Taurine/chenodeoxycholic acid ratio as a circulating biomarker of insidious vitamin B_{12} deficiency in humans

Madhu Baghel[1]*, Sting L. Shi[1]*, Himani Patel[1], Vidya Velagapudi[2], Abdullah Mahmood Ali[3#] and Vijay K. Yadav[1#]

[1]Systems Biology of Aging laboratory, Department of Genetics and Development, Columbia University, NY, USA

[2]Institute for Molecular Medicine Finland FIMM, University of Helsinki; Helsinki, Finland

[3]Department of Medicine, Columbia University Irving Medical Center, New York, NY, USA

*Joint First Authors

[#]Co-corresponding Authors:

Abdulla Mahmood Ali (ama2241@cumc.columbia.edu)

Vijay K. Yadav (vky2101@cumc.columbia.edu)

Conflict of interest: The authors have declared that no conflict of interest exists.

Key words: Vitamin B12; taurine; metabolism; aging; metabolomics; biomarkers

Short Title: Detection of vitamin B_{12} deficiency

ABSTRACT:

Deficiency of vitamin B_{12} (B_{12}), an essential water-soluble vitamin, leads to irreversible neurological damage, osteoporosis, cardiovascular diseases, and anemia. Clinical tests to detect B_{12} deficiency lack specificity and sensitivity. B_{12} deficiency is thus insidious because progressive decline in organ functions may go unnoticed until the damage is advanced or irreversible. Here, using targeted unbiased metabolomic profiling in the sera of B_{12}-deficient versus control individuals, we set out to identify biomarker(s) of B_{12} deficiency. Metabolomic profiling identified 77 metabolites, and Partial least squares discriminant-analysis (PLS-DA) and hierarchical clustering analysis (HCA) showed a differential abundance in B_{12}-deficient sera of taurine, xanthine, hypoxanthine, chenodeoxycholic acid, neopterin, and glycocholic acid. Random forest (RF) multivariate analysis identified a taurine/chenodeoxycholic acid ratio, with an AUC score of 1, to be the best biomarker to predict B_{12} deficiency. Mechanistically, B_{12} deficiency reshaped the transcriptomic and metabolomic landscape of the cell identifying a downregulation of methionine, taurine, urea cycle, and nucleotide metabolism, and an upregulation of Krebs cycle. Thus, we propose taurine/chenodeoxycholic acid ratio in serum as a potential biomarker of B_{12} deficiency in humans and elucidate cellular metabolic pathways regulated by B_{12} deficiency.

INTRODUCTION

Vitamin B_{12} (B_{12}) is an essential water-soluble vitamin derived from animal-based diets that regulates a multitude of cellular processes in humans such as one-carbon metabolism and Krebs cycle.(1-4) The absorption of dietary B_{12} requires gastric intrinsic factor (GIF), a stomach-specific protein.(4) Gif binds to B_{12} in the small intestine forming the GIF-B_{12} complex. This complex is endocytosed by the intestinal epithelial cells and B_{12} is released into the bloodstream.(4) In the bloodstream, B_{12} binds to the protein transcobalamin 2, which then carries it to the liver, the primary storage and recycling organ for B_{12} in mammals.(5) Once acquired, humans, for instance, can recycle B_{12} to maintain B_{12}-dependent cellular processes for up to a decade.(2) In the cells, B_{12}–derivatives function as cofactors for only two known enzymes: methylmalonyl-CoA mutase and methionine synthase, and through them affect a variety of downstream metabolic pathways such as Krebs cycle, amino acid synthesis, and DNA and histone methylation. (1, 6) In humans, decreased production of functional GIF protein or non-consumption of animal products causes B_{12} deficiency and results in various abnormalities, such as anemia, osteoporosis, and cognitive defects. (7-10)

In clinical practice, the diagnosis of B_{12} deficiency is typically established by the measurement of serum cobalamin (Cbl) levels.(11) Although B_{12} deficiency can be reflected by elevated methylmalonic acid (MMA) and homocysteine (Hcy) levels, these tests are not routinely used unless the initial Cbl levels are equivocal because MMA and Hcy can be elevated in conditions independent of B_{12} levels.(12-16) Despite the importance of B_{12} and its association with many physiological functions, many issues remain unresolved in the diagnosis of B_{12} deficiency, leading to poor diagnosis and irreversible consequences on the body.(17, 18) First, B_{12} is a very stable molecule and because 95-97% of B_{12} is stored in the liver, its serum levels do not accurately reflect its actual functional levels i.e., the amount of B_{12} required for maintaining body functions.(19)

Second, the cost of measurement of B_{12} in patient samples, despite being not able to accurately predict a B_{12}-deficient state, remains high and therefore is not the first line of measurement; clinicians measure B_{12} only when a patient presents signs of B_{12} deficiency such as anemia to confirm a deficient state. (20-23) These facts necessitate the need to identify molecules regulated by B_{12}, which can provide a functional readout of B_{12} deficiency in humans.

We recently created a transgenic mouse model of B_{12} deficiency by deleting the gene essential for B_{12} absorption from the gut, *Gif*, to understand the molecular consequences of B_{12} deficiency. These studies led to the identification that B_{12} stored in the liver regulates the production of taurine. Taurine is a semi-essential micronutrient that has recently been shown to be a driver of aging as its supplementation increases healthy lifespan in diverse species from worms to mice, and low taurine levels are associated with poor health in aged humans(24). In the B_{12} mode of action, taurine plays an important role as the reversal of taurine deficiency through daily oral taurine administration was shown to fully rescue the consequences of B_{12} deficiency(25). More importantly, the targeted metabolomics analysis of liver tissue collected from control and B_{12}-deficient mice showed changes in a multitude of metabolites besides taurine that are secreted from cells and could be detected in the serum(25). These studies suggested a plausible and testable hypothesis that certain metabolites or sets of metabolites may exist which could serve as a readout of, difficult to detect, B_{12}-deficient state in humans.

The present study was initiated to test the above hypothesis by performing a metabolomic analysis on serum samples collected from control and B_{12}-deficient individuals to identify which factor(s) could serve as a biomarker of B_{12}-deficient state. Results showed that serum levels of certain metabolites such as taurine, xanthine and hypoxanthine were dramatically downregulated in the B_{12}-deficient individuals. Using various downstream analyses, we suggest that taurine in

conjugation with chenodeoxycholic acid can serve as a biomarker of B_{12}-deficient state in humans. Furthermore, using mouse B_{12}-deficient tissues, we elucidate how despite only needed for 2 enzyme functions, B_{12} deficiency alters the metabolic and transcriptomic landscape in the cells, which will facilitate advances in further understanding biology of B_{12}.

RESULTS

Study population, sample classification, acquisition, pre-processing, and normalization of metabolomic data

A schematic diagram illustrating different steps of this study is presented in **Figure 1**. The samples utilized in this study are from the Kuopio Ischaemic Heart Disease Risk Factor (KIHD) study aimed at identifying the risk factors for coronary heart diseases, atherosclerosis, and other related conditions in the Eastern Finnish population.(26) Sera were classified in accordance with internationally established criterion into control subjects (n=13) with B_{12} levels >250 pmol/L, and into deficient subjects (n=8) with B_{12} levels <150 pmol/L.(1, 11, 17, 27) Samples were randomized before metabolite extraction and quantified using a ACQUITY UPLC-MS/MS system. Ninety-four metabolites could be detected in the sera, out of which 77 that passed quality control were selected for further downstream analysis. Imputation of one missing value with the minimum value in that cohort was done, and data was pre-processed by generalized log transformation (glog) and auto-scaling of metabolite concentration peaks in each sample to represent uniform distribution.

Identification of differentially expressed serum metabolites following B_{12} deficiency

We first performed a principal component analysis (PCA), an unsupervised multivariate analysis, to group/classify samples without any consideration of prior classification to detect any outliers in the two cohorts. The principal component 1 (PC1) accounted for 22.6% of the variance and PC2 accounted for 13.6% of the variance (**Figure 2A**). To identify differential concentration of each

metabolite between the control and B_{12}-deficient groups, we calculated the mean fold change and performed t-tests to compare the mean of each metabolite. A metabolite was considered significantly different between each group when the value of $p \leq 0.05$ and log2 fold change ±0.5. In the colvano plot the 3 blue dots in the upper left and 3 red dots in upper right quadrants represent the most significantly altered metabolites in B_{12}-deficient subjects compared to that in controls (**Figure 2B**). A hierarchical clustering analysis (HCA) of the metabolomic data using the top 3 downregulated and top 3 upregulated metabolites showed well-defined clustering of thirteen healthy subjects (pink, left cluster) versus eight B_{12}-deficient subjects (green, right cluster) (**Figure 2C**). The control group showed high abundance (shades of red colour) of taurine, hypoxanthine and xanthine compared to the B_{12}-deficient group, whereas the abundance of glycocholic acid, neopterin and chenodeoxycholic acid was significantly higher in the B_{12}-deficient group as compared to healthy controls (**Figure 2C**). Following the identification of differentially expressed metabolites (DEMs), we did Metabolite Set Enrichment Analysis (MSEA) and Metabolomic Pathway Analysis (MetPA) to determine the metabolic pathways that are associated with differences in the abundance of identified metabolites, and perturbations of which is associated with the B_{12} deficiency. The MSEA classified the 77 DEMs into 50 different metabolic pathways (**Figure 2D**) that include divergent cellular metabolism pathways such as bile acid biosynthesis, amino acid biosynthesis, glucose metabolism, and nucleic acid synthesis, which are listed in the order of descending fold enrichment (**Figure 2D**). Out of the 50 listed pathways, the taurine and hypotaurine metabolism pathway was the most enriched pathway with highest fold enrichment value (-logP value ~6). MetPA results revealed that taurine and hypotaurine metabolism pathway had the highest pathway impact value between the controls and B_{12}-deficient subjects, further validating the importance of this pathway (**Figure 2E**).

