Genome **Medicine**

REVIEW

Genome-wide association studies with metabolomics

Jerzy Adamski[1,2]*

Abstract

Genome-wide association studies (GWAS) analyze the genetic component of a phenotype or the etiology of a disease. Despite the success of many GWAS, little progress has been made in uncovering the underlying mechanisms for many diseases. The use of metabolomics as a readout of molecular phenotypes has enabled the discovery of previously undetected associations between diseases and signaling and metabolic pathways. In addition, combining GWAS and metabolomic information allows the simultaneous analysis of the genetic and environmental impacts on homeostasis. Most success has been seen in metabolic diseases such as diabetes, obesity and dyslipidemia. Recently, associations between loci such as *FADS1*, *ELOVL2* or *SLC16A9* and lipid concentrations have been explained by GWAS with metabolomics. Combining GWAS with metabolomics (mGWAS) provides the robust and quantitative information required for the development of specific diagnostics and targeted drugs. This review discusses the limitations of GWAS and presents examples of how metabolomics can overcome these limitations with the focus on metabolic diseases.

Keywords complex disease, genetics, metabolomics, metabolic traits, pathway inference

Complex diseases: omics and genome-wide association studies

Common, severe human diseases such as cancer, diabetes, asthma, or mental and cardiovascular disorders have complex etiologies and complex mechanisms. To uncover the causal events leading to these diseases, information on the factors that challenge human health and the immediate responses to these challenges is needed. Yet, unfortunately, the dataset is never complete. In most cases, studies of humans are restricted to observations after a disease has occurred, except in clinical cases when individuals with particular diseases are treated or take part in randomized controlled intervention trials. Outside clinical trials, longitudinal studies (observational studies tracking the same individuals) that analyze phenotypes can also be undertaken. Both of these types of studies are hampered by unknown and uncontrolled exposure to the environment (such as differences in nutrition, medication, environmental endocrine disruptors and lifestyle) even in well-phenotyped cohorts (where weight, height and health status, for example, are known).

Cohorts can be analyzed for specific features such as genomic variance (variants in the DNA sequence) or metric parameters (concentrations or comparative levels) of RNA, proteins or metabolites. If the features analyzed and disease phenotypes coincide (and the frequency of coincidence is biostatistically valid), then it would be possible to identify the pathways involved. Therefore, a current approach to unveiling the etiology and mechanism of complex diseases is to employ sophisticated analysis methodologies (omics) that allow for the integration of multiple layers of molecular and organismal data. Data acquired with omics have already contributed considerably to the understanding of homeostasis in health and disease. Genome-wide association studies (GWAS), in particular, have contributed substantially to the field in the past 6 years [1]. This approach has identified numerous genetic loci that are associated with complex diseases. However, the number of genetic mechanisms that have been identified to explain complex diseases has not increased significantly [2].

In this review, I will highlight the current limitations of GWAS and how issues such as the large sample size required can be overcome by adding metabolomics information to these studies. I will explain the principles behind the combination of metabolomics and GWAS (mGWAS) and how together they can provide a more

*Correspondence: Adamski@helmholtz-muenchen.de
[1]Helmholtz Zentrum München, German Research Center for Environmental Health, Institute of Experimental Genetics, Genome Analysis Center, 85764 Neuherberg, Germany
Full list of author information is available at the end of the article

BioMed Central

powerful analysis. I conclude by exploring how mGWAS has been used to identify the metabolic pathways involved in metabolic diseases.

Aims and limitations of GWAS

GWAS analyze the association between common genetic variants and specific traits (phenotypes). The phenotypes originally included weight (or body mass index), height, blood pressure or frequency of a disease. More recently, specific traits in the transcriptome, proteome or metabolome have been included, and these are usually quantitative (for example, concentration). GWAS can also be used to explore whether common DNA variants are associated with complex diseases (for example, cancer or type 2 diabetes mellitus). The common variants might be single nucleotide polymorphisms (SNPs), copy number polymorphisms (CNPs), insertions/deletions (indels) or copy number variations (CNVs), but most GWAS employ SNPs [3]. At present, SNPs are used most frequently because of coverage of a large fraction of genome, through-put of assay, quality assurance and cost-effectiveness. Because the concept of GWAS is hypothesis-free, the analyses of GWAS are generally genetically unbiased, but they assume a genetic cause that might not be the most significant contributor.

