The Consortium of Metabolomics Studies (COMETS): Metabolomics in 47 Prospective Cohort Studies

Bing Yu, Krista A. Zanetti, Marinella Temprosa, Demetrius Albanes, Nathan Appel, Clara Barrios Barrera, Yoav Ben-Shlomo, Eric Boerwinkle, Juan P. Casas, Clary Clish, Caroline Dale, Abbas Dehghan, Andriy Derkach, A. Heather Eliassen, Paul Elliott, Eoin Fahy, Christian Gieger, Marc J. Gunter, Sei Harada, Tamara Harris, Deron R. Herr, David Herrington, Joel N. Hirschhorn, Elise Hoover, Ann W. Hsing, Mattias Johansson, Rachel S. Kelly, Chin Meng Khoo, Mika Kivimäki, Bruce S. Kristal, Claudia Langenberg, Jessica Lasky-Su, Deborah A. Lawlor, Luca A. Lotta, Massimo Mangino, Loïc Le Marchand, Ewy Mathé, Charles E. Matthews, Cristina Menni, Lorelei A. Mucci, Rachel Murphy, Matej Oresic, Eric Orwoll, Jennifer Ose, Alexandre C. Pereira, Mary C. Playdon, Lucilla Poston, Jackie Price, Qibin Qi, Kathryn Rexrode, Adam Risch, Joshua Sampson, Wei Jie Seow, Howard D. Sesso, Svati H. Shah, Xiao-Ou Shu, Gordon C.S. Smith, Ulla Sovio, Victoria L. Stevens, Rachael Stolzenberg-Solomon, Toru Takebayashi, Therese Tillin, Ruth Travis, Ioanna Tzoulaki, Cornelia M. Ulrich, Ramachandran S. Vasan, Mukesh Verma, Ying Wang, Nick J. Wareham, Andrew Wong, Naji Younes, Hua Zhao, Wei Zheng, and Steven C. Moore

Correspondence to: Dr. Steven C. Moore, Division of Cancer Epidemiology and Genetics, National Cancer Institute, 9609 Medical Center Drive, Rockville, MD 20850 (e-mail: moorest@mail.nih.gov; phone: 1.240.276.7196).

Affiliations: Department of Epidemiology, Human Genetics, and Environmental Sciences, School of Public Health, the University of Texas Health Science Center at Houston, Houston, TX, USA (Bing Yu and Eric Boerwinkle); Epidemiology and Genomics Research Program, Division of Cancer Control and Population Sciences, National Cancer Institute, Rockville, MD, USA (Krista A. Zanetti, Elise Hoover, and Mukesh Verma); Department of Epidemiology and Biostatistics Milken Institute School of Public Health, George Washington University, Washington, DC, USA (Marinella Temprosa and Naji Younes); Division of Cancer Epidemiology and Genetics, National Cancer Institute, NIH, Rockville, MD, USA (Demetrius

Albanes, Andriy Derkach, Charles E. Matthews, Mary C. Playdon, Joshua Sampson, Rachael Stolzenberg-Solomon and Steven C. Moore); Information Management Services, Inc., Rockville, MD, USA (Nathan Appel and Adam Risch); Department of Nephrology, Hospital del Mar, Institut Mar d´Investigacions Mediques, Barcelona, Spain (Clara Barrios Barrera); Population Health Sciences, Bristol Medical School, University of Bristol, Bristol, UK (Yoav Ben-Shlomo and Deborah A. Lawlor); Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX, USA (Eric Boerwinkle); Institute of Health Informatics Research, UCL Institute of Health Informatics, University College London, UK (Juan P. Casas and Caroline Dale); Broad Institute of MIT and Harvard, Cambridge, MA, USA (Clary Clish and Joel N. Hirschhorn); MRC-PHE Centre for Environment and Health, Department of Epidemiology & Biostatistics, School of Public Health, Imperial College London, W2 1PG, UK (Abbas Dehghan, Paul Elliott and Ioanna Tzoulaki); Channing Division of Network Medicine, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston MA, USA (A. Heather Eliassen and Lorelei A. Mucci); Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston MA, USA (A. Heather Eliassen, Lorelei A. Mucci and Howard D. Sesso); National Institute for Health Research Imperial College Biomedical Research Center, London, SW7 2AZ, UK (Paul Elliott); Health Data Research-UK London Center at Imperial College London, SW7 2AZ (Paul Elliott); Department of Bioengineering, School of Engineering, University of California, San Diego, La Jolla, CA, USA (Eoin Fahy); Research Unit of Molecular Epidemiology, Helmholtz Zentrum München, German Research Center for Environmental Health, Neuherberg, Germany (Christian Gieger); Institute of Epidemiology, Helmholtz Zentrum München, German Research Center for Environmental Health, Neuherberg, Germany (Christian Gieger); German Center for Diabetes Research (DZD), Neuherberg, Germany (Christian Gieger); Section of Nutrition and Metabolism, International Agency for Research on Cancer, Lyon, France (Marc J. Gunter); Department of Preventive Medicine and Public Health, Keio University School of Medicine, Tokyo, 160-8582, Japan (Sei Harada and Toru Takebayashi); Institute for Advanced Biosciences, Keio University, Tsuruoka, 997-0052,

Japan (Sei Harada); Laboratory of Epidemiology & Population Science Laboratory (LEPS), National Institute on Aging, Bethesda, MD, USA (Tamara Harris); Department of Pharmacology, Yong Loo Lin School of Medicine, National University of Singapore, Singapore (Deron R. Herr); Department of Biology, San Diego State University, San Diego, CA, USA (Deron R. Herr); Department of Internal Medicine, Division of Cardiology, Wake Forest School of Medicine, Winston-Salem, NC, USA (David Herrington); Division of Endocrinology, Boston Children's Hospital, Harvard Medical School, Boston, MA, USA (Joel N. Hirschhorn); Department of Genetics, Harvard Medical School, Boston, MA, USA (Joel N. Hirschhorn); Stanford Prevention Research Center, Stanford Cancer Institute, Stanford, CA, USA (Ann W. Hsing); International Agency for Research on Cancer, Lyon, France (Mattias Johansson); Systems Genetics and Genomics Unit, Channing Division of Network Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA, USA (Rachel S. Kelly); Department of Medicine, Yong Loo Lin School of Medicine, National University of Singapore, Singapore (Chin Meng Khoo); Department of Medicine, National University Health System, Singapore (Chin Meng Khoo); Duke-National University of Singapore Graduate Medical School, Singapore (Chin Meng Khoo); Department of Epidemiology and Public Health, University College London, London, UK (Mika Kivimäki); Division of Sleep and Circadian Disorders, Department of Medicine, Brigham and Women's Hospital, Boston, MA, USA (Bruce S. Kristal); Division of Sleep Medicine, Department of Medicine, Harvard Medical School, Boston, MA, USA (Bruce S. Kristal); MRC Epidemiology Unit, University of Cambridge School of Clinical Medicine, Cambridge Biomedical Campus, Cambridge, UK (Claudia Langenberg, Luca A. Lotta and Nick J. Wareham); Channing Laboratory, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA, USA (Jessica Lasky-Su); MRC Integrative Epidemiology Unit at the University of Bristol, Bristol, UK (Deborah A. Lawlor); Department of Twin Research and Genetic Epidemiology, King's College London, London, SE1 7EH, UK (Massimo Mangino and Cristina Menni); University of Hawaii Cancer Center, Epidemiology Program, 701 Ilalo St., Honolulu, HI, USA (Loïc Le Marchand); Department of Biomedical Informatics,

College of Medicine, The Ohio State University, Columbus, OH, USA (Ewy Mathé); Faculty of Medicine, University of British Columbia, Vancouver, BC, Canada (Rachel Murphy); Turku Centre for Biotechnology, University of Turku and Åbo Akademi University, FI-20520 Turku, Finland (Matej Oresic); School of Medical Sciences, Örebro University, 702 81 Örebro, Sweden (Matej Oresic); Department of Medicine, Oregon Health & Science University, Portland, OR, USA (Eric Orwoll); Division of Cancer Population Sciences, Huntsman Cancer Institute, University of Utah, Salt Lake City, UT, USA (Jennifer Ose, Mary C. Playdon and Cornelia M. Ulrich); Department of Population Health Sciences, University of Utah, Salt Lake City, UT, USA (Jennifer Ose); Instituto de Pesquisas Rene Rachou, Fundação Oswaldo Cruz, Belo Horizonte, Brazil (Alexandre C. Pereira); Department of Nutrition and Integrative Physiology, University of Utah, Salt Lake City, UT, USA (Mary C. Playdon); Department of Women and Children's Health, School of Life Course Sciences, Faculty of Life Sciences and Medicine, King's College London, St Thomas' Hospital, London, UK (Lucilla Poston); Usher Institute of Population Health Sciences and Informatics, University of Edinburgh, Edinburgh, Scotland, UK (Jackie Price); Department of Epidemiology and Population Health, Albert Einstein College of Medicine, Bronx, NY, USA (Qibin Qi); Division of Preventive Medicine, Department of Medicine, Brigham and Women's Hospital, Boston, MA, USA (Kathryn Rexrode and Howard D. Sesso); Division of Women's Health, Department of Medicine, Brigham and Women's Hospital, Boston, MA, USA (Kathryn Rexrode); Saw Swee Hock School of Public Health, National University of Singapore and National University Health System, Singapore (Wei Jie Seow); Department of Medicine, Duke University School of Medicine, Durham, NC, USA (Svati H. Shah); Division of Cardiology, Department of Medicine, Duke University School of Medicine, Durham, NC, USA (Svati H. Shah); Duke Clinical Research Institute, Durham, NC (Svati H. Shah); Division of Epidemiology, Department of Medicine, Vanderbilt-Ingram Cancer Center, Vanderbilt Epidemiology Center, Vanderbilt University School of Medicine, Nashville, TN, USA (Xiao-Ou Shu and Wei Zheng); Department of Obstetrics and Gynecology, National Institute for Health Research Cambridge Comprehensive Biomedical Research

Center, and Center for Trophoblast Research, Department of Physiology, Development and Neuroscience, University of Cambridge, Cambridge, UK (Gordon C.S. Smith); Department of Obstetrics and Gynaecology, University of Cambridge, NIHR Cambridge Comprehensive Biomedical Research Centre, Cambridge, UK (Ulla Sovio); Epidemiology Research Program, American Cancer Society, Atlanta, GA, USA (Victoria L. Stevens and Ying Wang); Institute for Advanced Biosciences, Keio University, Tsuruoka, 997-0052, Japan (Toru Takebayashi); Institute of Cardiovascular Sciences, University College London, London, UK (Therese Tillin); Cancer Epidemiology Unit, Nuffield Department of Population Health, University of Oxford, Oxford, UK (Ruth Travis); Section of Preventive Medicine and Epidemiology, Department of Medicine, Boston University School of Medicine, Boston, MA, USA (Ramachandran S. Vasan); Section of Cardiovascular Medicine, Department of Medicine, Boston University School of Medicine, Boston, MA, USA (Ramachandran S. Vasan); Department of Epidemiology, Boston University School of Public Health, Boston, MA, USA (Ramachandran S. Vasan); Framingham Heart Study, Framingham, MA, USA (Ramachandran S. Vasan); MRC Unit for Lifelong Health and Ageing at UCL, London, UK (Andrew Wong); and Department of Epidemiology, University of Texas MD Anderson Cancer Center, Houston, TX, USA (Hua Zhao)

Funding: The **Airwave** Health Monitoring Study is funded by the UK Home Office (780-TETRA) with additional support from the National Institute for Health Research (NIHR), and the Imperial College Biomedical Research Centre in collaboration with Imperial College NHS Healthcare Trust. This work used computing resources of the UK MEDical BIOinformatics partnership (UK MED-BIO) supported by the Medical Research Council (MR/L01632X/1). Dr. Paul Elliott is supported by the UK Dementia Research Unit which receives funding from the UK Medical Research Council, Alzheimer's Society and Alzheimer's Research UK, the Medical Research Council and Public Health England (MR/L01341x/1) for the -MRC-PHE Centre of Environment and Health; the NIHR Health Protection Research Unit in Health Impact of

Environmental Hazards (hpru-2012-10141), and the Health Data Research UK London Centre funded by a consortium of funders led by the Medical Research Council. AD is supported by the Wellcome Trust [206046/Z/17/Z]. The **ATBC** study was supported by the NIH Intramural Research Program, Division of Cancer Epidemiology and Genetics, National Cancer Institute, NIH, Department of Health and Human Services. The Atherosclerosis Risk in Communities (**ARIC**) Study is carried out as a collaborative study supported by National Heart, Lung, and Blood Institute contracts (HHSN268201100005C, HHSN268201100006C, HHSN268201100007C, HHSN268201100008C, HHSN268201100009C, HHSN268201100010C, HHSN268201100011C, and HHSN268201100012C), R01HL087641, R01HL59367 and R01HL086694; National Human Genome Research Institute contract U01HG004402; and National Institutes of Health contract HHSN268200625226C. Dr. Bing Yu is supported by American Heart Association (17SDG33661228). The UK Medical Research Council and the Wellcome Trust (Grant ref: 102215/2/13/2) and the University of Bristol provide core support for **ALSPAC**. The British Heart Foundation (SP/07/008/24066), Wellcome Trust (WT092830/Z/10/ Z), and Joint UK Research Councils via the Lifelong Health and Wellbeing Programme (G1001357) funded follow-up of the women (ALSPAC mothers) currently contributing to COMETS, with metabolomic measurements funded by the US NIH (R01 DK10324) and European Research Council under the European Union's Seventh Framework Programme (FP7/2007-2013) / ERC grant agreement no 669545. **BiB** receives core infrastructure funding from the Wellcome Trust (WT101597MA), a joint grant from the UK Medical Research Council (MRC) and UK Economic and Social Science Research Council (ESRC) (MR/N024397/1), and the National Institute for Health Research (NIHR) under its Collaboration for Applied Health Research and Care (CLAHRC) for Yorkshire and Humber. Follow-up and metabolomic research is supported by the British Heart Foundation (CS/16/4/32482), US National Institute of Health (R01 DK10324), the European Research Council under the European Union's Seventh Framework Programme (FP7/2007-2013) / ERC grant agreement no 669545. The Baependi Heart Study (**BHS**) study was supported by awards from

