Human metabolic profiles are stably controlled by genetic and environmental variation

George Nicholson, Mattias Rantalainen, Anthony D Maher, Jia V Li, Daniel Malmodin, Kourosh R Ahmadi, Johan H Faber, Ingileif B Hallgrimsdottir, Amy Barrett, Henrik Toft, Maria Krestyaninova, Juris Viksna, Sudeshna Guha Neogi, Marc-Emmanuel Dumas, Ugis Sarkans, The MolPAGE Consortium, Bernard W Silverman, Peter Donnelly, Jeremy K Nicholson, Maxine Allen, Krina T Zondervan, John C Lindon, Tim D Spector, Mark I McCarthy, Elaine Holmes, Dorrit Baunsgaard, Chris C Holmes

Corresponding author: George Nicholson, University of Oxford

Review timeline:

Submission date: 24 December 2010
Editorial Decision: 08 February 2011
Revision received: 16 May 2011
Editorial Decision: 06 June 2011
Revision received: 28 June 2011
Accepted: 08 July 2011

Transaction Report:

(Note: With the exception of the correction of typographical or spelling errors that could be a source of ambiguity, letters and reports are not edited. The original formatting of letters and referee reports may not be reflected in this compilation.)

1st Editorial Decision 08 February 2011

Thank you again for submitting your work to Molecular Systems Biology. We have now heard back from the three referees whom we asked to evaluate your manuscript. As you will see from the reports below, the referees find the topic of your study of potential interest. However, they raise substantial concerns on your work, which should be convincingly addressed in a revision of this study. Thus, both reviewer #2 and #3 raise serious concerns with regard to the statistical analysis (ie. need to compare with existing methodologies) and some of the conclusions of the study.

On a more editorial level, please note our policies with regard to the "availability of published material, data and software" (http://www.nature.com/msb/authors). We would thus kindly ask you to provide in supplementary information the dataset(s) used in this study in a format that would allow others to reproduce the essential aspects of the analysis and reuse/compare/integrate your data in other studies (see also "Summary recommendations for standardization and reporting of metabolic analyses" <http://www.nature.com/nbt/journal/v23/n7/full/nbt0705-833.html>).

*** PLEASE NOTE *** As part of the EMBO Publications transparent editorial process initiative (see our Editorial at http://www.nature.com/msb/journal/v6/n1/full/msb201072.html), Molecular Systems Biology will publish online a Review Process File to accompany accepted manuscripts. When preparing your letter of response, please be aware that in the event of acceptance, your cover...
letter/point-by-point document will be included as part of this File, which will be available to the scientific community. More information about this initiative is available in our Instructions to Authors. If you have any questions about this initiative, please contact the editorial office msb@embo.org.

If you feel you can satisfactorily deal with these points and those listed by the referees, you may wish to submit a revised version of your manuscript. Please attach a covering letter giving details of the way in which you have handled each of the points raised by the referees. A revised manuscript will be once again subject to review and you probably understand that we can give you no guarantee at this stage that the eventual outcome will be favourable.

Yours sincerely,

Editor
Molecular Systems Biology

Reviewer #1 (Remarks to the Author):

This is an intriguing work with interesting novel aspects. The twin data set with double biofluid follow-up experimentation is novel and elegant (Table S3 is a nice summary). The manuscript is also well written and rather nicely and clearly visualised. It has some important (new/novel) messages incorporated and would certainly be of high interest in the area of metabonomics and its applications. I am not personally that keen on the rather extensive spectral issues incorporated (like Fig. 1) but the authors have in my opinion included enough quantitative metabolic discussion (like Fig. 2 which biologically is much more relevant than Fig. 1) and analysis to make the key point for the future studies (in my opinion the sometimes overly complicated spectral approach could have been replaced in much more simple terms via purely quantitative metabolic data - however, for the final message I don't think this really matters, ie., the conclusions would be the same). Of particular practical importance and value for the field is Figure 3 (and the statistical background of it) - hopefully this figure will awaken many people in the field to realise that it really is "thousands of individuals" we have to study in real epidemiological studies ("metabonomics" or not). I don't have any particular criticism that would need further consideration; I am highly in favour of this work.

Reviewer #2 (Remarks to the Author):

TITLE:
Human metabolic profiles are stably controlled by genetic and environmental variation.

AUTHORS:
Nicholson et al.

MSID:
MSB-10-2511R

GENERAL
The authors present the results of analyses of an impressive body of metabolomics data generated on the basis of two important biofluids obtained from twins. The authors exhibit creativity in their approach for the analysis of the data in this context. The manuscript is extensive and the biological interpretation of the findings is thorough and interesting. However the manuscript needs considerable improvement. Following my conclusion I will present my detailed comments.

CONCLUSION
In my opinion, in its current form the manuscript is unsuitable for publication.

DETAILED COMMENTS
The advantages of the presented statistical method for variance component estimation over existing
methods are not apparent from the paper. The authors should indicate how their method compares to existing methods (e.g., structural equation modeling as described in Martin et al Nature 1997:17, 387-392 and in Neale and Cardon 1992 as referenced by the authors) and why the presented method is to be preferred over these existing methods. Also, the authors should provide proof why separation of genetic and nongenetic sources of familial variation using existing methods is not possible on the basis of the data as presented in the manuscript. It is not clear whether every variance component included in the model could be estimated with sufficient statistical significance for each metabolite (is the model valid for all metabolites?).

Metabolomics data potentially contain important information for analyses in the context of systems biology. The authors should consider to use methods that are able to acknowledge the relationships among different metabolites in a systems biology context.

The organization of the manuscript makes reading very difficult. Notably, the authors should improve the organization of the Supporting Information (SI) section, also in combination with the main text. For example, there is partial redundancy in the description of the statistical model for the variance decomposition in both the main text and in the SI section. The authors should be critical about which information is essential to include in the manuscript including the SI, and which information can be left out.

The style of writing is inconsistent throughout the manuscript. The authors should avoid the use of informal language (e.g., p. 6 line 5 "... as large as 50% say"). There is considerable redundancy in the results presented in different tables and figures. The graphical representation of the variance components estimates for each NMR peak separately, as in Figures 1 and S1, is not practically useful and is information overload. Instead, the authors should consider listing the results of variance components analyses for each separate metabolite only, for example in tables or as box plots.

