Integrating Metabolomics Profiling Measurements Across Multiple Biobanks

ARTICLE *in* ANALYTICAL CHEMISTRY · MARCH 2014

Impact Factor: 5.64 · DOI: 10.1021/ac404191a · Source: PubMed

10 AUTHORS, INCLUDING:

Adrie Dane
Leiden University
26 PUBLICATIONS **524** CITATIONS
SEE PROFILE

Theo Reijmers
University of Amsterdam
51 PUBLICATIONS **948** CITATIONS
SEE PROFILE

Rob J. Vreeken
Janssen Pharmaceutica
95 PUBLICATIONS **1,809** CITATIONS
SEE PROFILE

Thomas Hankemeier
Leiden University
182 PUBLICATIONS **4,147** CITATIONS
SEE PROFILE

analytical chemistry

Letter

pubs.acs.org/ac

Integrating Metabolomics Profiling Measurements Across Multiple Biobanks

A. D. Dane,[†,‡,#] M. M. W. B. Hendriks,[†,‡,#] T. H. Reijmers,[†,‡] A. C. Harms,[†,‡] J. Troost,[†,‡] R. J. Vreeken,[†,‡] D. I. Boomsma,[§] C. M. van Duijn,[∥] E. P. Slagboom,[⊥] and T. Hankemeier*[,†,‡]

[†]Division Analytical Biosciences, Leiden Academic Center for Drug Research, Einsteinweg 55, 2333CC Leiden, The Netherlands

[‡]Netherlands Metabolomics Centre, Einsteinweg 55, 2333CC Leiden, The Netherlands

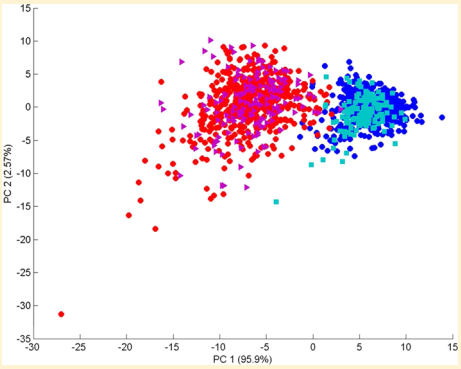
[§]Department of Biological Psychology, VU University Amsterdam, Amsterdam, The Netherlands

[∥]Department of Epidemiology, Erasmus University Medical School, Rotterdam, The Netherlands

[⊥]Leids Universitair Medisch Centrum, Albinusdreef 2, 2333 ZA Leiden, The Netherlands

Supporting Information

ABSTRACT: To optimize the quality of large scale mass-spectrometry based metabolomics data obtained from semiquantitative profiling measurements, it is important to use a strategy in which dedicated measurement designs are combined with a strict statistical quality control regime. This assures consistently high-quality results across measurements from individual studies, but semiquantitative data have been so far only comparable for samples measured *within* the same study. To enable comparability and integration of semiquantitative profiling data from different large scale studies over the time course of years, the measurement and quality control strategy has to be extended. We introduce a strategy to allow the integration of semiquantitative profiling data from different studies. We demonstrate that lipidomics data generated in samples from three different large biobanks acquired in the time course of 3 years can be effectively combined when using an appropriate measurement design and transfer model. This strategy paves the way toward an integrative usage of semiquantitative metabolomics data sets of multiple studies to validate biological findings in another study and/or to increase the statistical power for discovery of biomarkers or pathways by combining studies.



The application of metabolomics technology in large (epidemiological) studies is becoming common practice. The quest for biomarkers or pathways associated with certain diseases or traits has broadened from genomic to metabolomics measurements, with the metabolome being closer to the studied phenotype. In many cases, the metabolome is linked via genome-wide association studies (GWAS) to genetic markers, the metabolites acting as a quantitative intermediate phenotype. In addition, the metabolome is associated with phenotypic characteristics such as clinical end-point markers, disease subtypes, and the like. The number of subjects in such epidemiological or association studies is as a rule quite large; usually thousands of samples have to be measured to guarantee sufficient power to discover statistically significant associations.