FAPESP (grants 2007/58150-7, 2010/51010-8, 2011/05804-5, 2013/17368-0), from CNPq (150653/2008-5, 481304/2012-6, and 400791/2015-5), Fundação Zerbini and Proadi - Hospital Samaritano. The **BCFR** cohort was supported by NCU 1UM1CA164920. The BCFR Metabolomics study was funded by internal pilot grant from the Stanford Cancer Institute. **BWHHS** is supported by funding from the British Heart Foundation and the Department of Health Policy Research Programme (England). The British Heart Foundation had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication. **CaPS** was funded by the Medical Research Council and undertaken by the former MRC Epidemiology Unit (South Wales). The CaPS metabolomics was undertaken as part of the UCLEB consortium which is supported by a British Heart Foundation Programme Grant (RG/10/12/28456). The CaPS data archive is maintained by the University of Bristol. The American Cancer Society funds the creation, maintenance, and updating of the Cancer Prevention Study-II (**CPS-II**) cohort. Dr. Svati H. Shah (**CATHGEN** study) is supported by 5R01-HL127009; 1R56-HL129880; 5R01-HL095987; American Heart Association 17SFRN33590127 and American Heart Association 16SFRN31800000. The Childhood Asthma Management Program (**CAMP**) is supported by Contracts NO1-HR-16044, 16045, 16046, 16047, 16048, 16049, 16050, 16051, and 16052 with the National Heart, Lung, and Blood Institute. and the Metabolomic profiling was supported by 1R01HL123915-01 (PI: Jessica Lasky-Su). The **COLO** study was funded by the Lackas Foundation, the Division of Preventive Oncology (Dr. Ulrich) and the German Consortium of Translational Cancer Research (DKTK). Dr. Cornelia M. Ulrich received funding from the Huntsman Cancer Foundation. Research reported in this publication was supported by the National Cancer Institute of the National Institutes of Health under Award Number U01CA206110. The KORA study is funded by the Federal Ministry of Education and Research (Funding number BMBF 01KT1512). **KORA** was initiated and financed by the Helmholtz Zentrum München—German Research Center for Environmental Health,

which is funded by the German Federal Ministry of Education and Research and by the State of Bavaria. Furthermore, KORA research was supported within the Munich Center of Health Sciences (MC-Health), Ludwig-Maximilians-Universität, as part of LMUinnovativ. The KORA-Study Group consists of A. Peters, J.Heinrich, R.Holle, R.Leidl, C.Meisinger, K.Strauch and their co-workers, who are responsible for the design and conduct of the KORA studies. The **DIPP** study was supported by the Academy of Finland (Centre of Excellence in Molecular Systems Immunology and Physiology Research 2012-2017, Decision No. 250114) and Juvenile Diabetes Research Foundation (2-SRA-2014-159-Q-R). The Estonia cohort was supported by European Regional Development Fund, road-map grant no.3.2.0304.11-0312, grant "Center of Excellence in Genomics (EXCEGEN)", and by targeted financing from Estonian Government (IUT24-6, IUT20-60) and CTG grant (SP1GVARENG) from Development Fund of the University of Tartu. The metabolomic studies were funded by Hirschhorm R01 DK075787. During the **DPP** and DPPOS, the National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK) of the National Institutes of Health provided funding to the clinical centers and the Coordinating Center for the design and conduct of the study, and collection, management, analysis, and interpretation of the data (U01 DK048489). The Southwestern American Indian Centers were supported directly by the NIDDK, including its Intramural Research Program, and the Indian Health Service. The General Clinical Research Center Program, National Center for Research Resources, and the Department of Veterans Affairs supported data collection at many of the clinical centers. Funding was also provided by the National Institute of Child Health and Human Development, the National Institute on Aging, the National Eye Institute, the National Heart Lung and Blood Institute, the National Cancer Institute, the Office of Research on Women's Health, the National Institute on Minority Health and Health Disparities, the Centers for Disease Control and Prevention, and the American Diabetes Association. Bristol-Myers Squibb and Parke-Davis provided additional funding and material support during the DPP, Lipha (Merck-Sante) provided medication and LifeScan Inc. donated materials during the DPP and DPPOS. This research was

also supported, in part, by the intramural research program of the NIDDK. LifeScan Inc., Health O Meter, Hoechst Marion Roussel, Inc., Merck-Medco Managed Care, Inc., Merck and Co., Nike Sports Marketing, Slim Fast Foods Co., and Quaker Oats Co. donated materials, equipment, or medicines for concomitant conditions. McKesson BioServices Corp., Matthews Media Group, Inc., and the Henry M. Jackson Foundation provided support services under subcontract with the Coordinating Center. The sponsor for the **ET2DS** was the University of Edinburgh. The study was funded by the Medical Research Council (UK) (ProjectGrant G0500877), the Chief Scientist Office of the Scottish Executive (Programme Support Grant CZQ/1/38), Pfizer plc. and DiabetesUK (Clinical Research Fellowship 10/0003985). The funders had no other role in the design, analysis or writing of this manuscript. The Estonia cohort (**Estonia OE**) was supported by European Regional Development Fund, road-map grant no.3.2.0304.11-0312, grant "Center of Excellence in Genomics (EXCEGEN)", and by targeted financing from Estonian Government (IUT24-6, IUT20-60) and CTG grant (SP1GVARENG) from Development Fund of the University of Tartu. The metabolomic studies were funded by Hirschhorm R01 DK075787. The coordination of **EPIC** is financially supported by the European Commission (DG-SANCO) and the International Agency for Research on Cancer. The national cohorts are supported by Danish Cancer Society (Denmark); German Cancer Aid, German Cancer Research Center (DKFZ), Federal Ministry of Education and Research (BMBF), Deutsche Krebshilfe, Deutsches Krebsforschungszentrum and Federal Ministry of Education and Research (Germany); the Hellenic Health Foundation (Greece); Associazione Italiana per la Ricerca sul Cancro-AIRC-Italy and National Research Council (Italy); Dutch Ministry of Public Health, Welfare and Sports (VWS), Netherlands Cancer Registry (NKR), LK Research Funds, Dutch Prevention Funds, Dutch ZON (Zorg Onderzoek Nederland), World Cancer Research Fund (WCRF), Statistics Netherlands (The Netherlands); Health Research Fund (FIS) PI13/00061 (EPIC-Granada) and, PI13/01162 (EPIC-Murcia), Regional Governments of Andalucía, Asturias, Basque Country, Murcia and Navarra, ISCIII Health Research Funds RD12/0036/0018 (cofounded by FEDER funds/European Regional Development Fund

ERDF) (Spain); Swedish Cancer Society, Swedish Research Council and County Councils of Skåne and Västerbotten (Sweden); Cancer Research UK (14136 to EPIC-Norfolk; C570/A16491 for EPIC-Oxford), Medical Research Council (1000143 to EPIC-Norfolk, MR/M012190/1 to EPIC-Oxford) (UK). The ongoing metabolomics work in EPIC is funded by: Cancer Research UK, World Cancer Research Fund, European Commission and the French National Cancer Institute. The **FENLAND** study was funded by the United Kingdom's Medical Research Council through grants MC_UU_12015/1, MC_PC_13046, MC_PC_13048 and MR/L00002/1, and Cambridge Lipidomics Biomarker Research Initiative (G0800783). Dr. Nicholas J. Wareham is an NIHR Senior Investigator. Research and data from the Framingham studies (**FHS2**, **FHS3**) were supported by the contracts HHSN268201500001 and N01 HC 25195 from the NIH/NHLBI and grant R01 DK081572 from the NIDDK. The Genodiab-Mar (**GDM**) cohort is supporting by a Spanish grant: Instituto de Salud Carlos III (FIS-ISCIII). Project Ref: PI16/00620 and RedinRen RD12/0021/0024. The Health ABC study was supported by NIA Contracts N01-AG-6–2101, N01-AG-6–2103, and N01-AG-6–2106; NIA Grant R01-AG028050, National Institute of Nursing Research Grant R01-NR-012459, the Wake Forest University Claude D. Pepper Older Americans for Independence Center (1P30AG21332); and the Pittsburgh Claude D. Pepper Center (P30 AG024827). **HABC** work is also supported in part by the Intramural program of the National Institutes of Health. Dr. Rachel Murphy is supported by the Canadian Cancer Society (grant #704735). The Health Professionals Follow-up Study (**HPFS**) is funded by a grant from the National Cancer Institute U01 CA167552 and the metabolomics study is funded by P50 090381. The Mexican American Cohort (**MAC**) receives funds collected pursuant to the Comprehensive Tobacco Settlement of 1998 and appropriated by the 76th legislature to The University of Texas MD Anderson Cancer Center. Work in the Mexican American Cohort was supported in part by Center for Translational and Public Health Genomics, the Dan Duncan Family Institute for Risk Assessment and Cancer Prevention. The **MRC NSHD** is funded by the UK Medical Research Council (MC_UU_12019/1). The Multi-Ethnic cohort (**MEC**) was supported by grants P01 CA168530 and U01 CA164973. The **MESA** study

was supported by NIH grant R01 HL133932-01. The Nurses' Health Study (**NHS**) is funded by grants from the National Cancer Institute CA186107 and CA49449. The Nurses' Health Study II (**NHS-II**) is funded by grants from the National Cancer Institute CA176726 and CA067262. Metabolomics studies within the Nurses Health Studies and the Health Professionals Study cohorts are funded by grants from the National Institutes of Health NS045893, CA087969, CA050385, DK103720, CA163451, NS089619, CA090381, CA140790, AR049880; the Department of Defense W81XWH-13-1-0493; the American Society of Clinical Oncology Conquer Cancer Foundation; and the Howard Hughes Medical Institute and Promises for Purple. The Osteoporotic Fractures in Men (**MrOS**) Study is supported by National Institutes of Health funding. The following institutes provide support to MrOS: the National Institute on Aging (NIA), the National Institute of Arthritis and Musculoskeletal and Skin Diseases (NIAMS), Nathe National Center for Advancing Translational Sciences (NCATS), and NIH Roadmap for Medical Research under the following grant numbers: U01 AG027810, U01 AG042124, U01 AG042139, U01 AG042140, U01 AG042143, U01 AG042145, U01 AG042168, U01 AR066160, and UL1 TR000128. The Physicians' Health Study (**PHS**) is supported by grants CA 097193, CA 34944, CA 40360, HL 26490, and HL 34595 from the National Institutes of Health (Bethesda, MD USA). The work in **POPS** was supported by the National Institute for Health Research (NIHR) Cambridge Comprehensive Biomedical Research Centre (Women's Health theme) and a project grant from the Medical Research Council (United Kingdom; G1100221). The **PLCO** study was supported by the NIH Intramural Research Program of the National Cancer Institute, NIH, Department of Health and Human Services. The Shanghai Men's Health Study (**SMHS**), Shanghai Physical Activity Study (**SPA**) and Shanghai Women's Health studies (**SWHS**) are supported by grants from the US National Institutes of Health [R37 CA070867 and UM1 CA182910, R01 CA082729, UM1 CA173640, R01 HL079123 and R01DK108159] as well as Ingram Professorship Funds from the Vanderbilt-Ingram Cancer Center. With respect to the Singapore Prospective Study Program (**SP2**), Dr. Wei Jie Seow is supported by National University of Singapore Start-Up Grant, Dr. Derron R.

