Global Genomic
Medicine Collaborative

International Cohorts Summit

JB Duke Hotel
Durham, North Carolina, USA

March 26 – 27, 2018

International Cohorts Summit

Summit Objectives

The International Cohorts Summit was conceived by the **Heads of International Research Organizations (HIROs)** group co-chaired by Francis Collins at the National Institutes of Health (NIH) and Jeremy Farrar of the Wellcome Trust. Dr. Collins and Dr. Farrar reached out to the **Global Genomic Medicine Collaborative (G2MC, www.g2mc.org)**, to organize a first forum with the goal of enabling leaders of large-scale longitudinal cohorts worldwide to share best practices, discuss data sharing, explore standards, discuss common challenges, and explore the potential for a larg(er) collaborative sequencing strategy.

The primary objectives of this meeting will include:

- Improving prospects for interoperability and compatibility of instruments, data formats, phenotypic and clinical measures, etc
- Promoting data sharing and open access policies
- Broadening international cooperation through existing tools and resources
- Exploring the feasibility of a "digital" platform, or web-based, evolving registry of large-scale cohorts, in searchable format
- Examining the potential for a collaborative sequencing (and other -omics?) strategy
- Considering strategies for translating findings for health impact
- Advancing a collective vision: where do we want to be in ten years?

We thank you for joining us at this pivotal event, and for contributing to our shared goal of enhanced international collaboration across research projects.

Programme Committee

Geoff Ginsburg, *Duke University, USA*
Teri Manolio, *National Human Genome Research Institute (NHGRI)*
Rory Collins, *University of Oxford, UK*
Masayuki Yamamoto, *Tohoku Medical Megabank, Japan*

Support staff

Justina Chung, *G2MC, Canada*
Lena Dolman, *G2MC, Canada*
Ashley Hobb, *G2MC, Canada*
Teji Rakhra-Burris, *Duke University and G2MC, USA*

Table of Contents

Estonian Genome Project

European Prospective Investigation into Cancer & Nutrition (EPIC)

Finnish Maternity Cohort Serum Bank

Generations Study

Genomics England

Golestan Cohort Study

Japan Public Health Center-based Prospective Study (JPHC)

Japan Public Health Center-based Prospective Study for the Next generation (JPHC-NEXT)

Kaiser Permanente Research Program on Genes, Environment, and Health (RPGEH)

Korea Biobank Project

Korean Cancer Prevention Study-II (KCPS-II) Biobank

Korean Genome and Epidemiology Study (KoGES)

LifeGene (and EpiHealth)

LIFEPATH

Malaysian Cohort

Maule Cohort (MAUCO Study)

Mexico City Prospective Study

Million Veteran Program

Million Women Study

Multiethnic Cohort Study (MEC)

MyCode Community Health Initiative

Newfoundland and Labrador Genome Project

Northern Sweden Health and Disease Study

Norwegian Family Based Life Course Study

Norwegian Mother and Child Cohort Study (MoBa)

Nurses' Health Study (NHS)

Nurses' Health Study II (NHSII)

OHS (Ontario Health Study) and the Canadian Partnership for Tomorrow Project (CPTP)

PERSIAN Cohort Study

Prostate, Lung, Colorectal and Ovarian Cancer Screening Trial (PLCO)

Saudi Human Genome Program

Saudi National Biobank

Singapore National Precision Medicine Program

Taiwan Biobank

Tohoku Medical Megabank Project

UK Biobank

UKCTOCS (UK Collaborative Trial of Ovarian Cancer Screening) Longitudinal Women's Cohort – (UKLWC)

Women's Health Initiative (WHI)

Global Genomic Medicine Collaborative

International Cohorts Summit

International Cohorts Summit Agenda

Hosted by the Global Genomic Medicine Collaborative (G2MC)

JB Duke Hotel (Ballroom ABC on Level 3, i.e. lobby level), 230 Science Dr, Durham, North Carolina, USA
March 26th-27th, 2018

Meeting Objectives

- Improve prospects for interoperability and compatibility of instruments, data formats, phenotypic and clinical measures, etc.
- Promote data sharing and open access policies
- Broaden international cooperation through existing tools and resources
- Explore the feasibility of a "digital" platform, or web-based, evolving registry of large-scale cohorts, in searchable format
- Examine the potential for a collaborative sequencing (and other -omics?) strategy
- Consider strategies for translating findings for health impact
- Advance a collective vision: where do we want to be in ten years?

Day 1, Monday, March 26

7:30 – 8:30	Registration [location: JB Duke Hotel atrium, Level 3 (i.e. lobby level)]	

Session 1 – Introduction and Background
Chairs: Geoff Ginsburg & Teri Manolio

8:30 – 8:50	Welcome and Introductions Welcome from Chancellor for Health Affairs, Duke University	Geoff Ginsburg, Teri Manolio Eugene Washington
8:50 – 9:10	Vision for summit	Francis Collins, Jeremy Farrar
9:10 – 9:15	Summary of cohorts in attendance	Teri Manolio
9:15 – 9:35	Value and challenges of combining large cohorts	Rory Collins
9:35 – 9:55	Opportunities to enhance translation for discovery to health	Geoff Ginsburg
9:55 – 10:20	Discussion	
10:20 – 10:40	**Break**	

Session 2 – Opportunities for Collaboration Across Cohorts
Chairs: Nicola Mulder & Rory Collins

10:40 – 11:20	Obtaining phenotype and outcome data from electronic health records and digital platforms ● US ● UK ● Asia	 Josh Denny Cathie Sudlow Zhengming Chen
11:20 – 11:40	Value of biospecimen collection & biobanking	Nancy Pedersen
11:40 – 12:00	Value of genomic information and how to gather it	Matt Nelson
12:00 – 12:20	Value of other -omic information and how to gather it	John Danesh
12:20 – 12:45	Discussion	
12:45 – 1:45	**Lunch**	

Session 3 – Opportunities for Collaboration Across Cohorts (Cont'd)
Chairs: Francis Collins & Jeremy Farrar

1:45 – 2:05	Value of environmental information and how to gather it	David Hunter
2:05 – 2:25	Value of nutritional information and how to gather it	Walter Willett
2:25 – 2:45	Working with multi-ethnic data	Sekar Kathiresan
2:45 – 3:10	Discussion	
3:10 – 3:30	**Break**	

Session 4 – Data Standards and Privacy
Chair: Eric Dishman & Geoff Ginsburg

3:30 – 3:50	Data standards and global variant databases	Thomas Keane
3:50 – 4:10	Informed consent, data privacy	Laura Rodriguez
4:10 – 4:30	Quantitative science to optimize the value of cohort data	Robert Califf
4:30 – 5:00	Discussion	

Session 5 – Break Out Sessions Begin

5:00 – 6:00	Working sessions, leaders, and key topics to address	
Group 1 [Boardroom A, Level 3]	Creating a standardized database and registry—pros, cons, how best to do it ● What data are cohort investigators willing to share, with whom, and how? ● What challenges need to be addressed to optimize the value of sharing information?	Joyce Tung Daniel MacArthur
Group 2 [Ballroom E, Level 3]	IT considerations for enabling coordination, communication, centralization (include federated databases) ● What data are collected/available for each of your cohorts and what are their formats? ● What are sources of those data (e.g. electronic health record, laboratory, radiology, genomics, proteomics, metabolomics, others)? ● How are data collected in each of your cohorts stored? ● How are the data distributed, accessed, made discoverable? What are the challenges to enabling federated joint cohort analysis to benefit clinicians and research discoveries? ● Are there other IT considerations/challenges not covered above?	Teresa Zayas Cabán Thomas Keane
Group 3 [Meeting Room D, Level 2]	Scientific agenda with short- and long-term goals ● What enhancements to existing cohorts would most increase their utility and promote data sharing? ● What are the highest priority scientific questions that could be addressed by a cohort of cohorts?	Geoff Ginsburg Rory Collins
Group 4 [Boardroom B, Level 3]	Policy agenda to facilitate and optimize impact of assembling these cohorts (include MTAs, IRBs, consent, etc) ● What specific legal or regulatory barriers in each cohort, aside from ensuring confidentiality and appropriate consent, would impede data sharing of: a) genetic information; b) demographic and clinically-relevant data? What solutions or options to work through challenges could be pursued? ● What (if any) concerns might be raised by private industry involvement in a cohort of cohorts and how could they be dealt with?	Laura Rodriguez Gad Rennert

Group 5 [Ballroom ABC, Level 3]	Developing a collaborative genomic sequencing (and other -omics?) strategy ● What kinds of sequencing or other -omic data would be useful for individual cohorts? ● What aspects of a collaborative sequencing strategy, in addition to low cost, would facilitate obtaining and sharing these data? ● What methods/tools are optimal for data harmonization across different sites to address platform diversity/uniformity, batch effects and related issues	John Danesh Hakon Hakonarson
Group 6 [Meeting Room C, Level 2]	Translation / clinical impact ● What are the opportunities for translation of cohort findings to improved clinical care and population health? ● What are the major barriers to clinical and population health translation and how can they be dealt with?	Eric Green Dan Roden
6:00	**Adjourn**	
6:30	**Depart for dinner** [meet at front entry of JB Duke Hotel for group transportation]	
7:00	**Group Dinner at Mothers & Sons Restaurant**	

Day 2, Tuesday, March 27

8:00 – 8:15	Opening Day 2	Geoff Ginsburg, Teri Manolio

Session 6 – EU Cohorts and Break Out Working Sessions
Chair: Andres Metspalu & Teri Manolio

8:15 – 8:45	The EU Experience in Assembling Cohorts of Cohorts	Philippe Cupers
8:45 – 9:00	Welcome from the Dean of Medicine, Duke University	Mary Klotman
9:00 – 10:00	Break Out Group Working sessions (same room assignments as Day 1)	
10:00 – 10:20	**Break**	

Session 7 – Break Out Reports and Discussion

Chair: Camilla Stoltenberg & Rory Collins

10:20 – 10:40	Group 1: Standardized database	Joyce Tung Daniel MacArthur
10:40 – 11:00	Group 2: IT considerations	Teresa Zayas Cabán Thomas Keane
11:00 – 11:20	Group 3: Scientific agenda	Geoff Ginsburg Rory Collins
11:20 – 11:40	Group 4: Policy agenda	Laura Rodriguez Gad Rennert
11:40 – 12:00	Group 5: Collaborative sequencing strategy	John Danesh Hakon Hakonarson
12:00 – 12:20	Group 6: Translation / clinical impact	Eric Green Dan Roden

Session 8 – Working Lunch, Summary and Next Steps

12:20 – 12:50	**Lunch (pick up box lunch and return)**	
12:50 – 1:35	Summary, outline of 1-year plan	Geoff Ginsburg Teri Manolio
1:35– 2:15	Consensus vision and path forward	Francis Collins Jeremy Farrar
2:15	**Adjourn**	

Speakers and Breakout Chairs

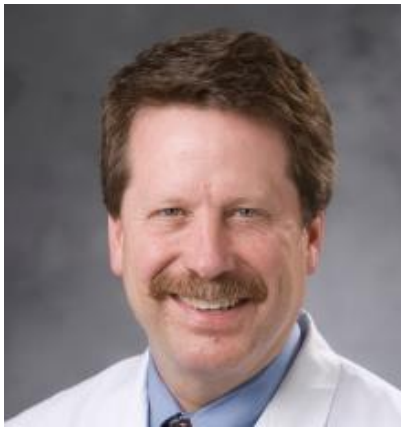
Teresa Zayas Cabán

Director, Office of the Chief Scientist (OCS)
Office of National Coordinator (ONC) for Health IT
USA

Teresa Zayas Cabán is the director of the Office of the Chief Scientist (OCS), which is responsible for developing and evaluating ONC's overall scientific efforts and activities. OCS develops, establishes, or recommends scientific policy to the national coordinator. OCS leads ONC's precision medicine initiative (PMI) activities and provides oversight of ONC's patient-centered outcomes research (PCOR) projects.

Dr. Zayas Cabán was previously the chief of health IT research and acting director of the division of health IT at the Agency for Healthcare Research and Quality (AHRQ). While at AHRQ, she set new directions for their funding opportunities and coordinated with federal partners, such as the National Science Foundation.

Before joining AHRQ, she served as a post-doctoral trainee in the computation and informatics in biology and medicine program at the University of Wisconsin-Madison. Dr. Zayas Cabán obtained her doctorate in industrial and systems engineering at the University of Wisconsin-Madison where she was a National Science Foundation graduate research fellow in industrial engineering.



Robert Califf

Duke University
USA

Robert Califf, MD MACC, is the Donald F. Fortin, MD, Professor of Cardiology. He is also Professor of Medicine in the Division of Cardiology and remains a practicing cardiologist. Dr. Califf was the Commissioner of Food and Drugs in 2016-2017 and Deputy Commissioner for Medical Products and Tobacco from February 2015 until his appointment as Commissioner in February 2016. Prior to joining the FDA, Dr. Califf was a professor of medicine and vice chancellor for clinical and translational research at Duke University. He also served as director of the Duke Translational Medicine Institute and founding director of the Duke Clinical Research Institute. A nationally and internationally recognized expert in cardiovascular medicine, health outcomes research, healthcare quality, and clinical research, Dr. Califf has led many landmark clinical trials and is one of the most frequently cited authors in biomedical science, with more than 1,200 publications in the peer-reviewed literature.

Zhengming Chen

Professor
University of Oxford
Co-Director
China Oxford Centre for
International Health Research
UK and China

Zhengming Chen is the lead principle investigator of the China Kadoorie Biobank, one of the world's largest blood-based prospective study ever established. He qualified in medicine at the Shanghai Medical University (SMU, now Fudan University) in 1983. He subsequently completed public health postgraduate training in the School of Public Health, SMU and gained his DPhil in Epidemiology at the University of Oxford in 1993. He currently holds the position of Professor of Epidemiology at the University of Oxford, and honorary professorships of Peking Union Medical College, Fudan University and Shanghai Institute of Biological Sciences, Chinese Academy of Sciences. He is the founding co-director of the China Oxford Centre for International Health Research.

Professor Chen's research has focused on the environmental, lifestyle and genetic determinants of chronic disease, development of evidence-based medicine and efficient strategies for chronic disease control in developing countries. Although based in Oxford, his research has mainly involved nationwide projects in China. In total, these probably represent the largest epidemiological collaboration in the world between China and other countries. They have provided, and will continue to do so, important results relevant to both Chinese and global health. Over the last 25 years, he has led large placebo-controlled trials involving in total 60,000 acute heart attacks, 20,000 strokes and 15,000 cancers, leading to major changes of international guidelines and new US FDA drug labelling. He has also led large observational epidemiologic studies of the relevance to health of tobacco, alcohol, adiposity, blood pressure, and diet. In particular, he initiated, and has led the prospective China Kadoorie Biobank study (www.ckbiobank.org) from its inception in 2002, which includes 512,000 adults enrolled during 2004-08 from 10 diverse areas across China.

Francis Collins

Director
National Institutes of Health (NIH)
USA

Francis S. Collins, M.D., Ph.D. was appointed the 16th Director of the National Institutes of Health (NIH) by President Barack Obama and confirmed by the Senate. He was sworn in on August 17, 2009. On June 6, 2017, President Donald Trump announced his selection of Dr. Collins to continue to serve as the NIH Director. In this role, Dr. Collins oversees the work of the largest supporter of biomedical research in the world, spanning the spectrum from basic to clinical research.

Dr. Collins is a physician-geneticist noted for his landmark discoveries of disease genes and his leadership of the international Human Genome Project, which culminated in April 2003 with the completion of a finished sequence of the human DNA instruction book. He served as director of the National Human Genome Research Institute at NIH from 1993-2008.
Before coming to NIH, Dr. Collins was a Howard Hughes Medical Institute investigator at the University of Michigan. He is an elected member of the National Academy of Medicine and the National Academy of Sciences, was awarded the Presidential Medal of Freedom in November 2007, and received the National Medal of Science in 2009.

Rory Collins

Principal Investigator, UK Biobank
Professor, University of Oxford
UK

Professor Sir Rory Collins studied Medicine at St Thomas's Hospital Medical School, London, and Statistics at George Washington University and Oxford University.

He came to Oxford in 1981 to run the ISIS "mega-trials" which showed that emergency treatment of heart attacks with streptokinase and aspirin halves mortality. Subsequently, his focus has involved showing that lowering LDL-cholesterol safely reduces the risk of having heart attacks and strokes.

In 1985, he became co-director (with Richard Peto) of the University of Oxford's Clinical Trial Service Unit & Epidemiological Studies Unit (CTSU). He was appointed BHF Professor of Medicine and Epidemiology in 1996, and Head of the Nuffield Department of Population Health in 2013.

Rory became Principal Investigator of the UK Biobank prospective study of 500,000 people in 2005. He was elected to the Fellowship of the UK Academy of Medical Science in 2004 and the Royal Society in 2015, and knighted by the Queen for services to Science in 2011.

Philippe Cupers

European Commission
Belgium

Philippe Cupers is Deputy Head of Unit for 'Health strategy' at the Directorate-General for Research and Innovation of the European Commission. His responsibilities focus on the development, promotion and monitoring of European health research and innovation strategy and policies (e.g. personalised medicine), including longterm vision and forward looking activities. It takes into account developments in the Member States (both in public and private sectors) and on the international scene. Philippe Cupers has a PhD in cell biology and biochemistry from the University of Louvain (de Duve Institute). He has worked at the Harvard University Medical School, at the Flemish Institute for Biotechnology (University of Leuven) and at GlaxoSmithKline. He joined the Directorate- General for Research and Innovation of the European Commission in 2001, where he worked as scientific officer and then Head of Sector for brain research, as well as scientific officer for setting up the Innovative Medicines Initiative (IMI).

John Danesh

Professor
University of Cambridge
UK

John Danesh FMedSci is British Heart Foundation Professor of Epidemiology and Medicine at the University of Cambridge, and Associate Faculty member at the Wellcome Sanger Institute.

Professor Danesh trained in medicine at the University of Otago in New Zealand and at the Royal Melbourne Hospital in Australia. During his time as a Rhodes scholar, he received an MSc in Epidemiology at the London School of Hygiene and Tropical Medicine and a DPhil in Epidemiology at the University of Oxford. He was elected a Fellow of the Royal College of Physicians in 2007.

Professor Danesh is the founder and director of the Cardiovascular Epidemiology Unit (CEU), a multi-disciplinary Unit of over 60 staff and students that aims to advance understanding and prevention of cardiovascular disease through population health research.

Josh Denny

Professor
Vanderbilt University Medical Centre
USA

Josh Denny is Professor of Biomedical Informatics and Medicine, Director of the Center for Precision Medicine and Vice President of Personalized Medicine at Vanderbilt University Medical Center. He is a Fellow of the American College of Medical Informatics and a member of the National Academy of Medicine.

He has substantial experience in the design, development, and implementation of electronic health record (EHR) data mining algorithms and was the primary author of several natural language processing systems to support phenotype extraction algorithms for genomic research projects, including development of the phenome-wide association study (PheWAS) method. He is principal investigator (PI) of nodes in the Electronic Medical Records and Genomics (eMERGE) Network, Pharmacogenomics Research Network (PGRN), and the Implementing Genomics into Practice (IGNITE) Network.

Dr. Denny is PI of the Data and Research Center of the Precision Medicine Initiative All of Us Research Program (previously called the Precision Medicine Initiative Cohort Program), which will eventually enroll at least 1 million Americans in an effort to understand the genetic, environmental, and behavioral factors that influence human health and disease. To date, he has led >100 genome-wide and candidate gene association studies using EHR data linked to genetic data.

Dr. Denny serves on a number of mentoring committees and has trained >30 postdoctoral and predoctoral trainees.

Jeremy Farrar

Director
Wellcome Trust
UK

Jeremy Farrar is Director of the Wellcome Trust, a global charitable foundation (both politically and financially independent) which exists to improve health for everyone by helping great ideas to thrive. Wellcome supports scientists and researchers, takes on big problems, fuels imaginations and sparks debate.

Jeremy is a clinician scientist who before joining Wellcome was, for eighteen years, Director of the Oxford University Clinical Research Unit in Viet Nam, where his research interests were in infectious diseases and global health with a focus on emerging infections. He has published over 500 peer-reviewed scientific papers, mentored many dozens of students and fellows and served as chair on several advisory boards for governments and global organisations including the World Health Organization. He was named 12th in the Fortune list of 50 World's Greatest Leaders in 2015.

Jeremy was appointed OBE in 2005 for services to tropical medicine, was awarded the Memorial Medal and the Ho Chi Minh City Medal by the Government of Vietnam, and has been honoured by the Royal College of Physicians in the UK and the American Society for Tropical Medicine and Hygiene. He is a Fellow of both the Academy of Medical Sciences and The Royal Society and an elected member of the European Molecular Biology Organisation (EMBO).

Geoff Ginsburg

Director
Center for Applied Genomics & Precision Medicine
Duke University Medical Center
USA

Geoff Ginsburg is the founding director for the Center for Applied Genomics & Precision Medicine at the Duke University Medical Center and for MEDx, a partnership between the Schools of Medicine and Engineering to spark and translate innovation. His research addresses the challenges for translating genomic information into medical practice and the integration of precision medicine into healthcare. In 2017 he received Duke's Translational Research Mentorship Award.

He is a member of the Advisory Council to the Director of NIH and is co-chair of the National Academies Roundtable on Genomic and Precision Health and is founder and president of the Global Genomic Medicine Collaborative, a not for profit organization aimed at creating international partnerships to advance the implementation of precision medicine. He has recently served as a member of the Board of External Experts for the NHLBI, the advisory council for the National Center for Accelerating Translational Science, the chair of the review for Genome Canada's Large Scale Applied Research Competition in Genomics and Precision Medicine, and the World Economic Forum's Global Agenda Council on the Future of the Health Sector.

Eric Green

Director
National Human Genome Research Institute (NHGRI)
USA

Eric Green is the Director of the National Human Genome Research Institute (NHGRI) at the National Institutes of Health (NIH), a position he has held since late 2009. Previously, he served as the NHGRI Scientific Director, Chief of the NHGRI Genome Technology Branch, and Director of the NIH Intramural Sequencing Center.

While directing an independent research program for almost two decades, Dr. Green was at the forefront of efforts to map, sequence, and understand eukaryotic genomes. His work included significant, start-to-finish involvement in the Human Genome Project.

As Director of NHGRI, Dr. Green is responsible for providing overall leadership of the Institute's research portfolio and other initiatives. In 2011, Dr. Green led NHGRI to the completion of a strategic planning process that yielded a new vision for the future of genomics research, entitled *Charting a course for genomic medicine from base pairs to bedside (Nature 470:204-213, 2011).* Since that time, he has led the Institute in broadening its research mission; this has included designing and launching a number of major programs to accelerate the application of genomics to medical care. Beyond NHGRI-specific programs, Dr. Green has also played an instrumental leadership role in the development of several high-profile efforts relevant to genomics, including the Smithsonian-NHGRI exhibition *Genome: Unlocking Life's Code*, the NIH Big Data to Knowledge (BD2K) program, the NIH Genomic Data Sharing Policy, and the U.S. Precision Medicine Initiative.

Hakon Hakonarson

Children's Hospital of
Philadelphia (CHOP)
USA

Hakon Hakonarson, M.D., Ph.D., is Director of the Center for Applied Genomics and is also an associate professor of pediatrics at The University of Pennsylvania School of Medicine. He leads a $40 million commitment from CHOP to genotype approximately 100,000 children, an initiative that has gained nationwide attention in the Wall Street Journal, New York Times, Time Magazine, Nature and Science.

Dr. Hakonarson has previously held several senior posts within the biopharmaceutical industry, directing a number of genomics and pharmacogenomics projects as vice president of Clinical Sciences and Development at deCODE genetics, Inc.

Dr. Hakonarson has been the principal investigator (PI) on several National Institute of Health-sponsored grants, and is currently co-PI on Neurodevelopmental Genomics: Trajectories of Complex Phenotypes, the largest project ever supported by the National Institute of Mental Health. He has published numerous high-impact papers on genomic discoveries and their translations in some of the most prestigious scientific medical journals, including Nature, Nature Genetics and The New England Journal of Medicine. Time Magazine listed Dr. Hakonarson's autism gene discovery project, reported in Nature in 2009, among the top 10 medical breakthroughs of that year.

David Hunter

Professor
Harvard T.H. Chan School of
Public Health
USA

David Hunter studied medicine at the University of Sydney, before moving to Harvard University for 33 years where he was the Vincent L. Gregory Professor of Cancer Prevention. He is the Richard Doll Professor of Epidemiology and Medicine, and director of the Harvard-Oxford Program in Epidemiology. His early research was on HIV transmission in East Africa, and subsequently he was involved in collaborative studies of nutrition and HIV pathogenesis, while also studying diet and cancer etiology in large scale prospective studies and founding the Pooling Project of Prospective Studies of Diet and Cancer.

As Director of the Harvard Center for Cancer Prevention he developed a sample handling and genotyping laboratory to explore genetic associations with cancer, and gene-environment interactions. He founded the Program in Genetic Epidemiology and Statistical Genetics at Harvard. He was co-chair of the steering committee of the NCI Breast and Prostate Cancer Cohort Consortium (BPC3) between 2003 and 2012, was co-director of the NCI Cancer Genetic Susceptibility Markers project focused on genome-wide association studies, and was an Eminent Scholar at the NCI between 2004 and 2009.

From 2009-2016 he was Dean for Academic Affairs at the Harvard TH Chan School of Public Health, and in 2015-2016 he was Acting Dean. He is one of about 3000 "highly cited researchers" worldwide according to Thomson-Reuters.

Sekar Kathiresan

Director
Center for Genomic Medicine,
Massachusetts General Hospital
USA

Sekar Kathiresan, a physician scientist and a human geneticist, is the Director of the Center for Genomic Medicine (CGM) at Massachusetts General Hospital (MGH), Ofer and Shelley Nemirovsky MGH Research Scholar, Director of the Cardiovascular Disease Initiative at the Broad institute, and an Associate Professor of Medicine at Harvard Medical School.

Dr. Kathiresan leverages human genetics to understand the root causes of heart attack and to improve preventive cardiac care. Among his scientific contributions, Dr. Kathiresan has helped highlight new biological mechanisms underlying heart attack, discovered mutations that protect against heart attack risk, and developed a genetic test for personalized heart attack prevention.

Dr. Kathiresan received his B.A. in history and graduated summa cum laude from the University of Pennsylvania in 1992 and received his M.D. from Harvard Medical School in 1997. He then completed his clinical training in internal medicine and cardiology at MGH, where he served as Chief Resident in Internal Medicine from 2002-2003. Dr. Kathiresan pursued research training in cardiovascular genetics through a combined experience at the Framingham Heart Study and the Broad Institute. In 2008, he joined the faculties of the MGH Cardiology Division, Cardiovascular Research Center, and Center for Genomic Medicin.

Thomas Keane

European Bioinformatics Institute
(EBI)
UK

Thomas Keane is head of the European Genome-phenome Archive (EGA), the European Variation Archive (EVA), and the European Nucleotide Archive (ENA) infrastructure team at EMBL-EBI. He is a member of the steering committee for the Global Alliance for Genomics and Health (GA4GH) and leads the Large Scale Genomics Work Stream.

Previously, he spent ten years at the Wellcome Sanger Institute managed the production for the 1000 Genomes Project, the UK10K project, and the Mouse Genomes Project. His research interests include mouse genetics and genomics, structural variation, and genome assembly.

Mary Klotman

Duke University
USA

Mary Klotman, MD, was named dean of the Duke University School of Medicine and vice chancellor for health affairs at Duke University in January 2017. She assumed her new role on July 1, 2017. Prior to her appointment as dean, Dr. Klotman served with distinction as chair of the Department of Medicine in the Duke University School of Medicine for seven years.

An accomplished clinician and scientist, Klotman's research interests are focused on the molecular pathogenesis of Human Immunodeficiency Virus 1 (HIV-1) infection.

Among many important contributions to this field, Klotman and her team demonstrated that HIV resides in and evolves separately in kidney cells, a critical step in HIV-associated kidney disease. Her research group also has determined the role of soluble host factors involved in an innate immune response to HIV in an effort to improve prevention strategies, topical microbicides that could be used to block sexual transmission of HIV.

Most recently, her group has been defining the role of integrase-defective lentiviral vectors for the delivery of an HIV vaccine.

Daniel MacArthur

Assistant Professor
Harvard Medical School
USA

Daniel MacArthur is an assistant professor at Harvard Medical School and Massachusetts General Hospital, and the co-director of Medical and Population Genetics at the Broad Institute of MIT and Harvard.

His research focuses on the use of large-scale genomic technologies and resources to improve the diagnosis of rare diseases. As part of this effort, he and his colleagues developed the Exome Aggregation Consortium resource and its successor, the Genome Aggregation Database, which make data from over 135,000 sequenced exomes and genomes publicly available. These resources are used widely in the genetics community and have contributed to many collaborative gene discovery projects.

He is also heavily involved in the application of genomic technologies to rare disease diagnosis and gene discovery, and co-directs the Broad Institute's Center for Mendelian Genomics as well as a large direct-to-patient rare disease study, the Rare Genomes Project. These projects involve the sequencing and analysis of more than 1,000 rare disease families each year.

Dr. MacArthur completed his Ph.D. at the Institute for Neuromuscular Research in Sydney, Australia. He later worked as a postdoctoral fellow at the Wellcome Trust Sanger Institute in Hinxton, UK, where he led the annotation of gene-disrupting ("loss-of-function") variants as part of the 1000 Genomes Project Consortium.

Teri Manolio

Director, Division of Genomic Medicine
National Human Genome Research Institute (NHGRI)
USA

Teri Manolio is a physician an epidemiologist at the NIH and Walter Reed National Military Medical Centre. She joined NHGRI in 2005 to lead efforts in applying genomic technologies to population research. She has authored over 270 research reports and has research interests in genome-wide association studies of complex diseases, ethnic differences in disease risk, and incorporating genomic findings into clinical care.

As a physician and epidemiologist, Teri has a deep interest in discovering genetic changes associated with diseases by conducting biomedical research on large groups of people. As the Director of the Division of Genomic Medicine, Dr. Manolio leads efforts to support research translating those discoveries into diagnoses, preventive measures, treatments and prognoses of health conditions.

Matt Nelson

Head of Genetics
GSK
USA

Matt Nelson is the head of Genetics at GSK, working in Philadelphia, PA, leading a group of scientists bringing genetic evidence to inform all drug discovery and drug development decisions.

Research activities of personal interest include investigating the role of growing body genome-wide association studies to inform drug target selection and validation, improving pharmacogenetics experiment design and developing methods and strategies for drawing inferences from both small- and large-scale genetic association studies.

Matt graduated from the University of Michigan in statistics and human genetics. He has previously worked for Sequenom and Esperion Therapeutics.

Nancy Pedersen

LifeGene
Sweden

Nancy Pedersen is a Professor of Genetic Epidemiology. Peterson has been at Karolinska Institutet for 36 years, served as the vice chair and chair of the Department of Medical Epidemiology and Biostatistics, and as the Vice Dean of research at KI. Pedersen's research is based on the Swedish Twin Registry

More recently, Petersen has been working with biomarkers of aging such as telomere length, and epigenetics (methylation). Pedersen conducts several studies on "normal" aging, and age-related disorders such as Alzheimer's and dementia, Parkinson's disease, and late-onset depression.

Richard Peto

Professor
University of Oxford
UK

Sir Richard Peto is Professor of Medical Statistics and Epidemiology at the University of Oxford. He is best known for his international epidemiological studies of smoking, alcohol, blood pressure, adiposity and other avoidable causes of premature death. He is a co-PI of the EBCTCG meta-analyses of all the breast cancer treatment trials in the world.

Richard Peto, Rory Collins and others in the Oxford Clinical Trial Service Unit (CTSU) have, by their large randomised trials, large prospective studies and worldwide meta-analyses, increased substantially the estimated importance of blood lipids, blood pressure and smoking as causes of premature death. Peto has recently collaborated in major studies of alcohol in Russia and of malaria in Africa and India. His investigations into the worldwide health effects of smoking and benefits of stopping at particular ages have helped to communicate effectively the vast and growing burden of disease from tobacco use, have helped change national and international attitudes about smoking and public health, and have helped many smokers to stop. He was the first to describe clearly the future worldwide health effects of current smoking patterns, predicting one billion deaths from tobacco in the present century if current smoking patterns persist, as against 'only' 100 million in the 20th century.

Gad Rennert

Director
Clalit National Cancer Control Center and National Personalized Medicine Program Israel

Gad Rennert's research interests are in molecular/genetic cancer epidemiology, cancer prevention and screening, and in molecular or personalized targeted medicine. Cancer chemoprevention studies are another focus of his scientific work. He is currently studying and coordinating the cancer screening activities on a national level in Israel and studying unique founder mutations leading to cancer through the Familial Cancer Consultation Service which cares for thousands of mutation carriers.

Dr. Rennert's work involves a deep epidemiological evaluation of more than 40,000 study participants combined with massive genetic testing using RT-PCR, Sanger sequencing, deep next-generation sequencing, microarray testing, multi-gene expression panels studies, Onco-chip multi-SNP panel, whole exome analysis and other laboratory methods.

Dan Roden

Advisory Council
National Human Genome Research Institute (NHGRI)
USA

Dan Roden received his medical degree and trained in Internal Medicine at McGill University in Montreal. He then went to Vanderbilt where, after fellowships in Clinical Pharmacology and Cardiology, he joined the faculty, and has remained there since. His research program studies the phenotypic consequences of genomic variation focusing on pharmacogenetics and on arrhythmia susceptibility. After serving as chief of the division of Clinical Pharmacology for 12 years, he was tasked in 2006 with leading Vanderbilt's efforts in Personalized Medicine. Under his leadership, Vanderbilt has become nationally- and internationally-recognized for cutting edge programs in this area: these include BioVU, the largest single-site collection (now >250,000) of DNA samples coupled to electronic health records in the world, and PREDICT, a program that puts genetic information on variable drug responses in patients' electronic health records, and uses that information when target drugs are prescribed. He has been principal investigator for the Vanderbilt sites of the Pharmacogenomics Research Network since 2001 and of the Electronic Medical Records and Genomics (eMERGE) Network since 2007.

Dr. Roden has received the Leon Goldberg Young Investigator Award and the Rawls Palmer Progress in Science Award from the American Society for Clinical Pharmacology and Therapeutics; the Distinguished Scientist Award and the Douglas Zipes lectureship from the Heart Rhythm Society; the Distinguished Scientist Award and the inaugural Functional Genomics and Translational Biology Medal of Honor from the American Heart Association; and the McGill Alumnus Lifetime Achievement Award. He currently serves on the Advisory Council to the National Human Genome Research Institute. He has been elected to membership in the American Society for Clinical

Investigation and the Association of American Physicians, and fellowship in the American Association for the Advancement of Science.

Laura Rodriguez

Director
Division of Policy, Communications, and Education
National Human Genome Research Institute (NHGRI)
USA

Laura Rodriguez is the Director of the Division of Policy, Communications, and Education. In this capacity, she works to develop and implement policy for research initiatives at the NHGRI, design communication and outreach strategies to engage the public in genomic science, and prepare health care professionals for the integration of genomic medicine into clinical care. Dr. Rodriguez is particularly interested in the policy and ethics questions related to the inclusion of human research participants in genomics and genetics research.

As a leader within the development of the National Institutes of Health (NIH) Policy for Data Sharing in Genome-Wide Association Studies (GWAS) and its follow-on the NIH Genomic Data Sharing (GDS) Policy, Dr. Rodriguez helped to shape the agency's approach to sharing NIH-supported genomic research data. She continues to play a significant role in overseeing implementation and on-going policy development for genomic data sharing across NIH.

Dr. Rodriguez' career has included positions in the legislative, advocacy, and non-governmental policy arenas where she focused on a range of topics including research ethics, intellectual property, and human research participant regulations.

Dr. Rodriguez received her bachelor's degree with honors in biology from Washington and Lee University in Lexington, Va. and earned a doctorate in cell biology at Baylor College of Medicine in Houston.

Cathie Sudlow

Chief Scientist
UK Biobank
UK

Cathie Sudlow, a clinician and scientist with a particular interest in understanding the causes and prevention of strokes, is UK Biobank's Chief Scientist and Senior Epidemiologist. Dr Sudlow is Clinical Reader and Honorary Consultant Neurologist at the University of Edinburgh, and continues to look after patients with strokes and other neurological disorders. She holds a number of key positions within stroke research, including chair of the British Association of Stroke Physicians scientific committee, and membership of the International Stroke Genetics Consortium. Cathie's role will include overseeing UK Biobank's linkages to health records and working with expert groups to ensure that these records are combined in the best way for health-related research on a range of different conditions. She also looks forward to developing new ideas for improving the resource to make it more useful to the health researchers.

Joyce Tung joined 23andMe in 2007 and leads the 23andMe research team, which is responsible for consumer health and ancestry research and development, academic and industry collaborations, computational analyses for therapeutics, and new research methods and tools development. While a postdoctoral fellow at Stanford University, Joyce studied the genetics of mouse and human pigmentation. She earned her Ph.D. in Genetics from the University of California, San Francisco where she was a National Science Foundation graduate research fellow.

Joyce Tung

Vice President, Research
23andMe
USA

Eugene Washington is the Chancellor for Health Affairs at Duke University and President and CEO of the Duke University Health System.

Previously, he served from 2010-2015 as Vice Chancellor of Health Sciences, Dean of the David Geffen School of Medicine at the University of California, Los Angeles (UCLA), and Chief Executive Officer of the UCLA Health System. He was also a Distinguished Professor of Gynecology and Health Policy at UCLA. Before UCLA, Washington worked at his alma mater, the University of California, San Francisco, where he took on positions of increasing responsibility, including serving as Executive Vice Chancellor and Provost from 2004 to 2010.

Eugene Washington

Chancellor for Health Affairs,
Duke University
President and CEO,
Duke University Health System
USA

Walter Willett

Professor
Harvard T.H. Chan School of Public Health
Harvard Medical School
USA

Walter Willett M.D., Dr. P.H., is Professor of Epidemiology and Nutrition at Harvard T.H. Chan School of Public Health and Professor of Medicine at Harvard Medical School. Dr. Willett studied food science at Michigan State University, and graduated from the University of Michigan Medical School before obtaining a Masters and Doctorate in Public Health from Harvard T.H. Chan School of Public Health. Dr. Willett has focused much of his work over the last 40 years on the development and evaluation of methods, using both questionnaire and biochemical approaches, to study the effects of diet on the occurrence of major diseases. He has applied these methods starting in 1980 in the Nurses' Health Studies I and II and the Health Professionals Follow-up Study. Together, these cohorts that include nearly 300,000 men and women with repeated dietary assessments, are providing the most detailed information on the long-term health consequences of food choices.

Dr. Willett has published over 1,700 original research papers and reviews, primarily on lifestyle risk factors for heart disease, cancer, and other conditions and has written the textbook, Nutritional Epidemiology, published by Oxford University Press, now in its third edition. He also has written four books for the general public. Dr. Willett is the most cited nutritionist internationally. He is a member of the National Academy of Medicine of the National Academy of Sciences and the recipient of many national and international awards for his research.

Note, the following bios and/or photos were retrieved from online sources: Robert Califf, Philippe Cupers, John Danesh, Hakon Hakonarson, Mary Klotman, Nancy Pedersen, Richard Peto, Cathie Sudlow, Walter Willett.

Location & Venue

JB Duke Hotel
230 Science Dr
Durham, NC 27708, USA
+1 919-660-6400

The JB Duke Hotel is the contemporary hotel located on the campus of Duke University, conveniently near both the university's athletic and academic facilities. Only 20 minutes from Raleigh-Durham International Airport and 10 minutes from Research Triangle Park, the JB Duke Hotel features 198 guestrooms, 25,000 square feet of meeting and event space and is home to a restaurant and two bars, all connected by a stately two-story glass corridor.

View the floor plan

Directions from I-40 West (Coming from the East - RDU Airport, Raleigh, Wilmington):

1. Follow I-40 West to exit 279B, which is Route 147 North (Durham Freeway).
2. Continue on 147 North for approx. 11 miles to exit 16B, which is 15/501 South Bypass.
3. Follow 15/501 South for approx. 1 mile to exit 107, Duke University West Campus.
4. At the bottom of the exit turn left onto NC-751 / Cameron Drive.
5. Continue on NC-751 / Cameron Blvd. to the 3rd traffic light, which is Science Drive.
6. Turn left at Science Drive.
7. Continue on Science Drive for ¼ mile and turn left onto Thomas Center Drive to reach the Entry Plaza of the JB Duke Hotel.

Sponsors

A big thank you to the sponsors of the International Cohorts Summit for their generous contribution and continued support!

