

PERSPECTIVE OPEN

Redefining environmental exposure for disease etiology

Stephen M. Rappaport¹

Etiological studies of human exposures to environmental factors typically rely on low-throughput methods that target only a few hundred chemicals or mixtures. In this Perspectives article, I outline how environmental exposure can be defined by the blood exposome—the totality of chemicals circulating in blood. The blood exposome consists of chemicals derived from both endogenous and exogenous sources. Endogenous chemicals are represented by the human proteome and metabolome, which establish homeostatic networks of functional molecules. Exogenous chemicals arise from diet, vitamins, drugs, pathogens, microbiota, pollution, and lifestyle factors, and can be measured in blood as subsets of the proteome, metabolome, metals, macromolecular adducts, and foreign DNA and RNA. To conduct ‘exposome-wide association studies’, blood samples should be obtained prospectively from subjects—preferably at critical stages of life—and then analyzed in incident disease cases and matched controls to find discriminating exposures. Results from recent metabolomic investigations of archived blood illustrate our ability to discover potentially causal exposures with current technologies.

npj Systems Biology and Applications (2018)4:30; doi:10.1038/s41540-018-0065-0

INTRODUCTION

The publication of the human genome in 2003 led to speculation^{1–3} that genomic technologies would identify the causes of major chronic diseases, particularly cancer and cardiovascular disease, and would lead to personalized strategies for disease prevention. However, most genome-wide-association studies (GWAS) have not detected large effects of common genetic variants on disease incidence.^{4,5} The small effect sizes identified from single nucleotide polymorphisms detected by GWAS (for example Pharoah et al.⁶ and Dehghan et al.⁷) are consistent with studies of monozygotic twins that point to contributions of entire genotypes toward cancer and cardiovascular disease of 8% and 22%, respectively.⁸ Thus, in weighing the relative influences of heritable genetics and environmental exposures on chronic diseases, the modest effects of heritable genetics suggest that exposures and/or gene–environment interactions ($G \times E$) are major causal factors. Indeed, roughly half of the 50 million global deaths in 2010 were attributed to 18 environmental exposures, led by tobacco smoking, particulate air pollution and indoor smoke, high plasma sodium, and alcohol use.⁹ The clear implication is that epidemiologists seeking unknown causes of chronic diseases should employ a balanced strategy that characterizes both heritable genetics and exposures at high resolution. However, because the human genome project focused exclusively on the genome, it did not motivate the discovery of causal exposures. Indeed, etiologic research still focuses on only a few hundred chemicals or mixtures that are quantified by combinations of questionnaires, deterministic models and some measurements.¹⁰ By continuing to explore such a small universe of exposures, we limit our chances to discover unknown causes of disease.

Defining exposure via the blood exposome

The conundrum, where scientists use high-throughput genomics to detect the effects of heritable genetics on disease incidence,

but rely upon low-technology methods to study the effects of exposures, motivated Christopher Wild to promote the concept of an ‘exposome’—representing the totality of exposures received by an individual during life—for etiologic investigations of cancer.¹¹ But unlike the genome, which is largely fixed at birth, the exposome has input from both exogenous and endogenous sources that change throughout life. This calls into question the very nature of ‘exposure’ as a variable in studies of disease etiology. Certainly environmental exposures can be related to levels of pollutants in air, water, and food. But do exposures also include input from nutrients, psycho-social stress, infections, and lifestyle factors? And do perturbations in levels of endogenous molecules, such as sterols and hormones, inflammatory proteins, and metabolites generated by intestinal microbiota constitute exposures? Based on results from the Global Burden of Disease Study,¹⁰ it is reasonable to speculate that all of these sources generate exposures that can contribute to disease risks. The challenge is to find a suitable avenue for investigating these myriad exposures collectively in etiologic research.