Reviewer #3 (Remarks to the Author):

This reviewer agrees with the authors that the study substantially extends our current knowledge of the sources of variation in the human 1H NMR metabolome. The study addresses the contributions of different sources of variation to the urine and plasma metabolomes as measured by proton NMR. These sources are central to the design of any metabolomics study and include familial, individual-specific variation (including individually stable and dynamic variation), and non-biological variation. By fitting a variance components model, the authors quantify the contribution of relevant, "stable" variation (familial+individual stable) to each sample type and to each metabolite. Extending this, the authors perform power calculations for study design.

The study is technically sound, concise, and clear. The technical impact and utility of the study is likely high, however this reviewer is not convinced of its biological insights and interpretation of findings.

Major points:
(1) The authors state that their main finding is the "remarkably high" proportion of stable variation in the urine and plasma metabolomes. The Discussion then refers to the stability level of urine as "somewhat surprising" while later stating that a direct comparison with a previous study isn't easy because of differences in design and analysis. This reviewer agrees with the latter interpretation: the previous study assesses only 6 individuals across only 10 metabolites (how many of which overlap with the 27 annotated metabolites here is unclear). The authors also compared plasma familial variation with the heritability estimates of Shah et al, finding that all but one fell within 95% credible intervals. Given the 6 comparisons, one falling outside the intervals could have been by chance. This reviewer is unconvinced that their main findings are of substantial impact.

(2) The authors only present power calculations for urine samples in Figure 3. Since plasma has different stability and precision, the authors should show the same plots for plasma if they are significantly different. This reviewer doesn't have a feel for how different the plots would be given the stability/precision estimates and whether they could be (c) and (d) panels of Figure 3 but it should at least be addressed.
(3) The authors should elaborate on what sort of fasting the subjects undertook. Was this overnight fasting? Were the subjects asked to estimate how long they had fasted? This will be important when evaluating statements about the persistence of stable variation. For example, one would expect that a sampling interval of several months would span seasons, where an individual's metabolite inputs may change dramatically.

(4) Will the data be made freely available for other researchers (as per MSB policy)?

1st Revision - authors’ response 16 May 2011

Details of changes made in response to referees

General note: for clarity in the context of renumbering of Figures and Tables between versions of the paper, we will refer to the Figure and Table numbers of the initially submitted manuscript as Figure #[old] and Table #[old].

Reviewer 1

We thank the reviewer for their careful comments. In response:

1.1
"I am not personally that keen on the rather extensive spectral issues incorporated (like Fig. 1)."

Reply to 1.1

We have removed Figure 1[old] and Figure S1[old] from the manuscript.

Reviewer 2

We thank the reviewer for their careful comments. We have grouped the points raised into sections so as to help in our response.

2.1
2.1.A. "The advantages of the presented statistical method for variance component estimation over existing methods are not apparent from the paper. The authors should indicate how their method compares to existing methods (e.g., structural equation modeling as described in Martin et al Nature 1997;17, 387-392 and in Neale and Cardon 1992 as referenced by the authors) and why the presented method is to be preferred over these existing methods."

2.1.B. "Also, the authors should provide proof why separation of genetic and nongenetic sources of familial variation using existing methods is not possible on the basis of the data as presented in the manuscript."

2.1.C. "It is not clear whether every variance component included in the model could be estimated with sufficient statistical significance for each metabolite (is the model valid for all metabolites?)."

Reply to 2.1

Reply to 2.1.A. We have now incorporated a new section in Materials and Methods entitled "Statistical Model for Twin Data," (pages 12-13) in which we related our mixed-modelling methods to conventional structural equation modelling of twin data (the two methods are based on the same covariance structure, but differ in their approach to dealing with parameter non-identifiability). The section also contains reasons why we preferred our approach to the SEM approach in the context of the current study.

Reply to 2.1.B. We have now incorporated a section in SI Text entitled "Sample Size for Heritability Estimation," (page 1 of SI Text) in which we described a simulation study demonstrating the unsuitably low precision of estimators of heritability using the sample size of the current study.
Reply to 2.1.C. The model is valid for all metabolites, but, as you rightly point out, the precision with which we can estimate the variance components varies across metabolites. By providing credible intervals for our estimates of the variance components (Figure 1 and Table S1) the reader can see the range of statistical precision obtainable from the data for each metabolite. We have now incorporated a sentence (at page 5, paragraph 2) pointing the reader towards this important feature of the data: "There was variation across metabolites in the statistical precision with which variance components could be estimated. We quantified this aspect of the results by providing Bayesian credible intervals (BCIs) for the variance parameters of each metabolite (Figure 1 and Table S1)."

2.2. "Metabolomics data potentially contain important information for analyses in the context of systems biology. The authors should consider to use methods that are able to acknowledge the relationships among different metabolites in a systems biology context."

Reply to 2.2 We have now incorporated new material into the manuscript in the form of a paragraph in the Discussion (page 9 paragraph 2), and a new table, Table S2. Specifically, we have used the MetaboAnalyst web server (Xia et al, 2009) to map 36 of the annotated metabolites from our study to KEGG pathways (Table S2), and looked for over-representation of highly familial or highly stable metabolites in each pathway (we did not find any instances of significant over-representation). We also included a discussion of potential future work on correlation-based approaches to the analysis of longitudinal twin data.

2.3. "The organization of the manuscript makes reading very difficult. Notably, the authors should improve the organization of the Supporting Information (SI) section, also in combination with the main text. For example, there is partial redundancy in the description of the statistical model for the variance decomposition in both the main text and in the SI section. The authors should be critical about which information is essential to include in the manuscript including the SI, and which information can be left out."

Reply to 2.3. We have now pruned and reorganized the SI section, and fully incorporated the description of the variance-component model into the main text. The model description is split over three sections, "Statistical Model for Twin Data," "Full Variance-Components Model," and "Robust Bayesian Implementation," appearing on pages 12-15. Furthermore, to facilitate interpretation of the notation, we have added Table II, which displays the correspondence between each variance parameter's mathematical notation and its textual description. These adjustments have improved the readability of the manuscript, and have removed the previously existing redundancy from the model description.