Attendees List

- Ada Al-Qunaibet, Saudi National Biobank, Saudi Arabia
- Adam Schlosser, World Economic Forum, USA
- Alexandre Pereira, ELSA-Brazil & Baependi Cohort, Brazil
- Alison Park, University College London, UK
- Andres Metspalu, Estonian Genome Project, Estonia
- Anthony Swerdlow, Generations Study, UK
- Arash Etemadi, Golestan Cohort Study & Persian Cohort Study, USA
- Arthur Holden, Genomic Resources Consortium, Ltd., US
- Beatrice Melin, Northern Sweden Health and Disease Study, Sweden
- Camilla Stoltenberg, Norwegian Mother and Child Cohort Study (MoBa) & Cohort of Norway (CONOR), Norway
- Carolina Haefliger, AstraZeneca Integrated Genomics Initiative, Sweden
- Catherine Schaefer, Kaiser Permanente Research Program on Genes, Environment, and Health, USA
- Cathie Sudlow, UK Biobank, UK
- Chen-Yang Shen, Taiwan Biobank, Taiwan
- Christopher Haiman, Multiethnic Cohort Study, USA
- Cori Bargmann, Chan-Zuckerberg Initiative, USA
- Dan Roden, BioVu Vanderbilt & eMERGE Network, USA
- Daniel MacArthur, Broad Institute of MIT and Harvard, USA
- David Hunter, Harvard University, USA
- David van Heel, East London Genes and Health, UK
- Eric Dishman, U.S. All of Us Research Program, USA
- Eric Green, National Human Genome Research Institute (NHGRI), USA
- Erica Pufall, Wellcome Trust, UK
- Eugene Washington, Duke University, USA
- Farin Kamangar, Golestan Cohort Study & Persian Cohort Study, USA
- Fowzan S Alkuraya, Saudi Human Genome Program, Saudi Arabia
- Francine Grodstein, Nurses' Health Study (NHS), NCI, USA
- Francis Collins, National Institutes of Health (NIH), USA
- Gabriela Repetto, Maule Cohort / MAUCO Study & Universidad del Desarrollo, Chile
- Gad Rennert, Clalit Israeli Genome Project, Israel
- Garnet Anderson, Women's Health Initiative (WHI), USA
- Geoff Ginsburg, Duke University Medical Center & Global Genomic Medicine Collaborative (G2MC), USA
- Goran Walldius, Apolipoprotein MORtality RISk study (AMORIS), Sweden
- Hakon Hakonarson, Children's Hospital of Philadelphia (CHOP) Biorepository, USA
- Heljä-Marja Surcel, Finnish Maternity Cohort Serum Bank, Finland
- J. Michael Gaziano, Million Veteran Program, USA
- Jae-Pil Jeon, Korea Biobank Project, Korea
- Jeremy Farrar, Wellcome Trust, UK
- Jeremy Grushcow, Newfoundland and Labrador Genome Project & Sequence Bio, Canada
- Jesus Alegre Diaz, Mexico City Prospective Study, Mexico

- Joe McNamara, Medical Research Council (MRC), UK
- John Danesh, University of Cambridge, UK
- John Gallacher, Medical Research Council (MRC) Dementia Platform, UK
- Jonathan Emberson, Mexico City Prospective Study, Mexico
- Joshua Denny, Vanderbilt University, USA
- Joyce Tung, 23andMe, USA
- Juan Pablo Casas, UK Collaborative Trial of Ovarian Cancer Screening (UKCTOCS), USA
- Justina Chung, Global Genomic Medicine Collaborative (G2MC) and Global Alliance for Genomics and Health (GA4GH), Canada
- Koichi Matsuda, Biobank Japan, Japan
- Laura Rodriguez, National Institutes of Health (NIH), USA
- Lena Dolman, Global Genomic Medicine Collaborative (G2MC) and Global Alliance for Genomics and Health (GA4GH), Canada
- Lixin Jiang, China PEACE (Patient-centered Evaluative Assessment of Cardiac Events) Million Persons Project, China
- Mads Melbye, Danish National Biobank, Denmark
- Marc Williams, MyCode Community Health Initiative, USA
- Marcel Goldberg, Constances Project, France
- Marie Zins, Constances Project, France
- Mark Caulfield, Genomics England / 100,000 Genomes Project, UK
- Martin McNamara, 45 and Up Study, Australia
- Masayuki Yamamoto, Tohoku Medical Megabank Project, Japan
- Matthew Gillman, Environmental influences on Child Health Outcomes (ECHO) Cohort, USA
- Matthew Nelson, GSK, UK
- Mattias Johansson, European Prospective Investigation into Cancer and Nutrition (EPIC), France
- Michael Phillips, Newfoundland and Labrador Genome Project & Sequence Bio, Canada
- Nancy Pedersen, LifeGene, Sweden
- Neal Freedman, Prostate, Lung, Colorectal and Ovarian Cancer Screening Trial (PLCO), USA
- Nicola Mulder, H3Africa and H3ABioNet, South Africa
- Norie Sawada, Japan Public Health Center-based Prospective Study (JPHC) and JPHC for the Next Generation (JPHC-Next), Japan
- Øyvind Næss, Norwegian Family Based Life Course Study, Norway
- Patrick Tan, Singapore National Precision Medicine Program, Singapore
- Paul Pinsky, Prostate, Lung, Colorectal and Ovarian Cancer Screening Trial (PLCO), USA
- Paulo Lotufo, ELSA-Brazil, Brazil
- Peter Goodhand, Global Alliance for Genomics and Health (GA4GH), Canada
- Philip Awadalla, Ontario Health Study (OHS) & Canadian Partnership for Tomorrow Project, Canada
- Philippe Cupers, European Commission, Belgium
- Rahman Jamal, Malaysian Cohort Study, Malaysia
- Rajesh Dikshit, Barshi Cohort & Tata Memorial Centre, India
- Richard Peto, University of Oxford, UK
- Robert Califf, Duke University, USA
- Robert Eiss, National Institutes of Health (NIH), USA
- Roger Glass, Fogarty International Center, USA
- Rory Collins, UK Biobank, UK

- Sekar Kathiresan, Center for Genomic Medicine, Massachusetts General Hospital, USA
- Stephanie Devaney, U.S. All of Us Research Program, USA
- Sun Ha Jee, Korean Cancer Prevention Study (KCPS), Korea
- Sung Soo Kim, Korean Genome and Epidemiological Study (KoGES) & Korea National Institute of Health, Korea
- Susan Gapstur, Cancer Prevention Study II (CPS-II), USA
- Tejinder Rakhra-Burris, Duke University, USA
- Teresa Zayas Cabán, Office of National Coordinator (ONC) for Health IT, USA
- Teri Manolio, Division of Genomic Medicine, National Human Genome Research Institute (NHGRI), USA
- Terrence Simmons, LIFEPATH (Lifecourse biological pathways underlying social differences in healthy aging), UK
- Thea Norman, Gates Foundation, USA
- Thomas Keane, European Bioinformatics Institute (EMBL-EBI), UK
- Valerie Beral, Million Women Study, UK
- Walter Willett, Nurses' Health Study II (NHSII), NCI, USA
- Yoshinori Murakami, Biobank Japan, Japan
- Zhengming Chen, China Kadoorie Biobank, UK & China

List of Participating Cohorts

23andMe 45 and Up Study All of Us Research Program Apolipoprotein-related MOrtality RISk (AMORIS) AstraZeneca Integrated Genomics Initiative Barshi Cohort Biobank Japan BioVU Vanderbilt Brazilian Longitudinal Study of Adult Health (ELSA-Brasil) Cancer Prevention Study-II (CPS-II) & CPS-II Nutrition subcohort Children's Hospital of Philadelphia (CHOP) Biorepository China Kadoorie Biobank China PEACE Million Persons Project Clalit Israeli Genome Project Cohort of Norway (CONOR) Constances Danish National Biobank East London Genes and Health Environmental influences on Child Health Outcomes (ECHO) Estonian Genome Project European Prospective Investigation into Cancer & Nutrition (EPIC) Finnish Maternity Cohort Serum Bank Generations Study Genomics England Golestan Cohort Study Japan Public Health Center-based Prospective Study (JPHC) Japan Public Health Center-based Prospective Study for the Next generation (JPHC-NEXT) Kaiser Permanente Research Program on Genes, Environment, and Health (RPGEH)	Korea Biobank Project Korean Cancer Prevention Study-II (KCPS-II) Biobank Korean Genome and Epidemiology Study (KoGES) LifeGene (and EpiHealth) LIFEPATH Malaysian Cohort Maule Cohort (MAUCO Study) Mexico City Prospective Study Million Veteran Program Million Women Study Multiethnic Cohort Study (MEC) MyCode Community Health Initiative Newfoundland and Labrador Genome Project Northern Sweden Health and Disease Study Norwegian Family Based Life Course Study Norwegian Mother and Child Cohort Study (MoBa) Nurses' Health Study (NHS) Nurses' Health Study II (NHSII) OHS (Ontario Health Study) and the Canadian Partnership for Tomorrow Project (CPTP) PERSIAN Cohort Study Prostate, Lung, Colorectal and Ovarian Cancer Screening Trial (PLCO) Saudi Human Genome Program Saudi National Biobank Singapore National Precision Medicine Program Taiwan Biobank Tohoku Medical Megabank Project UK Biobank UKCTOCS (UK Collaborative Trial of Ovarian Cancer Screening) Longitudinal Women's Cohort – (UKLWC) Women's Health Initiative (WHI)

Attendee Photos

 Dr. Jesus Alegre (Mexico)	 Rob Eiss (USA)	 Dr. Matthew Gillman (USA)	 Dr. Marcel Goldberg (France)
 Dr. Jeremy Grushcow (Canada)	 Dr. Carolina Haefliger (Sweden)	 Dr. Arthur Holden (USA)	 Dr. Farin Kamangar (USA)
 Dr. Rahman Jamal (Malaysia)	 Dr. Yoshinori Murukami (Japan)	 Dr. Sun Ha Jee (South Korea)	 Dr. Jae-Pil Jeon (South Korea)
 Dr. Mattias Johansson (France	 Dr. Sung Soo Kim (South Korea)	 Dr. Martin McNamara (UK)	 Dr. Andres Metspalu (Estonia)

 **Øyvind Næss (Norway)**	 **Thea Norman (USA)**	 **Dr. Alison Park (United Kingdom)**	 **Dr. Michael Phillips (Canada)**
 **Dr. Nori Sawada (Japan)**	 **Dr. Adam Schlosser (USA)**	 **Dr. Chen-Yang Shen (Taiwan)**	 **Dr. Camilla Stoltenberg (Norway)**
 **Dr. Anthony Swerdlow (UK)**	 **Dr. David van Heel (UK)**	 **Dr. Goran Walldius (Sweden)**	 **Dr. Marie Zins (France)**

Cohort Survey Responses

23andMe

Questions Relating to Cohort	
Name of study	23andMe
Principal Investigator/lead	Dr. Joyce Tung
Contact email	joyce@23andme.com
PubMed ID (or other information) for a protocol/marker paper on this study	Bibliography: https://www.23andme.com/publications/for-scientists/
Study website	For collaborators: https://researchers.23andme.org/collaborations For consumers: https://www.23andme.com/research/
Purpose or major Objectives of study	To help people benefit from the human genome by 1) better understanding, preventing, and treating disease, 2) helping people better understand who they are and how they fit in the larger global community.
Disease areas of focus	
Is your cohort selected for a specific disease (cancer, diabetes) or unselected for disease?	Unselected
Current size of population (and target number of participants)	3,000,000+ genotyped participants (80%+ consented for research)
Participating countries	Primarily US (to a lesser extent UK and other English-speaking countries).
Period of enrollment (and is enrollment on-going?)	Since May 2007. Enrollment ongoing.
Demographic characteristics of participants (age range, proportion male/female, national origin, race)	~54% female; ~77% European, 10% Latino, 5% East/South Asian, 4% African American, 4% other; majority 20-80 y/o
Major diseases or phenotypes collected to date.	Broad, across diseases, lifestyle habits, ancestry, personality, physical traits, etc.
Standardized clinical evaluation components measured, if applicable (e.g. height, weight, blood pressure, exercise testing, spirometry, etc.)	Primarily self-report, with a mix of novel and validated survey instruments. Also, online versions of validated cognitive assessment instruments (e.g., BART, DSST).
Electronic health/medical records or medical administrative data used to collect clinical phenotypes?	No.
Predominant type of electronic health records (e.g. Epic, Cerner, etc.)	
Other sources of clinical data	None broadly collected.

Environmental exposure data being obtained? What sort?	Some collected by self-report.
Other data collected	Data collected primarily by self-report (see phenotypes above).
Biological specimens collected? What sort?	DNA.
Is there a central biobank?	Saliva samples are biobanked (if chosen by participants).
DNA samples prepared (or available to be prepared) from each participant?	Yes
Is genotyping being done on some/all participants?	All.
Is genomic sequencing being done on some/all participants?	
Other molecular analyses performed	Not broadly. Some next-generation sequencing for specific projects.
Did participants provide consent regarding sharing of their data outside the initial study investigators?	Yes. Consenting to participate in the research program is an option given to 23andMe customers and not a requirement to access the 23andMe service.
How are data or specimens from the cohort made available for research? Any limitations on who can access the data (e.g. by country or sector?)	The majority of participants consent to sharing of aggregate/summary-level data with external investigators. Requests can be made through our academic collaborations program (https://researchers.23andme.org/collaborations). A smaller percentage of participants have consented to sharing of de-identified individual-level data with external investigators.
What study information or data are returned to or accessible by participants?	All participants are customers of 23andMe and have access to their raw genotype data as well as reports on genetic results relating to health and ancestry (disease-related reports are FDA-cleared). In addition, the following are also shared with participants: 1) scientific insights gleaned from the research, 2) a full bibliography of 23andMe publications, 3) participant-directed blog posts or newsletters describing results for many of the publications.
Follow-up occurring? (years of follow-up). Is recontact possible?	An ongoing relationship is maintained with participants and many of the surveys collect data longitudinally. Recontact is possible.
Notes/Comments	
Questions Relating to Sharing & Collaboration	
May we make the information you provided about your cohort available on an open website?	Yes
Are you willing to share data from your cohort? If so, would you share:	a) is available on a limited basis based on separate research consent. b, c, and d) are available on a case-by-case basis (we provide whatever is necessary for publication or collaboration).
a) individual data (redacted to protect confidentiality)?	
b) summary data (counts, distributions)?	

c) metadata (descriptive information on data collection methods)?	
d) case report forms and other data collection materials?	
What do you see as the values of sharing?	Value: to benefit people by accelerating scientific discovery in two ways 1) making more results available to the scientific community, 2) bringing our data together with investigators with complementary expertise.
What challenges do you anticipate with sharing?	Challenges: Participant privacy is of paramount importance to us. We need to ensure that their data is being used in a way consistent with their expectations and that their privacy is not being threatened by sharing more data than is necessary or poor data security practices.
What specific legal or regulatory barriers in your country or cohort (aside from ensuring confidentiality and appropriate consent) would prevent you from sharing data with other cohort of cohorts investigators, or openly with anyone who requests them?	We do have some participants from the EU; applicability of the GDPR on data sharing is currently being reviewed.
What aspects of your cohort are intended for translation to clinical care or population health?	We hope to use these data to help develop therapeutics through target discovery and validation, as well as to help people manage and prevent disease through personalized results and recommendations.
How might genomic sequencing add to/enhance your study objectives?	1) Better panels for imputation. 2) Enhanced fine mapping for target discovery. 3) Identification of variants that are private or more commonly found in non-European populations.
Might you be willing to contribute funding or other resources to support international collaboration?	Yes Comment: We share our summary data sets with academic investigators all over the world for free.

45 and Up Study

Questions Relating to Cohort	
Name of study	45 and Up Study
Principal Investigator/lead	Dr. Martin McNamara
Contact email	martin.mcnamara@saxinstitute.org.au
PubMed ID (or other information) for a protocol/marker paper on this study	PMID: 17881411
Study website	https://www.saxinstitute.org.au/our-work/45-up-study/
Purpose or major Objectives of study	To answer important health and quality-of-life questions and help manage and prevent illness through improved knowledge of conditions such as cancer, heart disease, depression, obesity and diabetes in aging individuals.
Disease areas of focus	Mixed

Is your cohort selected for a specific disease (cancer, diabetes) or unselected for disease?	Unselected
Current size of population (and target number of participants)	267,153
Participating countries	Australia
Period of enrollment (and is enrollment on-going?)	2006-2009
Demographic characteristics of participants (age range, proportion male/female, national origin, race)	M/F, 45+ y/o, race/ethnicity inclusive
Major diseases or phenotypes collected to date.	Broad - includes cancer, heart disease, depression, obesity, diabetes
Standardized clinical evaluation components measured, if applicable (e.g. height, weight, blood pressure, exercise testing, spirometry, etc.)	Subset include blood pressure, spirometry, anthropometry, and cognition (for subset)
Electronic health/medical records or medical administrative data used to collect clinical phenotypes?	Subset
Predominant type of electronic health records (e.g. Epic, Cerner, etc.)	Nil
Other sources of clinical data	Questionnaire, Medicare and Pharmaceutical benefits data, the NSW Central Cancer Registry, the NSW Admitted Patients Data Collection, death registration data.
Environmental exposure data being obtained? What sort?	Yes (not specified)
Other data collected	
Biological specimens collected? What sort?	Over 3,000 blood samples collected. Plans underway to collect 50,000 blood samples
Is there a central biobank?	Yes
DNA samples prepared (or available to be prepared) from each participant?	Yes for a subset of approximately 1,000
Is genotyping being done on some/all participants?	Yes for a subset of approximately 1,000
Is genomic sequencing being done on some/all participants?	Yes for a subset of approximately 1,000
Other molecular analyses performed	Nil
Did participants provide consent regarding sharing of their data outside the initial study investigators?	Yes - type not specified

How are data or specimens from the cohort made available for research? Any limitations on who can access the data (e.g. by country or sector?)	Research projects approved by data access committee
What study information or data are returned to or accessible by participants?	Participants receive regular updates on research arising from the study
Follow-up occurring? (years of follow-up). Is recontact possible?	Every 5 years. Contact occurs regularly as part of research projects
Notes/Comments	The largest cohort study ever conducted in Australia to understand ageing. Participant data is also collected retrospectively and prospectively through linked databases. Biospecimens collection has convened with 3,000 blood samples collected to date. Plans underway for the collection of 50,000 blood samples. Purpose: To answer important health and quality-of-life questions and help manage and prevent illness through improved knowledge of conditions such as cancer, heart disease, depression, obesity and diabetes in aging individuals.
Questions Relating to Sharing & Collaboration	
May we make the information you provided about your cohort available on an open website?	Yes
Are you willing to share data from your cohort? If so, would you share:	Yes Comment: The 45 and Up Study data is available for use by researchers. To date over 600 researchers have used the study. Access to study data is provided on a cost recovery basis.
a) individual data (redacted to protect confidentiality)?	Yes
b) summary data (counts, distributions)?	Yes
c) metadata (descriptive information on data collection methods)?	Yes
d) case report forms and other data collection materials?	Yes
What do you see as the values of sharing?	The 45 and Up Study is a collaborative research resource
What challenges do you anticipate with sharing?	TBD
What specific legal or regulatory barriers in your country or cohort (aside from ensuring confidentiality and appropriate consent) would prevent you from sharing data with other cohort of cohorts investigators, or openly with anyone who requests them?	TBD
What aspects of your cohort are intended for translation to clinical care or population health?	Participants data includes linked databases to answer important health questions and help manage and prevent chronic conditions

How might genomic sequencing add to/enhance your study objectives?	TBD
Might you be willing to contribute funding or other resources to support international collaboration?	Yes Comment: In kind contributions would be considered

All of Us Research Program

Questions Relating to Cohort	
Name of study	U.S. Precision Medicine Initiative / All of Us Research Program
Principal Investigator/lead	Dr. Stephanie Devaney
Contact email	stephanie.devaney@nih.gov
PubMed ID (or other information) for a protocol/marker paper on this study	N/A (working on marker paper now, not published yet)
Study website	https://allofus.nih.gov/
Purpose or major Objectives of study	The mission of All of Us is to accelerate health research and medical breakthroughs, enabling individualized prevention, treatment, and care for all of us
Disease areas of focus	None
Is your cohort selected for a specific disease (cancer, diabetes) or unselected for disease?	Unselected
Current size of population (and target number of participants)	1,000,000
Participating countries	USA
Period of enrollment (and is enrollment on-going?)	Enrollment ongoing (as of March 16, 2018 there are 18,906 participants enrolled); anticipate reaching 1 million within 5-6 years
Demographic characteristics of participants (age range, proportion male/female, national origin, race)	Currently only enrolling individuals 18yrs and older, will expand to children within the next year. All of Us is intended to reflect the diversity of the United States with an emphasis on reaching individuals that have been underrepresented in biomedical research (UBR)
Major diseases or phenotypes collected to date.	Unknown at this point
Standardized clinical evaluation components measured, if applicable (e.g. height, weight, blood pressure, exercise testing, spirometry, etc.)	Blood pressure, heart rate, height, hip circumference, waist circumference, weight
Electronic health/medical records or medical administrative data used to collect clinical phenotypes?	Yes

Predominant type of electronic health records (e.g. Epic, Cerner, etc.)	All vendors
Other sources of clinical data	
Environmental exposure data being obtained? What sort?	Self-report environmental data
Other data collected	Surveys on demographics, overall health, lifestyle, health care access and utilization; 2 additional surveys coming in spring: family and personal health history Some of our participants will be sharing activity data from their fitbit
Biological specimens collected? What sort?	Blood, urine
Is there a central biobank?	Yes
DNA samples prepared (or available to be prepared) from each participant?	Yes
Is genotyping being done on some/all participants?	Yes Comment: not yet begun, planning underway
Is genomic sequencing being done on some/all participants?	Yes Comment: WGS planned for 1 million participants
Other molecular analyses performed	Yes, but decisions still pending on the first assays we'll run beyond genomics
Did participants provide consent regarding sharing of their data outside the initial study investigators?	Yes
How are data or specimens from the cohort made available for research? Any limitations on who can access the data (e.g. by country or sector?)	No limitations on sectors that can access, participants signed a general use consent. All of Us has developed access policies that promote broad use of the data within the google cloud environment where the data will reside. The researcher portal is being built with an expected opening in Q1 of 2019
What study information or data are returned to or accessible by participants?	Currently participants receive their physical measures information back. We plan to return nearly all information that participants provide, including individual research results to participants.
Follow-up occurring? (years of follow-up). Is recontact possible?	We will have ongoing follow up with our participants; participants will also be able to set preferences around receiving information back about themselves and the opportunity to participate in other research studies.
Notes/Comments	The All of Us Research Program will seek to extend precision medicine to all diseases by building a national research cohort of one million or more U.S. participants. All of Us will be a participant-engaged, data-driven enterprise supporting research at the intersection of human biology, behavior, genetics, environment, data science and computation, and much more to produce new knowledge with the goal of developing more effective ways to prolong health and treat disease. The cohort will broadly reflect the diversity of the U.S. population by including participants from diverse social, racial/ethnic, and ancestral populations living in a variety of geographies, social environments, and economic circumstances, and from all age groups and health statuses. The longitudinal information gathered from the cohort will be a broad,

	powerful resource for researchers working on a variety of important health questions. Participants volunteer to share core data including their electronic health records, health questionnaire information, and mobile health data on lifestyle habits and environmental exposures. They also undergo a standard baseline evaluation and provide blood and urine samples. Participants will have access to their study information and results, along with summarized data from across the cohort. All of this will be accomplished with essential privacy and security safeguards in place. The All of Us Research Program will be a highly interactive research model, with participants as partners in the development and implementation of the research.
Questions Relating to Sharing & Collaboration	
May we make the information you provided about your cohort available on an open website?	Yes Comment: but would like the chance to update it and vet it more widely before anything goes online
Are you willing to share data from your cohort? If so, would you share:	Generally yes, but would like to understand more what this means and have a couple internal conversations.
a) individual data (redacted to protect confidentiality)?	
b) summary data (counts, distributions)?	
c) metadata (descriptive information on data collection methods)?	
d) case report forms and other data collection materials?	
What do you see as the values of sharing?	We need the various efforts from around the world to be able to leverage lessons learned, and have interoperable data to maximize the scientific value of the data
What challenges do you anticipate with sharing?	Issues in consent language, non-standardized data, any laws that prevent sharing, interoperable platforms.
What specific legal or regulatory barriers in your country or cohort (aside from ensuring confidentiality and appropriate consent) would prevent you from sharing data with other cohort of cohorts investigators, or openly with anyone who requests them?	I'm not aware of any barriers that are explicit prohibitions against us sharing
What aspects of your cohort are intended for translation to clinical care or population health?	
How might genomic sequencing add to/enhance your study objectives?	
Might you be willing to contribute funding or other resources to support international collaboration?	Yes Comment: and in fact we already have by co-sponsoring this meeting!

Apolipoprotein-related MOrtality RISk (AMORIS)

Questions Relating to Cohort	
Name of study	Apolipoprotein-related MOrtality RISk (AMORIS)
Principal Investigator/lead	Dr. Goran Walldius
Contact email	goran.walldius@ki.se
PubMed ID (or other information) for a protocol/marker paper on this study	PMID: 28158674
Study website	http://amoriscohort.imm.ki.se
Purpose or major Objectives of study	To study Metabolic and inflammatory mechanisms in chronic diseases (CVD, diabetes, cancer autoimmune diseases)
Disease areas of focus	Chronic diseases
Is your cohort selected for a specific disease (cancer, diabetes) or unselected for disease?	Unselected
Current size of population (and target number of participants)	812,073
Participating countries	Sweden
Period of enrollment (and is enrollment on-going?)	1985-1996
Demographic characteristics of participants (age range, proportion male/female, national origin, race)	Age range 10-100 plus years, male/female 50/50, about 5% foreigners
Major diseases or phenotypes collected to date.	Linkage to full population health registers in Sweden (hospitalizations, deaths, cancer incidence and more)
Standardized clinical evaluation components measured, if applicable (e.g. height, weight, blood pressure, exercise testing, spirometry, etc.)	In subpopulations
Electronic health/medical records or medical administrative data used to collect clinical phenotypes?	Linkage to full population health registers in Sweden (hospitalizations, deaths, cancer incidence and more). More than 20 years of follow up.
Predominant type of electronic health records (e.g. Epic, Cerner, etc.)	
Other sources of clinical data	Linkage to several other cohorts and quality registers and several research cohorts
Environmental exposure data being obtained? What sort?	Occupational and socioeconomic data available for all and lifestyle data in sub-cohorts
Other data collected	Linkage to Swedish twin register and multi - generation register, medical birth register

Biological specimens collected? What sort?	Fresh blood analyzed at the same CALAB laboratory, Stockholm all years. No frozen samples available.
Is there a central biobank?	No
DNA samples prepared (or available to be prepared) from each participant?	No
Is genotyping being done on some/all participants?	Could be available for some subpopulations i.e. Swedish Twin register
Is genomic sequencing being done on some/all participants?	
Other molecular analyses performed	No
Did participants provide consent regarding sharing of their data outside the initial study investigators?	No
How are data or specimens from the cohort made available for research? Any limitations on who can access the data (e.g. by country or sector?)	Secure Lan (electronic data system) at IMM, KI, Stockholm
What study information or data are returned to or accessible by participants?	None
Follow-up occurring? (years of follow-up). Is recontact possible?	20-30 years
Notes/Comments	The AMORIS study investigates the use of apolipoprotein and other lipids and many of the 500 different lab-analyses (36 million freshly analyzed samples) to predict myocardial infarction stroke, and other CVD diseases, diabetes, kidney disease, several cancer forms, ALS and other chronic diseases. The AMORIS database contains blood samples from mainly healthy individuals (as part of normal checkups) or outpatients referred for lab testing between 1985 to 1996. No samples were taken from hospital wards. More than 100 publications on CVD, diabetes, cancer and inflammatory diseases have been published.
Questions Relating to Sharing & Collaboration	
May we make the information you provided about your cohort available on an open website?	Yes
Are you willing to share data from your cohort? If so, would you share:	We are willing to share data according to our AMORIS Policy document and in accordance with Swedish data protection law and within the current ethical approval.
a) individual data (redacted to protect confidentiality)?	
b) summary data (counts, distributions)?	
c) metadata (descriptive information on data collection methods)?	

d) case report forms and other data collection materials?	
What do you see as the values of sharing?	To investigate if there are any risk markers that may predict the occurrence of the very severe, and often neglected or by clinicians unknown disease ME/CFS (Myalgic Encephalomyelitis/Chronic Fatigue Syndrome)
What challenges do you anticipate with sharing?	
What specific legal or regulatory barriers in your country or cohort (aside from ensuring confidentiality and appropriate consent) would prevent you from sharing data with other cohort of cohorts investigators, or openly with anyone who requests them?	This is a legal matter that needs further considerations and how policy is set up in the Swedish regulatory and health authorities holding registers
What aspects of your cohort are intended for translation to clinical care or population health?	
How might genomic sequencing add to/enhance your study objectives?	Not possible
Might you be willing to contribute funding or other resources to support international collaboration?	We are open to discuss collaboration according to our Policy document

AstraZeneca Integrated Genomics Initiative

Questions Relating to Cohort	
Name of study	AstraZeneca Integrated Genomics Initiative
Principal Investigator/lead	Dr. Ruth March (Representative at the meeting: Carolina Haefliger)
Contact email	Carolina.Haefliger@astrazeneca.com
PubMed ID (or other information) for a protocol/marker paper on this study	NA
Study website	https://www.astrazeneca.com/media-centre/press-releases/2016/AstraZeneca-launches-integrated-genomics-approach-to-transform-drug-discovery-and-development-22042016.html
Purpose or major Objectives of study	To generate genomic information associated with clinical data from AZ sponsored clinical studies
Disease areas of focus	Oncology, respiratory and cardiovascular, renal and metabolic diseases for which AZ is conducting clinical studies
Is your cohort selected for a specific disease (cancer, diabetes) or unselected for disease?	Individual cohorts are defined as per specific study protocol
Current size of population (and target number of participants)	Target in 2016 is up to 500,000

Participating countries	As per study protocol and local country regulations
Period of enrollment (and is enrollment on-going?)	Ongoing and until 2026
Demographic characteristics of participants (age range, proportion male/female, national origin, race)	As per study protocol
Major diseases or phenotypes collected to date.	NA
Standardized clinical evaluation components measured, if applicable (e.g. height, weight, blood pressure, exercise testing, spirometry, etc.)	As per study protocol
Electronic health/medical records or medical administrative data used to collect clinical phenotypes?	Data collected at the sites on electronic case report forms (CRFs)
Predominant type of electronic health records (e.g. Epic, Cerner, etc.)	NA
Other sources of clinical data	NA
Environmental exposure data being obtained? What sort?	As per study protocol
Other data collected	As per study protocol
Biological specimens collected? What sort?	Whole blood samples for DNA; other as per study protocol
Is there a central biobank?	yes
DNA samples prepared (or available to be prepared) from each participant?	Yes, for consenting patients
Is genotyping being done on some/all participants?	As per study protocol
Is genomic sequencing being done on some/all participants?	Only in patients who consent for exploratory genetic research
Other molecular analyses performed	As per study protocol
Did participants provide consent regarding sharing of their data outside the initial study investigators?	Yes (consent for exploratory genetics is optional)
How are data or specimens from the cohort made available for research? Any limitations on who can access the data (e.g. by country or sector?)	De-identified patient level data may be accessed by request after submission of a proposal through the Data Request Portal (for details see https://astrazenecagroup-dt.pharmacm.com//DT/Home/Index/).
What study information or data are returned to or accessible by participants?	As per study protocol. Exploratory genomics will not be returned to the patients

Follow-up occurring? (years of follow-up). Is recontact possible?	As per study protocol. No recontact after study finalization
Notes/Comments	AstraZeneca and its collaborators hope to unearth rare genetic sequences that are associated with disease and with responses to treatment. AstraZeneca will partner with research institutions including the Wellcome Trust Sanger Institute and Human Longevity. AstraZeneca also expects to draw on data from 500,000 participants in its own clinical trials, and medical samples that it has accrued over the past 15 years.
Questions Relating to Sharing & Collaboration	
May we make the information you provided about your cohort available on an open website?	No Comment: Data can be shared as per AZ policies https://astrazenecagrouptrials.pharmacm.com/ST/Submission/Search
Are you willing to share data from your cohort? If so, would you share:	Yes Comment: see specific requirements below
a) individual data (redacted to protect confidentiality)?	Comment: De-identified patient level data may be accessed by request after submission of a proposal through the Data Request Portal (for details see https://astrazenecagroup-dt.pharmacm.com//DT/Home/Index/).
b) summary data (counts, distributions)?	Comment: through a website where the public and trial participants can find published Trial Results Summaries for trials that started in late 2015 and beyond (https://www.trialsummaries.com/Home/LandingPage)
c) metadata (descriptive information on data collection methods)?	Comment: may be accessed by request after submission of a proposal through the Data Request Portal (for details see https://astrazenecagroup-dt.pharmacm.com//DT/Home/Index/
d) case report forms and other data collection materials?	Comment: may be accessed by request after submission of a proposal through the Data Request Portal (for details see https://astrazenecagroup-dt.pharmacm.com//DT/Home/Index/
What do you see as the values of sharing?	Advancement of medical research; increase understanding of underlying causes of diseases and patient subgroups; collaboration with groups with similar interests; increase cohort size and thus statistical power to find rare variants
What challenges do you anticipate with sharing?	Data privacy and confidentiality; acceptance of broad consent by EC/IRBs; pre-competitive vs competitive space
What specific legal or regulatory barriers in your country or cohort (aside from ensuring confidentiality and appropriate consent) would prevent you from sharing data with other cohort of cohorts investigators, or openly with anyone who requests them?	Depends on the local regulations of the countries in which individual studies are conducted
What aspects of your cohort are intended for translation to clinical care or population health?	Clinical studies are interventional
How might genomic sequencing add to/enhance your study objectives?	Patient selection; improve understanding of efficacy and safety; definition of patient subgroups

Might you be willing to contribute funding or other resources to support international collaboration?	Proposals will be assessed on a case by case basis regarding scientific merit and resource requirements, among other criteria

Barshi Cohort

Questions Relating to Cohort	
Name of study	Barshi Cohort
Principal Investigator/lead	Dr. Rajesh Dikshit
Contact email	dixr24@hotmail.com
PubMed ID (or other information) for a protocol/marker paper on this study	
Study website	
Purpose or major Objectives of study	To enroll 200,000 individuals in age group of 30-69 to understand risk for NCD related to dietary habits, smokeless tobacco, indoor air pollution, physical activities and obesity
Disease areas of focus	NCD
Is your cohort selected for a specific disease (cancer, diabetes) or unselected for disease?	Unselected
Current size of population (and target number of participants)	21,000 (target: 200,000)
Participating countries	Barshi rural (364 villages) and Barshi Town
Period of enrollment (and is enrollment on-going?)	Ongoing. Will finish enrolment by 2019.
Demographic characteristics of participants (age range, proportion male/female, national origin, race)	Yes
Major diseases or phenotypes collected to date.	CVD, Cancer, Stroke
Standardized clinical evaluation components measured, if applicable (e.g. height, weight, blood pressure, exercise testing, spirometry, etc.)	Height (standing and sitting) , weight, blood pressure , spirometry, hands grip strength, Body impedance, waist , hip measurement
Electronic health/medical records or medical administrative data used to collect clinical phenotypes?	Yes
Predominant type of electronic health records (e.g. Epic, Cerner, etc.)	
Other sources of clinical data	

Environmental exposure data being obtained? What sort?	Plan to keep monitors indoors and outdoors to measure PM2.5 levels
Other data collected	General Mood, Family history , Personal medical History, reproductive history for females, detailed FFQ
Biological specimens collected? What sort?	Blood samples, Nail Samples
Is there a central biobank?	Yes, Central automated bio bank
DNA samples prepared (or available to be prepared) from each participant?	For some of the samples, aliquots are ready to extract DNA.
Is genotyping being done on some/all participants?	Not on cohort participants, but method has been standardized using Illumina platform.
Is genomic sequencing being done on some/all participants?	
Other molecular analyses performed	Not yet
Did participants provide consent regarding sharing of their data outside the initial study investigators?	No
How are data or specimens from the cohort made available for research? Any limitations on who can access the data (e.g. by country or sector?)	
What study information or data are returned to or accessible by participants?	BMI and blood pressure data
Follow-up occurring? (years of follow-up). Is recontact possible?	Yes
Notes/Comments	
Questions Relating to Sharing & Collaboration	
May we make the information you provided about your cohort available on an open website?	Not yet
Are you willing to share data from your cohort? If so, would you share:	Summary data, but after obtaining clearance from ICMR
a) individual data (redacted to protect confidentiality)?	
b) summary data (counts, distributions)?	Yes (but after obtaining clearance from ICMR)
c) metadata (descriptive information on data collection methods)?	
d) case report forms and other data collection materials?	

What do you see as the values of sharing?	Sharing on Occupational and pollution data will be useful
What challenges do you anticipate with sharing?	
What specific legal or regulatory barriers in your country or cohort (aside from ensuring confidentiality and appropriate consent) would prevent you from sharing data with other cohort of cohorts investigators, or openly with anyone who requests them?	We need to take Health Ministry clearance.
What aspects of your cohort are intended for translation to clinical care or population health?	We intend to collect tissue of enrolled participants who are diagnosed with Cancer.
How might genomic sequencing add to/enhance your study objectives?	One of the objective of the study is also to get genomic information. We plan to prepare India specific chip.
Might you be willing to contribute funding or other resources to support international collaboration?	Yes

Biobank Japan

Questions Relating to Cohort	
Name of study	Biobank Japan
Principal Investigator/lead	Dr. Yoshinori Murakami
Contact email	ymurakam@ims.u-tokyo.ac.jp
PubMed ID (or other information) for a protocol/marker paper on this study	Hirata M et al. Overview of BioBank Japan Follow- up Data in 32 Diseases**. J Epidemiology**, 27, 22-28, 2017. (doi: 10.1016/j.je.2016.12.006.) Hirata M et al. Cross-sectional analysis of BioBank Japan Clinical Data: A Large Cohort of 200,000 Patients with 47 Common Diseases. **J Epidemiology**, 27, 9-21, 2017. (doi: 10.1016/j.je.2016.12.003.) Nagai A et al. Overview of the BioBank Japan Project: Study Design and Profile. **J Epidemiology**, 27:2-8, 2017. (doi: 10.1016/j.je.2016.12.005.) Kanai M et al. Genetic analysis of quantitative traits in the Japanese population links cell types to complex human diseases. **Nat Genet**, 2018. https://doi.org/10.1038/ s41588-018-0047-6.
Study website	https://biobankjp.org/english/index.html http://www.src.riken.jp/english/project/person/
Purpose or major Objectives of study	To apply genomic research results to healthcare by providing a detailed understanding of the mechanisms causing disease and its symptoms, and to develop new diagnostic methods and innovative drugs by targeting the genes, proteins or other biological components that are

	proven to be the cause of the disease or symptom.
Disease areas of focus	Mixed
Is your cohort selected for a specific disease (cancer, diabetes) or unselected for disease?	Broad - 47 diseases (1st cohort) , 38 diseases (2nd cohort), including 14 cancer, CVD, liver disease, kidney disease. Total 51 diseases.
Current size of population (and target number of participants)	270,000 (target: 270,000)
Participating countries	Japan
Period of enrollment (and is enrollment on-going?)	2003-2018. No additional enrollment after April, 2018.
Demographic characteristics of participants (age range, proportion male/female, national origin, race)	M/F, all age groups, Japanese population
Major diseases or phenotypes collected to date.	Broad - 47 diseases (1st cohort), 38 diseases (2nd cohort), including 14 cancer, CVD, liver disease, kidney disease. Total 51 diseases.
Standardized clinical evaluation components measured, if applicable (e.g. height, weight, blood pressure, exercise testing, spirometry, etc.)	All samples were collected from patients from specified 51 diseases at more than 50 hospitals in Japan. Clinical information of 2,472 issues for all patients, including height and weight, as well as some disease-specific issues, including results of clinical tests, were collected for each case.
Electronic health/medical records or medical administrative data used to collect clinical phenotypes?	No
Predominant type of electronic health records (e.g. Epic, Cerner, etc.)	
Other sources of clinical data	Questionnaire
Environmental exposure data being obtained? What sort?	Occupational questions are incuded.
Other data collected	Diet, lifestyle
Biological specimens collected? What sort?	DNA and serum for 1st cohort. DNA for 2nd cohort.
Is there a central biobank?	All samples are collected and stored in a central biobank.
DNA samples prepared (or available to be prepared) from each participant?	Yes
Is genotyping being done on some/all participants?	Yes for 1st cohort of 200, 000 cases. Not done for 2nd cohort of 70,000 cases.
Is genomic sequencing being done on some/all participants?	Done for 1,000 cases.
Other molecular analyses performed	No.
Did participants provide consent regarding sharing of their data outside the initial study investigators?	Yes - for research purpose

How are data or specimens from the cohort made available for research? Any limitations on who can access the data (e.g. by country or sector?)	Committee approval to institutes with subsidiary in Japan
What study information or data are returned to or accessible by participants?	No
Follow-up occurring? (years of follow-up). Is recontact possible?	Clinical data are followed up for 11 years on average and serums are additionally collected in some cases through each hospital and stocked in BioBank Japan for1st cohort. Death information are collected for 1st and 2nd cohort.
Notes/Comments	To apply genomic research results to healthcare by providing a detailed understanding of the mechanisms causing disease and its symptoms, and to develop new diagnostic methods and innovative drugs by targeting the genes, proteins or other biological components that are proven to be the cause of the disease or symptom. Sample collection will be ended in March 2018, whereas distributions of DNA, serum, clinical data and genotyping data will be continued for 1st and 2nd cohort.
Questions Relating to Sharing & Collaboration	
May we make the information you provided about your cohort available on an open website?	_x_ Yes ___ No Comment:
Are you willing to share data from your cohort? If so, would you share:	___ Yes ___ No Comment:
a) individual data (redacted to protect confidentiality)?	___ Yes _x__ No Comment: Due to limitation of participants' agreement, ELSI and governmental policy
b) summary data (counts, distributions)?	_x_ Yes ___ No Comment:
c) metadata (descriptive information on data collection methods)?	_x__ Yes ___ No Comment:
d) case report forms and other data collection materials?	___ Yes __x_ No Comment:
What do you see as the values of sharing?	We recognize the importance of data sharing and try to discuss with members of ELSI committee in Japan and the Japanese government.
What challenges do you anticipate with sharing?	Summary data would be no problem to be shared. Individual genomic information with clinical data would be relatively difficult to be shared at present due to ELSI limitation.
What specific legal or regulatory barriers in your country or cohort (aside from ensuring confidentiality and appropriate consent) would prevent you from sharing data with other cohort of cohorts investigators, or openly with anyone who requests them?	In Japan, we have an amendment to the Act on the Protection of Personal Information since, 2017. In this Act, genomic information of more than 41 SNPs are recognized as a personal information, while medical records are recognized as a personal information required to be considered, even in a field of medical science and genome science.
What aspects of your cohort are intended for translation to clinical care or population health?	We are planning to feedback some of genomic information involved in or associated with various diseases to relevant participants in next 5 years. To do this, we need to overcome ELS issues in Japan.

