Ruth, K.S., Day, F.R., Hussain, J. *et al.* Genetic insights into biological mechanisms governing human ovarian ageing. *Nature* **596,** 393–397 (2021).
https://doi.org/10.1038/s41586-021-03779-7

Contents:

Supplementary Tables are provided separately.

Title: Genetic insights into biological mechanisms governing human ovarian ageing

Katherine S Ruth*[1], Felix R Day*[2], Jazib Hussain*[3], Ana Martínez-Marchal*[4,5], Catherine E Aiken[6,7], Ajuna Azad[3], Deborah J Thompson[8], Lucie Knoblochova[9,10], Hironori Abe[11], Jane L Tarry-Adkins[6,7], Javier Martin Gonzalez[12], Pierre Fontanillas[13], Annique Claringbould[14], Olivier B Bakker[15], Patrick Sulem[16], Robin G Walters[17,18], Chikashi Terao[19,20,21], Sandra Turon[22], Momoko Horikoshi[23], Kuang Lin[17], N Charlotte Onland-Moret[24], Aditya Sankar[3], Emil Peter Thrane Hertz[3,25], Pascal N Timshel[26], Vallari Shukla[3], Rehannah Borup[3], Kristina W Olsen[27,3], Paula Aguilera[3,28], Mònica Ferrer-Roda[4,5], Yan Huang[4,5], Stasa Stankovic[2], Paul RHJ Timmers[29,30], Thomas U Ahearn[31], Behrooz Z Alizadeh[32], Elnaz Naderi[32], Irene L Andrulis[33,34], Alice M Arnold[35], Kristan J Aronson[36], Annelie Augustinsson[37], Stefania Bandinelli[38], Caterina M Barbieri[39], Robin N Beaumont[1], Heiko Becher[40], Matthias W Beckmann[41], Stefania Benonisdottir[16], Sven Bergmann[42,43], Murielle Bochud[44], Eric Boerwinkle[45], Stig E Bojesen[46,47,48], Manjeet K Bolla[8], Dorret I Boomsma[49,50], Nicholas Bowker[2], Jennifer A Brody[51], Linda Broer[52], Julie E Buring[53,54], Archie Campbell[55], Harry Campbell[29], Jose E Castelao[56], Eulalia Catamo[57], Stephen J Chanock[31], Georgia Chenevix-Trench[58], Marina Ciullo[59,60], Tanguy Corre[44,42,43], Fergus J Couch[61], Angela Cox[62], Laura Crisponi[63], Simon S Cross[64], Francesco Cucca[63,65], Kamila Czene[66], George Davey Smith[67,68], Eco JCN de Geus[49,50], Renée de Mutsert[69], Immaculata De Vivo[70,71], Ellen W Demerath[72], Joe Dennis[8], Alison M Dunning[73], Miriam Dwek[74], Mikael Eriksson[75], Tõnu Esko[76,77], Peter A Fasching[78,41], Jessica D Faul[79], Luigi Ferrucci[80], Nora Franceschini[81], Timothy M Frayling[1], Manuela Gago-Dominguez[82,83], Massimo Mezzavilla[84], Montserrat García-Closas[31], Christian Gieger[85,86,87], Graham G Giles[88,89,90], Harald Grallert[85,86,87], Daniel F Gudbjartsson[16], Vilmundur Gudnason[91,92], Pascal Guénel[93], Christopher A Haiman[94], Niclas Håkansson[95], Per Hall[66], Caroline Hayward[30], Chunyan He[96,97], Wei He[75], Gerardo Heiss[81], Miya K Høffding[3], John L Hopper[89], Jouke J Hottenga[49,50], Frank Hu[98,71,70], David Hunter[17,99,71,100], Mohammad A Ikram[101], Rebecca D Jackson[102], Micaella DR Joaquim[1], Esther M John[103,104], Peter K Joshi[29], David Karasik[105,54], Sharon LR Kardia[106], Christiana Kartsonaki[17,18], Robert Karlsson[107], Cari M Kitahara[108], Ivana Kolcic[109], Charles Kooperberg[110], Peter Kraft[70,111], Allison W Kurian[104,103], Zoltan Kutalik[44,43], Martina La Bianca[57], Genevieve LaChance[112], Claudia Langenberg[2], Lenore J Launer[113], Joop SE Laven[114], Deborah A Lawlor[67,68], Loic Le Marchand[115], Jingmei Li[66], Annika Lindblom[116,117], Sara Lindstrom[118], Tricia Lindstrom[119], Martha Linet[108], YongMei Liu[120], Simin Liu[121], Jian'an Luan[2], Reedik Mägi[77], Patrik KE Magnusson[107], Massimo Mangino[112,122], Arto Mannermaa[123,124,125], Brumat Marco[84], Jonathan Marten[30], Nicholas G Martin[126], Hamdi Mbarek[49,50], Barbara McKnight[35], Sarah E Medland[126], Christa Meisinger[86,127], Thomas Meitinger[128], Cristina Menni[112], Andres Metspalu[77], Lili Milani[77], Roger L Milne[88,89,90], Grant W Montgomery[129], Dennis O Mook-Kanamori[69,130], Antonella Mulas[63], Anna M Mulligan[131,132], Alison Murray[133], Mike A Nalls[134], Anne Newman[135], Raymond Noordam[136], Teresa Nutile[59], Dale R Nyholt[137], Andrew F Olshan[138], Håkan Olsson[37], Jodie N Painter[126], Alpa V Patel[139], Nancy L Pedersen[107], Natalia Perjakova[77], Annette Peters[86,87], Ulrike Peters[110], Paul DP Pharoah[73,8], Ozren Polasek[140,141], Eleonora Porcu[63], Bruce M Psaty[51], Iffat Rahman[142], Gad Rennert[143], Hedy S Rennert[143], Paul M Ridker[53,54], Susan M Ring[67,68], Antonietta Robino[57], Lynda M Rose[53], Frits R Rosendaal[69], Jacques Rossouw[144], Igor Rudan[29], Rico Rueedi[42,43], Daniela Ruggiero[59,60], Cinzia F Sala[39], Emmanouil Saloustros[145], Dale P Sandler[146], Serena Sanna[63], Elinor J Sawyer[147], Chloé Sarnowski[148], David