Once we identified the most significant DEMs and major pathways to which these DEMs belonged to, we wanted to check the consistency of identified DEMs as most discriminant variables for classifying healthy controls versus B_{12}-deficient subjects. For this purpose, we performed a PLS-DA analysis that helps in highlighting whether a metabolite is upregulated or downregulated in a group/sample by creating a latent structure, and the values of variable importance projection (VIP) score which represent the importance of the metabolite in the PLS-DA model (**Figure 2F**). The VIP score plot (threshold of >1.0) revealed that taurine had the maximum score with low abundance in B_{12}-deficient samples versus controls (**Figure 2F**). The other metabolites that were identified in volcano plot i.e., xanthine, hypoxanthine, chenodeoxycholic acid, neopterin, and glycocholic acid also came up in PLS-DA plot, suggesting the consistency of these metabolites as important DEMs in controls versus B_{12}-deficient subjects. Further, we performed univariate analysis (t-test) on individual DEMs to determine the significant difference in the abundance of each metabolite between the two groups. Based on the analysis, abundance of taurine (p=0.002), xanthine (p=0.019) and hypoxanthine (p=0.000) was significantly lower whereas the levels of chenodeoxycholic acid (p= 0.063), neopterin (p= 0.023), and glycocholic acid (p= 0.027) was significantly higher in sera of B_{12}-deficient subjects (green bars) compared to healthy controls (pink bars) (**Figure 2G**).

Metabolites that belong to the same pathway tend to work in coherence. To this end, we subjected the metabolite data to Pearson's correlation matrix analysis to reveal any correlation that might exist between the 77 identified metabolites or between 21 study subjects (**Figure S1A-B**). Between the two cohorts, metabolites such as taurine, xanthine, and hypoxanthine were positively correlated (red color) to each other and negatively correlated (blue color) to chenodeoxycholic acid, neopterin, and glycocholic acid (**Figure S1A**). Moreover, there was a high positive correlation observed between all the essential amino acids. This suggests a strong inter-relationship

between these metabolites which could be expected as these belong to same metabolic pathway such as amino acid biosynthesis. Pearson's correlation matrix analysis on the different cohort subjects, however, revealed no significant trends (**Figure S1B**), suggesting no inter-relationship or correlation between the samples, which negates the possibility of any biases in the sample workflow.

Taken together, these multiple lines of evidence suggest that taurine, hypoxanthine, xanthine, chenodeoxycholic acid, neopterin, and glycocholic acid are the most significant DEMs in the sera of healthy controls versus B_{12}-deficient subjects. Pathway enrichment analysis further confirmed that the alteration in taurine and hypoxanthine metabolic pathway is strongly associated with B_{12} deficiency.

Selection and identification of metabolite and/or metabolite ratio as biomarker

To identify the best metabolite and/or metabolites ratio that could serve as a sensitive biomarker for prediction of B_{12} deficiency, we subjected the data to two statistical analysis tools: Partial least squares discriminant-analysis (PLS-DA) (**Figure 3A and 3E**) and Random forest (RF) analysis (**Figure 3C and 3G**). Multiple statistical models generated by these analyses were validated and compared for their ability to identify the metabolite or metabolites ratio which can serve as the best biomarker to predict B_{12} deficiency. All models generated by PLS-DA or RF were validated using Receiver Operating Characteristic (ROC) analysis, in which Area Under the Curve (AUC) score was used to monitor the sensitivity and specificity of a model (variable) in predicting the B_{12} deficiency. Although both are predictive modelling tools, PLS-DA analysis has a tendency to overfit even on completely random data as compared to RF analysis. Thus, the quality of the models was further assessed using Monte-Carlo cross validation (MCCV) to create ROC curve for every model generated from both PLS-DA and RF analysis. These models use a combination of

the most important features to build classification models, ranging from a minimum of 2 to a maximum of 100. Since MCCV uses defined sub-sampling, 2/3 of the samples were used to evaluate the feature importance and 1/3 of the samples were used for validation. This iterative procedure was used to calculate the performance (AUC) and confidence interval of each model and the one with AUC closest to 1 with low variability (CI) was considered to be the best model. The software gave output in the form of ROC curves of top 6 models, referred to as variables, based on the CV performance. we used the most significant DEMs (**Figure 3A & C**) or metabolite ratio (**Figure 3E & G**) as top features to generate best 6 models for prediction of B_{12} deficiency. Note that the nomenclature of models (referred to as variables, hereinafter) is representative of the number of features used to create the model. **Figure 3B, D, F, and H** represent the ROC curve for the top 6 models obtained following PLS-DA and RF analysis, whereas the model numbers 1, 2, 3, 4, 5 and 6 represent the variables (Var.) 3 (red), 5 (green), 10 (blue), 20 (cyan), 28 (pink) and 77 (yellow), respectively, signifying that model 1 was created using 2 metabolites of top importance, whereas model 6 used top 77 metabolites.

Both PLS-DA (**Figure 3A**) and RF (**Figure 3C**) analysis, using singular metabolites as features, showed that models with more than 20 metabolites (38 and 77) have high AUC (> 7) and tight CI, suggesting their potential to be better models, compared to those with fewer than 20 metabolites. A higher score suggests better predictive ability of a model to identify the B_{12}-deficient state. The feature ranking plot for both PLS-DA (**Figure 3B**) and RF (**Figure 3D**) analysis showed the top 15 metabolites arranged in descending order of average importance scores contributing to the model accuracy. The average importance scores of hypoxanthine and taurine were among the top three metabolites in both analyses, with hypoxanthine having the maximum score. Both models showed lower (blue) abundance of taurine in B_{12}-deficient cohort, but the same was not true for hypoxanthine. This was consistent with PLS-DA analysis done in **Figure 2F**. It is important to

note that (a) 7 of 15 top metabolites were different between the models generated by PLS-DA and RF and (b) the individual average importance score for the 8 identical metabolites varied in the two analyses. This suggested that both analyses work on independent algorithms and there was no bias in the selection of hypoxanthine and taurine as top metabolite biomarkers for predicting B_{12} deficiency.

Next, we investigated whether abundance ratios of metabolite pairs could increase the sensitivity of PLS-DA and RF models to detect B_{12} deficiency (**Figure 3C, 3D**). Ratios of all possible metabolite pairs were computed, and top ranked ratios (based on p values) and top 20 were included for biomarker analysis. Using abundance ratios of metabolite pair as a feature, both PLS-DA (**Figure 3E**) and RF (**Figure 3G**) models showed that all the top 6 models have high AUC (> 9) and high CI which were comparable, suggesting any model with more than 3 features was a good model with high specificity and sensitively but high variability (scattered CI) as well. One-to-one comparison of AUC and CI scores for both the PLS-DA and RF models based on the abundance ratios of metabolite pair versus singular metabolites revealed that the former can serve as better biomarkers in predicting B_{12} deficiency. The feature ranking plot for models in **Figure 3F** and **Figure 3H** listed 13 identical sets of metabolite pairs with taurine/chenodeoxycholic acid gaining the highest average importance score in both (**Figure 3G-H**). The abundance for taurine/chenodeoxycholic acid ratio however was reversed in the two models, being low (blue) in PLS-DA and high (red) in RF for B_{12}-deficient group (**Figure 3E, 3G**). It is important to note that this analysis was consistent with the previous analysis shown in **Figure 2** (PCA, volcano plot, PLS-DA and univariate analysis).

Together, results suggest that out of the metabolites identified to be differentially expressed between healthy controls and B_{12}-deficient group taurine, hypoxanthine and the ratio of taurine/chenodeoxycholic acid could serve as biomarkers for B_{12} deficiency.

Comparison of the abilities of taurine, hypoxanthine and taurine/chenodeoxycholic acid ratio to predict B_{12}-deficient state

We performed ROC analysis to further characterise the predictive ability of taurine alone, hypoxanthine and taurine/chenodeoxycholic acid ratio, which were shortlisted from previous PLS-DA and RF analysis. The sensitivity and significance of taurine, hypoxanthine and taurine/chenodeoxycholic acid in predicting B_{12} deficiency is represented using AUC score from ROC analysis (**Figures 4A-C**). The scaled concentration of the indicated metabolites are shown in **Figures 4D-F**. This analysis showed that AUC for taurine/chenodeoxycholic abundance ratio was 1, which is equivalent to being a perfect diagnostic biomarker (**Figure 4C**). Furthermore, the AUC and *p*-values for taurine/chenodeoxycholic acid ratio were the lowest (*p*-value=5.3193E-7) in comparison to hypoxanthine (AUC = 0.885, *p*-value =7.0513E-4) and taurine alone (AUC = 0.885, *p*-value =0.002), suggesting that taurine/chenodeoxycholic ratio was the best variable as a biomarker to predict B_{12} deficiency compared to others. Between taurine and hypoxanthine, the AUC scores were comparable, but hypoxanthine was significant in differentiating the two groups because of lower *p*-value.

These results suggest that serum taurine/chenodeoxycholic acid abundance ratio can serve as a diagnostic biomarker for predicting B_{12} deficiency with high specificity and sensitivity.

To further test the ability of RF using taurine alone or and in combination with other metabolites as biomarker to predict B_{12} deficiency, we trained a RF model on train data using cross validation and predicted on the test data. For unbiased assessment, equal number of samples (n=4/group) were randomly selected from control and B_{12}-deficient group as hold-out samples. These samples were

not used for fitting process in the model but used as testing samples. The rest of the samples were used as training samples to predict B_{12} deficiency. We compared predictive ability of taurine alone, taurine and hypoxanthine, and ratio of taurine/chenodeoxycholic acid using AUC score (ROC analysis), predicted class probabilities, and cross validation (CV) prediction (**Figure 5**). Amongst these model (**Figure 5A, 5C, 5E**) comparisons, taurine/chenodeoxycholic acid showed the highest margin of separation between the control (empty grey circles, left edge of x-axis) and B_{12}-deficient (filled grey circles, right edge of x-axis) group in training set, (**Figure 5E**). Also, the hold-out samples from both groups (control = empty red circles, B_{12}-deficient = red filled circles) fit perfectly well with the corresponding group in testing data set. Moreover the ROC-AUC curve showed that taurine/chenodeoxycholic abundance ratio had the highest accuracy (AUC CV=1, AUC holdout =1, **Figure 5F**) in predicting B_{12} deficiency compared to taurine alone (AUC CV = 0.665, AUC holdout=0.938, **Figure 5B**) or hypoxanthine (AUC CV= 0.809, holdout=0.938, **Figure 5D).** Overall, this analysis was consistent with previous RF analysis, suggesting towards great potential of taurine/chenodeoxycholic acid to serve as serum biomarker for predicting B_{12} deficiency.

Metabo-transcriptomic network analysis linked B_{12}-dependent reactions with taurine/chenodeoxycholic acid.