Herr is supported by National University of Singapore: NUHSRO/2014/085/AF-Partner/01, and Dr. Chin M. Khoo is supported by National Medical Research Council, Clinician Scientist Award, Ministry of Health Alignment Fund, Janssen Pharmaceuticals Inc, and the National Kidney Foundation. **SABRE** is funded at baseline by the UK Medical Research Council, Diabetes UK, British Heart Foundation, Metabolomics analyses funded by Diabetes UK (13/0004774) and follow- up is funded by the Wellcome Trust (WT082464) and British Heart Foundation (SP/07/001/23603 and CS/13/1/30327). The Tsuroka Metabolomics Cohort study (**TMCS**) is supported in part by research funds from the Yamagata Prefectural Government and the city of Tsuruoka, and by the Grant-in-Aid for Scientific Research (B) (grant numbers JP24390168, JP15H04778), Grant-in-Aid for Challenging Exploratory Research (grant number 25670303), and Grant-in-Aid for Young Scientists(B) (grant number JP15K19231) from the Japan Society for the Promotion of Science. The **TwinsUK** was funded by the Wellcome Trust; European Community's Seventh Framework Programme (FP7/2007-2013). TwinsUK also receives support from the National Institute for Health Research (NIHR)- funded BioResource, Clinical Research Facility and Biomedical Research Centre based at Guy's and St Thomas' NHS Foundation Trust in partnership with King's College London. **UPBEAT** was funded by NIHR (RP-0407-104522), MRC (MR/L002477/1), Diabetes UK, CSO (CZB/A/680), the Biomedical Research Centre at Guys & St Thomas NHS Foundation Trust & King's College London and Tommy's Charity. **VDAART** was supported by U01HL091528 and the metabolomics work in VDAART was supported by 1R01HL123915-01 from the National Heart, Lung, and Blood Institute. Additional support was provided by U54TR001012 from the National Centers for Advancing Translational Sciences (NCATS) for participant visits at Boston Medical Center. VDAART clinical trial registration number: NCT00920621. The Whitehall II study (**WH-II**) is supported by grants from the US National Institutes on Aging (R56AG056477; R01AG034454; R01AG013196), the Heart, Lung, and Blood Institute (R01HL036310), the UK Medical Research Council (MRC, K013351 and R024227) and the British Heart Foundation (PG/29605; RG/13/2/30098). M.K. is supported by the U.K.

MRC (S011676), NordForsk and the Academy of Finland (311492). The Metabolomic analysis in the **WHI** was funded by the National Heart, Lung, and Blood Institute, National Institutes of Health, U.S. Department of Health and Human Services through contract HHSN268201300008C. The WHI program is funded by the National Heart, Lung, and Blood Institute, National Institutes of Health, U.S. Department of Health and Human Services through contracts HHSN268201600018C, HHSN268201600001C, HHSN268201600002C, HHSN268201600003C, and HHSN268201600004C. The WIHS is funded primarily by the National Institute of Allergy and Infectious Diseases (NIAID), with additional co-funding from the Eunice Kennedy Shriver National Institute of Child Health and Human Development (NICHD), the National Cancer Institute (NCI), the National Institute on Drug Abuse (NIDA), and the National Institute on Mental Health (NIMH). Targeted supplemental funding for specific projects is also provided by the National Institute of Dental and Craniofacial Research (NIDCR), the National Institute on Alcohol Abuse and Alcoholism (NIAAA), the National Institute on Deafness and other Communication Disorders (NIDCD), and the NIH Office of Research on Women's Health. WIHS data collection is also supported by UL1-TR000004 (UCSF CTSA), UL1-TR000454 (Atlanta CTSA), and P30-AI-050410 (UNC CFAR). WIHS (Principal Investigators): UAB-MS WIHS (Mirjam-Colette Kempf and Deborah Konkle-Parker), U01-AI-103401; Atlanta WIHS (Ighovwerha Ofotokun and Gina Wingood), U01-AI-103408; Bronx WIHS (Kathryn Anastos and Anjali Sharma), U01-AI-035004; Brooklyn WIHS (Howard Minkoff and Deborah Gustafson), U01-AI-031834; Chicago WIHS (Mardge Cohen and Audrey French), U01-AI-034993; Metropolitan Washington WIHS (Seble Kassaye), U01-AI-034994; Miami WIHS (Margaret Fischl and Lisa Metsch), U01-AI-103397; UNC WIHS (Adaora Adimora), U01-AI-103390; Connie Wofsy Women's HIV Study, Northern California (Ruth Greenblatt, Bradley Aouizerat, and Phyllis Tien), U01-AI-034989; WIHS Data Management and Analysis Center (Stephen Gange and Elizabeth Golub), U01-AI-042590; Southern California WIHS (Joel Milam), U01-HD-032632 (WIHS I – WIHS IV). The **WIHS** is funded primarily by the National Institute of Allergy and Infectious Diseases (NIAID), with additional co-funding from the Eunice

Kennedy Shriver National Institute of Child Health and Human Development (NICHD), the National Cancer Institute (NCI), the National Institute on Drug Abuse (NIDA), and the National Institute on Mental Health (NIMH). Targeted supplemental funding for specific projects in WIHS is also provided by the National Institute of Dental and Craniofacial Research (NIDCR), the National Institute on Alcohol Abuse and Alcoholism (NIAAA), the National Institute on Deafness and other Communication Disorders (NIDCD), and the NIH Office of Research on Women's Health. WIHS data collection is also supported by UL1-TR000004 (UCSF CTSA), UL1-TR000454 (Atlanta CTSA), and P30-AI-050410 (UNC CFAR). The WIHS metabolomics was supported by the National Heart, Lung, and Blood Institute (NHLBI) K01HL129892 to Q.Q.

The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Conflict of interest: Dr. Deborah A Lawlor has received support from numerous government and charitable funding, as well as Roche Diagnostics and Medtronic for research unrelated to this paper. Dr. Bruce S. Kristal is the inventor on general metabolomics-related IP that has been licensed to Metabolon via Weill Medical College of Cornell University and for which he receives royalty payments via Weill Medical College of Cornell University. He also consults for and has a small equity interest in the company. Metabolon offers biochemical profiling services and is developing molecular diagnostic assays detecting and monitoring disease. Metabolon has no rights or proprietary access to the research results presented and/or new IP generated under these grants/studies. Dr. Bruce S. Kristal interests were reviewed by the Brigham and Women's Hospital and Partners Healthcare in accordance with their institutional policy. Accordingly, upon review, the institution determined that Dr. Bruce S. Kristal's financial interest in Metabolon does not create a significant financial conflict of interest (FCOI) with this

research. The addition of this statement where appropriate was explicitly requested and approved by BWH.

Running head: The Consortium of Metabolomics Studies (COMETS)

ABSTRACT

The Consortium of Metabolomics Studies (COMETS) was established in 2014 to facilitate large-scale collaborative research on the human metabolome and its relationship with disease etiology, diagnosis, and prognosis. COMETS comprised 47 cohorts from Asia, Europe, North America and South America that together include 137, 000+ participants with blood metabolomics data on samples collected from 1985-2017. Metabolomics data were provided by 17 different platforms with the most frequently used labs being Metabolon Inc. (14 cohorts), the Broad Institute (15 cohorts), and Nightingale Health (11 cohorts). Participants were followed for a median 23 years for health outcomes, including death, cancer, cardiovascular disease, diabetes and others. Available exposure-related data include common clinical measurements, behavioral factors, as well as genome-wide genotype data. Two feasibility studies were conducted to evaluate the comparability of metabolomics platforms used by COMETS cohorts. The first study showed that the overlap between any two different laboratories ranged from 6 to 121 metabolites at five leading laboratories. The second study showed that the median Spearman correlation comparing 111 overlapping metabolites captured by Metabolon and the Broad Institute was 0.79 (interquartile range: 0.56-0.89). We welcome new cohorts and proposals from interested investigators. Forms and other information about COMETS are at our website (https://epi.grants.cancer.gov/comets/).

Keywords: cancer, cohort, diabetes, epidemiology, genetics, heart disease, metabolomics, prospective

Metabolomics is the systematic study of the small molecule constituents of a biological system, typically involving the measurement of 100s to 1000s of metabolites. Metabolomics analyses currently employ a variety of platforms and analytical technologies, none of which measure all metabolites. Recently, metabolomics platforms have improved remarkably in sensitivity and metabolite coverage, leading many researchers, including epidemiologists, to take increased interest in this research area. Metabolomics studies have led to the discovery of new metabolic aspects of complex chronic diseases like diabetes (1-4), cardiovascular disease (5-7), renal disease (8), cancer (9-13), and has yielded new insights into the genome (14-20). Metabolomics studies have also identified biomarkers of blood pressure (21), obesity (22-24), diet and nutrition (25-35), physical activity/sedentary behavior (36), reproductive factors (37, 38), and pharmacological therapies (39).

These studies provided important insights about the human metabolome, but, because metabolomics is expensive ($200-300/sample), they have been small (e.g. <1,000 participants) and with limited demographic and/or socioeconomic diversity. One means to address these issues is to aggregate datasets and resources within a metabolomics consortium. Such a consortium could rapidly attain large sample sizes and increase demographic and geographic diversity. In addition, a consortium can pool expertise from multiple disciplines—such as metabolomics, chemistry, epidemiology, bioinformatics, computational biology, and biostatistics—to improve the conduct of such research.

We describe herein the development of such a consortium, the COnsortium of METabolomics Studies (COMETS). The objectives of this report are to introduce COMETS to the research community at large and describe its participant characteristics, metabolomics assays, and available questionnaire/clinical data. In addition, we describe two feasibility studies: 1) a study that enumerates the metabolites measured by five leading platforms and establishes their overlap; and, 2) a study that compares blood metabolite levels obtained by two leading platforms when tested on split samples.

METHODS

Design of the COMETS consortium

The COMETS consortium was initiated at the "Think Tank on Metabolomics and Prospective Cohorts" on October 28-29, 2014 in Rockville, Maryland, which was supported and convened by the U.S. National Cancer Institute. Invitees were identified by searching the literature (including hand-search of citations) for cohort studies with blood metabolomics data (identified metabolites only) and through discussions with invitees to determine whether we missed key cohorts or investigators. In total, thirty-four investigators representing 23 prospective cohorts and two existing research consortia attended and ultimately agreed to initiate the COMETS consortium.

COMETS includes prospective cohort studies that meet two criteria: (1) the cohort includes 100+ participants with metabolites of known chemical identity measured in blood (plasma or serum) using mass spectrometry (MS), nuclear magnetic resonance spectroscopy, or other multi-analyte analytical technology (e.g. coularray), and (2) cohort participants are followed after blood collection for outcomes (e.g. mortality, cardiovascular disease, diabetes and/or cancer).

COMETS has employed a rolling enrollment and, as of April 2018, included 47 prospective cohorts from Asia, Europe, North America and South America (**Figure 1, Web Table 1**). Participants in these cohort studies were recruited for varying purposes and from different source populations, as follows: (1) eight cohorts initiated as randomized clinical trials (40-48); (2) sixteen cohorts that were population-based or representative of a given geographical area (4, 18, 36, 49-64); (3) three cohorts consisting of volunteers from defined geographical areas (65-67); (4) six cohorts recruited from participants with colorectal cancer (68), cardiovascular disease (69), diabetes (70, 71) or families of persons with these diseases (72, 73); (5) one study of participants with human immunodeficiency virus or at high risk of human immunodeficiency virus (74, 75); (6) four cohorts recruited from specific

occupational groups (76-78); (7) six cohorts—including two of the randomized clinical trials above—that recruited pregnant mothers and/or their recently-born children (41, 45, 79-82); and, (8) four cohorts based on other participant factors, namely elderly participants—including one of the studies above (53, 83), twins (84), and Mexican-Americans residing in Houston, Texas (85).

COMETS research projects are initiated when interested investigators submit a formal proposal describing the aims, outcomes, exposures, covariates, and analytical approach of a proposed study. If the COMETS Steering Committee approves the proposal, it is forwarded to cohort representatives who can then "opt-in" for analysis. These projects will cover a wide scope of topics and require diverse analytical strategies. Initially, however, we will focus on meta-analyses conducted through aggregate results sharing, i.e. each cohort will evaluate metabolite-outcome associations individually and send results centrally for meta-analysis. In addition to producing meta-analysis effect estimates, we will evaluate heterogeneity by study, platform, and participant characteristics (e.g. gender, race, age), and we will account for participant sampling (e.g., selection of twins or case-control risk sets) through mixed effects modeling.

Survey

We ascertained cohort data by e-mailing a survey to each cohort's representative asking about participant characteristics, metabolomics measurements, and measurements from questionnaires or clinical assessments. All cohorts completed the survey. Missing results on the survey led to a recontact and/or telephone call until all items were complete. For determining eligibility, follow-up for disease outcomes was confirmed by literature search. Cohort representatives verified cohort details prior to submission.

Feasibility studies

One key challenge in COMETS is that different cohorts used different metabolomics platforms, and these platforms vary in which metabolites they measure. A second key challenge is that platforms may measure metabolites dissimilarly, i.e. the relative concentrations may differ, ultimately leading to heterogenous study-specific estimates and attenuation of overall meta-analysis estimates. To better understand platform comparability and its implications for future COMETS projects, we conducted two feasibility studies in which we: 1) assessed metabolite overlap for five widely-used metabolomics platforms; and 2) compared the metabolite values measured by the two most widely-used metabolomics platforms (Broad Institute and Metabolon) when tested on split samples.

Assessment of metabolite overlap for five widely-used metabolomics platforms

Currently, no single platform comprehensively assays all metabolites in blood; instead, platforms use customized instrumentation and sample extraction protocols to optimize measurement of broad classes of metabolites. Consequently, different platforms measure different metabolites. The extent of platform overlap, however, has not been systematically evaluated. Most likely, this reflects the difficulty in collating 100s to 1000s of metabolite names in a field that still lacks a standardized nomenclature.