2.4
2.4.A. "The style of writing is inconsistent throughout the manuscript."
2.4.B. "The authors should avoid the use of informal language (e.g., p. 6 line 5 "... as large as 50% say")."
2.4.C. "There is considerable redundancy in the results presented in different tables and figures. The graphical representation of the variance components estimates for each NMR peak separately, as in Figures 1 and S1, is not practically useful and is information overload. Instead, the authors should consider listing the results of variance components analyses for each separate metabolite only, for example in tables or as box plots."

Reply to 2.4
Reply to 2.4.A. We have revised the style of the manuscript to be as homogeneous as possible. In particular, we have adapted the manuscript so that now the research description consistently employs the past tense.

Reply to 2.4.B. We went carefully through the manuscript removing instances of informal language. For example:

- Page 6 paragraph 1: replaced "and so might be as large as 50% say," with "which can be large (e.g. exceeding 50%), as the current study has demonstrated."
- Page 4, paragraph 2: replaced "we discovered a remarkably high proportion of stable variation" with "identification and quantification of a substantive proportion of stable variation."
- Page 4, paragraph 5: replaced "The proportion of familial variation is notably high in both biofluids, but somewhat higher in plasma (42% is the mean across all peaks) than in urine (30%)."
Aggregating the familial and individual environment effects reveals a high level of stable biological variation: inter-peak means of 60% (IQR: 51-72) and 47% (IQR: 35-60) for plasma and urine respectively," with "The proportion of familial variation was found to be substantive in both biofluids, and somewhat higher in plasma (42% is the mean across all peaks) than in urine (30%). Finally, we aggregated the familial and individual environment effects to estimate the total proportion of biological variation that was longitudinally stable. We found the inter-peak average percentage of stable variation to be 60% (IQR: 51-72) and 47% (IQR: 35-60) for plasma and urine respectively."

Page 6 paragraph 5: replaced "the high average level of stability we observe in urine metabolites (47%) is a somewhat surprising finding" with "we have quantified the relative contributions of stable and unstable sources to population variation in urine metabolite concentration, and identified a substantive average level of stability (47%)."

Reply to 2.4.C. In response to 2.4.C, we have removed Figure 1[old], Figure S1[old], Figure S2[old], and Table S2[old] from the manuscript. The description of results now has considerably less redundancy now than it did in the initially submitted version. Specifically, the revised paper's representation of the VC decomposition comprises: Figure 1 [visual representation of variance decomposition, with estimates of precision, for each annotated metabolite]; and Table S1 [table of variance decomposition, and estimates of precision, across all metabolite peaks analysed]. The only remaining redundancy is minor, and is to be found in the representation of metabolite variance decompositionsóthey are represented visually in Figure 1, and numerically in a subset of the rows in Table S1.

By removing Figure 1[old] and S1[old], we have also addressed reviewer 2's concerns on the impracticality and information overload of the spectral representation of results.

Reviewer 3

We thank the reviewer for their careful comments. We have grouped the points raised into sections so as to help in our response.

3.1
3.1.A. "The authors state that their main finding is the "remarkably high" proportion of stable variation in the urine and plasma metabolomes. The Discussion then refers to the stability level of urine as "somewhat surprising" while later stating that a direct comparison with a previous study isn't easy because of differences in design and analysis. This reviewer agrees with the latter interpretation: the previous study assesses only 6 individuals across only 10 metabolites (how many of which overlap with the 27 annotated metabolites here is unclear)."

3.1.B. "The authors also compared plasma familial variation with the heritability estimates of Shah et al, finding that all but one fell within 95% credible intervals. Given the 6 comparisons, one falling outside the intervals could have been by chance. This reviewer is unconvinced that their main findings are of substantial impact."

Reply to 3.1

Reply to 3.1.A. We thank the reviewer for pointing this out. We were being rather subjective in some of the statements highlighted, such as "remarkably high" and "somewhat surprising" (they were made in the context of us personally finding the results remarkable). We have replaced these statements with objective ones:

Page 4, paragraph 2: replaced "we discovered a remarkably high proportion of stable variation" with "[our] identification and quantification of a substantive proportion of stable variation."

Page 6 paragraph 5: replaced "the high average level of stability we observe in urine metabolites (47%) is a somewhat surprising finding" with "we have quantified the relative contributions of stable and unstable sources to population variation in urine metabolite concentration, and identified a substantive average level of stability (47%)."

In continuation of our reply to 3.1.A, we agree with reviewer 3 that a direct comparison with the results of (Saude et al, 2007) is not feasible due to fundamental differences in design and data analysis.

In the Discussion, we have re-written the paragraph spanning pages 7-8, so as to place our
results on urine metabolomic stability in the context of pre-existing work. In that paragraph, we summarize the work of (Saude et al, 2007), as this comprises the most comprehensive existing work on the longitudinal stability of the urine metabolome. We also list the metabolites from our study that overlap with those reported in (Saude et al, 2007). Saude et al demonstrated that urine metabolite concentrations were capable of substantive longitudinal instability, but did not quantify the proportion of population variability attributable to such short-term dynamic effects. This is why our quantitative analysis of stable variation is novel and important; previous work had flagged instances of urine metabolomic instability, yet had not yet quantified the contribution of unstable/stable variation to population variance in metabolite concentration. As we feel this point is important, we have touched on it in our overview at the start of the Discussion (page 6 final paragraph) as well as in the Discussion paragraph spanning pages 7-8.

Reply to 3.1.B. What we intended to convey was exactly as you say, and we apologize that it was not clear in the original version of the paper. We have made some clarifications on page 8 paragraph 2, specifically including the sentences: "Some of the metabolites in our study overlapped with those examined by Shah et al, and hence we were able to check the consistency of a number of our findings against pre-existing work" and "It is reassuring that our plasma familiality findings are consistent with previous work."

3.2
"The authors only present power calculations for urine samples in Figure 3. Since plasma has different stability and precision, the authors should show the same plots for plasma if they are significantly different. This reviewer doesn't have a feel for how different the plots would be given the stability/precision estimates and whether they could be (c) and (d) panels of Figure 3 but it should at least be addressed."

Reply to 3.2. We have incorporated an analogous plot for the plasma data (Figure S3). The sample sizes required for plasma-based studies are estimated to be very close to (if slightly larger than) the sample sizes for urine-based studies.

3.3
"The authors should elaborate on what sort of fasting the subjects undertook. Was this overnight fasting? Were the subjects asked to estimate how long they had fasted? This will be important when evaluating statements about the persistence of stable variation. For example, one would expect that a sampling interval of several months would span seasons, where an individual's metabolite inputs may change dramatically."

