Author's response to reviews

Title: Extensive Alterations of the Whole-Blood Transcriptome are associated with Body Mass Index: Results of an mRNA Profiling Study involving two large Population-based Cohorts

Authors:

Georg Homuth (georg.homuth@uni-greifswald.de)
Simone Wahl (simone.wahl@helmholtz-muenchen.de)
Christian Müller (christian_m@gmx.net)
Claudia Schurmann (claudia.schurmann@mssm.edu)
Ulrike Mäder (ulrike.maeder@uni-greifswald.de)
Stefan Blankenberg (s.blankenberg@uke.de)
Maren Carstensen (maren.carstensen@ddz.uni-duesseldorf.de)
Marcus Dörr (mdoerr@uni-greifswald.de)
Karlhans Endlich (karlhans.endlich@uni-greifswald.de)
Christian Englbrecht (che74@gmx.de)
Stephan B Felix (felix@uni-greifswald.de)
Christian Gieger (christian.gieger@helmholtz-muenchen.de)
Harald Grallert (harald.grallert@helmholtz-muenchen.de)
Christian Herder (Christian.Herder@ddz.uni-duesseldorf.de)
Thomas Illig (Illig.Thomas@mh-hannover.de)
Jochen Krupa (Jochen.Kruppa@med.uni-goettingen.de)
Carola S Marzi (carola.marzi@helmholtz-muenchen.de)
Julia Mayerle (mayerle@uni-greifswald.de)
Thomas Meitinger (Meitinger@helmholtz-muenchen.de)
Andres Metspalu (andres.metspalu@ut.ee)
Matthias Nauck (matthias.nauck@uni-greifswald.de)
Annette Peters (peters@helmholtz-muenchen.de)
Wolfgang Rathmann (rathmann@ddz.uni-duesseldorf.de)
Eva Reinmaa (eva.reinmaa@ut.ee)
Rainer Rettig (rettig@uni-greifswald.de)
Michael Roden (Michael.Roden@ddz.uni-duesseldorf.de)
Arne Schillert (arne.schillert@imbs.uni-luebeck.de)
Katharina Schramm (katharina.schramm@helmholtz-muenchen.de)
Leif Steil (steil@uni-greifswald.de)
Konstantin Strauch (strauch@helmholtz-muenchen.de)
Alexander Teumer (ateumer@uni-greifswald.de)
Henry Völzke (voelzke@uni-greifswald.de)
Henri Wallaschofski (henri.wallaschofski@hotmail.com)
Philipp S Wild (philipp.wild@unimedizin-mainz.de)
Andreas Ziegler (ziegler@imbs.uni-luebeck.de)
Uwe Völker (voelker@uni-greifswald.de)
Holger Prokisch (Prokisch@helmholtz-muenchen.de)
Tanja Zeller (t.zeller@uke.de)
Author’s response to reviews: see over
To

The Editors, BMC Medical Genomics

Dear Editor,

Please find enclosed our revised manuscript (MS: 1786618450166530) entitled "Extensive Alterations of the Whole-Blood Transcriptome are associated with Body Mass Index: Results of an mRNA Profiling Study involving two large Population-based Cohorts ".

Please also find enclosed our responses to the reviewers. We addressed all the concerns raised by the two reviewers and hope that the current version is now acceptable for publication in BMC Medical Genomics. We would like to emphasize that the two newly provided tables KEGG_pathways_WebGestalt and WikiPathways_pathways_WebGestalt are only meant for reviewing purposes and not to be part of the final publication.

Sincerely yours,

Georg Homuth, Ph. D.
Uwe Völker, Ph. D.

On behalf of the MetaXpress Consortium
OUR RESPONSES TO THE REVIEWERS COMMENTS:

REVIEWER 1 (André Scherag):

REVIEWER'S MAJOR COMPULSORY REVISIONS:

REVIEWER'S COMMENT:

Regarding the first part, the authors should also provide the results for all transcripts in the adjusted and unadjusted analysis (as supplement).

OUR ANSWER:

To address this point we added two new tables to Additional_file2 of the Supplement: The first table (Supplementary Table 31) provides our annotation of all 48,803 probes present on the HumanHT-12 v3 Beadchip array as well as the respective detection rate (based on Illumina’s detection p-value < 0.05) in SHIP-TREND and KORA F4. Probes with a detection rate > 50% in both cohorts were considered to indicate significant transcript levels and were included in further analyses (N=12,778). The second table (Supplementary Table 32) contains the meta-analysis results for all 12,778 probes with and without adjustment for HOMA-IR, as described in the Supplementary Table Overview.