Recognizing that disease processes involve chemicals that alter normal function inside the body, Martyn Smith and I suggested in 2010 that the exposome could be considered as the totality of chemicals that can be measured in blood.¹² We reasoned that fundamental processes of life rely on chemical communication via circulating molecules from both genetic and environmental sources, and that these chemicals can be interrogated in blood. Thus the ‘blood exposome’ offers an efficient means to integrate exposures from all sources.¹³

As shown in Fig. 1, endogenous chemicals are generated in the pathway: genome (G), epigenome (G_E), transcriptome (R), proteome (P), and metabolome (M). The genome interacts with molecules and cells via proteins (for example enzymes, cytokines, receptors, transcription factors, and post-translational modifications) and small molecules (for example amino acids, hormones, lipids, neurotransmitters, human metabolites, and reactive

¹Program in Environmental Health Sciences, School of Public Health, University of California, Berkeley, CA 94720, USA

Correspondence: Stephen M. Rappaport (srappaport@berkeley.edu)

Received: 18 April 2018 Revised: 14 May 2018 Accepted: 23 May 2018

Published online: 1 September 2018

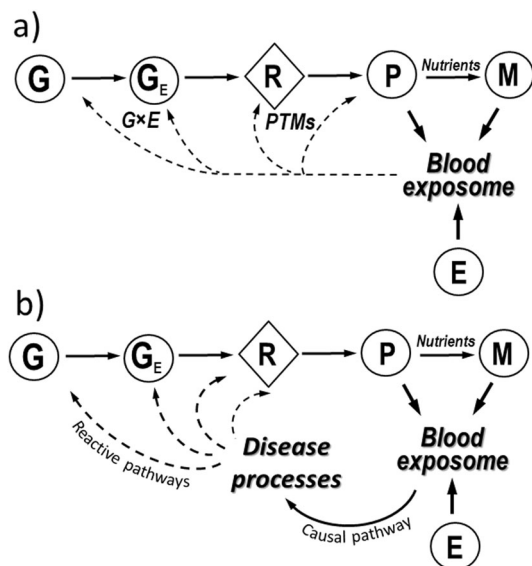


Fig. 1 **a** Inputs to the blood exposome from endogenous sources (G, genome; G_E, epigenome; R, transcriptome; P, proteome; M, metabolome), exogenous exposures (E), post-translational modifications (PTMs) and gene–environment interactions (G × E). **b** Pathways connecting the blood exposome to disease processes (causal pathways) and subsequent feedback to G, G_E, R, and P (via reactive pathways)

oxygen, and carbonyl species) that are distributed throughout the body by the blood. Indeed, modern medicine relies on surveillance of genome-related factors in blood to evaluate disease risks; for example blood levels of C-reactive protein, fibrinogen, and homocysteine have been used as biomarkers of heart disease for more than a decade.¹⁴ Careful curation of factors from the genome to the metabolome (Fig. 1) in observational studies can link circulating molecules with genetic loci and reinforces the idea that the proteome and metabolome contribute to the molecular events that underlie disease associations in GWAS.^{15–17}

The environmental-exposure component (E) in Fig. 1 represents chemicals from exogenous sources, such as diet, vitamins, drugs, pathogens, microbiota, pollution, and lifestyle factors^{12,18} that can be measured in blood as small molecules,¹⁹ metals,²⁰ antigenic proteins,²¹ and foreign DNA and RNA.²² Furthermore, after exogenous chemicals enter the systemic circulation via inhalation, ingestion, or infection, they generate more chemicals via metabolism to reactive intermediates and end products that also enter the blood. Stable adducts of circulating proteins, particularly hemoglobin and human serum albumin, offer avenues for studying the distribution of reactive intermediates that cannot be measured directly in blood.²³

To glimpse a portion of the blood exposome, Rappaport et al.¹³ examined blood concentrations of 1561 small molecules and inorganic species that had been compiled from healthy individuals (mostly adults) by the National Health and Nutrition Examination Survey (NHANES, www.cdc.gov/nchs/nhanes/index.htm) in samples from the U.S. and the Human Metabolome Database (HMDB, www.hmdb.ca) in samples from throughout the world. These molecules and inorganic species comprised more than 100 chemical classes and displayed an extraordinary 10¹¹-fold range of blood concentrations (from fM to mM). Distributions of chemical concentrations derived from food, drugs, and endogenous sources were very similar, whereas blood concentrations of chemicals that were likely results of exposure to pollutants were typically 1000-fold lower than those from the other three sources. Of these 1561 chemicals, 336 had at least one PubMed

citation that associated them with a major chronic disease (cardiovascular disease, cancer, or respiratory disease).¹³ Median numbers of PubMed citations per chemical varied significantly across sources of exposure (endogenous, food, drugs, and pollutants) with a typical chemical derived from food being cited about twice as often as one from another source.