Reply to 3.3. We have inserted details of the fasting protocol in Materials and Methods section, "Recruitment and Sample Collection," on page 10, in paragraph 5. Twins making morning visits, usually at 10am, had fasted overnight, since midnight; twins making afternoon visits, usually at 2pm, had fasted since 6am. Participants were not asked to estimate how long they had fasted.

3.4
"Will the data be made freely available for other researchers (as per MSB policy)?"

Reply to 3.4. Yes. The data have been uploaded to an FTP server, from which they will be made freely available to other researchers. The server login details are:

Host: svilpaste.mii.lu.lv
Login: Moltwin_NMR
Password: Moltwin_NMR1
Path: /home/George/NMR

For each of the four data sets analysed in the current paper, the following data formats are available for download: (a) raw frequency domain spectral data; (b) preprocessed spectral data (denoised, baseline corrected and normalized); (c) extracted peak heights. Sample metadata are also available.

References
Saude E, Adamko D, Rowe B, Marrie T, Sykes B (2007) Variation of metabolites in normal human
urine. Metabolomics 3: 439-451

Xia J, Psychogios N, Young N, Wishart D (2009) MetaboAnalyst: a web server for metabolomic data analysis and interpretation. Nucleic Acids Res 37

2nd Editorial Decision 06 June 2011

Thank you again for submitting your revised work to Molecular Systems Biology. We have now heard back from the two referees who accepted to evaluate your revision. As you will see, the referees are now supportive. Referee #1 raises however a series of minor points which we would kindly ask you to address.

We would also ask you to include in Materials and Method the exact link to the datasets used in this study.

Finally, we would kindly ask you to prepare supplementary material according to our instructions to authors.

Thank you for submitting this paper to Molecular Systems Biology.

Yours sincerely,

Section Editor
Molecular Systems Biology

referee reports

Reviewer #2 (Remarks to the Author):

In my opinion, the manuscript has improved considerably, and all issues raised on the basis of the previous submission have been addressed accurately by the authors. However, still some minor issues remain that in my opinion should be resolved to make the manuscript suitable for publication. Below I have listed these points, together with a number of points that could be regarded 'nitpicking'.

Minor issues:
1.) At various instances in the text the authors make statements about the relevance of their findings for the design of 1H-NMR based biomarker detection studies. It is not clear, however, to what type of study design these findings pertain: do the authors recommend repeated sampling, or focusing at longitudinally stable metabolites? Do we need a genetically informative sample (e.g., twins) for this? To which type of study design do their power calculations (revealing that sample sizes of 'a few thousand' should be sufficient) pertain?

2.) On pages 7 and further of the main document, the authors have amended their discussion of existing literature on the subject; in particular a study by Saude et al is now discussed in detail but the description is somewhat unclear. "This work has focused..." "← change to "This previous work has focused..."? The sentences "Saude et al reported results...in the concentrations of some urine metabolites" are somewhat unclear: did Saude et al leave out the results for the other four metabolites? The authors suggest that for five metabolites, a comparison is possible between the results of Saude et al and the results of the current study, however this comparison is not made. "Saude et al observed some instances..." "← in some individuals/ in some metabolites? "...one to two times the standard deviation of the entire population,..." "← at a single time point? "Against this background, an important goal..." "← of Saude et al? Or of the current study?"
3.) On page 9, the authors state that Table S2 shows the KEGG pathway mapping alongside the familial and stable variance component estimates of the current paper; however these variance component estimates are not shown in Table S2.

4.) Sample Sizes for MWASs, p.5:
"Metabolite concentrations reflect...given a study with sufficient sample size" <--- This sentence needs improvement. I agree that metabolite concentrations reflect both "molecular" (genetic?) and environmental risk factors on the one hand, and that on the other hand they provide relatively strong disease associations that are readily detectable and interpretable. I do not, however, see the direct connection between the two as the authors suggest.

5.) Discussion, p.7 "...the current study's design did not address the dynamics of those metabolites that varied diurnally about a stationary long-term mean" <--- This sentence needs improvement, as metabolites that vary diurnally about a stationary long-term mean are not interesting as biomarkers as well (long-term mean should change for a useful biomarker?)

6.) Statistical Model for Twin Data, p.12 "The structural-equation model (SEM)...and a different form for DZ twins" <--- This suggests that in SEM the phenotypic data are changed due to the structural equation model, which is not the case. Please rewrite, e.g. "The structural-equation model (SEM) for the classical twin study (e.g., Rijsdijk and Sham, 2002) provides a model for the covariance structures in phenotypic data obtained from MZ and DZ twin pairs."

7.) Full Variance-Components Model, p.14 "The familial variance...combined effects of genetics and common environment" <--- Is it true that it was assumed that these variance components were stable across the sampling period (this is particularly important for the common environment)?

8.) Table I, p. 23 Please indicate that this table lists variance proportions as percentages of total variance

With regard to the Supporting Information:
9.) SI Text, "Sample Size for Heritability Estimation" (p.1), "...our simulations are indicative of usefully precise estimation taking place (Figure S5)" <--- should refer to Figure S4 instead?
Model Validation: Comparison with Alternative Method
p.2 "Figure S4 compares estimates...the two methods of analysis" <--- should refer to Figure S5 instead?

10.) SI Text p.1, "Model Validation: Comparison with Alternative Method": The aim of this analysis is not quite clear: do the authors want to show that the Bayesian method is a viable alternative for mapping of the data to the standard normal distribution? Also the authors should indicate in what sense this alternative method differs from the method they use in the paper.

11.) Table S3:
Variation/lifestyle shared/non-shared by twin --> Variation/lifestyle shared/non-shared by twins of the same pair (family).
From descriptions of "Individual visit" and "Common visit", it is not clear whether these factors (also) are a source of variation between the two visits, i.e. whether they are a source of "inter-visit" variation.