Metabolomics data can be classified as either quantitative when absolute concentrations are provided or as semiquantitative when no absolute concentrations can be reported. The distinction is especially important for mass-spectrometry based metabolomics measurements, as when suitable reference compounds are unavailable no absolute concentrations can be obtained. It should be added that the sample preparation procedure has to be taken into account too as it determines whether metabolites are extracted in a quantitative manner.

Quantitative and semiquantitative metabolomics studies have demonstrated to provide the required quality to analyze large cohort studies. NMR-based metabolomics was applied in a large GWAS study on more than 8000 Finnish individuals,[1] while the 3000 samples of the Husermet project were analyzed with gas chromatography/mass spectrometry (GC/MS) and liquid chromatography−mass spectrometry (LC−MS) profiling platforms.[2] Commercial quantitative and semiquantitative LC−MS platforms have also been used in large scale metabolomics studies, such as the Metabolon platform in the GWAS analysis of the KORA and the TwinsUK cohorts.[3]

If these dedicated standardized platforms deliver quantitative metabolomics data, data obtained for a certain cohort in one study can directly be combined with metabolomics data from the same platform for another cohort and study. In most cases, however, the quantitative platforms comprise a limited set of metabolites. Contrary to the quantitative metabolomics measurements, semiquantitative metabolic mass spectrometry based platforms provide a broader coverage of the metabolome.

Received: December 20, 2013
Accepted: March 20, 2014
Published: March 20, 2014

dx.doi.org/10.1021/ac404191a | *Anal. Chem.* 2014, 86, 4110−4114

However, these metabolic profiling methods do not provide absolute quantitative data for all metabolites measured. With an appropriate measurement design and an adequate quality control and correction strategy, the within-study data are of high quality, with a high reproducibility. As the semi-quantitation is not absolute, additional between-studies variation is introduced. In this paper, we show that the use of an adequate within-studies measurement design and quality control procedure also serves as the basis of controlling and reducing the between-studies variability. Extending the measurement design with so-called transfer samples and adding a transfer procedure to the within-studies quality control and correction process enables matching and integration of semiquantitative profiling data between different studies and cohorts measured in different time periods.

In the Leiden Longevity Study (LLS[4]), a large biobanking cohort consisting of middle-aged offspring of nonagenarians and their spouses acting as controls, novel lipid markers were found to associate with familial longevity.[5] This observation encouraged the use of the same lipidomics profiling platform to enrich other existing Dutch biobanks. All was done as a subproject of the European Biobanking and BioMolecular Resources Research Infrastructure (BBMRI), where Dutch biobanks are collaborating to harmonize and further enrich their repositories. At this point in the project two additional large scale epidemiological cohorts are profiled, The Netherlands Twin Register (NTR[6]) and the Erasmus Rucphen Family (ERF[7]) cohort. For a short description of the cohorts see the Supporting Information.

Vaughan[8] presents a methodology for fusing metabolomics data of the same samples from different ultra performance liquid chromatography–mass spectrometry (UPLC–MS) platforms, but to our knowledge the combination of metabolomics profiling data of (different) samples from multiple large-scale studies using a proper study design and algorithms has not yet been described. In this letter we illustrate the integration of semiquantitative profiling data for different large-scale studies, measured over the time course of 3 years on the lipidomics profiling data generated in this project that was aimed at enriching existing biobanks with metabolomics data. The major challenge of combining profiling data measured over a large period of time is the introduction of variation due to changing experimental circumstances in sample preparation, data acquisition, and data processing. These circumstances are not always controllable, especially in the case when the combination of multiple studies was not intended from the start. We show that with a proper workflow, including the appropriate measurement design and quality control, comparability of studies can be achieved.