REVIEWER'S COMMENT:

For the main text I suggest that the authors provide a table showing the unadjusted and unadjusted analysis results for the most recent GIANT meta-analysis results (Locke et al., 2015; all genes implied by the DEPICT method).
OUR ANSWER:

We thank the reviewer for this comment. Indeed, differential expression of the genes prioritized by the DEPICT method (according to Supplementary Table 24 in Locke et al., 2015) as causative candidate genes of the SNP-BMI associations detected in the GIANT consortium can be hypothesized to affect BMI. Even if tissues and organs like liver, adipose tissue, muscle or brain are certainly more relevant in the context of the control of body fat deposition and blood glucose control, the mRNA levels determined in whole blood might at least partially reflect these effects. Therefore, as suggested, we additionally analysed the BMI-transcript level associations for those 989 genes prioritized by DEPICT. Only 388 out of the 989 genes prioritized by DEPICT could be mapped to probes in our list of 12,778 probes indicating significant transcript levels, and 151 out of 388 genes mapped to probes that were significantly associated with BMI in our analysis. These results are now provided in Supplementary Table 33 of the Additional_file2, as described in the Supplementary Table Overview. In addition, we added the following sentences to the General results paragraph of the RESULTS AND DISCUSSION section: “Recently, the Data-driven Expression Prioritized Integration for Complex Traits (DEPICT) method was used to prioritize BMI-related genes [5]. Of the 989 genes prioritized by DEPICT as BMI-related, 388 exhibited significant mRNA levels in our analysis. For 151 of these transcripts, we could confirm association of the their expression levels with BMI (Supplementary Table S33).”

REVIEWER'S MINOR ESSENTIAL REVISIONS:

REVIEWER'S COMMENT:

*Regrettably large parts of the paper focus on gene-set enrichment analyses (GSEA) which...*
As requested, we repeated the pathway over-representation analysis using the web-based toolkit WebGestalt (bioinfo.vanderbilt.edu/webgestalt/) and the online database resources KEGG (genome.jp/kegg/) and WikiPathways (wikipathways.org) to validate robustness across platforms. For the attention of the reviewers and editors, significantly enriched pathways (FDR ≤ 0.05) for both databases are attached (Tables KEGG_pathways_WebGestalt and WikiPathways_pathways_WebGestalt). Two out of the three BMI-related gene expression signatures defined in the initial submission were also observed in KEGG and WikiPathways based on the enriched pathways (e.g. Ribosome, Oxidative phosphorylation, Insulin signaling pathway, Type II diabetes mellitus, mTOR signaling pathway in KEGG; Insulin Signaling, AMPK signaling, MAPK Cascade, MAPK signaling pathway, TOR signaling, Electron Transport Chain, Heme Biosynthesis, Cytoplasmic Ribosomal Proteins, Translation Factors, Oxidative phosphorylation in WikiPathways). This was not the case for the “Reduced protection against oxidative stress” signature, represented by the “NRF2-mediated oxidative stress response” pathway in the Ingenuity Pathway Analysis (IPA). Even if many target genes of the transcription factor NRF2 as well as genes encoding known regulators of the NRF2 system, the major defense system against oxidative stress in eukaryotes, exhibited significantly BMI-associated transcript
levels in our analysis (NFE2L2, GSTM2, MGST2, NQO2, SOD2, TXNRD1, CREBBP, JUN, JUNB, MAFG, MAFF…) KEGG and WikiPathway analyses did not reproduce this IPA finding. Presumably, this is due to a partially differing pathway definition in WikiPathways and the absence of this pathway in KEGG. In this context, it is also essential to mention that the publicly available resources for KEGG/WikiPathways were updated for the last time in 2011/2012 (!), whereas the IPA database is updated regularly (last update 2015-06-17), indicating that the results obtained with IPA might be superior compared to the other two approaches.

As suggested by the reviewer, we validated the robustness of our findings by randomly sampling gene-specific transcript sets of the same size as the set of transcripts significantly associated with BMI in our analysis (3,762) for 5,000 times, collecting enrichment statistics for each BMI-associated pathway and calculating the probability for each pathway to be enriched by chance. This analysis indicated that only one out of the top 25 significantly enriched pathways, namely PDGF Signaling, might represent a false positive result, suggesting that the three signatures described in our work, including “Reduced protection against oxidative stress”, represent robust findings. We now extended Table 2 by adding the corresponding corrected p-values. In the corresponding RESULTS AND DISCUSSION section we added the following sentence “Permutation analysis (for details see Supplementary Materials) indicated that over-representation of genes in all pathways except PDGF Signaling is indeed true.”. The methodology is described as follows in the section SUPPLEMENTARY METHODS: EVALUATING ROBUSTNESS OF INGENIUM PATHWAY ANALYSIS BY PERMUTATION STATISTICS in SUPPLEMENTARY MATERIALS of Additional File 1: “Permutation analysis was performed for evaluating robustness of Ingenuity Pathway Analysis (IPA) and identifying potentially false positive over-representations. All gene-pathway mappings from the IPA online database were downloaded. According to the number of significantly BMI
associated genes, 3,762 genes were randomly selected 5,000 times from the list of all genes expressed in whole-blood and mapped to IPA pathways. Only pathways with more than two BMI associated genes were considered in further analysis. For each pathway, over-representation Z scores, a permutation p-value and Benjamini-Hochberg corrected p-values were calculated as described by Zambon et al. [1].”

