

RESEARCH HIGHLIGHT

Look, no hands! Spectral biomarkers from genetic association studies

Michael Inouye* and Gad Abraham

See related research article: <http://genomemedicine.com/content/5/2/13>

Abstract

Recent advances in our understanding of the genomics of the human metabolome have shed light on the pathways involved in metabolic and cardiovascular disease. Such studies crucially depend on the interpretation of complex molecular spectra. A recent study by Suhre and colleagues provides a way to identify potentially clinically relevant biomarkers without *a priori* information, such as reference spectra, thus aiding the discovery of additional spectral features and corresponding genomic loci associated with metabolism and disease.

Spectral genome-wide association studies as a tool for understanding pathogenesis

Before the widespread application of genome-wide association studies (GWASs) in the mid-2000s, techniques such as linkage analysis of families and candidate-gene studies had largely failed to identify robust and replicable loci associated with diseases that are common in the population. As judged by the criteria of replication, GWASs have been among the most successful epidemiological study designs to date, in no small measure due to large sample sizes, stringent quality control, simplicity of experimental design, and collaborative transparency between researchers. Yet identifying common disease loci, even when they explain a large proportion of heritability, only goes so far in advancing our understanding of pathogenesis.

Many GWASs employ a case-control design, where a set of individuals carrying the disease is compared with a set of non-diseased or population-based individuals. This is a useful strategy for finding loci associated with

disease; however, categorizing patients into two classes (such as disease/no-disease) ignores the biological intricacies of the disease at hand, and provides only a rough guide to the underlying etiology. To create more detailed and accurate models of pathogenesis, it is important to look in more detail at the potential intermediate phenotypes, for example, by measuring concentrations of cellular products and enzymes that underlie the processes of disease. The ready availability of the relevant tissues and accurate, high-throughput technology have allowed researchers to leverage metabolomic profiling to elucidate the genomics of one such class of intermediate phenotypes, namely metabolites, which play an important role in metabolic and cardiovascular diseases [1]. Recent GWASs of the metabolome have identified scores of loci associated with metabolites [2-5], some of which (both loci and metabolites) have been shown to be associated with disease. Furthermore, given known pathway relationships between metabolites and the high dimensionality of the phenotype data, researchers have begun using novel approaches such as phenotype ratios and multivariate analysis of phenotype networks [6] to increase statistical power and interpretation.

In this issue of *Genome Medicine*, Suhre and colleagues [7] side-step a fundamental challenge in previous GWASs, the decomposition of nuclear magnetic resonance (NMR) spectra into known metabolite concentrations, to expand the power of spectral association studies. In doing so, they present a novel method for identifying previously uncharacterized spectral features that may prove to be important biomarkers of disease.

Unbiased assessment of NMR spectra

A large contributing factor to the success of GWASs has been that they are relatively unbiased, in the sense that they assess marker variables that are roughly evenly drawn from across the genome rather than focusing only on specific loci or variants of interest. This lack of bias has enabled detection of previously unknown signals that would not have been found by methods such as candidate-gene studies. Analogously, the new study [7] shows the benefit of considering phenotypes in an

*Correspondence: minouye@unimelb.edu.au
Medical Systems Biology, Department of Pathology and Department of Microbiology & Immunology, The University of Melbourne, Parkville, Victoria 3010, Australia