In the past, candidate gene and pedigree analyses were very successful in the study of diseases of monogenetic origin: heritable dysregulation of certain metabolomic traits (inborn errors of metabolism) were among the first to be associated with specific genes [4]. However, these approaches are not useful in complex diseases because candidate regions contain too many genes or there are no groups of related individuals with a clear inheritance pattern of the disease phenotype. Inspired by the success of the Mendelian inheritance (genetic characteristics passed from the parent organism to offspring) approach, a great effort was undertaken to generate a human reference database of common genetic variant patterns based on a haplotype survey - the haplotype map (HapMap) [5]. This resource indeed improved, through linkage disequilibrium (LD) analyses, both the quality and the speed of GWAS, but it has not solved the major issue of study outcome. The common limitation of GWAS is that they do not provide mechanisms for disease; in other words, GWAS are unable to detect causal variants. Specifically, a GWAS provides information about an association between a variant (for example, SNP) and a disease, but the connection between a SNP and a gene is sometimes unclear. This is because annotated genes in the vicinity of a SNP are used in an attempt to explain the association functionally. However, proximity to a gene (without any functional analyses) should not be taken as the only sign that the identified gene contributes to a disease.

It should be further noted that the current analysis tools for SNPs do not include all possible variants, but rather only common ones with a major allele frequency greater than 0.01. SNPs with frequencies of less than 1% are not visible (or hardly discernible) in GWAS at present [3], and therefore some genetic contributions might remain undiscovered. So far, associations discovered by GWAS have had almost no relevance to clinical prognosis or treatment [6], although they might have contributed to risk stratification in the human population. However, common risk factors fail to explain the heritability of human disease [7]. For example, a heritability of 40% had been estimated for type 2 diabetes mellitus [8,9], but only 5 to 10% of the type 2 diabetes mellitus heritability can be explained by the more than 40 confirmed diabetes loci identified by GWAS [9,10].

Overcoming the limitations

There are several ways to improve GWAS performance. Instead of searching for a single locus, multiple independent DNA variants are being selected to identify those responsible for the occurrence of a disease [2]. Odds ratios could be more useful than *P*-values for the associations [6] in the interpretation of mechanisms and the design of replication or functional studies. This is especially true if highly significant (but spurious) associations are observed in a small number of samples, which might originate from a stratified population. The design of GWAS is also moving from tagging a single gene as a cause of disease to illuminating the pathway involved. This pathway might then be considered as a therapeutic target. In this way, GWAS comes back to its roots. The term 'post-GWAS' is used to describe GWAS-inspired experiments designed to study disease mechanisms. This usually involves exploration of expression levels of genes close to the associated variants, or knockout experiments in cells or animals [11]. In other words, post-GWAS analyses bring functional validation to associations [12].

Although omics approaches are powerful, they do not provide a complete dataset. Each omic technology provides a number of specific features (for example, transcript level fold change, protein identity or metabolite concentration, concentration ratios). At present, experimental datasets consisting of thousands of features unfortunately do not encompass all the features present *in vivo*. With incomplete data, only imperfect conclusions can be expected. However, the coverage of different omics features is expanding rapidly to overcome both genetic and phenotypic limitations of GWAS. As for the genetic aspects, progress in whole genome sequencing (for example, the 1000 Genomes Project [13,14]) is beginning to provide more in-depth analyses for less frequent (but still significant), and multiple, co-existing

disease loci. In addition, epigenetic features (for example, methylation, histone deacetylation) will soon be expanded in GWAS [15-17].

Improvements in the interpretation of phenotypes are likely to come from causal DNA variants showing significant and multiple associations with different omics data [11]. GWAS can be applied to intermediate phenotypes (including traits measured in the transcriptome, proteome or metabolome). The resulting associations can identify SNPs related to molecular traits and provide candidate loci for disease phenotypes related to such traits. Disease-associated alleles might modulate distinct traits such as transcript levels and splicing, thus acting on protein function, which can be monitored directly (for example, by proteomics) or by metabolite assays. This leads to the conclusion that another way to improve the outcomes of GWAS is the application of versatile and unbiased molecular phenotyping. The choice of molecular phenotyping approach will be driven by its quality regarding feature identification, coverage, throughput and robustness.

Metabolomic phenotyping for GWAS

Metabolomics deals with metabolites with molecular masses below 1,500 Da that reflect functional activities and transient effects, as well as endpoints of biological processes, that are determined by the sum of a person's or tissue's genetic features, regulation of gene expression, protein abundance and environmental influences. Ideally all metabolites will be detected by metabolomics. Metabolomics is a very useful tool that complements classical GWAS for several reasons. These include quantification of metabolites, unequivocal identification of metabolites, provision of longitudinal (time-resolved) dynamic datasets, high throughput (for example, 500 samples a week, with 200 metabolites for each sample), implementation of quality measures [18-21] and standardized reporting [22].

Enhancing classical GWAS for disease phenotypes with metabolomics is better than metabolomics alone for unequivocal description of individuals, stratification of test persons, and provision of multiparametric datasets with independent metabolites or identification of whole pathways affected (including co-dependent metabolites). It is also instrumental in quantitative trait locus (QTL) or metabolite quantitative trait locus (mQTL) analyses. In these studies quantitative traits (for example, weight or concentrations of specific metabolites) are linked to DNA stretches or genes. This information is important for assessing the extent of the genetic contribution to the observed changes in phenotypes.