How might genomic sequencing add to/enhance your study objectives?	Genomic sequencing is essential to understand the genomic background of various diseases. We are planning to start genomic sequencing of a portion of 1st and 2nd cohort by disease-based collaboration and funding.
Might you be willing to contribute funding or other resources to support international collaboration?	_x_ Yes ___ No Comment:

BioVU Vanderbilt

Questions Relating to Cohort	
Name of study	BioVU Vanderbilt
Principal Investigator/lead	Dr. Dan Roden
Contact email	dan.roden@vanderbilt.edu
PubMed ID (or other information) for a protocol/marker paper on this study	PMIDs: 18500243; 20443953; 24786321
Study website	https://victr.vanderbilt.edu/pub/biovu/
Purpose or major Objectives of study	To study the links between genes and disease and between genes and how a patient might respond to a prescribed medication.
Disease areas of focus	Mixed
Is your cohort selected for a specific disease (cancer, diabetes) or unselected for disease?	Unselected
Current size of population (and target number of participants)	>244,000 (undefined target)
Participating countries	US
Period of enrollment (and is enrollment on-going?)	2007-ongoing
Demographic characteristics of participants (age range, proportion male/female, national origin, race)	M/F, all ages, race/ethnicity inclusive 57% female 13% AA
Major diseases or phenotypes collected to date.	Broad
Standardized clinical evaluation components measured, if applicable (e.g. height, weight, blood pressure, exercise testing, spirometry, etc.)	None. All phenotypes derived from a deidentified image of the EHR.
Electronic health/medical records or medical administrative data used to collect clinical phenotypes?	Yes

Predominant type of electronic health records (e.g. Epic, Cerner, etc.)	Star panel (VUMC-developed). Converted to Epic on 11/02/17.
Other sources of clinical data	None
Environmental exposure data being obtained? What sort?	No. Only as represented in the EHR
Other data collected	No. Only as represented in the EHR
Biological specimens collected? What sort?	Plasma samples being piloted for selected phenotypes
Is there a central biobank?	Yes
DNA samples prepared (or available to be prepared) from each participant?	Yes
Is genotyping being done on some/all participants?	Yes
Is genomic sequencing being done on some/all participants?	Yes
Other molecular analyses performed	
Did participants provide consent regarding sharing of their data outside the initial study investigators?	Yes. Included in consent
How are data or specimens from the cohort made available for research? Any limitations on who can access the data (e.g. by country or sector?)	Access only to VUMC investigators given that EHR is the source of phenotype information. Each project must be approved by data access committee
What study information or data are returned to or accessible by participants?	None
Follow-up occurring? (years of follow-up). Is recontact possible?	As new information accrues in the EHR, it is made available to the deidentified image. No recontact.
Notes/Comments	To study the links between genes and disease and between genes and how a patient might respond to a prescribed medication.
Questions Relating to Sharing & Collaboration	
May we make the information you provided about your cohort available on an open website?	Yes
Are you willing to share data from your cohort? If so, would you share:	Yes
a) individual data (redacted to protect confidentiality)?	
b) summary data (counts, distributions)?	Yes
c) metadata (descriptive information on data collection methods)?	Yes

d) case report forms and other data collection materials?	N/A
What do you see as the values of sharing?	We have been part of eMERGE since the beginning. Bigger datasets enable more robust genotypeàphenotype and phenotypeàgenotype analyses
What challenges do you anticipate with sharing?	
What specific legal or regulatory barriers in your country or cohort (aside from ensuring confidentiality and appropriate consent) would prevent you from sharing data with other cohort of cohorts investigators, or openly with anyone who requests them?	
What aspects of your cohort are intended for translation to clinical care or population health?	That has to be the ultimate goal
How might genomic sequencing add to/enhance your study objectives?	Greater coverage. Identify rare variants with large effect sizes
Might you be willing to contribute funding or other resources to support international collaboration?	Unknown

Brazilian Longitudinal Study of Adult Health (ELSA-Brasil)

Questions Relating to Cohort	
Name of study	Brazilian Longitudinal Study of Adult Health (ELSA-Brasil)
Principal Investigator/lead	Dr. Paulo A. Lotufo
Contact email	palotufo@usp.br
PubMed ID (or other information) for a protocol/marker paper on this study	https://www.ncbi.nlm.nih.gov/pubmed/22234482 https://www.ncbi.nlm.nih.gov/pubmed/24585730 https://www.ncbi.nlm.nih.gov/pubmed/24346715 https://www.ncbi.nlm.nih.gov/pubmed/24346723
Study website	www.elsa.org.br
Purpose or major Objectives of study	To improve prevention, diagnosis, and treatment of adulthood/elderly diseases
Disease areas of focus	Noncommunicable diseases, mainly heart, stroke, renal, and diabetes.
Is your cohort selected for a specific disease (cancer, diabetes) or unselected for disease?	Unselected
Current size of population (and target number of participants)	15,105
Participating countries	Brazil: six cities

Period of enrollment (and is enrollment on-going?)	Baseline: 2008-10; second visit: 2012-14; third visit: ongoing since 2017
Demographic characteristics of participants (age range, proportion male/female, national origin, race)	Sex Men 6887 (45.6%); Women 8218 (54.4%)
Major diseases or phenotypes collected to date.	Cardiovascular, diabetes by OGTT, renal, thyroid, musculoskeletal, hearing & vision, mental health, cognitive function, common symptoms as headache.
Standardized clinical evaluation components measured, if applicable (e.g. height, weight, blood pressure, exercise testing, spirometry, etc.)	All cohort: height (standing/sitting), weight with body impedance; circumferences (neck/waist/hip); blood pressure (arm and brachial); retinoscopy; hepatic sonogram; echocardiogram; hand-grip strength; ECG at resting and heart rate variability; pulse wave velocity; IMT carotid a. and femoral a.; exercise testing.
Electronic health/medical records or medical administrative data used to collect clinical phenotypes?	No, only to support events adjudication
Predominant type of electronic health records (e.g. Epic, Cerner, etc.)	
Other sources of clinical data	Mortality system; high-cost procedures system (CABG, dialysis).
Environmental exposure data being obtained? What sort?	Neighborhood air pollution with PM2, NOx, O3; urban traffic maps; green areas maps.
Other data collected	
Biological specimens collected? What sort?	Fasting: serum, plasma, plasma-EDTA, plasma-heparin, plasm-citrate. After OGTT: serum, plasma-heparin. Urine. Subsetting: cells.
Is there a central biobank?	Yes
DNA samples prepared (or available to be prepared) from each participant?	Yes
Is genotyping being done on some/all participants?	Yes, genetic ancestry for all participants. GWAS and WES data are ongoing for a subset of participants. Current funds request to extend to the entire sample.
Is genomic sequencing being done on some/all participants?	Yes, WES on a subset. No current resources to extend sequencing to the entire cohort.
Other molecular analyses performed	Lipid subfractions by NMR
Did participants provide consent regarding sharing of their data outside the initial study investigators?	No, but IRBs can grant sharing data permission.
How are data or specimens from the cohort made available for research? Any limitations on who can access the data (e.g. by country or sector?)	All data is stored in datacenter located in Porto Alegre, and all biosamples in Sao Paulo. All access should be authorized by the P&P Committee or the Steering Committee.
What study information or data are returned to or accessible by participants?	Broad. By default, a panel of test results is sent, but by request, all information can be returned.

Follow-up occurring? (years of follow-up). Is recontact possible?	Ten years of yearly by phone. There is a contact in case of ancillary studies. So far, we have approximately 99% of follow-up
Notes/Comments	ELSA-Brasil enrolled 15105 apparently healthy men and women aged 35-74 years-old between 2008-10. Further, a second (2012-14) and third visits (ongoing) have been done. There are six sites in different cities (Sao Paulo, Rio de Janeiro, Belo Horizonte, Porto Alegre, Salvador, Vitoria). ELSA-Brasil is the first large prospective study performed in Latin American countries addressing non-communicable diseases. The planning design is ambitious with exposures since de ZIP code to the genetic code, including social, psychological, and clinical variables. Another characteristic was to follow mid-aged adults to study detailed the determinants of weight gain, mental health, and cognition and their relationship with cardiovascular outcomes. We have collaboration with US cohorts as MESA, ARIC, FHS, NOMAS, and WHS.
Questions Relating to Sharing & Collaboration	
May we make the information you provided about your cohort available on an open website?	Yes
Are you willing to share data from your cohort? If so, would you share:	Yes
a) individual data (redacted to protect confidentiality)?	Yes Comment: We are doing this procedure with researchers who work with the same processes as no charge for access and interchange (reciprocity).
b) summary data (counts, distributions)?	Yes
c) metadata (descriptive information on data collection methods)?	Yes
d) case report forms and other data collection materials?	Yes
What do you see as the values of sharing?	All possible advantages.
What challenges do you anticipate with sharing?	One of the issues will be the personnel cost associated with data extraction and harmonization.
What specific legal or regulatory barriers in your country or cohort (aside from ensuring confidentiality and appropriate consent) would prevent you from sharing data with other cohort of cohorts investigators, or openly with anyone who requests them?	There is no extra regulatory burden to share summary-level statistics. These can be made available to collaborators if the project for their derivation is locally approved.
What aspects of your cohort are intended for translation to clinical care or population health?	
How might genomic sequencing add to/enhance your study objectives?	Sharing individual-level data with international collaborators requires specific approval from the Brazilian Federal IRB. This procedure is usually on a case-by-case basis (project and collaborator-wise) and

	poses a significant burden regarding processing time. It is the way to go for specific types of data-sharing when there is a particular hypothesis being tested (for example, this has been the chosen mechanism for current on-going collaborations with the TOPMED or other NHGRI initiatives where individual-level genetic data is being made available).
Might you be willing to contribute funding or other resources to support international collaboration?	

Cancer Prevention Study-II (CPS-II) & CPS-II Nutrition subcohort

Questions Relating to Cohort	
Name of study	Cancer Prevention Study-II (CPS-II)
Principal Investigator/lead	Dr. Susan Gapstur
Contact email	susan.gapstur@cancer.org
PubMed ID (or other information) for a protocol/marker paper on this study	PMID: 4047150 (Natl Cancer Inst Monogr. 1985 May;67:49-52.) PMID: 12015775 (Cancer. 2002 May 1;94(9):2490-501)
Study website	http://www.cancer.org/research/researchtopreventcancer/currentcancerpreventionstudies/index
Purpose or major Objectives of study	To identify, and/or confirm cancer risk factors, and improve relative risk estimates
Disease areas of focus	Cancer incidence and mortality, total mortality
Is your cohort selected for a specific disease (cancer, diabetes) or unselected for disease?	Unselected
Current size of population (and target number of participants)	CPS-II Mortality Cohort: 1,185,106 CPS-II Nutrition Subcohort: 184,194
Participating countries	United States
Period of enrollment (and is enrollment on-going?)	CPS-II Mortality Cohort: 1982/83 CPS-II Nutrition Subcohort: 1992/93 (followed for cancer incidence as well)
Demographic characteristics of participants (age range, proportion male/female, national origin, race)	CPS-II Mortality Cohort: 30-111 years; 43% male/57% female; N/A; 93% white/4% Black/3% other CPS-II Nutrition Subcohort: 40-92 years; 47% male/53% female; N/A; 97% white/1% Black/2% other
Major diseases or phenotypes collected to date.	CPS-II Mortality Cohort: Specific cause of death (99% of known deaths) CPS-II Nutrition Subcohort: verified cancer incidence; self-reported other diseases (e.g., diabetes, stroke)
Standardized clinical evaluation components measured, if applicable (e.g. height, weight, blood pressure, exercise testing, spirometry, etc.)	See other data collection under for self-reported information

Electronic health/medical records or medical administrative data used to collect clinical phenotypes?	CPS-II Nutrition Subcohort: Medical record or tumor registries
Predominant type of electronic health records (e.g. Epic, Cerner, etc.)	
Other sources of clinical data	None
Environmental exposure data being obtained? What sort?	Some geo spacial data are available. Please inquire for specific information if interested.
Other data collected	CPS-II Mortality Cohort: Self-reported lifestyle, behavioral, reproductive history, medical history, height and weight, medication use questionnaires CPS-II Nutrition Subcohort: Self-reported lifestyle, behavioral, reproductive history, medical history, weight and weight history, medication use questionnaires
Biological specimens collected? What sort?	Blood (approximately 40,000 participants) with plasma, serum, RBCs, and lymphocytes stored at -80C Mouthwash samples (approximately 70,000 participants without a blood sample) with cell pellet for DNA/other; and supernatant stored.= at -80C tumor tissue samples for approximately 30% of breast, prostate, colorectal and hematologic cancer cases.
Is there a central biobank?	Yes
DNA samples prepared (or available to be prepared) from each participant?	Extracted DNA available for approximately 40,000 participants
Is genotyping being done on some/all participants?	GWAS data available on approximately 40,000
Is genomic sequencing being done on some/all participants?	
Other molecular analyses performed	Some tissue molecular markers for breast, colorectal, and hematologic cancers
Did participants provide consent regarding sharing of their data outside the initial study investigators?	We are able to share data
How are data or specimens from the cohort made available for research? Any limitations on who can access the data (e.g. by country or sector?)	Please see our Cancer Prevention Studies Data Access Policies and Procedures which can be found on our website at: https://www.cancer.org/content/dam/cancer-org/research/epidemiology/cancer-prevention-study-data-access-policies.pdf
What study information or data are returned to or accessible by participants?	None
Follow-up occurring? (years of follow-up). Is recontact possible?	20+
Notes/Comments	The American Cancer Society's cancer prevention studies help researchers identify cancer risk factors by allowing them to study large groups of people over long periods of time.

Questions Relating to Sharing & Collaboration	
May we make the information you provided about your cohort available on an open website?	Yes
Are you willing to share data from your cohort? If so, would you share:	We share individual and summary data as described in our Cancer Prevention Studies Data Access Policies and Procedures which can be found on our website at: https://www.cancer.org/content/dam/cancer-org/research/epidemiology/cancer-prevention-study-data-access-policies.pdf Meta-data are also available on our website (i.e., surveys) and methods of data collected described in papers references above.
a) individual data (redacted to protect confidentiality)?	
b) summary data (counts, distributions)?	
c) metadata (descriptive information on data collection methods)?	
d) case report forms and other data collection materials?	
What do you see as the values of sharing?	Extending expertise; utilizing a valuable resource to its fullest
What challenges do you anticipate with sharing?	
What specific legal or regulatory barriers in your country or cohort (aside from ensuring confidentiality and appropriate consent) would prevent you from sharing data with other cohort of cohorts investigators, or openly with anyone who requests them?	Conflicts of interest; research or researcher supported by the tobacco industry. Other considerations will depend on staff availability; support from collaboration; overlap; expertise of collaborator.
What aspects of your cohort are intended for translation to clinical care or population health?	All
How might genomic sequencing add to/enhance your study objectives?	Our main objective is cancer prevention and identifying germline or somatic mutations would allow us to contribute to the study of gene and environment interactions as well as the characterization of penetrance and prevalence of these mutations in large collaborative studies.
Might you be willing to contribute funding or other resources to support international collaboration?	This is decided on an individual basis and may depend on availability of funds, mission relevance, or other organizational considerations.

Children’s Hospital of Philadelphia (CHOP) Biorepository

Questions Relating to Cohort	
Name of study	Children's Hospital of Philadelphia (CHOP) Biorepository

Principal Investigator/lead	Dr. Hakon Hakonarson
Contact email	hakonarson@email.chop.edu
PubMed ID (or other information) for a protocol/marker paper on this study	PMIDs: 16778028, 23743551, 28190457
Study website	https://caglab.org/index.php/for-researchers/biorepository.html.
Purpose or major Objectives of study	Large scale pediatric biobank for the study of complex pediatric disorders, both common and rare. Over 500 publications have resulted from this effort over the past 10 years.
Disease areas of focus	All pediatric disorders, both common and rare
Is your cohort selected for a specific disease (cancer, diabetes) or unselected for disease?	Unselected for disease, random recruitment, common diseases building up most rapidly (e.g., over 17,000 asthma cases in the biobank)
Current size of population (and target number of participants)	Over 500,000 unique samples, each in multiple aliquots, are currently in the biobank (including 120,000 pediatric samples from CHOP). Goal is to reach 1M. Over 95% are blood samples (less than 5% are saliva); isolation is made of DNA, PBMCs and plasma from all samples
Participating countries	The key pediatric samples are from CHOP. Collaborative samples are from all over the world, including, US, Europe, South America, Canada, Saudi Arabia and Australia
Period of enrollment (and is enrollment on-going?)	Enrollment began in July 2006. Enrollment is ongoing with over 15k new samples added every year from CHOP alone
Demographic characteristics of participants (age range, proportion male/female, national origin, race)	0-21 year old; 50/50 M/F, 51% Caucasian, 38% AA, 7% Asian and 4% other
Major diseases or phenotypes collected to date.	All pediatric diseases both common and rare
Standardized clinical evaluation components measured, if applicable (e.g. height, weight, blood pressure, exercise testing, spirometry, etc.)	EPIC EHR data abstraction updated every 6 months. Recontact permission of 90%; 2 page survey done on all subjects with standard questions on diseases and health and medications. VS and height/weight available on multiple measures
Electronic health/medical records or medical administrative data used to collect clinical phenotypes?	Yes EHR (EPIC)
Predominant type of electronic health records (e.g. Epic, Cerner, etc.)	Epic
Other sources of clinical data	Various other databases accessed (EKG, ECHO, EEG, PFTs, Sleep and other) plus standard survey on all
Environmental exposure data being obtained? What sort?	Not standardized; some available in Epic. Recontact to 90% so can be collected
Other data collected	
Biological specimens collected? What sort?	Paediatric biological samples (blood, PBMCs, plasma)
Is there a central biobank?	CHOP has a central biobank that is independent of the CAG biobank which was put in place for a specific purpose. CAG samples make up

	>90% of all samples at CHOP
DNA samples prepared (or available to be prepared) from each participant?	Yes, already prepared
Is genotyping being done on some/all participants?	About ¾ completed
Is genomic sequencing being done on some/all participants?	Done on about 20k
Other molecular analyses performed	Yes, project specific (team of 8 translational scientist at CAG)
Did participants provide consent regarding sharing of their data outside the initial study investigators?	Yes
How are data or specimens from the cohort made available for research? Any limitations on who can access the data (e.g. by country or sector?)	Upon request – access controlled by CAG
What study information or data are returned to or accessible by participants?	We send news letters. We have Return-Of-Results program in place through eMERGE. We also run a CLIA based clinical program and return results through there.
Follow-up occurring? (years of follow-up). Is recontact possible?	Recontact for 90%. For example, the Philadelphia Neurodevelopmental Cohort (PNC) study brought back 10,000 subjects already enrolled in CAG for detailed neurocognitive assessment and imaging.
Notes/Comments	The world's largest pediatric biorepository connects DNA to the hospital's health records for studies of childhood diseases. Associated with the Children's Brain Tumour Tissue Consortium (CBTTC). From Hakon: "we have 112,000+ CHOP patient samples recruited (almost all are GWAS genotyped and about 20% are sequenced), we have blood, DNA, PBMCs and plasma from those subjects all of whom are linked to EHR that we update every 6 months and about 90% have signed off for recontact permission. In addition, I have almost 400k collaborative samples in the CHOP biobank from all over the world (so over 500k unique individuals in total) we have used for replication/validation of CHOP findings. I have published over 550 papers from this assed in the past 10 years"
Questions Relating to Sharing & Collaboration	
May we make the information you provided about your cohort available on an open website?	Yes
Are you willing to share data from your cohort? If so, would you share:	Yes
a) individual data (redacted to protect confidentiality)?	Yes Comment: In proper collaborative setting
b) summary data (counts, distributions)?	Yes
c) metadata (descriptive information on data collection methods)?	Yes

d) case report forms and other data collection materials?	Yes
What do you see as the values of sharing?	Collaborative opportunities and expansion of results, career development of junior staff
What challenges do you anticipate with sharing?	It requires substantive effort to handle and manage sharing – funding is always an issue
What specific legal or regulatory barriers in your country or cohort (aside from ensuring confidentiality and appropriate consent) would prevent you from sharing data with other cohort of cohorts investigators, or openly with anyone who requests them?	None
What aspects of your cohort are intended for translation to clinical care or population health?	We have multiple programs in clinical development and other in late stage of translation for same – this is our main goal – to move our discoveries into the clinic
How might genomic sequencing add to/enhance your study objectives?	Would transform our abilities to make new discoveries
Might you be willing to contribute funding or other resources to support international collaboration?	Yes Comment: If I have funding I would, currently

China Kadoorie Biobank

Questions Relating to Cohort	
Name of study	China Kadoorie Biobank
Principal Investigator/lead	Professor Zhengming Chen (Oxford) & Professor Liming Li (Beijing)
Contact email	zhengming.chen@ctsu.ox.ac.uk; lmleeph@vip.163.com
PubMed ID (or other information) for a protocol/marker paper on this study	
Study website	http://www.ckbiobank.org/site/
Purpose or major Objectives of study	To investigate the main genetic and environmental causes of common chronic diseases in the Chinese population.
Disease areas of focus	Mixed
Is your cohort selected for a specific disease (cancer, diabetes) or unselected for disease?	Unselected
Current size of population (and target number of participants)	512,891
Participating countries	China

Period of enrollment (and is enrollment on-going?)	2004-2008
Demographic characteristics of participants (age range, proportion male/female, national origin, race)	M/F, 30-79 y/o, of Chinese descent
Major diseases or phenotypes collected to date.	Broad, including heart attack, stroke, diabetes, cancer, COPD, chronic liver and renal diseases. After 10 years' follow-up, ~0.9 million episodes of hospitalisation have been recorded, involving with >1300 different diseases
Standardized clinical evaluation components measured, if applicable (e.g. height, weight, blood pressure, exercise testing, spirometry, etc.)	Whole cohort: Height (sitting and standing), weight (measured and recalled at age 25), waist & hip circumferences, body fat percentage, blood pressure, heart rate, lung function, Exhaled CO.
Electronic health/medical records or medical administrative data used to collect clinical phenotypes?	Yes
Predominant type of electronic health records (e.g. Epic, Cerner, etc.)	National health insurance record
Other sources of clinical data	Questionnaire + detailed medical notes for several major diseases (e.g. stroke, IHD, cancer, CKD)
Environmental exposure data being obtained? What sort?	Indoor air pollution, ambient air pollution, ambient temperature, built environment
Other data collected	Education, occupation, demographics, healthcare coverage, tea and alcohol consumption, tobacco use, diet, medical history, physical activities, reproductive history, sleep, mental health
Biological specimens collected? What sort?	Whole cohort: Blood (plasma and buffy coat) Subset (25,000) : urine, faecal (starting in 2019)
Is there a central biobank?	Yes
DNA samples prepared (or available to be prepared) from each participant?	Yes (DNA extracted for 330,000 samples and remaining 180,000 will be done over the next few years)
Is genotyping being done on some/all participants?	Yes (done for 102,000 samples)
Is genomic sequencing being done on some/all participants?	2000 samples will be done in 2018
Other molecular analyses performed	Whole cohort: Blood glucose, HBsAg Subset: blood chemistry, metabolomics, infectious antigens, inflammation biomarkers
Did participants provide consent regarding sharing of their data outside the initial study investigators?	Yes - type not specified
How are data or specimens from the cohort made available for research? Any limitations on who can access the data (e.g. by country or sector?)	Currently, all the baseline data are open and first resurvey and cause-specific mortality (7 years) are being released during early 2018. Other data (but not samples) will be released in a staged manner to research communities in China and elsewhere.

What study information or data are returned to or accessible by participants?	No
Follow-up occurring? (years of follow-up). Is recontact possible?	10+; Yes
Notes/Comments	Known previously as the Kadoorie Study of Chronic Disease in China (KSCDC), the China Kadoorie Biobank is a prospective cohort study of over 0.5 million adults from across China; participants were enrolled between 2004 and 2008 and have been followed up ever since for morbidities and mortality. The CKB is an open-ended study with very broad research aims. The main objectives of the study are: 1) To assess reliably the effects of both established and emerging risk factors for many diseases, not only overall but also under various circumstances (e.g. at different ages and at different levels of other risk factors); 2) To determine the complex interplay between genes and environmental factors and between different genes on the risks of common chronic diseases. Purpose: To investigate the main genetic and environmental causes of common chronic diseases in the Chinese population.
Questions Relating to Sharing & Collaboration	
May we make the information you provided about your cohort available on an open website?	Yes
Are you willing to share data from your cohort? If so, would you share:	Yes
a) individual data (redacted to protect confidentiality)?	Yes
b) summary data (counts, distributions)?	Yes
c) metadata (descriptive information on data collection methods)?	Yes
d) case report forms and other data collection materials?	Yes
What do you see as the values of sharing?	Increase the study power; promote scientific collaboration; help to generate new findings; facilitate transethnic comparative studies
What challenges do you anticipate with sharing?	Duplication; inadequate use of the data; poor quality research; lack of support for managing the process
What specific legal or regulatory barriers in your country or cohort (aside from ensuring confidentiality and appropriate consent) would prevent you from sharing data with other cohort of cohorts investigators, or openly with anyone who requests them?	None.
What aspects of your cohort are intended for translation to clinical care or population health?	All

How might genomic sequencing add to/enhance your study objectives?	Our ultimate goal will be to whole genome sequence all 512,891 participants, which will greatly enhance the resource and help to detect large number of rare mutation unique to Chinese and relevant for a wide range of phenotypic traits and disease outcomes. It will also help to develop and optimise regional-specific imputation panel for Chinese.
Might you be willing to contribute funding or other resources to support international collaboration?	No

China PEACE Million Persons Project

Questions Relating to Cohort	
Name of study	China PEACE (Patient-centered Evaluative Assessment of Cardiac Events) Million Persons Project
Principal Investigator/lead	Prof. Linxin Jiang
Contact email	lixin.jiang@fwoxford.org
PubMed ID (or other information) for a protocol/marker paper on this study	Lu J, et al. BMJ Open 2016;6:e010200. doi:10.1136/bmjopen-2015-010200
Study website	None
Purpose or major Objectives of study	This public health effort can serve as a powerful research-grade database for future precision medicine investigations into the biological, environmental, behavioral and other contextual factors associated with CVD in the Chinese population.
Disease areas of focus	Major chronic diseases: cardiovascular disease, stroke, COPD and cancer.
Is your cohort selected for a specific disease (cancer, diabetes) or unselected for disease?	Unselected
Current size of population (and target number of participants)	2,000,000 (target: 4,000,000)
Participating countries	China
Period of enrollment (and is enrollment on-going?)	2014-on-going
Demographic characteristics of participants (age range, proportion male/female, national origin, race)	35-75 y/o, Chinese community-dwelling residents
Major diseases or phenotypes collected to date.	Cardiovascular diseases, stroke, COPD and cancer.
Standardized clinical evaluation components measured, if applicable (e.g. height, weight, blood pressure, exercise testing, spirometry, etc.)	Blood pressure, height, weight and blood lipids in all enrolled participants. From 2015, waist circumference, blood glucose and peak expiratory flow are also measured in all enrolled participants.

Electronic health/medical records or medical administrative data used to collect clinical phenotypes?	Yes (hospitals)
Predominant type of electronic health records (e.g. Epic, Cerner, etc.)	
Other sources of clinical data	Questionnaire on cardiovascular-related health status. For individuals at high risk of CVD, additional ECG, cardiac and carotid ultrasound scan, blood and urine analysis, and a questionnaire on lifestyle and medical history are performed.
Environmental exposure data being obtained? What sort?	Air pollution data.
Other data collected	Physical activity and diet.
Biological specimens collected? What sort?	Yes. Blood and urine.
Is there a central biobank?	Yes
DNA samples prepared (or available to be prepared) from each participant?	DNA samples from more than 40,000 participants were prepared, and DNA abstraction is on-going.
Is genotyping being done on some/all participants?	No
Is genomic sequencing being done on some/all participants?	Yes, 3000 participants (whole exome sequencing)
Other molecular analyses performed	No
Did participants provide consent regarding sharing of their data outside the initial study investigators?	No
How are data or specimens from the cohort made available for research? Any limitations on who can access the data (e.g. by country or sector?)	Approved by the research guidance committee.
What study information or data are returned to or accessible by participants?	All results of physical measurements and ultrasound imaging in the baseline and follow-up assessment.
Follow-up occurring? (years of follow-up). Is recontact possible?	2-3 years. Yes
Notes/Comments	China PEACE (Patient-centered Evaluative Assessment of Cardiac Events) Million Persons Project (China PEACE MPP), funded by the Chinese government, has been being conducted since 2014 in 31 provinces. It is designed to screen about 4 million community-dwelling residents aged 35-75 years with measurements of blood pressure, height and weight, a lipid blood test, and a questionnaire on cardiovascular- related health status. Collection of blood and urine samples is used to establish a biobank.
Questions Relating to Sharing & Collaboration	

May we make the information you provided about your cohort available on an open website?	Yes
Are you willing to share data from your cohort? If so, would you share:	Yes Comment: we should follow the corresponding law and regulations about human genetic resources and data sharing released by the Chinese government.
a) individual data (redacted to protect confidentiality)?	Yes Comment: we should follow the corresponding law and regulations about human genetic resources and data sharing released by the Chinese government.
b) summary data (counts, distributions)?	Yes
c) metadata (descriptive information on data collection methods)?	Yes
d) case report forms and other data collection materials?	Yes
What do you see as the values of sharing?	Increasing use of the data and promoting output of health research
What challenges do you anticipate with sharing?	None
What specific legal or regulatory barriers in your country or cohort (aside from ensuring confidentiality and appropriate consent) would prevent you from sharing data with other cohort of cohorts investigators, or openly with anyone who requests them?	Regulations of human genetic resources (We need to get the approval for data sharing of human genetic resources and related phenotype data from government).
What aspects of your cohort are intended for translation to clinical care or population health?	All
How might genomic sequencing add to/enhance your study objectives?	The addition of genomic sequencing to the currently available phenotype data would allow to detect the effects of rare variations on a variety of health conditions.
Might you be willing to contribute funding or other resources to support international collaboration?	No

Clalit Israeli Genome Project

Questions Relating to Cohort	
Name of study	Clalit Israeli Genome Project
Principal Investigator/lead	Dr. Gad Rennert
Contact email	rennert@technion.ac.il

PubMed ID (or other information) for a protocol/marker paper on this study	
Study website	
Purpose or major Objectives of study	Demonstrate benefits from introducing genomic medicine into usual practice in primary care units
Disease areas of focus	All diseases
Is your cohort selected for a specific disease (cancer, diabetes) or unselected for disease?	Unselected
Current size of population (and target number of participants)	140,000 (100,000 expected to agree to consent)
Participating countries	Israel
Period of enrollment (and is enrollment on-going?)	5 years from start of sample collection (hopefully to start in 6 months). Study has other population and clinic education phases which have already started
Demographic characteristics of participants (age range, proportion male/female, national origin, race)	Age 18+, both sexes, all ethnic groups (Jews/Arabs)
Major diseases or phenotypes collected to date.	At least 130 different diseases known in target population in different volumes
Standardized clinical evaluation components measured, if applicable (e.g. height, weight, blood pressure, exercise testing, spirometry, etc.)	Clinical records of participants are available, including clinical diagnoses, data on medication use, blood test results and more
Electronic health/medical records or medical administrative data used to collect clinical phenotypes?	Clalit EMRs are available to the study
Predominant type of electronic health records (e.g. Epic, Cerner, etc.)	Local make
Other sources of clinical data	No
Environmental exposure data being obtained? What sort?	None
Other data collected	
Biological specimens collected? What sort?	Yes (expected for 100,000 people). Collecting bloods (for WGA, when relevant), buccal swabs(for WEA), feces for microbiome testing
Is there a central biobank?	Yes, in Haifa, Israel
DNA samples prepared (or available to be prepared) from each participant?	Pending
Is genotyping being done on some/all participants?	Planned for all
Is genomic sequencing being done on some/all participants?	Planned for all

Other molecular analyses performed	Possibly 850K array
Did participants provide consent regarding sharing of their data outside the initial study investigators?	Consent form permits other studies if approved by local IRB
How are data or specimens from the cohort made available for research? Any limitations on who can access the data (e.g. by country or sector?)	Not discussed yet
What study information or data are returned to or accessible by participants?	Participants get their FastQs on media. Clinical results are returned via primary care physician with full genetic support
Follow-up occurring? (years of follow-up). Is recontact possible?	Yes
Notes/Comments	
Questions Relating to Sharing & Collaboration	
May we make the information you provided about your cohort available on an open website?	
Are you willing to share data from your cohort? If so, would you share:	
a) individual data (redacted to protect confidentiality)?	Unknown yet
b) summary data (counts, distributions)?	Yes
c) metadata (descriptive information on data collection methods)?	Yes
d) case report forms and other data collection materials?	Yes
What do you see as the values of sharing?	Effect augmentation
What challenges do you anticipate with sharing?	Data security
What specific legal or regulatory barriers in your country or cohort (aside from ensuring confidentiality and appropriate consent) would prevent you from sharing data with other cohort of cohorts investigators, or openly with anyone who requests them?	Very regulatory. Beyond local IRB committee also a national committee which operates under the Israeli genetics law and which limits its freedom of allowing scientific freedom
What aspects of your cohort are intended for translation to clinical care or population health?	All of the components: 1. population education and study of barriers, 2. Clinical teams education, 3. Practicing return of actionable results, 4. Measuring clinics for clinical and financial outcomes in a randomized controlled study model
How might genomic sequencing add to/enhance your study objectives?	Will not only build a genomic atlas of the ethnicities in Israel and other Ashkenazi or Arab populations, but will measure the value of

	incorporating genomic medicine into usual care decisions.
Might you be willing to contribute funding or other resources to support international collaboration?	Still searching for money for genetic analysis of cohort.

Cohort of Norway (CONOR)

[Note: information for this cohort was drawn from publicly available sources and has not been verified]

Questions Relating to Cohort	
Name of study	Cohort of Norway (CONOR)
Principal Investigator/lead	Dr. Camilla Stoltenberg
Contact email	Camilla.Stoltenberg@fhi.no
PubMed ID (or other information) for a protocol/marker paper on this study	
Study website	https://www.fhi.no/en/studies/conor/
Purpose or major Objectives of study	To investigate the causes of disease by developing a unique database with health data and biological samples of about 200,000 individuals.
Disease areas of focus	Mixed
Is your cohort selected for a specific disease (cancer, diabetes) or unselected for disease?	
Current size of population (and target number of participants)	200,000
Participating countries	Norway
Period of enrollment (and is enrollment on-going?)	1994-2008
Demographic characteristics of participants (age range, proportion male/female, national origin, race)	Varies by cohort
Major diseases or phenotypes collected to date.	Broad
Standardized clinical evaluation components measured, if applicable (e.g. height, weight, blood pressure, exercise testing, spirometry, etc.)	Blood pressure, height, weight, and waist and hip circumference are measured
Electronic health/medical records or medical administrative data used to collect clinical phenotypes?	No
Predominant type of electronic health records (e.g. Epic, Cerner, etc.)	

Other sources of clinical data	Questionnaire
Environmental exposure data being obtained? What sort?	
Other data collected	Yes (not specified)
Biological specimens collected? What sort?	Blood
Is there a central biobank?	
DNA samples prepared (or available to be prepared) from each participant?	Yes
Is genotyping being done on some/all participants?	No
Is genomic sequencing being done on some/all participants?	
Other molecular analyses performed	Lipid profile, blood sugar
Did participants provide consent regarding sharing of their data outside the initial study investigators?	Yes - type not specified
How are data or specimens from the cohort made available for research? Any limitations on who can access the data (e.g. by country or sector?)	Project approved by data access committee
What study information or data are returned to or accessible by participants?	
Follow-up occurring? (years of follow-up). Is recontact possible?	
Notes/Comments	To investigate the causes of disease by developing a unique database with health data and biological samples of about 200,000 individuals.