To assess overlap, we collected metabolite names from volunteer COMETS cohorts that used one of five metabolomics platforms (Metabolon, the Broad Institute, Biocrates, the West Coast Metabolomics Center, and Nightingale Health). We also collected relevant meta-data provided by these labs, especially unique identifiers from online metabolite databases such as the Human Metabolome Database (HMDB)(86), Pubchem (87), or Chemspider(88). We used these identifiers and metabolite names to link metabolite identities from different labs. Metabolites with multiple isomers, such as D- and L-glutamate, were adjudicated using InChIKey values, if available, or (as a last resort) original

reported names. The final product was a table that cross-references metabolites assessed by each cohort and is easily queried to show metabolite overlap for any given combination of cohorts.

Comparing blood metabolite levels between two metabolomics platforms

Few studies have examined the comparability of metabolite measurements across different metabolomics platforms when tested against split samples. To our knowledge, only two platforms used by COMETS cohorts (Metabolon and Biocrates) have had their metabolomics measures compared against one another this way. These two platforms had forty overlapping metabolites and moderately intercorrelated metabolites values (median correlation of ~0.5) (89, 90).

To expand our understanding of platform comparability, we sent split samples from the Health ABC cohort (83) to both Metabolon and the Broad Institute. In brief, each study participant had multiple vials of ethylenediaminetetraacetic acid (EDTA) plasma aliquoted during initial blood collection and stored at -80°C. We sent never-thawed aliquots from 40 African-American men in Health ABC to Metabolon for analysis on their Orbitrap Elite liquid chromatography MS platform (positive and negative ion mode) and gas chromatography MS. We also sent identically prepared aliquots from these men to the Broad Institute for analysis on their MS platforms (C8-positive ultra-performance liquid chromatography MS, hydrophilic interaction ultra-performance liquid chromatography positive ion mode MS, and hydrophilic interaction ultra-performance liquid chromatography negative ion mode MS). From Metabolon Inc., we received data on 610 named metabolites and from the Broad Institute, we received data on 347 named metabolites. We linked metabolite names across platforms using the HMDB identifiers, which Metabolon provided for 385 of its metabolites and the Broad Institute provided for 332 of its metabolites. To ascertain other potential overlapping metabolites, we separately evaluated all pairwise correlations across platforms and flagged metabolite pairs with high correlations

(i.e. Spearman correlation$\geq$ 0.7). More complete details on study participants, sample extractions, and instrumentation are provided in **Web Appendix.**

RESULTS

In total, the 47 cohorts included 137,047 participants with blood metabolomics measurements (**Table 1**), with numbers still likely to grow further. For most cohorts, participant enrollment and blood sample collection occurred during the 1990s, although some cohorts collected samples earlier (e.g. the Nurses' Health Study in 1989) (77). Follow-up for disease outcomes is still ongoing for nearly all studies.

Selected baseline characteristics of participants of each cohort are summarized in **Table 2**. Of the 137,047 participants with metabolome profiles, 82,142 (59.9%) were women. The distribution of different groups of ethnic ancestry was 70.2% European, 17.6% Asian (13.7% East Asian and 3.9% South Asian), 5.8% African, 1.8% Hispanic, 0.5% Native Hawaiian and 4.1% other mixed population. Study participants ranged from 0 (newborn) to 100 years of age at the time of blood collection, with a median age of 51 years.

COMETS cohorts use both active and passive follow-up methods to track participants longitudinally for disease outcomes like diabetes mellitus, heart-disease and cancer (**Web Table 2**). Forty-six of 47 COMETS cohorts use active follow-up methods, including tracking outcomes through mailed questionnaires, phone calls, or during follow-up visits. For active follow-up methods, each cohort further verifies outcomes through medical record review. Thirty-four of the 47 cohorts also utilize passive follow-up methods, such as linkages to electronic health records from hospitalization, or registries for cancer or death (e.g. National Death Index), which helps to ensure complete and objective follow-up. Our review of passive follow-up methods indicates that U.S. cancer registries for our cohorts are 95+% complete (91) and that the U.S. National Death Index is 93-98% complete (92, 93). For

European cohorts, cancer registries are 90+% complete for 90% of registries (self-audited) (94) and vital status is ~98% complete, according to European Prospective Investigation into Cancer and Nutrition (EPIC) data (95). Across the 47 participating cohorts in COMETS, the median follow-up for disease outcomes was 23 years.

Details on the blood samples and metabolomics platforms used for each cohort are presented in **Table 3**. Blood samples for metabolomics profiling primarily include serum (23 out of 47 cohorts) or plasma (31 out of 47 cohorts). Samples were predominantly collected at study baseline and include fasted-only samples (23 cohorts), non-fasted samples (10 cohorts), or a mix of fasted and non-fasted samples (14 cohorts). Seventeen metabolomics labs were used by COMETS cohorts, with the most heavily-used platforms being those of Metabolon Inc. (14 cohorts), the Broad Institute (15 cohorts), and Nightingale Health (11 cohorts). After accounting for use of multiple platforms, 34 of the 47 cohorts in total used at least one of these three platforms. Other platforms include, but are not limited to, Biocrates, Imperial College London National Phenome Centre, Duke Molecular Physiology Institute, and the West Coast Metabolomics Center.

Each cohort study collected data on demographics and health-related participant characteristics during study visits and/or through questionnaires (**Table 4**). Overall, 46 cohorts in COMETS assessed smoking status, 45 cohorts asked about alcohol intake, 35 cohorts inquired about leisure-time physical activity, and 36 cohorts ascertained diet. Forty-four cohorts evaluated educational levels and/or other measures of socioeconomic position. All forty-seven cohorts collected body mass index and 38 cohorts assessed waist circumference. Many cohort studies also included clinical measurements, such as systolic and diastolic blood pressure (n=41), lipid profiles (n=39), inflammation markers (n=38) and fasting glucose (n=37) (**Table 5)**. In addition to traditional clinical information, genome-wide single nucleotide polymorphism data are available for about 68% COMETS participants (93,082 out of 137,047).

In our first feasibility study, there were 1,874 metabolites measured across the five platforms tested. Of these, 1,550 had assigned identities, and 1,111 also had unique identifiers from HMDB, Pubchem, or other online databases that allowed us to match across platforms. A complete listing of each metabolite, the platforms it was measured on, and other details is in **Web Table 3**.

The specific numbers of metabolites by platform (cohort) were as follows: Metabolon, Inc. (the Atherosclerosis Risk in Communities Study): 1158 (includes 293 unidentified metabolites); Broad Institute (the Health, Aging and Body Composition Study): 350; Biocrates (Fenland Study): 187; West Coast Metabolomics Center WCMC (the ColoCare Study): 439; and Nightingale Health (the Prostate, Lung, Colorectal, and Ovarian Cancer Screening Trial): 236 (**Table 6**). The overlap in metabolites between platforms ranged from moderate (e.g. ~100 metabolites) to modest (e.g. ~20 metabolites). For example, for Metabolon Inc. the overlap in metabolites with other platforms was as follows: Broad Institute (HABC): 121; Biocrates (Fenland): 24; West Coast Metabolomics Center (ColoCare): 92; Nightingale Health (PLCO): 16. For the three platforms used most often by COMETS cohorts—Metabolon, the Broad Institute and Nightingale Health—14 metabolites were measured in common by all three.

Only two of the five metabolomics platforms, Nightingale Health and Biocrates, quantified any metabolites in terms of absolute concentrations. In total, they quantified 31 metabolites: Nightingale Health quantified 25 metabolites, Biocrates quantified 14 metabolites, and eight of these metabolites were quantified in common on both platforms (as listed in Web Table 3).

In our second feasibility study, 111 metabolites overlapped between Metabolon and the Broad Institute and their values were moderately to strongly correlated. Specifically, over the 111 metabolites, the median Spearman correlation across platforms was 0.79 and the interquartile range was 0.56 to 0.89 (**Figure 2**; **Web Table 4**). Pearson correlations were similar (median=0.78; interquartile range=0.65, 0.91). Given some minor measurement error and different techniques for each platform, these

correlations are high. Beyond the 111 overlapping metabolites, we found another 37 metabolite pairs with strongly correlated values (**Web Table 5**). These were biologically-interrelated metabolites (e.g. lactose and maltose) rather than identical metabolites, suggesting our match on HMDB identifiers was reasonably complete.

DISCUSSION

In this report, we described key details of COMETS which, with more than 137,000 participants, is the world's largest metabolomics consortium. Our survey found that COMETS captures a broad range of demographics, with many women (59.9% of participants), younger and older participants (range of 0 to 100 years), and diverse geography (many participants from each of North America, South America, Europe, and Asia). Key questionnaire data needed for epidemiologic research, e.g. smoking status, were collected by nearly all COMETS cohorts and many also assessed physical or clinical measures of interest, as well as genome-wide association study data. The breadth of demographics and available exposure data provide a strong foundation for the conduct of epidemiologic research.

With respect to the metabolomics assays, three labs in particular predominated: Metabolon Inc., the Broad Institute, and/or Nightingale Health. Each lab was used by 10 or more cohorts, and consequently tens of thousands of COMETS participants have data for the metabolites that each of these platforms measure.

In our comparative assessment, we found that platforms overlapped only modestly in the metabolites measured. For example, of the aggregate 1,421 metabolites measured by Metabolon, the Broad Institute and Nightingale Health, only 126 metabolites were measured by at least two platforms, and only 14 metabolites were measured by all three. For many metabolites, then, meta-analyses will be restricted to participants analyzed on a single specific platform, resulting in lower sample size and statistical power than if all platforms had measured all metabolites. We also found that few metabolites

were measured on a fully quantitative basis (i.e. as absolute concentrations)—just 31 across all five platforms. This precludes comparing metabolite levels across cohorts, or direct pooling of data, though meta-analyses are still possible.

One challenge we faced in this comparative assessment was that 28% of identified metabolites (439 of 1,550 entries) did not have assigned identifiers in public databases like the HMDB. Lacking this key information, we were unable to match these metabolites to others, possibly resulting in an undercount of platform overlap. Additionally, some platforms make distinctions between biochemically similar metabolites that other platforms do not, which can complicate match attempts. For example, Metabolon measures two different forms of 3-methylglutarylcarnitine, Biocrates measures one generic 3-methylglutarylcarnitine, and all three measures link to the same HMDB identifier. Consequently, none of three measures "matched", and they are recorded as three separate entries in our metabolite table.

As metabolomics platforms develop, we anticipate that metabolite linkage will improve, and platform overlap will grow. The establishment of data repositories such as Metabolomics Workbench(96) and MetaboLights (97) have accelerated the rate of data- and meta-data sharing between labs, fostering greater standardization in metabolomics analyses (98) and improving metabolite coverage. Additionally, as labs move toward newer, higher-sensitivity analytical technologies, such as Q Exactive Mass Spectrometry, more metabolites will be measured, resulting in more overlap.

To mitigate issues arising from lack of full quantitation, COMETS is developing a reference sample set of serum and EDTA plasma samples from each of 40 people, including 10 Hispanics, 10 Asians, 10 Blacks, and 10 Whites. We intend to embed one aliquot per person among any new large COMETS studies (e.g. 1000+ samples), with the resulting metabolomics data to be deposited in a central repository. These common samples should facilitate comparisons of metabolite levels (99) across

studies and enable pooled analyses for some metabolites, particularly those measured on a fully quantitative basis.

For two metabolomics platforms—Metabolon and the Broad Institute—we compared the values for 111 metabolites obtained in split samples and found them to be highly intercorrelated. This suggests that these platforms should yield comparable results in statistical analyses based on ranked levels of metabolites, such as Spearman correlations or quantile-based analyses. Such high correlations do not guarantee agreement of absolute concentrations, however (100), which may be a prerequisite for performing some kinds of statistical analyses. We could not evaluate agreement directly in this comparison because the units of measurement differ between platforms (neither provides absolute concentrations). In the future, we will continue evaluating comparability of other metabolomics platforms used by COMETS cohorts, such as by using the reference sample set discussed above.

As a consortium, COMETS has several distinctive strengths. First, to our knowledge, it is the world's largest consortium of metabolomics cohort studies. The large sample size will enable well-powered statistical analyses and/or permit rapid replication of study findings, helping to minimize false positives in this research area. Second, COMETS is a multi-ethnic international consortium that includes populations from Asia, Europe, North America and South America, and both children and adults. The diversity of study populations increases the range of exposures that can be studied within COMETS and makes it possible to assess associations within a wide range of demographic and socioeconomic groups. Additionally, because confounding patterns vary by population, evidence that associations consistently replicate across diverse populations may reassure researchers that results do not simply reflect confounding (101). Third, the large scope of COMETS makes it possible to flag associations that vary by platform and may therefore be influenced by measurement error (e.g. random noise) or more fundamental errors (e.g. misidentified metabolites). By communicating this information to the laboratories, it may help them to improve the quality and consistency of their measurements. Lastly,

COMETS brings together expertise from multiple disciplines relevant to conducting successful metabolomics research, which could help to drive forward methodologic advances in this field.

COMETS has limitations as well. The metabolomics platforms used by participating studies vary in their sample preparation, instrumentation, and, consequently, in the metabolites they measure. Additionally, metabolite levels in many COMETS cohorts, and indeed in most high-throughput metabolomics profiling studies, are semi-quantitative rather than fully quantitative concentrations. For important associations identified in COMETS, further follow-up work will be needed to establish clinically meaningful concentration thresholds. Also, COMETS is restricted to analyses based on identified metabolites, as raw nuclear magnetic resonance or MS peak data may not consistently align across platforms or different studies (102). Future efforts will aim to integrate data from unidentified metabolites and/or raw nuclear magnetic resonance or MS peaks. At present, COMETS is restricted to blood metabolomics data. Metabolomics data from urine and other biospecimen types will be added in the future, once initial blood-based analyses are complete. Another limitation is that our comparison of metabolite values is valid for only the two platforms tested. Whether other platforms would also provide comparable results still needs to be empirically tested. Finally, COMETS cohorts vary in their depth and breadth of coverage for health-related characteristics, thus for proposals requiring unusual data, some cohorts may be unable to contribute.