Moving towards exposome-wide-association studies

Untargeted-omics analysis of chemicals in blood samples from patients with disease and healthy controls allows what have been termed exposome-wide-association studies,²⁴ which seek to discover discriminating molecular features that may ultimately be linked to causal exposures.^{18,24} Since the proteome, metabolome, and environmental exposures all contribute to the blood exposome (Fig. 1), examples of this type of analysis include proteomics (endogenous and foreign proteins),²⁵ metabolomics (small molecules),¹⁹ metallomics (metals),²⁰ adductomics (products of reactive intermediates),^{26,27} and metagenomics (foreign DNA and RNA).²² Indeed, it is now feasible to conduct studies that focus on each of these chemical classes separately in human blood or other available biofluids, such as urine or saliva.

The functional genome (genome, epigenome, transcriptome, proteome, and metabolome in Fig. 1) translates genetic information into homeostatic networks of proteins and small molecules. Some of these molecules are causally related to disease processes²⁸ ('causal pathway' in Fig. 1b). But as a disease progresses, it affects the systems biology in ways that disrupt normal homeostasis, thereby altering the composition of the proteome and metabolome (for example Liddy et al.²⁹ and Sekula et al.³⁰). These feedback loops, where disease processes alter the functional genome, have been termed 'reactive pathways'³¹ (Fig. 1b) and can lead to reverse causality in observed associations.²⁴ That is, when blood is obtained from disease cases after diagnosis, a protein or small molecule that discriminates between blood samples from cases and controls could have resulted from a reactive pathway rather than a causal pathway. One way to correctly identify the influence of causal environmental exposures is to conduct exposome-wide-association studies with archived blood from disease cases and controls that are nested in prospective cohorts. By using specimens from these cohorts that were collected prior to diagnosis, causal signals are less affected by metabolic dysregulation and the interval between blood collection and diagnosis can be used as a covariate to determine whether a given association is likely to involve reactive pathways.³²

Metabolomics-based exposome-wide-association studies

Of the various 'omics' methods that can be used to discover environmental exposures associated with disease, metabolomics has received the greatest attention. The current generation of high-resolution liquid chromatography-mass spectrometry (LC-MS) can routinely quantify more than 20,000 small-molecule features in a few microliters of blood,³³ and online databases facilitate annotation of many analytes.³⁴ Nuclear magnetic resonance spectroscopy (NMR) can also be used for untargeted analysis of a much smaller set of abundant small molecules and lipoproteins.^{35,36} When coupled with multivariate analyses to find discriminating small molecules in prediagnostic blood from disease cases and controls, metabolomics³⁷ can be regarded as an important subset of exposome-wide-association studies for disease etiology. Table 1 summarizes results from 13 studies that measured small molecules in plasma or serum from incident cases and controls to discover possible causes of cardiovascular disease,³⁸ diabetes,^{39–43} and a host of cancers.^{44–52} Periods of follow-up ranged from 2 to 9.6 years and 10 of the 14 studies were performed with LC-MS. Interestingly, only three of the LC-MS studies employed untargeted designs^{38,50,51} and thus many did

Table 1. Recent metabolomics studies that investigated disease associations with small-molecule features in plasma or serum from prospective cohorts