Constances

Questions Relating to Cohort	
Name of study	Constances
Principal Investigator/lead	Dr. Marie Zins
Contact email	marie.zins@inserm.fr
PubMed ID (or other information) for a protocol/marker paper on this study	Zins et al. The French CONSTANCES population-based cohort: design, inclusion and follow-up. Eur J Epidemiology. 2015, 30:1317-1328. DOI 10.1007/s10654-015-0096-4.
Study website	http://www.constances.fr/index_EN.php

Purpose or major Objectives of study	Open general-purpose research infrastructure to study the causes of a wide range of diseases
Disease areas of focus	Mixed. Focus on chronic diseases and aging, women health, social, environmental and occupational epidemiology
Is your cohort selected for a specific disease (cancer, diabetes) or unselected for disease?	Yes; subjects selected only on age
Current size of population (and target number of participants)	162,000 (target: 200,000)
Participating countries	France
Period of enrollment (and is enrollment on-going?)	Mid-2012-mid 2019 (recruitment still ongoing)
Demographic characteristics of participants (age range, proportion male/female, national origin, race)	French residents Age % 18-29 11.3 30-39 17.0 40-49 22.1 50-59 23.7 60+ 25.8 Gender Men 46.1 Women 53.9 Education No diploma or lower than high school 27.4 High school 16.6 College 23.5 University 30.5 Missing 2.0 Marital status Single 23.7 Married, civil partnerships 60.1 Divorced, separated 10.9 Widower 2.4 French laws forbid to collect data on race. Geographic origin: Metropolitan France 87,6 % French overseas territories 0,9 % Europe 3,8 % North Africa 3,0 % Sub-Saharan Africa 1,2 % Asia 0,8 % Other 1,0 %
Major diseases or phenotypes collected to date.	Broad - includes cancer, Alzheimer's, mental health, cardiovascular diseases, musculoskeletal, respiratory diseases, cognitive function, specific women diseases
Standardized clinical evaluation components measured, if applicable (e.g. height, weight, blood pressure, exercise testing, spirometry, etc.)	Anthropometric, blood pressure, ECG, vision, hearing, spirometry and lung function

Electronic health/medical records or medical administrative data used to collect clinical phenotypes?	Yes (social security, hospital and primary care)
Predominant type of electronic health records (e.g. Epic, Cerner, etc.)	
Other sources of clinical data	For some diseases, medical record and adjudication committees; causes of death
Environmental exposure data being obtained? What sort?	Residential history allowing for residential air pollution assessment, area deprivation index, green space Focus on occupational exposures through questionnaires and linkage for various job-exposure matrices (chemicals, biomechanical factors…)
Other data collected	Health scales (CES-D, SF-12, COPD, asthma and rhinitis, oral health, skin diseases, transit and urinary incontinence, sleep, physical activity …), diet (FFQ), alcohol (AUDIT), smoking, E-cigarette, cannabis, limitations, contraception, menopause, family life, education, income, sexual life, life events, sick leave, handicap, healthcare utilization and services provided
Biological specimens collected? What sort?	Blood, serum, plasma, urine will be collected on a sub-sample of the cohort (n=85,000) starting in 2019
Is there a central biobank?	Yes (under construction: will start in 2019)
DNA samples prepared (or available to be prepared) from each participant?	Will available to be prepared when the biobank is ready
Is genotyping being done on some/all participants?	GWAS planned for a subsample
Is genomic sequencing being done on some/all participants?	
Other molecular analyses performed	Planned for a subsample
Did participants provide consent regarding sharing of their data outside the initial study investigators?	Yes; broad (for any type of health-related research)
How are data or specimens from the cohort made available for research? Any limitations on who can access the data (e.g. by country or sector?)	Project approved by the external scientific committee; rules defined in a Charter No limitations.
What study information or data are returned to or accessible by participants?	Individual data: results of the clinical and biological of the baseline assessment returned to participants; no feedback of any subsequent results from analyses of samples or data
Follow-up occurring? (years of follow-up). Is recontact possible?	Annual questionnaire, linkage to the national health and economic administrative databases, health examination every 4-5 years Follow-up started in 2013 (one year after the beginning of cohort inception). In addition to the regular follow-up, recontact is possible for collection of additional data.
Notes/Comments	The CONSTANCES general-purpose cohort with a focus on chronic diseases and aging and on occupational and social factors intended to serve as a research infrastructure accessible to the research community. CONSTANCES is designed as a randomly selected

	representative sample of French adults aged 18-69 years at inception. To take into account non-participation at inclusion and attrition throughout the longitudinal follow-up, a cohort of non-participants was set up and will be followed through the same national databases as participants.
Questions Relating to Sharing & Collaboration	
May we make the information you provided about your cohort available on an open website?	Yes
Are you willing to share data from your cohort? If so, would you share:	Yes for any kind of data; for individual data, confidentiality must be kept according to French regulations. Constances is already open to researchers who had an approval from our Scientific Committee and cleared all legal aspects.
a) individual data (redacted to protect confidentiality)?	
b) summary data (counts, distributions)?	
c) metadata (descriptive information on data collection methods)?	
d) case report forms and other data collection materials?	
What do you see as the values of sharing?	Increasing power of analyses, allowing for international comparisons, fostering scientific collaborations.
What challenges do you anticipate with sharing?	Defining common scientific objectives, data harmonization, costs
What specific legal or regulatory barriers in your country or cohort (aside from ensuring confidentiality and appropriate consent) would prevent you from sharing data with other cohort of cohorts investigators, or openly with anyone who requests them?	No barrier (except confidentiality aspects). The main difficulty is to cover the costs of data sharing which could involve an important workload
What aspects of your cohort are intended for translation to clinical care or population health?	All
How might genomic sequencing add to/enhance your study objectives?	Detection of large effects of rare mutations on a variety of health conditions; description of the genetic variability in the French population
Might you be willing to contribute funding or other resources to support international collaboration?	Yes, provided that specific funding is made available.

Danish National Biobank

[Note: information for this cohort was drawn from publicly available sources and has not been verified]

Questions Relating to Cohort	
Name of study	Danish National Biobank

Principal Investigator/lead	Dr. Mads Melbye
Contact email	Mads: MME@ssi.dk General contact: mail@nationalbiobank.dk
PubMed ID (or other information) for a protocol/marker paper on this study	
Study website	http://www.biobankdenmark.dk/
Purpose or major Objectives of study	To give scientists from Denmark and abroad overview and access to more than 16 million biological samples in both existing and future collections.
Disease areas of focus	Mixed
Is your cohort selected for a specific disease (cancer, diabetes) or unselected for disease?	
Current size of population (and target number of participants)	54,000,000
Participating countries	Denmark
Period of enrollment (and is enrollment on-going?)	1976-ongoing
Demographic characteristics of participants (age range, proportion male/female, national origin, race)	M/F, all ages, Danish population
Major diseases or phenotypes collected to date.	None
Standardized clinical evaluation components measured, if applicable (e.g. height, weight, blood pressure, exercise testing, spirometry, etc.)	No
Electronic health/medical records or medical administrative data used to collect clinical phenotypes?	No
Predominant type of electronic health records (e.g. Epic, Cerner, etc.)	
Other sources of clinical data	Administrative data
Environmental exposure data being obtained? What sort?	
Other data collected	
Biological specimens collected? What sort?	Blood
Is there a central biobank?	Yes
DNA samples prepared (or available to be prepared) from each participant?	Yes

Is genotyping being done on some/all participants?	
Is genomic sequencing being done on some/all participants?	
Other molecular analyses performed	
Did participants provide consent regarding sharing of their data outside the initial study investigators?	Varies by cohort
How are data or specimens from the cohort made available for research? Any limitations on who can access the data (e.g. by country or sector?)	Project approved by data access committee
What study information or data are returned to or accessible by participants?	No
Follow-up occurring? (years of follow-up). Is recontact possible?	
Notes/Comments	To give scientists from Denmark and abroad overview and access to more than 16 million biological samples in both existing and future collections.

Danish National Birth Cohort

Questions Relating to Cohort	
Name of study	Danish National Birth Cohort
Principal Investigator/lead	Mads Melbye (biobank)
Contact email	mme@ssi.dk
PubMed ID (or other information) for a protocol/marker paper on this study	PMID: 11775787
Study website	dnbc.dk
Purpose or major Objectives of study	To study the influence of exposures in the period from conception to early childhood - including fetal growth, cell division, and organ functioning - on long-lasting health and disease susceptibility
Disease areas of focus	All that may be influenced by in utero life exposures
Is your cohort selected for a specific disease (cancer, diabetes) or unselected for disease?	Unselected for disease

Current size of population (and target number of participants)	101,042 pregnant women recruited in first trimester at first antenatal visit at the GP's and their 96,986 children resulting from the pregnancies
Participating countries	Denmark
Period of enrollment (and is enrollment on-going?)	1998-2002
Demographic characteristics of participants (age range, proportion male/female, national origin, race)	Pregnant women in Denmark that reads and speak Danish
Major diseases or phenotypes collected to date.	Many. Please be referred to home page
Standardized clinical evaluation components measured, if applicable (e.g. height, weight, blood pressure, exercise testing, spirometry, etc.)	Many. Please be referred to home page where all questionnaires can be found translated into english
Electronic health/medical records or medical administrative data used to collect clinical phenotypes?	National databases, including all hospital diagnosis, pathology records, prescribed medication, etc.
Predominant type of electronic health records (e.g. Epic, Cerner, etc.)	National health records establihed in Denmark. Epic available since 2017 in parts of the country
Other sources of clinical data	
Environmental exposure data being obtained? What sort?	Many, please be referred to questionnaiores. A separate food questionnaire exists for all enrolled
Other data collected	
Biological specimens collected? What sort?	At enrollment (week 10-12) again around week 24, and cord blood at time of birth, Gutheri cards (day 2-5 after birth) also exist
Is there a central biobank?	Yes, all sample are placed in the Dnaish National Biuobank (danishnationalbiobank.com)
DNA samples prepared (or available to be prepared) from each participant?	From all mothers and soon also all children
Is genotyping being done on some/all participants?	Approx. 10.000
Is genomic sequencing being done on some/all participants?	none

Other molecular analyses performed	Performed in nested studies, please be referred to publication lsit from our home page.
Did participants provide consent regarding sharing of their data outside the initial study investigators?	There is a broad informed consent
How are data or specimens from the cohort made available for research? Any limitations on who can access the data (e.g. by country or sector?)	Application form is found on our web site. No limitation as long as the project protocol is approved by the comitee and the data protection agency in Denmark
What study information or data are returned to or accessible by participants?	
Follow-up occurring? (years of follow-up). Is recontact possible?	All participants and their children are followed by linking to national registries. In addition new data collections have taken place at age 7, 11, 18 years of the children
Notes/Comments	

East London Genes and Health

Questions Relating to Cohort	
Name of study	East London Genes and Health
Principal Investigator/lead	Dr. David van Heel
Contact email	D.VANHEEL@QMUL.AC.UK
PubMed ID (or other information) for a protocol/marker paper on this study	
Study website	http://www.genesandhealth.org/
Purpose or major Objectives of study	ELGH as a long-term resource for population genomic medicine research in British South Asians, delivering high quality basic science, translational, clinical and population health research. Areas of health importance identified by the local East London community are prioritised, including type 2 diabetes, cardiovascular disease and mental health. In addition to a health-deprived community, ELGH has a unique combination of features: high rates of consanguinity such that rare genetic variants occur as homozygotes, including homozygous predicted loss of function genotypes (knockouts); local excellence in e--health record access & analysis (enhanced by a new Health Data Research UK - London site); recall for further research by genotype and/or phenotype to excellent London clinical research facilities; and understudied (yet highly engaged) minority ethnic groups.
Disease areas of focus	

Is your cohort selected for a specific disease (cancer, diabetes) or unselected for disease?	Unselected
Current size of population (and target number of participants)	27,806 (target: 100,000)
Participating countries	UK
Period of enrollment (and is enrollment on-going?)	on-going to 2023
Demographic characteristics of participants (age range, proportion male/female, national origin, race)	British Bangladeshi & British Pakistani (target: equal recruitment). Target: equal gender recruitment. Current age range at 23 Jan 2018: median year of birth 1977, IWR 1967 - 1985, range 1915 - 2001
Major diseases or phenotypes collected to date.	All data in primary care (UK NHS GP), secondary care (East London Hospitals), and national NHS datasets (e.g. Hospital Episode Statistics).
Standardized clinical evaluation components measured, if applicable (e.g. height, weight, blood pressure, exercise testing, spirometry, etc.)	None by the research study. However 96.5% have clinician measured BMI available in e-health records, 72.7% have lab measured blood lipid data, and lots of other data.
Electronic health/medical records or medical administrative data used to collect clinical phenotypes?	YES. Have already linked to 16,955 primary care NHS GP e-health records and 20,072 Barts Health secondary care NHS e-health records.
Predominant type of electronic health records (e.g. Epic, Cerner, etc.)	Primary care: varied, mainly EMIS-Web and SystmOne. Secondary care: varied, mainly Cerner.
Other sources of clinical data	None at present. Within study recall for cardiometabolic health assessment (anthropometrics, retinal photo, iDXA, blood, Fibroscan) is planned.
Environmental exposure data being obtained? What sort?	None by the research study. Linkage to small area statistics within London and UK is possible.
Other data collected	
Biological specimens collected? What sort?	YES - oragene saliva DNA on all. Within study recall programmes on selected individuals for: human knockouts (n ~5000) - blood and urine - planned; cardiometabolic health assessment - planned - see above; other.
Is there a central biobank?	NOT YET, but would use UK Biocentre if funding can be obtained.
DNA samples prepared (or available to be prepared) from each participant?	YES. All have oragene saliva, currently ~20% samples have been DNA extracted.
Is genotyping being done on some/all participants?	NOT YET, but would be (e.g. Illumina GSA array or Affy UK Biobank chip, or similar) if funding can be obtained.
Is genomic sequencing being done on some/all participants?	Exome sequencing to 20-40X depth on ~20% with self-stated autozygosity. IDEALLY, exome sequencing to 60X depth on all if funding can be obtained.
Other molecular analyses performed	Planned: blood metabolic, lipidomic, transcriptomics on within-study subset of ~2500, if funding can be obtained.

Did participants provide consent regarding sharing of their data outside the initial study investigators?	YES. yes to genotype data being shared widely under standard data access agreement (& deposition at EBI-EGA). limited yes to phenotype including e-health record data - access within data safe haven.
How are data or specimens from the cohort made available for research? Any limitations on who can access the data (e.g. by country or sector?)	Application to the study. We have an application form similar to UK Biobank. No limitations on access by country, academia or industry. Currently access is limited by not having funding or resource to support external applications.
What study information or data are returned to or accessible by participants?	Return to participants only of immediately clinical actionable results (e.g. anaemic or leukemic blood sample, severe hypertension, etc). No return of results available at a later date, including no return of genetic data.
Follow-up occurring? (years of follow-up). Is recontact possible?	Current 20 year follow up. Recontact, and recall by genotype or phenotype is possible. We have performed our first recall by genotype studies successfully.
Notes/Comments	East London Genes & Health is one of the world's largest community-based genetics studies, aiming to improve health among people of Pakistani and Bangladeshi heritage in East London by analysing the genes and health of 100,000 local people.
Questions Relating to Sharing & Collaboration	
May we make the information you provided about your cohort available on an open website?	Yes
Are you willing to share data from your cohort? If so, would you share:	Yes, but individual level phenotype data only within data safe haven.
a) individual data (redacted to protect confidentiality)?	Yes, but individual level phenotype data only within data safe haven.
b) summary data (counts, distributions)?	Yes
c) metadata (descriptive information on data collection methods)?	Yes
d) case report forms and other data collection materials?	All study forms available on study website.
What do you see as the values of sharing?	Very beneficial, especially genetic data. Phenotype data, especially NHS e-health records, however requires a higher level of security to protect community/participant confidence in the study.
What challenges do you anticipate with sharing?	None, but individual level phenotype data only within data safe haven.
What specific legal or regulatory barriers in your country or cohort (aside from ensuring confidentiality and appropriate consent) would prevent you from sharing data with other cohort of cohorts investigators, or openly with anyone who requests them?	None

What aspects of your cohort are intended for translation to clinical care or population health?	Many aspects - south asians are an understudied population. Within cohort sub-studies are possible aimed at clinical or population health questions/translation.
How might genomic sequencing add to/enhance your study objectives?	High depth exome sequencing (in addition to SNP array genotyping) on all would hugely help our study objectives. It is unclear what added value whole genome sequencing would currently offer (& we have saliva DNA not blood).
Might you be willing to contribute funding or other resources to support international collaboration?	No, as we are currently very limited in funding for even our main study.

Environmental influences on Child Health Outcomes (ECHO)

Questions Relating to Cohort	
Name of study	Environmental influences on Child Health Outcomes (ECHO): ECHO-wide Cohort
Principal Investigator/lead	Dr. Matt Gillman
Contact email	matthew.gillman@nih.gov (assistant Keianna.beckett@nih.gov)
PubMed ID (or other information) for a protocol/marker paper on this study	In press.
Study website	https://www.nih.gov/echo; http://www.echochildren.org/
Purpose or major Objectives of study	To address solution-oriented questions about effects of broad range of early environmental exposures on child health and development
Disease areas of focus	Pre/peri/postnatal, upper & lower airway, obesity, neurodevelopment, positive child health
Is your cohort selected for a specific disease (cancer, diabetes) or unselected for disease?	83 cohorts in a single consortium: Most unselected
Current size of population (and target number of participants)	Goal 100,000: 50,000 children + 50,000 parents
Participating countries	USA
Period of enrollment (and is enrollment on-going?)	1980s to present, depending on the cohort. All cohorts existed at time of ECHO funding in 2016. All are following the children, some are still recruiting
Demographic characteristics of participants (age range, proportion male/female, national origin, race)	Age-strata (years). 35–44: 3340 (22.1%);45–54: 5939 (39.3%);55–64: 4234 (28.0%);65–74: 1592 (10.6%).
Major diseases or phenotypes collected to date.	Pre/peri/postnatal, upper & lower airway, obesity, neurodevelopment, positive child health

Standardized clinical evaluation components measured, if applicable (e.g. height, weight, blood pressure, exercise testing, spirometry, etc.)	
Electronic health/medical records or medical administrative data used to collect clinical phenotypes?	Most EMR data are from pregnancy or neonatal period
Predominant type of electronic health records (e.g. Epic, Cerner, etc.)	
Other sources of clinical data	Questionnaire, interview, biospecimen, environmental specimen, geocode
Environmental exposure data being obtained? What sort?	Broad range of exposures: physical/chemical, societal, psychosocial, behavioral, biological
Other data collected	Many covariates
Biological specimens collected? What sort?	Blood, urine, placenta, hair, nails, teeth, nasal swab, buccal swab, saliva, stool; environmental specimens, e.g., dust. Not all specimens in all cohorts
Is there a central biobank?	Yes, within ~1 year
DNA samples prepared (or available to be prepared) from each participant?	Yes, in the future
Is genotyping being done on some/all participants?	Yes, to be done on all children and as many parents as possible
Is genomic sequencing being done on some/all participants?	
Other molecular analyses performed	Epigenetics, transcriptomics, metabolomics, “microbiomics” (not all in all cohorts)
Did participants provide consent regarding sharing of their data outside the initial study investigators?	Yes for some now. Yes in the future for all.
How are data or specimens from the cohort made available for research? Any limitations on who can access the data (e.g. by country or sector?)	Through specified publications and data/biospecimen sharing policies and processes. Data will be available starting ~2019. Ongoing discussions re: any access limitations
What study information or data are returned to or accessible by participants?	Not yet specified.
Follow-up occurring? (years of follow-up). Is recontact possible?	Yes, followup is occurring, within all 6 life course stages depending on the cohort. Funding is through 2022 to continue followup. Cohorts are re-contacting participants as needed to address ECHO science.
Notes/Comments	The Environmental influences on Child Health Outcomes (ECHO) Program, a seven-year research initiative, aims to determine what factors give children the highest probability of achieving the best health outcomes over their lifetimes. To find answers to vital unanswered questions, ECHO taps into existing cohorts totaling about 50,000 children from racially, socioeconomically and geographically diverse backgrounds. ECHO research will focus primarily on pre-, peri-, and

	post-natal outcomes, upper and lower airway health and development, obesity, and development of cognition, emotion and behavior, as well as positive health. The ECHO-wide Cohort knits together existing harmonizable, and new standardized/harmonized, data from 83 cohorts into a single data platform. ECHO cohorts are supported by a Coordinating Center, Data Analysis Center, Children's Health Exposure Analysis Resource, Person-reported Outcomes Core, and (in the next 1-2 years) Genetics Core. ECHO also includes the IDeA States Pediatric Clinical Trials Network for supporting intervention studies among rural and underserved children. ECHO is governed by Executive and Steering Committees representing all components, with oversight by US National Institutes of Health (NIH), Office of the Director. To support ECHO, in 2016 NIH awarded cooperative agreement grants to 62 institutions, representing 110 principal investigators and almost 1300 key personnel.
Questions Relating to Sharing & Collaboration	
May we make the information you provided about your cohort available on an open website?	Yes
Are you willing to share data from your cohort? If so, would you share:	There will be two forms of data sharing in ECHO: 1.). The ECHO-wide Cohort data platform within a restrictive enclave, with individual data, to which cohorts regularly contribute. High security and publication controls, available first to ECHO investigators and 6 months later to non-ECHO investigators; 2.) In the future, a "public use" data set with anonymized data. Fewer security or publication controls. Metadata available broadly—TBD process. ECHO Data Analysis Center holds ECHO-wide Cohort data and metadata.
a) individual data (redacted to protect confidentiality)?	
b) summary data (counts, distributions)?	
c) metadata (descriptive information on data collection methods)?	
d) case report forms and other data collection materials?	
What do you see as the values of sharing?	Values: Large sample size for 1.) heterogeneity 2.) generalizability, 3.) rare exposures or outcomes. Diversity. Testing big data approaches. Team science. Opportunities for junior scientists.
What challenges do you anticipate with sharing?	Challenges: Investigator hegemony. Data standardization/harmonization. Blithely applying big data methods to etiologic questions without conceptual models.
What specific legal or regulatory barriers in your country or cohort (aside from ensuring confidentiality and appropriate consent) would prevent you from sharing data with other cohort of cohorts investigators, or openly with anyone who requests them?	We are not sure yet.

What aspects of your cohort are intended for translation to clinical care or population health?	ECHO's watchword is solution-oriented research, i.e, addressing questions with impact on programs, policies, practices. Perspective is mostly primordial prevention, less on risk stratification. More on etiology, less on prediction.
How might genomic sequencing add to/enhance your study objectives?	It might uncover otherwise vague pathways from exposure to outcome. It might help interpret gene X epigene interactions. It might help with mendelian randomization.
Might you be willing to contribute funding or other resources to support international collaboration?	Alas, ECHO is unable to contribute tangible resources.

Estonian Genome Project

Questions Relating to Cohort	
Name of study	Estonian Genome Project
Principal Investigator/lead	Dr. Andres Metspalu
Contact email	andres.metspalu@ut.ee
PubMed ID (or other information) for a protocol/marker paper on this study	PMID: 24518929
Study website	www.biobank.ee
Purpose or major Objectives of study	1) promote the development of genetic research; 2) collect information on the health of the Estonian population and genetic information concerning the Estonian population; 3) use the results of genetic research to improve public health. 4) collect information on the health of the Estonian population and genetic information concerning the Estonian population; 5) use the results of genetic research to improve public health.
Disease areas of focus	population based
Is your cohort selected for a specific disease (cancer, diabetes) or unselected for disease?	Unselected
Current size of population (and target number of participants)	52 000 currently, but recruitment is ongoing to add 100 000 individuals in 2018
Participating countries	Estonia
Period of enrollment (and is enrollment on-going?)	First phase: 2002-2010 recruited 52 000 subjects Second phase: 2018 - additional 100 000
Demographic characteristics of participants (age range, proportion male/female, national origin, race)	From 18 years of age, 65% female 35% male, 85% estonians
Major diseases or phenotypes collected to date.	we have all medical events from the EMR and other databases incl. prescription drugs

Standardized clinical evaluation components measured, if applicable (e.g. height, weight, blood pressure, exercise testing, spirometry, etc.)	Yes, basic (e.g. EKG or spirometry is not done on all)
Electronic health/medical records or medical administrative data used to collect clinical phenotypes?	Yes, at baseline and longitudinally updated on regular basis
Predominant type of electronic health records (e.g. Epic, Cerner, etc.)	e-Health is using the following: HL7CDA Ver 3.0 This is for the whole country
Other sources of clinical data	hospital databases, central insurance fund, registries etc in total 9 sources
Environmental exposure data being obtained? What sort?	Smoking, alcohol, physical activity, sleep, food frequency q., work, education
Other data collected	
Biological specimens collected? What sort?	whole EDTA blood from which plasma, DNA and WBC are stored in liquid N2
Is there a central biobank?	Yes, we have one central biobank in Estonia
DNA samples prepared (or available to be prepared) from each participant?	DNA samples are available and all are genotyped with Illumina GSA array
Is genotyping being done on some/all participants?	All 52 000 and the next 100 000 will be genotyped also by 31.01.2019.
Is genomic sequencing being done on some/all participants?	yes, 2500 WGS 30X and 2500 WES 80X
Other molecular analyses performed	NMR on 12 000 samples
Did participants provide consent regarding sharing of their data outside the initial study investigators?	Yes, broad consent
How are data or specimens from the cohort made available for research? Any limitations on who can access the data (e.g. by country or sector?)	Project has to get approval from the ethical review committee and government approval to send samples outside of Estonia
What study information or data are returned to or accessible by participants?	We are returning the genetic risk scores for some diseases (e.g. T2D, CAD), incidental findings and likely drug response
Follow-up occurring? (years of follow-up). Is recontact possible?	Recontact is possible and frequently used, follow-up is part of it
Notes/Comments	The Estonian Biobank, housed at the Estonian Genome Center in Tartu, currently holds 52,000 samples collected during the first phase of the Estonian Genome Project, which commenced in 2000. As part of this new phase of the project, organizers will not only collect genetic data on participants, but report back genetic risks via the country's national health information system.
Questions Relating to Sharing & Collaboration	

May we make the information you provided about your cohort available on an open website?	Yes
Are you willing to share data from your cohort? If so, would you share:	Yes
a) individual data (redacted to protect confidentiality)?	Yes
b) summary data (counts, distributions)?	Yes
c) metadata (descriptive information on data collection methods)?	Yes
d) case report forms and other data collection materials?	Yes
What do you see as the values of sharing?	Improving the variant database with the phenotypes
What challenges do you anticipate with sharing?	to keep confidentiality
What specific legal or regulatory barriers in your country or cohort (aside from ensuring confidentiality and appropriate consent) would prevent you from sharing data with other cohort of cohorts investigators, or openly with anyone who requests them?	partners have to have the same ethical and legal standards as in Estonia
What aspects of your cohort are intended for translation to clinical care or population health?	We are piloting the return of the data and precision medicine already in hospital settings
How might genomic sequencing add to/enhance your study objectives?	WGS would be much more informative to generate the GRS and the precision and the scope of the info would be much better. This is our goal - to sequence them all
Might you be willing to contribute funding or other resources to support international collaboration?	No Comment: we do not have funding to share, but we would be happy to help in any other way and in fact, doing it all the time

European Prospective Investigation into Cancer & Nutrition (EPIC)

Questions Relating to Cohort	
Name of study	European Prospective Investigation into Cancer and nutrition (EPIC)
Principal Investigator/lead	Dr. Elio Riboli, Dr. Paul Brennan, and Dr. Marc Gunter (Representative at this meeting: Mattias Johansson)
Contact email	e.riboli@imperial.ac.uk; brennanp@iarc.fr; gunterm@iarc.fr; johanssonm@iarc.fr

PubMed ID (or other information) for a protocol/marker paper on this study	PMID: 12639222
Study website	http://epic.iarc.fr/
Purpose or major Objectives of study	To investigate the relationship between lifestyle, genetics and chronic disease
Disease areas of focus	Mixed (primary cancer and subset CVD and type 2 diabetes)
Is your cohort selected for a specific disease (cancer, diabetes) or unselected for disease?	Population-based with general long-term follow-up for cancer, and follow-up of diabetes and cardiovascular disease until 2008
Current size of population (and target number of participants)	521,000
Participating countries	10 European (UK, Italy, France, Germany, Norway, Netherlands, Denmark, Spain, Greece, Sweden)
Period of enrollment (and is enrollment on-going?)	1992-1999
Demographic characteristics of participants (age range, proportion male/female, national origin, race)	M/F (majority women), aged 35-70 years, race : predominantly white
Major diseases or phenotypes collected to date.	Broad - Cancers, coronary heart disease, stroke, diabetes, neurodegenerative disorders, obesity and weight change
Standardized clinical evaluation components measured, if applicable (e.g. height, weight, blood pressure, exercise testing, spirometry, etc.)	anthropometric (height, weight, waist/hip circumference); blood pressure (subset)
Electronic health/medical records or medical administrative data used to collect clinical phenotypes?	Available in a subset of centers.
Predominant type of electronic health records (e.g. Epic, Cerner, etc.)	
Other sources of clinical data	Depending on country, linkage-based follow-up with registries or active follow-up
Environmental exposure data being obtained? What sort?	Tobacco smoke, air pollution
Other data collected	diet, education SES, occupation, previous illness, tobacco and alcohol use, reproductive history
Biological specimens collected? What sort?	Blood from 385,000 participants
Is there a central biobank?	Yes (except for Swedish & Danish samples)
DNA samples prepared (or available to be prepared) from each participant?	DNA has been extracted, or can be extracted, from all participants who originally provided a blood sample
Is genotyping being done on some/all participants?	Yes, for subsets, generally nested case-control studies

Is genomic sequencing being done on some/all participants?	No
Other molecular analyses performed	Yes, for subsets, generally nested case-control studies
Did participants provide consent regarding sharing of their data outside the initial study investigators?	Varies with cohort (Broad for EPIC-CVD)
How are data or specimens from the cohort made available for research? Any limitations on who can access the data (e.g. by country or sector?)	Individual projects are reviewed and approved by the EPIC steering committee. Access procedures are described on the EPIC website.
What study information or data are returned to or accessible by participants?	None
Follow-up occurring? (years of follow-up). Is recontact possible?	Participants have been followed for incident cancer over 20 years. Re-contact is feasible assuming ethics approval in the individual centers.
Notes/Comments	EPIC is a prospective cohort with study participants enrolled from 23 centers in 10 western European countries. Detailed information on diet, lifestyle characteristics, anthropometric measurements, and medical history was collected at recruitment (1992—1999). The cohort is governed by a steering committee involving investigators from each center. Imperial college (London) and the International Agency for Research on Cancer (IARC, Lyon) act as focal points.
Questions Relating to Sharing & Collaboration	
May we make the information you provided about your cohort available on an open website?	Yes (Upon approval from the steering committee)
Are you willing to share data from your cohort? If so, would you share:	Yes (Upon approval from the steering committee)
a) individual data (redacted to protect confidentiality)?	Yes (Upon approval from the steering committee, ethics approval, and compliance with data protection)
b) summary data (counts, distributions)?	Yes (Upon approval from the steering committee)
c) metadata (descriptive information on data collection methods)?	Yes (Upon approval from the steering committee)
d) case report forms and other data collection materials?	Yes (Upon approval from the steering committee)
What do you see as the values of sharing?	Accelerate scientific discovery and ensure validity of study results.
What challenges do you anticipate with sharing?	Legal and regulatory barriers (see below).
What specific legal or regulatory barriers in your country or cohort (aside from ensuring confidentiality and appropriate consent) would prevent you from sharing data with other cohort of cohorts investigators, or openly with anyone who requests them?	Compliance the EU General Data Protection Regulation (GDPR), as well as ethics standards.

What aspects of your cohort are intended for translation to clinical care or population health?	The EPIC cohort enables a wide range of observational studies intended to improve our understanding of disease etiology and population health.
How might genomic sequencing add to/enhance your study objectives?	Sequencing the entire EPIC cohort with follow-up of over 20 years would enable far-reaching studies on chronic disease that are not otherwise feasible.
Might you be willing to contribute funding or other resources to support international collaboration?	Not clear

Finnish Maternity Cohort Serum Bank

Questions Relating to Cohort	
Name of study	Finnish Maternity Cohort Serum Bank
Principal Investigator/lead	Dr. Surcel Heljä-Marja
Contact email	helja-marja.surcel@oulu.fi
PubMed ID (or other information) for a protocol/marker paper on this study	
Study website	www.esis.fi AND https://www.ppshp.fi/Tutkimus-ja-opetus/Biopankki/Pages/Biobank-Borealis-briefly-in-English.aspx
Purpose or major Objectives of study	As a population based cohort representing nearly all pregnant women since 1983, the FMC is the world's largest continuously gathered serum bank. Combined with an uniform health care system and several high-quality health registers in Finland, use of the FMC enables follow-up and longitudinal research of risk factors and biomarkers associated to a disease of interest in the mother or in the child born after the pregnancy in question.
Disease areas of focus	Mixed
Is your cohort selected for a specific disease (cancer, diabetes) or unselected for disease?	Unselected
Current size of population (and target number of participants)	950,000 females; 2M samples.
Participating countries	Finland
Period of enrollment (and is enrollment on-going?)	1983-2016 systematic enrollment of pregnant women, since 2018 partly ongoing
Demographic characteristics of participants (age range, proportion male/female, national origin, race)	Mean age of all pregnant women is 28-31 years; Median number of pregnancies 2,0 (min 1, max 24);Number of pregnancies: one in 17.1 % ; two in 30,8 %, > 3 in 8,1 %; Race:mainly caucasians, no data available.
Major diseases or phenotypes collected to date.	Disease data is not included in the biobank but the subjects nested in the FMC can be combined to Finnish health registers or Biobanks using a personal identification number of a citizen.

Standardized clinical evaluation components measured, if applicable (e.g. height, weight, blood pressure, exercise testing, spirometry, etc.)	No data included in the biobank. Some data available in the health registers.
Electronic health/medical records or medical administrative data used to collect clinical phenotypes?	No data included in the biobank. Some data available in the health registers.
Predominant type of electronic health records (e.g. Epic, Cerner, etc.)	
Other sources of clinical data	Health registers and hospital archives. Combining FMC with existing Biobanks enables research of genetic data in association with biomarkers of interest.
Environmental exposure data being obtained? What sort?	None. POP expose can be analysed in the samples.
Other data collected	Combining FMC with major research projects in Finland (Northern Finland Birth Cohort 1986 or Finnish Birth Cohort 1987) enables research of combined biomarkers and longitudinal follow up and impact of lifestyle and environmental factors.
Biological specimens collected? What sort?	Serum
Is there a central biobank?	Not yet - under planning
DNA samples prepared (or available to be prepared) from each participant?	Not from all participants, but linking the FMC data with existing biobanks enables search of samples for DNA preparation or of existing DNA material.
Is genotyping being done on some/all participants?	Not done for all participants, but linking the FMC data with existing biobanks enables search of existing and available genomic preparation or DNA material.
Is genomic sequencing being done on some/all participants?	none
Other molecular analyses performed	Yes
Did participants provide consent regarding sharing of their data outside the initial study investigators?	A project approved by Biobank's Scientific Committee of the biobank
How are data or specimens from the cohort made available for research? Any limitations on who can access the data (e.g. by country or sector?)	None.
What study information or data are returned to or accessible by participants?	
Follow-up occurring? (years of follow-up). Is recontact possible?	No possibility to contact participants.
Notes/Comments	The Finnish Maternity Cohort (FMC) serum bank is a nationwide biorepository of prenatal serum samples accumulated since 1983.

Questions Relating to Sharing & Collaboration	
May we make the information you provided about your cohort available on an open website?	Yes
Are you willing to share data from your cohort? If so, would you share:	Yes
a) individual data (redacted to protect confidentiality)?	Yes Comment:Individual data can only be shared following the personal data protection restrictions by Finnish legislation and that of EU (https://gdpr-info.eu/chapter-2/)
b) summary data (counts, distributions)?	Yes
c) metadata (descriptive information on data collection methods)?	Yes
d) case report forms and other data collection materials?	Yes
What do you see as the values of sharing?	Increased awareness and use of the data for an increasingly wide and imaginative set of purposes.
What challenges do you anticipate with sharing?	The personal data protection legislation in Finland and in EU may sometimes be challenging and hold up a well planned project of high scientific quality.
What specific legal or regulatory barriers in your country or cohort (aside from ensuring confidentiality and appropriate consent) would prevent you from sharing data with other cohort of cohorts investigators, or openly with anyone who requests them?	None
What aspects of your cohort are intended for translation to clinical care or population health?	All
How might genomic sequencing add to/enhance your study objectives?	Combined biomarker data and genomic data would enhance understanding of the role and interaction between different factors in terms of health and disease.
Might you be willing to contribute funding or other resources to support international collaboration?	No

Generations Study

Questions Relating to Cohort	
Name of study	Generations Study (GS)
Principal Investigator/lead	Prof. Anthony Swerdlow
Contact email	anthony.swerdlow@icr.ac.uk

PubMed ID (or other information) for a protocol/marker paper on this study	PMID: 21897394 (Swerdlow et al., Br J Ca 2011)
Study website	http://www.breakthroughgenerations.org.uk/home
Purpose or major Objectives of study	A cohort study focused particularly on finding the causes of breast cancer, with comprehensive data, but also investigating causes of other cancers and causes of death.
Disease areas of focus	Site-specific cancers; cause-specific mortality.
Is your cohort selected for a specific disease (cancer, diabetes) or unselected for disease?	Unselected
Current size of population (and target number of participants)	113,000 (target: >100,000)
Participating countries	UK (England, Scotland, Wales, Northern Ireland, Isle of Man, Channel Islands)
Period of enrollment (and is enrollment on-going?)	2003-2015
Demographic characteristics of participants (age range, proportion male/female, national origin, race)	100% F; ages 16-102 years at entry: 98.8% white, 0.2% black, 0.4% S. Asian, 0.6% other and mixed.
Major diseases or phenotypes collected to date.	Site-specific cancer incidence, cause-specific mortality, diabetes, reproductive-related diseases, anthropometrics.
Standardized clinical evaluation components measured, if applicable (e.g. height, weight, blood pressure, exercise testing, spirometry, etc.)	Accelerometry, anthropometrics, mammographic density, on subsets.
Electronic health/medical records or medical administrative data used to collect clinical phenotypes?	Yes – cancer and mortality data.
Predominant type of electronic health records (e.g. Epic, Cerner, etc.)	UK systems.
Other sources of clinical data	Questionnaires, data from physicians, medical records, record linkage.
Environmental exposure data being obtained? What sort?	No.
Other data collected	Large range of questionnaire data including on behaviours, exposures, phenotype, past health conditions and operations, occupation, exercise, diet, tobacco and alcohol use, menstrual and reproductive, HRT, OCs.
Biological specimens collected? What sort?	Blood, urine, tumour.
Is there a central biobank?	Yes.
DNA samples prepared (or available to be prepared) from each participant?	Yes, available or prepared for 92%.

Is genotyping being done on some/all participants?	Yes, on a nested case-control basis currently.
Is genomic sequencing being done on some/all participants?	Not yet.
Other molecular analyses performed	No.
Did participants provide consent regarding sharing of their data outside the initial study investigators?	Yes – when anonymised.
How are data or specimens from the cohort made available for research? Any limitations on who can access the data (e.g. by country or sector?)	On application. No country limits.
What study information or data are returned to or accessible by participants?	Overall results.
Follow-up occurring? (years of follow-up). Is recontact possible?	Yes, active re-contact every ~3 years.
Notes/Comments	A long-term comprehensive cohort study focused on finding the causes of breast cancer, but also investigating the causes of other cancers and fatal conditions. The study involves 113,000 women from the general population from whom questionnaire information, and for 92% blood samples, have been gathered at recruitment. Highly complete active follow-up.
Questions Relating to Sharing & Collaboration	
May we make the information you provided about your cohort available on an open website?	Comment: Yes, the information above, not the information below.
Are you willing to share data from your cohort? If so, would you share:	Yes Comment: in principle.
a) individual data (redacted to protect confidentiality)?	Yes Comment: in principle.
b) summary data (counts, distributions)?	Yes Comment: in principle.
c) metadata (descriptive information on data collection methods)?	Yes Comment: in principle.
d) case report forms and other data collection materials?	
What do you see as the values of sharing?	More powerful, better research materials and scientific input, ideas.
What challenges do you anticipate with sharing?	Harmonisation, transfer agreements, resources, confidentiality (e.g. if genotyping data shared).

What specific legal or regulatory barriers in your country or cohort (aside from ensuring confidentiality and appropriate consent) would prevent you from sharing data with other cohort of cohorts investigators, or openly with anyone who requests them?	The key constraints are need to follow consent, ethics, and priorities, rather than legal barriers.
What aspects of your cohort are intended for translation to clinical care or population health?	Risk prediction, screening.
How might genomic sequencing add to/enhance your study objectives?	Genetics of cancer.
Might you be willing to contribute funding or other resources to support international collaboration?	We don't have funds that we could send, if that is what you mean. We could potentially contribute data, time, analyses, if the collaboration were in line with our funded objectives.