The primary objective of COMETS is to engage researchers in collaborative efforts to advance knowledge of the metabolome and its relationship with disease etiology, diagnosis, treatment and prognosis. In that spirit, we invite cohort studies with metabolomics data to join COMETS and we welcome data analysis proposals from interested scientific investigators, including those without data of their own. Information about how to join COMETS, and/or how to propose a data analysis can be found at our website (103).

Acknowledgments

The authors thank the staff and participants of the ARIC study for their important contributions. The ALSPAC investigators are extremely grateful to all the families who took part in this study, the midwives for their help in recruiting them, and the whole ALSPAC team, which includes interviewers, computer and laboratory technicians, clerical workers, research scientists, volunteers, managers, receptionists and nurses. Born in Bradford is only possible because of the enthusiasm and commitment of the Children and Parents in BiB. We are grateful to all the participants, practitioners and researchers who have made Born in Bradford happen. We thank all BWHHS participants, the general practitioners and their staff who have supported data collection since the study inception. The authors express sincere appreciation to all Cancer Prevention Study II participants and to each member of the study and biospecimen management group. We gratefully acknowledge the contribution of all members of field staff conducting the KORA study. The authors gratefully acknowledge the help of the MRC Epidemiology Unit, Field Teams, Laboratory Team, Data Management Team, and all other Support Teams. The authors are grateful to NSHD study members for their continuing support. We thank the residents of Tsuruoka City for their interest in our study and the members of the Tsuruoka Metabolomic Cohort Study team for their commitment to the project. The NHS, NHSII and HPFS investigators would like to thank the participants and staff of the studies for their valuable contributions as well as the following state cancer registries for their help: AL, AZ, AR, CA, CO, CT, DE, FL, GA, ID, IL, IN, IA, KY, LA, ME, MD, MA, MI, NE, NH, NJ, NY, NC, ND, OH, OK, OR, PA, RI, SC, TN, TX, VA, WA, WY. The Whitehall II research team thanks all of the participating civil service departments and their welfare, personnel, and establishment officers, study coordinators, nurses, data managers, administrative assistants and data entry staff, who make the study possible. The authors assume full responsibility for analyses and interpretation of these data.

References

1. Wang TJ, Larson MG, Vasan RS, et al. Metabolite profiles and the risk of developing diabetes. *Nat Med* 2011;17(4):448-53.
2. Floegel A, Stefan N, Yu Z, et al. Identification of serum metabolites associated with risk of type 2 diabetes using a targeted metabolomic approach. *Diabetes* 2013;62(2):639-48.
3. Menni C, Fauman E, Erte I, et al. Biomarkers for type 2 diabetes and impaired fasting glucose using a nontargeted metabolomics approach. *Diabetes* 2013;62(12):4270-6.
4. Yu D, Moore SC, Matthews CE, et al. Plasma metabolomic profiles in association with type 2 diabetes risk and prevalence in Chinese adults. *Metabolomics* 2016;12.
5. Tang WH, Wang Z, Levison BS, et al. Intestinal microbial metabolism of phosphatidylcholine and cardiovascular risk. *N Engl J Med* 2013;368(17):1575-84.
6. Shah SH, Bain JR, Muehlbauer MJ, et al. Association of a peripheral blood metabolic profile with coronary artery disease and risk of subsequent cardiovascular events. *Circ Cardiovasc Genet* 2010;3(2):207-14.
7. Kraus WE, Muoio DM, Stevens R, et al. Metabolomic Quantitative Trait Loci (mQTL) Mapping Implicates the Ubiquitin Proteasome System in Cardiovascular Disease Pathogenesis. *PLoS Genet* 2015;11(11):e1005553.
8. Sekula P, Goek ON, Quaye L, et al. A Metabolome-Wide Association Study of Kidney Function and Disease in the General Population. *J Am Soc Nephrol* 2016;27(4):1175-88.
9. Mayers JR, Wu C, Clish CB, et al. Elevation of circulating branched-chain amino acids is an early event in human pancreatic adenocarcinoma development. *Nat Med* 2014;20(10):1193-8.
10. Mondul AM, Moore SC, Weinstein SJ, et al. Metabolomic analysis of prostate cancer risk in a prospective cohort: The alpha-tocolpherol, beta-carotene cancer prevention (ATBC) study. *Int J Cancer* 2015;137(9):2124-32.

11. Kuhn T, Floegel A, Sookthai D, et al. Higher plasma levels of lysophosphatidylcholine 18:0 are related to a lower risk of common cancers in a prospective metabolomics study. *BMC Med* 2016;14:13.

12. Huang J, Weinstein SJ, Kitahara CM, et al. A prospective study of serum metabolites and glioma risk. *Oncotarget* 2017;8(41):70366-77.

13. Moore SC, Playdon MC, Sampson JN, et al. A Metabolomics Analysis of Body Mass Index and Postmenopausal Breast Cancer Risk. *J Natl Cancer Inst* 2018;110(6):588-97.

14. Shin SY, Fauman EB, Petersen AK, et al. An atlas of genetic influences on human blood metabolites. *Nat Genet* 2014;46(6):543-50.

15. Rhee EP, Ho JE, Chen MH, et al. A genome-wide association study of the human metabolome in a community-based cohort. *Cell Metab* 2013;18(1):130-43.

16. Long T, Hicks M, Yu HC, et al. Whole-genome sequencing identifies common-to-rare variants associated with human blood metabolites. *Nat Genet* 2017;49(4):568-78.

17. Yu B, Zheng Y, Alexander D, et al. Genetic determinants influencing human serum metabolome among African Americans. *PLoS Genet* 2014;10(3):e1004212.

18. Illig T, Gieger C, Zhai G, et al. A genome-wide perspective of genetic variation in human metabolism. *Nat Genet* 2010;42(2):137-41.

19. Yu B, Li AH, Metcalf GA, et al. Loss-of-function variants influence the human serum metabolome. *Sci Adv* 2016;2(8):e1600800.

20. Yu B, de Vries PS, Metcalf GA, et al. Whole genome sequence analysis of serum amino acid levels. *Genome Biol* 2016;17(1):237.

21. Menni C, Graham D, Kastenmuller G, et al. Metabolomic identification of a novel pathway of blood pressure regulation involving hexadecanedioate. *Hypertension* 2015;66(2):422-9.

22. Newgard CB, An J, Bain JR, et al. A branched-chain amino acid-related metabolic signature that differentiates obese and lean humans and contributes to insulin resistance. *Cell Metab* 2009;9(4):311-26.

23. Cheng S, Rhee EP, Larson MG, et al. Metabolite profiling identifies pathways associated with metabolic risk in humans. *Circulation* 2012;125(18):2222-31.

24. Moore SC, Matthews CE, Sampson JN, et al. Human metabolic correlates of body mass index. *Metabolomics* 2014;10(2):259-69.

25. Scalbert A, Brennan L, Manach C, et al. The food metabolome: a window over dietary exposure. *The American journal of clinical nutrition* 2014;99(6):1286-308.

26. Guertin KA, Moore SC, Sampson JN, et al. Metabolomics in nutritional epidemiology: identifying metabolites associated with diet and quantifying their potential to uncover diet-disease relations in populations. *The American journal of clinical nutrition* 2014;100(1):208-17.

27. Zheng Y, Yu B, Alexander D, et al. Human metabolome associates with dietary intake habits among African Americans in the atherosclerosis risk in communities study. *Am J Epidemiol* 2014;179(12):1424-33.

28. Playdon MC, Ziegler RG, Sampson JN, et al. Nutritional metabolomics and breast cancer risk in a prospective study. *The American journal of clinical nutrition* 2017;106(2):637-49.

29. Schmidt JA, Rinaldi S, Ferrari P, et al. Metabolic profiles of male meat eaters, fish eaters, vegetarians, and vegans from the EPIC-Oxford cohort. *The American journal of clinical nutrition* 2015;102(6):1518-26.

30. Pallister T, Jennings A, Mohney RP, et al. Characterizing Blood Metabolomics Profiles Associated with Self-Reported Food Intakes in Female Twins. *PLoS One* 2016;11(6):e0158568.

31. Schmidt JA, Rinaldi S, Scalbert A, et al. Plasma concentrations and intakes of amino acids in male meat-eaters, fish-eaters, vegetarians and vegans: a cross-sectional analysis in the EPIC-Oxford cohort. *Eur J Clin Nutr* 2016;70(3):306-12.

32. Mondul AM, Sampson JN, Moore SC, et al. Metabolomic profile of response to supplementation with beta-carotene in the Alpha-Tocopherol, Beta-Carotene Cancer Prevention Study. *The American journal of clinical nutrition* 2013;98(2):488-93.

33. Playdon MC, Sampson JN, Cross AJ, et al. Comparing metabolite profiles of habitual diet in serum and urine. *The American journal of clinical nutrition* 2016;104(3):776-89.

34. Akbaraly T, Wurtz P, Singh-Manoux A, et al. Association of circulating metabolites with healthy diet and risk of cardiovascular disease: analysis of two cohort studies. *Sci Rep* 2018;8(1):8620.

35. Nelson SM, Panagiotou OA, Anic GM, et al. Metabolomics analysis of serum 25-hydroxy-vitamin D in the Alpha-Tocopherol, Beta-Carotene Cancer Prevention (ATBC) Study. *Int J Epidemiol* 2016;45(5):1458-68.

36. Xiao Q, Moore SC, Keadle SK, et al. Objectively measured physical activity and plasma metabolomics in the Shanghai Physical Activity Study. *Int J Epidemiol* 2016;45(5):1433-44.

37. Wang Q, Wurtz P, Auro K, et al. Metabolic profiling of pregnancy: cross-sectional and longitudinal evidence. *BMC Med* 2016;14(1):205.

38. Wang Q, Wurtz P, Auro K, et al. Effects of hormonal contraception on systemic metabolism: cross-sectional and longitudinal evidence. *Int J Epidemiol* 2016;45(5):1445-57.

39. Wurtz P, Wang Q, Soininen P, et al. Metabolomic Profiling of Statin Use and Genetic Inhibition of HMG-CoA Reductase. *J Am Coll Cardiol* 2016;67(10):1200-10.

40. The alpha-tocopherol, beta-carotene lung cancer prevention study: design, methods, participant characteristics, and compliance. The ATBC Cancer Prevention Study Group. *Ann Epidemiol* 1994;4(1):1-10.

41. The Childhood Asthma Management Program (CAMP): design, rationale, and methods. Childhood Asthma Management Program Research Group. *Controlled clinical trials* 1999;20(1):91-120.

42. Diabetes Prevention Program Research G. Long-term effects of lifestyle intervention or metformin on diabetes development and microvascular complications over 15-year follow-up: the Diabetes Prevention Program Outcomes Study. *Lancet Diabetes Endocrinol* 2015;3(11):866-75.

43. Gaziano JM, Sesso HD, Christen WG, et al. Multivitamins in the prevention of cancer in men: the Physicians' Health Study II randomized controlled trial. *JAMA* 2012;308(18):1871-80.

44. Prorok PC, Andriole GL, Bresalier RS, et al. Design of the Prostate, Lung, Colorectal and Ovarian (PLCO) Cancer Screening Trial. *Controlled clinical trials* 2000;21(6 Suppl):273S-309S.

45. Litonjua AA, Lange NE, Carey VJ, et al. The Vitamin D Antenatal Asthma Reduction Trial (VDAART): rationale, design, and methods of a randomized, controlled trial of vitamin D supplementation in pregnancy for the primary prevention of asthma and allergies in children. *Contemp Clin Trials* 2014;38(1):37-50.

46. Cheng TY, Makar KW, Neuhouser ML, et al. Folate-mediated one-carbon metabolism genes and interactions with nutritional factors on colorectal cancer risk: Women's Health Initiative Observational Study. *Cancer* 2015;121(20):3684-91.

47. Oresic M, Simell S, Sysi-Aho 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* 2008;205(13):2975-84.

48. Briley AL, Barr S, Badger S, et al. A complex intervention to improve pregnancy outcome in obese women; the UPBEAT randomised controlled trial. *BMC Pregnancy Childbirth* 2014;14:74.

49. The Atherosclerosis Risk in Communities (ARIC) Study: design and objectives. The ARIC investigators. *Am J Epidemiol* 1989;129(4):687-702.

50. Clifton EA, Day FR, De Lucia Rolfe E, et al. Associations between body mass index-related genetic variants and adult body composition: The Fenland cohort study. *Int J Obes (Lond)* 2017;41(4):613-9.

51. Kolonel LN, Henderson BE, Hankin JH, et al. A multiethnic cohort in Hawaii and Los Angeles: baseline characteristics. *Am J Epidemiol* 2000;151(4):346-57.