We performed a network analysis of differentially expressed genes and metabolites between controls and B_{12}-deficient livers in a mouse model of B_{12} deficiency reported previously by us.(28) Liver is a suitable tissue to investigate effects of B_{12} deficiency since it is one of the principal site of B_{12} storage, and we demonstrated earlier that B_{12} deficiency compromises its functions.(28) In the cells, B_{12} is known thus far to be converted into two cofactors (methyl-B_{12} and adenosyl-B_{12}), which are required for the functioning of two known enzymes, methionine synthase and methyl-malonyl CoA mutase.(29, 30) Thus, we focused our attention on metabolic

pathways that are interconnected with the B_{12}-derived cofactor-dependent reactions such as Krebs cycle, amino acid metabolism, urea cycle, and nucleotide metabolism.

The network visualization of differentially expressed transcriptome showed that transcripts encoding the enzymes that catalyze metabolite conversions in these pathways were overall downregulated (in blue), except for the Krebs cycle, in which expression of 5 out of 9 enzymes was upregulated (in red) (**Figure 6**). This upregulation in the expression levels of Krebs cycle enzymes could be linked to decreased activity of methyl-malonyl CoA mutase (Mut), which is dependent on the adenosyl-B_{12} for its activity. Mut catalyzes the synthesis of Succinyl-CoA, an intermediate in the Krebs cycle that plays a critical role in providing protons for the OXPHOS system, and thus, energy production in the cells. B_{12} deficiency leads to an energy deficit in the cells, and consequently likely, a compensatory increase in the expression levels of enzymes in the Krebs cycle. However, no reactions surrounding the adenosyl-B_{12}-dependent Mut enzyme and Krebs cycle could relate to known taurine biosynthetic machinery in B_{12}-deficient cells.

An analysis of reactions surrounding methionine synthase (Mtr), the second enzyme that is dependent on the methyl-B_{12} as a cofactor, showed that the concentrations of methionine, the downstream product, were decreased while concentrations of its precursor, homocysteine, were increased (**Figure 6**). Expression levels of the enzymes in the methionine cycle were either not affected or were decreased. The methionine cycle is linked to cysteine synthesis in the cells and through a relay of changes, to taurine biosynthesis. Most of the enzymes and their downstream products in this pathway were downregulated, consequently leading to deficiency of multiple metabolites in taurine metabolic pathway (taurine, taurocholate, tauro-chenodeoxycholate) (**Figure 6**). The expression levels of the enzyme, *Csad*, that catalyzes the rate limiting step in taurine

biosynthesis, was increased likely as a compensatory mechanism due to deficiency of taurine (**Figure 6**).

Further analysis of gene-metabolite networks interconnected with B_{12}-dependent reactions showed that gene expression of enzymes and metabolite intermediates in the urea cycle were downregulated. In the amino acid metabolism pathway, barring tryptophan metabolite, HIAA and NAD^+ pathways, all enzyme expressions and metabolite intermediates were downregulated. In the nucleotide metabolism pathways, metabolite intermediates were either downregulated or not affected, and apart from a few enzymes, most of the enzyme expressions were downregulated.

Together, these integrated metabolomic and transcriptomic analyses in the WT and B_{12}-deficient liver samples revealed global downregulation of metabolic networks upon B_{12} deficiency and identified a hitherto unanticipated connectivity between B_{12}-dependent reactions and taurine metabolism.

DISCUSSION:

By using metabolomic analysis of serum from controls and B_{12}-deficient subjects, we were able to identify that a ratio of taurine/chenodeoxycholic acid levels can serve as a biomarker of, difficult to detect, B_{12} deficiency. The quantitative metabolomic analysis of 77 relevant metabolites in the sera of B_{12}-deficient patients revealed that most of the metabolites were downregulated and are involved in metabolism of amino acids, betaine, glutathione, bile acid, and purines (**Figure 2**). Metabolite set enrichment analysis on the perturbed metabolite profiles showed alterations in the metabolic pathways associated with amino acid and methionine metabolism (**Figure 1**). Downregulation in methionine levels in this metabolome is consistent with the role of B_{12} as an essential cofactor of methionine synthase, while homocysteine accumulated from the dysfunction

of methionine synthase was 1.8-fold elevated. Furthermore, univariate analysis of the B_{12}-deficient metabolome identified a differential abundance of taurine, hypoxanthine, and xanthine between the two groups. The multivariate random forest (RF) analysis aimed towards identifying which metabolite(s) contributed to the separation of the two groups with higher specificity and sensitivity showed taurine/chenodeoxycholic ratio as the metabolic parameter that could separate the two groups with 99% accuracy. Thus, we propose taurine/chenodeoxycholic acid ratio as a potential biomarker of a B_{12}-deficient state in humans.

Previous studies have characterized the human serum metabolome in B_{12}-deficient subjects in an attempt to reveal connections between B_{12}-deficient state and serum metabolic markers. Alex et al., performed metabolomic profiles in sera of Chilean older adults with subclinical borderline B_{12} deficiency (defined by serum B_{12} <148 pmol/L, holotranscobalamin <35 pmol/L, tHcy >15 μmol/L, or MMA >271 nmol/L).(31) Although, this study showed perturbations in multiple metabolite such as acylcarnitine and plasmalogens Authors did not subject their data to downstream algorithms to identify potential biomarkers of B_{12} levels. Moreover, the previous study did not include a control group, whereas our study has a well-defined control group. Although, these studies provide evidence that serum metabolome is altered by B_{12} deficiency it was unknown whether any of the metabolites of set of metabolites could serve as a biomarker of B_{12}-deficient state. Our study fills this gap in our knowledge and elucidates the effect of B_{12} deficiency on the cellular, metabolic and transcriptomic landscape of the cell using liver biopsies from a B_{12}-deficient mouse model. Together, these studies pave a way towards better understanding of the cellular defects caused by B_{12} deficiency.

We acknowledge that our study has certain limitations. Firstly, the small sample size limits the statistical power of the RF models. Repeating the same study in a larger sample size may allow a greater number of metabolites to pass quality control for downstream analysis. Secondly, the

current study population was only tested for B_{12} deficiency, which does not rule out the possibility of deficiency of other vitamins or nutrients in the study population. These and other questions will need to be addressed in future studies.

Vitamin B_{12} deficiency leads to perturbed levels of taurine, hypoxanthine, xanthine, chenodeoxycholic acid, neopterin, and glycocholic acid. We show that taurine levels alone and taurine/chenodeoxycholic acid ratio are promising candidates for serum metabolite-based biomarkers to identify B_{12} deficiency. The two critical metabolites identified in this study regulated by B_{12}, taurine and chenodeoxycholic acid, belong to the taurine metabolic pathway. Taurine metabolism gets compromised with age and leads to taurine deficiency in humans, however, the cause of this deficiency is unknown(24). The present study identifies vitamin B_{12} as the very first upstream regulator of taurine metabolism in aged humans and illustrates the transcriptomic and metabolomic changes through which B_{12} regulates this process. These results are significant given that taurine deficiency has recently been shown to be a driver of aging in diverse species, and is associated with poor health in humans. This study paves a way for future clinical work to streamline diagnostic tools to detect B_{12} deficiency through a simple blood test and perhaps other age-associated diseases.

DISCLOSURE STATEMENT

Acknowledgements: We thank Research Support Facility staff at Sanger Institute and National Institute of Immunology especially Dr. P. Nagarajan for assistance with animal experiments.

Financial Support: This work was supported by Wellcome Trust grant (09851) to VKY and and a core Grant from National Institute of Immunology to VKY.

Conflict of Interest: Authors declare no conflict of interest

Authorship: All authors have seen and approved the manuscript

MATERIAL AND METHODS:

Chemicals and reagents

All the metabolite standards, ammonium formate, ammonium acetate and ammonium hydroxide were obtained from Sigma-Aldrich (Helsinki, Finland). Formic acid (FA), 2-proponol, acetonitrile (ACN), and methanol (all HiPerSolv CHROMANORM, HPLC grade, BDH Prolabo) were purchased from VWR International (Helsinki, Finland). Isotopically labelled internal standards were obtained from Cambridge Isotope Laboratory. Inc., USA (Ordered from Euriso-Top, France). Deionized Milli-Q water up to a resistivity of 18 MΩ cm was purified with a purification system (Barnstead EASYpure RoDi ultrapure water purification system, Thermo scientific, Ohio, USA).

Metabolite extraction protocol

The working calibration solutions were prepared in 96-well plate by serial dilution of the stock calibration mix using Hamilton's MICROLAB® STAR line (Hamilton, Bonaduz AG, Switzerland) liquid handling robot system. Starting from a stock solution mix, 10 additional lower working solutions were prepared using water as the diluent to build the calibration curves.

Clinical serum samples:

Clinical samples used for assessing the changes in vitamin B_{12} levels and metabolites in blood were obtained from the Kuopio Ischaemic Heart Disease Risk Factor Study (KIHD study), a population-based cohort study described previously (25, 32), and were donated by J. Kauhanen and T. Nurmi (University of Eastern Finland, Kuopio, Finland). Ten microliters of labelled internal standard mixture was added to 100 µL of serum sample. Metabolites were extracted by adding 4 parts (1:4, sample: extraction solvent) of the 100% ACN + 1% FA solvent. The collected extracts

were dispensed in OstroTM 96-well plate (Waters Corporation, Milford, USA) and filtered by applying vacuum at a delta pressure of 300-400 mbar for 2.5 min on robot's vacuum station. This resulted a cleaner extract to the 96-well collection plate, which was placed under the OstroTM plate. The collection plate was sealed with the cap map and placed in auto-sampler of the LC system for the injection.

Instrumentation and analytical conditions

Sample analysis was performed on an ACQUITY UPLC-MS/MS system (Waters Corporation, Milford, MA, USA). The auto-sampler was set at 5°C, and the column, 2.1 × 100 mm Acquity 1.7um BEH amide HILIC column (Waters Corporation, Milford, MA, USA), temperature was maintained at 45°C. The total run time is 14.5 min including 2.5 min of equilibration step at a flow rate of 600 µL/min. Initially the gradient started with a 2.5 min isocratic step at 100% mobile phase B (ACN/ H2O, 90/10 (v/v), 20 mM ammonium formate, pH at 3), and then rising to 100% mobile phase A (ACN/H2O, 50/50 (v/v), ammonium formate, pH at 3) over the next 10 min and maintained for 2min at 100% A and finally equilibrated to the initial conditions for 2.5 min. An injection volume of 5 µL of sample extract was used and two cycles of 300 µL of strong wash (methanol/isopropanol/ACN/H2O, 25/25/25/25, 0.5% FA) and 900 µL of weak wash (methanol/isopropanol/ACN/H2O, 25/25/25/25, 0.5% ammonium hydroxide) and in addition 2 min of seal wash (90/10, methanol/H2O) were carried out. The auto-sampler was used to perform partial loop with needle overfill injections for the samples and standards.