Genomics England

Questions Relating to Cohort	
Name of study	Genomics England / 100,000 Genomes Project
Principal Investigator/lead	Dr. Mark Caulfield
Contact email	m.j.caulfield@qmul.ac.uk
PubMed ID (or other information) for a protocol/marker paper on this study	https://doi.org/10.6084/m9.figshare.4530893.v4
Study website	https://www.genomicsengland.co.uk/
Purpose or major Objectives of study	Four main aims: 1. to create an ethical and transparent programme based on consent 2. to bring benefit to NHS patients and set up a genomic medicine service for the NHS; 3. to enable new scientific discovery and medical insights; 4. to kick start the development of a UK genomics industry.
Disease areas of focus	Rare Disease, Cancer and Infection
Is your cohort selected for a specific disease (cancer, diabetes) or unselected for disease?	Unselected
Current size of population (and target number of participants)	75,000 participants but 100,000 whole genomes
Participating countries	England
Period of enrollment (and is enrollment on-going?)	December 2013 to December 2018

Demographic characteristics of participants (age range, proportion male/female, national origin, race)	This is complex as this is both children and older people. Last check median age is 34 years of age Covers the demography of England with excess representation of ethnic groups in which consanguinity is more likely.
Major diseases or phenotypes collected to date.	300 rare disease categories reaching 1200 disorders 15 cancer types
Standardized clinical evaluation components measured, if applicable (e.g. height, weight, blood pressure, exercise testing, spirometry, etc.)	Standardised primary data collection models for rare disease and cancer using standardised data commons from Human Phenotyping Ontology or SNOMED CT
Electronic health/medical records or medical administrative data used to collect clinical phenotypes?	Electronic health data from primary care, hospital episodes, cancer registries, mental health and disability NHS Digital • Hospital Episodes • Office of National Stats death details • Diagnostic Imaging • Patient recorded outcomes • Mental health & intellectual disability Public Health England • Cancer registry & datasets • Other disease registries GP data • Prescribing/dispensing • Coding/quantitative data Reports/letters • Notes (free text)
Predominant type of electronic health records (e.g. Epic, Cerner, etc.)	
Other sources of clinical data	Primary enrolment data from the NHS (see above)
Environmental exposure data being obtained? What sort?	Not at present
Other data collected	
Biological specimens collected? What sort?	DNA, RNA plasma on probands in rare disease Germline DNA, Somatic DNA and samples for RNA and cfTDNA
Is there a central biobank?	Yes
DNA samples prepared (or available to be prepared) from each participant?	Yes
Is genotyping being done on some/all participants?	Whole genome sequence but also a GWAS array
Is genomic sequencing being done on some/all participants?	Whole genome sequence but also a GWAS array
Other molecular analyses performed	
Did participants provide consent regarding sharing of their data outside the initial study investigators?	Yes

How are data or specimens from the cohort made available for research? Any limitations on who can access the data (e.g. by country or sector?)	Data are accessible by the Genomics England Clinical Interpretation Partnership where 2612 scientists from 352 institutions in 24 countries have volunteered to undertake research to drive up clinical value for patients. They are self-organised in 42 domains facing rare disease, cancer and cross-cutting themes
What study information or data are returned to or accessible by participants?	In rare disease results about their main condition All participants can opt to receive additional findings about genomic findings against a limited list of diseases. Parents can also opt to receive a limited list of findings which if they came together in a reproductive event could led to a child with rare disease In cancer they receive potentially actionable somatic and germline changes
Follow-up occurring? (years of follow-up). Is recontact possible?	Longitudinal life course follow-up be electronic health records Recall for research is possible
Notes/Comments	
Questions Relating to Sharing & Collaboration	
May we make the information you provided about your cohort available on an open website?	Yes
Are you willing to share data from your cohort? If so, would you share:	We will place those on our website
a) individual data (redacted to protect confidentiality)?	
b) summary data (counts, distributions)?	
c) metadata (descriptive information on data collection methods)?	
d) case report forms and other data collection materials?	
What do you see as the values of sharing?	Enormous value but this must occur inside our data centre it is a reading library not a lending library
What challenges do you anticipate with sharing?	
What specific legal or regulatory barriers in your country or cohort (aside from ensuring confidentiality and appropriate consent) would prevent you from sharing data with other cohort of cohorts investigators, or openly with anyone who requests them?	We have managed those by asking researchers to work in our data centre and by participating in a GA4GH driver project with EBI around application programming interfaces
What aspects of your cohort are intended for translation to clinical care or population health?	As many as possible

How might genomic sequencing add to/enhance your study objectives?	We're doing this already
Might you be willing to contribute funding or other resources to support international collaboration?	We cannot commit funds but otherwise depending on the nature yes

Golestan Cohort Study

Questions Relating to Cohort	
Name of study	Golestan Cohort Study
Principal Investigator/lead	Dr. Reza Malekzadeh, Dr. Christian Abnet, Dr. Paolo Boffetta, Dr. Paul Brennan, Dr. Farin Kamangar, Dr. Arash Etemadi
Contact email	farinkamangar@gmail.com, arash.etemadi@nih.gov
PubMed ID (or other information) for a protocol/marker paper on this study	Int J Epidemiol. 2010 Feb; 39(1): 52–59. doi: 10.1093/ije/dyp161
Study website	https://epi.grants.cancer.gov/Consortia/members/gcs.html
Purpose or major Objectives of study	To identify the risk factors associated with esophageal cancer and other chronic diseases
Disease areas of focus	Upper gastrointestinal cancers, non-communicable diseases
Is your cohort selected for a specific disease (cancer, diabetes) or unselected for disease?	Unselected
Current size of population (and target number of participants)	50,000
Participating countries	Iran
Period of enrollment (and is enrollment on-going?)	2004-2008, closed.
Demographic characteristics of participants (age range, proportion male/female, national origin, race)	Male and females from Northeast Iran (Golestan), 40-75 years of age at baseline, Turkmen (~80%) and non-Turkmen ethnicities.
Major diseases or phenotypes collected to date.	All cancer outcomes, cardiovascular diseases and mortality
Standardized clinical evaluation components measured, if applicable (e.g. height, weight, blood pressure, exercise testing, spirometry, etc.)	Height, weight, waist, hip and wrist circumferences, blood pressure
Electronic health/medical records or medical administrative data used to collect clinical phenotypes?	Yes—during follow up if necessary

Predominant type of electronic health records (e.g. Epic, Cerner, etc.)	Local cancer registry (in progress)
Other sources of clinical data	Hospital information
Environmental exposure data being obtained? What sort?	Fuel exposure, housing, type of heating used, water source, animal contact
Other data collected	Questionnaire Data: Food frequency questionnaire and dietary habits, physical activity, past medical history, family history, medication history, oral health, smoking, drug and alcohol use Laboratory Data (complete blood count, fasting blood sugar, total cholesterol, HDL cholesterol, triglycerides, alanine transaminase, aspartate transaminase, alkaline phosphatase, gamma- glutamyl transpeptidase, blood urea nitrogen, and creatinine levels) in a 20% subpopulation after 5 years
Biological specimens collected? What sort?	Yes, blood (serum, whole blood, buffy coat, plasma), urine, hair, and nail
Is there a central biobank?	Yes
DNA samples prepared (or available to be prepared) from each participant?	Yes, buffy coat is prepared and stored for each participant
Is genotyping being done on some/all participants?	No
Is genomic sequencing being done on some/all participants?	No
Other molecular analyses performed	Several ongoing exposure biomarker studies
Did participants provide consent regarding sharing of their data outside the initial study investigators?	Yes
How are data or specimens from the cohort made available for research? Any limitations on who can access the data (e.g. by country or sector?)	Data and specimens are available to all interested researchers. Proposals can be submitted via an online access portal and evaluated by the central committee. No specific limitations.
What study information or data are returned to or accessible by participants?	Results of laboratory tests performed are given back to individuals after enrollment. If an individual is found to be a disease carrier or at risk of a disease, they will be informed to follow it up with their primary care physician.
Follow-up occurring? (years of follow-up). Is recontact possible?	Yes, active annual follow up is performed for all participants; a 20% subsample is re-evaluated every 5 years, and re-contact is possible.
Notes/Comments	
Questions Relating to Sharing & Collaboration	
May we make the information you provided about your cohort available on an open website?	Yes

Are you willing to share data from your cohort? If so, would you share:	Yes
a) individual data (redacted to protect confidentiality)?	Yes (subject to proposal submission and approval)
b) summary data (counts, distributions)?	Yes
c) metadata (descriptive information on data collection methods)?	Yes
d) case report forms and other data collection materials?	Yes
What do you see as the values of sharing?	The Golestan Cohort Study started as a collaborative effort, and the leadership consider the expansion of their collaborative efforts as part of the Cohort objectives.
What challenges do you anticipate with sharing?	
What specific legal or regulatory barriers in your country or cohort (aside from ensuring confidentiality and appropriate consent) would prevent you from sharing data with other cohort of cohorts investigators, or openly with anyone who requests them?	The Cohort does not make data publicly available, but there are no barriers to sharing in well-planned protocols.
What aspects of your cohort are intended for translation to clinical care or population health?	All data collected, especially those found to be significant risk factors for common diseases, will be translated to preventative programs and/or clinical care.
How might genomic sequencing add to/enhance your study objectives?	
Might you be willing to contribute funding or other resources to support international collaboration?	Yes

Japan Public Health Center-based Prospective Study (JPHC)

Questions Relating to Cohort	
Name of study	Japan Public Health Center-based Prospective Study (JPHC Study)
Principal Investigator/lead	Dr. Shoichiro Tsugane
Contact email	jphcadmin@ml.res.ncc.go.jp
PubMed ID (or other information) for a protocol/marker paper on this study	PMID: 25104790
Study website	http://epi.ncc.go.jp/en/jphc/
Purpose or major Objectives of study	JPHC Study is conducted as a joint research by a designated study group for research to establish evidence to benefit health maintenance

	and improvement including cancer prevention based on multipurpose cohort studies.
Disease areas of focus	Cancer, cardiovascular, diabetes, mortality
Is your cohort selected for a specific disease (cancer, diabetes) or unselected for disease?	Selected
Current size of population (and target number of participants)	13,000
Participating countries	Japan
Period of enrollment (and is enrollment on-going?)	1990-1994
Demographic characteristics of participants (age range, proportion male/female, national origin, race)	49%Male, 51%Female, 40-69y/o, Japanese
Major diseases or phenotypes collected to date.	Broad - includes cancer, cardiovascular, diabetes, etc
Standardized clinical evaluation components measured, if applicable (e.g. height, weight, blood pressure, exercise testing, spirometry, etc.)	Data of health checkup;Height(standing), weight, blood pressure, Total cholesterol, Triglyceride, HDL-cholesterol, AST, ALT,
Electronic health/medical records or medical administrative data used to collect clinical phenotypes?	No
Predominant type of electronic health records (e.g. Epic, Cerner, etc.)	
Other sources of clinical data	Death registry, cancer registries
Environmental exposure data being obtained? What sort?	Residential air pollution (contracting)
Other data collected	
Biological specimens collected? What sort?	Blood (50,000 samples)
Is there a central biobank?	No
DNA samples prepared (or available to be prepared) from each participant?	Yes –prepared from about 20,000 participants, and are available to be prepared from remaining participants
Is genotyping being done on some/all participants?	Yes –done on about 20,000 participants.
Is genomic sequencing being done on some/all participants?	No
Other molecular analyses performed	Yes –metabolomic analysis done on about 1,000 participants.
Did participants provide consent regarding sharing of their data outside the initial study investigators?	No. After research project is permitted by the steering committee and is approved by IRB, we can share our data as the collaborative study.

How are data or specimens from the cohort made available for research? Any limitations on who can access the data (e.g. by country or sector?)	If research protocol is permitted by the steering committee and is approved by IRB, we can provide data or specimens as collaborative study based on DTA or MTA.
What study information or data are returned to or accessible by participants?	Data are not returned to or not accessible by participants. Study information is publicly opened in homepage site.
Follow-up occurring? (years of follow-up). Is recontact possible?	20+ (until 2023). Recontact is not possible.
Notes/Comments	This study is being conducted to determine what lifestyle habits are relevant to the incidence of diseases, by collecting information about lifestyle habits from approximately 100,000 people living in various parts of Japan, and performing a long-term follow-up of over 20 years regarding the development of their diseases. Areas of interest: Cancer information for general public; Causes and prevention of cancer, CVD and other diseases which interact healthy life. Began in 1990, active at present.
Questions Relating to Sharing & Collaboration	
May we make the information you provided about your cohort available on an open website?	Yes
Are you willing to share data from your cohort? If so, would you share:	Yes
a) individual data (redacted to protect confidentiality)?	Yes Comment: After research project is permitted by the steering committee and is approved by IRB, we can only share our data as the collaborative study.
b) summary data (counts, distributions)?	Yes Comment: After research project is permitted by the steering committee and is approved by IRB, we can only share our data as the collaborative study.
c) metadata (descriptive information on data collection methods)?	Yes Comment: After research project is permitted by the steering committee and is approved by IRB, we can only share our data as the collaborative study.
d) case report forms and other data collection materials?	Yes Comment: After research project is permitted by the steering committee and is approved by IRB, we can only share our data as the collaborative study.
What do you see as the values of sharing?	It is necessary to have large dataset and various ideas towards personalized prevention.
What challenges do you anticipate with sharing?	To establish system and harmonize data from each cohort seems to be challenging.

What specific legal or regulatory barriers in your country or cohort (aside from ensuring confidentiality and appropriate consent) would prevent you from sharing data with other cohort of cohorts investigators, or openly with anyone who requests them?	Ethical guideline and privacy protection law in Japan are very strict. Therefore, our data in not openly accessed with anyone who requests them, even though our data is de-identified.
What aspects of your cohort are intended for translation to clinical care or population health?	Individual-based guideline for healthy life
How might genomic sequencing add to/enhance your study objectives?	The addition of whole genome sequencing would allow detection of large effects of rare mutations on a variety of health conditions.
Might you be willing to contribute funding or other resources to support international collaboration?	No

Japan Public Health Center-based Prospective Study for the Next generation (JPHC-NEXT)

Questions Relating to Cohort	
Name of study	Japan Public Health Center-based Prospective Study for the Next generation (JPHC-NEXT)
Principal Investigator/lead	Dr. Shoichiro Tsugane
Contact email	jphcadmin@ml.res.ncc.go.jp
PubMed ID (or other information) for a protocol/marker paper on this study	
Study website	http://epi.ncc.go.jp/jphcnext/ (in Japanese)
Purpose or major Objectives of study	JPHC-NEXT Study is conducted to establish evidence to benefit health maintenance and to contribute toward personalized healthcare based on multipurpose cohort studies.
Disease areas of focus	Cancer, cardiovascular, diabetes, depression, dementia, mortality
Is your cohort selected for a specific disease (cancer, diabetes) or unselected for disease?	Unselected
Current size of population (and target number of participants)	115,405
Participating countries	Japan
Period of enrollment (and is enrollment on-going?)	2011-2016
Demographic characteristics of participants (age range, proportion male/female, national origin, race)	46%Male, 54%Female, 40-74y/o, Japanese

Major diseases or phenotypes collected to date.	Broad - includes cancer, cardiovascular, respiratory, musculoskeletal, diabetes, mental health, cognitive function, etc
Standardized clinical evaluation components measured, if applicable (e.g. height, weight, blood pressure, exercise testing, spirometry, etc.)	Data of health checkup;Height(standing), weight, blood pressure, Total cholesterol, Triglyceride, HDL-cholesterol, AST, ALT, HbA1c
Electronic health/medical records or medical administrative data used to collect clinical phenotypes?	Yes
Predominant type of electronic health records (e.g. Epic, Cerner, etc.)	Type of electronic health records depends on hospital. Therefore, we collect medical administrative data from National Health Insurance or DPC (diagnosis procedure combination) database which the data form is unified.
Other sources of clinical data	Death registry, cancer registries
Environmental exposure data being obtained? What sort?	
Other data collected	
Biological specimens collected? What sort?	Blood (50,000 samples), Urine (50,000samples)
Is there a central biobank?	No
DNA samples prepared (or available to be prepared) from each participant?	No
Is genotyping being done on some/all participants?	No
Is genomic sequencing being done on some/all participants?	No
Other molecular analyses performed	No
Did participants provide consent regarding sharing of their data outside the initial study investigators?	Yes. After research project is permitted by the steering committee and is approved by IRB, we can share our data as the collaborative study.
How are data or specimens from the cohort made available for research? Any limitations on who can access the data (e.g. by country or sector?)	If research protocol is permitted by the steering committee and is approved by IRB, we can provide data or specimens as collaborative study based on DTA or MTA.
What study information or data are returned to or accessible by participants?	Data are not returned to or not accessible by participants. Study information is publicly opened in homepage site.
Follow-up occurring? (years of follow-up). Is recontact possible?	6 years (until 2036). Recontact is possible.
Notes/Comments	This study is being conducted to elucidate risk factors for lifestyle-related disease and to contribute toward personalized healthcare, by collecting information about lifestyle habits from approximately 100,000 people living in various parts of Japan, and beginning follow-up

	regarding the development of their diseases. Areas of interest: Causes and prevention of lifestyle-related diseases. Began in 2011, active at present.
Questions Relating to Sharing & Collaboration	
May we make the information you provided about your cohort available on an open website?	Yes
Are you willing to share data from your cohort? If so, would you share:	Yes
a) individual data (redacted to protect confidentiality)?	Yes Comment: After research project is permitted by the steering committee and is approved by IRB, we can only share our data as the collaborative study.
b) summary data (counts, distributions)?	Yes Comment: After research project is permitted by the steering committee and is approved by IRB, we can only share our data as the collaborative study.
c) metadata (descriptive information on data collection methods)?	Yes Comment: After research project is permitted by the steering committee and is approved by IRB, we can only share our data as the collaborative study.
d) case report forms and other data collection materials?	Yes Comment: After research project is permitted by the steering committee and is approved by IRB, we can only share our data as the collaborative study.
What do you see as the values of sharing?	It is necessary to have large dataset and various ideas towards personalized prevention.
What challenges do you anticipate with sharing?	To establish system and harmonize data from each cohort seems to be challenging.
What specific legal or regulatory barriers in your country or cohort (aside from ensuring confidentiality and appropriate consent) would prevent you from sharing data with other cohort of cohorts investigators, or openly with anyone who requests them?	Ethical guideline and privacy protection law in Japan are very strict. Therefore, our data in not openly accessed with anyone who requests them, even though our data is de-identified.
What aspects of your cohort are intended for translation to clinical care or population health?	Individual-based guideline for healthy life
How might genomic sequencing add to/enhance your study objectives?	The addition of whole genome sequencing would allow detection of large effects of rare mutations on a variety of health conditions.
Might you be willing to contribute funding or other resources to support international collaboration?	No

Kaiser Permanente Research Program on Genes, Environment, and Health (RPGEH)

Questions Relating to Cohort	
Name of study	Kaiser Permanente Research Program on Genes, Environment, and Health (RPGEH)
Principal Investigator/lead	Dr. Catherine Schaefer
Contact email	cathy.schaefer@kp.org
PubMed ID (or other information) for a protocol/marker paper on this study	
Study website	http://www.dor.kaiser.org/external/DORExternal/rpgeh/index.aspx
Purpose or major Objectives of study	Create a resource for research on the genetic and environmental factors that influence common diseases such as heart disease, cancer, diabetes, high blood pressure, psychiatric disorders, Alzheimer's disease, asthma and other conditions.
Disease areas of focus	Adult diseases, risk factors, and health-related traits and behaviors
Is your cohort selected for a specific disease (cancer, diabetes) or unselected for disease?	Unselected, it is a general cohort
Current size of population (and target number of participants)	210,000 (target: 500,000)
Participating countries	United States; Northern California members of Kaiser Permanente Health Plan
Period of enrollment (and is enrollment on-going?)	2008- ongoing
Demographic characteristics of participants (age range, proportion male/female, national origin, race)	Age Range: 18 – 107 years, avg age = 63 yrs; Gender: 58% F, 42% M; Race-ethnicity: 4% African-American/Black; 7% Asian; 8% Hispanic-Latino; 1% Other; 80% Non-Hispanic White
Major diseases or phenotypes collected to date.	Broad
Standardized clinical evaluation components measured, if applicable (e.g. height, weight, blood pressure, exercise testing, spirometry, etc.)	Routine clinical measures extracted from EMR, including height, weight, blood pressure and laboratory tests
Electronic health/medical records or medical administrative data used to collect clinical phenotypes?	Yes, complete inpatient and outpatient records and administrative data, beginning 1995 and ongoing
Predominant type of electronic health records (e.g. Epic, Cerner, etc.)	
Other sources of clinical data	Specialized exams (echocardiography, spirometry, imaging, etc.)
Environmental exposure data being obtained? What sort?	Yes. Baseline Survey of demographic, behavioral factors; link to U.S. census and GIS databases; serum/plasma collected for environmental exposures

Other data collected	Yes. Baseline survey of demographic, behavioral and reproductive risk factors
Biological specimens collected? What sort?	Saliva, blood (EDTA-preserved and serum)
Is there a central biobank?	Yes
DNA samples prepared (or available to be prepared) from each participant?	Yes
Is genotyping being done on some/all participants?	Yes, 103,000 Genotyped
Is genomic sequencing being done on some/all participants?	Exome and WGS on selected subsamples
Other molecular analyses performed	
Did participants provide consent regarding sharing of their data outside the initial study investigators?	Yes - broad
How are data or specimens from the cohort made available for research? Any limitations on who can access the data (e.g. by country or sector?)	Application to Access Review Committee at Kaiser; dbGaP; Commercial access limited
What study information or data are returned to or accessible by participants?	None currently
Follow-up occurring? (years of follow-up). Is recontact possible?	Follow-up is ongoing using Kaiser records; consent permits recontact
Notes/Comments	All participants in the RPGEH are/were members of the Kaiser Permanente Medical Care Plan in Northern California at entry to the cohort. The RPGEH has recruited about 210,000 adults, including 25,000 women comprising a Pregnancy Cohort. The purpose of the cohort and associated resources are to enable research on genetic and environmental factors that influence a broad range of diseases and health-related traits. The overall goal is to enable and contribute research that improves the health and well-being of KP members and the public by preventing disease, reducing severity, and improving outcomes of medical care.
Questions Relating to Sharing & Collaboration	
May we make the information you provided about your cohort available on an open website?	Yes
Are you willing to share data from your cohort? If so, would you share:	We have shared and are willing to share a) individual-level de-identified data; b) summary data; c) meta-data; d) surveys
a) individual data (redacted to protect confidentiality)?	Yes (de-identified)
b) summary data (counts, distributions)?	Yes

c) metadata (descriptive information on data collection methods)?	Yes
d) case report forms and other data collection materials?	Yes, surveys
What do you see as the values of sharing?	Value of sharing is principally in increasing sample size and diversity.
What challenges do you anticipate with sharing?	Some challenges include a) harmonization of data, incomplete data (lowest common denominator), and misinterpretation of data; b) heavily resourced centers (e.g., Broad) concentrate “shared” data and have lion’s share of research opportunities; c) discouragement of programmatic research that depends on refinement of hypotheses and data; NIH study sections don’t fund projects that utilize “shared” data
What specific legal or regulatory barriers in your country or cohort (aside from ensuring confidentiality and appropriate consent) would prevent you from sharing data with other cohort of cohorts investigators, or openly with anyone who requests them?	Ensuring privacy and confidentiality is not simple. Phenotypic data come from active medical records. Uncertainty about continuing health insurance and coverage of pre-existing conditions has a chilling effect on governance of cohort in regard to sharing
What aspects of your cohort are intended for translation to clinical care or population health?	All, potentially. Initial focus is on pharmacogenetics and avoidance of iatrogenic/adverse events
How might genomic sequencing add to/enhance your study objectives?	Limited utility for research on common diseases at present; may add information on rare variants useful for some conditions. Imputation of genotypic data has become so good, with imputation of even very rare variants, that the extra expense and logistical problems engendered by WGS data should be reserved for specifically selected and focused problems. Exome sequence data may have some limited uses. The costs of sequencing will suck resources away from development of much more useful types of data.
Might you be willing to contribute funding or other resources to support international collaboration?	No Comment: Resources that support maintenance of cohort are not controlled by cohort. No funding to contribute.

Korea Biobank Project

Questions Relating to Cohort	
Name of study	Korea Biobank Project
Principal Investigator/lead	Dr. Jae-Pil Jeon, PhD
Contact email	jaepiljeon@korea.kr
PubMed ID (or other information) for a protocol/marker paper on this study	PMIDs: 24465232; 24159511
Study website	http://www.nih.go.kr/NIH/eng/contents/NihEngContentView.jsp?cid=65660&menuIds=HOME004-MNU2210-MNU2327-MNU2346

Purpose or major Objectives of study	To establish the nation-wide biobank infrastructure for biomedical and public health research. The Korea Biobank Project(KBP) supports both population-based biobanking of KNIH and KCDC (KoGES, KNHANES, etc) and patient-based biobanking of university hospitals.
Disease areas of focus	Mixed
Is your cohort selected for a specific disease (cancer, diabetes) or unselected for disease?	Unselected
Current size of population (and target number of participants)	830,000 (including 245,000 of KoGES participants)
Participating countries	Republic of Korea
Period of enrollment (and is enrollment on-going?)	On-going
Demographic characteristics of participants (age range, proportion male/female, national origin, race)	Korean
Major diseases or phenotypes collected to date.	Broad - includes general population and disease patients
Standardized clinical evaluation components measured, if applicable (e.g. height, weight, blood pressure, exercise testing, spirometry, etc.)	Varies by types of cohorts or patient-based biobanks. KoGES collects anthropometric(height, weight, waist and hip circumference), blood pressure, blood glucose, blood lipid profile, Food Frequency Questionnaire, 24-hour diet recall.
Electronic health/medical records or medical administrative data used to collect clinical phenotypes?	Yes (hospital and health examination center)
Predominant type of electronic health records (e.g. Epic, Cerner, etc.)	NA
Other sources of clinical data	-KoGES provides questionnaire(self-report of diagnosed disease, medication), national health insurance, national cancer registry, national death registry, etc.
Environmental exposure data being obtained? What sort?	No
Other data collected	Physical measurements, lifestyle, etc.
Biological specimens collected? What sort?	KoGES: serum, plasma, DNA, urine, lymphocytes KNHANES: serum, plasma, DNA, urine Regional biobanks: specific types of biospecimens varies by biobanks.
Is there a central biobank?	Yes. The Korea Biobank Network consists of a central biobank (National Biobank of Korea) and 17 regional biobanks.
DNA samples prepared (or available to be prepared) from each participant?	Almost (Depending on the cohort or disease group.)
Is genotyping being done on some/all participants?	Some . (130,000 participants of KoGES)
Is genomic sequencing being done on some/all participants?	Some . (1,700 participants of KoGES)

Other molecular analyses performed	Some
Did participants provide consent regarding sharing of their data outside the initial study investigators?	Broad consent (comply with the Bioethics and Safety Act)
How are data or specimens from the cohort made available for research? Any limitations on who can access the data (e.g. by country or sector?)	Available for research projects approved by KNIH Biobank and Data Sharing Committee
What study information or data are returned to or accessible by participants?	Clinical exam data, Nutritional status
Follow-up occurring? (years of follow-up). Is recontact possible?	Varies by cohorts and biobanks.
Notes/Comments	In 2008, the Korea Biobank Project was initiated to advance biomedical and public health research by establishing the nation-wide biobank network (Korea Biobank Network) to collect, manage, and provide high-quality human biospecimens and related data. Currently, the Korea Biobank Project supports biorepositories of one central population-based biobank (supporting the Korea Genome and Epidemiology Study, Korea National Health and Nutritional Examination Survey, etc.) and 17 regional patient-based biobanks (set-up in university hospitals) across the country.
Questions Relating to Sharing & Collaboration	
May we make the information you provided about your cohort available on an open website?	Yes
Are you willing to share data from your cohort? If so, would you share:	Yes Comment: samples and data are accessible to all projects approved by distribution review committee
a) individual data (redacted to protect confidentiality)?	Yes Comment: Data sharing policy varies by biobank members of the Korea Biobank Network. The KoGES resource is already accessible to domestic researchers for health research, subject to approval of an application by the KNIH Biobank Data Sharing Committee. Individual data sharing with international researchers is restrictive, but international collaboration using meta-analysis is possible.
b) summary data (counts, distributions)?	Yes
c) metadata (descriptive information on data collection methods)?	Yes
d) case report forms and other data collection materials?	Yes
What do you see as the values of sharing?	Responsible research & innovation
What challenges do you anticipate with sharing?	Obtaining high-quality bioresources with appropriate high-quality consent. In addition, many researchers may lack the funds or access to data to answer the questions that they are interested in.

What specific legal or regulatory barriers in your country or cohort (aside from ensuring confidentiality and appropriate consent) would prevent you from sharing data with other cohort of cohorts investigators, or openly with anyone who requests them?	Legal restrictions on the use of personal information
What aspects of your cohort are intended for translation to clinical care or population health?	All
How might genomic sequencing add to/enhance your study objectives?	Enhancing detection of rare mutations
Might you be willing to contribute funding or other resources to support international collaboration?	Yes Comment: No funds available now.

Korean Cancer Prevention Study-II (KCPS-II) Biobank

Questions Relating to Cohort	
Name of study	Korean Cancer Prevention Study-II (KCPS-II) Biobank
Principal Investigator/lead	Dr. Sun Ha Jee
Contact email	jsunha@yuhs.ac
PubMed ID (or other information) for a protocol/marker paper on this study	PMID: 29186422
Study website	
Purpose or major Objectives of study	To examine the determinants and long-term consequences of the metabolic syndrome that is associated with increased risk of cardiovascular diseases and cancers.
Disease areas of focus	Mixed
Is your cohort selected for a specific disease (cancer, diabetes) or unselected for disease?	Selected
Current size of population (and target number of participants)	156,701
Participating countries	Korea
Period of enrollment (and is enrollment on-going?)	2004-2013
Demographic characteristics of participants (age range, proportion male/female, national origin, race)	61% M/ 39% F; 42 y/o; Korea residents 20 – 84 y/o

Major diseases or phenotypes collected to date.	Broad –includes cardiovascular, cancers, diabetes, dementia, pulmonary disease, tuberculosis
Standardized clinical evaluation components measured, if applicable (e.g. height, weight, blood pressure, exercise testing, spirometry, etc.)	Anthropometry (Body mass index, Waist circumference), blood pressure, pulmonary function Blood measurements: lipid profile, fasting glucose test, liver function test, kidney function, albumin, bilirubin, tumour markers (CEA, CA19-9, CA125), uric acid, adiponectin, insulin, Cr
Electronic health/medical records or medical administrative data used to collect clinical phenotypes?	Yes
Predominant type of electronic health records (e.g. Epic, Cerner, etc.)	
Other sources of clinical data	National cancer registry, NHIS (hospital admissions and outpatients), and death registry
Environmental exposure data being obtained? What sort?	POPs, BPA, Phthalate
Other data collected	
Biological specimens collected? What sort?	Whole Blood, Serum
Is there a central biobank?	Yes (located in Yonsei University)
DNA samples prepared (or available to be prepared) from each participant?	Yes (DNA or whole blood)
Is genotyping being done on some/all participants?	Yes (13,937 K-chip available)
Is genomic sequencing being done on some/all participants?	
Other molecular analyses performed	Partially (metabolomics)
Did participants provide consent regarding sharing of their data outside the initial study investigators?	Yes
How are data or specimens from the cohort made available for research? Any limitations on who can access the data (e.g. by country or sector?)	Data are not available to be shared. However, collaborative studies are encouraged.
What study information or data are returned to or accessible by participants?	Any information are returned to participants by request.
Follow-up occurring? (years of follow-up). Is recontact possible?	8+, partial re-contact
Notes/Comments	The Korean Cancer Prevention Study-II (KCPS-II) Biobank is a long-term prospective study of 156,701 Koreans (94,840 men and 61,861 women) who undertook routine health assessments, provided blood samples and gave informed consent in 18 health promotion centres across South Korea.

Questions Relating to Sharing & Collaboration	
May we make the information you provided about your cohort available on an open website?	Yes
Are you willing to share data from your cohort? If so, would you share:	Yes, we can share a full summary data and a partial individual data for collaborative study.
a) individual data (redacted to protect confidentiality)?	Yes (partial individual data for collaborative study)
b) summary data (counts, distributions)?	Yes (full summary data for collaborative study)
c) metadata (descriptive information on data collection methods)?	
d) case report forms and other data collection materials?	
What do you see as the values of sharing?	Enlargement of sample size
What challenges do you anticipate with sharing?	
What specific legal or regulatory barriers in your country or cohort (aside from ensuring confidentiality and appropriate consent) would prevent you from sharing data with other cohort of cohorts investigators, or openly with anyone who requests them?	None
What aspects of your cohort are intended for translation to clinical care or population health?	More likely for the population health
How might genomic sequencing add to/enhance your study objectives?	It would definitely improve our study objectives.
Might you be willing to contribute funding or other resources to support international collaboration?	No; no funds available to do so

Korean Genome and Epidemiology Study (KoGES)

Questions Relating to Cohort	
Name of study	Korean Genome and Epidemiology Study (KoGES)
Principal Investigator/lead	Dr. Sung Soo Kim
Contact email	ksungsoo@korea.kr
PubMed ID (or other information) for a protocol/marker paper on this study	PMID 27085081
Study website	Korea National Institutes of Health Homepage

	(http://www.nih.go.kr/NIH/eng/main.jsp) >> Korean Genome and Epidemiology Study(KoGES) (http://www.nih.go.kr/NIH/eng/contents/NihEngContentView.jsp?cid=65199&menuIds=HOME004-MNU2261-MNU2262-MNU2263-MNU2264)
Purpose or major Objectives of study	To improve the prediction, prevention, diagnosis and treatment of major chronic diseases by establishing a basis for domestic epidemiologic and genome research
Disease areas of focus	Mixed
Is your cohort selected for a specific disease (cancer, diabetes) or unselected for disease?	Unselected
Current size of population (and target number of participants)	245,000
Participating countries	South Korea, Vietnam, Cambodia, Japan, China
Period of enrollment (and is enrollment on-going?)	Varies; 2001-2013(completed)
Demographic characteristics of participants (age range, proportion male/female, national origin, race)	Varies by cohort : Ansan & Ansung - 47.4%M/52.6%F(40-69 yr at baseline), Health Examinee cohort - 34.2%M/65.8%F(40-79yr at baseline), Rural-based cohort - 38.2%M/61.8%F(40yr+ at baseline)
Major diseases or phenotypes collected to date.	Broad - All types of cancer, type 2 diabetes, hypertension, cardiovascular diseases, cerebrovascular diseases, obesity, metabolic syndrome, all types of underlying causes of death etc
Standardized clinical evaluation components measured, if applicable (e.g. height, weight, blood pressure, exercise testing, spirometry, etc.)	Anthropometric(height, weight, waist and hip circumference), blood pressure, blood glucose, blood lipid profile etc. Food Frequency Questionnaire (210,000 person subset), 24-hour diet recall (62,000 person subset)
Electronic health/medical records or medical administrative data used to collect clinical phenotypes?	No
Predominant type of electronic health records (e.g. Epic, Cerner, etc.)	NA
Other sources of clinical data	Questionnaire(self-report of diagnosed disease, medication), national health insurance, national cancer registry, national death registry
Environmental exposure data being obtained? What sort?	Residential air pollution
Other data collected	SES, education, lifestyle(smoking, alcohol, sleep), family disease history, reproductive factors(for women), blood test results, physical activity, Diet(FFQ, 24hr recall)
Biological specimens collected? What sort?	Blood, urine
Is there a central biobank?	Yes
DNA samples prepared (or available to be prepared) from each participant?	Yes

Is genotyping being done on some/all participants?	Yes, for subset (130,000 participants, SNP Chip: 830,000 SNPs)
Is genomic sequencing being done on some/all participants?	Yes (1,700 subset)
Other molecular analyses performed	Biomarkers(subset)
Did participants provide consent regarding sharing of their data outside the initial study investigators?	Yes (for health-related research)
How are data or specimens from the cohort made available for research? Any limitations on who can access the data (e.g. by country or sector?)	Project approved by data sharing committee
What study information or data are returned to or accessible by participants?	Clinical exam data, Nutritional status
Follow-up occurring? (years of follow-up). Is recontact possible?	Yes (5~17); Yes
Notes/Comments	KoGES, initiated in 2001, studied more than 260 traits through epidemiological surveys, physical examinations, and laboratory tests. KoGES consists of 7 cohorts: The population-based cohorts, which are Ansan cohort (n = 5,012), Ansung cohort (n = 5,018), Rural-based cohort (n = 28,338) and Health Examinee cohort (n = 173,346), recruited aged 40-69 years in 2001-2013. The gene-environmental model studies include the Marriage-based Immigrant cohort (n = 7,191), Korean Emigrant cohort (n = 1063), and Twins and Family cohort (n = 3,202). Follow-up has been conducted through periodic survey and data linkage to the national database (death records, cancer registry, national health insurance).
Questions Relating to Sharing & Collaboration	
May we make the information you provided about your cohort available on an open website?	Yes
Are you willing to share data from your cohort? If so, would you share:	Yes
a) individual data (redacted to protect confidentiality)?	No Comment:: It is already accessible to domestic researchers for health research, subject to approval of an application by the KNIH Biobank data sharing committee. Individual data sharing with international researchers is restrictive, but international collaboration using meta-analysis is possible.
b) summary data (counts, distributions)?	Yes
c) metadata (descriptive information on data collection methods)?	Yes
d) case report forms and other data collection materials?	Yes

What do you see as the values of sharing?	Expand the scope and use of the study data; enhance statistical power to study relatively rare exposures and outcomes; expect new approaches of diverse expertise
What challenges do you anticipate with sharing?	Data standardization, protecting privacy of participants
What specific legal or regulatory barriers in your country or cohort (aside from ensuring confidentiality and appropriate consent) would prevent you from sharing data with other cohort of cohorts investigators, or openly with anyone who requests them?	Legal restrictions on the use of personal information
What aspects of your cohort are intended for translation to clinical care or population health?	All
How might genomic sequencing add to/enhance your study objectives?	Detection of rare variants, development of precise prediction model
Might you be willing to contribute funding or other resources to support international collaboration?	No Comment: No funds available to do so yet.

LifeGene (and EpiHealth)

Questions Relating to Cohort	
Name of study	LifeGene (and sister cohort, EpiHealth)
Principal Investigator/lead	Dr. Nancy Pedersen
Contact email	Nancy.Pedersen@ki.se
PubMed ID (or other information) for a protocol/marker paper on this study	PMID: 21104112
Study website	https://www.lifegene.se/For-scientists/About-LifeGene/
Purpose or major Objectives of study	To build up a prospective cohort as a resource for research in all medical and behavioral disciplines.
Disease areas of focus	Mixed
Is your cohort selected for a specific disease (cancer, diabetes) or unselected for disease?	Unselected
Current size of population (and target number of participants)	51,300 in LifeGene (target: 300,000)
Participating countries	Sweden
Period of enrollment (and is enrollment on-going?)	Sept 2009 – Dec 2016

Demographic characteristics of participants (age range, proportion male/female, national origin, race)	60.3% F, index population 18-50 y/o, participants newborn - >100 y/o, Swedish population
Major diseases or phenotypes collected to date.	Broad - includes asthma, allergies, infections, obesity, chronic fatigue and pain, eating disorders, mental health, hearing, respiratory, cardiovascular
Standardized clinical evaluation components measured, if applicable (e.g. height, weight, blood pressure, exercise testing, spirometry, etc.)	Anthropometry (e.g., weight, height, bioimpedence), blood pressure , spirometry, pure tone audiometry
Electronic health/medical records or medical administrative data used to collect clinical phenotypes?	Not to collect clinical phenotypes but linkage to national health records is possible.
Predominant type of electronic health records (e.g. Epic, Cerner, etc.)	
Other sources of clinical data	Inpatient and outpatient registries, death, cancer and prescribed drug registries
Environmental exposure data being obtained? What sort?	Residential air and noise pollution
Other data collected	Extensive questionnaire including info on diet, work history, substance use, stress, trauma, etc.
Biological specimens collected? What sort?	Blood, urine
Is there a central biobank?	Yes
DNA samples prepared (or available to be prepared) from each participant?	Yes – prepared on all.
Is genotyping being done on some/all participants?	Yes
Is genomic sequencing being done on some/all participants?	Yes
Other molecular analyses performed	Hemoglobin, HbA1C, CRP, creatinine, cholesterol, Apolipoprotein A1, Apolipoprotein B on all. Proteomics and metabolomics on subsets
Did participants provide consent regarding sharing of their data outside the initial study investigators?	Yes - broad
How are data or specimens from the cohort made available for research? Any limitations on who can access the data (e.g. by country or sector?)	Applicants must be associated Sweden institutions or collaborate with Swedish researchers; ethical approval necessary;project approved by a data access committee
What study information or data are returned to or accessible by participants?	clinical exam results including blood chemistries
Follow-up occurring? (years of follow-up). Is recontact possible?	Annual followup of questionnaires. Recontact is possible

Notes/Comments	Sister cohort in Sweden: EpiHealth - index participants were 45-70 years old (contact person Peter Nilsson: peter.nilsson@med.lu.se) and together we now have approximately 60,000 with biobanked samples and nearly identical data collection. https://www.epihealth.se/For-scientists/
Questions Relating to Sharing & Collaboration	
May we make the information you provided about your cohort available on an open website?	Yes
Are you willing to share data from your cohort? If so, would you share:	Yes
a) individual data (redacted to protect confidentiality)?	Yes/No Comment: Not on an open website
b) summary data (counts, distributions)?	Yes
c) metadata (descriptive information on data collection methods)?	Yes Comment: already available @ Maelstrom
d) case report forms and other data collection materials?	Yes
What do you see as the values of sharing?	Increases collaboration opportunities, novel uses of data
What challenges do you anticipate with sharing?	New European regulations (GDPR)
What specific legal or regulatory barriers in your country or cohort (aside from ensuring confidentiality and appropriate consent) would prevent you from sharing data with other cohort of cohorts investigators, or openly with anyone who requests them?	We are not allowed to “openly share with anyone”. New legislation is coming into place which may make sharing data burdensome but not impossible.
What aspects of your cohort are intended for translation to clinical care or population health?	All
How might genomic sequencing add to/enhance your study objectives?	Would provide information on rare variants
Might you be willing to contribute funding or other resources to support international collaboration?	No Comment: don’t even have enough funds to maintain own cohort

LIFEPATH

Questions Relating to Cohort	
Name of study	LIFEPATH (Life-course biological pathways underlying social differences in healthy aging)

Principal Investigator/lead	Dr. Terrence Simmons
Contact email	t.simmons@imperial.ac.uk
PubMed ID (or other information) for a protocol/marker paper on this study	PMID: 28159391
Study website	http://www.lifepathproject.eu/
Purpose or major Objectives of study	To investigate the biological pathways underlying social differences in healthy ageing.
Disease areas of focus	Mixed
Is your cohort selected for a specific disease (cancer, diabetes) or unselected for disease?	Unselected
Current size of population (and target number of participants)	>235,000
Participating countries	Intercontinental - 7 European; Australian; USA
Period of enrollment (and is enrollment on-going?)	Varies by cohort
Demographic characteristics of participants (age range, proportion male/female, national origin, race)	Different cohorts have different age ranges including children
Major diseases or phenotypes collected to date.	Very broad, cancer, diabetes, CVD, frailty, mental health, healthy aging
Standardized clinical evaluation components measured, if applicable (e.g. height, weight, blood pressure, exercise testing, spirometry, etc.)	Yes-ish (yes for all or subset of participants for 15/17 cohorts)
Electronic health/medical records or medical administrative data used to collect clinical phenotypes?	Varies by cohort
Predominant type of electronic health records (e.g. Epic, Cerner, etc.)	Only in some cohorts
Other sources of clinical data	Registries, medical administrative data
Environmental exposure data being obtained? What sort?	Varies by cohort - some cohorts have air pollution and noise
Other data collected	functional outcomes (walking speed, grip strength), sleep, cognitive function
Biological specimens collected? What sort?	Blood
Is there a central biobank?	No
DNA samples prepared (or available to be prepared) from each participant?	For subsets

Is genotyping being done on some/all participants?	Some
Is genomic sequencing being done on some/all participants?	No
Other molecular analyses performed	Methylation, inflammatory markers, metabolomics
Did participants provide consent regarding sharing of their data outside the initial study investigators?	Yes - Broad for 14/17 cohorts
How are data or specimens from the cohort made available for research? Any limitations on who can access the data (e.g. by country or sector?)	Varies with cohort
What study information or data are returned to or accessible by participants?	No
Follow-up occurring? (years of follow-up). Is recontact possible?	Yes, depending on cohort
Notes/Comments	To investigate the biological pathways underlying social differences in healthy ageing.
Questions Relating to Sharing & Collaboration	
May we make the information you provided about your cohort available on an open website?	Yes
Are you willing to share data from your cohort? If so, would you share:	Yes
a) individual data (redacted to protect confidentiality)?	No
b) summary data (counts, distributions)?	Yes
c) metadata (descriptive information on data collection methods)?	Yes
d) case report forms and other data collection materials?	Yes
What do you see as the values of sharing?	Increase power of analyses, increase variability in exposures
What challenges do you anticipate with sharing?	
What specific legal or regulatory barriers in your country or cohort (aside from ensuring confidentiality and appropriate consent) would prevent you from sharing data with other cohort of cohorts investigators, or openly with anyone who requests them?	