Schlessinger[149], Marjanka K Schmidt[150,151], Minouk J Schoemaker[152], Katharina E Schraut[153,29], Christopher Scott[119], Saleh Shekari[1], Amruta Shrikhande[3], Albert V Smith[91,92], Blair H Smith[154], Jennifer A Smith[106], Rossella Sorice[59], Melissa C Southey[90,155,88], Tim D Spector[112], John J Spinelli[156,157], Meir Stampfer[70,71,98], Doris Stöckl[86,158], Joyce BJ van Meurs[52], Konstantin Strauch[159,160,161], Unnur Styrkarsdottir[16], Anthony J Swerdlow[152,162], Toshiko Tanaka[80], Lauren R Teras[139], Alexander Teumer[163], Unnur Þorsteinsdottir[16,164], Nicholas J Timpson[67,68], Daniela Toniolo[39], Michela Traglia[39], Melissa A Troester[138], Thérèse Truong[93], Jessica Tyrrell[1], André G Uitterlinden[52,101], Sheila Ulivi[57], Celine M Vachon[165], Veronique Vitart[30], Uwe Völker[166], Peter Vollenweider[167], Henry Völzke[163], Qin Wang[8], Nicholas J Wareham[2], Clarice R Weinberg[168], David R Weir[79], Amber N Wilcox[31], Ko Willems van Dijk[169,170,171], Gonneke Willemsen[49,50], James F Wilson[29,30], Bruce HR Wolffenbuttel[172], Alicja Wolk[95,173], Andrew R Wood[1], Wei Zhao[106], Marek Zygmunt[174], Biobank-based Integrative Omics Study (BIOS) Consortium[#], eQTLGen Consortium[#], The Biobank Japan Project[#], China Kadoorie Biobank Collaborative Group[#], kConFab Investigators[#], The LifeLines Cohort Study[#], The InterAct consortium[#], 23andMe Research Team[#], Zhengming Chen[17,18], Liming Li[175,176], Lude Franke[15,177], Stephen Burgess[178,179], Patrick Deelen[180,15], Tune H Pers[26], Marie Louise Grøndahl[27], Claus Yding Andersen[181], Anna Pujol[22], Andres J Lopez-Contreras[3,28], Jeremy A Daniel[25], Kari Stefansson[16,164], Jenny Chang-Claude[182,183], Yvonne T van der Schouw[24], Kathryn L Lunetta[148,184], Daniel I Chasman[53,54], Douglas F Easton[73,8], Jenny A Visser[52], Susan E Ozanne[185], Satoshi H Namekawa[11], Petr Solc[§9], Joanne M Murabito[186,184], Ken K Ong[2,187], Eva R Hoffmann*[3], Anna Murray*[1], Ignasi Roig*[4,5] and John RB Perry*[2,52]