A part of the metabolome could be computed from the genome [23], but the information would be static and hardly usable in biological systems except for annotation purposes. The time dynamics of the metabolome provides a means to identify the relative contributions of genes and environmental impact in complex diseases. Therefore, combining mGWAS expands the window of phenotypes that can be analyzed to multiple quantitative features, namely total metabolite concentrations.

Metabolomic approaches

Metabolomics mostly uses two major technological approaches: non-targeted metabolomics by nuclear magnetic resonance (NMR) or mass spectrometry (MS) [24]; and targeted metabolomics by MS [20,25].

Non-targeted metabolomics provides information on the simultaneous presence of many metabolites or features (for example, peaks or ion traces). Sample throughput may reach 100 samples a week on a single NMR spectrometer, gas chromatography-mass spectrometer (GC-MS) or liquid chromatography-tandem mass spectrometer (LC-MS/MS) [20,25]. The number of metabolites identified varies depending on the tissue and is usually between 300 (blood plasma) and 1,200 (urine) [26]. The major advantage of non-targeted metabolomics is its unbiased approach to the metabolome. The quantification is a limiting issue in non-targeted metabolomics as it provides the differences in the abundance of metabolites rather than absolute concentrations. *In silico* analyses (requiring access to public [27-30] or proprietary [31,32] reference databanks) are required to annotate the NMR peaks, LC peaks or ion traces to specific metabolites. Therefore, if a metabolite mass spectrum is not available in the databases, the annotation is not automatic but requires further steps. These may include analyses under different LC conditions, additional mass fragmentation or high-resolution (but slow) NMR experiments.

Targeted metabolomics work with a defined set of metabolites and can reach a very high throughput (for example, 1,000 samples per week on a single LC-MS/MS). The set might range from 10 to 200 metabolites in a specific (for example, only for lipids, prostaglandins, steroids or nucleotides) GC-MS or LC-MS/MS assay [33-37]. To cover more metabolites, samples are divided into aliquots and parallel assays are run under different conditions for GC- or LC-MS/MS. In each of the assays the analyzing apparatus is tuned for one or more specific chemical classes and stable isotope labeled standards are used to facilitate concentration determination. The major advantages of targeted metabolomics are the throughput and absolute quantification of metabolites.

Both approaches (that is, targeted and non-targeted) reveal a large degree of common metabolite coverage [38] or allow for quantitative comparisons of the same metabolites [21,39]. Metabolomics generates large-scale datasets, in the order of thousands of metabolites, which are easily included in bioinformatics processing [40,41].

GWAS with metabolomics traits

The outcome of GWAS depends very much on the sample size and the power of the study, which increases with the sample size. Some criticisms of GWAS have addressed this issue by questioning whether GWAS are theoretically big enough to overcome the threshold of *P*-values and associated odds ratios. Initial GWAS for a single metabolic trait (that is, plasma high-density lipoprotein (HDL) concentration [42]) were unable to detect the genetic component even with 100,000 samples. This indicates low genetic penetrance for this trait and suggests that another approach should be used to delineate the underlying mechanism. More recently, metabolomics was found to reveal valuable information when combined with GWAS. Studies with a much smaller sample size (284 individuals) but with a larger metabolic set (364 featured concentrations) demonstrated the advantage of GWAS combined with targeted metabolomics [34]. In this study the genetic variants were able to explain up to 28% of the metabolic ratio variance (that is, the presence or absence of a genetic variant coincided with up to 28% of changes in concentration ratios of metabolites from the same pathway). Moreover, the SNPs in metabolic genes were indeed functionally linked to specific metabolites converted by the enzymes, which are gene products of the associated genes.

In another study on the impact of genetics in human metabolism [35], involving 1,809 individuals but only 163 metabolic traits, followed by targeted metabolomics (LC-MS/MS), it was shown that in loci with previously known clinical relevance in dyslipidemia, obesity or diabetes (*FADS1*, *ELOVL2*, *ACADS*, *ACADM*, *ACADL*, *SPTLC3*, *ETFDH* and *SLC16A9*) the genetic variant is located in or near genes encoding enzymes or solute carriers whose functions match the associating metabolic traits. For example, variants in the promoter of *FADS1*, a gene that encodes a fatty acid desaturase, coincided with changes in the conversion rate of arachidonic acid. In this study, the metabolite concentration ratios were used as proxies for enzymatic reaction rates, and this yielded very robust statistical associations, with a very small *P*-value of 6.5×10^{-179} for *FADS1*. The loci explained up to 36% of the observed variance in metabolite concentrations [35]. In a recent fascinating study on the genetic impact on the human metabolome and its pharmaceutical implications with GWAS and non-targeted metabolomics (GC or LC-MS/MS), 25 genetic loci showed unusually high penetrance in a population of 1,768 individuals (replicated in another cohort of 1,052 individuals) and accounted for up to 60% of the difference in metabolite levels per allele copy. The study generated many new hypotheses for biomedical and pharmaceutical research [21] for indications such as cardiovascular and kidney disorders, type 2 diabetes, cancer, gout, venous thromboembolism and Crohn's disease.