METHODS

Measurements, Processing, and Quality Control. The lipidomics LC–MS measurement platform used is a full-scan, global profiling method with targeted processing of the data (see ref 5 for a description of the platform). Relative quantification is obtained by using a limited set of internal standards that can roughly account for sample-to-sample variation caused by, e.g., differences in sample preparation and variation in the response. Only very few lipid standards are commercially available making absolute quantification of these lipid measurements using reference lipids unfeasible. This quantification problem is seen with most global profiling methods, even more so if the processing of the data is untargeted.

The measurement design follows a standard in-house strategy providing means for correction of between and within (run order) batch variation using quality control samples, study samples randomization and blocking (if necessary), validation and evaluation of measurements, preprocessing, and (relative) quantification procedures with analytical and technical replicates. Calibration samples are added to assess linear ranges, while blank samples are added to (eventually) correct for the (constant) background signal. Data on lipids were processed using MassHunter Quantitative Analysis Software.[9] Relative quantification is obtained following an in-house developed procedure, largely similar to the workflow described in ref 2. After background subtraction (the background obtained from blank samples), internal standard corrected relative lipid peak area ratios were calculated and corrected for intra- and interbatch variation using QC samples according to ref 10. A short description of this single study workflow is given in the Supporting Information. We intend to publish this workflow in more detail soon and a software tool will be made available to the community. In total, 15% of the study samples were prepared in duplicate and included and distributed over the various batches to assess the quality of the obtained corrected lipidomics data. Following the Standard Operating Procedures of the Biomedical Metabolomics Facility Leiden, data were visually inspected. Data were reported when the relative standard deviation (RSD) of QC corrected samples as well as the relative standard deviation of the QC corrected replicate samples was smaller than 20%. The in-house developed software tool automatically takes care of the whole correction and the quality control workflow, starting from the MassHunter integrated data. In the remainder of this paper we will use the term IS ratios when addressing the response values that are calculated from the integrated peak areas after the whole correction workflow.

Lipid profiles in citrate plasma from the three studies were measured in positive ion mode using the same protocol, apart from a change that was made in the extraction method between the first and the last two studies. The LLS lipidomics measurements were performed after two-phase liquid–liquid extraction of the citrate plasma samples. This favors the most nonpolar lipids, the triglycerides, since the extraction solvents are stronger. Since the two-phase liquid–liquid extraction was considered to be too time-consuming, for the other studies a change was made to the single-phase liquid–liquid extraction, using the more polar solvent isopropylalcohol of the citrate plasma samples. This gave better results for the more polar phospholipids but gave a poorer recovery for the triglycerides.

Combining Studies. Combining studies is challenging since multiple sources of variation introduce changes in the measured lipid abundances. These sources of variation include differences in sampling procedures, sample preparation procedures, instrumental settings, and environmental circumstances. The studies, that were to be combined, consisted of 27, 35, and 36 batches respectively, taking approximately 3–6 months per study to measure each. See the Supporting Information Table S1 for some statistics on the number of samples analyzed.

To be able to combine studies, the measurement design was extended by adding at the end of each batch so-called transfer samples. A transfer sample is a sample from an earlier measured study that is added to a newer study to make the first and

dx.doi.org/10.1021/ac404191a | Anal. Chem. 2014, 86, 4110–4114

second study comparable allowing integration of the two studies. In our strategy, approximately 6−7% of the total number of samples of the first study was added to the measurements of the second study. The added transfer samples from (older) studies are treated in the same way in the sample preparation, measurement, preprocessing, and correction procedures. In our case we have a third study to which both transfer samples of the first and of the second study were added. A large number (128 of the 149 and 163 transfer samples) of the first study (LLS) transfer samples were added both to the NTR and ERF measurements. The corrected data of all the transfer samples used in this manuscript can be obtained on request from the corresponding author.