52. Bild DE, Bluemke DA, Burke GL, et al. Multi-Ethnic Study of Atherosclerosis: objectives and design. *Am J Epidemiol* 2002;156(9):871-81.

53. Orwoll E, Blank JB, Barrett-Connor E, et al. Design and baseline characteristics of the osteoporotic fractures in men (MrOS) study--a large observational study of the determinants of fracture in older men. *Contemp Clin Trials* 2005;26(5):569-85.

54. Shu XO, Li H, Yang G, et al. Cohort Profile: The Shanghai Men's Health Study. *Int J Epidemiol* 2015;44(3):810-8.

55. Shah SH, Newgard CB. Integrated metabolomics and genomics: systems approaches to biomarkers and mechanisms of cardiovascular disease. *Circ Cardiovasc Genet* 2015;8(2):410-9.

56. Dale CE, Bowling A, Adamson J, et al. Predictors of patterns of change in health-related quality of life in older women over 7 years: evidence from a prospective cohort study. *Age Ageing* 2013;42(3):312-8.

57. Bainton D, Miller NE, Bolton CH, et al. Plasma triglyceride and high density lipoprotein cholesterol as predictors of ischaemic heart disease in British men. The Caerphilly and Speedwell Collaborative Heart Disease Studies. *Br Heart J* 1992;68(1):60-6.

58. Tillin T, Forouhi NG, McKeigue PM, et al. Southall And Brent REvisited: Cohort profile of SABRE, a UK population-based comparison of cardiovascular disease and diabetes in people of European, Indian Asian and African Caribbean origins. *Int J Epidemiol* 2012;41(1):33-42.

59. Kuh D, Pierce M, Adams J, et al. Cohort profile: updating the cohort profile for the MRC National Survey of Health and Development: a new clinic-based data collection for ageing research. *Int J Epidemiol* 2011;40(1):e1-9.

60. Marmot M, Brunner E. Cohort Profile: the Whitehall II study. *Int J Epidemiol* 2005;34(2):251-6.

61. Nang EE, Khoo CM, Tai ES, et al. Is there a clear threshold for fasting plasma glucose that differentiates between those with and without neuropathy and chronic kidney disease?: the Singapore Prospective Study Program. *Am J Epidemiol* 2009;169(12):1454-62.

62. Leitsalu L, Haller T, Esko T, et al. Cohort Profile: Estonian Biobank of the Estonian Genome Center, University of Tartu. *Int J Epidemiol* 2015;44(4):1137-47.

63. Kannel WB, Feinleib M, McNamara PM, et al. An investigation of coronary heart disease in families. The Framingham offspring study. *Am J Epidemiol* 1979;110(3):281-90.

64. Tsao CW, Vasan RS. The Framingham Heart Study: past, present and future. *Int J Epidemiol* 2015;44(6):1763-6.

65. Calle EE, Rodriguez C, Jacobs EJ, et al. The American Cancer Society Cancer Prevention Study II Nutrition Cohort: rationale, study design, and baseline characteristics. *Cancer* 2002;94(2):500-11.

66. Riboli E, Hunt KJ, Slimani N, et al. European Prospective Investigation into Cancer and Nutrition (EPIC): study populations and data collection. *Public Health Nutr* 2002;5(6B):1113-24.

67. Harada S, Takebayashi T, Kurihara A, et al. Metabolomic profiling reveals novel biomarkers of alcohol intake and alcohol-induced liver injury in community-dwelling men. *Environ Health Prev Med* 2016;21(1):18-26.

68. Liesenfeld DB, Grapov D, Fahrmann JF, et al. Metabolomics and transcriptomics identify pathway differences between visceral and subcutaneous adipose tissue in colorectal cancer patients: the ColoCare study. *The American journal of clinical nutrition* 2015;102(2):433-43.

69. Kraus WE, Granger CB, Sketch MH, Jr., et al. A Guide for a Cardiovascular Genomics Biorepository: the CATHGEN Experience. *J Cardiovasc Transl Res* 2015;8(8):449-57.

70. Price JF, Reynolds RM, Mitchell RJ, et al. The Edinburgh Type 2 Diabetes Study: study protocol. *BMC Endocr Disord* 2008;8:18.

71. Barrios C, Zierer J, Wurtz P, et al. Circulating metabolic biomarkers of renal function in diabetic and non-diabetic populations. *Sci Rep* 2018;8(1):15249.

72. John EM, Hopper JL, Beck JC, et al. The Breast Cancer Family Registry: an infrastructure for cooperative multinational, interdisciplinary and translational studies of the genetic epidemiology of breast cancer. *Breast Cancer Res* 2004;6(4):R375-89.

73. de Oliveira CM, Pereira AC, de Andrade M, et al. Heritability of cardiovascular risk factors in a Brazilian population: Baependi Heart Study. *BMC Med Genet* 2008;9:32.

74. Bacon MC, von Wyl V, Alden C, et al. The Women's Interagency HIV Study: an observational cohort brings clinical sciences to the bench. *Clin Diagn Lab Immunol* 2005;12(9):1013-9.

75. Qi Q, Hua S, Clish CB, et al. Plasma Tryptophan-Kynurenine Metabolites Are Altered in Human Immunodeficiency Virus Infection and Associated With Progression of Carotid Artery Atherosclerosis. *Clin Infect Dis* 2018;67(2):235-42.

76. Wilson KM, Kasperzyk JL, Rider JR, et al. Coffee consumption and prostate cancer risk and progression in the Health Professionals Follow-up Study. *J Natl Cancer Inst* 2011;103(11):876-84.

77. Colditz GA, Hankinson SE. The Nurses' Health Study: lifestyle and health among women. *Nat Rev Cancer* 2005;5(5):388-96.

78. Elliott P, Vergnaud AC, Singh D, et al. The Airwave Health Monitoring Study of police officers and staff in Great Britain: rationale, design and methods. *Environ Res* 2014;134:280-5.

79. Wright J, Small N, Raynor P, et al. Cohort Profile: the Born in Bradford multi-ethnic family cohort study. *Int J Epidemiol* 2013;42(4):978-91.

80. Pasupathy D, Dacey A, Cook E, et al. Study protocol. A prospective cohort study of unselected primiparous women: the pregnancy outcome prediction study. *BMC Pregnancy Childbirth* 2008;8:51.

81. Boyd A, Golding J, Macleod J, et al. Cohort Profile: the 'children of the 90s'--the index offspring of the Avon Longitudinal Study of Parents and Children. *Int J Epidemiol* 2013;42(1):111-27.

82. Nanto-Salonen K, Kupila A, Simell S, et al. Nasal insulin to prevent type 1 diabetes in children with HLA genotypes and autoantibodies conferring increased risk of disease: a double-blind, randomised controlled trial. *Lancet* 2008;372(9651):1746-55.

83. Murphy RA, Moore SC, Playdon M, et al. Metabolites Associated With Lean Mass and Adiposity in Older Black Men. *J Gerontol A Biol Sci Med Sci* 2017;72(10):1352-9.

84. Moayyeri A, Hammond CJ, Valdes AM, et al. Cohort Profile: TwinsUK and healthy ageing twin study. *Int J Epidemiol* 2013;42(1):76-85.

85. Chow WH, Chrisman M, Daniel CR, et al. Cohort Profile: The Mexican American Mano a Mano Cohort. *Int J Epidemiol* 2017;46(2):e3.

86. Wishart DS, Feunang YD, Marcu A, et al. HMDB 4.0: the human metabolome database for 2018. *Nucleic Acids Res* 2018;46(D1):D608-D17.

87. Kim S, Chen J, Cheng T, et al. PubChem 2019 update: improved access to chemical data. *Nucleic Acids Res* 2019;47(D1):D1102-D9.

88. Pence HE, Williams A. ChemSpider: An Online Chemical Information Resource. *J Chem Educ* 2010;87(11):1123-4.

89. Yet I, Menni C, Shin SY, et al. Genetic Influences on Metabolite Levels: A Comparison across Metabolomic Platforms. *PLoS One* 2016;11(4):e0153672.

90. Suhre K, Meisinger C, Doring A, et al. Metabolic footprint of diabetes: a multiplatform metabolomics study in an epidemiological setting. *PLoS One* 2010;5(11):e13953.

91. North American Association of Central Cancer Registries. https://www.naaccr.org/certified-registries/. Published July 7, 2017. Accessed November 1, 2018.

92. Rich-Edwards JW, Corsano KA, Stampfer MJ. Test of the National Death Index and Equifax Nationwide Death Search. *Am J Epidemiol* 1994;140(11):1016-9.

93. Calle EE, Terrell DD. Utility of the National Death Index for ascertainment of mortality among cancer prevention study II participants. *Am J Epidemiol* 1993;137(2):235-41.

94. Zanetti R, Schmidtmann I, Sacchetto L, et al. Completeness and timeliness: Cancer registries could/should improve their performance. *Eur J Cancer* 2015;51(9):1091-8.

95. Rohrmann S, Overvad K, Bueno-de-Mesquita HB, et al. Meat consumption and mortality--results from the European Prospective Investigation into Cancer and Nutrition. *BMC Med* 2013;11:63.

96. Sud M, Fahy E, Cotter D, et al. Metabolomics Workbench: An international repository for metabolomics data and metadata, metabolite standards, protocols, tutorials and training, and analysis tools. *Nucleic Acids Res* 2016;44(D1):D463-70.

97. Haug K, Salek RM, Conesa P, et al. MetaboLights--an open-access general-purpose repository for metabolomics studies and associated meta-data. *Nucleic Acids Res* 2013;41(Database issue):D781-6.

98. Sansone SA, Rocca-Serra P, Field D, et al. Toward interoperable bioscience data. *Nat Genet* 2012;44(2):121-6.

99. Siskos AP, Jain P, Romisch-Margl W, et al. Interlaboratory Reproducibility of a Targeted Metabolomics Platform for Analysis of Human Serum and Plasma. *Anal Chem* 2017;89(1):656-65.

100. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* 1986;1(8476):307-10.

101. Lawlor DA, Tilling K, Davey Smith G. Triangulation in aetiological epidemiology. *Int J Epidemiol* 2016;45(6):1866-86.

102. Tzoulaki I, Ebbels TM, Valdes A, et al. Design and analysis of metabolomics studies in epidemiologic research: a primer on -omic technologies. *Am J Epidemiol* 2014;180(2):129-39.

103. Division of Cancer Control and Population Sciences, National Cancer Institute. Consortium of Metabolomics Studies. https://epi.grants.cancer.gov/comets/. Accessed January 28, 2019.

104. Blank JB, Cawthon PM, Carrion-Petersen ML, et al. Overview of recruitment for the osteoporotic fractures in men study (MrOS). *Contemp Clin Trials* 2005;26(5):557-68.

105. Miller JW, Beresford SA, Neuhouser ML, et al. Homocysteine, cysteine, and risk of incident colorectal cancer in the Women's Health Initiative observational cohort. *The American journal of clinical nutrition* 2013;97(4):827-34.

106. Padilha K, Venturini G, de Farias Pires T, et al. Serum metabolomics profile of type 2 diabetes mellitus in a Brazilian rural population. *Metabolomics* 2016;12(10):156.

107. Menni C, Zhai G, Macgregor A, et al. Targeted metabolomics profiles are strongly correlated with nutritional patterns in women. *Metabolomics* 2013;9(2):506-14.

108. Gathungu RM, Bird SS, Sheldon DP, et al. Identification of metabolites from liquid chromatography-coulometric array detection profiling: gas chromatography-mass spectrometry and refractionation provide essential information orthogonal to LC-MS/microNMR. *Anal Biochem* 2014;454:23-32.

109. Zhu J, Djukovic D, Deng L, et al. Colorectal cancer detection using targeted serum metabolic profiling. *J Proteome Res* 2014;13(9):4120-30.

110. Edmands WM, Ferrari P, Rothwell JA, et al. Polyphenol metabolome in human urine and its association with intake of polyphenol-rich foods across European countries. *The American journal of clinical nutrition* 2015;102(4):905-13.

111. Chan Q, Loo RL, Ebbels TM, et al. Metabolic phenotyping for discovery of urinary biomarkers of diet, xenobiotics and blood pressure in the INTERMAP Study: an overview. *Hypertens Res* 2017;40(4):336-45.

112. Soga T, Igarashi K, Ito C, et al. Metabolomic profiling of anionic metabolites by capillary electrophoresis mass spectrometry. *Anal Chem* 2009;81(15):6165-74.

113. Evans AM BB, Liu Q, Mitchell MW, Robinson RJ, Dai H, Stewart SJ, DeHaven CD, Miller LAD. High Resolution Mass Spectrometry Improves Data Quantity and Quality as Compared to Unit Mass Resolution Mass Spectrometry in High-Throughput Profiling Metabolomics. *Metabolomics* 2014;4(2):132.

114. Soininen P, Kangas AJ, Wurtz P, et al. Quantitative serum nuclear magnetic resonance metabolomics in cardiovascular epidemiology and genetics. *Circ Cardiovasc Genet* 2015;8(1):192-206.

115. Wikoff WR, Hanash S, DeFelice B, et al. Diacetylspermine Is a Novel Prediagnostic Serum Biomarker for Non-Small-Cell Lung Cancer and Has Additive Performance With Pro-Surfactant Protein B. *J Clin Oncol* 2015;33(33):3880-6.