Phenotype	Cohort	Cases/controls	Follow-up (y)	Analytical platform	Design	Exposure variables	Likely associations	Refs.
Cardiovascular disease	Gene Bank	75/75	≤3	LC-MS	Untargeted	40 Metabolites (out of >2000 detected features) that met 'acceptance criteria'	18 Small-molecule features of which choline, betaine, and TMAO were annotated	38
Type 1 diabetes	DIPP	56/73	3.7	LC-MS	Targeted	53 Lipids	Children who progressed to T1D were deficient in triglycerides and phosphatidylcholines (possible choline deficiency)	39
Type 2 diabetes	FHS	189/189	12	LC-MS	Targeted	61 Polar metabolites and >100 Lipids	Branched-chain and aromatic amino acids increased risk; sets of lipids increased or decreased risk depending on chain length and double bonds	40,41
Type 2 diabetes	SCHS	197/197	6	LC-MS and GC-MS	Untargeted	4859 Polar and nonpolar metabolites	35 Significant associations including branched-chain amino acids & nonesterified fatty acids and lysophosphatidylinositols	43
Pre-diabetes	KORA	IFG:102/866, IGT:238/866	7	LC-MS	Targeted	140 Lipids, amino acids, and biogenic amines	26 Associations, with glycine, LPC (18:2), and acetylcarnitine being the strongest	42
Gastric cancer	EPIC	238/626	3.2	GC-MS	Targeted	22 Phospholipid fatty acids	Oleic acid, α-linolenic acid, and di-homo-γ-linolenic acid	44
Breast cancer	EPIC	363/702	7	GC-MS	Targeted	22 Phospholipid fatty acids	Trans-palmitoleic and elaidic acids	45
Hepatocellular carcinoma	EPIC	114/122	>2	NMR	Untargeted	8500 NMR bins reduced to 285 clusters of variables	Clusters of sugars, amino acids, lipids and nutrients	46
Colorectal cancer	WHI-OS	835/835	5.2	LC-MS	Targeted	Choline and its metabolites	TMAO and betaine/choline ratio	47
Colorectal cancer	EPIC	1367/2323	3.7	LC-MS	Targeted	Methionine, choline, betaine and dimethylglycine	Weak associations with methionine, choline and betaine	48
Colorectal cancer	PLCO	254/254	7.8	LC-MS and GC-MS	Untargeted	268 Annotated metabolites detected in >80% of specimens	Glycochenodeoxycholate in women but not men	50
Pancreatic cancer	HPFS, NHS, PHS, WHI-OS	454/908	8.7	LC-MS	Targeted	83 Polar metabolites	Branched-chain amino acids	49
Prostate cancer	ATBC	200/200	≤20	LC-MS & GC-MS	Untargeted	626 Annotated metabolites detected in >95% of specimens	None after Bonferroni correction	51
Hepatobiliary cancers	EPIC	HCC:147/147, IHBC:43/43, GBTC:134/134	9.6	LC-MS	Targeted	28 Amino acids, biogenic amines, and total hexoses	HCC: 14 Molecules, mainly branched-chain and aromatic amino acids	52

ATBC alpha-tocopherol beta-carotene cancer prevention study; DIPP Type-1 diabetes prediction and prevention study (birth cohort); EPIC European Prospective Investigation into Cancer; FHS Framingham Health Study; GBTC gallbladder and biliary tract cancers; GC-MS gas chromatography-mass spectrometry; HCC hepatocellular carcinoma; HPFS Health Professionals Follow-up Study; IFG impaired fasting glucose; IGT impaired glucose tolerance; IHBC intrahepatic bile duct cancer; GBTC gallbladder and biliary-tract cancers; KORA Cooperative Health Research in the Region of Augsburg cohort; LC-MS liquid chromatography-mass spectrometry; LPC lysophosphatidylcholine; NHS Nurses' Health Study; NMR nuclear mass resonance spectroscopy; PHS Physicians' Health Study; PLCO prostate, lung, colorectal, and ovarian cancer screening trial; SCHS Singapore Chinese Health Study; TMAO trimethylamine-N-oxide; WHI-OS Women's Health Initiative-Observational Study

not take full advantage of the omics capabilities of the analytical platforms. Nonetheless, these studies demonstrate that metabolomics can readily characterize complex mixtures of small molecules in a few microliters of archived blood from incident cases and matched controls. Indeed, most of the studies in Table 1 found significant disease associations with particular molecules. If exposure-related covariates are available from questionnaires or environmental measurements, then a 'meet-in-the-middle' strategy can be used to connect discriminating features from untargeted metabolomics with possible exposure sources,^{46,53,54} and such correlations can assist with annotations.

Although the literature summarized in Table 1 is dominated by targeted designs, hypothesis-free exposome-wide-association studies can be performed with untargeted analyses that focus on those features, whose signatures (for example LC-MS peaks defined by accurate molecular mass and chromatographic retention time) differ in abundance between cases and controls.³³ After highly associated features from this analysis have been identified, the molecules can be targeted in follow-up studies to identify environmental sources or reactive pathways, to establish exposure-response relationships and other evidence supporting causality,²⁴ and to direct interventions and predictive modeling. These follow-up studies can employ high-throughput methods to quantify selected analytes in thousands of biospecimens using, for example, triple-quadrupole LC-MS^{38,55,56} or NMR.^{35,36}

The untargeted exposome-wide association study conducted by Wang et al.³⁸ (Table 1) is noteworthy because the authors found 18 features (out of more than 2000 detected by LC-MS) that were associated with cardiac events in plasma samples from only 75 incident cases and matched controls. Three highly discriminating features were choline (a nutrient) and its metabolites, betaine, and trimethylamine-*N*-oxide (TMAO). As TMAO is a product of joint microbial and human metabolism of choline and carnitine (another nutrient),^{55,56} the positive association between plasma TMAO and cardiac events points to the involvement of dietary factors combined with the gut microbiota in the etiology of cardiovascular disease. Indeed, the early associations detected between plasma TMAO and cardiac events³⁸ spawned an extensive set of follow-up studies that employed targeted methods to replicate the findings and to explore contributions of TMAO and the gut microbiota towards development of cardiovascular disease.⁵⁷ It is also interesting that the study by Bae et al.⁴⁷ (Table 1) found a positive association between plasma TMAO and colorectal-cancer incidence, again suggesting involvement of the gut microbiota.