A specific subset of the metabolome dealing with lipids termed lipidomics has provided important insights into how genetics contributes to modulated lipid levels. This area is of particular interest for cardiovascular disease research, as about 100 genetic loci (without causal explanation as yet) are associated with serum lipid concentrations [42]. Lipidomics increases the resolution of mGWAS over that with complex endpoints such as total serum lipids (for example, HDL only). For example, a NMR study showed that eight loci (*LIPC*, *CETP*, *PLTP*, *FADS1*, *-2*, and *-3*, *SORT1*, *GCKR*, *APOB*, *APOA1*) were associated with specific lipid subfractions (for example, chylomicrons, low-density lipoprotein (LDL), HDL), whereas only four loci (*CETP*, *SORT1*, *GCKR*, *APOA1*) were associated with serum total lipids [43]. GWAS have already enabled tracing of the impact of human ancestry on *n*-3 polyunsaturated fatty acid (PUFA) levels. These fatty acids are an important topic in nutritional science in trying to explain the impact of PUFA levels on immunological responses, cholesterol biosynthesis and cardiovascular disease [44-47]. It has been shown that the common variation in *n*-3 metabolic pathway genes and in the *GCKR* locus, which encodes the glucose kinase regulator protein, influences the levels of plasma phospholipid of *n*-3 PUFAs in populations of European ancestry, whereas in other ancestries (for example, African or Chinese) there is an impact on the influences in the *FADS1* locus [48]. This explains the mechanisms of different responses to diet in these populations. GWAS with NMR-based metabolomics can also be applied to large cohorts. An example is the analysis of 8,330 individuals in whom significant associations ($P < 2.31 \times 10^{-10}$) were identified at 31 loci, including 11 new loci for cardiometabolic disorders (among these most were allocated to the following genes: *SLC1A4*, *PPM1K*, *F12*, *DHDPSL*, *TAT*, *SLC2A4*, *SLC25A1*, *FCGR2B*, *FCGR2A*) [49]. A comparison of 95 known loci with 216 metabolite concentrations uncovered 30 new genetic or metabolic associations ($P < 5 \times 10^{-8}$) and provides insights into the underlying processes involved in the modulation of lipid levels [50].mGWAS can also be used in the assignment of new functions to genes. In metabolite quantitative trait locus (mQTL) analyses with non-targeted NMR-based metabolomics, a previously uncharacterized familial component of variation in metabolite levels, in addition to the heritability contribution from the corresponding mQTL effects, was discovered [38]. This study demonstrated that the so-far functionally unannotated genes *NAT8* and *PYROXD2* are new candidates for the mediation of changes in the metabolite levels of triethylamine and dimethylamine. Serum-based GWAS with LC/MS targeted metabolomics

has also contributed to field of function annotation: *SLC16A9*, *PLEKHH1* and *SYNE2* have been assigned to transport of acylcarnitine C5 and metabolism of phosphatidylcholine PCae36:5 and PCaa28:1, respectively [34,35].

mGWAS has recently contributed to knowledge on how to implement personalized medicine by analysis of the background of sexual dimorphism [51]. In 3,300 independent individuals 131 metabolite traits were quantified, and this revealed profound sex-specific associations in lipid and amino acid metabolism - for example, in the *CPS1* locus (carbamoyl-phosphate synthase 1; $P = 3.8 \times 10^{-10}$) for glycine. This study has important implications for strategies concerning the development of drugs for the treatment of dyslipidemia and their monitoring; an example would be statins, for which different predispositions should now be taken into account for women and men.

GWAS and metabolic pathway identification

By integrating genomics, metabolomics and complex disease data, we may be able to gain important information about the pathways that are involved in the development of complex diseases. These data are combined in systems biology [52] and systems epidemiology evaluations [53,54]. For example, SNP rs1260326 in *GCKR* lowers fasting glucose and triglyceride levels and reduces the risk of type 2 diabetes [55]. In a recent mGWAS [35], this locus was found to be associated with different ratios between phosphatidylcholines, thus providing new insights into the functional background of the original association. The polymorphism rs10830963 in the melatonin-receptor gene *MTNR1B* has been found to be associated with fasting glucose [56], and the same SNP associates with tryptophan:phenylalanine ratios in mGWAS [35]: this is noteworthy because phenylalanine is a precursor of melatonin. This may indicate a functional relationship between the phenylalanine-melatonin pathway and the regulation of glucose homeostasis. The third example is SNP rs964184 in the apolipoprotein cluster APOA1-APOC3-APOA4-APOA5, which associates strongly with blood triglyceride levels [57]. The same SNP associates with ratios between different phosphatidylcholines in mGWAS [35]: these are biochemically connected to triglycerides by only a few enzymatic reaction steps.

