY=en\_US.UTF-8 LC\_MESSAGES=en\_US.UTF-8   
## [7] LC\_PAPER=en\_US.UTF-8 LC\_NAME=C   
## [9] LC\_ADDRESS=C LC\_TELEPHONE=C   
## [11] LC\_MEASUREMENT=en\_US.UTF-8 LC\_IDENTIFICATION=C   
##   
## attached base packages:  
## [1] stats graphics grDevices utils datasets methods base   
##   
## other attached packages:  
## [1] tmod\_0.37 sm\_2.2-5.4 segmented\_0.5-3.0   
## [4] randomForest\_4.6-12 plotwidgets\_0.5 pander\_0.6.1   
## [7] OptimalCutpoints\_1.1-3 limma\_3.28.21 gplots\_3.0.1   
## [10] beeswarm\_0.2.3 rmarkdown\_1.6 epiR\_0.9-93   
## [13] survival\_2.41-3 pROC\_1.10.0 pca3d\_0.10   
## [16] myfuncs\_1.8 setwidth\_1.0-4 colorout\_1.1-2   
##   
## loaded via a namespace (and not attached):  
## [1] tagcloud\_0.6 TH.data\_1.0-8 colorspace\_1.3-2   
## [4] class\_7.3-14 rprojroot\_1.2 estimability\_1.2   
## [7] DRR\_0.0.3 prodlim\_1.6.1 mvtnorm\_1.0-7   
## [10] lubridate\_1.7.2 codetools\_0.2-15 splines\_3.4.3   
## [13] mnormt\_1.5-5 robustbase\_0.92-8 knitr\_1.20   
## [16] RcppRoll\_0.2.2 jsonlite\_1.5 caret\_6.0-78   
## [19] broom\_0.4.3 ddalpha\_1.3.1.1 kernlab\_0.9-25   
## [22] sfsmisc\_1.1-1 shiny\_1.0.5 compiler\_3.4.3   
## [25] lsmeans\_2.27-61 backports\_1.1.1 assertthat\_0.2.0   
## [28] Matrix\_1.2-12 lazyeval\_0.2.1 htmltools\_0.3.6   
## [31] tools\_3.4.3 bindrcpp\_0.2 coda\_0.19-1   
## [34] gtable\_0.2.0 glue\_1.2.0 reshape2\_1.4.3   
## [37] dplyr\_0.7.4 Rcpp\_0.12.15 gdata\_2.18.0   
## [40] nlme\_3.1-131.1 iterators\_1.0.9 crosstalk\_1.0.0   
## [43] psych\_1.7.8 timeDate\_3042.101 gower\_0.1.2   
## [46] stringr\_1.2.0 mime\_0.5 gtools\_3.5.0   
## [49] XML\_3.98-1.10 DEoptimR\_1.0-8 MASS\_7.3-48   
## [52] zoo\_1.8-1 scales\_0.5.0 ipred\_0.9-6   
## [55] parallel\_3.4.3 sandwich\_2.4-0 RColorBrewer\_1.1-2  
## [58] yaml\_2.1.16 ggplot2\_2.2.1 rpart\_4.1-12   
## [61] stringi\_1.1.6 foreach\_1.4.4 caTools\_1.17.1   
## [64] lava\_1.6 rlang\_0.2.0 pkgconfig\_2.0.1   
## [67] bitops\_1.0-6 rgl\_0.99.9 evaluate\_0.10.1   
## [70] lattice\_0.20-35 purrr\_0.2.4 bindr\_0.1   
## [73] recipes\_0.1.2 htmlwidgets\_1.0 CVST\_0.2-1   
## [76] tidyselect\_0.2.3 plyr\_1.8.4 magrittr\_1.5   
## [79] R6\_2.2.2 multcomp\_1.4-8 dimRed\_0.1.0   
## [82] pillar\_1.1.0 foreign\_0.8-69 withr\_2.1.1   
## [85] nnet\_7.3-12 tibble\_1.4.2 KernSmooth\_2.23-15  
## [88] ellipse\_0.4.1 grid\_3.4.3 ModelMetrics\_1.1.0  
## [91] digest\_0.6.15 xtable\_1.8-2 tidyr\_0.8.0   
## [94] httpuv\_1.3.5 stats4\_3.4.3 munsell\_0.4.3   
## [97] BiasedUrn\_1.07 tcltk\_3.4.3