Time-varying exposures

The blood exposome is dynamic with concentrations of chemicals varying throughout life due to changes in location, physiology, diet, lifestyle, and other factors.⁴² Given the impact of cumulative exposures ('exposure memory'⁵⁸) on chronic diseases, it is important that exposure monitoring begin in early life. Birth cohorts provide a perfect avenue for obtaining repeated measurements of the blood exposome—beginning at birth and continuing through critical stages of life—that can be used to detect disease associations and windows of susceptibility (for example Oresic et al.³⁹). However, any cohort with repeated collection of blood can provide critical information regarding the timing of disease progression (for example Soininen et al.³⁵). Neonatal blood spots that are collected at birth to screen for congenital errors in metabolism could be archived for subsequent exposome-wide-association studies to find effects of in utero environmental exposures on pediatric (or later) diseases.⁵⁹

Temporal variability of individuals' exposomes leads to exposure-measurement errors that attenuate case-control comparisons of blood levels⁶⁰ and thereby reduce the power to detect disease associations. The magnitudes of exposure-measurement

errors depend, in part, on the residence times of omics features in the body. Small molecules, which tend to have residence times of less than one day, can have much greater measurement error than longer-lived biomarkers, such as adducts of human serum albumin or hemoglobin which reside in the body for 1–2 months.⁶⁰ However, other factors influence temporal variability in blood concentrations; for example, levels of small molecules under homeostatic control can be quite stable over time.⁶¹ Cohorts with repeated collection of blood permit cumulative exposures of omics features to be estimated with concomitant reduction in exposure-measurement errors.^{60,61} Although exposure-measurement errors tend to bias case-control comparisons towards the null and thus result in false negatives, associations detected in exposome-wide association studies with single biospecimens from each subject are unlikely to be false positives after adjustment for multiple testing and should be followed up with validation samples.

CONCLUSIONS

Transformative research generally happens once in a generation. Over the last quarter of a century, epidemiologists have emphasized genetic factors as the putative causes of chronic diseases. Because the human genome project planted the seeds for genome sequencing and large-scale GWAS, it was inevitable that these methods would be used to search for disease causes and, in fact, almost 2000 GWAS have been reported.⁶² Yet, virtually all disease-associated variants individually contributed very small risks.⁶³ This outcome should not be taken to mean that the totality of genetic risks is trivial. After all, studies of monozygotic twins in Western Europe point to attributable genetic risks of about 8% overall for cancer and 22% for coronary heart disease and accounted for around 250,000 deaths in the year 2000.⁸ Nor do I discount the possibility that genes mainly exert their influence through gene-environment interactions, including epigenetics.^{64,65} But, based on current evidence, there can be little doubt that the next generation of etiological research should move towards environmental exposures as causes of chronic diseases, possibly operating in tandem with genetic factors.

In the age of GWAS it is difficult to reconcile the crude state of knowledge about environmental exposures that has been gleaned from traditional methods.⁶⁶ Indeed, a compelling reason for embracing the blood exposome is the potential to perform exposome-wide-association studies that comprehensively characterize environmental exposures with biospecimens from nested case-control studies or from surveillance of individuals' blood exposomes via routine screening.³³ By heightening awareness of the enormous diversity of environmental exposures, the blood exposome should promote the coalescing of etiological research that has been fractured along lines related to exposure sources, for example air, water, diet, microbiota, infections, and psychosocial stress.¹² To reach their full potential, applications employing human blood or other biofluids for exposome-wide-association studies will require standardization of methods and rigorous multi-step replication in order to find unknown causes of chronic diseases.

ACKNOWLEDGEMENTS

The author is indebted to Dean P. Jones for insights regarding the concept of the functional genome and metabolomics-based exposome-wide-association studies and for editorial suggestions by Rebecca Kirk. This work was supported by the U.S. National Institutes of Health through the National Institute for Environmental Health Sciences (grants P01E5018172, P50E5018172, and P42E5004705) and the National Cancer Institute (grant R33CA191159), and by the U.S. Environmental Protection Agency (grant RD83451101).