Conclusions

By combining metabolomics as a phenotyping tool with GWAS, the studies gain more precision, standardization, robustness and sensitivity. Published records worldwide illustrate the power of mGWAS. They provide new insights into the genetic mechanisms of diseases that is required for personalized medicine.

Author details
¹Helmholtz Zentrum München, German Research Center for Environmental Health, Institute of Experimental Genetics, Genome Analysis Center, 85764 Neuherberg, Germany. ²Lehrstuhl für Experimentelle Genetik, Technische Universität München, 85350 Freising-Weihenstephan, Germany.

References
1. Ku CS, Loy EY, Pawitan Y, Chia KS: The pursuit of genome-wide association studies: where are we now? *J Hum Genet* 2010, **55**:195-206.
2. Visscher PM, Brown MA, McCarthy MI, Yang J: Five years of GWAS discovery. *Am J Hum Genet* 2012, **90**:7-24.
3. Altshuler D, Daly MJ, Lander ES: Genetic mapping in human disease. *Science* 2008, **322**:881-888.
4. Palosaari PM, Kilponen JM, Hiltunen JK: Peroxisomal diseases. *Ann Med* 1992, **24**:163-166.
5. de Bakker PI, Yelensky R, Pe'er I, Gabriel SB, Daly MJ, Altshuler D: Efficiency and power in genetic association studies. *Nat Genet* 2005, **37**:1217-1223.
6. McClellan J, King MC: Genetic heterogeneity in human disease. *Cell* 2010, **141**:210-217.
7. Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorff LA, Hunter DJ, McCarthy MI, Ramos EM, Cardon LR, Chakravarti A, Cho JH, Guttmacher AE, Kong A, Kruglyak L, Mardis E, Rotimi CN, Slatkin M, Valle D, Whittemore AS, Boehnke M, Clark AG, Eichler EE, Gibson G, Haines JL, Mackay TF, McCarroll SA, Visscher PM: Finding the missing heritability of complex diseases. *Nature* 2009, **461**:747-753.
8. So HC, Yip BH, Sham PC: Estimating the total number of susceptibility variants underlying complex diseases from genome-wide association studies. *PLoS One* 2010, **5**:e13898.
9. Park KS: The search for genetic risk factors of type 2 diabetes mellitus. *Diabetes Metabol J* 2011, **35**:12-22.
10. McCarthy MI: Genomics, type 2 diabetes, and obesity. *N Engl J Med* 2010, **363**:2339-2350.
11. Freedman ML, Monteiro AN, Gayther SA, Coetzee GA, Risch A, Plass C, Casey G, De Biasi M, Carlson C, Duggan D, James M, Liu P, Tichelaar JW, Vikis HG, You M, Mills IG: Principles for the post-GWAS functional characterization of cancer risk loci. *Nat Genet* 2011, **43**:513-518.
12. Doucleff M: Select: GWAS gets functional. *Cell* 2010, **143**:177-178.
13. Via M, Gignoux C, Burchard EG: The 1000 Genomes Project: new opportunities for research and social challenges. *Genome Med* 2010, **2**:3.
14. Huang J, Ellinghaus D, Franke A, Howie B, Li Y: 1000 Genomes-based imputation identifies novel and refined associations for the Wellcome Trust Case Control Consortium phase 1 Data. *Eur J Hum Genet* 2012. doi: 10.1038/ejhg.2012.3
15. Rakyan VK, Down TA, Balding DJ, Beck S: Epigenome-wide association studies for common human diseases. *Nat Rev Genet* 2011, **12**:529-541.
16. Bataille V, Lens M, Spector TD: The use of the twin model to investigate the genetics and epigenetics of skin diseases with genomic, transcriptomic and methylation data. *J Eur Acad Dermatol Venereol* 2012, doi: 10.1111/j.1468-3083.2011.04444.x
17. Krueger F, Kreck B, Franke A, Andrews SR: DNA methylome analysis using short bisulfite sequencing data. *Nat Methods* 2012, **9**:145-151.

18. Griffin JL, Nicholls AW: **Metabolomics as a functional genomic tool for understanding lipid dysfunction in diabetes, obesity and related disorders.** *Pharmacogenomics* 2006, **7**:1095-1107.

19. Grebe SK, Singh RJ: **LC-MS/MS in the clinical laboratory - where to from here?** *Clin Biochem Rev* 2011, **32**:5-31.

20. Griffiths WJ, Koal T, Wang Y, Kohl M, Enot DP, Deigner HP: **Targeted metabolomics for biomarker discovery.** *Angew Chem Int Ed Engl* 2010, **49**:5426-5445.