Nitpicking:

-Results
Variation Landscape of the 1H NMR Metabolome, p.4:
Unclear sentence "For each of 526 common peaks... (Table S1 and Materials and Methods)" <--- "present in >80% of spectra", was this 80% within a data set or over all four data sets?
"(Table S1 and Materials and Methods)", please describe what is given here

"The observation of a higher proportion...plasma metabolite concentrations" <--- "greater baseline variability", explain: baseline as in spectral baseline?
"broad peaks arising from proteins": in plasma?
Variance Decomposition for Annotated Metabolites, p.5:
"The biological variance decomposition...(the underlying numbers are in Table S1)" <-- The text suggests that Figure 1 and Table S1 are redundant, but Table S1 gives the numbers for all peaks (not only for those shown in Figure 1) and this could be emphasized in the text

p.6 "We took the estimated stable proportion...can contribute to (x,y) covariation" <-- what is meant here, did the authors take the estimated stable proportion of variation for ALL urine metabolites (as opposed to, e.g., the mean)

p.8 "Nonetheless, Illig et al's (2010) findings...substantial single-locus genetic control" <-- unclear why this sentence is given here

p.8 "Analyses of 1H NMR metabolic profiles...between geographic regions" <-- replace "discovered" by e.g. "revealed" (discovery is human action?))

p.9 "A previous twin study, (Sahai & Vogel, 1983), found this enzyme..." <-- found the activity of this enzyme?

-Materials and Methods
Preprocessing and Feature Extraction
p.12 The authors describe that in order to prevent the analyses of the plasma data to be dominated by a single metabolite, only one glucose peak in each plasma data set was retained. As the analyses were carried out for separate peaks and finally the results for only one peak per metabolite were listed in the results, why is this important? Only for normalization of the spectra?

Robust Bayesian Implementation
p.14 It should be explained that in the formula for the definition of q(), delta defines the mixture proportions

-Figure Legends
p.21 "a Bonferroni-corrected significance level of 0.05, corrected for 500 tests" <-- given 500 phenotypes (metabolites)?

p.22 "The probability distribution on p..." <-- The probability distribution on p (not shown)...p.22 "...the current paper's estimates of stable variation..." <-- "...the current paper's estimates of proportion of stable variation..."

p.22 References to "Bottom panel", "Main panel", "Left panel" are unclear, replace by references to axis annotation e.g. "Bottom panel (annotated "Vpq(=|cor(x,y)|)")); and "Main panel (annotated "...")"

p.22 "both on logarithmic scale" <-- please transfer this to the description of the respective panels

-SI Text
Sample Size for Heritability Estimation
p.1: "In order to provide evidence...(as in the current study)" <-- replace "as in the current study" by "consistent with the composition of our dataset consisting of actual measured data"
p.1 "...our simulations are indicative of usefully precise estimation taking place (Figure S5)" <-- "...our simulations suggest that in our study indeed we should be able to estimate these two variance components with useful precision (Figure S4)" (also note the probably incorrect current figure numbering)

-SI Figure Legends
It should be indicated in the figure legends that the meaning of the "peak identifiers" is given in Table S1.
Figure titles are missing for Figures S2, S4 and S5.

-SI Table Legends
p.4 Table S2: More explanation is desired: the text in the main document, for example, suggests that in this table also familial and stable variance components are listed.
Reviewer #3 (Remarks to the Author):

The authors have addressed this referee's concerns.

2nd Revision - authors' response 28 June 2011

We are pleased to submit a revised version of our manuscript. We hope to have addressed the further minor issues raised by reviewer #2. The changes that we have made are detailed over the following pages.

Thank you once again for your time spent in consideration of our manuscript.

Reviewer #2

We are extremely grateful to Reviewer #2 for their time and effort in creating this extensive list of constructive comments.

Minor issues

1.) At various instances in the text the authors make statements about the relevance of their findings for the design of 1H-NMR based biomarker detection studies. It is not clear, however, to what type of study design these findings pertain: do the authors recommend repeated sampling, or focusing at longitudinally stable metabolites? Do we need a genetically informative sample (e.g., twins) for this? To which type of study design do their power calculations (revealing that sample sizes of 'a few thousand' should be sufficient) pertain?

Response to 1.) We have now added the following clarifying text at the end of the section "Sample Sizes for MWAS" at page 6, paragraph 3:

"The sample-size calculations are applicable to molecular epidemiological studies (not necessarily involving twins) in which the underlying disease model is assumed to be one where persistent over- or under-expression of an individual's baseline molecular level, relative to that of the general population, is associated with an increase or decrease in disease susceptibility relative to the background disease prevalence. We further assume that each participant donates a sample at a single time point. In this situation, variation due to longitudinal instability will reduce the precision in the estimate of the true baseline level and hence affect power to detect systematic differences between baseline measurements in cases versus controls. Studies with repeated longitudinal sampling of individuals could estimate the within-individual baseline level with greater precision, by averaging over the longitudinal variation, and so could increase power to detect disease associations with the same number of participants, but at increased numbers of samples and assays."

2.) On pages 7 and further of the main document, the authors have amended their discussion of existing literature on the subject; in particular a study by Saude et al is now discussed in detail but the description is somewhat unclear: "This work has focused..." change to "This previous work has focused..." The sentences "Saude et al reported results...in the concentrations of some urine metabolites" are somewhat unclear: did Saude et al leave out the results for the other four metabolites? The authors suggest that for five metabolites, a comparison is possible between the results of Saude et al and the results of the current study, however this comparison is not made. "Saude et al observed some instances..." in some individuals/ in some metabolites? "...one to two times the standard deviation of the entire population..." at a single time point? "Against this background, an important goal..." of Saude et al? Or of the current study?

Response to 2.) We have now revised this section and hope to have addressed all of the useful comments made above. The revised text (on page 8, paragraph 2) reads as follows:
"This previous work has focused on low-dimensional subspaces of the metabolome defined by pattern recognition methods (Bollard et al, 2005; Lenz et al, 2003), or on a restricted subset of metabolites, as did Saude et al (2007), who measured daily concentrations of 10 urine metabolites in six subjects over 30 days. Saude et al reported results for six randomly selected metabolites (they omitted results for four of the 10 metabolites). Of these, five are annotated and analysed in the current study: alanine (54%), citrate (76%), creatine (70%), hippurate (57%), and lactate (35%); parenthesized percentages are our estimates of the stable proportion of biological variation. We are unable to make a direct quantitative comparison between our results and those of Saude et al due to fundamental differences between the two studies in design and data analysis. Instead, we describe how our results develop knowledge of longitudinal stability of urine metabolites against the background of Saude et al's study. Saude et al reported some instances of within-individual longitudinal fluctuations (specifically, for citrate and tyrosine in a subset of individuals) that were of the same magnitude as one to two times the inter-individual standard deviation (i.e. the standard deviation, across individuals in the population, of the within-individual baseline mean concentration). They thus demonstrated the existence of substantive within-individual longitudinal variation (relative to population variation) in the concentrations of some urine metabolites in some individuals. Against this background created by the results of Saude et al, an important next goal was to quantify the relative contributions of stable and unstable variation to population variation in urine metabolite concentrations. Our research has done this, providing a formal and comprehensive treatment of longitudinal variation in the urine and plasma 1H NMR metabolomes. In contrast to previous work, we have explicitly modelled and estimated the proportional contribution of longitudinally fluctuating variation to population variance in metabolite concentration. We have demonstrated the importance of such results to the design and interpretation of MWASs."