What aspects of your cohort are intended for translation to clinical care or population health?	
How might genomic sequencing add to/enhance your study objectives?	
Might you be willing to contribute funding or other resources to support international collaboration?	No: We are a group of research institutions without resources to provide funding

Malaysian Cohort

Questions Relating to Cohort	
Name of study	Malaysian Cohort
Principal Investigator/lead	Dr. Rahman Jamal
Contact email	rahmanj@ppukm.ukm.edu.my
PubMed ID (or other information) for a protocol/marker paper on this study	PMID: 24729425
Study website	http://mycohort.gov.my/site/
Purpose or major Objectives of study	1. Identify risk factors for non-communicable diseases 2. Discovery biomarkers for early detection of diseases such as cancers 3. Set-up a large resource for research through big data and a large biobank 4. Participate in collaborative research with other countries or research institutions
Disease areas of focus	Non-communicable diseases including cancers
Is your cohort selected for a specific disease (cancer, diabetes) or unselected for disease?	Unselected
Current size of population (and target number of participants)	106,527
Participating countries	Malaysia
Period of enrollment (and is enrollment on-going?)	2007-2013
Demographic characteristics of participants (age range, proportion male/female, national origin, race)	Age range: 35-70 years Male: Female = 49:51 Malaysians Malays: Chinese: Indians
Major diseases or phenotypes collected to date.	Cancers, Diabetes, Hypercholesterolemia, Cardiovascular Diseases

Standardized clinical evaluation components measured, if applicable (e.g. height, weight, blood pressure, exercise testing, spirometry, etc.)	Height, Weight, BMI, Waist Circumference, Hip Circumference Blood pressure Spirometry Body composition index
Electronic health/medical records or medical administrative data used to collect clinical phenotypes?	No
Predominant type of electronic health records (e.g. Epic, Cerner, etc.)	NA
Other sources of clinical data	NA
Environmental exposure data being obtained? What sort?	Self reported exposure from questionnaire including occupation etc. GIS mapping of each participant address (which can be used to extract air pollution index, location of factories erc)
Other data collected	Medical and surgical history, medication history, smoking and alcohol
Biological specimens collected? What sort?	Blood, urine
Is there a central biobank?	Yes
DNA samples prepared (or available to be prepared) from each participant?	Yes
Is genotyping being done on some/all participants?	Yes (some participants)
Is genomic sequencing being done on some/all participants?	Yes (whole genome sequencing on 26 participants)
Other molecular analyses performed	Whole exome sequencing on patients with hypercholesterolemia
Did participants provide consent regarding sharing of their data outside the initial study investigators?	Not specifically
How are data or specimens from the cohort made available for research? Any limitations on who can access the data (e.g. by country or sector?)	There are guidelines for these published in our website. Limitations: We prefer not to transfer samples out of our biobank unless the analyses cannot be performed by us
What study information or data are returned to or accessible by participants?	The health screening data: the biophysical assessment and the baseline blood tests (blood sugar, lipids, etc)
Follow-up occurring? (years of follow-up). Is recontact possible?	10+
Notes/Comments	primary objectives: i) to study the roles and interaction of genes, environment and lifestyle in various diseases through a large-scale population cohort study; ii) to discover biomarkers for cancers/ diseases using the genomics and proteomics approach; iii) to consolidate and sustain the initiative for research in life sciences through a systematic discovery programme and also international collaborative research; iv) to establish a rich database of info and a bank of biospecimens which will become a national resource for

	research.
Questions Relating to Sharing & Collaboration	
May we make the information you provided about your cohort available on an open website?	Yes
Are you willing to share data from your cohort? If so, would you share:	Yes but depends on the proposed project: with proper data transfer agreement
a) individual data (redacted to protect confidentiality)?	No
b) summary data (counts, distributions)?	Yes
c) metadata (descriptive information on data collection methods)?	Yes
d) case report forms and other data collection materials?	Yes
What do you see as the values of sharing?	Collaborative research, co-authorship, larger sample size and study power
What challenges do you anticipate with sharing?	Local and institutional policies
What specific legal or regulatory barriers in your country or cohort (aside from ensuring confidentiality and appropriate consent) would prevent you from sharing data with other cohort of cohorts investigators, or openly with anyone who requests them?	Mainly institutional policies to protect biospecimens and data from leaving the country. This can be overcome through data transfer agreement.
What aspects of your cohort are intended for translation to clinical care or population health?	Discovery of risk factors or biomarkers for early detection of various diseases
How might genomic sequencing add to/enhance your study objectives?	Identify variants which are unique to our Malaysian population
Might you be willing to contribute funding or other resources to support international collaboration?	Yes but this will not come from my institution but from the funding agencies. Transfer of funds are currently not encouraged but there have been exceptions.

Maule Cohort (MAUCO Study)

Questions Relating to Cohort	
Name of study	Maule Cohort (MAUCO Study)
Principal Investigator/lead	Dr. Catterina Ferreccio
Contact email	cferrec@med.puc.cl

PubMed ID (or other information) for a protocol/marker paper on this study	PMID: 26847446 https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4743396/
Study website	http://www.accdis.cl/eng/en/
Purpose or major Objectives of study	To understand the natural history and factors involved in chronic diseases and ageing
Disease areas of focus	Cardiovascular and cancer
Is your cohort selected for a specific disease (cancer, diabetes) or unselected for disease?	Unselected
Current size of population (and target number of participants)	7,000 (target: 10,000)
Participating countries	Chile
Period of enrollment (and is enrollment on-going?)	2015-2018.
Demographic characteristics of participants (age range, proportion male/female, national origin, race)	adults ages 38-74 years, Chileans residing in the town of Molina, Currently 58% female
Major diseases or phenotypes collected to date.	Cardiovascular (CV) diseases/cancer/any major health event
Standardized clinical evaluation components measured, if applicable (e.g. height, weight, blood pressure, exercise testing, spirometry, etc.)	Clinical history, complete physical exam, anthropometric measures, impedanciometry, grip test, gait and balance, psychometric tests, ECG, chem panel, lipid profile, abdominal ultrasound
Electronic health/medical records or medical administrative data used to collect clinical phenotypes?	Yes
Predominant type of electronic health records (e.g. Epic, Cerner, etc.)	RedCap
Other sources of clinical data	Local outpatient and hospital records
Environmental exposure data being obtained? What sort?	Household conditions, exposure to animals, pesticides, heavy metals. Air, water and land pollution measurements
Other data collected	Outpatient consults and admissions during the period of enrollment
Biological specimens collected? What sort?	Blood, saliva and urines
Is there a central biobank?	Yes
DNA samples prepared (or available to be prepared) from each participant?	Yes, for 4000 participants
Is genotyping being done on some/all participants?	Yes, ancestry informative markers and approx 3000 SNPs associated with CV diseases and cancer
Is genomic sequencing being done on some/all participants?	Not currently, requires additional funding

Other molecular analyses performed	No
Did participants provide consent regarding sharing of their data outside the initial study investigators?	Yes
How are data or specimens from the cohort made available for research? Any limitations on who can access the data (e.g. by country or sector?)	Research committee evaluates proposals to access samples or data. Proposals need to be aligned with the purposes of the study and not be redundant
What study information or data are returned to or accessible by participants?	Lab test results, with interpretation and recommendations
Follow-up occurring? (years of follow-up). Is recontact possible?	Yes. Currently funded for 10 years, with applications for renewal afterwards
Notes/Comments	Maule Cohort (MAUCO Study) in Chile. It is prospective, aims to include 10,000 unselected adult participants from a single town in Chile and has a biobank. It was not originally conceived to include genomic data, but has incorporated it now (at least, ancestry). Here is a reference on its methodology: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4743396/ The director is Catterina Ferreccio, MD, MPH, Director of the Dept of Public Health at Universidad Católica in Santiago. Her email is cferrec@med.puc.cl
Questions Relating to Sharing & Collaboration	
May we make the information you provided about your cohort available on an open website?	Yes
Are you willing to share data from your cohort? If so, would you share:	Yes
a) individual data (redacted to protect confidentiality)?	Yes
b) summary data (counts, distributions)?	Yes
c) metadata (descriptive information on data collection methods)?	Yes
d) case report forms and other data collection materials?	Yes
What do you see as the values of sharing?	Improvements in the capacities for Chilean researchers; additional results and interpretation
What challenges do you anticipate with sharing?	Assurance that requests are aligned with the purposes of the study. Adequate recognition of the source

What specific legal or regulatory barriers in your country or cohort (aside from ensuring confidentiality and appropriate consent) would prevent you from sharing data with other cohort of cohorts investigators, or openly with anyone who requests them?	There are currently no regulations regarding international sharing of anonymized research data
What aspects of your cohort are intended for translation to clinical care or population health?	Identification of preventable/modifiable factor related to chronic disease
How might genomic sequencing add to/enhance your study objectives?	Identification of risk/protective factors and their relationship with environment
Might you be willing to contribute funding or other resources to support international collaboration?	Yes Comment: No funding available for this, but can consider other resources

Mexico City Prospective Study

Questions Relating to Cohort	
Name of study	Mexico City Prospective Study
Principal Investigator/lead	Jesus Alegre (Mexico) and Jonathan Emberson (UK)
Contact email	inypcjad@gmail.com; jonathan.emberson@ndph.ox.ac.uk
PubMed ID (or other information) for a protocol/marker paper on this study	PMID: 16556648
Study website	https://www.ctsu.ox.ac.uk/research/prospective-blood-based-study-of-150-000-individuals-in-mexico
Purpose or major Objectives of study	To gain a clearer understanding of the major determinants of morbidity and premature mortality in Mexico.
Disease areas of focus	Mixed (but with a particular focus on vascular/metabolic diseases).
Is your cohort selected for a specific disease (cancer, diabetes) or unselected for disease?	Unselected
Current size of population (and target number of participants)	159,755
Participating countries	Mexico
Period of enrollment (and is enrollment on-going?)	1998-2004
Demographic characteristics of participants (age range, proportion male/female, national origin, race)	Aged ≥35 years at recruitment; two-thirds female

Major diseases or phenotypes collected to date.	Cause-specific mortality through linkage to national mortality register
Standardized clinical evaluation components measured, if applicable (e.g. height, weight, blood pressure, exercise testing, spirometry, etc.)	Anthropometry (height, weight, waist and hip circumference) Blood pressure
Electronic health/medical records or medical administrative data used to collect clinical phenotypes?	No
Predominant type of electronic health records (e.g. Epic, Cerner, etc.)	Linkage to national mortality register
Other sources of clinical data	Baseline questionnaire (socioeconomic characteristics [eg, education, occupation, income]; lifestyle [eg, smoking, drinking, diet, physical activity, sleep]; prior diseases and medications; reproductive history) Baseline blood sample (plasma and buffy coat)
Environmental exposure data being obtained? What sort?	None
Other data collected	None
Biological specimens collected? What sort?	Baseline plasma and buffy coat samples for nearly all participants
Is there a central biobank?	Samples are in long-term storage in Oxford (UK)
DNA samples prepared (or available to be prepared) from each participant?	Yes Comment: Samples available to be prepared
Is genotyping being done on some/all participants?	No Comment: Not yet.
Is genomic sequencing being done on some/all participants?	No Comment: Not yet.
Other molecular analyses performed	HbA1c has been measured from all baseline buffy coat samples (see N Engl J Med 2016;375:1961-71). Baseline plasma samples have not yet been assayed for any biomarkers.
Did participants provide consent regarding sharing of their data outside the initial study investigators?	Participants provided consent for their data to be used for future (unspecified) health research.
How are data or specimens from the cohort made available for research? Any limitations on who can access the data (e.g. by country or sector?)	Requests for the sharing of data from this study will be considered according to the principles set out in NDPHs data sharing policy (www.ndph.ox.ac.uk/about/data-access-policy) and would need to be approved by both the Mexican and UK principal investigators.
What study information or data are returned to or accessible by participants?	None
Follow-up occurring? (years of follow-up). Is recontact possible?	Indefinite follow-up. Median follow-up among survivors is currently 14 years.

Notes/Comments	A prospective study of cause-specific mortality in 160 000 adults (previously called 'Proyecto Coyoacan'); Goal: Analysis of the main avoidable causes of chronic diseases in developing countries; Date commenced 1998; Cohort size: 160 000 adult men and women aged >35; Cohort sample: Representative sample of adults in the Coyoacan and Iztapalapa districts of Mexico City; Data collection: Baseline assessment (including a questionnaire, simple physical examination and blood collection) of 160 000 participants from 1998 to 2004; ongoing follow-up for cause-specific mortality now underway; repeat assessment of 10,000 survivors nearly complete; Other measurements: Questionnaire sought information on personal health behaviours (alcohol, smoking, diet, etc.), disease history and socioeconomic status; physical examination assessed height, weight, waist and hip circumferences, and blood pressure; Biological samples: Blood (now in storage for 155 000 participants) Purpose: To gain a clearer understanding of the major determinants of morbidity and premature mortality in Mexico.
Questions Relating to Sharing & Collaboration	
May we make the information you provided about your cohort available on an open website?	No Comment: The study cohort profile paper / website gives a much better background to the study / up-to-date information.
Are you willing to share data from your cohort? If so, would you share:	Requests for the sharing of data from this study will be considered according to the principles set out in NDPHs data sharing policy (www.ndph.ox.ac.uk/about/data-access-policy) and would need to be approved by both the Mexican and UK principal investigators.
a) individual data (redacted to protect confidentiality)?	See above
b) summary data (counts, distributions)?	See above
c) metadata (descriptive information on data collection methods)?	See above
d) case report forms and other data collection materials?	See above
What do you see as the values of sharing?	Collaboration Sharing expertise / new ideas Increased scientific output
What challenges do you anticipate with sharing?	Integrity of the science
What specific legal or regulatory barriers in your country or cohort (aside from ensuring confidentiality and appropriate consent) would prevent you from sharing data with other cohort of cohorts investigators, or openly with anyone who requests them?	Not currently aware of any specific barriers.
What aspects of your cohort are intended for translation to clinical care or population health?	Our findings on diabetes (N Engl J Med 2016;375:1961-71) have already translated to changes in national disease prevention policy. The availability of baseline blood samples for nearly all participants offers the potential for an improved understanding of biology (eg, related to diabetes) with subsequent translational opportunities.

How might genomic sequencing add to/enhance your study objectives?	Increased likelihood for genetic discovery / biological insight.
Might you be willing to contribute funding or other resources to support international collaboration?	Yes Comment: If study research funds permitted.

Million Veteran Program

Questions Relating to Cohort	
Name of study	Million Veteran Program
Principal Investigator/lead	Dr. Mike Gaziano
Contact email	michael.gaziano@va.gov; jmgaziano@bwh.harvard.edu
PubMed ID (or other information) for a protocol/marker paper on this study	
Study website	http://www.research.va.gov/mvp/
Purpose or major Objectives of study	To partner with Veterans receiving their care in the VA Healthcare System to study how genes affect health.
Disease areas of focus	Mixed
Is your cohort selected for a specific disease (cancer, diabetes) or unselected for disease?	Unselected
Current size of population (and target number of participants)	650,000 (target: 1,000,000)
Participating countries	United States
Period of enrollment (and is enrollment on-going?)	2011-ongoing
Demographic characteristics of participants (age range, proportion male/female, national origin, race)	M/F, all ages, race/ethnicity inclusive, all veterans. 18% African American, 7% hispanic
Major diseases or phenotypes collected to date.	All disease available in comprehensive EHR. Current projects included mental health including PTSD, schizophrenia, bipolar disorder, cardiovascular disease, cancer, dyslipidemia, tinnitus/ hearing loss, diabetes, eye disease among others.
Standardized clinical evaluation components measured, if applicable (e.g. height, weight, blood pressure, exercise testing, spirometry, etc.)	None
Electronic health/medical records or medical administrative data used to collect clinical phenotypes?	Yes including both VA and DoD health records

Predominant type of electronic health records (e.g. Epic, Cerner, etc.)	VISTA
Other sources of clinical data	Questionnaire
Environmental exposure data being obtained? What sort?	Yes
Other data collected	Yes
Biological specimens collected? What sort?	Blood
Is there a central biobank?	Yes
DNA samples prepared (or available to be prepared) from each participant?	Yes
Is genotyping being done on some/all participants?	Yes, genotyping is done on all participants
Is genomic sequencing being done on some/all participants?	Yes, subset
Other molecular analyses performed	Yes
Did participants provide consent regarding sharing of their data outside the initial study investigators?	Yes
How are data or specimens from the cohort made available for research? Any limitations on who can access the data (e.g. by country or sector?)	Project approved by data access committee
What study information or data are returned to or accessible by participants?	No
Follow-up occurring? (years of follow-up). Is recontact possible?	Follow up is through the EHR and will be done with follow-up questionnaires
Notes/Comments	Genomic discoveries are critically dependent on having a well-defined cohort with a sufficiently large number of subjects to ensure statistically and clinically meaningful associations. The Department of Veterans Affairs Million Veteran Program (MVP) was launched in 2011 to create a cohort of Veterans, one-million strong to leverage the excellent electronic medical record resource and combine it with genomic data and self-reported survey data to allow for both original and replication studies on diseases and conditions prevalent in Veterans. The Veterans Affairs (VA) Research and Development program is launching the Million Veteran Program (MVP), an important partnership between VA and Veterans to learn more about how genes affect health, to improve health care for Veterans. In order to do this, MVP will establish one of the largest databases of genetic, military exposure, lifestyle, and health information. Research findings based on MVP may lead to new ways of preventing and treating illnesses in Veterans. Such findings may help answer questions like "Why does a treatment work well for some Veterans but not for others?"; "Why are some Veterans at

	a greater risk for developing an illness?"; and "How can we prevent certain illnesses in the first place?" With the expected enrollment of one million Veterans over the next five to seven years, MVP aims to be one of the largest databases of its kind in the United States. Purpose: To partner with Veterans receiving their care in the VA Healthcare System to study how genes affect health.

Million Women Study

Questions Relating to Cohort	
Name of study	Million Women Study
Principal Investigator/lead	Prof. Valerie Beral
Contact email	valerie.beral@ceu.ox.ac.uk
PubMed ID (or other information) for a protocol/marker paper on this study	
Study website	http://www.millionwomenstudy.org/introduction/
Purpose or major Objectives of study	To study how various reproductive and lifestyle factors affect women's health.
Disease areas of focus	Hormones, reproductive factors, smoking, adiposity, alcohol
Is your cohort selected for a specific disease (cancer, diabetes) or unselected for disease?	Unselected
Current size of population (and target number of participants)	1,320,000
Participating countries	England, Scotland
Period of enrollment (and is enrollment on-going?)	1996-2001
Demographic characteristics of participants (age range, proportion male/female, national origin, race)	1 in 4 UK women born in 1935-1950. Over 95% White.
Major diseases or phenotypes collected to date.	99% linked to national health death databases. Includes all outcomes eg cardiovascular, respiratory, musculoskeletal, cancer, reproductive, mental health, cognitive function, neurodegenerative disease
Standardized clinical evaluation components measured, if applicable (e.g. height, weight, blood pressure, exercise testing, spirometry, etc.)	Height/weight, hip & waist circumferences, blood pressure
Electronic health/medical records or medical administrative data used to collect clinical phenotypes?	Yes, 99% survivors still linked after 20 years
Predominant type of electronic health records (e.g. Epic, Cerner, etc.)	

Other sources of clinical data	Death and cancer registries, hospital inpatient, cancer screening, primary care (for subset), 5 questionnaires 3-5 years apart
Environmental exposure data being obtained? What sort?	No
Other data collected	Over 4,000 variables, many with repeat measures eg diet, alcohol use, physical activity, smoking
Biological specimens collected? What sort?	Blood (for subset)
Is there a central biobank?	Yes
DNA samples prepared (or available to be prepared) from each participant?	Some
Is genotyping being done on some/all participants?	Yes - breast cancer SNPs, CVD SNPs
Is genomic sequencing being done on some/all participants?	No
Other molecular analyses performed	Not at present
Did participants provide consent regarding sharing of their data outside the initial study investigators?	Broad consent for medical research.
How are data or specimens from the cohort made available for research? Any limitations on who can access the data (e.g. by country or sector?)	See link to data access policy http://www.millionwomenstudy.org/files/MWS-DataAccessPolicy.pdf
What study information or data are returned to or accessible by participants?	None
Follow-up occurring? (years of follow-up). Is recontact possible?	Annual follow-up for deaths, cancers, hospital admissions for 20 years; re-contacted every 3-5 years
Notes/Comments	A national study of women's health, involving more than 1 million UK women aged 50+. Purpose: To study how various reproductive and lifestyle factors affect women's health.
Questions Relating to Sharing & Collaboration	
May we make the information you provided about your cohort available on an open website?	Yes
Are you willing to share data from your cohort? If so, would you share:	Yes Comment: Subject to approval by study data access committee
a) individual data (redacted to protect confidentiality)?	Yes
b) summary data (counts, distributions)?	Yes
c) metadata (descriptive information on data collection methods)?	Yes

d) case report forms and other data collection materials?	Yes
What do you see as the values of sharing?	Increased use of the data for increasingly wide set of purposes, as stand-alone dataset and through consortia.
What challenges do you anticipate with sharing?	Given complexity of data and varying level and focus of expertise of those using data, it is likely some analyses will be flawed.
What specific legal or regulatory barriers in your country or cohort (aside from ensuring confidentiality and appropriate consent) would prevent you from sharing data with other cohort of cohorts investigators, or openly with anyone who requests them?	Current data governance agreements mean we can share individual level linked follow-up data only through collaboration and at our department. This is under review. Lack of explicit consent for data sharing has to be taken into account in access decisions.
What aspects of your cohort are intended for translation to clinical care or population health?	All
How might genomic sequencing add to/enhance your study objectives?	Whole genome/exome sequencing would allow our cohort to contribute to studies of large effects of rare mutations, as well as to GWAS.
Might you be willing to contribute funding or other resources to support international collaboration?	No Comment: Funds not available for this.

Multiethnic Cohort Study (MEC)

Questions Relating to Cohort	
Name of study	Multiethnic Cohort Study (MEC, NCI)
Principal Investigator/lead	Dr. Loic Le Marchand
Contact email	loic@cc.hawaii.edu
PubMed ID (or other information) for a protocol/marker paper on this study	PMID: 10695593 PMC4482109
Study website	http://www.uhcancercenter.org/research/the-multiethnic-cohort-study-mec http://www.crch.org/multiethniccohort/
Purpose or major Objectives of study	To examine lifestyle risk factors, especially diet and nutrition, as well as genetic susceptibility in relation to the causation of cancer.
Disease areas of focus	Cancer; other chronic diseases
Is your cohort selected for a specific disease (cancer, diabetes) or unselected for disease?	Unselected
Current size of population (and target number of participants)	215,251 at recruitment; Blood was collected on 77,000 mostly in 2001-2006
Participating countries	US (Hawaii, California)

Period of enrollment (and is enrollment on-going?)	1993-1997. Cohort is closed to enrollment.
Demographic characteristics of participants (age range, proportion male/female, national origin, race)	45% M, 55% F, 45-75 y/o at baseline, Native Hawaiian; Japanese American; African American; Latino; Non-Hispanic white
Major diseases or phenotypes collected to date.	Broad categories of phenotypes: - cancer, death, obesity, diabetes, heart and vascular diseases, digestive and/or genitourinary diseases, autoimmune diseases
Standardized clinical evaluation components measured, if applicable (e.g. height, weight, blood pressure, exercise testing, spirometry, etc.)	Self-reported anthropometry (weight, height, waist and hip circumferences)
Electronic health/medical records or medical administrative data used to collect clinical phenotypes?	Medicare; California hospital discharge diagnosis; Kaiser Permanente Hawaii
Predominant type of electronic health records (e.g. Epic, Cerner, etc.)	
Other sources of clinical data	Questionnaires, Tumor registries, SEER, death certificates, NDI
Environmental exposure data being obtained? What sort?	Yes. Geocoding of address history with annotation from census and other environmental data sources
Other data collected	Diet history, medication, tobacco use, physical activity and female reproductive histories
Biological specimens collected? What sort?	Blood, urine, tumor tissue
Is there a central biobank?	biorepository at UHCC and USC
DNA samples prepared (or available to be prepared) from each participant?	Yes, for subset
Is genotyping being done on some/all participants?	Yes, for subset
Is genomic sequencing being done on some/all participants?	Yes, on subset
Other molecular analyses performed	metabolomic and methylation on subsets
Did participants provide consent regarding sharing of their data outside the initial study investigators?	Yes
How are data or specimens from the cohort made available for research? Any limitations on who can access the data (e.g. by country or sector?)	Project approved by data access committee
What study information or data are returned to or accessible by participants?	None

Follow-up occurring? (years of follow-up). Is recontact possible?	20+years; recontact possible
Notes/Comments	To examine lifestyle risk factors, especially diet and nutrition, as well as genetic susceptibility in relation to the causation of cancer.
Questions Relating to Sharing & Collaboration	
May we make the information you provided about your cohort available on an open website?	Yes
Are you willing to share data from your cohort? If so, would you share:	Yes
a) individual data (redacted to protect confidentiality)?	Yes
b) summary data (counts, distributions)?	Yes
c) metadata (descriptive information on data collection methods)?	Yes
d) case report forms and other data collection materials?	Yes Comment: already available on MEC website
What do you see as the values of sharing?	accelerate discovery; having a multiethnic population like MEC would help in fine mapping; to study rare diseases.
What challenges do you anticipate with sharing?	no funding to support sharing of data and specimen.
What specific legal or regulatory barriers in your country or cohort (aside from ensuring confidentiality and appropriate consent) would prevent you from sharing data with other cohort of cohorts investigators, or openly with anyone who requests them?	None anticipated based on current experience with GWAS
What aspects of your cohort are intended for translation to clinical care or population health?	Multiethnic composition
How might genomic sequencing add to/enhance your study objectives?	
Might you be willing to contribute funding or other resources to support international collaboration?	No Comment: None available

MyCode Community Health Initiative

Questions Relating to Cohort	
Name of study	MyCode Community Health Initiative

Principal Investigator/lead	Dr. Marc Williams
Contact email	mswilliams1@geisinger.edu
PubMed ID (or other information) for a protocol/marker paper on this study	PMID:28008009 PMID:26866580
Study website	https://www.geisinger.org/mycode
Purpose or major Objectives of study	Study the impact of the return of pathogenic variants in clinically actionable genes on health and economic outcomes of a large scale population-based exome sequencing project.
Disease areas of focus	Population based, not disease focused. Multiple health conditions are involved with prioritization of the three CDC Tier 1 conditions: Hereditary Breast/Ovarian Cancer syndrome; Lynch syndrome; Familial Hypercholesterolemia. Full list of conditions available at: https://www.geisinger.org/mycode/mycode-conditions
Is your cohort selected for a specific disease (cancer, diabetes) or unselected for disease?	Mostly unselected although the cohort includes a number of patients accrued from the cardiac cath lab, but this is no more than 4% of the entire cohort.
Current size of population (and target number of participants)	181,117 (as of Jan 1, 2018). Enrollment updated monthly at: https://www.geisinger.org/-/media/OneGeisinger/pdfs/ghs/research/mycode/scorecardinfographic-jan2018.pdf?la=en&hash=4B444A3649C7A849DC29DE5C56DA9E55C8FE9383 Target: Minimum 250000
Participating countries	US
Period of enrollment (and is enrollment on-going?)	Enrollment began in 2007 and is ongoing
Demographic characteristics of participants (age range, proportion male/female, national origin, race)	Full details available in PMID:26866580 While this is a subset of the current consented group, it is reflective of the entire group. However we are now consenting in Atlantic City, so we anticipate some change in the demographics over the next year, particularly in race and ethnic diversity.
Major diseases or phenotypes collected to date.	See condition list at: https://www.geisinger.org/mycode/mycode-conditions and the number of returns per condition at:https://www.geisinger.org/-/media/OneGeisinger/pdfs/ghs/research/mycode/ror-table-jan2018.pdf?la=en&hash=A2E0B02C174B5DA4EFCB8B84953B78C4CFC46151 Both of these are update monthly
Standardized clinical evaluation components measured, if applicable (e.g. height, weight, blood pressure, exercise testing, spirometry, etc.)	No information is obtained at the time of enrollment. All participants consent to use of all electronic health record data. Participants have a median of 14 years of EHR data. A high level view of available information is included in PMID:26866580
Electronic health/medical records or medical administrative data used to collect clinical phenotypes?	Yes as noted previously
Predominant type of electronic health records (e.g. Epic, Cerner, etc.)	Epic for the majority of enrollees, although enrollment from New Jersey will be on Cerner.
Other sources of clinical data	Can use the KeyHIE health information exchange. Claims data from Geisinger Health Plan is available on about 30-40% of participants

	which can capture medical care taking place outside the Geisinger system.
Environmental exposure data being obtained? What sort?	On a subset of individuals. Geisinger has an Environmental Health Institute. All MyCode participants are Geocoded allowing matching with exposure data from the EHI. Details of ongoing projects and data types are available at: https://www.geisinger.edu/research/departments-and-centers/environmental-health-institute
Other data collected	
Biological specimens collected? What sort?	Blood, Saliva. Other biospecimens are available for a subset of participants depending on there clinical care or participation in other studies (for example, bariatric surgery patients have liver and adipose tissue samples)
Is there a central biobank?	Yes
DNA samples prepared (or available to be prepared) from each participant?	Yes
Is genotyping being done on some/all participants?	Yes, all participants
Is genomic sequencing being done on some/all participants?	Exome sequencing all participants
Other molecular analyses performed	HLA typing
Did participants provide consent regarding sharing of their data outside the initial study investigators?	Broad research consent that allow sharing with outside investigators under the policies and procedures set forth by the MyCode program.
How are data or specimens from the cohort made available for research? Any limitations on who can access the data (e.g. by country or sector?)	Potential collaborators can contact either Geisinger or the Regeneron Genetics Center for information. There is a vetting and approval process for research involving outside investigators. No restrictions on potential collaborators
What study information or data are returned to or accessible by participants?	Pathogenic and likely pathogenic variants in the clinically actionable genes defined by the project. Pharmacogenomic information will be returned to a subset of individuals in a pilot project.
Follow-up occurring? (years of follow-up). Is recontact possible?	Recontact is allowed. Follow-up is ongoing and can occur in theory as long as the patient is alive and a Geisinger patient and hasn't withdrawn from the study.
Notes/Comments	The MyCode® Community Health Initiative is a precision health project at Geisinger that includes a system-wide biobank designed to store blood and other samples for research use by Geisinger and Geisinger collaborators. Samples and information in the biobank are used to do health research. Geisinger & Regeneron have agreed to an initial 5-year partnership, which may be extended to 10 years.
Questions Relating to Sharing & Collaboration	
May we make the information you provided about your cohort available on an open website?	Yes
Are you willing to share data from your cohort? If so, would you share:	Comment: As noted above we evaluate collaboration and data sharing on a case by case basis. We do publicly share allele frequency data on a website: http://www.discovehrshare.com/

a) individual data (redacted to protect confidentiality)?	Comment: case by case
b) summary data (counts, distributions)?	Yes Comment: Publicly available on http://www.discovehrshare.com/
c) metadata (descriptive information on data collection methods)?	Comment: case by case
d) case report forms and other data collection materials?	Comment: case by case
What do you see as the values of sharing?	Ability to increase knowledge at a faster pace. Availability of other perspectives to inform studies. Replication to confirm preliminary findings from other studies.
What challenges do you anticipate with sharing?	Balancing the inherent value of the sequences that were generated through the use of private funds, with the value to the broader scientific community of allowing full access.
What specific legal or regulatory barriers in your country or cohort (aside from ensuring confidentiality and appropriate consent) would prevent you from sharing data with other cohort of cohorts investigators, or openly with anyone who requests them?	None that I'm aware of.
What aspects of your cohort are intended for translation to clinical care or population health?	Any aspect that passes a threshold of utility for study in clinical care/population health.
How might genomic sequencing add to/enhance your study objectives?	N/A
Might you be willing to contribute funding or other resources to support international collaboration?	This would be difficult however I can't say no absolutely as there could be a case where the value of the project could justify contributing funding either in kind or financial.

Newfoundland and Labrador Genome Project

Questions Relating to Cohort	
Name of study	Newfoundland and Labrador Genome Project
Principal Investigator/lead	Dr. Michael Phillips
Contact email	info@sequencebio.com
PubMed ID (or other information) for a protocol/marker paper on this study	N/A
Study website	https://www.sequencebio.com/
Purpose or major Objectives of study	Discover drug targets and cure diseases within the Newfoundland founder population.
Disease areas of focus	General population and targeted cohorts

Is your cohort selected for a specific disease (cancer, diabetes) or unselected for disease?	Both
Current size of population (and target number of participants)	>520,000
Participating countries	Canada (Province of Newfoundland and Labrador)
Period of enrollment (and is enrollment on-going?)	2018 onward
Demographic characteristics of participants (age range, proportion male/female, national origin, race)	95% Caucasian and 5% other.
Major diseases or phenotypes collected to date.	Ascertainment underway.
Standardized clinical evaluation components measured, if applicable (e.g. height, weight, blood pressure, exercise testing, spirometry, etc.)	Phenotypic information being collected from baseline surveys and Provincial EMR data.
Electronic health/medical records or medical administrative data used to collect clinical phenotypes?	Yes (hospital, primary care and drug usage information)
Predominant type of electronic health records (e.g. Epic, Cerner, etc.)	Newfoundland and Labrador Centre for Health Information
Other sources of clinical data	Patient charts, supplemental phenotypic data
Environmental exposure data being obtained? What sort?	N/A
Other data collected	Baseline survey data
Biological specimens collected? What sort?	Saliva, blood, biopsies
Is there a central biobank?	No
DNA samples prepared (or available to be prepared) from each participant?	Yes
Is genotyping being done on some/all participants?	Yes - all participants will be genotyped using a genome wide array.
Is genomic sequencing being done on some/all participants?	Yes- WGS on selected participants
Other molecular analyses performed	Yes - selected technology for specific diseases
Did participants provide consent regarding sharing of their data outside the initial study investigators?	Broad consent of Sequence Bio uses and collaborations.

How are data or specimens from the cohort made available for research? Any limitations on who can access the data (e.g. by country or sector?)	Sequence Bio research oversight committee.
What study information or data are returned to or accessible by participants?	Research findings will be made available to participants' physicians for clinical confirmation where actionable and other findings will be returned over time to participants and their physicians to encourage engagement and participation.
Follow-up occurring? (years of follow-up). Is recontact possible?	10+ years, recontact is possible.
Notes/Comments	The NL Genome Project studies the genetic makeup of Newfoundland and Labrador residents; securely connects that data to their health data; and returns certain genetic guidance to them and their physicians (if they so choose). We aim to recontact relevant individuals, and direct recruiting toward families and large cohorts, ultimately to identify new drug targets and biomarkers.
Questions Relating to Sharing & Collaboration	
May we make the information you provided about your cohort available on an open website?	No Comment: Not while study protocol is awaiting ethics approval.
Are you willing to share data from your cohort? If so, would you share:	Yes Comment: Sequence Bio is open to discuss collaborations
a) individual data (redacted to protect confidentiality)?	Yes
b) summary data (counts, distributions)?	Yes
c) metadata (descriptive information on data collection methods)?	Yes
d) case report forms and other data collection materials?	Yes
What do you see as the values of sharing?	Sequence Bio is committed to working with partners who support our goals of returning value to the population of Newfoundland and Labrador. We recognize that we benefit from public and shared resources and will support the development of such resources to the maximum extent possible.
What challenges do you anticipate with sharing?	Any collaborative policies will need to address sensitivities expressed by Sequence Bio's participants and stakeholders concerning control and sovereignty of data and samples.
What specific legal or regulatory barriers in your country or cohort (aside from ensuring confidentiality and appropriate consent) would prevent you from sharing data with other cohort of cohorts investigators, or openly with anyone who requests them?	Access to personal information is governed by The Personal Information Protection and Electronic Documents Act (Canada) and The Personal Health Information Act (Newfoundland and Labrador)
What aspects of your cohort are intended for translation to clinical care or population health?	It is our intent to make our research grade data available for translation.