Description of the Transfer Model. The relation between the concentration of a metabolite and its recorded response is assumed to be linear and is defined by the response factor. The assumption underlying the transfer model is that what changes between studies is this response factor. In preliminary calculations we confirmed that this assumption was true for blank corrected data and that we could do satisfactory data transfer without including an offset in the model. The transfer samples are used to calculate this change in response factor (using simple linear regression). Since metabolite abundance data are usually heteroscedastic, meaning that there is a nonconstant relation between the variance and the mean or expected abundance, the data were first transformed using square root transformation (a square root transformation is preferred over a logarithmic transformation since the first can handle zeros). A simple transfer model remains:

$$\sqrt{y_1} = \mathrm{drf}\sqrt{y_2} + e$$

Here y_1 and y_2 are the response values (IS ratios) of a specific metabolite in a certain transfer sample in, respectively, the reference study (ERF in this case) and the original study, e is the error representing the part of the variation in y_1 that cannot be described by y_2. In this formula drf represents the change in response factor and has to be estimated from the transfer samples, by means of linear regression. The estimated differential response factor is then applied on all samples by multiplying all responses with the factor drf squared to make the studies comparable:

$$\hat{y}_2 = \mathrm{drf}^2 y_2$$

In this formula $\hat{y}_2$ represents the adjusted response value (IS ratio) for a certain metabolite in a certain sample. This transfer model is fitted and the factor drf is estimated independently for each metabolite.

RESULTS

The three studies that were included, LLS, NTR, and ERF, consist of approximately 2500 samples each. In this paper we chose to take the most recently measured study, ERF as the reference study, meaning that all other studies are projected onto ERF.

In the Supporting Information variation in lipid measurements of transfer samples in the original study and the reference study are presented for all lipids. Validation of the transferred data was assessed by evaluating the reproducibility that was defined as the relative standard deviation (RSD_t) of all metabolites obtained from the transfer samples. The reproducibility after transfer is only slightly lowered as compared to the within study reproducibility for most of the lipids (for detailed information see the Supporting Information).

Figures 1 and 2 represent principal component analysis (PCA) scores plots of the three studies, both before and after

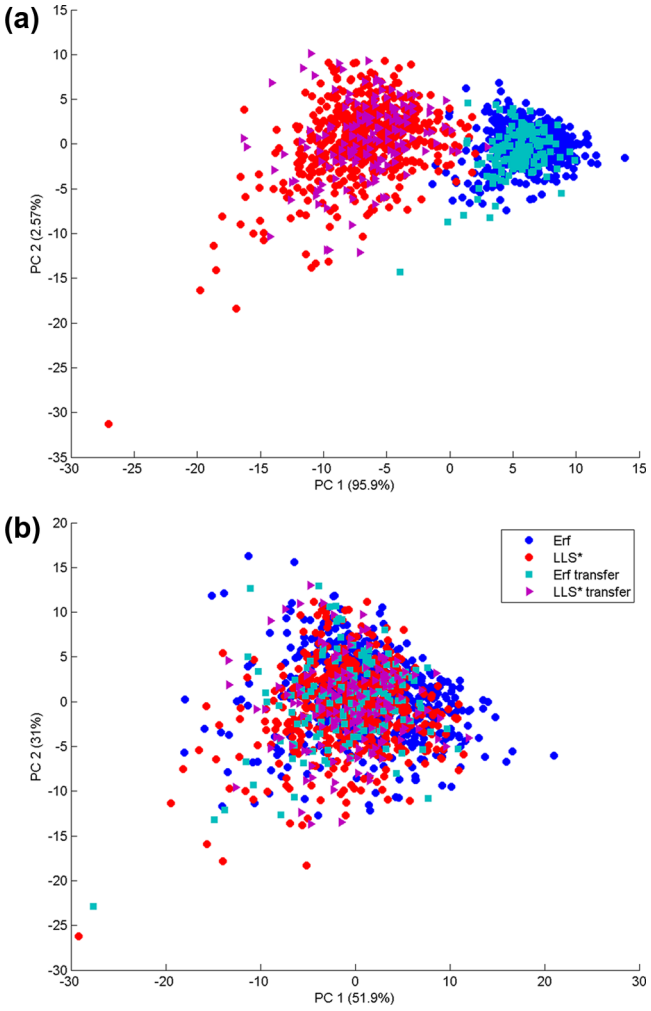


Figure 1. Principal component analysis of IS ratios of combined data sets (LLS with the ERF reference data set) illustrates the transfer process. (a) The top panel shows the spacing of samples before transfer, and (b) the bottom panel shows spacing of samples after the transfer.

application of the transfer models to the LLS and NTR data. Although this is a very global representation of the data, it illustrates that studies are more comparable, e.g., cover the same lipid space after the transfer.