The detection system, a Xevo® TQ-S tandem triple quadrupole mass spectrometer (Waters, Milford, MA, USA), was operated in both positive and negative polarities with a polarity switching time of 20 msec. Electro spray ionization (ESI) was chosen as the ionization mode with a capillary voltage at 0.6 KV in both polarities. The source temperature and desolvation

449 temperature of 120°C and 650°C, respectively, were maintained constantly throughout the
450 experiment. Declustering potential (DP) and collision energy (CE) were optimized for each
451 compound. High pure nitrogen and argon gas were used as desolvation gas (1000 L/hr) and
452 collision gas (0.15 ml/min), respectively. Multiple Reaction Monitoring (MRM) acquisition mode
453 was selected for quantification of metabolites with individual span time of 0.1 sec given in their
454 individual MRM channels. The dwell time was calculated automatically by the software based on
455 the region of the retention time window, number of MRM functions and depending on the number
456 of data points required to form the peak. MassLynx 4.1 software was used for data acquisition,
457 data handling and instrument control. Data processing was done using TargetLynx software and
458 metabolites were quantified by using labelled internal standards and external calibration curves.
459
460 ***Data analysis using MetaboAnalyst 5.0 software and downstream analysis.***
461 The raw data was analyzed using MetaboAnalyst 5.0 software (https://www.metaboanalyst.ca/).
462 (33, 34) Metabolite raw values were generalized log (glog) transformed and auto-scaled (mean-
463 centered and divided by the standard deviation of each variable).(35) Missing values for any
464 metabolites in the sample below the limit of detection were inputted with 1/5 of the minimum
465 positive value for each variable. Unsupervised Principal component analysis (PCA) was done to
466 differentially cluster the two groups.(36, 37) Hierarchical clustering and Pearson's correlation
467 analysis were also performed to cluster the metabolite and sample data in the form of a heatmap to
468 easily identify patterns in metabolite concentrations across samples. Metabolite Set Enrichment
469 Analyses (MSEA)(38) were performed on all metabolites with a VIP ≥1.5 that matched the
470 database using the "Pathway-associated metabolite sets (SMPDB)" database in the MetaboAnalyst
471 software . Pathway analysis was performed using the "Homo sapiens (KEGG(39, 40))" database in
472 the MetaboAnalyst software. Interactive scatter plot with 'Enrichment Factor' as *x* axis and
473 '$-\log_{10}(P)$' as *y* axis was generated for functional analysis to show the significance of top 50

metabolic pathways involving the metabolites identified. The variable importance to projection (VIP) score for each metabolite was calculated to quantitatively represent metabolite feature importance in the model. A volcano plot scatterplot that shows statistical significance (-log10(p-value) versus magnitude of change (log 2-fold change) of metabolites. Metabolites that show significant ($p \leq 0.05$) change (log 2-fold change ± 0.5) are highlighted. Multivariate supervised Partial least squares discriminant analysis (PLS-DA) and Random-forest (RF) analysis were performed to assess the difference between the abundance of top metabolites or metabolite ratio between the two groups. The area under the curve (AUC) of the receiver operating characteristic (ROC) curve was also calculated for each metabolite to determine its predictive ability as a biomarker. The ROC curve is a plot of false positive rate (FPR) vs the true positive rate (TPR). The higher the AUC value, the better the measurements are at classifying between the two groups.

REFERENCES:

1. Stabler SP. Vitamin B12 deficiency. New England Journal of Medicine. 2013;368(2):149-60.
2. Lazar MA, Birnbaum MJ. De-meaning of metabolism. Science. 2012;336(6089):1651-2.
3. Hunt A, Harrington D, Robinson S. Vitamin B12 deficiency. Bmj. 2014;349.
4. Nielsen MJ, Rasmussen MR, Andersen CB, et al. Vitamin B12 transport from food to the body's cells—a sophisticated, multistep pathway. Nature reviews Gastroenterology & hepatology. 2012;9(6):345-54.
5. Quadros EV, Regec AL, Khan KF, et al. Transcobalamin II synthesized in the intestinal villi facilitates transfer of cobalamin to the portal blood. American Journal of Physiology-Gastrointestinal and Liver Physiology. 1999;277(1):G161-G6.
6. Kalhan SC, Marczewski SE. Methionine, homocysteine, one carbon metabolism and fetal growth. Reviews in Endocrine and Metabolic Disorders. 2012;13:109-19.
7. Stabler SP, Allen RH. Vitamin B12 deficiency as a worldwide problem. Annu Rev Nutr. 2004;24:299-326.
8. Van Meurs JB, Dhonukshe-Rutten RA, Pluijm SM, et al. Homocysteine levels and the risk of osteoporotic fracture. New England Journal of Medicine. 2004;350(20):2033-41.
9. McLean RR, Jacques PF, Selhub J, et al. Homocysteine as a predictive factor for hip fracture in older persons. New England Journal of Medicine. 2004;350(20):2042-9.
10. Toh B-H, van Driel IR, Gleeson PA. Pernicious anemia. New England Journal of Medicine. 1997;337(20):1441-8.

11. Stabler SP, Allen RH, Savage DG, et al. Clinical spectrum and diagnosis of cobalamin deficiency. Blood. 1990;76(5):871-81.
12. Stabler SP, Marcell PD, Podell ER, et al. Elevation of total homocysteine in the serum of patients with cobalamin or folate deficiency detected by capillary gas chromatography-mass spectrometry. The Journal of clinical investigation. 1988;81(2):466-74.
13. Stabler S. Megaloblastic anemias: pernicious anemia and folate deficiency. Clinical hematology Philadelphia: Mosby. 2006:242-51.
14. Klee GG. Cobalamin and folate evaluation: measurement of methylmalonic acid and homocysteine vs vitamin B(12) and folate. Clin Chem. 2000;46(8 Pt 2):1277-83.
15. Lindenbaum J, Savage DG, Stabler SP, et al Diagnosis of cobalamin deficiency: II. Relative sensitivities of serum cobalamin, methylmalonic acid, and total homocysteine concentrations. Am J Hematol. 1990;34(2):99-107.
16. Nygard O, Nordrehaug JE, Refsum H, et al. Plasma homocysteine levels and mortality in patients with coronary artery disease. N Engl J Med. 1997;337(4):230-6.
17. Lindenbaum J, Healton EB, Savage DG, et al. Neuropsychiatric disorders caused by cobalamin deficiency in the absence of anemia or macrocytosis. New England journal of medicine. 1988;318(26):1720-8.
18. Savage DG, Lindenbaum J, Stabler SP, et al. Sensitivity of serum methylmalonic acid and total homocysteine determinations for diagnosing cobalamin and folate deficiencies. The American journal of medicine. 1994;96(3):239-46.
19. Lloyd-Wright Z, Hvas A-M, Møller J, et al. Holotranscobalamin as an indicator of dietary vitamin B12 deficiency. Clinical Chemistry. 2003;49(12):2076-8.
20. Dhonukshe□Rutten RA, Pluijm SM, De Groot LC, et al. Van Staveren WA. Homocysteine and vitamin B12 status relate to bone turnover markers, broadband ultrasound attenuation, and fractures in healthy elderly people. Journal of Bone and Mineral Research. 2005;20(6):921-9.
21. Goerss JB, Kim CH, Atkinson EJ, et al. Risk of fractures in patients with pernicious anemia. Journal of Bone and Mineral Research. 1992;7(5):573-9.
22. Tucker KL, Hannan MT, Qiao N, et al. Low plasma vitamin B12 is associated with lower BMD: the Framingham Osteoporosis Study. Journal of Bone and Mineral Research. 2005;20(1):152-8.
23. Morris MC, Evans DA, Bienias JL, et al. Dietary folate and vitamin B12 intake and cognitive decline among community-dwelling older persons. Archives of neurology. 2005;62(4):641-5.
24. Singh P, Gollapalli K, Mangiola S, et al. Taurine deficiency as a driver of aging. Science. 2023;380(6649):eabn9257.
25. Roman-Garcia P, Quiros-Gonzalez I, Mottram L, et al. Vitamin B(1)(2)-dependent taurine synthesis regulates growth and bone mass. J Clin Invest. 2014;124(7):2988-3002.
26. Jussi K. Kuopio Ischemic Heart Disease Risk Factor Study. Encyclopedia of Behavioral Medicine. 2020:1271-2.
27. Pennypacker LC, Allen RH, Kelly JP, et al. High prevalence of cobalamin deficiency in elderly outpatients. J Am Geriatr Soc. 1992;40(12):1197-204.
28. Roman-Garcia P, Quiros-Gonzalez I, Mottram L, et al. Vitamin B 12–dependent taurine synthesis regulates growth and bone mass. The Journal of clinical investigation. 2014;124(7):2988-3002.
29. Nielsen MJ, Rasmussen MR, Andersen CB, et al. Vitamin B12 transport from food to the body's cells--a sophisticated, multistep pathway. Nat Rev Gastroenterol Hepatol. 2012;9(6):345-54.