21. Suhre K, Shin SY, Petersen AK, Mohney RP, Meredith D, Wagele B, Altmaier E, Deloukas P, Erdmann J, Grundberg E, Hammond CJ, de Angelis MH, Kastenmüller G, Köttgen A, Kronenberg F, Mangino M, Meisinger C, Meitinger T, Mewes HW, Milburn MV, Prehn C, Raffler J, Ried JS, Römisch-Margl W, Samani NJ, Small KS, Wichmann HE, Zhai G, Illig T, Spector TD, Adamski J, Soranzo N, Gieger C: **Human metabolic individuality in biomedical and pharmaceutical research.** *Nature* 2011, **477**:54-60.

22. Goodacre R, Broadhurst D, Smilde AK, Kristal BS, Baker JD, Beger RD, Bessant C, Connor SC, Capuani G, Craig A, Ebbels T, Kell DB, Manetti C, Newton J, Paternostro G, Somorjai R, Sjöström M, Trygg J, Wulfert F: **Proposed minimum reporting standards for data analysis in metabolomics.** *Metabolomics* 2007, **3**:231-241.

23. Weckwerth W: **Unpredictability of metabolism - the key role of metabolomics science in combination with next-generation genome sequencing.** *Anal Bioanal Chem* 2011, **400**:1967-1978.

24. Malet-Martino M, Holzgrabe U: **NMR techniques in biomedical and pharmaceutical analysis.** *J Pharm Biomed Anal* 2011, **55**:1-15.

25. Koal T, Deigner HP: **Challenges in mass spectrometry based targeted metabolomics.** *Curr Mol Med* 2010, **10**:216-226.

26. Sreekumar A, Poisson LM, Rajendiran TM, Khan AP, Cao Q, Yu J, Laxman B, Mehra R, Lonigro RJ, Li Y, Nyati MK, Ahsan A, Kalyana-Sundaram S, Han B, Cao X, Byun J, Omenn GS, Ghosh D, Pennathur S, Alexander DC, Berger A, Shuster JR, Wei JT, Varambally S, Beecher C, Chinnaiyan A: **Metabolomic profiles delineate potential role for sarcosine in prostate cancer progression.** *Nature* 2009, **457**:910-914.

27. Williams AJ: **Public chemical compound databases.** *Curr Opin Drug Discov Devel* 2008, **11**:393-404.

28. Babushok VI, Linstrom PJ, Reed JJ, Zenkevich IG, Brown RL, Mallard WG, Stein SE: **Development of a database of gas chromatographic retention properties of organic compounds.** *J Chromatogr* 2007, **1157**:414-421.

29. Wishart DS, Knox C, Guo AC, Eisner R, Young N, Gautam B, Hau DD, Psychogios N, Dong E, Bouatra S, Mandal R, Sinelnikov I, Xia J, Jia L, Cruz JA, Lim E, Sobsey CA, Shrivastava S, Huang P, Liu P, Fang L, Peng J, Fradette R, Cheng D, Tzur D, Clements M, Lewis A, De Souza A, Zuniga A, Dawe M, *et al.*: **HMDB: a knowledgebase for the human metabolome.** *Nucleic Acids Res* 2009, **37** (Database issue):D603-610.

30. Little JL, Williams AJ, Pshenichnov A, Tkachenko V: **Identification of "known unknowns" utilizing accurate mass data and ChemSpider.** *J Am Soc Mass Spectrom* 2012, **23**:179-185.

31. Evans AM, DeHaven CD, Barrett T, Mitchell M, Milgram E: **Integrated, nontargeted ultrahigh performance liquid chromatography/electrospray ionization tandem mass spectrometry platform for the identification and relative quantification of the small-molecule complement of biological systems.** *Anal Chem* 2009, **81**:6656-6667.

32. Lawton KA, Berger A, Mitchell M, Milgram KE, Evans AM, Guo L, Hanson RW, Kalhan SC, Ryals JA, Milburn MV: **Analysis of the adult human plasma metabolome.** *Pharmacogenomics* 2008, **9**:383-397.

33. Unterwurzacher I, Koal T, Bonn GK, Weinberger KM, Ramsay SL: **Rapid sample preparation and simultaneous quantitation of prostaglandins and lipoxygenase derived fatty acid metabolites by liquid chromatography-mass spectrometry from small sample volumes.** *Clin Chem Lab Med* 2008, **46**:1589-1597.

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

35. Illig T, Gieger C, Zhai G, Römisch-Margl W, Wang-Sattler R, Prehn C, Altmaier E, Kastenmüller G, Kato BS, Mewes HW, Meitinger T, de Angelis MH, Kronenberg F, Soranzo N, Wichmann HE, Spector TD, Adamski J, Suhre K: **A genome-wide perspective of genetic variation in human metabolism.** *Nat Genet* 2010, **42**:137-141.