[#] Lists of participants and their affiliations appear in the Supplementary Information.
* Denotes equal contribution
[§] Deceased
Correspondence to Eva R Hoffmann (eva@sund.ku.dk), Anna Murray (A.Murray@exeter.ac.uk), Ignasi Roig (ignasi.roig@uab.cat) and John R B Perry (john.perry@mrc-epid.cam.ac.uk).

Affiliations

[1]Genetics of Human Complex Traits, University of Exeter Medical School, University of Exeter, Exeter, U.K, [2]MRC Epidemiology Unit, University of Cambridge School of Clinical Medicine, Box 285 Institute of Metabolic Science, Cambridge Biomedical Campus, Cambridge CB2 0QQ, UK, [3]DNRF Center for Chromosome Stability, Department of Cellular and Molecular Medicine, Faculty of Health and Medical Sciences, University of Copenhagen, Denmark, [4]Genome Integrity and Instability Group, Institut de Biotecnologia i Biomedicina, Universitat Autònoma de Barcelona, Cerdanyola del Vallès, Spain., [5]Department of Cell Biology, Physiology and Immunology, Universitat Autònoma de Barcelona, Cerdanyola del Vallès, Spain, [6]University of Cambridge Metabolic Research Laboratories and MRC Metabolic Diseases Unit, Institute of Metabolic Science, Addenbrooke's Hospital, Cambridge CB2 0QQ, United Kingdom, [7]Department of Obstetrics and Gynaecology, University of Cambridge, Box 223, The Rosie Hospital and NIHR Cambridge Biomedical Research Centre, Cambridge CB2 0SW, UK, [8]Centre for Cancer Genetic Epidemiology, Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK., [9]Institute of Animal Physiology and Genetics of the Czech Academy of Sciences, Libechov, Czech Republic, [10]Faculty of Science, Charles University, Prague, Czech Republic, [11]Division of Reproductive Sciences, Cincinnati Children's Hospital Medical Center, Cincinnati, Ohio, 45229, USA, [12]Transgenic Core Facility, Department of Experimental Medicine, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark, [13]23andMe Inc., 223 N Mathilda Ave, Sunnyvale, CA, 94086, USA., [14]Structural and Computational Biology Unit, EMBL, Heidelberg, Germany, [15]University of Groningen, University Medical Center Groningen, Department of Genetics, Groningen, The Netherlands, [16]deCODE genetics/Amgen, Reykjavik, Iceland, [17]Nuffield Department of Population Health, University of Oxford, Oxford OX3 7LF, UK, [18]MRC Population Health Research Unit, University of Oxford, Oxford OX3 7LF, UK, [19]Laboratory for Statistical and Translational Genetics, RIKEN Center for Integrative Medical Sciences, Yokohama, Japan., [20]Clinical Research Center, Shizuoka General Hospital, Shizuoka, Japan., [21]Department of Applied Genetics, School of Pharmaceutical Sciences, University of Shizuoka, Shizuoka, Japan., [22]Transgenic Animal Unit, Center of Animal Biotechnology and Gene Therapy, Universitat Autònoma de Barcelona, Cerdanyola del Vallès, Spain., [23]Laboratory for Genomics of Diabetes and Metabolism, RIKEN Center for Integrative Medical Sciences, Yokohama, Japan, [24]Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Utrecht University, the Netherlands., [25]The Novo Nordisk Foundation Center for Protein Research, Faculty of Health and Medical Sciences, University of Copenhagen, Denmark, [26]The Novo Nordisk Foundation Center for Basic Metabolic Research, Faculty of Health and Medical Sciences, University of Copenhagen, København Ø, Denmark., [27]Department of Obstetrics and Gynaecology, Department of Reproductive Medicine, Copenhagen University Hospital Herlev, Denmark, [28]Centro Andaluz de Biología Molecular y Medicina Regenerativa (CABIMER), Consejo Superior de Investigaciones Científicas (CSIC) - Universidad de Sevilla - Universidad Pablo de Olavide, Seville, Spain., [29]Centre for Global Health Research, Usher Institute, University of Edinburgh, Teviot Place, Edinburgh, EH8 9AG, UK, [30]MRC Human Genetics Unit, Institute of Genetics and Molecular Medicine, University of Edinburgh, Western General Hospital, Edinburgh EH4 2XU, Scotland, [31]Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health, Department of Health and Human Services, Bethesda, MD, USA., [32]University of Groningen, University Medical Center Groningen, Department of Epidemiology, Groningen, The Netherlands, [33]Fred A. Litwin Center for Cancer Genetics, Lunenfeld-Tanenbaum Research Institute of Mount Sinai Hospital, Toronto, ON, Canada., [34]Department of Molecular Genetics, University of Toronto, Toronto, ON, Canada.,