30. Clemens TL. Vitamin B12 deficiency and bone health. N Engl J Med. 2014;371(10):963-4.
31. Brito A, Grapov D, Fahrmann J, et al. The human serum metabolome of vitamin B-12 deficiency and repletion, and associations with neurological function in elderly adults. The Journal of Nutrition. 2017;147(10):1839-49.
32. Vanharanta M, Voutilainen S, Rissanen TH, et al. Risk of cardiovascular disease-related and all-cause death according to serum concentrations of enterolactone: Kuopio Ischaemic Heart Disease Risk Factor Study. Arch Intern Med. 2003;163(9):1099-104.
33. Pang Z, Zhou G, Ewald J, et al. Using MetaboAnalyst 5.0 for LC-HRMS spectra processing, multi-omics integration and covariate adjustment of global metabolomics data. Nat Protoc. 2022;17(8):1735-61.
34. Pang Z, Chong J, Zhou G, et al. MetaboAnalyst 5.0: narrowing the gap between raw spectra and functional insights. Nucleic acids research. 2021;49(W1):W388-W96.
35. Saccenti E, Hoefsloot HC, Smilde AK, et al. Reflections on univariate and multivariate analysis of metabolomics data. Metabolomics. 2014;10:361-74.
36. Worley B, Powers R. Multivariate analysis in metabolomics. Current metabolomics. 2013;1(1):92-107.
37. Jolliffe IT. Principal component analysis for special types of data: Springer; 2002.
38. Xia J, Wishart DS. Metabolomic data processing, analysis, and interpretation using MetaboAnalyst. Current protocols in bioinformatics. 2011;34(1):14.0. 1-.0. 48.
39. Ogata H, Goto S, Sato K, et al. KEGG: Kyoto encyclopedia of genes and genomes. Nucleic acids research. 1999;27(1):29-34.
40. Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. Nucleic acids research. 2000;28(1):27-30.

FIGURE LEGENDS:

FIGURE 1:

Study population, sample classification, acquisition, pre-processing, and normalization of metabolomic data. Schematic diagram illustrating the steps for metabolomic analysis of serum samples from B_{12}-deficient (B_{12} levels <150 pmol/L) versus the healthy control group. (1) In this study, 8 and 13 subjects were grouped in B_{12}-deficient and control groups (age- and gender-matched), respectively, (2) blood samples were collected and processed, (3) metabolomics data was acquired from serum samples using ACQUITY UPLC-MS/MS system (Waters Corporation, Milford, MA, USA), data was pre-processed and analyzed using MetaboAnalyst 5.0 to identify (4) differentially expressed metabolites between 2 study groups, (5) serum metabolic biomarker for Vitamin B_{12} deficiency followed by (6) pathway analysis.

FIGURE 2:

Identification of differentially expressed serum metabolites following B_{12} deficiency. (A) Unsupervised multivariate PCA plot showing the spread of control (pink dots) versus B_{12}-deficient (green dots) cohort based on the serum metabolic profile. The horizontal and vertical coordinates are the first and second principal components, respectively. Each dot represents a sample. (B) Volcano plot showing six (blue and red dots) most significant differentially expressed metabolites between the B_{12}-deficient patients versus controls, with a p-value < 0.05 and a log2 fold change ±0.5. X-axis corresponds to log2(Fold Change) and Y-axis to −log10(p-value). (C) Hierarchical clustering analysis sorted the control (pink) versus B_{12}-deficient (green) group based on differential abundance of six metabolites (taurine, hypoxanthine, xanthine, glycocholic acid, neopterin, and chenodeoxycholic acid). Relative abundance scored from 4 (highest, red color) to -4 (lowest, blue). (D) MSEA plot with top 50 enriched metabolic pathways (vertical-axis) to which the 77 identified metabolites belong. The pathways are arranged in descending order of fold enrichment score (horizontal axis) where the highest is 6 (red color) and lowest is 0 (yellow color) (E) MetPA plot showing most enriched pathways with significance (-logP) values for each of the pathway as dots of red (high significance) or yellow (low significance). X-axis corresponds to pathway impact and Y-axis to -logP values. The size of the dot represents its impact value. (F) VIP score plot from PLS-DA analysis showing the top 20 differentially expressed metabolites in serum of control versus B_{12}-deficient group scored from 1 to 2. Relative abundance is depicted with red (highest) and green (lowest) color. (G) Box plots showing normalized concentrations of individual metabolites following univariate analysis: taurine (p=0.002), xanthine (p=0.019) and hypoxanthine

(p=0.000), chenodeoxycholic acid (p=0.063), neopterin (p=0.023), and glycocholic acid (p=0.027) in the sera of control (red) versus B_{12}-deficient (green) groups.

FIGURE 3:

Selection and identification of metabolite and/or metabolite ratio as a biomarker. The top 6 predictive models (Var.) generated by various multivariant analyses were compared for their performance as metabolite biomarker predictors for B_{12} deficiency using ROC-AUC curves based on the MCCV method. ROC-AUC curve for (A) PLS-DA and (C) RF models using singular metabolites as features. ROC-AUC curve for € PLS-DA and (G) RF models using abundance ratio of metabolite pairs as features. Feature ranking plot for (B) PLS-DA and (D) RF models representing the top 15 metabolites arranged in descending value of average importance score. The average importance scores range from 1 to 2 for PLS-DA and 0 to 2 for RF. Feature ranking plot for (F) PLS-DA and (H) RF models representing top 15 abundance ratio of metabolite pairs arranged in descending value of average importance score. The average importance score ranges from 1 to 2 for PLS-DA and 1 to 4 for RF. In all the feature ranking plots the relative abundance of each feature between the control and B_{12}-deficient group was graded with red and blue colors representing high and low abundance, respectively.

FIGURE 4:

Comparison of the abilities of taurine, hypoxanthine and taurine/chenodeoxycholic acid ratio to predict B_{12}-deficient state. ROC-AUC curve showing performance of (A) taurine, (B) hypoxanthine and (C) taurine/chenodeoxycholic acid ratio as biomarker to predict B_{12} deficiency based on AUC (sensitivity, specificity) and CI (variability) values. Each ROC curve is a plot between false positive rate (x-axis) and true positive rate (y-axis). Box plots showing normalized

644 concentration of (D) taurin€(E) hypoxanthine and (F) taurine/chenodeoxycholic acid ratio between
645 control (pink) versus B_{12}-deficient (green) group. Each dot represents a sample. Y-axis represents
646 fold change values. P value <0.05.
647
648 **FIGURE 5:**
649 **Statistical Model to test predictive ability of taurine alone and in combination as biomarker.**
650 Random forest was used as a model to test the predictive abilities of taurine, taurine and
651 hypoxanthine together, and taurine/chenodeoxycholic acid ratio to predict B_{12} deficiency.
652 Predicted class probability plot for (A) taurine, (B) taurine and hypoxanthine together, and (C)
653 taurine/ chenodeoxycholic acid ratio showing the classification accuracy of each factor to
654 differentiate between control (grey dots) and B_{12}-deficient (red dots) samples. The solid dots are
655 training data sets and the empty dots are test data sets. ROC-AUC curve analysis showing cross-
656 validation (pink) and hold-out (blue) scores to determine the performance of (D) tau€e, (E) taurine
657 and hypoxanthine, and (F) taurine/chenodeoxycholic acid ratio as a biomarker to predict B_{12}
658 deficiency. Each ROC curve is a plot between the false positive rate (specificity) on the x-axis and
659 true positive rate (sensitivity) on the y-axis.
660
661 **FIGURE 6:**
662 **Metabo-transcriptomic network analysis links B_{12} dependent reactions with**
663 **taurine/chenodeoxycholic acid.** Network analysis showing the differentially expressed genes and
664 metabolites between controls and B_{12}-deficient livers in a mouse model of B_{12} deficiency reported
665 previously(25). The network shows interactions between enzymes (italics font) and metabolites
666 (normal font) across various metabolic pathways in the liver such as Krebs cycle, urea cycle,
667 amino acid metabolism, nucleotide metabolism, etc. The arrows represent the direction of the

668 reaction. The downregulation and upregulation of enzyme transcript or metabolite concentrations
669 are represented by blue and red color, respectively.
670
671 **FIGURE S1:**
672 **Correlation analysis between metabolites and samples.** Pearson's correlation matrix to identify
673 highly correlated (A) metabolites and (B) samples in two groups. Correlation score ranged from 1
674 (highest, red) to -1 (lowest, blue).
675