AUTHOR CONTRIBUTIONS

S.M.R. conceived and wrote the paper.

ADDITIONAL INFORMATION

Competing interests: The author declares no competing interests.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

REFERENCES

- Guttmacher, A. E. & Collins, F. S. Welcome to the genomic era. *N. Engl. J. Med.* **349**, 996–998 (2003).
- Haga, S. B., Khoury, M. J. & Burke, W. Genomic profiling to promote a healthy lifestyle: not ready for prime time. *Nat. Genet.* **34**, 347–350 (2003).
- Hood, L., Heath, J. R., Phelps, M. E. & Lin, B. Systems biology and new technologies enable predictive and preventative medicine. *Science* **306**, 640–643 (2004).
- Goldstein, D. B. Common genetic variation and human traits. *N. Engl. J. Med.* **360**, 1696–1698 (2009).
- Manolio, T. A. et al. Finding the missing heritability of complex diseases. *Nature* **461**, 747–753 (2009).
- Pharoah, P. D. et al. GWAS meta-analysis and replication identifies three new susceptibility loci for ovarian cancer. *Nat. Genet.* **45**, 370e1–372e1 (2013).
- Dehghan, A. et al. Genome-wide association study for incident myocardial infarction and coronary heart disease in prospective cohort studies: the CHARGE Consortium. *PLoS One* **11**, e0144997 (2016).
- Rappaport, S. M. Genetic factors are not the major causes of chronic diseases. *PLoS One* **11**, e0154387 (2016).
- Lim, S. S. et al. A comparative risk assessment of burden of disease and injury attributable to 67 risk factors and risk factor clusters in 21 regions, 1990–2010: a systematic analysis for the Global Burden of Disease Study 2010. *Lancet* **380**, 2224–2260 (2012).
- Collaborators, G. B. D. R. F. et al. Global, regional, and national comparative risk assessment of 79 behavioural, environmental and occupational, and metabolic risks or clusters of risks in 188 countries, 1990–2013: a systematic analysis for the Global Burden of Disease Study 2013. *Lancet* **386**, 2287–2323 (2015).
- Wild, C. P. Complementing the genome with an “exposome”: the outstanding challenge of environmental exposure measurement in molecular epidemiology. *Cancer Epidemiol. Biomark. Prev.* **14**, 1847–1850 (2005).
- Rappaport, S. M. & Smith, M. T. Epidemiology. *Environ. Dis. Risks Sci.* **330**, 460–461 (2010).
- Rappaport, S. M., Barupal, D. K., Wishart, D., Vineis, P. & Scalbert, A. The blood exposome and its role in discovering causes of disease. *Environ. Health Perspect.* **122**, 769–774 (2014).
- Gerszten, R. E. & Wang, T. J. The search for new cardiovascular biomarkers. *Nature* **451**, 949–952 (2008).
- Ruedi, R. et al. Genome-wide association study of metabolic traits reveals novel gene-metabolite-disease links. *PLoS Genet.* **10**, e1004132 (2014).
- Adamski, J. Genome-wide association studies with metabolomics. *Genome Med.* **4**, 34 (2012).
- Hartiala, J. A. et al. Genome-wide association study and targeted metabolomics identifies sex-specific association of CPS1 with coronary artery disease. *Nat. Commun.* **7**, 10558 (2016).
- Wild, C. P. The exposome: from concept to utility. *Int. J. Epidemiol.* **41**, 24–32 (2012).
- Johnson, C. H., Ivanisevic, J. & Siuzdak, G. Metabolomics: beyond biomarkers and towards mechanisms. *Nat. Rev. Mol. Cell Biol.* **17**, 451–459 (2016).
- Vogiatzis, C. G. & Zachariadis, G. A. Tandem mass spectrometry in metallomics and the involving role of ICP-MS detection: a review. *Anal. Chim. Acta* **819**, 1–14 (2014).
- Burbelo, P. D., Ching, K. H., Bush, E. R., Han, B. L. & Iadarola, M. J. Antibody-profiling technologies for studying humoral responses to infectious agents. *Expert. Rev. Vaccin.* **9**, 567–578 (2010).
- Dinakaran, V. et al. Elevated levels of circulating DNA in cardiovascular disease patients: metagenomic profiling of microbiome in the circulation. *PLoS One* **9**, e105221 (2014).
- Rubino, F. M., Pitton, M., Di Fabio, D. & Colombi, A. Toward an “omic” physiopathology of reactive chemicals: thirty years of mass spectrometric study of the protein adducts with endogenous and xenobiotic compounds. *Mass. Spectrom. Rev.* **28**, 725–784 (2009).
- Rappaport, S. M. Biomarkers intersect with the exposome. *Biomarkers* **17**, 483–489 (2012).
- Lindsey, M. L. et al. Transformative impact of proteomics on cardiovascular health and disease: a scientific statement from the American Heart Association. *Circulation* **132**, 852–72 (2015).
- Balbo, S., Turesky, R. J. & Villalta, P. W. DNA adductomics. *Chem. Res. Toxicol.* **27**, 356–366 (2014).
- Grigoryan, H. et al. Adductomics pipeline for untargeted analysis of modifications to Cys34 of human serum albumin. *Anal. Chem.* **88**, 10504–10512 (2016).