How might genomic sequencing add to/enhance your study objectives?	Whole genome sequencing is a core component of our research and discovery strategy.
Might you be willing to contribute funding or other resources to support international collaboration?	No Comment: No funds available to do so at this time

Northern Sweden Health and Disease Study

Questions Relating to Cohort	
Name of study	Northern Sweden Health and Disease Study
Principal Investigator/lead	Dr. Beatrice Melin
Contact email	beatrice.melin@onkologi.umu.se
PubMed ID (or other information) for a protocol/marker paper on this study	
Study website	http://www.biobank.umu.se/biobank/biobank---for-researchers/biobank-research---sample-collections/
Purpose or major Objectives of study	Public Health intervention at 40 and 50 years of age
Disease areas of focus	Many different outcomes, cancer , cardiovascular diseases and more
Is your cohort selected for a specific disease (cancer, diabetes) or unselected for disease?	Unselected
Current size of population (and target number of participants)	135,000
Participating countries	Sweden
Period of enrollment (and is enrollment on-going?)	1985 - ongoing
Demographic characteristics of participants (age range, proportion male/female, national origin, race)	Ages 18-83, Majority 40, 50, and 60 years old. 53.1% Female 46,9% Male. Origin and race representing inhabitants of the region, predominantly Swedish and Caucasian.
Major diseases or phenotypes collected to date.	Various, Cancer, cardiovascular disease, neurological diseases-disease specific coordinators
Standardized clinical evaluation components measured, if applicable (e.g. height, weight, blood pressure, exercise testing, spirometry, etc.)	Height, Weight, Waistline, Blood pressure, Cholesterol, Triglyceride, Blood sugar test
Electronic health/medical records or medical administrative data used to collect clinical phenotypes?	Quality registers, cancer registry, some medical register abstraktion through research groups at Umeå University.
Predominant type of electronic health records (e.g. Epic, Cerner, etc.)	NCSCross

Other sources of clinical data	
Environmental exposure data being obtained? What sort?	No
Other data collected	Lifestyle and dietary data
Biological specimens collected? What sort?	Blood
Is there a central biobank?	Yes in Umeå
DNA samples prepared (or available to be prepared) from each participant?	DNA samples have currently been prepared for 48 758 individuals, and can be prepared for anyone else who have left blood samples at a per sample cost.
Is genotyping being done on some/all participants?	Some approximately 20 000 individuals.
Is genomic sequencing being done on some/all participants?	Not systematically. A smaller control population of 300 individuals performed within a research project.
Other molecular analyses performed	
Did participants provide consent regarding sharing of their data outside the initial study investigators?	Yes, but information of transferring data to third country is lacking in the patient information
How are data or specimens from the cohort made available for research? Any limitations on who can access the data (e.g. by country or sector?)	Researchers can apply to use specimens or data from the cohort to the expert committee, to be able to access phenotype data of disease and obtain ethical and data regulation. Permissions, a collaboration with the disease coordinator is necessary. To be able to access phenotype data of disease and obtain ethical and data regulation.
What study information or data are returned to or accessible by participants?	Participants can ask to receive information about which studies their samples are used for. Specific results are not reported back to the cohort participants.
Follow-up occurring? (years of follow-up). Is recontact possible?	20+
Notes/Comments	NSHDS Cohort contains 3 sub cohorts: the Vasterbotten intervention programme cohort, the MONICA cohort, and the mammary (mammography) screening cohort. All individuals 40, 50, 60 years of age in the county's population are invited for screening. They are asked to complete a questionnaire concerning various lifestyle factors including diet and to donate a blood sample for frozen storage for later research purposes. The project started in 1985 and by December 2002 the cohort included 74,000 individuals, of whom 67,000 had donated blood samples.
Questions Relating to Sharing & Collaboration	
May we make the information you provided about your cohort available on an open website?	No
Are you willing to share data from your cohort? If so, would you share:	

a) individual data (redacted to protect confidentiality)?	Comment: Sharing for individual projects
b) summary data (counts, distributions)?	Yes Comment: Sharing for individual projects
c) metadata (descriptive information on data collection methods)?	Yes
d) case report forms and other data collection materials?	Yes
What do you see as the values of sharing?	Methods of data collection and storing of samples. Information is available the cohort webpage.
What challenges do you anticipate with sharing?	Data protection compliance to Swedish law, makes still collaborations possible but it can take time to access samples.
What specific legal or regulatory barriers in your country or cohort (aside from ensuring confidentiality and appropriate consent) would prevent you from sharing data with other cohort of cohorts investigators, or openly with anyone who requests them?	Data regulatory according to Swedish law.
What aspects of your cohort are intended for translation to clinical care or population health?	The health exams are used for intervention at inclusion.
How might genomic sequencing add to/enhance your study objectives?	Sequencing would be beneficial as the genetic substructures differ between different parts, even in Sweden.
Might you be willing to contribute funding or other resources to support international collaboration?	Yes Comment: It is possible but difficult to apply for national funding, for a consortium protocol.

Norwegian Family Based Life Course Study

Questions Relating to Cohort	
Name of study	Norwegian Family Based Life Course Study
Principal Investigator/lead	Dr. Øyvind Næss
Contact email	oyvind.nass@medisin.uio.no
PubMed ID (or other information) for a protocol/marker paper on this study	PMID: 22735991
Study website	
Purpose or major Objectives of study	To take advantage of full scale Norwegian register based family data and establish comprehensive linkages involving multiple data sources, eg. registers and survey data
Disease areas of focus	Non-communicable diseases and comorbidity with other health states such as mental health

Is your cohort selected for a specific disease (cancer, diabetes) or unselected for disease?	Unselected
Current size of population (and target number of participants)	5,266,270
Participating countries	Norway
Period of enrollment (and is enrollment on-going?)	1960-2011. All Norwegian inhabitants being residence during the period 1960-2011 (time point for last census). Follow-up of death and other endpoints are ongoing.
Demographic characteristics of participants (age range, proportion male/female, national origin, race)	Born 1855-2011. Conditioned on being alive in 1960 or later, male/female 50 %, predominantly Norwegian born caucasian
Major diseases or phenotypes collected to date.	Cardiovascular diseases from hospital discharge data Cause of death registry
Standardized clinical evaluation components measured, if applicable (e.g. height, weight, blood pressure, exercise testing, spirometry, etc.)	Cardiovascular survey data: height, weight, blood pressure, systolic and diastolic blood pressure, cholesterol and triglycerides, daily smoking, coffee consumption, physical activity
Electronic health/medical records or medical administrative data used to collect clinical phenotypes?	Linked to various endpoint registers: Cause of death register, Hospital discharge for cardiovascular diseases, Cancer Register, Birth Register and Disability Register (major mental disorders)
Predominant type of electronic health records (e.g. Epic, Cerner, etc.)	
Other sources of clinical data	
Environmental exposure data being obtained? What sort?	
Other data collected	
Biological specimens collected? What sort?	Blood in Cohort of Norway (CONOR) which is one of the surveys linked to NFLC https://www.fhi.no/en/studies/conor/ and presentation of CONOR: https://academic.oup.com/ije/article/37/3/481/742090
Is there a central biobank?	Yes
DNA samples prepared (or available to be prepared) from each participant?	Yes
Is genotyping being done on some/all participants?	Yes (some)
Is genomic sequencing being done on some/all participants?	
Other molecular analyses performed	
Did participants provide consent regarding sharing of their data outside the initial study investigators?	Yes, consent not limited to specific research group in CONOR. For register data ethical committee needs to approve projects

How are data or specimens from the cohort made available for research? Any limitations on who can access the data (e.g. by country or sector?)	Data server located in Norway
What study information or data are returned to or accessible by participants?	
Follow-up occurring? (years of follow-up). Is recontact possible?	40+
Notes/Comments	The NFLC is one of the largest follow-up of individuals over several decades in their life course. The comprehensive multigenerational, family linkage within the database contributes to large-scale use of various designs for investigating life course determinants.
Questions Relating to Sharing & Collaboration	
May we make the information you provided about your cohort available on an open website?	Yes
Are you willing to share data from your cohort? If so, would you share:	Yes
a) individual data (redacted to protect confidentiality)?	Yes
b) summary data (counts, distributions)?	Yes
c) metadata (descriptive information on data collection methods)?	Yes
d) case report forms and other data collection materials?	Yes
What do you see as the values of sharing?	NFLC is underused in research. With sharing more researchers will learn about the specific potentials of the study which in itself may generate new projects.
What challenges do you anticipate with sharing?	Challenges in communicating a realistic timeframe. Technical challenges. Challenges in communicating specific details about designing substudies. Challenges in futile planning
What specific legal or regulatory barriers in your country or cohort (aside from ensuring confidentiality and appropriate consent) would prevent you from sharing data with other cohort of cohorts investigators, or openly with anyone who requests them?	Individual data owned by the Statistics Norway cannot be exported. This includes family data and sociodemographic and socioeconomic data from registers. However, they can be analysed from abroad using a Norwegian server. If research questions are not covered by the ethics approval for the current NFLC projects, researchers will need a new approval.
What aspects of your cohort are intended for translation to clinical care or population health?	Establishing better evidence on how behavioural determinants of non-communicable diseases have genetic foundation can support prevention efforts and guide future studies on mechanisms for NCD
How might genomic sequencing add to/enhance your study objectives?	Establish novel evidence on transgenerational mechanisms or the impact of genetic factors for life course determinants of non-communicable diseases

Might you be willing to contribute funding or other resources to support international collaboration?	Yes (to some extent for time spent in planning studies)

Norwegian Mother and Child Cohort Study (MoBa)

Questions Relating to Cohort	
Name of study	Norwegian Mother and Child Cohort Study (MoBa)
Principal Investigator/lead	Dr. Per Magnus
Contact email	Camilla.Stoltenberg@fhi.no per.magnus@fhi.no
PubMed ID (or other information) for a protocol/marker paper on this study	PMID: 27063603 (Magnus P et al. Cohort Profile Update: The Norwegian Mother and Child Cohort Study (MoBa). Int J Epidemiol 2016;45:382-8)
Study website	https://www.fhi.no/en/studies/moba/
Purpose or major Objectives of study	To identify causes of diseases among mothers, fathers and children.
Disease areas of focus	All diseases
Is your cohort selected for a specific disease (cancer, diabetes) or unselected for disease?	Unselected
Current size of population (and target number of participants)	284,000 (114,000 children, 95,000 mothers and 75,000 fathers)
Participating countries	Norway
Period of enrollment (and is enrollment on-going?)	1999-2008. This is a closed cohort
Demographic characteristics of participants (age range, proportion male/female, national origin, race)	Present age range for the children is 10 to 18 years. For the mothers, the mean age is 43 years. For the fathers the mean age is 45 years.
Major diseases or phenotypes collected to date.	A wide range of diseases is captured from questionnaire data. For the children, the spectrum of diseases is limited by their relatively young age. Linkage to disease registries will add to the list of diseases
Standardized clinical evaluation components measured, if applicable (e.g. height, weight, blood pressure, exercise testing, spirometry, etc.)	No clinical examination for all, only in subcohorts
Electronic health/medical records or medical administrative data used to collect clinical phenotypes?	Yes, disease registries
Predominant type of electronic health records (e.g. Epic, Cerner, etc.)	
Other sources of clinical data	Questionnaires, national registries

Environmental exposure data being obtained? What sort?	Through measures of contaminants in biological samples and through estimates from place of living
Other data collected	Yes, a large variety of background and exposure data
Biological specimens collected? What sort?	Blood, urine, teeth
Is there a central biobank?	Yes
DNA samples prepared (or available to be prepared) from each participant?	Yes
Is genotyping being done on some/all participants?	Yes, for about 60,000 subjects, but planned for all
Is genomic sequencing being done on some/all participants?	Only for small subsamples
Other molecular analyses performed	Biomarkers from plasma and RNA samples and methylation patterns and telomere length from DNA
Did participants provide consent regarding sharing of their data outside the initial study investigators?	Yes
How are data or specimens from the cohort made available for research? Any limitations on who can access the data (e.g. by country or sector?)	They are made available after approval from the MoBa administration and from the regional committee for ethics. There is no limitation concerning country or sector as long as the research question is within the consent given by the participants
What study information or data are returned to or accessible by participants?	Only general information in newsletters
Follow-up occurring? (years of follow-up). Is recontact possible?	Yes
Notes/Comments	
Questions Relating to Sharing & Collaboration	
May we make the information you provided about your cohort available on an open website?	Yes
Are you willing to share data from your cohort? If so, would you share:	Yes
a) individual data (redacted to protect confidentiality)?	Yes
b) summary data (counts, distributions)?	Yes
c) metadata (descriptive information on data collection methods)?	Yes
d) case report forms and other data collection materials?	Yes

What do you see as the values of sharing?	Potential new knowledge on causes of disease
What challenges do you anticipate with sharing?	Practical and legal issues
What specific legal or regulatory barriers in your country or cohort (aside from ensuring confidentiality and appropriate consent) would prevent you from sharing data with other cohort of cohorts investigators, or openly with anyone who requests them?	Approval from the ethics committee is needed. Data will be available through remote access to a secure server in Norway
What aspects of your cohort are intended for translation to clinical care or population health?	Any relevant new findings on causes or progression of disease
How might genomic sequencing add to/enhance your study objectives?	It would increase the possibility of finding new causes of diseases
Might you be willing to contribute funding or other resources to support international collaboration?	Yes Comment: Would require grants from external sources

Nurses' Health Study (NHS)

Questions Relating to Cohort	
Name of study	Nurses' Health Study (NHS, NCI)
Principal Investigator/lead	Dr. Francine Grodstein
Contact email	fran.grodstein@channing.harvard.edu
PubMed ID (or other information) for a protocol/marker paper on this study	
Study website	www.nurseshealthstudy.org
Purpose or major Objectives of study	A large prospective investigation into the risk factors for major chronic diseases in women. Cancer is a primary focus, but the study has also produced landmark data on cardiovascular disease, diabetes and many other conditions.
Disease areas of focus	mixed
Is your cohort selected for a specific disease (cancer, diabetes) or unselected for disease?	Unselected
Current size of population (and target number of participants)	121,700
Participating countries	US
Period of enrollment (and is enrollment on-going?)	1976-1976, closed cohort

Demographic characteristics of participants (age range, proportion male/female, national origin, race)	In 1976: F, 30-55 y/o, race/ethnicity inclusive, US
Major diseases or phenotypes collected to date.	Very broad - cancer, cardiovascular, respiratory, musculoskeletal, hearing and vision, reproductive, mental health, kidney, neurodegenerative, incontinence, type 2 diabetes, mental health, GI, and other
Standardized clinical evaluation components measured, if applicable (e.g. height, weight, blood pressure, exercise testing, spirometry, etc.)	Questionnaire: weight, height, waist circumference, BMI, blood pressure, diet, employment, telephone assessment of cognitive function (subset of 20,000)
Electronic health/medical records or medical administrative data used to collect clinical phenotypes?	Yes
Predominant type of electronic health records (e.g. Epic, Cerner, etc.)	Medicare Claims available since 2006
Other sources of clinical data	Questionnaires, tumor registries, death registry
Environmental exposure data being obtained? What sort?	Yes, linked residential address (GIS) - air pollution, radon, built environment, etc
Other data collected	Many types of data - diet, vitamin supplements, physical activity, night shifts, work history, medications, SF-36, psychosocial items (eg, social networks, optimism, etc), and many others
Biological specimens collected? What sort?	Blood, saliva, tumor tissue, urine, toenails
Is there a central biobank?	Yes
DNA samples prepared (or available to be prepared) from each participant?	Yes, for subset
Is genotyping being done on some/all participants?	Yes, for subset
Is genomic sequencing being done on some/all participants?	yes, for subset
Other molecular analyses performed	In subsets: gene expression, exome, metabolomics, epigenetics
Did participants provide consent regarding sharing of their data outside the initial study investigators?	Yes
How are data or specimens from the cohort made available for research? Any limitations on who can access the data (e.g. by country or sector?)	Project approved by data access committee
What study information or data are returned to or accessible by participants?	No study results returned to participants

Follow-up occurring? (years of follow-up). Is recontact possible?	follow-up ongoing, 42 years of follow-up to date, recontact continues every two years
Notes/Comments	The Nurses Health Study is one of the largest and longest running investigations of factors that influence women's health. Started in 1976, the information provided by the 121,700 dedicated nurse-participants has led to many new insights on health and disease. While the prevention of cancer is one primary focus, the study has also produced landmark data on cardiovascular disease, diabetes and many other conditions. Most importantly, these studies have shown that diet, physical activity and other lifestyle factors can powerfully promote better health. Purpose: A large prospective investigation into the risk factors for major chronic diseases in women. Cancer is a primary focus, but the study has also produced landmark data on cardiovascular disease, diabetes and many other conditions.
Questions Relating to Sharing & Collaboration	
May we make the information you provided about your cohort available on an open website?	Yes
Are you willing to share data from your cohort? If so, would you share:	Yes
a) individual data (redacted to protect confidentiality)?	Yes
b) summary data (counts, distributions)?	Yes
c) metadata (descriptive information on data collection methods)?	Yes
d) case report forms and other data collection materials?	Yes
What do you see as the values of sharing?	The values are infinite, and range from new scientific discoveries to new avenues of investigation to improved efficiencies in data use
What challenges do you anticipate with sharing?	Minimal
What specific legal or regulatory barriers in your country or cohort (aside from ensuring confidentiality and appropriate consent) would prevent you from sharing data with other cohort of cohorts investigators, or openly with anyone who requests them?	None
What aspects of your cohort are intended for translation to clinical care or population health?	All aspects of cohort are intended for translation to clinical care and population health
How might genomic sequencing add to/enhance your study objectives?	This would tremendously enhance research in the cohort. In particular, the wealth of data throughout the adult lifecourse on behavior, lifestyle, and health -- combined with genomic data -- can yield tremendous opportunities for evaluating interactions of behavioral, genetic, and biologic pathways in disease development and prognosis

Might you be willing to contribute funding or other resources to support international collaboration?	Yes Comment: no clear funds available to contribute, but happy to contribute resources

Nurses' Health Study II (NHSII)

Questions Relating to Cohort	
Name of study	Nurses' Health Study II (NHSII, NCI)
Principal Investigator/lead	Dr. Walter Willett
Contact email	wwillett@hsph.harvard.edu and assistant Debbie Flynn at dosulliv@hsph.harvard.edu
PubMed ID (or other information) for a protocol/marker paper on this study	
Study website	http://www.channing.harvard.edu/nhs/
Purpose or major Objectives of study	A large prospective investigations into the risk factors for major chronic diseases in women. Cancer is a primary focus, but the study has also produced landmark data on cardiovascular disease, diabetes and many other conditions.
Disease areas of focus	Comprehensive
Is your cohort selected for a specific disease (cancer, diabetes) or unselected for disease?	Unselected
Current size of population (and target number of participants)	116,430
Participating countries	US
Period of enrollment (and is enrollment on-going?)	1989-1989
Demographic characteristics of participants (age range, proportion male/female, national origin, race)	F, 25-45 y/o, race/ethnicity inclusive
Major diseases or phenotypes collected to date.	Broad - diabetes, heart and vascular diseases, cancer-related phenotypes
Standardized clinical evaluation components measured, if applicable (e.g. height, weight, blood pressure, exercise testing, spirometry, etc.)	Anthropometry (e.g., weight, height, waist circumference, BMI
Electronic health/medical records or medical administrative data used to collect clinical phenotypes?	Yes
Predominant type of electronic health records (e.g. Epic, Cerner, etc.)	

Other sources of clinical data	Questionnaires, tumor registries
Environmental exposure data being obtained? What sort?	Yes (not specified)
Other data collected	Diet, physical activity, smoking, sleep, shift work, hormone therapy, oral contraceptive use
Biological specimens collected? What sort?	Blood, saliva, tumor tissue, nails,urine. Fecal samples and oral samples for microbiome work are being collected
Is there a central biobank?	Yes
DNA samples prepared (or available to be prepared) from each participant?	Yes, for subset
Is genotyping being done on some/all participants?	Yes, for subset
Is genomic sequencing being done on some/all participants?	
Other molecular analyses performed	
Did participants provide consent regarding sharing of their data outside the initial study investigators?	Yes - type not specified
How are data or specimens from the cohort made available for research? Any limitations on who can access the data (e.g. by country or sector?)	Review by research group or external advisory committee
What study information or data are returned to or accessible by participants?	None
Follow-up occurring? (years of follow-up). Is recontact possible?	Follow-up by questionnaire every two years
Notes/Comments	The Nurses' Health Study (NHS) II, established in 1989, is an expansion of the original NHS (established in 1976). The primary motivation for developing the NHSII was to study oral contraceptives, diet, and lifestyle risk factors in a population younger than the original NHS cohort. The NHSII enrolled 116,430 women aged 25 to 42 years (the upper age was to correspond with the lowest age group in the NHS). Every two years, cohort members receive a follow-up questionnaire with questions about diseases and health-related topics such as smoking, hormone use, pregnancy history, and menopausal status. Food-frequency questionnaires have been administered at four-year intervals since 1991. Blood and urine samples have been collected from approximately 30,000 participants; cheek cell samples have been collected from an additional approximately 30,000 participants. Purpose: A large prospective investigations into the risk factors for major chronic diseases in women. Cancer is a primary focus, but the study has also produced landmark data on cardiovascular disease, diabetes and many other conditions. **Note: "The Health Professional's Follow-up Study of 52,000 men that we conduct does not qualify by numbers, but can be combined with Nurses' Health Study

	II, as we often do in analyses."
Questions Relating to Sharing & Collaboration	
May we make the information you provided about your cohort available on an open website?	Yes
Are you willing to share data from your cohort? If so, would you share:	Yes, we have active data sharing. All forms of data sharing are possible
a) individual data (redacted to protect confidentiality)?	
b) summary data (counts, distributions)?	
c) metadata (descriptive information on data collection methods)?	
d) case report forms and other data collection materials?	
What do you see as the values of sharing?	The main value is to maximize the scientific value of the cohort. Challenges include the constrained resources to manage data sharing
What challenges do you anticipate with sharing?	
What specific legal or regulatory barriers in your country or cohort (aside from ensuring confidentiality and appropriate consent) would prevent you from sharing data with other cohort of cohorts investigators, or openly with anyone who requests them?	None
What aspects of your cohort are intended for translation to clinical care or population health?	All aspects are intended for translation to clinical and population health
How might genomic sequencing add to/enhance your study objectives?	This would add more detailed information than we have from GWAS analyses to integrate with diet and other nongenetic variables in analyses of disease risk
Might you be willing to contribute funding or other resources to support international collaboration?	We are very actively involved in many international collaborations and would welcome more

OHS (Ontario Health Study) and the Canadian Partnership for Tomorrow Project (CPTP)

Questions Relating to Cohort	
Name of study	OHS (Ontario Health Study) and the Canadian Partnership for Tomorrow Project (CPTP)
Principal Investigator/lead	Dr. Philip Awadalla

Contact email	Philip.awadalla@oicr.on.ca
PubMed ID (or other information) for a protocol/marker paper on this study	http://www.partnershipfortomorrow.ca
Study website	https://www.ontariohealthstudy.ca/, http://www.partnershipfortomorrow.ca
Purpose or major Objectives of study	To help researchers better understand the causes of chronic diseases like cancer, heart disease and diabetes, and to develop new ways to prevent and treat them.
Disease areas of focus	Mixed
Is your cohort selected for a specific disease (cancer, diabetes) or unselected for disease?	Unselected
Current size of population (and target number of participants)	315,000
Participating countries	Canada
Period of enrollment (and is enrollment on-going?)	2008-ongoing
Demographic characteristics of participants (age range, proportion male/female, national origin, race)	M/F, 18 + y/o, drawn from Canadian residents including Quebec, Ontario, British Columbia, Alberta, Nova Scotia, Prince Edward Island, New Brunswick, Newfoundland, including new recruitment in Manitoba.
Major diseases or phenotypes collected to date.	Broad - includes cancer, heart disease, diabetes, asthma, Alzheimer's
Standardized clinical evaluation components measured, if applicable (e.g. height, weight, blood pressure, exercise testing, spirometry, etc.)	Anthropometry (e.g., weight, height, BMI), blood pressure, spirometry, CBCs, blood biochemistry (lipids, blood sugar-HbA1c,) grip strength, cytokines etc.
Electronic health/medical records or medical administrative data used to collect clinical phenotypes?	Participants were collected from the population and provide identifiers such that linkages to administrative health record data can be linked allowing for followups and retrospectives
Predominant type of electronic health records (e.g. Epic, Cerner, etc.)	
Other sources of clinical data	Questionnaire, Clinical Site Visits, linkages to administrative health record sites
Environmental exposure data being obtained? What sort?	Yes, air pollution (eig. PM2.5, SO2 etc), water quality, built environment, walkability, noise, food desert, green space etc.
Other data collected	Magnetic Resonance Imaging (MRIs), Recontact Biologics, Followup Medical Questionnaire, Food Frequencies Questionnaires, Residential History and related linkages,
Biological specimens collected? What sort?	Blood and Urine
Is there a central biobank?	There are 4 Central biobanks across Canada
DNA samples prepared (or available to be prepared) from each participant?	DNA samples are prepared for a majority

Is genotyping being done on some/all participants?	Yes active
Is genomic sequencing being done on some/all participants?	Yes active
Other molecular analyses performed	RNAseq, ATACseq and in some cases single cell analyses
Did participants provide consent regarding sharing of their data outside the initial study investigators?	Yes - type not specified
How are data or specimens from the cohort made available for research? Any limitations on who can access the data (e.g. by country or sector?)	Project approved by data access committee
What study information or data are returned to or accessible by participants?	Research ethics board will determine on a case-by-case basis
Follow-up occurring? (years of follow-up). Is recontact possible?	Annual
Notes/Comments	The Canadian Partnership for Tomorrow Project (CPTP) is Canada's largest group of volunteer research participants (population cohort), built to address key questions about what causes cancer and chronic disease. Over 300,000 Canadians aged 30-74 years have joined CPTP; they were recruited from five regional cohorts—BC Generations Project, Alberta's Tomorrow Project, Ontario Health Study, CARTaGENE, and Atlantic PATH.
Questions Relating to Sharing & Collaboration	
May we make the information you provided about your cohort available on an open website?	Yes
Are you willing to share data from your cohort? If so, would you share:	Yes
a) individual data (redacted to protect confidentiality)?	Yes Comment: We have an access approval process
b) summary data (counts, distributions)?	Yes
c) metadata (descriptive information on data collection methods)?	Yes
d) case report forms and other data collection materials?	Comment: not relevant
What do you see as the values of sharing?	
What challenges do you anticipate with sharing?	

What specific legal or regulatory barriers in your country or cohort (aside from ensuring confidentiality and appropriate consent) would prevent you from sharing data with other cohort of cohorts investigators, or openly with anyone who requests them?	
What aspects of your cohort are intended for translation to clinical care or population health?	While we consider the cohort to be critical to the science of prevention medicine and population health, all aspects - including genomics and linkages to admin record, are critical in these translations..
How might genomic sequencing add to/enhance your study objectives?	Genome sequencing would support numerous baseline and followup activities as we pursue longitudinal research in this space. Standard cross-sectional studies could be extended to longitudinal trajectories. Rare (and common) mutations can be tracked with respect to health outcomes. Further, integrated with other functional genomic and molecular data.
Might you be willing to contribute funding or other resources to support international collaboration?	Yes Comment: We could contribute resources to support international funding and scientific initiatives.

PERSIAN Cohort Study

Questions Relating to Cohort	
Name of study	PERSIAN Cohort Study
Principal Investigator/lead	Dr. Reza Malekzadeh, Dr. Hossein Poustchi, Dr. Farin Kamangar, Dr. Arash Etemadi
Contact email	malek@tums.ac.ir, h.poustchi@gmail.com, farinkamangar@gmail.com, arash.etemadi@nih.gov
PubMed ID (or other information) for a protocol/marker paper on this study	PMID: 29145581 (Am J Epidemiol. 2017 Nov 14. doi: 10.1093/aje/kwx314)
Study website	http://persiancohort.com
Purpose or major Objectives of study	To find the incidence, prevalence and risk factors associated with common non-communicable diseases for better prevention and treatment
Disease areas of focus	Non-communicable diseases
Is your cohort selected for a specific disease (cancer, diabetes) or unselected for disease?	Unselected
Current size of population (and target number of participants)	180,000
Participating countries	Iran

Period of enrollment (and is enrollment on-going?)	The PERSIAN Cohort is being conducted in 18 different centers. Since September 2014, enrollment began at the cohort sites at various time points. All sites have currently started enrollment. Seven centers have finished, while it is ongoing at the rest.
Demographic characteristics of participants (age range, proportion male/female, national origin, race)	Iranian male and females, 35-70 years of age, from all the major ethnic groups in Iran are enrolled (Fars, Turk, Lur, Kurd, Balouch, Arab, Tabari, etc). Since enrollment has not yet finished, we do not have proportions at this time. Although ideal proportions are 50/50 for each site, 45% male, 55% female are also acceptable as men tend to participate less in such studies in Iran.
Major diseases or phenotypes collected to date.	All NCDs including cardiovascular diseases, cancers of all types
Standardized clinical evaluation components measured, if applicable (e.g. height, weight, blood pressure, exercise testing, spirometry, etc.)	In all cohort sites: Height, weight, waist, hip and wrist circumferences, blood pressure In some cohort sites: electrocardiograms, bioelectrical impedance analysis, spirometry, eye examinations
Electronic health/medical records or medical administrative data used to collect clinical phenotypes?	Yes—during follow up if necessary
Predominant type of electronic health records (e.g. Epic, Cerner, etc.)	National disease registries (Cancer, Multiple sclerosis, etc), Sepas (hospital medical record system). A nationwide electronic medical record program is also being designed, which will be used in the future.
Other sources of clinical data	No
Environmental exposure data being obtained? What sort?	Yes, occupational exposures, fuel exposure, housing, type of heating used, pesticide use, mobile use, water source, animal contact
Other data collected	Questionnaire Data: Food frequency questionnaire and dietary habits, circadian rhythm, physical activity, past medical history, family history, medication history, oral health, smoking, drug and alcohol use, physical examination on disabilities Laboratory Data: Blood: complete blood count, fasting blood sugar, total cholesterol, HDL cholesterol, triglycerides, alanine transaminase, aspartate transaminase, alkaline phosphatase, gamma- glutamyl transpeptidase, blood urea nitrogen, and creatinine levels Urine: urine pH, specific gravity and the presence of blood, protein, glucose, bilirubin, nitrates, ketones, ascorbic acid, leukocytes, and microalbumin
Biological specimens collected? What sort?	Yes, blood (serum, whole blood, buffy coat, plasma), urine, hair, nail for all participants Stool samples to be collected in a subgroup
Is there a central biobank?	There is a biobank in each cohort site, but not centrally for all 18 sites
DNA samples prepared (or available to be prepared) from each participant?	Yes, buffy coat is prepared and stored for each participant
Is genotyping being done on some/all participants?	Not currently, but it is a future goal if funding becomes available

Is genomic sequencing being done on some/all participants?	Not currently, but it is a future goal if funding becomes available
Other molecular analyses performed	Not currently, but it is a future goal if funding becomes available
Did participants provide consent regarding sharing of their data outside the initial study investigators?	Yes
How are data or specimens from the cohort made available for research? Any limitations on who can access the data (e.g. by country or sector?)	Data and specimens are available to all interested researchers. An online access portal is being designed on the PERSIAN Cohort websites, where proposals can be sent for evaluation by the central committee. Specific guidelines for data/specimen use as well as data/specimen transfer agreements will also be made available on the PERSIAN cohort website in the near future. Studies meeting all requirements will be granted access.
What study information or data are returned to or accessible by participants?	Results of laboratory tests performed are given back to individuals immediately after enrollment. If an individual is found to be a disease carrier or at risk of a disease, they will be informed to follow it up with their primary care physician.
Follow-up occurring? (years of follow-up). Is recontact possible?	Yes, annual follow up is performed for all participants. Currently, follow up information is available for approximately 30,000 individuals. Contact is possible with all participants
Notes/Comments	
Questions Relating to Sharing & Collaboration	
May we make the information you provided about your cohort available on an open website?	Yes
Are you willing to share data from your cohort? If so, would you share:	Yes
a) individual data (redacted to protect confidentiality)?	Yes; subject to proposal submission and approval
b) summary data (counts, distributions)?	Yes
c) metadata (descriptive information on data collection methods)?	Yes
d) case report forms and other data collection materials?	Yes
What do you see as the values of sharing?	Sharing data is very valuable as it adds to the knowledge we are trying to gather, gives us other research ideas and makes evidences found much stronger.
What challenges do you anticipate with sharing?	One important challenge is in the accurate use of data. The PERSIAN Cohort investigators do not agree with any public sharing of data unless it is absolutely clear how the data will be used. This is because when complex data is publicly available, it is very much open to misinterpretation, or use in analyses that are not well-planned. The result may be inaccurate conclusions, sometimes contradictory to what a careful analysis can show. We are open to collaboration requests, and sharing data for well-planned studies, but not with public data sharing.

What specific legal or regulatory barriers in your country or cohort (aside from ensuring confidentiality and appropriate consent) would prevent you from sharing data with other cohort of cohorts investigators, or openly with anyone who requests them?	No barriers
What aspects of your cohort are intended for translation to clinical care or population health?	All data collected, especially those found to be significant risk factors for common diseases, will be translated to preventative programs and/or clinical care.
How might genomic sequencing add to/enhance your study objectives?	It is a future goal of the PERSIAN Cohort to perform genetic studies since they will provide us with valuable data on personalization of medical care.
Might you be willing to contribute funding or other resources to support international collaboration?	Yes

Prostate, Lung, Colorectal and Ovarian Cancer Screening Trial (PLCO)

Questions Relating to Cohort	
Name of study	Prostate, Lung, Colorectal and Ovarian Cancer Screening Trial (PLCO, NCI)
Principal Investigator/lead	Dr. Paul Pinsky and Dr. Neal Freedman
Contact email	pinskyp@mail.nih.gov, freedmanne@mail.nih.gov
PubMed ID (or other information) for a protocol/marker paper on this study	
Study website	http://prevention.cancer.gov/major-programs/prostate-lung-colorectal https://biometry.nci.nih.gov/cdas/studies/plco/
Purpose or major Objectives of study	To determine the effects of screening on cancer-related mortality and secondary endpoints in men and women aged 55 to 74.
Disease areas of focus	Cancer
Is your cohort selected for a specific disease (cancer, diabetes) or unselected for disease?	Unselected, except anyone with history of prostate, lung, colorectal or ovarian cancer was excluded.
Current size of population (and target number of participants)	154,907
Participating countries	US
Period of enrollment (and is enrollment on-going?)	1993-2001

Demographic characteristics of participants (age range, proportion male/female, national origin, race)	M/F, 55-74 y/o, race/ethnicity based on demographic region
Major diseases or phenotypes collected to date.	All cancers. Selected self-reported conditions by questionnaire. Chronic diseases on a subset (about one third) based on Medicare claims data.
Standardized clinical evaluation components measured, if applicable (e.g. height, weight, blood pressure, exercise testing, spirometry, etc.)	For subset (about half): Chest x-rays, flexible sigmoidoscopy, transvaginal ultrasound and CA-125 levels (women), and digital rectal exam and PSA levels (men)
Electronic health/medical records or medical administrative data used to collect clinical phenotypes?	Yes
Predominant type of electronic health records (e.g. Epic, Cerner, etc.)	
Other sources of clinical data	Questionnaires, tumor registries
Environmental exposure data being obtained? What sort?	Geo-coding of addresses to allow for linkage to some environmental exposures
Other data collected	No
Biological specimens collected? What sort?	Blood, buccal cells, lymphocytes, tumor tissue
Is there a central biobank?	Yes
DNA samples prepared (or available to be prepared) from each participant?	Yes, for majority of subjects
Is genotyping being done on some/all participants?	Yes, for majority of subjects
Is genomic sequencing being done on some/all participants?	
Other molecular analyses performed	Many, on various subsets
Did participants provide consent regarding sharing of their data outside the initial study investigators?	Yes - sharing with outside researchers allowed.
How are data or specimens from the cohort made available for research? Any limitations on who can access the data (e.g. by country or sector?)	
What study information or data are returned to or accessible by participants?	Only screening test results performed as part of the PLCO trial
Follow-up occurring? (years of follow-up). Is recontact possible?	Yes, 20+ ; Recontact possible for subset

Notes/Comments	The Prostate, Lung, Colorectal, and Ovarian (PLCO) Cancer Screening Trial is a large-scale clinical trial to determine whether certain cancer screening tests reduce deaths from prostate, lung, colorectal, and ovarian cancer. Sponsored and run by NCI s Division of Cancer Prevention (DCP), in collaboration with the Division of Cancer Epidemiology and Genetics (DCEG), the PLCO Trial is taking place at ten screening centers across the country. Between 1993, when the trial opened, and by 2001, when enrollment ended, 155,000 women and men aged 55-74 who had no previous diagnosis of cancer of the prostate, lung, colon/rectum, or ovary had joined. At entry, participants were randomly divided into two study arms; one received routine health care from their health providers while the other received a series of exams to screen for prostate, lung, colorectal, and ovarian cancers. Screening of participants ended in late 2006. Active follow-up is continuing for up to 14 years as originally scheduled, with an additional follow-up for at least five years being planned. New findings are noted as they are published on the PLCO News Updates and Publications. With more than 2.6 million biological samples collected, PLCO is a rich and unique resource for etiologic and early marker studies. The Etiology and Early Marker Studies (EEMS) program is a critical component of the PLCO Trial. By collecting biologic materials and risk factor information from trial participants before the diagnosis of disease, the EEMS adds substantial value by providing a resource for ancillary research projects investigating cancer etiology and early markers for detection of cancer.
Questions Relating to Sharing & Collaboration	
May we make the information you provided about your cohort available on an open website?	Yes
Are you willing to share data from your cohort? If so, would you share:	Yes
a) individual data (redacted to protect confidentiality)?	Yes; already doing this
b) summary data (counts, distributions)?	Yes; already doing this
c) metadata (descriptive information on data collection methods)?	Yes; already doing this
d) case report forms and other data collection materials?	Yes; already doing this
What do you see as the values of sharing?	Allow outside researchers to explore their own hypotheses and also allow them to do re-analyses of data analyzed and published by PLCO researchers
What challenges do you anticipate with sharing?	We currently have five years of experience with data sharing, which has been summarized in a publication in PLOS Medicine; we can share this publication at the meeting.
What specific legal or regulatory barriers in your country or cohort (aside from ensuring confidentiality and appropriate consent) would prevent you from sharing data with other cohort of cohorts	None at the moment. There are some potential issues with sharing linked data obtained from U.S. state cancer registries; such data are currently just being received.

investigators, or openly with anyone who requests them?	
What aspects of your cohort are intended for translation to clinical care or population health?	The screening trial aspect; also, studies on cancer etiology
How might genomic sequencing add to/enhance your study objectives?	Allow further analyses of genetic risk of cancer and other conditions
Might you be willing to contribute funding or other resources to support international collaboration?	Yes

Saudi Human Genome Program

Questions Relating to Cohort	
Name of study	Saudi Human Genome Program
Principal Investigator/lead	Dr. Fowzan S Alkuraya
Contact email	falkuraya@kfshrc.edu.sa
PubMed ID (or other information) for a protocol/marker paper on this study	
Study website	http://shgp.kacst.edu.sa/site/
Purpose or major Objectives of study	
Disease areas of focus	
Is your cohort selected for a specific disease (cancer, diabetes) or unselected for disease?	No; skewed towards individuals with Mendelian diseases
Current size of population (and target number of participants)	100,000
Participating countries	Saudi Arabia
Period of enrollment (and is enrollment on-going?)	
Demographic characteristics of participants (age range, proportion male/female, national origin, race)	
Major diseases or phenotypes collected to date.	

Standardized clinical evaluation components measured, if applicable (e.g. height, weight, blood pressure, exercise testing, spirometry, etc.)	
Electronic health/medical records or medical administrative data used to collect clinical phenotypes?	
Predominant type of electronic health records (e.g. Epic, Cerner, etc.)	
Other sources of clinical data	
Environmental exposure data being obtained? What sort?	NA
Other data collected	
Biological specimens collected? What sort?	Currently the standard protocol is blood. However, the goal is to establish a tissue Biobank in the near future.
Is there a central biobank?	The Biobank facility is located in Riyadh, the capital of Saudi Arabia.
DNA samples prepared (or available to be prepared) from each participant?	DNA Samples are available for analysis.
Is genotyping being done on some/all participants?	We aim to conduct genotyping once we have 5000 participants.
Is genomic sequencing being done on some/all participants?	
Other molecular analyses performed	Will be done based on research questions.
Did participants provide consent regarding sharing of their data outside the initial study investigators?	Yes
How are data or specimens from the cohort made available for research? Any limitations on who can access the data (e.g. by country or sector?)	A research proposal should be submitted to obtain the approval of the institute's IRB and the Biobank director.
What study information or data are returned to or accessible by participants?	None as for now.
Follow-up occurring? (years of follow-up). Is recontact possible?	The goals is for this cohort to be a prospective study. However, the interval for follow-up visits have not been defined yet.
Notes/Comments	The Saudi Bio-bank is a longitudinal study of the combined effects of genes, environment, and lifestyle on common diseases of adult life. The Saudi bio-bank plans to recruit a group of 200,000 subjects that form the population of the catchments areas of NGHA housing and clinics. Of these patients, 100,000 will be a sample from a prospective family-based study group, and 100,000 will be from a disease-specific bio-banking group. The diseases represented in the latter group will include diabetes, cancer, coronary artery disease, hepatitis, obesity, bronchial asthma, chronic renal impairment and failure, stroke and

	more.
Questions Relating to Sharing & Collaboration	
May we make the information you provided about your cohort available on an open website?	Yes
Are you willing to share data from your cohort? If so, would you share:	Sharing of data should follow the above stated process.
a) individual data (redacted to protect confidentiality)?	
b) summary data (counts, distributions)?	
c) metadata (descriptive information on data collection methods)?	
d) case report forms and other data collection materials?	
What do you see as the values of sharing?	Sharing would enrich the population available for research. This will increase the study power. In addition, sharing would provide the opportunity to compare different populations to each other.
What challenges do you anticipate with sharing?	Obtaining the approval of the institute's IRB.
What specific legal or regulatory barriers in your country or cohort (aside from ensuring confidentiality and appropriate consent) would prevent you from sharing data with other cohort of cohorts investigators, or openly with anyone who requests them?	The institute's IRB.
What aspects of your cohort are intended for translation to clinical care or population health?	All aspects can be used for translation to clinical care or population health if applicable.
How might genomic sequencing add to/enhance your study objectives?	It advance our understanding of the health problems that affect the Saudi population.
Might you be willing to contribute funding or other resources to support international collaboration?	This depends on the annual budget and higher management. That being said, the institute encourages international collaborations.