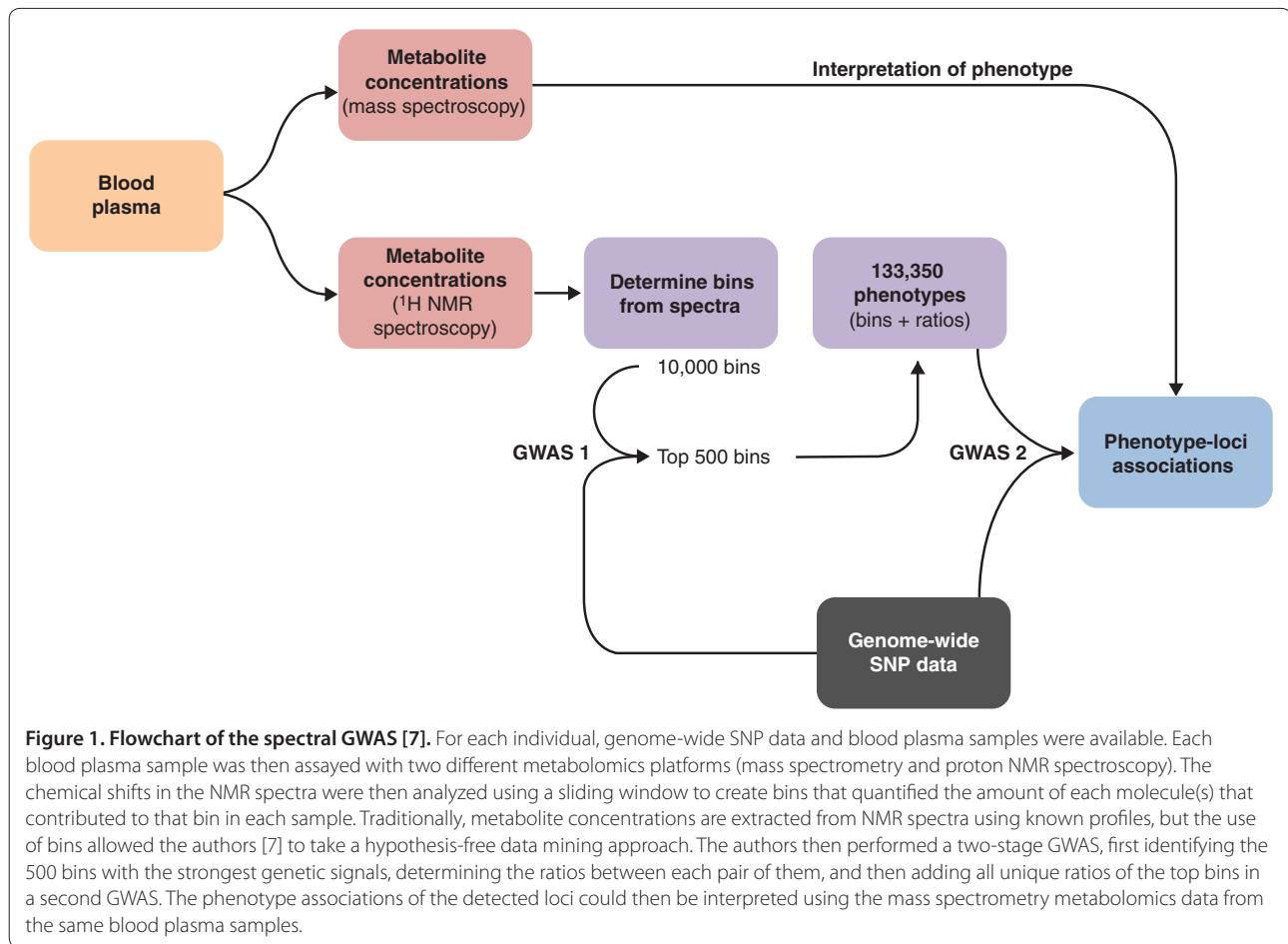


Figure 1. Flowchart of the spectral GWAS [7]. For each individual, genome-wide SNP data and blood plasma samples were available. Each blood plasma sample was then assayed with two different metabolomics platforms (mass spectrometry and proton NMR spectroscopy). The chemical shifts in the NMR spectra were then analyzed using a sliding window to create bins that quantified the amount of each molecule(s) that contributed to that bin in each sample. Traditionally, metabolite concentrations are extracted from NMR spectra using known profiles, but the use of bins allowed the authors [7] to take a hypothesis-free data mining approach. The authors then performed a two-stage GWAS, first identifying the 500 bins with the strongest genetic signals, determining the ratios between each pair of them, and then adding all unique ratios of the top bins in a second GWAS. The phenotype associations of the detected loci could then be interpreted using the mass spectrometry metabolomics data from the same blood plasma samples.

unbiased way as well. Instead of searching for previously characterized metabolites in the NMR spectra and testing for association of these metabolites with genotype, the authors examined all available signals in the molecular spectra and associated each one with genotypes in a GWAS-style approach [7]. Similar to unbiased GWASs, the main premise of the unbiased NMR search is that by expanding testing beyond previously known metabolites, some novel classes and associations may be discovered and characterized.

To this end, Suhre and colleagues [7] used NMR measurements of plasma samples from more than 1,700 individuals in the KORA study [8]. A workflow of their study is presented in Figure 1. The same individuals were also genotyped using a genome-wide array, covering more than 600,000 genome-wide single nucleotide polymorphisms (SNPs). They binned the NMR spectra into 10,000 bins (spectral features), where each bin represents a potentially different metabolite. Binning is a simple procedure where an NMR spectra is split into windows of equal width (in parts per million (ppm)) and the signal intensity in a bin represents a quantification of

the molecule(s) in that window for that sample. Often ratios of metabolites are more biologically informative than metabolite concentrations themselves, as these ratios better reflect enzymatic reactions, in which one metabolite is converted into another at a certain rate. However, exploring all unique pairs of bins for association with each SNP is computationally difficult. Therefore, the authors [7] took a two-stage approach: first all spectral features were examined for association with the SNPs, and then the top 500 spectral features were used to compute pairwise ratios, yielding a total of 133,350 phenotypes. The association between the genotypes and the NMR-based phenotypes (either spectral features or ratios thereof) was tested using a linear model adjusted for age and gender, followed by Bonferroni adjustment for multiple testing.