36. Ceglarek U, Shackleton C, Stanczyk FZ, Adamski J: **Steroid profiling and analytics: going towards sterome.** *J Steroid Biochem Mol Biol* 2010, **121**:479-480.

37. Dunn WB, Broadhurst D, Begley P, Zelena E, Francis-McIntyre S, Anderson N, Brown M, Knowles JD, Halsall A, Haselden JN, Nicholls AW, Wilson ID, Kell DB, Goodacre R; Human Serum Metabolome (HUSERMET) Consortium: **Procedures for large-scale metabolic profiling of serum and plasma using gas chromatography and liquid chromatography coupled to mass spectrometry.** *Nat Protoc* 2011, **6**:1060-1083.

38. Nicholson G, Rantalainen M, Li JV, Maher AD, Malmodin D, Ahmadi KR, Faber JH, Barrett A, Min JL, Rayner NW, Toft H, Krestyaninova M, Viksna J, Neogi SG, Dumas ME, Sarkans U; MolPAGE Consortium, Donnelly P, Illig T, Adamski J, Suhre K, Allen M, Zondervan KT, Spector TD, Nicholson JK, Lindon JC, Baunsgaard D, Holmes E, McCarthy MI, Holmes CC: **A genome-wide metabolic QTL analysis in Europeans implicates two loci shaped by recent positive selection.** *PLoS Genet* 2011, **7**:e1002270.

39. Suhre K, Meisinger C, Doring A, Altmaier E, Belcredi P, Gieger C, Chang D, Milburn MV, Gall WE, Weinberger KM, Mewes HW, Hrabé de Angelis M, Wichmann HE, Kronenberg F, Adamski J, Illig T: **Metabolic footprint of diabetes: a multiplatform metabolomics study in an epidemiological setting.** *PLoS One* 2010, **5**:e13953.

40. Oresic M: **Metabolomics, a novel tool for studies of nutrition, metabolism and lipid dysfunction.** *Nutr Metab Cardiovasc Dis* 2009, **19**:816-824.

41. Krumsiek J, Suhre K, Illig T, Adamski J, Theis FJ: **Gaussian graphical modeling reconstructs pathway reactions from high-throughput metabolomics data.** *BMC Syst Biol* 2011, **5**:21.

42. Teslovich TM, Musunuru K, Smith AV, Edmondson AC, Stylianou IM, Koseki M, Pirruccello JP, Ripatti S, Chasman DI, Willer CJ, Johansen CT, Fouchier SW, Isaacs A, Peloso GM, Barbalic M, Ricketts SL, Bis JC, Aulchenko YS, Thorleifsson G, Feitosa MF, Chambers J, Orho-Melander M, Melander O, Johnson T, Li X, Guo X, Li M, Shin Cho Y, Jin Go M, Jin Kim Y, *et al*: **Biological, clinical and population relevance of 95 loci for blood lipids.** *Nature* 2010, **466**:707-713.

43. Petersen AK, Stark K, Musameh MD, Nelson CP, Römisch-Margl W, Kremer W, Raffler J, Krug S, Skurk T, Rist MJ, Daniel H, Hauner H, Adamski J, Tomaszewski M, Döring A, Peters A, Wichmann HE, Kaess BM, Kalbitzer HR, Huber F, Pfahlert V, Samani NJ, Kronenberg F, Dieplinger H, Illig T, Hengstenberg C, Suhre K, Gieger C, Kastenmüller G: **Genetic associations with lipoprotein subfractions provide information on their biological nature.** *Hum Mol Genet* 2012, **21**:1433-1443.

44. Moodley T, Vella C, Djahanbakhch O, Branford-White CJ, Crawford MA: **Arachidonic and docosahexaenoic acid deficits in preterm neonatal mononuclear cell membranes. Implications for the immune response at birth.** *Nutr Health* 2009, **20**:167-185.

45. Choi YS, Goto S, Ikeda I, Sugano M: **Effect of dietary n-3 polyunsaturated fatty acids on cholesterol synthesis and degradation in rats of different ages.** *Lipids* 1989, **24**:45-50.

46. Misra A, Khurana L, Isharwal S, Bhardwaj S: **South Asian diets and insulin resistance.** *Br J Nutr* 2009, **101**:465-473.

47. Singer P, Berger I, Moritz V, Forster D, Taube C: **N-6 and N-3 PUFA in liver lipids, thromboxane formation and blood pressure from SHR during diets supplemented with evening primrose, sunflowerseed or fish oil.** *Prostaglandins Leukot Essent Fatty Acids* 1990, **39**:207-211.