3.) On page 9, the authors state that Table S2 shows the KEGG pathway mapping alongside the familial and stable variance component estimates of the current paper; however these variance component estimates are not shown in Table S2.

Response to 3.) We have now corrected this text (page 9, paragraph 5) to refer to Table S2, without the incorrect suggestion that the Table contains estimates of variance components.

4.) Sample Sizes for MWASs, p.5: "Metabolite concentrations reflect...given a study with sufficient sample size" <-- This sentence needs improvement. I agree that metabolite concentrations reflect both "molecular" (genetic?) and environmental risk factors on the one hand, and that on the other hand they provide relatively strong disease associations that are readily detectable and interpretable. I do not, however, see the direct connection between the two as the authors suggest.

Response to 4.) We have now changed this paragraph (page 5, paragraph 4) to read:

"The MWAS has emerged as an interesting 'top-down' approach for the characterization of disease aetiology (Chadeau-Hyam et al, 2010; Nicholson et al, 2008). Physiological concentrations of metabolites reflect both genetic and environmental risk factors, and can thus offer a relatively comprehensive and accurate assessment of complex-disease susceptibility, compared to molecular markers that are mechanistically closer to the genome (e.g. mRNA-transcript or protein levels). We examined the implications of our findings for the effective design of an MWAS in search of such disease-susceptibility metabolite biomarkers."

5.) Discussion, p.7 ".the current study's design did not address the dynamics of those metabolites that varied diurnally about a stationary long-term mean" <-- This sentence needs improvement, as metabolites that vary diurnally about a stationary long-term mean are not interesting as biomarkers as well (long-term mean should change for a useful biomarker?)

Response to 5.) We have changed the text (paragraph spanning pages 6-7) to read:

"...the current study's design did not address the dynamics of those metabolites that varied diurnally about a relatively stable baseline."
6.) Statistical Model for Twin Data, p.12 “The structural-equation model (SEM)...and a different form for DZ twins” -- This suggests that in SEM the phenotypic data are changed due to the structural equation model, which is not the case. Please rewrite, e.g. "The structural-equation model (SEM) for the classical twin study (e.g., Rijsdijk and Sham, 2002) provides a model for the covariance structures in phenotypic data obtained from MZ and DZ twin pairs."

Response to 6.) We have re-written as suggested (page 13, paragraph 2):

"The structural-equation model (SEM) for the classical twin study (e.g., Rijsdijk & Sham, 2002) provides a model for the covariance structure in phenotypic data obtained from MZ and DZ twin pairs. The covariance matrix of the phenotype measurements, and , from a pair of MZ twins is..."

7.) Full Variance-Components Model, p.14 "The familial variance...combined effects of genetics and common environment" -- Is it true that it was assumed that these variance components were stable across the sampling period (this is particularly important for the common environment)?

Response to 7.) Yes, we do assume this. For a given twin pair with two visits, the random effects representing common-environment and genetic effects (i.e. the and the ) are the same at both visits. If there is variation between visits that is shared by both twins in a pair, then that variation will be captured by the common-visit variance component, whose random effects, the , are visit specific.

8.) Table I, p. 23 Please indicate that this table lists variance proportions as percentages of total variance

Response to 8.) We have changed the title of Table 1 to "Percentage decomposition of biological population variation: summary of results."

9.) SI Text, "Sample Size for Heritability Estimation" (p.1), "...our simulations are indicative of usefully precise estimation taking place (Figure S5)" -- should refer to Figure S4 instead?

Model Validation: Comparison with Alternative Method p.2 "Figure S4 compares estimates...the two methods of analysis" -- should refer to Figure S5 instead?

Response to 9.) We have now corrected the references to these Figures and double-checked Figure and Table references throughout the manuscript.

10.) SI Text p.1, "Model Validation: Comparison with Alternative Method": The aim of this analysis is not quite clear: do the authors want to show that the Bayesian method is a viable alternative for mapping of the data to the standard normal distribution? Also the authors should indicate in what sense this alternative method differs from the method they use in the paper.

Response to 10.) We have now changed this section (Supplementary Information, spans pages 3-4) in an effort to clarify the rationale for the methodological comparison, and the effective differences between the two methods:

"For the current study, we developed bespoke robust Bayesian methods to model the data more effectively than was possible with existing methods (we refer to this approach to analysis as Analysis A). For comparison, we also performed a parallel analysis (accessible using existing software), referred to as Analysis B, and conducted as follows:

1. The same preprocessing, feature extraction, and normalization methods as for the main analysis (analysis A) were applied.
2. In order to mitigate the effect of outliers, in analysis B the quantiles of the data at each peak were transformed to the corresponding quantiles of a standard Gaussian distribution.
3. The linear mixed model defined in Eq. S1 was fitted by restricted maximum likelihood (Searle et al, 2006) using the lmer() R function in the lme4 package (version 0.999375-32).
4. Proportions of variance were calculated by substituting the variance estimates from lmer() directly into the corresponding ratio formulas defined in Materials and Methods."
Our Bayesian robust method (Analysis A) had a number of advantages over Analysis B in the current context, in which artefactually extreme observations were present for some metabolites, and in which estimates of precision for ratios of variance components were desired. Analysis A automatically down-weighted extreme observations (whilst Analysis B dealt with outliers by mapping the data to the quantiles of a Gaussian). Analysis A also readily provided estimates of precision for variance components and ratios thereof (in the form of posterior Bayesian credible intervals).