1 Study design

Healthy patients

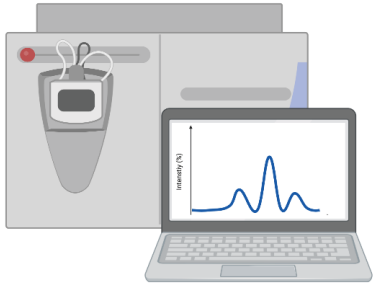
Patients with vitamin B12 deficiency

2 Sample preparation

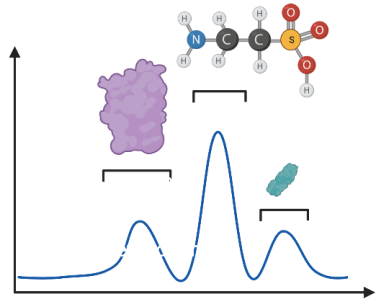
Blood/ serum preparation

3 Data acquisition

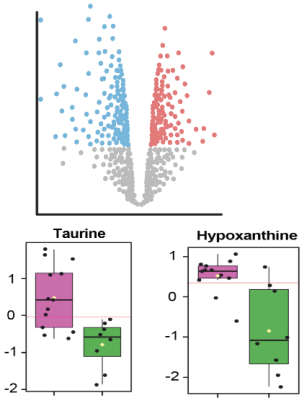
Mass spectrometry

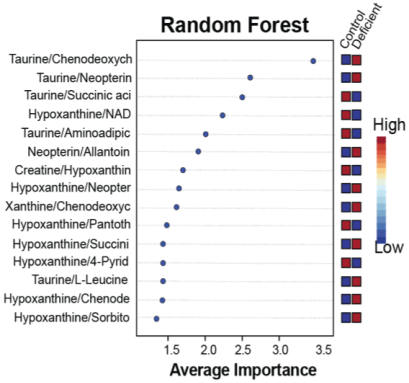


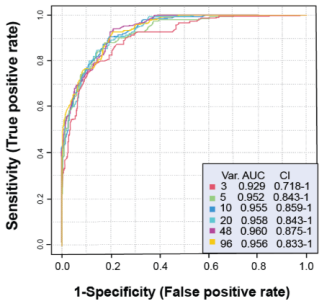
4 Identification of metabolites

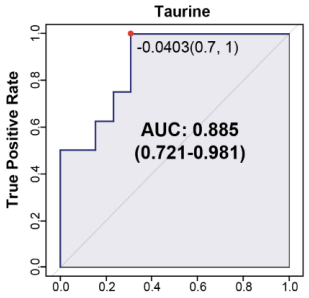


5 Biomarker identification









6 Pathway analysis

Metabolomic and transcriptomic analysis in mouse liver

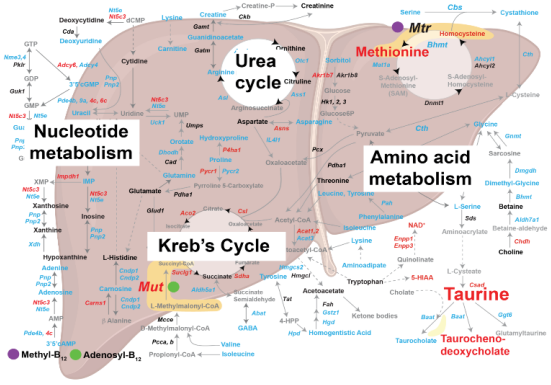


Figure 1

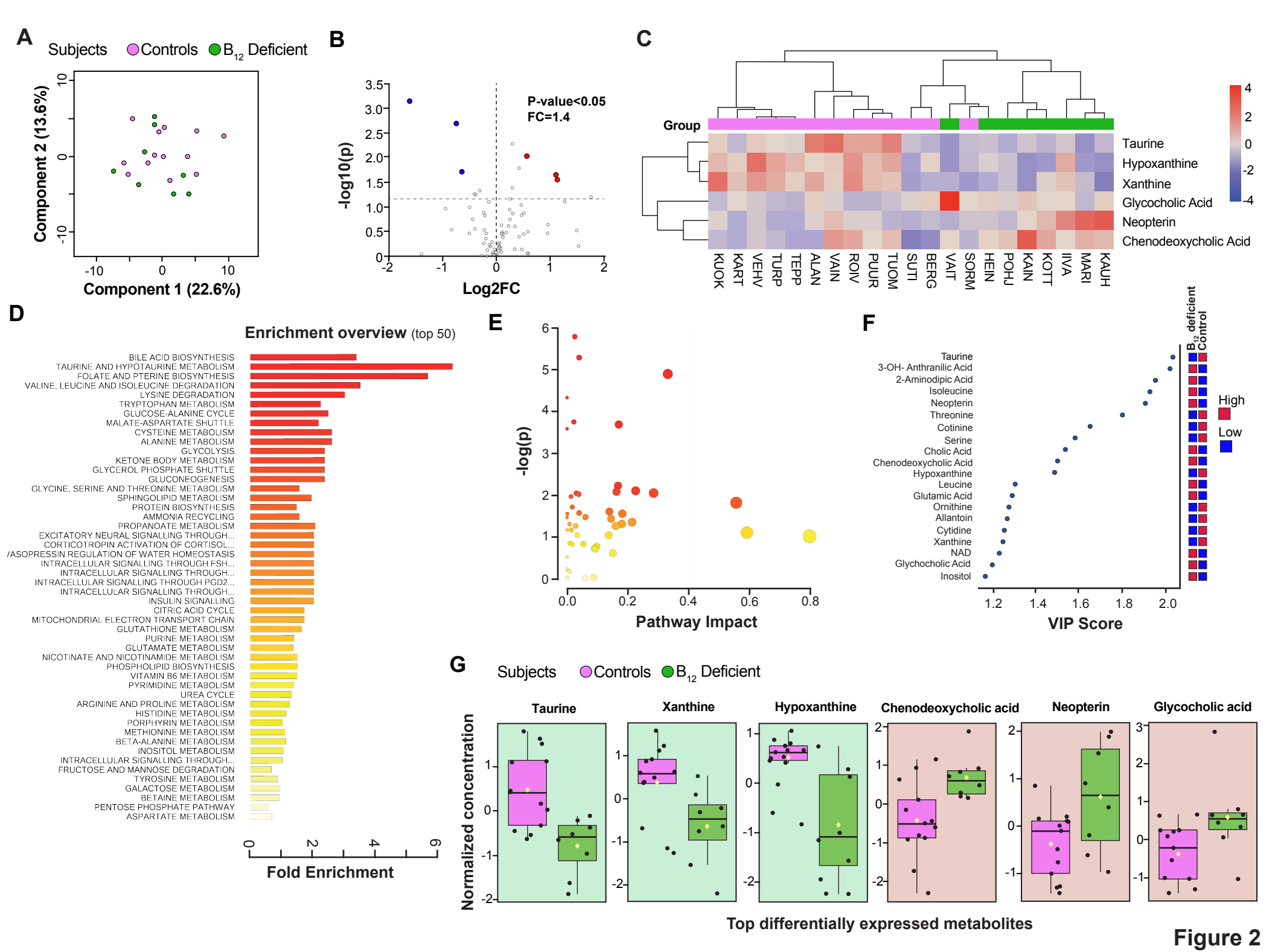

A
Subjects
Controls
B12 Deficient
Component 2 (13.6%)
Component 1 (22.6%)
B
P-value<0.05
FC=1.4
-log10(p)
Log2FC
C
Group
Taurine
Hypoxanthine
Xanthine
Glycocholic Acid
Neopterin
Chenodeoxycholic Acid
KUOK
KART
VEHV
TURP
TEPP
ALAN
VAIN
ROIV
PUUR
TUOM
SUTI
BERG
VAIT
SORM
HEIN
POHJ
KAIN
KOTT
IIVA
MARI
KAUH
D
Enrichment overview (top 50)
BILE ACID BIOSYNTHESIS
TAURINE AND HYPOTAURINE METABOLISM
FOLATE AND PTERINE BIOSYNTHESIS
VALINE, LEUCINE AND ISOLEUCINE DEGRADATION
LYSINE DEGRADATION
TRYPTOPHAN METABOLISM
GLUCOSE-ALANINE CYCLE
MALATE-ASPARTATE SHUTTLE
CYSTEINE METABOLISM
ALANINE METABOLISM
GLYCOLYSIS
KETONE BODY METABOLISM
GLYCEROL PHOSPHATE SHUTTLE
GLUCONEOGENESIS
GLYCINE, SERINE AND THREONINE METABOLISM
SPHINGOLIPID METABOLISM
PROTEIN BIOSYNTHESIS
AMMONIA RECYCLING
PROPANOATE METABOLISM
EXCITATORY NEURAL SIGNALLING THROUGH...
CORTICOTROPIN ACTIVATION OF CORTISOL...
/ASOPRESSIN REGULATION OF WATER HOMEOSTASIS
INTRACELLULAR SIGNALLING THROUGH FSH...
INTRACELLULAR SIGNALLING THROUGH...
INTRACELLULAR SIGNALLING THROUGH PGD2...
INTRACELLULAR SIGNALLING THROUGH...
INSULIN SIGNALLING
CITRIC ACID CYCLE
MITOCHONDRIAL ELECTRON TRANSPORT CHAIN
GLUTATHIONE METABOLISM
PURINE METABOLISM
GLUTAMATE METABOLISM
NICOTINATE AND NICOTINAMIDE METABOLISM
PHOSPHOLIPID BIOSYNTHESIS
VITAMIN B6 METABOLISM
PYRIMIDINE METABOLISM
UREA CYCLE
ARGININE AND PROLINE METABOLISM
HISTIDINE METABOLISM
PORPHYRIN METABOLISM
METHIONINE METABOLISM
BETA-ALANINE METABOLISM
INOSITOL METABOLISM
INTRACELLULAR SIGNALLING THROUGH...
FRUCTOSE AND MANNOSE DEGRADATION
TYROSINE METABOLISM
GALACTOSE METABOLISM
BETAINE METABOLISM
PENTOSE PHOSPHATE PATHWAY
ASPARTATE METABOLISM
Fold Enrichment
E
-log(p)
Pathway Impact
F
Taurine
3-OH- Anthranilic Acid
2-Aminodipic Acid
Isoleucine
Neopterin
Threonine
Cotinine
Serine
Cholic Acid
Chenodeoxycholic Acid
Hypoxanthine
Leucine
Glutamic Acid
Ornithine
Allantoin
Cytidine
Xanthine
NAD
Glychocholic Acid
Inositol
B12 deficient
Control
High
Low
VIP Score
G
Subjects
Controls
B12 Deficient
Taurine
Xanthine
Hypoxanthine
Chenodeoxycholic acid
Neopterin
Glycocholic acid
Normalized concentration
Top differentially expressed metabolites