- Yazdani, A., Yazdani, A., Samiei, A. & Boerwinkle, E. Identification, analysis, and interpretation of a human serum metabolomics causal network in an observational study. *J. Biomed. Inform.* **63**, 337–343 (2016).
- Liddy, K. A., White, M. Y. & Cordwell, S. J. Functional decorations: post-translational modifications and heart disease delineated by targeted proteomics. *Genome Med.* **5**, 20 (2013).
- Sekula, P. et al. A metabolome-wide association study of kidney function and disease in the general population. *J. Am. Soc. Nephrol.* **27**, 1175–1188 (2016).
- Schadt, E. E. et al. An integrative genomics approach to infer causal associations between gene expression and disease. *Nat. Genet.* **37**, 710–717 (2005).
- Perttula, K. et al. Evaluating ultra-long-chain fatty acids as biomarkers of colorectal cancer risk. *Cancer Epidemiol. Biomark. Prev.* **25**, 1216–1223 (2016).
- Jones, D. P. Sequencing the exposome: a call to action. *Toxicol. Rep.* **3**, 29–45 (2016).
- Edmands, W. M. et al. compMS2Miner: an automatable metabolite identification, visualization and data-sharing R package for high-resolution LC-MS datasets. *Anal. Chem.* **89**, 3919–3928 (2017).
- Soininen, P., Kangas, A. J., Wurtz, P., Suna, T. & Ala-Korpela, M. Quantitative serum nuclear magnetic resonance metabolomics in cardiovascular epidemiology and genetics. *Circ. Cardiovasc. Genet.* **8**, 192–206 (2015).
- Mallol, R., Rodriguez, M. A., Brezmes, J., Masana, L. & Correig, X. Human serum/plasma lipoprotein analysis by NMR: application to the study of diabetic dyslipidemia. *Prog. Nucl. Magn. Reson Spectrosc.* **70**, 1–24 (2013).
- Nicholson, J. K. & Lindon, J. C. Systems biology: metabolomics. *Nature* **455**, 1054–1056 (2008).
- Wang, Z. et al. Gut flora metabolism of phosphatidylcholine promotes cardiovascular disease. *Nature* **472**, 57–63 (2011).
- Oresic, M. et al. Dysregulation of lipid and amino acid metabolism precedes islet autoimmunity in children who later progress to type 1 diabetes. *J. Exp. Med.* **205**, 2975–2984 (2008).
- Rhee, E. P. et al. Lipid profiling identifies a triacylglycerol signature of insulin resistance and improves diabetes prediction in humans. *J. Clin. Invest.* **121**, 1402–1411 (2011).
- Wang, T. J. et al. Metabolite profiles and the risk of developing diabetes. *Nat. Med.* **17**, 448–453 (2011).
- Wang-Sattler, R. et al. Novel biomarkers for pre-diabetes identified by metabolomics. *Mol. Syst. Biol.* **8**, 615 (2012).
- Lu, Y. et al. Metabolic signatures and risk of type 2 diabetes in a Chinese population: an untargeted metabolomics study using both LC-MS and GC-MS. *Diabetologia* **59**, 2349–2359 (2016).
- Chajes, V. et al. Plasma phospholipid fatty acid concentrations and risk of gastric adenocarcinomas in the European Prospective Investigation into Cancer and Nutrition (EPIC-EURGAST). *Am. J. Clin. Nutr.* **94**, 1304–1313 (2011).
- Chajes, V. et al. Association between serum trans-monounsaturated fatty acids and breast cancer risk in the E3N-EPIC Study. *Am. J. Epidemiol.* **167**, 1312–1320 (2008).
- Assi, N. et al. A statistical framework to model the meeting-in-the-middle principle using metabolomic data: application to hepatocellular carcinoma in the EPIC study. *Mutagenesis* **30**, 743–53 (2015).
- Bae, S. et al. Plasma choline metabolites and colorectal cancer risk in the Women's Health Initiative Observational Study. *Cancer Res.* **74**, 7442–7452 (2014).
- Nitter, M. et al. Plasma methionine, choline, betaine, and dimethylglycine in relation to colorectal cancer risk in the European Prospective Investigation into Cancer and Nutrition (EPIC). *Ann. Oncol.* **25**, 1609–1615 (2014).
- Mayers, J. R. et al. Elevation of circulating branched-chain amino acids is an early event in human pancreatic adenocarcinoma development. *Nat. Med.* **20**, 1193–1198 (2014).
- Cross, A. J. et al. A prospective study of serum metabolites and colorectal cancer risk. *Cancer* **120**, 3049–3057 (2014).
- Mondul, A. M. et al. Metabolomic analysis of prostate cancer risk in a prospective cohort: The alpha-tocopherol, beta-carotene cancer prevention (ATBC) study. *Int. J. Cancer* **137**, 2124–2132 (2015).
- Stepien, M. et al. Alteration of amino acid and biogenic amine metabolism in hepatobiliary cancers: Findings from a prospective cohort study. *Int. J. Cancer* **138**, 348–60 (2015).
- Chadeau-Hyam, M. et al. Meeting-in-the-middle using metabolic profiling—a strategy for the identification of intermediate biomarkers in cohort studies. *Biomarkers* **16**, 83–88 (2011).