Despite our preference for Analysis A, we compared its results with Analysis B in order to demonstrate that qualitatively similar results could be obtained using existing software (such as the lme4 R package). Figure S5 compares estimates of variance proportions between the robust Bayesian methodology described in Materials and Methods (Analysis A), and Analysis B. There is a high level of agreement between estimates of variance proportions across the two methods of analysis.

11.) Table S3: Variation/lifestyle shared/non-shared by twin --> Variation/lifestyle shared/non-shared by twins of the same pair (family).

From descriptions of "Individual visit" and "Common visit", it is not clear whether these factors (also) are a source of variation between the two visits, i.e. whether they are a source of "inter-visit" variation.

Response to 11.) We have made Table S3 more precise in its terminology, and added to the descriptions of the two visit effects to clarify that they model inter-visit variation:

Familiality
- Heritability (additive and dominant genetic effects)
- Common environment (developmental conditions shared by twins of the same pair from conception until end of childhood cohabitation; elements of long-term lifestyle that are shared by twins of the same pair)

Individual environment
- Elements of long-term lifestyle (diet, cultural and social factors) that are not shared by twins of the same pair

Individual visit
- Inter-visit variation that is not shared by twins of the same pair
- Sample treatment-induced effects that are not shared by the two samples donated by a twin pair at a visit (e.g. such variation might be created if one twin's blood sample were to be mishandled, whilst the other twin's sample was correctly handled)
- Short-term lifestyle that is not shared by twins of the same pair

Common visit
- Inter-visit variation that is shared by twins of the same pair
- Sample treatment-induced effects that are shared by the two samples donated by a twin pair at a visit (e.g. each twin's blood sample being mishandled might create such variation)
- Short-term lifestyle that is shared by twins of the same pair (e.g. common diet over previous 24 hours)

Nitpicking

12.) Results. Variation Landscape of the 1H NMR Metabolome. p.4:
Unclear sentence "For each of 526 common peaks... (Table S1 and Materials and Methods)" <-- "present in >80% of spectra", was this 80% within a data set or over all four data sets? "(Table S1 and Materials and Methods)", please describe what is given here

Response to 12.) We have changed the text to read:

"For each of 526 common peaks (a peak was defined to be common if it was present in >80% of spectra in its corresponding data set), we quantified its height as a proxy for area in each spectrum in its data set, and fitted a variance-components model to the resulting data (see Materials and Methods for methodological details; Table S1 shows peak-specific variance..."
13.) Results. Variation Landscape of the 1H NMR Metabolome, p.4:
"The observation of a higher proportion...plasma metabolite concentrations" <-- "greater baseline variability", explain: baseline as in spectral baseline?
"broad peaks arising from proteins": in plasma?

Response to 13.) We have changed the text to read:

"The observation of a higher proportion of non-biological variation in plasma relative to urine was partially attributable to there being more variation across spectra in the spectral baseline (caused by a collection of broad peaks in plasma spectra arising from proteins), as well as to the presence of less population variation in (homeostatically controlled) plasma metabolite concentrations."

14.) Results. Variance Decomposition for Annotated Metabolites, p.5:
"The biological variance decomposition...(the underlying numbers are in Table S1)" <-- The text suggests that Figure 1 and Table S1 are redundant, but Table S1 gives the numbers for all peaks (not only for those shown in Figure 1) and this could be emphasized in the text

Response to 14.) We have a pointer to this in the section entitled "Variation Landscape of the 1H NMR Metabolome" reading:

"...Table S1 shows peak-specific variance decompositions for all 526 common peaks."

We have also made this clearer in the section entitled "Variance Decomposition for Annotated Metabolites":

"The biological variance decomposition for each such representative peak is represented in Figure 1 (the underlying numbers are in a subset of the rows of Table S1)."

15.) Results. p.6 "We took the estimated stable proportion...can contribute to (x,y) covariation" <-- what is meant here, did the authors take the estimated stable proportion of variation for ALL urine metabolites (as opposed to, e.g., the mean)

Response to 15.) We have now clarified this statement in the Results text:

"We created a distribution for that quantified the stability of common 1H NMR-detectable urine metabolites. The probability distribution on was constructed using (for upper bounds) the current paper's estimates of stable variation for peaks in the urine data. Specifically, we defined the distribution on to be a non-weighted mixture of the set of uniform densities , where denotes the estimate of the stable proportion of total phenotypic variance for the th peak."

16.) p.8 "Nonetheless, Illig et al's (2010) findings...substantial single-locus genetic control" <-- unclear why this sentence is given here

Response to 16.) We have adapted the text slightly so that the argument makes more sense:

"An estimate of heritability or familiality draws on variation from a potentially large number of genetic loci. Contrastingly, Illig et al (2010) searched for single-locus genetic drivers of metabolite levels. They quantified the strength of association in a human population between serum metabolite concentration and genetic variation at each of many single-nucleotide polymorphisms spanning the genome (see also Gieger et al (2008)). They reported nine loci, each of which exhibited a significant, replicable association with either a metabolite's concentration or with a concentration ratio (i.e. the ratio of one metabolite's concentration to another's), with the loci explaining between 5.6%-36.3% of the observed variance in concentration ratios. The MS-based Biocrates platform used by these authors was largely non-overlapping with 1H NMR in the subset of
the metabolome it targeted (it targeted mostly amino acids and lipids). Some of the strongly familial 1H NMR-detectable metabolites of our study may also be driven substantively by single-locus variation.

17.) p.8 "Analyses of 1H NMR metabolic profiles...between geographic regions" <-- replace "discovered" by e.g. "revealed" (discovery is human action?)

Response to 17.) We have changed this as suggested.

18.) p.9 "A previous twin study, (Sahai & Vogel, 1983), found this enzyme..." <-- found the activity of this enzyme?

Response to 18.) Yes, the text should have read as suggested. It has now been changed.

19.) Materials and Methods, Preprocessing and Feature Extraction, p.12. The authors describe that in order to prevent the analyses of the plasma data to be dominated by a single metabolite, only one glucose peak in each plasma data set was retained. As the analyses were carried out for separate peaks and finally the results for only one peak per metabolite were listed in the results, why is this important? Only for normalization of the spectra?