Using this approach, seven loci achieved genome-wide significance: *LIPC*, *CETP*, *FADS1*, *GCKR*, *APOA1*, *CPS1*, and *PYROXD2*. Of these, five are well-known loci that also had been previously reported using a targeted approach on the same data (examining 15 known lipoprotein subclasses). The use of ratios of NMR shifts

rather than the individual shifts themselves resulted in lower phenotypic variance and substantially lower *P*-values for four of these loci than were achievable using the previously reported lipid subclasses.

As further validation of the NMR spectra, the authors [7] compared the results from NMR with those obtained from mass spectroscopy, showing that NMR spectra for the detected loci generally correlated with concentrations for the same metabolites determined by mass spectrometry. Although the possible applications of these methods are exciting, one future challenge for phenotypically and genotypically unbiased studies will be the interpretation of the associations detected, as correlation with a known variable is confounded by other cross-correlations.

The future of metabolic trait associations

This study has highlighted two concepts that may prove useful in further genetic association studies of many phenotypes: large-scale unbiased screening of phenotypes and trying to account for inter-phenotype relationships (such as ratios). There is potential to expand the types of relationships modeled, for example, using phenotype correlation networks [6,9] that capture potential pleiotropy of loci affecting a group of correlated metabolites. More generally, this work is part of a trend towards a systems-level analysis of disease, based on multivariate data analysis of multiple complementary datasets such as gene expression, metabolites, and genetic variation data [10], leading not just to detection of genotype-phenotype associations as in standard GWASs but ultimately to better mechanistic understanding of the pathways and molecular networks involved in the architecture of human traits and disease.

Abbreviations

GWAS, genome-wide association study; NMR, nuclear magnetic resonance; SNP, single nucleotide polymorphism.

Competing interests

The authors declare that they have no competing interests.

Acknowledgements

MI is supported by the NHMRC (fellowship no. 637400).

Published: 15 February 2013

References

1. Suhre K, Gieger C: **Genetic variation in metabolic phenotypes: study designs and applications.** *Nat Rev Genet* 2012, **13**:759-769.
2. Gieger C, Geistlinger L, Altmaier E, Hrabce 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.
3. 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.
4. Suhre K, Shin SY, Petersen AK, Mohny RP, Meredith D, Wägele B, Altmaier E, CARDIoGRAM, 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, et al.: **Human metabolic individuality in biomedical and pharmaceutical research.** *Nature* 2011, **477**:54-60.
5. 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.
6. Inouye M, Ripatti S, Kettunen J, Lyytikäinen LP, Oksala N, Laurila PP, Kangas AJ, Soininen P, Savolainen MJ, Viikari J, Kähönen M, Perola M, Salomaa V, Raitakari O, Lehtimäki T, Taskinen MR, Järvelin MR, Ala-Korpela M, Palotie A, de Bakker PI: **Novel loci for metabolic networks and multi-tissue expression studies reveal genes for atherosclerosis.** *PLoS Genet* 2012, **8**:e1002907.
7. Raffler J, Römisch-Margl W, Petersen A-K, Pagel P, Blöchl F, Hengstenberg C, Illig T, Meisinger C, Stark K, Wichmann H-E, Adamski J, Gieger C, Kastenmüller G, Suhre K: **Identification and MS assisted interpretation of genetically influenced NMR signals in human plasma.** *Genome Med* 2013, **5**:13.
8. Wichmann HE, Gieger C, Illig T: **KORA-gen--resource for population genetics, controls and a broad spectrum of disease phenotypes.** *Gesundheitswesen* 2005, **67** Suppl 1:S26-S30.
9. Kim S, Xing EP: **Statistical estimation of correlated genome associations to a quantitative trait network.** *PLoS Genet* 2009, **5**:e1000587.
10. Inouye M, Kettunen J, Soininen P, Silander K, Ripatti S, Kumpula LS, Hämmäläinen E, Jousilahti P, Kangas AJ, Männistö S, Savolainen MJ, Jula A, Leiviskä J, Palotie A, Salomaa V, Perola M, Ala-Korpela M, Peltonen L: **Metabonomic, transcriptomic, and genomic variation of a population cohort.** *Mol Syst Biol* 2010, **6**:441.

doi:10.1186/gm418

Cite this article as: Inouye M, Abraham G: **Look, no hands! Spectral biomarkers from genetic association studies.** *Genome Medicine* 2013, **5**:14.