Figure 2

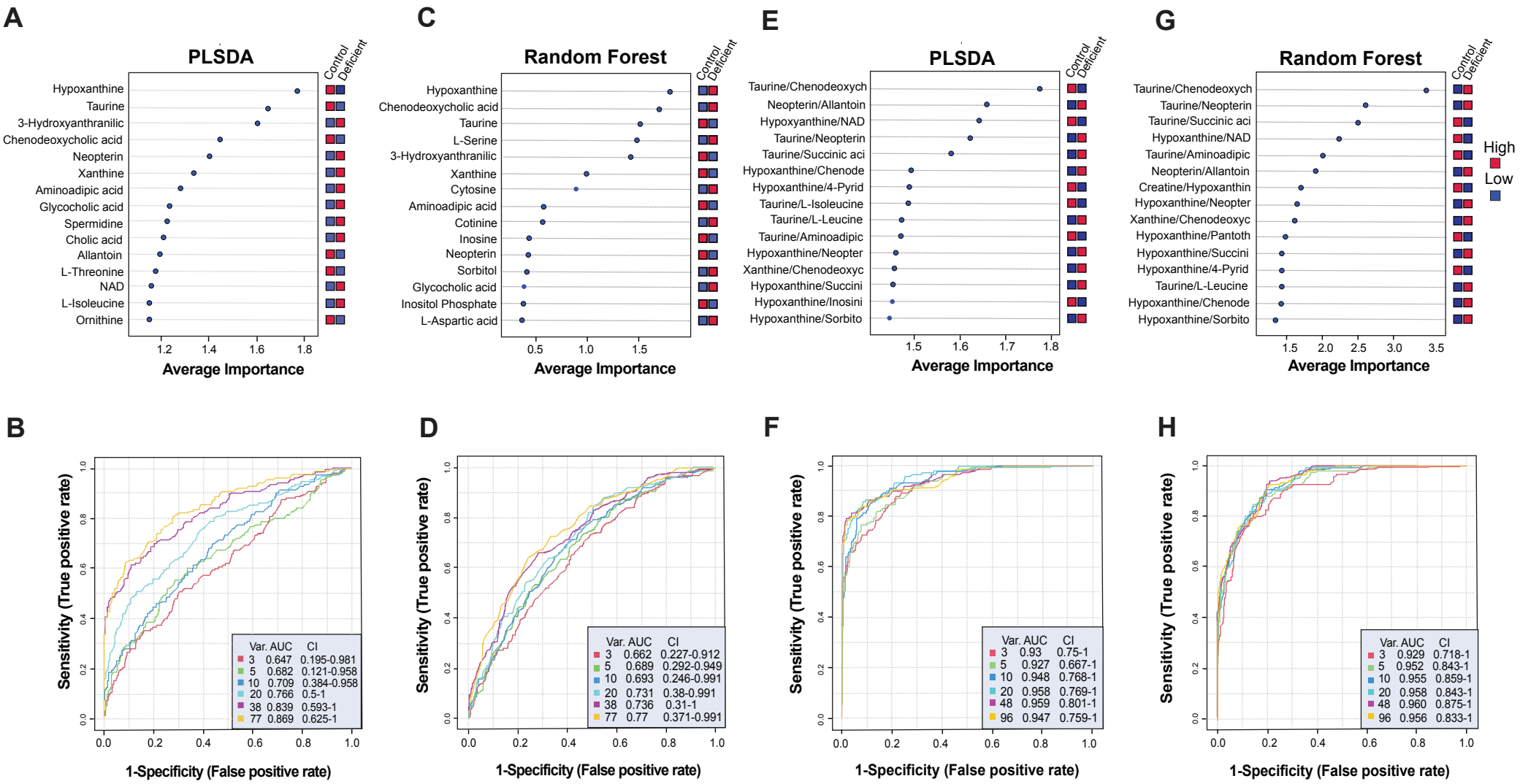

A
PLSDA
Hypoxanthine
Taurine
3-Hydroxyanthranilic
Chenodeoxycholic acid
Neopterin
Xanthine
Aminoadipic acid
Glycocholic acid
Spermidine
Cholic acid
Allantoin
L-Threonine
NAD
L-Isoleucine
Ornithine
Control
Deficient
1.2 1.4 1.6 1.8
Average Importance
C
Random Forest
Hypoxanthine
Chenodeoxycholic acid
Taurine
L-Serine
3-Hydroxyanthranilic
Xanthine
Cytosine
Aminoadipic acid
Cotinine
Inosine
Neopterin
Sorbitol
Glycocholic acid
Inositol Phosphate
L-Aspartic acid
Control
Deficient
0.5 1.0 1.5
Average Importance
E
PLSDA
Taurine/Chenodeoxych
Neopterin/Allantoin
Hypoxyanthine/NAD
Taurine/Neopterin
Taurine/Succinic aci
Hypoxanthine/Chenode
Hypoxanthine/4-Pyrid
Taurine/L-Isoleucine
Taurine/L-Leucine
Taurine/Aminoadipic
Hypoxanthine/Neopter
Xanthine/Chenodeoxyc
Hypoxanthine/Succini
Hypoxanthine/Inosini
Hypoxanthine/Sorbito
Control
Deficient
1.5 1.6 1.7 1.8
Average Importance
G
Random Forest
Taurine/Chenodeoxych
Taurine/Neopterin
Taurine/Succinic aci
Hypoxanthine/NAD
Taurine/Aminoadipic
Neopterin/Allantoin
Creatine/Hypoxanthin
Hypoxanthine/Neopter
Xanthine/Chenodeoxyc
Hypoxanthine/Pantoth
Hypoxanthine/Succini
Hypoxanthine/4-Pyrid
Taurine/L-Leucine
Hypoxanthine/Chenode
Hypoxanthine/Sorbito
Control
Deficient
1.5 2.0 2.5 3.0 3.5
Average Importance
High
Low
B
Sensitivity (True positive rate)
1-Specificity (False positive rate)
Var. AUC CI
3 0.647 0.195-0.981
5 0.682 0.121-0.958
10 0.709 0.384-0.958
20 0.766 0.5-1
38 0.839 0.593-1
77 0.869 0.625-1
D
Sensitivity (True positive rate)
1-Specificity (False positive rate)
Var. AUC CI
3 0.662 0.227-0.912
5 0.689 0.292-0.949
10 0.693 0.246-0.991
20 0.731 0.38-0.991
38 0.736 0.31-1
77 0.77 0.371-0.991
F
Sensitivity (True positive rate)
1-Specificity (False positive rate)
Var. AUC CI
3 0.93 0.75-1
5 0.927 0.667-1
10 0.948 0.768-1
20 0.958 0.769-1
48 0.959 0.801-1
96 0.947 0.759-1
H
Sensitivity (True positive rate)
1-Specificity (False positive rate)
Var. AUC CI
3 0.929 0.718-1
5 0.952 0.843-1
10 0.955 0.859-1
20 0.958 0.843-1
48 0.960 0.875-1
96 0.956 0.833-1

Figure 3

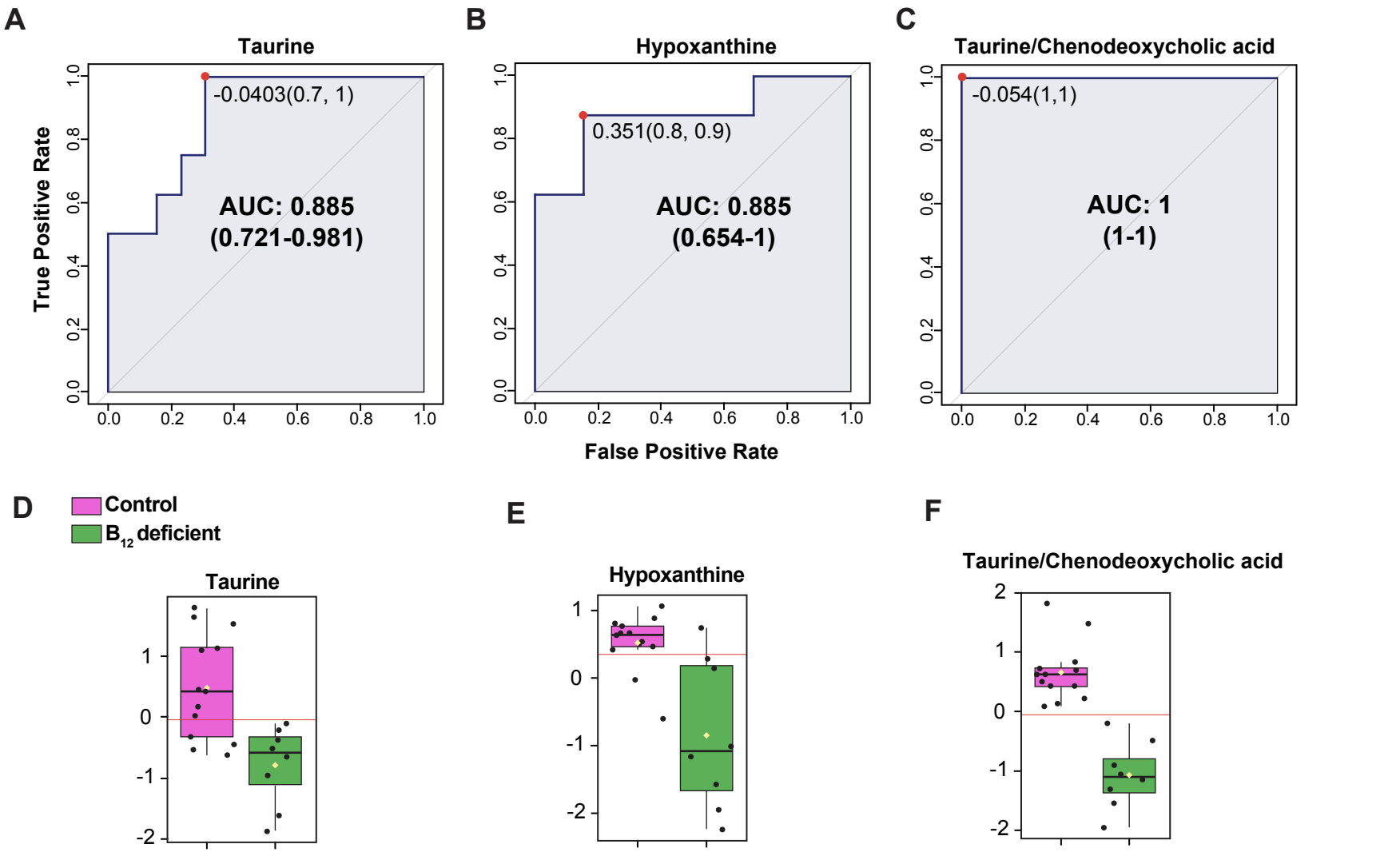

A
Taurine
-0.0403(0.7, 1)
AUC: 0.885
(0.721-0.981)
True Positive Rate
B
Hypoxanthine
0.351(0.8, 0.9)
AUC: 0.885
(0.654-1)
False Positive Rate
C
Taurine/Chenodeoxycholic acid
-0.054(1,1)
AUC: 1
(1-1)
D
Control
B12 deficient
Taurine
E
Hypoxanthine
F
Taurine/Chenodeoxycholic acid