Table 1. Studies Participating in COMETS and the Number of Participants with Metabolomics Data

First Author, Year (Reference No.)	Study Name[a]	Study Acronym	Region	Baseline Examination Date[b]	Latest Follow-up Year	No. With Metabolomics[c]
Elliott, 2014 (78)	Airwave Health Monitoring Study	AIRWAVE	Europe	2004	Ongoing	4,000
The ATBC cancer prevention study group, 1994 (40)	Alpha-Tocopherol, Beta-Carotene Cancer Prevention Study	ATBC	Europe	1985-1988	Ongoing	950
The ARIC investigators, 1989 (49)	Atherosclerosis Risk in Communities Study	ARIC	North America	1987-1989	Ongoing	4,032
Boyd, 2013 (81)	Avon Longitudinal Study of Parents and Children	ALSPAC	Europe	1990-1993	Ongoing	4,572 mothers 7,178 offspring
de Oliveira, 2008 (73)	Baependi Heart Study	BHS	South America	2010-2013	Ongoing	939
Wright, 2013 (79)	Born in Bradford	BIB	Europe	2007-2011	Ongoing	10,000 mothers
John, 2004 (72)	Breast Cancer Family Registry	BCFR	North America	1995	2017	100
Dale, 2013 (56)	British Women's Heart & Health Study	BWHHS	Europe	1999-2001	Ongoing	3,780
Bainton, 1992 (57)	Caerphilly Prospective Study	CaPS	Europe	1989-1993	Ongoing	1,230
Calle, 2002 (65)	Cancer Prevention Study-II	CPS-II	North America	1992-1993	Ongoing	2,266
Kraus, 2015 (69)	Catherization Genetics	CATHGEN	North America	2001-2010	Ongoing	3,869
Childhood Asthma Management Program Research Group, 1999 (41)	Childhood Asthma Management Program	CAMP	North America	1991	1999	1,041
Liesenfeld, 2015 (68)	ColoCare	COLO	Europe & North America	2010-2017	Ongoing	359
Illig, 2010 (18)	Cooperative Health Research in the Region of Augsburg	KORA	Europe	1986	2009	3,000
Oresic, 2008 (47)	Diabetes Prediction and Prevention Birth Cohort	DIPP	Europe	1994-2017	Ongoing	534
Diabetes Prevention Program Research Group, 2015 (42)	Diabetes Prevention Program and Diabetes Prevention Program Outcomes Study	DPP	North America	1996-1999	Ongoing	2,015

Price, 2008 (70)	Edinburgh Type 2 Diabetes Study	ET2DS	Europe	2006-2007	Ongoing	1,060
Leitsalu, 2015 (62)	Estonia Biobank Obesity Extremes	Estonia OE	Europe	2003-2010	Ongoing	298
Riboli, 2002 (66)	European Prospective Investigation into Cancer and Nutrition	EPIC	Europe	1992-2000	Ongoing	15,000
Clifton, 2017 (50)	Fenland Study	Fenland	Europe	2005/2015	Ongoing	10,555
Framingham Heart Study, Gen 2 (63, 64)	Framingham Heart Study, Generation 2	FHS2	North America	1971	Ongoing	2,526
Kannel, 1979 and Tsao, 2015 (63, 64)	Framingham Heart Study, Generation 3	FHS3	North America	2002	Ongoing	998
Barrios, 2018 (71)	GenodiabMar	GDM	Europe	2012-2014	Ongoing	656
Murphy, 2017 (83)	Health, Aging and Body Composition	HABC	North America	1997-1998	Ongoing	319
Wilson, 2011 (76)	Health Professionals Follow-up Study	HPFS	North America	1993-1995	Ongoing	1,059
Chow, 2015 (85)	Mano-A-Mano, the Mexican American Cohort	MAC	North America	2001-2017	Ongoing	300
Kuh, 2011 (59)	MRC National Survey of Health & Development	MRC NSHD	Europe	2006-2010	Ongoing	1,790
Orwoll, 2005 and Blank, 2005 (53, 104)	MrOS-Osteoporotic Fractures in Men	MrOS	North America	2000-2002	Ongoing	1,400
Kolonel, 2000 (51)	The Multiethnic Cohort	MEC	North America	1993-1996	Ongoing	5,436
Bild, 2002 (52)	Multi-ethnic Study of Atherosclerosis	MESA	North America	2000-2002	Ongoing	3,831
Colditz, 2005 (77)	Nurses' Health Study	NHS	North America	1989-1990	Ongoing	1,200
Colditz, 2005 (77)	Nurses' Health Study II	NHS-II	North America	1996-1999	Ongoing	693
Gaziano, 2012 (43)	Physicians' Health Study	PHS	North America	1982-1984	Ongoing	224
Pasupathy, 2008 (80)	Pregnancy Outcome Prediction study	POPS	Europe	2008-2012	Ongoing	923
Prorok, 2000 (44)	Prostate, Lung, Colorectal, and Ovarian Cancer Screening Trial	PLCO	North America	1993-2001	Ongoing	1,742
Shu, 2015 (54)	Shanghai Men's Health Study	SMHS	Asia	1987-2000	Ongoing	1,006
Xiao, 2016 (36)	Shanghai Physical Activity Study	SPA	Asia	2005-2007	Ongoing	339

Yu, 2016 (4)	Shanghai Women's Health Study	SWHS	Asia	2001-2006	Ongoing	1,990
Nang, 2009 (61)	Singapore Prospective Study Program	SP2	Asia	2004-2007	2016	2,334
Tillin, 2012 (58)	Southall And Brent Revisited	SABRE	Europe	1988-1990	Ongoing	3,304
Harada, 2016 (67)	Tsuruoka Metabolomics Cohort Study	TMCS	Asia	2012-2015	Ongoing	10,957
Moayyeri, 2013 (84)	Twins United Kingdom	TwinsUK	Europe	1992	Ongoing	7,234
Briley, 2014 (48)	United Kingdom Pregnancies Better Eating and Activity Trial	UPBEAT	Europe	2009	2012-2013	1,303
Litonjua, 2014 (45)	Vitamin D Antenatal Asthma Reduction Trial	VDAART	North America	2009-2011	Ongoing	651
Marmot, 2005 (60)	Whitehall II	WH-II	Europe	1997-1999	Ongoing	4,762
Cheng, 2015 and Miller, 2013 (46, 105)	Women's Health Initiative	WHI	North America	1993-1998	Ongoing	2,706
Bacon, 2005 and Qi, 2018 (74, 75)	Women's Interagency HIV Study	WIHS	North America	2004-2005	Ongoing	411

[a] Studies are listed in alphabetical order by their study full name. [b] Baseline for the assessment of metabolomics. This is the time period for which a blood sample was used to generate the metabolomic data but may or may not be the first assessment of the cohort. [c] In some studies, metabolomics data is available at multiple timepoints on the same individuals. For these studies, we report details at the earliest timepoint for which data are available.

Table 2. Descriptive Characteristics of Participants with Metabolomics Data in COMETS[a]

Study	Median age (year) at blood collection (range)	No. of Women (*n* = 82,142)	No. of Men (*n* = 54,905)	No. of European Ancestry (*n* = 96,143)	No. of African Ancestry (*n* = 7,898)	No. of Asian (*n* = 24,165)	No. of Other Ancestry (*n* = 8,841)
AIRWAVE	42 (19-65)	1,497	2,503	3,835	29	0	136
ATBC	57 (50-69)	0	950	950	0	0	0
ARIC	53 (44-66)	2420	1,612	1,553	2479	0	0
ALSPAC (mothers)	48 (45-51)	4,572	0	4,318	31	34	366
ALSPAC (offspring)	14 (8-18)	3,732	3,444	6315	50	51	760
BHS	45 (18-90)	557	382	0	0	0	939
BIB	27 (15-40)	10,000	0	4,200	220	5170	410
BCFR	52 (26-80)	100	0	100	0	0	0
BWHHS	69 (67-71)	3,780	0	3,780	0	0	0
CaPS	57 (45-59)	0	1,230	1,230	0	0	0
CPS-II	68 (53-83)	1,710	556	2,225	17	0	24
CATHGEN	60 (21-94)	1,577	2,292	2,755	802	0	312
CAMP	9 (5-13)	420	621	711	138	0	192
COLO	63 (51-75)	143	216	359	0	0	0
KORA	56 (25-74)	1,500	1,500	3,000	0	0	0
DIPP	0 (0-15)	294	240	534	0	0	0
DPP	52 (25-85)	1,336	679	1,158	376	0	481
ET2DS	68 (60-75)	530	530	1,060	0	0	0
Estonia OE	39 (20-64)	149	149	298	0	0	0
EPIC	58 (45-80)	7,000	8,000	15,000	0	0	0
Fenland	45 (30-60)	4,905	5,650	10,555	0	0	0
FHS2	55 (26-84)	1,320	1,206	2,526	0	0	0
FHS3	41 (19-72)	529	469	998	0	0	0
GDM	66 (44-94)	257	399	656	0	0	0
HABC	74 (70-79)	0	319	0	319	0	0
HPFS	52 (40-75)	0	1,059	1,006	32	0	21
MAC	38 (20-72)	300	0	0	0	0	300
MRC NSHD	53 (53-53)	895	895	1,790	0	0	0
MrOS	74 (65-100)	0	1,400	1,321	24	0	55
MEC	68 (47-86)	3,579	1,857	1,066	915	1,748	1,707
MESA	63 (44-84)	1,933	1,898	1,482	934	536	879
NHS	56 (43-69)	1,200	0	1,164	30	0	6
NHS-II	43 (32-54)	693	0	658	21	0	14
PHS	54 (40-85)	0	224	224	0	0	0

POPS	30 (16-48)	923	0	923	0	0	0
PLCO	65 (55-74)	1,492	250	1,700	42	0	0
SMHS	56 (40-75)	0	1,006	0	0	1,006	0
SPA	60 (40-74)	200	139	0	0	339	0
SWHS	56 (40-71)	1,990	0	0	0	1,990	0
SP2	47 (24-79)	1,247	1,087	0	0	2,334	0
SABRE	52 (40-70)	467	2,837	1,572	192	0	1,540
TMCS	62 (34-75)	5,844	5,113	0	0	10,957	0
TwinsUK	50 (16-82)	6,531	703	7,065	69	0	100
UPBEAT	31 (31-31)	1,303	0	820	311	0	172
VDAART	1 (1-1)	304	347	211	315	0	125
WH-II	65 (50-79)	1,619	3,143	4,762	0	0	0
WHI	68 (62-72)	2,706	0	2,235	295	0	176
WIHS	42 (38-47)	411	0	28	257	0	126

[a] Descriptive data are provided specifically for participants as of the date of blood sample collection. Number of participants in each study is shown in Table 1.

Table 3. Blood Samples and Laboratories Used for Metabolomics in COMETS[a]

Cohort	Type of blood specimen	Year of blood collection[b]	Fasted status	Lab(s) used[c]	Analytical technology
AIRWAVE	Serum + EDTA plasma	Baseline	Non-fasted	Metabolon, Inc., ICL NPC	LC-MS, NMR
ATBC	Serum	Baseline	Fasted	Metabolon, Inc.	GC-MS, LC-MS
ARIC	Serum	Baseline	Fasted	Metabolon, Inc.	GC-MS, LC-MS
ALSPAC	Serum	Baseline	Mostly Fasted (offspring at age 7 non-fasted)	Nightingale Health	NMR
BHS	Serum	Baseline	Fasted	Agilent COE	GC-MS
BIB	Serum + EDTA plasma	2007-2010	Fasted	Nightingale Health	NMR
BCFR	EDTA plasma	Baseline	Non-fasted	Metabolon, Inc.	LC-MS
BWHHS	Serum	1999-2001	Fasted	Nightingale Health	NMR
CaPS	Serum	1989-1993	Fasted	Nightingale Health	NMR
CPS-II	Serum + EDTA Plasma	1998-2001	Non-fasted	Metabolon, Inc.	LC-MS
CATHGEN	EDTA plasma	Baseline	Fasted	Duke University	GC-MS, LC-MS
CAMP	Serum	Baseline	Non-fasted	Broad Institute	LC-MS
COLO	EDTA plasma	Baseline	Fasted + Non-fasted	IARC, WCMC	GC-MS, LC-MS
KORA	Serum	Baseline	Fasted + Non-fasted	Metabolon, Inc., Biocrates	GC-MS, LC-MS
DIPP	Serum + EDTA plasma	Baseline	Non-fasted	Örebro University	GC-MS, LC-MS
DPP	EDTA plasma	Baseline	Fasted	Broad Institute, Mass. General	LC-MS
ET2DS	Serum	Baseline	Fasted	Nightingale Health	NMR
Estonia OE	EDTA plasma	Baseline	Fasted + Non-fasted	Broad Institute	LC-MS
EPIC	Serum + citrated plasma	Baseline	Fasted + Non-fasted	IARC	LC-MS
Fenland	Heparin plasma	Baseline	Fasted	Biocrates	LC-MS
FHS2	EDTA plasma	1991-1995	Fasted	Broad Institute	LC-MS
FHS3	EDTA plasma	2002-2005	Fasted	Broad Institute	LC-MS
GDM	Serum	Baseline	Fasted	Nightingale Health	NMR
HABC	EDTA plasma	1999-2000	Fasted	Broad Institute	LC-MS
HPFS	EDTA plasma	1993-1995	Fasted + Non-fasted	Broad Institute	LC-MS
MAC	EDTA plasma	Baseline	Non-fasted	Fred Hutch	LC-MS, NMR
MRC NSHD	Serum	2006-2010	Fasted + Non-fasted	Nightingale Health	NMR
MrOS	Serum	Baseline	Fasted	Pacific Northwest National Labs, WCMC	GC-MS, LC-MS