The difference in the centers of the clouds of the NTR and LLS study samples as compared to the centers of the ERF samples represents the combined effect of the changes in response factor between studies (Figures 1 and 2, top panels). After application of the transfer model to the LLS data, the ERF and LLS transfer and study samples are nicely on top of each other (Figure 1, bottom panel).

In the second set of PCA scores plots (Figure 2), the study samples of ERF and NTR (after transfer, bottom panel) describe a slightly different part of the PCA scores space, as can be deduced from the observation that the blue and red clouds of scores are not fully overlapping. However, in the before-transfer scores plot, the green NTR transfer samples as

dx.doi.org/10.1021/ac404191a | Anal. Chem. 2014, 86, 4110−4114

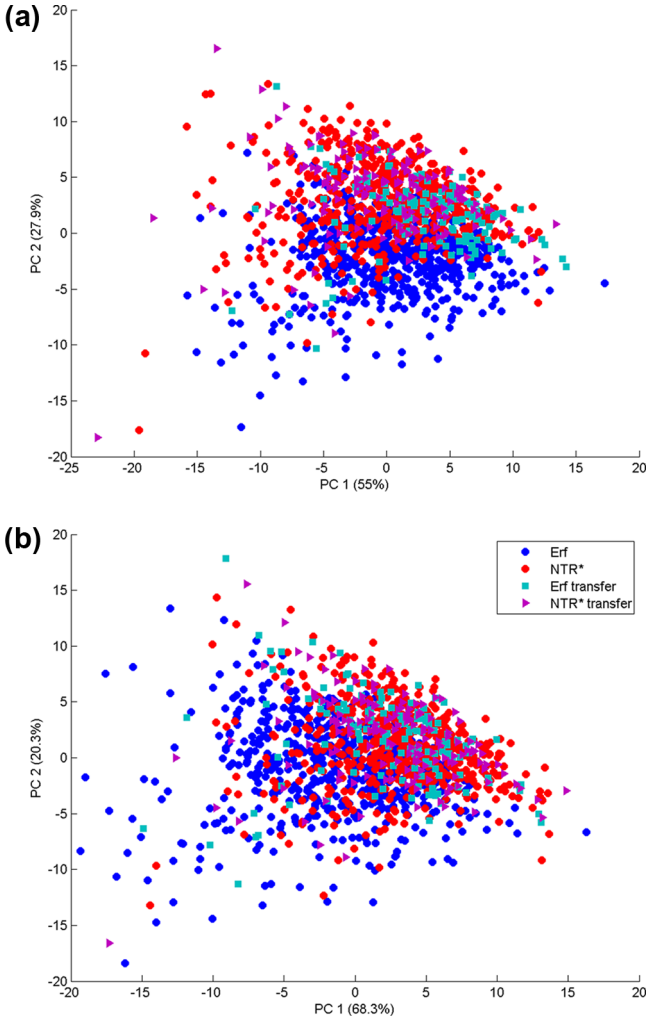


Figure 2. Principal component analysis of IS ratios of combined data sets (NTR with ERF reference data set) illustrates the transfer process. (a) The top panel shows the spacing of samples before transfer, and (b) the bottom panel shows the spacing of samples after the transfer.

measured in the ERF study are found in a noncentric subspace of the blue ERF sample-scores cloud. Pairwise PCA (PCA after per paired sample and per metabolite mean centering of the data) revealed that IS ratios of the NTR transfer samples alone, as measured in the ERF study, and the transferred IS ratios of the same samples as measured in the NTR study cover the same sample subspace (see the Supporting Information). These observations indicate that the transfer itself was adequate. Need for more complex transfer models, for instance including an offset, would have resulted in nonoverlapping clouds in the PCA plot. Therefore, we conclude that the variation caused by the difference in response factors is removed. The difference that is still observed (e.g., NTR and ERF sample clouds not fully overlapping in the PCA space) represents differences between the ERF and NTR samples themselves.