**REVIEWER'S COMMENT:**

Finally, I strongly recommend deleting any potential clinical implications (e.g. in the conclusions) derived from such kind of analyses.

**OUR ANSWER:**

We believe that we indicated a potential clinical implication in the ABSTRACT only. Thus, following the recommendation of the reviewer we shortened the last sentence of the ABSTRACT “BMI-associated negative transcriptional regulation of insulin signaling and oxidative stress management provide new insights into the pathogenesis of metabolic syndrome and T2D, and may thus help to identify new therapeutic targets.” to “BMI-associated negative transcriptional regulation of insulin signaling and oxidative stress management provide new insights into the pathogenesis of metabolic syndrome and T2D.”
REVIEWER 2 (Luke Pilling):

REVIEWER'S MAJOR COMPULSORY REVISIONS:

REVIEWER'S COMMENT:

On page 10/11 you mention additional adjustment for HOMA-IR – this is very interesting for this study as this indicates whether you are actually finding type-2 diabetes (T2D) results or obesity-specific signals. I feel these two sentences deserve more prominence and detail, e.g. “The corresponding R2 for effect size... were as high as 99%”. Please indicate the actual values and what effect (if any) this has on the number of obesity-associated genes and mention this in the discussion. This relates to the section on page 13 (fasting glucose is associated with BMI in the study), page 14, and in particular page 15 during the discussion about “The mRNA signature of attenuated insulin signalling” – if empirically measured insulin is available this could be further investigated. I am suggesting a small sensitivity analysis (not re-running the whole analysis) of these insulin-related genes to see if the effect is attenuated by adjusting for insulin levels or T2D status – this would at least suggest whether the effect is obesity-specific or a result of T2D (which is presumably more common in the obese participants). Either way it would be interesting and useful to report, particularly for the conclusions where emphasis is placed on the relationship between BMI, leptin, glucose, insulin and HOMA-IR.

OUR ANSWER:

First, we want to thank the reviewer for his constructive comments. As suggested, we expanded the corresponding paragraph in the METHODS section as follows: “The effect of an additional adjustment for homeostasis model assessment – insulin resistance (HOMA-IR) on the results was negligible. The corresponding R^2-values for effect size (beta), standard error, and -
log10 p-value of the meta-analysis were 99.56, 99.98, and 99.39%, respectively.” In the RESULTS AND DISCUSSION section, we added the following paragraph: “Adjustment for HOMA-IR resulted in the detection of 1836 and 2602 transcripts exhibiting positive and negative BMI-correlations, respectively, with an overlap of 90.2% and 94.1% to the non-HOMA-IR-adjusted analysis, demonstrating that associations of BMI with transcript levels were largely independent of HOMA-IR.”

As insulin resistance represents the most common clinical consequence of a pathologically increased BMI, in other words of overweight, gene expression signatures that are associated with BMI can be predicted to be also associated with measures of insulin resistance. Indeed, BMI and measures of insulin resistance are for their part strongly associated, which was demonstrated in a plethora of epidemiological studies as well as in the present study. Type-2 diabetes (T2D) represents only the final sequelae of a long pathophysiological process with preceding sub-clinical insulin resistance. Therefore, increased serum levels of insulin or/and glucose represent clearly better parameters to determine the extent of insulin resistance as compared to the diagnosis of manifest T2D. However, an even better measure of insulin resistance is the Homeostasis Model Assessment (HOMA)-IR because it takes serum glucose as well as insulin into account. Indeed, adjusting BMI-associated transcripts for HOMA-IR did not cause pronounced differences in the number of associated transcripts in our study.