Figure 4

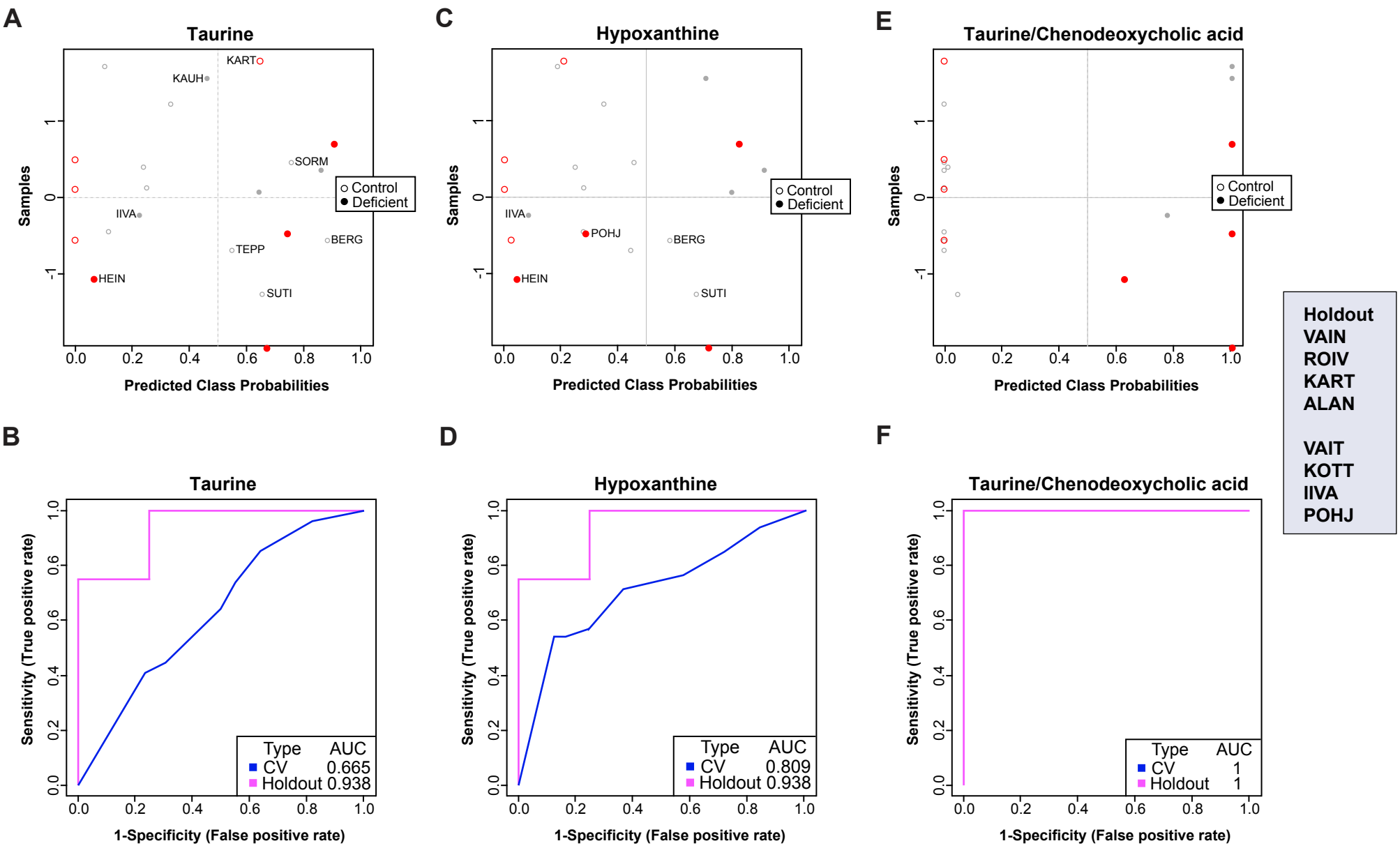

A
Taurine
KART
KAUH
SORM
IIVA
TEPP
BERG
HEIN
SUTI
Samples
Predicted Class Probabilities
Control
Deficient
C
Hypoxanthine
IIVA
POHJ
BERG
HEIN
SUTI
E
Taurine/Chenodeoxycholic acid
B
Taurine
Sensitivity (True positive rate)
1-Specificity (False positive rate)
Type AUC
CV 0.665
Holdout 0.938
D
Hypoxanthine
Type AUC
CV 0.809
Holdout 0.938
F
Taurine/Chenodeoxycholic acid
Type AUC
CV 1
Holdout 1
Holdout
VAIN
ROIV
KART
ALAN
VAIT
KOTT
IIVA
POHJ

Figure 5

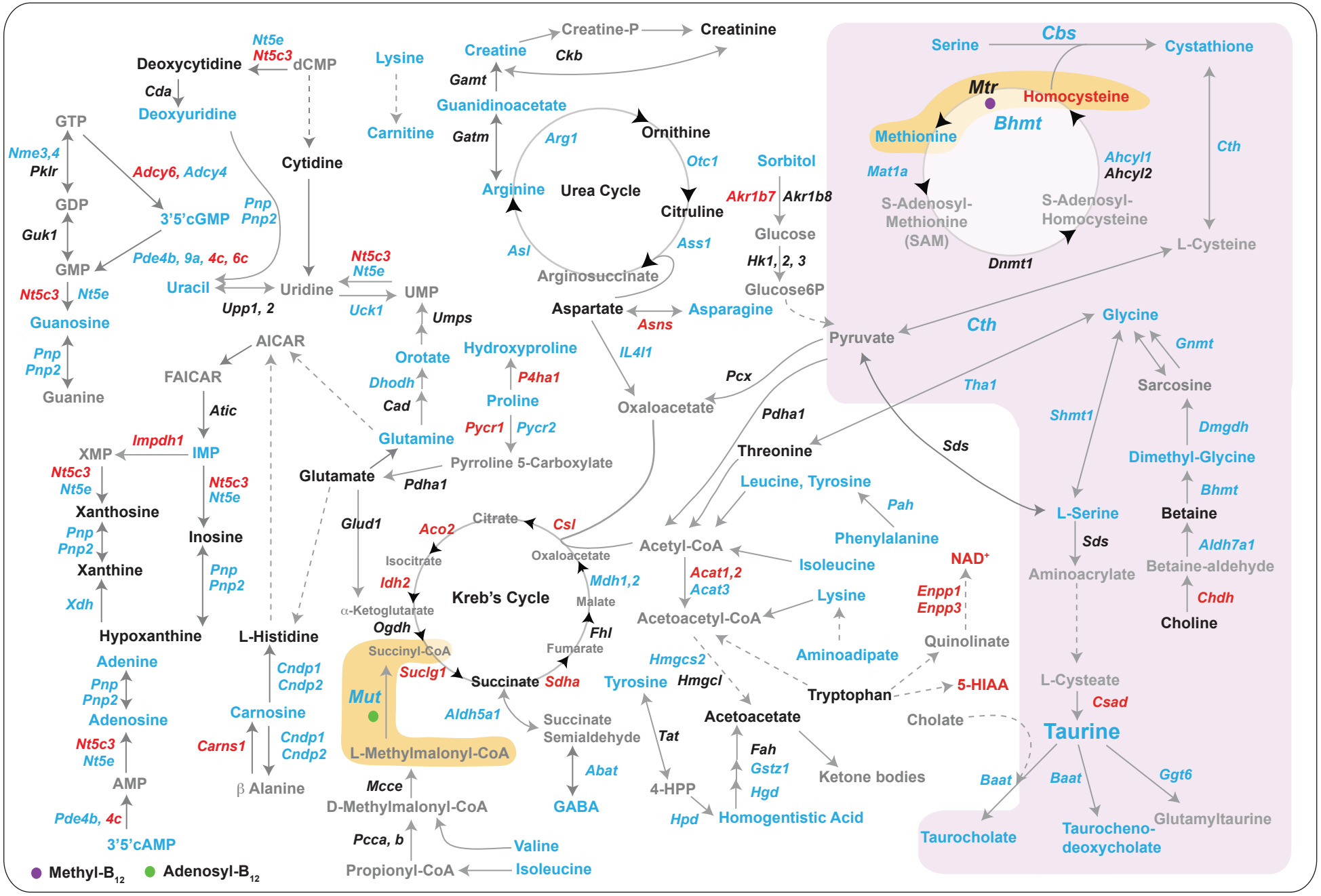

Deoxycytidine
dCMP
Nt5e
Nt5c3
Cda
Deoxyuridine
Cytidine
GTP
Nme3,4
Pklr
GDP
Guk1
GMP
Nt5c3
Nt5e
Guanosine
Pnp
Pnp2
Guanine
Adcy6, Adcy4
3'5'cGMP
Pde4b, 9a, 4c, 6c
Pnp
Pnp2
Uracil
Uridine
Upp1, 2
Nt5c3
Nt5e
UMP
Uck1
Umps
Lysine
Carnitine
Creatine
Creatine-P
Creatinine
Ckb
Gamt
Guanidinoacetate
Gatm
Arginine
Arg1
Urea Cycle
Ornithine
Otc1
Citruline
Asl
Ass1
Arginosuccinate
Aspartate
Asns
Asparagine
Sorbitol
Akr1b7
Akr1b8
Glucose
Hk1, 2, 3
Glucose6P
Serine
Cbs
Cystathione
Mtr
Homocysteine
Bhmt
Methionine
Mat1a
S-Adenosyl-Methionine (SAM)
Ahcyl1
Ahcyl2
S-Adenosyl-Homocysteine
Dnmt1
Cth
L-Cysteine
Pyruvate
Cth
Glycine
Gnmt
Sarcosine
Dmgdh
Dimethyl-Glycine
Bhmt
Betaine
Aldh7a1
Betaine-aldehyde
Chdh
Choline
Tha1
Shmt1
Pcx
Oxaloacetate
Pdha1
Threonine
Sds
L-Serine
Sds
Aminoacrylate
L-Cysteate
Csad
Taurine
Baat
Baat
Ggt6
Taurocholate
Taurocheno-deoxycholate
Glutamyltaurine
Cholate
AICAR
FAICAR
Atic
IMP
Impdh1
XMP
Nt5c3
Nt5e
Xanthosine
Pnp
Pnp2
Xanthine
Xdh
Hypoxanthine
Nt5c3
Nt5e
Inosine
Pnp
Pnp2
Adenine
Pnp
Pnp2
Adenosine
Nt5c3
Nt5e
AMP
Pde4b, 4c
3'5'cAMP
Orotate
Dhodh
Cad
Glutamine
Glutamate
Pdha1
Glud1
Hydroxyproline
P4ha1
Proline
Pycr1
Pycr2
Pyrroline 5-Carboxylate
IL4I1
Leucine, Tyrosine
Pah
Phenylalanine
Isoleucine
Lysine
Aminoadipate
Tryptophan
Tyrosine
Hmgcs2
Hmgcl
Acetoacetate
Fah
Gstz1
Hgd
Homogentistic Acid
4-HPP
Hpd
Ketone bodies
5-HIAA
NAD+
Enpp1
Enpp3
Quinolinate
Acetyl-CoA
Acat1,2
Acat3
Acetoacetyl-CoA
Citrate
Csl
Aco2
Isocitrate
Idh2
α-Ketoglutarate
Ogdh
Succinyl-CoA
Suclg1
Succinate
Sdha
Fumarate
Fhl
Malate
Mdh1,2
Oxaloacetate
Kreb's Cycle
Mut
L-Methylmalonyl-CoA
Aldh5a1
Succinate Semialdehyde
Abat
GABA
Tat
Mcce
D-Methylmalonyl-CoA
Pcca, b
Propionyl-CoA
Valine
Isoleucine
Carnosine
Cndp1
Cndp2
Carns1
β Alanine
Methyl-B12
Adenosyl-B12

Figure 6