Two possible explanations can be given for this non-centralized positioning of the NTR transfer samples and the nonoverlapping scores clouds of the NTR and ERF samples after transfer of the first. A first explanation may be a difference of storage and sampling conditions of the biobanking samples. It is known that sample handling differences between clinics may introduce dissimilarities in measurement results in otherwise similar samples. In our case, for instance, we know that the NTR samples were stored at −30 °C, while the other studies' samples were stored at −80 °C. A second explanation may be found in the difference of the study populations of the two studies. Small differences in demographic and clinical characteristics exist in the three study populations (see Table S2 in the Supporting Information), as well as differences in time of sample taking, etc. In both cases, the differences should express themselves in nonlinear changes in lipid abundances or the need for an offset, since if the changes would be linear it can be corrected for by the transfer function. To investigate if the data can reveal which of the two explanations is most plausible a PLS-DA (see among many others ref 11) model was built to investigate which lipids are mostly responsible for the difference between the two studies, ERF on the one hand and NTR after transfer to the ERF domain on the other. The PLS-DA results could not reveal any systematic difference in lipid abundances between ERF and the transferred NTR data that could refer to sample acquisition and storage (see the Supporting Information).

Comparing the LLS transfer samples as measured during the LLS study, transferred to the ERF domain, with the responses of the same samples in the ERF study, shows that the difference in response induced by the difference in extraction method can fully be compensated for by the linear transfer function.

■ CONCLUDING REMARKS

This letter illustrates how the quality of a well-designed lipidomics data acquisition strategy can be assessed and illustrates how comparability and integration of studies measured over time can be obtained. The quality of the resultant lipidomics data is presented and varies among lipid classes. Apart from the TGs, the transfer RSDs are only a few percentages larger than the within study RSDs. For the TGs we see that some of the transfer RSDs are quite high, compared to the within study RSD. However, this is mainly due to only a few bad performing samples (see the Supporting Information for a visual representation of the data and a more thorough discussion).

For NMR-based and quantitative metabolomics (e.g., the Biocrates platform), combination of studies and comparability of studies is resolved by either using standards or by utilizing the actual concentrations. We showed that for a profiling method, where absolute concentrations are not being reported, combining studies results in metabolomics data of similar quality as the data of single studies.

Additionally, we showed that this procedure also works when combining studies was not foreseen from the start. This is especially valuable when interesting biomarkers are found in one study, and the aim is to validate these biomarkers in a new independent study (a so-called replicate cohort). Remarkably, even with a distinct change in extraction solvent such as occurred in this case, this procedure enables valid combination of data obtained from multiple studies.

We used 6−7% of the samples as transfer samples. These transfer samples are a random draw from all the study samples. Theoretically, drf could be estimated from a single transfer sample. This would, however, result in a very inaccurate estimate of drf. The minimal number of samples needed to get accurate estimates of drf is the subject of a future study. The principal requirement of the set of transfer samples is that they cover a considerable part of the total range of abundances of every metabolite measured. For large studies with many metabolites measured, this will in most cases imply that the

dx.doi.org/10.1021/ac404191a | Anal. Chem. 2014, 86, 4110−4114

number of transfer samples when using only a limited set of samples will have to be well chosen, to keep the number relatively low. The question then rises what the number of transfer samples at least should be. This, however, depends very much on the diversity and number of metabolites included and the heterogeneity of the study samples.