48. Lemaitre RN, Tanaka T, Tang W, Manichaikul A, Foy M, Kabagambe EK, Nettleton JA, King IB, Weng LC, Bhattacharya S, Bandinelli S, Bis JC, Rich SS, Jacobs DR Jr, Cherubini A, McKnight B, Liang S, Gu X, Rice K, Laurie CC, Lumley T, Browning BL, Psaty BM, Chen YD, Friedlander Y, Djousse L, Wu JH, Siscovick DS, Uitterlinden AG, *et al.*: **Genetic loci associated with plasma phospholipid n-3 fatty acids: a meta-analysis of genome-wide association studies from the CHARGE Consortium.** *PLoS Genet* 2011, **7**:e1002193.

49. Kettunen J, Tukiainen T, Sarin AP, Ortega-Alonso A, Tikkanen E, Lyytikäinen LP, Kangas AJ, Soininen P, Würtz P, Silander K, Dick DM, Rose RJ, Savolainen MJ, Viikari J, Kähönen M, Lehtimäki T, Pietiläinen KH, Inouye M, McCarthy MI, Jula A, Eriksson J, Raitakari OT, Salomaa V, Kaprio J, Järvelin MR, Peltonen L, Perola M, Freimer NB, Ala-Korpela M, Palotie A, Ripatti S: **Genome-wide association study identifies multiple loci influencing human serum metabolite levels.** *Nat Genet* 2012, **44**:269-276.

50. Tukiainen T, Kettunen J, Kangas AJ, Lyytikäinen LP, Soininen P, Sarin AP, Tikkanen E, O'Reilly PF, Savolainen MJ, Kaski K, Pouta A, Jula A, Lehtimäki T, Kähönen M, Viikari J, Taskinen MR, Jauhiainen M, Eriksson JG, Raitakari O, Salomaa V, Järvelin MR, Perola M, Palotie A, Ala-Korpela M, Ripatti S: **Detailed metabolic and genetic characterization reveals new associations for 30 known lipid loci.** *Hum Mol Genet* 2012, **21**:1444-1455.

51. Mittelstrass K, Ried JS, Yu Z, Krumsiek J, Gieger C, Prehn C, Roemisch-Margl W,

Polonikov A, Peters A, Theis FJ, Meitinger T, Kronenberg F, Weidinger S, Wichmann HE, Suhre K, Wang-Sattler R, Adamski J, Illig T: **Discovery of sexual dimorphisms in metabolic and genetic biomarkers.** *PLoS Genet* 2011, **7**:e1002215.

52. Hood L, Heath JR, Phelps ME, Lin B: **Systems biology and new technologies enable predictive and preventative medicine.** *Science* 2004, **306**:640-643.

53. Adourian A, Jennings E, Balasubramanian R, Hines WM, Damian D, Plasterer TN, Clish CB, Stroobant P, McBurney R, Verheij ER, Bobeldijk I, van der Greef J, Lindberg J, Kenne K, Andersson U, Hellmold H, Nilsson K, Salter H, Schuppe-Koistinen I: **Correlation network analysis for data integration and biomarker selection.** *Mol Biosyst* 2008, **4**:249-259.

54. Haring R, Wallaschofski H: **Diving through the "-omics": the case for deep phenotyping and systems epidemiology.** *OMICS* 2012. doi: 10.1089/omi.2011.0108

55. Vaxillaire M, Cavalcanti-Proenca C, Dechaume A, Tichet J, Marre M, Balkau B, Froguel P: **The common P446L polymorphism in GCKR inversely modulates fasting glucose and triglyceride levels and reduces type 2 diabetes risk in the DESIR prospective general French population.** *Diabetes* 2008, **57**:2253-2257.

56. Prokopenko I, Langenberg C, Florez JC, Saxena R, Soranzo N, Thorleifsson G, Loos RJ, Manning AK, Jackson AU, Aulchenko Y, Potter SC, Erdos MR, Sanna S, Hottenga JJ, Wheeler E, Kaakinen M, Lyssenko V, Chen WM, Ahmadi K, Beckmann JS, Bergman RN, Bochud M, Bonnycastle LL, Buchanan TA, Cao A, Cervino A, Coin L, Collins FS, Crisponi L, de Geus EJ, *et al*: **Variants in MTNR1B influence fasting glucose levels.** *Nat Genet* 2009, **41**:77-81.

57. Kathiresan S, Melander O, Guiducci C, Surti A, Burtt NP, Rieder MJ, Cooper GM, Roos C, Voight BF, Havulinna AS, Wahlstrand B, Hedner T, Corella D, Tai ES, Ordovas JM, Berglund G, Vartiainen E, Jousilahti P, Hedblad B, Taskinen MR, Newton-Cheh C, Salomaa V, Peltonen L, Groop L, Altshuler DM, Orho-Melander M: **Six new loci associated with blood low-density lipoprotein cholesterol, high-density lipoprotein cholesterol or triglycerides in humans.** *Nat Genet* 2008, **40**:189-197.