Saudi National Biobank

Questions Relating to Cohort	
Name of study	Saudi National Biobank
Principal Investigator/lead	Dr. Ada Al-Qunaibet
Contact email	alqunaibetad@ngha.med.sa

PubMed ID (or other information) for a protocol/marker paper on this study	Has not been published yet.
Study website	http://www.ksau-hs.edu.sa/English/Research/Pages/KAIMRC.aspx
Purpose or major Objectives of study	1. To enable to quantify disease incidences in various populations and subpopulations, 2. To enable to understand natural histories and risk factors for these diseases including genome-environment interactions.
Disease areas of focus	As the proposal stands the disease areas of focus are: Cancer, Diabetes, Coronary arteries Diseases, Hepatitis, obesity, Bronchial Asthma, Chronic renal impairment, and Stroke.
Is your cohort selected for a specific disease (cancer, diabetes) or unselected for disease?	
Current size of population (and target number of participants)	2,000
Participating countries	Saudi Arabia
Period of enrollment (and is enrollment on-going?)	Enrollment is on-going. The plans is for it to continue until a total of 200,000 participants are recruited.
Demographic characteristics of participants (age range, proportion male/female, national origin, race)	We will be doing initial descriptive analysis. We can provided once completed.
Major diseases or phenotypes collected to date.	We will be doing initial descriptive analysis. We can provided once completed.
Standardized clinical evaluation components measured, if applicable (e.g. height, weight, blood pressure, exercise testing, spirometry, etc.)	Standardized Clinical Evaluation Components Measured include: height, weight, blood pressure, waist circumference, hip circumference, body fat percentage, peak flow (for participants with respiratory disease).
Electronic health/medical records or medical administrative data used to collect clinical phenotypes?	All enrolled participants have active Electronic Health records. Thus, collected data and specimen analysis will be merged with data from Electronic Health records.
Predominant type of electronic health records (e.g. Epic, Cerner, etc.)	
Other sources of clinical data	
Environmental exposure data being obtained? What sort?	N/A
Other data collected	
Biological specimens collected? What sort?	Currently the standard protocol is blood. However, the goal is to establish a tissue Biobank in the near future.
Is there a central biobank?	The Biobank facility is located in Riyadh, the capital of Saudi Arabia.
DNA samples prepared (or available to be prepared) from each participant?	DNA Samples are available for analysis.

Is genotyping being done on some/all participants?	We aim to conduct genotyping once we have 5000 participants.
Is genomic sequencing being done on some/all participants?	
Other molecular analyses performed	Will be done based on research questions.
Did participants provide consent regarding sharing of their data outside the initial study investigators?	Yes
How are data or specimens from the cohort made available for research? Any limitations on who can access the data (e.g. by country or sector?)	A research proposal should be submitted to obtain the approval of the institute's IRB and the Biobank director.
What study information or data are returned to or accessible by participants?	None as for now.
Follow-up occurring? (years of follow-up). Is recontact possible?	The goals is for this cohort to be a prospective study. However, the interval for follow-up visits have not been defined yet.
Notes/Comments	The Saudi Bio-bank is a longitudinal study of the combined effects of genes, environment, and lifestyle on common diseases of adult life. The Saudi bio-bank plans to recruit a group of 200,000 subjects that form the population of the catchments areas of NGHA housing and clinics. Of these patients, 100,000 will be a sample from a prospective family-based study group, and 100,000 will be from a disease-specific bio-banking group. The diseases represented in the latter group will include diabetes, cancer, coronary artery disease, hepatitis, obesity, bronchial asthma, chronic renal impairment and failure, stroke and more.
Questions Relating to Sharing & Collaboration	
May we make the information you provided about your cohort available on an open website?	Yes
Are you willing to share data from your cohort? If so, would you share:	Sharing of data should follow the above stated process.
a) individual data (redacted to protect confidentiality)?	
b) summary data (counts, distributions)?	
c) metadata (descriptive information on data collection methods)?	
d) case report forms and other data collection materials?	
What do you see as the values of sharing?	Sharing would enrich the population available for research. This will increase the study power. In addition, sharing would provide the opportunity to compare different populations to each other.
What challenges do you anticipate with sharing?	Obtaining the approval of the institute's IRB.

What specific legal or regulatory barriers in your country or cohort (aside from ensuring confidentiality and appropriate consent) would prevent you from sharing data with other cohort of cohorts investigators, or openly with anyone who requests them?	The institute's IRB.
What aspects of your cohort are intended for translation to clinical care or population health?	All aspects can be used for translation to clinical care or population health if applicable.
How might genomic sequencing add to/enhance your study objectives?	It advance our understanding of the health problems that affect the Saudi population.
Might you be willing to contribute funding or other resources to support international collaboration?	This depends on the annual budget and higher management. That being said, the institute encourages international collaborations.

Singapore National Precision Medicine Program

Questions Relating to Cohort	
Name of study	Singapore National Precision Medicine Program
Principal Investigator/lead	Prof. Patrick Tan and Prof. E Shyong Tai
Contact email	patrick_tan@a-star.edu.sg OR mdctes@nus.edu.sg
PubMed ID (or other information) for a protocol/marker paper on this study	None at present. To be prepared in 2018.
Study website	In preparation
Purpose or major Objectives of study	Establish genomic-clinical data vault for the nation of Singapore
Disease areas of focus	Infectious diseases Diabetes mellitus and related metabolic/ endocrine disorders Cancers Cardiovascular diseases Neurological and sense disorders
Is your cohort selected for a specific disease (cancer, diabetes) or unselected for disease?	Combination of healthy controls and disease cohorts
Current size of population (and target number of participants)	10,000 with extension to 1,000,000 over three phases
Participating countries	Singapore
Period of enrollment (and is enrollment on-going?)	10 year program

Demographic characteristics of participants (age range, proportion male/female, national origin, race)	Reflective of Singapore population Ethnicity primarily Chinese, Malay, Indian
Major diseases or phenotypes collected to date.	See above
Standardized clinical evaluation components measured, if applicable (e.g. height, weight, blood pressure, exercise testing, spirometry, etc.)	Real-time linkage to all national electronic health records with plans to integrate environmental, behavioural and exposure information through Smart Nation initiative
Electronic health/medical records or medical administrative data used to collect clinical phenotypes?	See above
Predominant type of electronic health records (e.g. Epic, Cerner, etc.)	
Other sources of clinical data	All public funded cohort data will be consolidated in the national database
Environmental exposure data being obtained? What sort?	See above
Other data collected	See above
Biological specimens collected? What sort?	Initial 10,000 already collected
Is there a central biobank?	Leveraging on existing biobanks at academic medical centres, supported by national funding
DNA samples prepared (or available to be prepared) from each participant?	Initial 10,000 available
Is genotyping being done on some/all participants?	Already done for some
Is genomic sequencing being done on some/all participants?	
Other molecular analyses performed	Specific cohorts already have epigenetic, imaging, lipidomic data available
Did participants provide consent regarding sharing of their data outside the initial study investigators?	Yes for initial 10,000
How are data or specimens from the cohort made available for research? Any limitations on who can access the data (e.g. by country or sector?)	For specimens, sequencing to be done in Singapore For data, Ministry of Health IT group will be developing an access platform
What study information or data are returned to or accessible by participants?	Under discussion by National Precision Medicine Clinical Adoption workgroup and Regulation and Ethics workgroup
Follow-up occurring? (years of follow-up). Is recontact possible?	In perpetuity. Standardized consent process for recontact being developed by Public and Community Trust workgroup

Notes/Comments	Specific focus in early phases on defining good use-cases to show cost effectiveness and clinical value of population-based precision medicine efforts (in collaboration with MOH Chief Health Scientist's office)
Questions Relating to Sharing & Collaboration	
May we make the information you provided about your cohort available on an open website?	Yes in principle. However this should be cleared with Public and Community Trust workgroup.
Are you willing to share data from your cohort? If so, would you share:	National policies for data sharing are in progress. Open to discussion.
a) individual data (redacted to protect confidentiality)?	
b) summary data (counts, distributions)?	
c) metadata (descriptive information on data collection methods)?	
d) case report forms and other data collection materials?	
What do you see as the values of sharing?	Value : Access to variant allele frequencies across populations for accurate diagnosis.
What challenges do you anticipate with sharing?	Challenge : Correct attribution to data depositors and tracking of data usage
What specific legal or regulatory barriers in your country or cohort (aside from ensuring confidentiality and appropriate consent) would prevent you from sharing data with other cohort of cohorts investigators, or openly with anyone who requests them?	To be discussed
What aspects of your cohort are intended for translation to clinical care or population health?	Entire cohort to be representative of the Singapore nation
How might genomic sequencing add to/enhance your study objectives?	Quantify baseline genetic risk of Singapore population for MOH-prioritized therapeutic areas
Might you be willing to contribute funding or other resources to support international collaboration?	Currently in discussions with several countries and agencies.

Taiwan Biobank

Questions Relating to Cohort	
Name of study	Taiwan Biobank
Principal Investigator/lead	Dr. Chen-Yang Shen
Contact email	bmcys@ibms.sinica.edu.tw

PubMed ID (or other information) for a protocol/marker paper on this study	
Study website	https://www.twbiobank.org.tw/new_web/
Purpose or major Objectives of study	To improve the health promotion, chronic disease prevention and the prognosis and treatment of diseases.
Disease areas of focus	chronic diseases
Is your cohort selected for a specific disease (cancer, diabetes) or unselected for disease?	Yes. There are two parts of our cohort. The first one is individuals from the community, and the second part is the patients of common chronic diseases, including cancer, DM, stroke, endometriosis and chronic kidney diseases.
Current size of population (and target number of participants)	92,371 (200,000), and 700 patients of the common chronic disease, and our goal is 100,000 common chronic diseases.
Participating countries	Taiwan
Period of enrollment (and is enrollment on-going?)	2012 - up to date (on-going)
Demographic characteristics of participants (age range, proportion male/female, national origin, race)	35.89% M/ 64.11% F; 50-59 y/o with no history of cancer.
Major diseases or phenotypes collected to date.	Broad - includes stroke, hypertension, hyper-lipedema, cardiovascular disease, arrhythmia, coronary artery disease, valve heart disease, asthma, emphysema or bronchitis, gout, arthritis, osteoporosis, allergy, diabetes (the community part). For the second part, i.e. patients-based cohort, we have collected specimens from cancer, cardiovascular diseases, DM, etc.
Standardized clinical evaluation components measured, if applicable (e.g. height, weight, blood pressure, exercise testing, spirometry, etc.)	height, weight, body fat, waist and hip circumference, blood pressure, pulse, bone density, lung function, bone density, routine biochemical tests.
Electronic health/medical records or medical administrative data used to collect clinical phenotypes?	Yes
Predominant type of electronic health records (e.g. Epic, Cerner, etc.)	No
Other sources of clinical data	National Health Insurance Research Database (NHIRD)
Environmental exposure data being obtained? What sort?	No
Other data collected	basic information, lifestyle, diet, exposed environments, reproductive history, medical history, family medical history.
Biological specimens collected? What sort?	Urine, blood
Is there a central biobank?	Yes
DNA samples prepared (or available to be prepared) from each participant?	Yes

Is genotyping being done on some/all participants?	Yes- some participants (SNP chip: 653,291 SNPs) (24,000 cases)
Is genomic sequencing being done on some/all participants?	Yes- some participants (1500 cases)
Other molecular analyses performed	DNA methylation (1500 cases), HLA typing (1100 cases), Blood metabolome (390 cases), Urine metabolome (390 cases)
Did participants provide consent regarding sharing of their data outside the initial study investigators?	Yes.
How are data or specimens from the cohort made available for research? Any limitations on who can access the data (e.g. by country or sector?)	Projects have to be approved by Ethical Governance Council
What study information or data are returned to or accessible by participants?	Only incidental findings thought to be serious during the baseline assessment.
Follow-up occurring? (years of follow-up). Is recontact possible?	3+; yes
Notes/Comments	The Taiwan Biobank is a scientific infrastructure accessible to biomedical researchers aimed at furthering understanding of the relationships between environmental exposure, diet, genetics, and the aetiology and progression of chronic disease. Through the recruitment and follow-up of a cohort of 200,000 individuals from the general population with no history of cancer, the Taiwan Biobank aims to improve the health of future generations and facilitate genomic/epigenomic research in the post-genomic era.
Questions Relating to Sharing & Collaboration	
May we make the information you provided about your cohort available on an open website?	Yes
Are you willing to share data from your cohort? If so, would you share:	Yes
a) individual data (redacted to protect confidentiality)?	Yes
b) summary data (counts, distributions)?	Yes
c) metadata (descriptive information on data collection methods)?	Yes
d) case report forms and other data collection materials?	Yes
What do you see as the values of sharing?	The data sharing from other cohorts will give the heterogeneity of the data and the varying level of the comparisons should shed light on public health promotion and disease prevention.
What challenges do you anticipate with sharing?	It is likely that some of these analyses will require subsequent reanalysis, in order to compare among cohorts. Furthermore, the transfer of large data will become an issue.

What specific legal or regulatory barriers in your country or cohort (aside from ensuring confidentiality and appropriate consent) would prevent you from sharing data with other cohort of cohorts investigators, or openly with anyone who requests them?	According to the Human Biobank Management Act (Taiwan), international data transfer need to be reviewed and approved by specific committee of the Ministry of Health and Welfare of our government.
What aspects of your cohort are intended for translation to clinical care or population health?	The information and specimens contained in the Taiwan Biobank have been made publicly available. The goal is to develop personalised and precision medicine in which progressive elucidation of risk factors and molecular pathogenesis of disease will improve disease prevention and facilitate therapy development for individuals and generations to come.
How might genomic sequencing add to/enhance your study objectives?	The whole genome sequencing currently available would allow detection of large effects of rare mutations on a variety of health conditions. In addition, we have completed 24,000 whole-genome genotypings, which would be very helpful to identify disease susceptibility loci.
Might you be willing to contribute funding or other resources to support international collaboration?	No Comment: No funds available

Tohoku Medical Megabank Project

[Note: information for this cohort was drawn from publicly available sources and has not been verified]

Questions Relating to Cohort	
Name of study	Tohoku Medical Megabank Project
Principal Investigator/lead	Dr. Masi Yamamoto
Contact email	masiyamamoto@med.tohoku.ac.jp
PubMed ID (or other information) for a protocol/marker paper on this study	PMID: 27374138 Kuriyama S et al., The Tohoku Medical Megabank Project: Design and Mission. J Epidemiol 26, 493-511 (2016) / Epub 2016 Jul 2
Study website	http://www.megabank.tohoku.ac.jp/english/ http://iwate-megabank.org/en/
Purpose or major Objectives of study	To develop a biobank that combines medical and genome information during the process of rebuilding the community medical system and supporting health and welfare in the Tohoku area.
Disease areas of focus	Mixed (natural disaster focus)
Is your cohort selected for a specific disease (cancer, diabetes) or unselected for disease?	
Current size of population (and target number of participants)	>142,000 (target: 150,000)
Participating countries	Japan - 1 region in Japan (Tohoku)
Period of enrollment (and is enrollment on-going?)	2013-ongoing

Demographic characteristics of participants (age range, proportion male/female, national origin, race)	M/F, 1. community-based cohort ≥ 20 years; 2. Birth and three generation cohort families, of Japanese descent
Major diseases or phenotypes collected to date.	Gynecological phenotypes
Standardized clinical evaluation components measured, if applicable (e.g. height, weight, blood pressure, exercise testing, spirometry, etc.)	Height, weight, blood tests (routine biochemical analysis, CBC, etc.), urine tests (routine biochemical analysis)
Electronic health/medical records or medical administrative data used to collect clinical phenotypes?	Yes
Predominant type of electronic health records (e.g. Epic, Cerner, etc.)	
Other sources of clinical data	Questionnaires, disease registries (not specified), medical insurance records
Environmental exposure data being obtained? What sort?	No
Other data collected	Lifestyle, nutrition, SES, medical history
Biological specimens collected? What sort?	All participants - DNA, plasma, serum, buffy coat, and mononuclear cells from peripheral blood, urine, (saliva; for people who cannot provide blood for some reason) Newborns - DNA, plasma, serum, buffy coat, and mononuclear cells from cord blood Pregnant women - breast milk
Is there a central biobank?	Yes
DNA samples prepared (or available to be prepared) from each participant?	DNA samples from 63,000 participants; buffy coat samples from 142,000 participants
Is genotyping being done on some/all participants?	Yes, for subset; 3,000 WGS
Is genomic sequencing being done on some/all participants?	
Other molecular analyses performed	Omics (DNA methylation, gene expression, protein and low-molecular metabolites using NMR, mass spectrometer)
Did participants provide consent regarding sharing of their data outside the initial study investigators?	Yes - for research purpose
How are data or specimens from the cohort made available for research? Any limitations on who can access the data (e.g. by country or sector?)	Project approved by data access committee
What study information or data are returned to or accessible by participants?	Health check-up data, mental health data, nutrition data; (in preparation) A part of genome analysis data for limited participants

Follow-up occurring? (years of follow-up). Is recontact possible?	
Notes/Comments	To develop a biobank that combines medical and genome information during the process of rebuilding the community medical system and supporting health and welfare in the Tohoku area.

UK Biobank

Questions Relating to Cohort	
Name of study	UK Biobank
Principal Investigator/lead	Prof. Rory Collins
Contact email	ukbiobank@ukbiobank.ac.uk; access@ukbiobank.ac.uk
PubMed ID (or other information) for a protocol/marker paper on this study	
Study website	http://www.ukbiobank.ac.uk/
Purpose or major Objectives of study	To improve the prevention, diagnosis, and treatment of a wide range of diseases.
Disease areas of focus	Mixed
Is your cohort selected for a specific disease (cancer, diabetes) or unselected for disease?	Unselected
Current size of population (and target number of participants)	502,713 (target: 500,000)
Participating countries	England, Scotland, Wales
Period of enrollment (and is enrollment on-going?)	2006-2010 (completed)
Demographic characteristics of participants (age range, proportion male/female, national origin, race)	46% M/ 54% F; 49-69 y/o; UK residents; 94% White, 2% Asian/Asian British, 2% Black/Black British, 1% Other
Major diseases or phenotypes collected to date.	Broad - includes cardiovascular, respiratory, musculoskeletal, hearing and vision, reproductive, cancer, mental health, cognitive function
Standardized clinical evaluation components measured, if applicable (e.g. height, weight, blood pressure, exercise testing, spirometry, etc.)	Height (standing and sitting), Weight, Body impedance, Blood pressure, Hand grip strength, Bone-densitometry of heel, Spirometry (full cohort) Exercise test, Arterial stiffness, Eye measures (100-200,000 person subsets) Web questionnaires: Diet, Employment, Cognitive function, Modd, Gastrointestinal symptoms (150,000 person subsets) Imaging: Abdominal MRI, Brain MRI, Heart MRI, Carotid ultrasound, DXA assessment, ECG at rest (12-lead) (for 100,000 person subset)

Electronic health/medical records or medical administrative data used to collect clinical phenotypes?	Yes (hospital and primary care)
Predominant type of electronic health records (e.g. Epic, Cerner, etc.)	
Other sources of clinical data	Death registry, cancer registries
Environmental exposure data being obtained? What sort?	Residential air pollution, Residential noise pollution, UK Biobank Urban Morphometric Platform (https://biobank.ctsu.ox.ac.uk/crystal/label.cgi?id=100115)
Other data collected	Physical activity, diet, cognitive function, work history
Biological specimens collected? What sort?	Blood, urine, saliva
Is there a central biobank?	Yes
DNA samples prepared (or available to be prepared) from each participant?	Yes
Is genotyping being done on some/all participants?	Yes - all participants (SNP Chip; 820,000 SNPs)
Is genomic sequencing being done on some/all participants?	Yes – exome sequencing all participants ongoing
Other molecular analyses performed	Biomarker panel
Did participants provide consent regarding sharing of their data outside the initial study investigators?	Broad (for any type of health-related research)
How are data or specimens from the cohort made available for research? Any limitations on who can access the data (e.g. by country or sector?)	Project approved by data access committee
What study information or data are returned to or accessible by participants?	Only incidental findings thought to be serious during the baseline assessment or the ongoing imaging visits. No feedback of any subsequent results from analyses of samples or data
Follow-up occurring? (years of follow-up). Is recontact possible?	10+; yes (provided it does not constitute feedback of information not already known by participant)
Notes/Comments	UK Biobank recruited 500,000 people aged between 40-69 years in 2006-2010 from across the country to take part in this project. UK Biobank is a major long-term initiative to produce a national database that will improve the prevention, diagnosis and treatment of many serious illnesses. The prospective nature of the UK Biobank study is important because the effects of risk factors can be assessed before a disease or its treatment affects a participant. The study design also allows for a wide range of conditions to be investigated, including those that are difficult if not impossible to study retrospectively (for example, dementia and rapidly fatal conditions such as pancreatic or lung cancer). Moreover, both the beneficial and adverse effects of a specific factor on the risk of

	developing disease can be considered simultaneously, and this information can then be used to provide evidence-based public health guidance for primary prevention. Prospective studies must, however, involve large numbers of participants because only a relatively small proportion of the cohort will develop any particular condition, and the effect of any one risk factor on overall disease risk is likely to be small.
Questions Relating to Sharing & Collaboration	
May we make the information you provided about your cohort available on an open website?	Yes
Are you willing to share data from your cohort? If so, would you share:	Yes
a) individual data (redacted to protect confidentiality)?	Yes Comment: It is already accessible to all registered researchers for any type of health research, subject to approval of an application by UK Biobank's Access Sub-Committee
b) summary data (counts, distributions)?	Yes
c) metadata (descriptive information on data collection methods)?	Yes
d) case report forms and other data collection materials?	Yes
What do you see as the values of sharing?	Increased use of the data for an increasingly wide and imaginative set of purposes (albeit, given the complexity of the data and the varying level and focus of the expertise of the people using the data, it is likely that some of these analyses will be flawed and require subsequent correction by other researchers re-analysing the data appropriately).
What challenges do you anticipate with sharing?	The much larger scale of sequence data is likely to require a change to the way in which we have, up to now, been sharing data since it may well not be possible to send the data to the researchers (as we have been doing with the genotyping data) and may instead need to establish systems that allow researchers to go to the data and carry out their analyses
What specific legal or regulatory barriers in your country or cohort (aside from ensuring confidentiality and appropriate consent) would prevent you from sharing data with other cohort of cohorts investigators, or openly with anyone who requests them?	None
What aspects of your cohort are intended for translation to clinical care or population health?	All
How might genomic sequencing add to/enhance your study objectives?	The addition of whole genome sequencing to the currently available genotype data (and, when they become available over the next few years, exome sequence data) would allow detection of large effects of rare mutations (particularly those located outside the exome close to previously identified loci from GWAS analyses) on a variety of health conditions

Might you be willing to contribute funding or other resources to support international collaboration?	No Comment: No funds available to do so

UKCTOCS (UK Collaborative Trial of Ovarian Cancer Screening) Longitudinal Women's Cohort – (UKLWC)

Questions Relating to Cohort	
Name of study	UKCTOCS (UK Collaborative Trial of Ovarian Cancer Screening) Longitudinal Women's Cohort – (UKLWC)
Principal Investigator/lead	Prof. Usha Menon
Contact email	u.menon@ucl.ac.uk
PubMed ID (or other information) for a protocol/marker paper on this study	Design: https://www.ncbi.nlm.nih.gov/pubmed/19008269 Follow up: https://www.ncbi.nlm.nih.gov/pubmed/26707054 Biobanking details: https://www.ncbi.nlm.nih.gov/pubmed/25964255 Examples of -omics studies in secondary studies Epidemiological risk factors available: https://www.ncbi.nlm.nih.gov/pubmed/28264823 Genomics in nested sub-sample: https://www.ncbi.nlm.nih.gov/pubmed/28500271 Epigenomics in nested sub-sample: https://www.ncbi.nlm.nih.gov/pubmed/29268762 Proteomics in nested sub-sample: https://www.ncbi.nlm.nih.gov/pubmed/26815306 Metabolomics in nested sub-sample: https://www.ncbi.nlm.nih.gov/pubmed/29237687
Study website	http://www.isrctn.com/ISRCTN22488978 http://www.ucl.ac.uk/womens-health/research/womens-cancer/gynaecological-cancer-research-centre/ukctocs
Purpose or major Objectives of study	Create an 'open research' platform for the study of common disease and delivery of precision medicine, specifically: (i) Identification of new drug targets for common conditions to enable drug development, repositioning and indication expansion (ii) Discovery of novel risk prediction, screening, diagnostic and prognostic disease markers.
Disease areas of focus	All diseases affecting post-menopausal women
Is your cohort selected for a specific disease (cancer, diabetes) or unselected for disease?	Unselected. Participants were recruited from the general population between 2001-2005 (16.3% acceptance rate), and at the time were cancer-free.
Current size of population (and target number of participants)	202,638
Participating countries	England, Wales, Northern Ireland
Period of enrollment (and is enrollment on-going?)	2001-2005 (completed)

<table>
<tr><td>Demographic characteristics of participants (age range, proportion male/female, national origin, race)</td><td>All female post-menopausal; 50-74 years (at recruitment); UK residents; 96.5% White, 1.4% Black, 0.9% Asian, 0.8% Other, 0.5% No info</td></tr>
<tr><td>Major diseases or phenotypes collected to date.</td><td>Broad: based on Electronic Health Records (EHR) and questionnaire data, the following “research ready” disease phenotypes are available: cancer (ovarian, endometrial, breast, pancreatic, lung, colorectal) and cardiovascular (acute coronary syndrome components, and stroke including sub-types).
Through linkage with secondary care electronic health records (in-patient, outpatient and A&E), death and cancer registries, we have the capacity to recreate a wider number of health-related phenotypes following the work done by Data Lab team at the Institute of Health Informatics UCL (http://www.ucl.ac.uk/health-informatics/caliber). Table describes the total number of incident cases, with serum samples available, by major ICD-10 chapter.
<table>
<tr><th>ICD10 Disease type</th><th>Total</th></tr>
<tr><td>Malignant Neoplasms (ICD10 - C)</td><td>24001</td></tr>
<tr><td>Nervous System (ICD10 - G)</td><td>20649</td></tr>
<tr><td>Circulatory System (ICD10 - I)</td><td>73037</td></tr>
<tr><td>Respiratory System (ICD10 - J)</td><td>33786</td></tr>
<tr><td>Digestive System (ICD10 - K)</td><td>66801</td></tr>
<tr><td>Musculoskeletal System (ICD10 - M)</td><td>62143</td></tr>
<tr><td>Genitourinary System (ICD10 - M)</td><td>42285</td></tr>
</table>
</td></tr>
<tr><td>Standardized clinical evaluation components measured, if applicable (e.g. height, weight, blood pressure, exercise testing, spirometry, etc.)</td><td>Self–reported health questionnaires:
Data at baseline (202,638 women)
Two follow-up waves: 1st 2005-2009 (144,450 responses); and 2nd in 2014 (83,529 responses).
https://www.ucl.ac.uk/womens-health/research/womens-cancer/gynaecological-cancer-research-centre/ukctocs

Collectively the questionnaires hold information on:
Demographic characteristics, lifestyle parameters (smoking, alcohol consumption), education, height/ weight/ skirt size, reproductive history, HRT management, medical history (non-cancer diseases; cancer(s) diagnoses; gynaecological or breast operations), outlook on life.</td></tr>
<tr><td>Electronic health/medical records or medical administrative data used to collect clinical phenotypes?</td><td>Yes: Currently linked to the National Cancer Registries (NHS Digital and Northern Ireland), Hospital Episodes Statistics for both England and Wales (HES and PEDW), and Death Certificates (NHS Digital and Northern Ireland Death Registry). All above EHR sources are available from 2000 to 2017. In addition, data from national registries is available for Cancer (NCIN, 2000-2014) and acute coronary syndrome (Myocardial Ischaemia National Audit Project, 2003-2010).

Aim to link to primary care records (via CPRD, EMIS and Vision) and national disease registries (e.g. NICOR) but currently not funded.</td></tr>
<tr><td>Predominant type of electronic health records (e.g. Epic, Cerner, etc.)</td><td>The current available EHR (HES and the registries (death and cancers) are national resources that use ICD (International Classification of Diseases) coding.

Currently, we do not have access to within hospital EHR but have consent to retrieve copies of medical records.</td></tr>
<tr><td>Other sources of clinical data</td><td>1. Serial annual CA125 serum assays (~350,000) in 50,640 women (median 9 annual samples)</td></tr>
</table>

	2. Detailed reports of 370,419 pelvic ultrasounds including DICOM imaging data from ~260,000 exams in 50,639 women (median 9 scans).
Environmental exposure data being obtained? What sort?	No - currently not available.
Other data collected	Available in nested cohorts: 1. Medical records: Diagnostic reports, treatment modalities for ovarian (1,323 cases) and endometrial (1,492 cases) cancers diagnosed till 2014; Physician reported stage, histology and treatment for 1,690 breast, 850 lung, 200 pancreatic and 500 colorectal cancer cases 2. Detailed information on menopause and its management (10,607 women) 3. Ovarian cancer symptoms (approx. 50,000 women) 4. Quality-of-life questionnaires (689 women with endometrial cancer) Details can be obtained at: https://www.ucl.ac.uk/womens-health/research/womens-cancer/gynaecological-cancer-research-centre/ukctocs
Biological specimens collected? What sort?	Serum (> 544,000 serum samples collected/stored; baseline samples from 189,642 participants. Unique aspect is availability of serial annual samples (median 9) in 50,262 participants
Is there a central biobank?	Yes, all serum samples were collected using a standard protocol and are stored in 10 x 500µL barcoded straws in liquid nitrogen at the commercial HTA licensed, ISO accredited Fisher BioServices facility, UK. The biobank will transfer to the NIHR UK Biocentre in April 2018.
DNA samples prepared (or available to be prepared) from each participant?	No - serum collected; DNA has been extracted in subsets for the purpose of collaborative nested case control projects. A pilot is underway to assess the most cost-effective method to extract DNA from the entire cohort and proceed to whole genome array based genotyping (and eventually sequencing).
Is genotyping being done on some/all participants?	Yes – subset of participants with genetic data obtained using various arrays/platforms: Illumina Cardio-Metabochip (200,000 SNPs) in 1472 women OCAC/OncoArray derived 96 SNPs in 2178 women
Is genomic sequencing being done on some/all participants?	None
Other molecular analyses performed	Biomarker panel analyses using a variety of technologies including: 1) DNA methylation using Illumina's HiSeq 2500 array in 1097 participants 2) miRNA in 888 samples 2) Proteomics with Mass spectrometry in 2,458 samples and with SWATH in 482 samples 3) Quantitative NMR metabolomics using Nightingale platform in 4,841 participants (nested case-control for cardiovascular disease) 4) Lipidomics using Multiple Reaction Monitoring in 452 samples. 5) ELISA in more than 3,000 samples. 6) Autoantibody profile in 5600 samples

Did participants provide consent regarding sharing of their data outside the initial study investigators?	Broad (for any type of health-related research) - participants provided informed consent for the use of their samples and data in ethically approved secondary studies (with academic and commercial partners) Since 2015, Section-251 approval from the Confidentiality Advisory Group has also been obtained that allows processing of confidential patient information without consent.
How are data or specimens from the cohort made available for research? Any limitations on who can access the data (e.g. by country or sector?)	Project approved by data access committee The study is open to any investigator and industry partner with the only exception being industry proposals that plan to use biological samples (genomics is not included) for early detection/screening for Cancers, which needs to go Abcodia Pvt Ltd.
What study information or data are returned to or accessible by participants?	As per consent, no feedback of any results from analyses of samples or data in secondary studies.
Follow-up occurring? (years of follow-up). Is recontact possible?	Yes (currently 14 years follow-up; ongoing) Yes (provided it does not constitute feedback of information not already known by participant)
Notes/Comments	The UKCTOCS Longitudinal Women's Cohort (UKLWC) was originally established in 2001-2005 during the course of the United Kingdom Collaborative Trial of Ovarian Cancer Screening, one of the world's largest randomised controlled trials: 202,638 post-menopausal women from the general population aged 50-74 years were randomised (2:1:1) to routine care, or 7-11 rounds of annual blood tests or ultrasound to evaluate the impact of ovarian cancer screening on disease mortality. As a result, a high-quality serum biorepository is set up comprising of >544,000 serum samples including a unique set of longitudinal sampling in 50,262 participants (median of 9 samples). The latter (serial sampling) enables discovery of new multi-omics biomarkers and longitudinal algorithms of disease onset and prognosis. Furthermore, available comprehensive electronic health record linkage of the cohort's participants allows exploiting clinical phenotyping and diagnoses made during routine healthcare in the NHS. UKLWC has the potential to provide a national resource for the study of chronic disease that complements UK Biobank. The overall goal is to maximise the value of this comprehensive resource by transforming it into a cost-effective, digital, open-research bioresource for chronic conditions following the UK Biobank model by incorporating whole genome array based genotyping (and eventually sequencing), serum proteomics and metabolomics, and disease outcome ascertainment through health record linkage. This platform aims to be accessible to all bona fide scientists from academia and industry.
Questions Relating to Sharing & Collaboration	
May we make the information you provided about your cohort available on an open website?	Yes
Are you willing to share data from your cohort? If so, would you share:	Yes

a) individual data (redacted to protect confidentiality)?	Yes Comment: Individual data can only be shared with registered researchers for specific projects following an application to the UKLWC Committee and documented ethical approval
b) summary data (counts, distributions)?	Yes
c) metadata (descriptive information on data collection methods)?	Yes
d) case report forms and other data collection materials?	Yes
What do you see as the values of sharing?	Sharing of summary and meta data will allow bona fide scientists from a wide range of expertise to more easily assess suitability of the resource for their project Sharing of IPD if participants have provided specific consent will result in more rapid and transparent analysis and translation
What challenges do you anticipate with sharing?	The data ultimately belongs to the participant – there is need for clear and transparent consent before IPD is shared There is risk of identification of individual participants if significant amounts of data on an individual are shared There is a need to acknowledge appropriately the huge amount of work done by researchers (not only the PI) who gather and share the data. Significant challenges and costs to establish the infrastructure and platforms that allow storage, sharing and analyses of datasets.
What specific legal or regulatory barriers in your country or cohort (aside from ensuring confidentiality and appropriate consent) would prevent you from sharing data with other cohort of cohorts investigators, or openly with anyone who requests them?	None
What aspects of your cohort are intended for translation to clinical care or population health?	All
How might genomic sequencing add to/enhance your study objectives?	Adding genome profiling to the already available comprehensive electronic health record linkage and a unique biorepository will offer the opportunity to unravel the genome-phenome connections on a variety of medical conditions and would contribute significantly to: • closing the knowledge gap in diseases with limited/no GWAS data particularly those more prevalent in women • identification of additional GWAS-hits in disease/traits with mature GWAS datasets • augmenting power for analysis of low frequency/rare variants identified by ongoing sequencing efforts in the UK (Genomics England and UK Biobank) and elsewhere
Might you be willing to contribute funding or other resources to support international collaboration?	No Comment: No funds are available to do so

Women's Health Initiative (WHI)

Questions Relating to Cohort	
Name of study	Women's Health Initiative (WHI)
Principal Investigator/lead	Dr. Garnet Anderson
Contact email	garnet@whi.org
PubMed ID (or other information) for a protocol/marker paper on this study	PMID: 14575938, PubMed ID: 9492970
Study website	www.whi.org
Purpose or major Objectives of study	A long term study focused on strategies for preventing heart disease, breast and colorectal cancer, and osteoporotic fractures in postmenopausal women.
Disease areas of focus	Cancer, Cardiovascular Diseases, Fractures
Is your cohort selected for a specific disease (cancer, diabetes) or unselected for disease?	Unselected
Current size of population (and target number of participants)	161,808
Participating countries	US
Period of enrollment (and is enrollment on-going?)	1993-1998
Demographic characteristics of participants (age range, proportion male/female, national origin, race)	F, 50-79 y/o, race/ethnicity inclusive
Major diseases or phenotypes collected to date.	Broad - includes cancer, cardiovascular, bone, cognition, aging, eye disease, diabetes, autoimmune, mental health
Standardized clinical evaluation components measured, if applicable (e.g. height, weight, blood pressure, exercise testing, spirometry, etc.)	Anthropometry, blood pressure, CBC, pulse, and in specific subsets: ECG, physical function (gait speed, chair stands, grip strength), cognitive testing, accelerometry
Electronic health/medical records or medical administrative data used to collect clinical phenotypes?	Yes, but only through Medicare linkage
Predominant type of electronic health records (e.g. Epic, Cerner, etc.)	N/A
Other sources of clinical data	Questionnaires, medical record retrieval/adjudication
Environmental exposure data being obtained? What sort?	Yes--linkage to air pollution databases
Other data collected	Medical, family and reproductive histories, cancer screening, medication use, dietary supplement use, diet/physical activity, other lifestyle factors, psychosocial factors, demographics and

	socioeconomic factors, cancer treatment and recurrence for selected cancer sites
Biological specimens collected? What sort?	Blood, buffy coat, RNA (subsample), urine (subsample), and tumor tissue from selected cancers
Is there a central biobank?	Yes
DNA samples prepared (or available to be prepared) from each participant?	Yes
Is genotyping being done on some/all participants?	Yes, for some
Is genomic sequencing being done on some/all participants?	Yes, for some
Other molecular analyses performed	In representative subsamples of main study: lipids, creatinine, glucose, insulin, fat-soluble vitamins. In case-control or case-cohort subsamples: oxysterols, immune, coagulation, inflammation markers, adipokines, sex steroids, LH, FSH, thyroid hormonescarcinogens, cancer markers, metabolomics, cystatin, antibodies (hepatitis, H. pylori), fatty acids, folate, vitamin B12, HbA1c, telomere length, whole genome methylation, miRNA. Details can be found at https://deino.whiops.org/go/www.whi.org~ssl/SitePages/WHI%20Home.aspx
Did participants provide consent regarding sharing of their data outside the initial study investigators?	Yes - tiered
How are data or specimens from the cohort made available for research? Any limitations on who can access the data (e.g. by country or sector?)	Access through collaboration with a WHI investigator. Requires a scientific proposal to be approved by WHI and IRB with full funding. The only structural limitations are for commercial use (restricted to data and participant who signed the supplemental consent administered in 2004-5).
What study information or data are returned to or accessible by participants?	Complete blood count results were returned. Consent precludes return of other results.
Follow-up occurring? (years of follow-up). Is recontact possible?	Yes, annual questionnaires administered to the ~70,000 who remain active.
Notes/Comments	A long term study focused on strategies for preventing heart disease, breast and colorectal cancer, and osteoporotic fractures in postmenopausal women. Increasing focus on aging.
Questions Relating to Sharing & Collaboration	
May we make the information you provided about your cohort available on an open website?	Yes
Are you willing to share data from your cohort? If so, would you share:	Yes
a) individual data (redacted to protect confidentiality)?	Yes (Already shared through NHLBI BioLINCC, dbGaP)

b) summary data (counts, distributions)?	Yes (Already shared through WHI website, NHLBI BioLINCC)
c) metadata (descriptive information on data collection methods)?	Yes (Already shared through WHI website, NHLBI BioLINCC)
d) case report forms and other data collection materials?	Yes (Already shared through WHI website, NHLBI BioLINCC)
What do you see as the values of sharing?	Sharing promotes transparency; offers the opportunity to examine rare diseases or subtypes for which a single study is not adequately powered; provides the opportunity to validate findings in independent studies; allows for greater precision in evaluating the shape and strength of known associations; also allows one to look for negligible effect sizes.
What challenges do you anticipate with sharing?	Logistics are tedious and more costly than anticipated. It can be challenging to generate interest among WHI investigators to participate in consortia.
What specific legal or regulatory barriers in your country or cohort (aside from ensuring confidentiality and appropriate consent) would prevent you from sharing data with other cohort of cohorts investigators, or openly with anyone who requests them?	None
What aspects of your cohort are intended for translation to clinical care or population health?	WHI is a population-based study of older women. Results are directly generalizable to US postmenopausal women.
How might genomic sequencing add to/enhance your study objectives?	Genomic sequencing and whole genome methylation of WHI buffy coat samples would allow for more informative analyses of genetic and epigenetic influences on divsere disease outcomes. Genomic sequencing of WHI tumor samples would create a much enhanced version of the TCGA as these samples are fully annotated with high quality pre-diagnostic information, clinical characteristics, treatment (first course) and long-term follow-up.
Might you be willing to contribute funding or other resources to support international collaboration?	No (WHI is funded by a contract from NHLBI that provides funds only for the existing workscope)