In this manuscript we describe the use of remeasuring a subset of samples from different studies to obtain correct transfer models. As far as we know, this approach was not described before in the literature. When the addition of this type of transfer samples is a problem, for instance due to limited sample volume, alternatives may be found in using nonrelated but similar transfer samples. Also in this case, the transfer samples have to fulfill the requirements that the abundances of the measured compounds cover the same concentration range as in the study samples. Reference samples such as the recently introduced NIST sample[12] may serve this purpose to some extent. However, as discussed before, to warrant high quality transfer models, a whole set of samples is needed. The use of such reference sample sets will allow for community-wide comparison of semiquantitative metabolomics data and obviously also for quantitative data.

■ ASSOCIATED CONTENT

Supporting Information

Additional information as noted in text. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: hankemeier@lacdr.leidenuniv.nl.

Author Contributions

#A.D.D. and M.M.W.B.H. are shared first authors of this work.

Notes

The authors declare no competing financial interest.

■ REFERENCES

(1) Kettunen, J.; Tukiainen, T.; Sarin, A.-P.; Ortega-Alonso, A.; Tikkanen, E.; Lyytikainen, L.-P.; Kangas, A. J.; Soininen, P.; Wurtz, P.; Silander, K.; Dick, E. M.; Rose, R. J.; et al. *Nat. Genet.* **2012**, *44*, 269−276.

(2) Dunn, W. B.; Broadhurst, D.; Begley, P.; Zelena, E.; Francis-McIntyre, S.; Anderson, N.; Brown, M.; Knowles, J. D.; et al. *Nat. Protoc.* **2011**, *6*, 1060−1083.

(3) Suhre, K.; Shin, S.-Y.; Petersen, A.-K.; Mohney, R. P.; Meredith, D.; Wagele, B.; Altmaier, E.; CARDIoGRAM; Deloukas, P.; Erdmann, J.; Grundberg, E.; Hammond, C. J.; et al. *Nature* **2011**, *477*, 54−60.

(4) Schoenmaker, M.; de Craen, A. J. M.; de Meijer, P. H. E. M.; et al. *Eur. J. Human Genet.* **2005**, *14*, 79−84.

(5) Gonzalez-Covarrubias, V.; Beekman, M.; Uh, H.-W; Dane, A.; Troost, J.; Paliukhovich, I.; van der Kloet, F. M.; Houwing-Duistermaat, J.; Vreeken, R. J.; Hankemeier, T.; Slagboom, E. P. *Aging Cell* **2013**, *12*, 426−434.

(6) Willemsen, G.; de Geus, E. J. C.; Bartels, M. *Twin Res. Human Genet.* **2010**, *13*, 231−245.

(7) Hofman, A.; van Duijn, C. M.; Franco, O. H. *Eur. J. Epidemiol.* **2011**, *26*, 657−686.

(8) Vaughan, A. A.; Dunn, W. B.; Allwood, J. W.; Wedge, D. C.; Blackhall, F. H.; Whetton, A. D.; Dive, C.; Goodacre, R. *Anal. Chem.* **2012**, *84*, 9848−9857.

(9) *MassHunter Quantitative Analysis*; Agilent Technologies: Santa Clara, CA, http://www.agilent.com.

(10) van der Kloet, F. M.; Bobeldijk, I.; Verheij, E. R.; Jellema, R. H. *J. Proteome Res.* **2009**, *8*, 5132−5141.

(11) van Velzen, E. J. J.; Westerhuis, J. A.; van Duynhoven, J. P. M.; van Dorsten, F. A.; Hoefsloot, H. C. J.; Jacob, D. M.; Smit, S.; Draijer, R.; Kroner, C. I.; Smilde, A. K. *J. Proteome Res.* **2008**, *7*, 4483−4491.

(12) Simón-Manso, Y.; Lowenthal, M. S.; Kilpatrick, L. E.; Sampson, M. L.; Telu, K. H.; Rudnick, P. A.; Mallard, W. G.; Bearden, D. W.; Schock, T. B.; Tchekhovskoi, D. V.; Blonder, N.; Yan, X.; Liang, Y.; Zheng, Y.; Wallace, W. E.; Neta, P.; Phinney, K. W.; Remaley, A. T.; Stein, S. E. *Anal. Chem.* **2013**, *85*, 11725−11731.