Furthermore, we performed a sensitivity analysis using the 39 genes comprising the Attenuated Insulin Signaling signature. In the subset of the 12,778 probes detecting significant transcript levels, we found 57 probes mapping to the mRNAs specified by these 39 genes. Out of those 57 probes, 47 and 46 detected mRNA levels significantly associated with BMI in our main meta-analyses and in the meta-analyses after additional adjustment for HOMA-IR, respectively. In the SHIP-TREND sample individuals with a self-reported diagnosis of manifest T2D were
excluded, but this was not the case for the KORA F4 sample which included 203 T2D patients. Therefore, we also performed a look-up of the *Attenuated Insulin Signaling* signature transcripts in the KORA-specific results alone. Here, 20 out of the 47 probes significantly associated in the meta-analysis also demonstrated significant association after both i) adjustment for BMI only and ii) adjustment for BMI and HOMA-IR. Nineteen probes were common to both analyses and the two probes only found significant in either of both analyses (detecting *ATM* and *MAP2K1* transcripts) showed nominal association in the respective other analysis. When adjusting for the T2D status but not for HOMA-IR in KORA F4, 17 of the 20 probes remained significantly associated with BMI (BH FDR < 0.01). The effect direction of all 47 probes was consistent across the different analyses. Overall, the mRNA levels of the *Attenuated Insulin Signaling* signature genes were correlated with BMI with and without adjustment for HOMA-IR as the most useful and reliable measure of insulin resistance available. This suggests that indeed the inverse association between BMI and the transcript levels of the genes comprising this signature is not mediated by insulin resistance.

**REVIEWER'S MINOR ESSENTIAL REVISIONS:**

**REVIEWER'S COMMENT:**

*On page 9 the age of participants is noted - KORA (n=988, 62-81), SHIP (n=989, 20-81) – I wonder what significance this has on the results, as the participants are predominantly quite old by this point (mean 50 in SHIP and 70 in KORA, from Table 1). The relationship between obesity and mortality changes as we age, so this is an important point to mention in the discussion, at the very least, as the results (e.g. insulin resistance implications) may be different in a younger population (or not, but we cannot see from this study).*
OUR ANSWER:

The following paragraph was added to the CONCLUSIONS section: “It also has to be mentioned that with a mean age of 50 and 70 years in KORA and SHIP, respectively, these cohorts are relatively old and thus, we cannot exclude the possibility that at least partially differing results would be obtained with clearly younger individuals. As the relationship between obesity and mortality changes with increasing age, the identified gene expression signatures, in particular those related to insulin signaling and oxidative stress defense, might also be affected in a younger cohort.”

REVIEWER'S COMMENT:

On page 10 please include the exact method/package used for the meta-analysis “... A sample size-weighted z-score based meta-analysis ...”

OUR ANSWER:

The meta-analyses were performed with an in-house R-script using the metafor meta-analysis package. We added the following paragraph in the METHODS section: “A sample size-weighted z-score based meta-analysis was used to combine the gene expression data from both cohorts (n=1,977) using the metafor meta-analysis package for R, version 1.4-0 (http://www.jstatsoft.org/v36/i03/).”

REVIEWER'S COMMENT:

On page 12 please clearly explain what is meant by “... 82 genes (2.2%) exhibited inconsistent effect directions on the probe level”. Do you mean that for 82 genes there were
multiple probes on the array, significantly associated with BMI in different directions? If so, could you have a quick look to see if the probes map to different exons (so potentially different isoforms of the genes), as this could be very interesting.

OUR ANSWER:

Indeed, as the arrays used for our whole-blood transcriptome analysis represented Illumina arrays, the different probes are designed to target different exon-specific transcripts of individual genes. Therefore, we now added the following phrase to the RESULTS AND DISCUSSION section: “In the latter case, the different probes target different exons of the corresponding genes, indicating that individual mRNA isoforms specified by the same genes were inversely associated with BMI.”

REVIEWER'S DISCRETIONARY REVISIONS:

REVIEWER'S COMMENT:

On page 9 there is a possible typo “… after overnight fasting between 8:00 am and noon” – should this be 8 pm? – repeated again in the conclusion, so maybe the time is correct but the “overnight” is incorrect?

OUR ANSWER:

The two corresponding sentences were changed to: “Briefly, whole-blood samples were collected from participants of both studies between 8 a.m. and noon after overnight fasting and stored in PAXgene Blood RNA Tubes (BD).” and “As circadian rhythms in whole-blood gene expression patterns have been described [26], it has to be emphasized that in SHIP-TREND as well in KORA F4 all blood samples were collected between 8 a.m. and noon from fasting
individuals.”

REVIEWER'S COMMENT:

It is interesting to note that of the 3,762 genes associated with BMI only 396 were in one of the ‘top’ 25 pathways described in the discussion. I appreciate the difficulties in interpreting collections of genes that do not appear to have pathways enriched, but feel other pathway programs could have been tried, or at the very least this point mentioned in the discussion; that the mRNA signature of obesity includes the three major pathways discussed, but appears to be predominantly not from these pathways.

OUR ANSWER:

The following sentence was added to the CONCLUSIONS section: “Finally, we would like to emphasize the fact that the three extracted signatures are based on a subset of around 400 from the more than 3700 BMI-associated gene-specific transcripts annotated in IPA. This demonstrates that for most of the BMI-associated whole blood transcriptome, adequate physiological interpretation is still not available. Further hypotheses about the mechanisms underlying the observed associations might be generated by using more sophisticated bioinformatical approaches in future analyses.”