MEC	Heparin plasma	1994-2016	Fasted	Brigham and Women's Hospital	CoulArray
MESA	EDTA plasma	Baseline	Fasted	ICL NPC	LC-MS, NMR
NHS	Heparin plasma	1989-1990	Fasted + Non-fasted	Broad Institute	LC-MS
NHS-II	Heparin plasma	1996-1999	Fasted + Non-fasted	Broad Institute	LC-MS
PHS	EDTA plasma	Baseline	Fasted + Non-fasted	Broad Institute	LC-MS
POPS	Serum	Baseline	Non-fasted	Metabolon, Inc.	LC-MS
PLCO	Serum	Baseline	Non-fasted	Metabolon, Inc., Broad Institute, Mass. General	GC-MS, LC-MS
SMHS	EDTA plasma	Baseline	Fasted + Non-fasted	Metabolon, Inc., Broad Institute, Metabo-Profile R&D Lab	GC-MS, LC-MS
SPA	EDTA plasma	Baseline	Fasted + Non-fasted	Metabolon, Inc.	GC-MS, LC-MS
SWHS	EDTA plasma	Baseline	Fasted + Non-fasted	Metabolon, Inc., Broad Institute, Metabo-Profile R&D Lab	GC-MS, LC-MS
SP2	EDTA plasma	Baseline	Fasted	Duke-NUS, NUS SLING	GC-MS, LC-MS
SABRE	Serum	1988-1990	Fasted + non-fasted (post OGTT)	Nightingale Health	NMR
TMCS	Serum + EDTA plasma	Baseline	Fasted	Keio University	CE-MS, LC-MS
TwinsUK	Serum + EDTA plasma	1995-2013	Fasted	Metabolon, Inc., Biocrates, Nightingale Health	GC-MS, LC-MS, NMR
UPBEAT	Serum + EDTA plasma	2009-2013	Non-fasted	Nightingale Health	NMR
VDAART	EDTA plasma	Baseline	Non-fasted	Metabolon, Inc.	LC-MS
WH-II	Serum	1997-1999	Fasted + Non-fasted	Nightingale Health	NMR
WHI	EDTA plasma	Baseline	Fasted	Broad Institute, Metabolon, Inc.	LC-MS
WIHS	Citrated plasma	Baseline	Fasted	Broad Institute	LC-MS

Ethylenediaminetetraacetic acid (EDTA); liquid chromatography–mass spectrometry (LC-MS); nuclear magnetic resonance (NMR); gas chromatography–mass spectrometry (GC-MS); capillary electrophoresis–mass spectrometry (CE-MS). [a] Number of participants in each study is shown in Table 1. [b] For those studies with metabolomics at multiple timepoints on the same individuals, we report details at the earliest timepoint for which data are available. [c] Details on metabolomics platforms are as follows: Agilent COE refers to the Agilent Center of Excellence, Brazil (106), Biocrates refers to commercial AbsoluteIDQ^{tm} kits sold by BIOCRATES Life Sciences AG (Innsbruck, Austria) and various academic laboratories use (107), Brigham and Women's Hospital refers to the laboratory of Bruce Kristal (108), Broad Institute refers to the lab of Clary Clish (9), Duke University refers to the Duke Molecular Physiology Institute Metabolomics Core (22), Duke-NUS refers to Duke Metabolomics Core Facility—National University of Singapore, Fred Hutch refers to the laboratory of Daniel Raftery at the Northwest Metabolomics Research Center in the University of Washington (109), IARC refers to the laboratory of Augustin Scalbert at the International Agency for Research on Cancer (110), ICL NPC—Imperial College London National Phenome Centre—refers to the laboratory of Jeremy Nicolson, Elaine Holmes and colleagues (111), Keio University refers to the laboratory of Tomoyoshi Soga (112), Mass General—Massachusetts General Hospital—refers to the former laboratory of Robert Gerszten(1), Metabolon, Inc. refers to the commercial lab Metabolon, Inc. located in North Carolina (113), Nightingale Health refers to the commercial laboratory formerly known as Brainshake Inc. and is the same as the Biocentre Oulu platform of the Mika Ala-Korpela lab in Finland (114), NUS SLING refers to National University of Singapore Singapore Lipidomics Incubator, Örebro University refers to the laboratory and platforms by Matej Oresic and Tuulia Hyötyläinen

(previously at VTT, Finland and Steno Diabetes Center, Denmark), Pacific Northwest Labs refers to the laboratory of Tom Metz, WCMC—West Coast Metabolomics Center—refers to the laboratory of Oliver Fiehn (115).

Table 4. Measurements Available for Participants with Metabolomics Data in COMETS[a]

Cohort	Smoking Status (*n* = 46)	Alcohol intake (*n* = 44)	BMI (*n* = 47)	Waist size (*n* = 37)	LTPA (*n* = 35)	Diet (FFQ) (*n* = 36)	Education Level (*n* = 44)
AIRWAVE	yes	yes	yes	yes	yes	yes	yes
ATBC	yes	yes	yes	no	yes	yes	yes
ARIC	yes	yes	yes	yes	yes	yes	yes
ALSPAC	yes	yes	yes	yes	yes	yes	yes
BHS	yes	yes	yes	yes	yes	yes	yes
BIB	yes	yes	yes	no	yes	yes	yes
BCFR	yes	yes	yes	no	no	no	yes
BWHHS	yes	yes	yes	yes	yes	yes	yes
CaPS	yes	yes	yes	yes	yes	yes	yes
CPS-II	yes	yes	yes	yes	yes	yes	yes
CATHGEN	yes	no	yes	no	no	no	no
CAMP	yes	yes	yes	yes	no	no	yes
COLO	yes	yes	yes	yes	yes	yes	yes
KORA	yes	yes	yes	yes	no	no	yes
DIPP	no	no	yes	no	no	yes	no
DPP	yes	yes	yes	yes	yes	yes	yes
ET2DS	yes	yes	yes	yes	no	no	yes
Estonia OE	yes	yes	yes	yes	yes	yes	yes
EPIC	yes	yes	yes	yes	yes	yes	yes
Fenland	yes	yes	yes	yes	yes	yes	yes
FHS2	yes	yes	yes	yes	yes	yes	yes
FHS3	yes	yes	yes	yes	yes	yes	yes
GDM	yes	no	yes	no	no	no	no
HABC	yes	yes	yes	yes	yes	yes	yes
HPFS	yes	yes	yes	yes	yes	yes	yes
MAC	yes	yes	yes	yes	no	no	yes
MRC NSHD	yes	yes	yes	yes	yes	yes	yes
MrOS	yes	yes	yes	no	yes	yes	yes
MEC	yes	yes	yes	yes	yes	yes	yes
MESA	yes	yes	yes	yes	yes	yes	yes
NHS	yes	yes	yes	yes	yes	yes	yes
NHS-II	yes	yes	yes	yes	yes	yes	yes
PHS	yes	yes	yes	yes	yes	no	yes
POPS	yes	yes	yes	no	no	no	yes
PLCO	yes	yes	yes	no	yes	yes	yes
SMHS	yes	yes	yes	yes	yes	yes	yes
SPA	yes	yes	yes	yes	yes	yes	yes

SWHS	yes	yes	yes	yes	yes	yes	yes
SP2	yes	yes	yes	yes	yes	yes	yes
SABRE	yes	yes	yes	yes	yes	yes	yes
TMCS	yes	yes	yes	yes	yes	yes	yes
TwinsUK	yes	yes	yes	yes	yes	yes	yes
UPBEAT	yes	yes	yes	yes	yes	yes	yes
VDAART	yes	yes	yes	yes	no	yes	yes
WH-II	yes	yes	yes	yes	no	no	yes
WHI	yes	yes	yes	yes	yes	yes	yes
WIHS	yes	yes	yes	yes	no	no	yes

Body mass index (BMI), Leisure-time physical activity (LTPA), Food Frequency Questionnaires (FFQ). [a] yes indicates that the measurement is available in all participants, no indicates that the measurement is not available in any of the participants.

Table 5. Available[a] Clinical Measurements of Participants with Metabolomics Data in COMETS

Cohort	SBP (*n* = 41)	DBP (*n* = 41)	HDL (*n* =39)	LDL (*n* = 38)	TG (*n* = 37)	TC (*n* = 38)	CRP (*n* = 38)	IL-6 (*n* = 32)	HbA1c (*n* = 33)	Fasting glucose (*n* = 37)	Fasting insulin (*n* = 30)	No. with GWAS data (*n* = 93,082)
AIRWAVE	yes	yes	yes	yes	yes	yes	yes	no	yes	P	P	4,000
ATBC	yes	yes	yes	no	no	yes	no	no	no	P	P	475
ARIC	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	3,650
ALSPAC	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	7,176
BHS	yes	yes	yes	yes	yes	yes	no	no	yes	yes	no	939
BIB	yes	yes	yes	yes	yes	yes	yes	no	no	yes	yes	10,000
BCFR	no	no	no	no	no	no	no	no	no	no	no	0
BWHHS	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	3,800
CaPS	yes	yes	yes	yes	yes	yes	yes	yes	no	yes	yes	1,000
CPS-II	no	no	P	P	no	P	P	no	no	no	no	1,450
CATHGEN	yes	yes	yes	yes	yes	yes	P	no	P	P	no	3,255
CAMP	yes	yes	no	no	no	no	no	no	yes	no	no	1,041
COLO	P	P	P	P	P	no	yes	no	no	no	no	408
KORA	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	3,000
DIPP	no	no	no	no	no	no	no	no	yes	yes	no	0
DPP	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	1,815
ET2DS	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	no	1,060
Estonia OE	yes	yes	P	P	P	P	P	P	no	P	no	298
EPIC	P	P	P	P	P	P	P	P	P	P	P	5,000
Fenland	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	9,851
FHS2	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	2,526
FHS3	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	998
GDM	yes	yes	yes	yes	yes	yes	no	no	yes	yes	no	656
HABC	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	160
HPFS	yes	yes	P	P	P	P	P	P	no	no	no	953
MAC	no	no	no	no	no	no	P	P	yes	no	no	0
MRC NSHD	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	0
MrOS	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	1,391
MEC	P	P	yes	yes	yes	yes	yes	P	no	yes	yes	4,431
MESA	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	3,772
NHS	P	P	P	P	P	P	P	P	P	no	P	1,000
NHS-II	P	P	P	P	P	P	P	P	P	no	P	100
PHS	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	224
POPS	yes	yes	no	no	no	no	no	no	no	no	no	0
PLCO	no	no	no	no	no	no	no	no	no	no	no	530
SMHS	yes	yes	P	P	P	P	P	P	P	P	P	656
SPA	yes	yes	no	no	no	no	no	no	no	no	no	295
SWHS	yes	yes	P	P	P	P	P	P	P	P	no	1,300

SP2	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	1,705
SABRE	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	3,000
TMCS	yes	yes	yes	yes	yes	yes	yes	no	yes	yes	P	12,00
TwinsUK	yes	yes	yes	yes	yes	yes	yes	yes	no	yes	yes	6,232
UPBEAT	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	1,303
VDAART	no	no	no	no	no	no	no	no	no	no	no	651
WH-II	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	0
WHI	yes	yes	yes	P	P	yes	P	P	P	P	P	1,781
WIHS	yes	yes	yes	yes	yes	yes	P	P	yes	yes	yes	0

Systolic blood pressure (SBP), diastolic blood pressure (DBP), high-density lipoprotein cholesterol (HDL), low-density lipoprotein cholesterol (LDL), triglycerides (TG), C-reactive protein (CRP), interleukin-6 (IL-6), glycated hemoglobin (HbA1c), genome-wide association study (GWAS). [a] yes indicates that the measurement is available in all participants, P indicates that the measurement is available in a portion of participants, no indicates that the measurement is not available in any of the participants. SBP and DBP levels are self-reported in NHS and NHSII.

Table 6. Number of Identified Metabolites for Five Different Metabolomics Platforms in Five Different Studies Participating in COMETS, and the Overlap across Platforms/Studies.

Platform (Study)	Metabolon, Inc. (ARIC)	Broad Institute (HABC)	Biocrates (Fenland)	WCMC (ColoCare)	Nightingale Health[a] (PLCO)
Metabolon, Inc. (ARIC)	1,158				
Broad Institute (HABC)	121	350			
Biocrates (Fenland)	24	33	187		
WCMC (ColoCare)	92	82	20	439	
Nightingale Health (PLCO)	16	14	6	12	25[b]

West Coast Metabolomics Center (WCMC). [a] Formerly known as Brainshake Inc. [b] Excluding metabolite ratios and sums that are routinely included as part of the platform results.

Figure 1 legend. Geographical locations of studies participating in COMETS (acronyms defined in Table 1).

Figure 2 legend. Spearman correlations between metabolite values measured at the Broad Institute and Metabolon, Inc. for 111 overlapping metabolites. The median correlation across the 111 metabolites was 0.79.