54. Varbo, A. et al. Remnant cholesterol, low-density lipoprotein cholesterol, and blood pressure as mediators from obesity to ischemic heart disease. *Circ. Res.* **116**, 665–673 (2015).
55. Tang, W. H. et al. Intestinal microbial metabolism of phosphatidylcholine and cardiovascular risk. *N. Engl. J. Med.* **368**, 1575–1584 (2013).
56. Koeth, R. A. et al. Intestinal microbiota metabolism of L-carnitine, a nutrient in red meat, promotes atherosclerosis. *Nat. Med.* **19**, 576–585 (2013).
57. Tang, W. H. & Hazen, S. L. The gut microbiome and its role in cardiovascular diseases. *Circulation* **135**, 1008–1010 (2017).
58. Go, Y. M. & Jones, D. P. Redox theory of aging: implications for health and disease. *Clin. Sci.* **131**, 1669–1688 (2017).
59. Petrick, L. et al. An untargeted metabolomics method for archived newborn dried blood spots in epidemiologic studies. *Metabolomics* **13**, 27–37 (2017).
60. Lin, Y. S., Kupper, L. L. & Rappaport, S. M. Air samples versus biomarkers for epidemiology. *Occup. Environ. Med.* **62**, 750–760 (2005).
61. Kukuljan, S. et al. Independent and combined effects of calcium-vitamin D3 and exercise on bone structure and strength in older men: an 18-month factorial design randomized controlled trial. *J. Clin. Endocrinol. Metab.* **96**, 955–963 (2011).
62. Welter, D. et al. The NHGRI GWAS Catalog, a curated resource of SNP-trait associations. *Nucleic Acids Res.* **42**, D1001–D1006 (2014).
63. Chang, C. Q. et al. A systematic review of cancer GWAS and candidate gene meta-analyses reveals limited overlap but similar effect sizes. *Eur. J. Hum. Genet.* **22**, 402–408 (2014).
64. Gluckman, P. D., Hanson, M. A., Cooper, C. & Thornburg, K. L. Effect of in utero and early-life conditions on adult health and disease. *N. Engl. J. Med.* **359**, 61–73 (2008).
65. Roberts, N. J. et al. The predictive capacity of personal genome sequencing. *Sci. Transl. Med.* **4**, 133ra58 (2012).
66. Rappaport, S. M. Implications of the exposome for exposure science. *J. Expo. Sci. Environ. Epidemiol.* **21**, 5–9 (2011).



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2018