Response to 19.) We have now clarified the sections of the analysis for which the filtering of multiple glucose peaks was important:

"In order to prevent the analysis of the plasma data from being dominated by a single metabolite, we retained just one representative glucose peak in each plasma data set (the parts of the analysis to which the glucose peak-omission is relevant are: the normalization of each of the three plasma data sets; the summary of variance-decomposition results for all metabolite peaks in Table 1; and the calculation of sample sizes for biomarker discovery presented in Figure S3)."

20.) Robust Bayesian Implementation, p.14 It should be explained that in the formula for the definition of \( q(.) \), delta defines the mixture proportions

Response to 20.) We have now amended this as suggested.

21.) Figure Legends. p.21 "a Bonferroni-corrected significance level of 0.05, corrected for 500 tests" <-- given 500 phenotypes (metabolites)?

Response to 21.) We have now clarified this to read: "...assuming that 500 metabolite peaks were tested for disease association)."

22.) Figure Legends. p.22 "The probability distribution on p..." <-- The probability distribution on p (not shown)

Response to 22.) We have made this suggested change.

23.) Figure Legends. p.22 "...the current paper's estimates of stable variation..." <-- "...the current paper's estimates of proportion of stable variation..."

Response to 23.) We have changed the text to read: "...the current paper's estimates of the stable proportion of variation for peaks in the urine data..."

24.) Figure Legends. p.22 References to "Bottom panel", "Main panel", "Left panel" are unclear, replace by references to axis annotation e.g. "Bottom panel (annotated "Ypq(=xcor(x,y))")"; and "Main panel (annotated ")"

Response to 24.) We have now included these clarifying references.

25.) p.22 "both on logarithmic scale" <-- please transfer this to the description of the respective panels
Response to 25.) We have adapted the two-panel descriptions accordingly.

26.) SI Text, Sample Size for Heritability Estimation section, p. 1: "In order to provide evidence...(as in the current study)" <-- replace "as in the current study" by "consistent with the composition of our dataset consisting of actual measured data"

Response to 26.) We have now made this suggested clarification.

27.) p.1 "...our simulations are indicative of useful as the precision estimation taking place (Figure S5)" <-- "...our simulations suggest that in our study indeed we should be able to estimate these two variance components with useful precision (Figure S4)" (also note the probably incorrect current figure numbering)

Response to 27.) We have made this improvement to the SI Text.

28.) SI Figure Legends. It should be indicated in the figure legends that the meaning of the "peak identifiers" is given in Table S1.

Response to 28.) We have clarified the legend to include "... (the meaning of each peak identifier is given in Table S1)."

29.) Figure titles are missing for Figures S2, S4 and S5.

Response to 29.) We have now entitled these Figures. Figure S2: "Choice of representative peak for each metabolite." Figure S4: "Sample size for heritability estimation results of our simulation study." Figure S5: "Qualitative insensitivity of results to choice of statistical method."

30.) p.4 Table S2: More explanation is desired: the text in the main document, for example, suggests that in this table also familial and stable variance components are listed.

Response to 30.) We have included more detail in Table S2's legend:

"Mapping of metabolites to KEGG pathways. We mapped 33 annotated metabolites to KEGG pathways (see Discussion for details). The metabolites' KEGG-compound IDs and HMDB IDs (Wishart et al, 2009) are shown in the correspondingly labelled columns. For each of several KEGG pathways (labelled by column), an 'X' in a row indicates that the metabolite in that row maps to the corresponding KEGG pathway."

The text in the main document previously incorrectly suggested that Table S2 contained variance components (and has been corrected).

References

Bollard M, Stanley E, Lindon J, Nicholson J, Holmes E (2005) NMR-based metabolomic approaches for evaluating physiological influences on biofluid composition. NMR Biomed 18: 143-162

Chadeau-Hyam M, Ebbels T, Brown I, Chan Q, Stamler J, Huang CC, Daviglus M, Ueshima H, Zhao L, Holmes E, Nicholson J, Elliott P, De Iorio M (2010) Metabolic profiling and the metabolome-wide association study: significance level for biomarker identification. Journal of proteome research 9: 4620-4627

Gieger C, Geistlinger L, Altmaier E, HrabÈ de Angelis M, Kronenberg F, Meitinger T, Mewes H-W, Wichmann HE, Weinberger K, Adamski J, Illig T, Suhre K (2008) Genetics meets metabolomics: a genome-wide association study of metabolite profiles in human serum. PLoS Genetics 4: e1000282

Illig T, Gieger C, Zhai G, Romisch-Margl W, Wang-Sattler R, Prehn C, Altmaier E, Kastenmuller
G, Kato B, Mewes H-W, Meitinger T, de Angelis M, Kronenberg F, Soranzo N, Wichmann HE, Spector T, Adamski J, Suhré K (2010) A genome-wide perspective of genetic variation in human metabolism. Nat Genet 42: 137-141

Lenz EM, Bright J, Wilson ID, Morgan SR, Nash AF (2003) A 1H NMR-based metabonomic study of urine and plasma samples obtained from healthy human subjects. J Pharm Biomed Anal 33: 1103-1115

Nicholson J, Holmes E, Elliott P (2008) The Metabolome-Wide Association Study: A New Look at Human Disease Risk Factors. Journal of proteome research 7: 3637-3638

Rijsdijk F, Sham P (2002) Analytic approaches to twin data using structural equation models. Briefings in bioinformatics 3: 119-133

Saude E, Adamko D, Rowe B, Marrie T, Sykes B (2007) Variation of metabolites in normal human urine. Metabolomics 3: 439-451

Searle SR, Casella G, McCulloch CE (2006) Variance components: Wiley.

Wishart D, Knox C, Guo AC, Eisner R, Young N, Gautam B, Hau D, Psychogios N, Dong E, Bouatra S, Mandal R, Sinelnikov I, Xia J, Jia L, Cruz J, Lim E, Sobsey C, Shrivastava S, Huang P, Liu P et al (2009) HMDB: a knowledgebase for the human metabolome. Nucleic Acids Res 37: D603-610

Acceptance letter 08 July 2011

Thank you again for sending us your revised manuscript. We are now satisfied with the modifications made and I am pleased to inform you that your paper has been accepted for publication.

Proofs will be forwarded to you within the next 2-3 weeks.

Thank you very much for submitting your work to Molecular Systems Biology.

Sincerely,

Editor
Molecular Systems Biology