[35]Department of Biostatistics, University of Washington, Seattle, WA, USA, [36]Department of Public Health Sciences, and Cancer Research Institute, Queen's University, Kingston, ON, Canada., [37]Department of Cancer Epidemiology, Clinical Sciences, Lund University, Lund, Sweden., [38]Geriatric Unit, Azienda Sanitaria Firenze (ASF), Florence, Italy, [39]Genetics of Common Disorders Unit, IRCCS San Raffaele Scientific Institute, Milan, Italy, [40]Institute of Medical Biometry and Epidemiology, University Medical Center Hamburg-Eppendorf, Hamburg, Germany., [41]Department of Gynecology and Obstetrics, Comprehensive Cancer Center ER-EMN, University Hospital Erlangen, Friedrich-Alexander-University Erlangen-Nuremberg, Erlangen, Germany., [42]Department of Computational Biology, University of Lausanne, Switzerland, [43]Swiss Institute of Bioinformatics, [44]University Center for Primary Care and Public Health, University of Lausanne, Switzerland, [45]Human Genetics Center, School of Public Health, The University of Texas Health Science Center at Houston, Houston, TX 77030, USA, [46]Copenhagen General Population Study, Herlev and Gentofte Hospital, Copenhagen University Hospital, Herlev, Denmark., [47]Department of Clinical Biochemistry, Herlev and Gentofte Hospital, Copenhagen University Hospital, Herlev, Denmark., [48]Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark., [49]Department of Biological Psychology, Vrije Universiteit Amsterdam, Amsterdam, The Netherlands, [50]Amsterdam Public Health (APH) and Amsterdam Reproduction and Development (AR&D) research institutes, The Netherlands, [51]Cardiovascular Health Research Unit, Departments of Medicine, Epidemiology, and Health Services, University of Washington, Seattle, WA, USA, [52]Department of Internal Medicine, Erasmus MC, Rotterdam, The Netherlands, [53]Brigham and Women's Hospital, Boston MA, [54]Harvard Medical School, Boston, Massachusetts, USA, [55]Medical Genetics Section, Centre for Genomic and Experimental Medicine, Institute of Genetics and Molecular Medicine, University of Edinburgh, Edinburgh EH4 2XU, UK, [56]Oncology and Genetics Unit, Instituto de Investigacion Sanitaria Galicia Sur (IISGS), Xerencia de Xestion Integrada de Vigo-SERGAS, Vigo, Spain., [57]Institute for Maternal and Child Health - IRCCS "Burlo Garofolo", Trieste, Italy, [58]Department of Genetics and Computational Biology, QIMR Berghofer Medical Research Institute, Brisbane, Queensland, Australia., [59]Institute of Genetics and Biophysics - CNR, via Pietro Castellino 111, 80131, Naples, Italy, [60]IRCCS Neuromed, Pozzilli, Isernia, Italy, [61]Department of Laboratory Medicine and Pathology, Mayo Clinic, Rochester, MN, USA., [62]Sheffield Institute for Nucleic Acids (SInFoNiA), Department of Oncology and Metabolism, University of Sheffield, Sheffield, UK., [63]Institute of Genetics and Biomedical Research, National Research Council, Cagliari, Italy, [64]Academic Unit of Pathology, Department of Neuroscience, University of Sheffield, Sheffield, UK., [65]University of Sassari, Dept. Of Biomedical Sciences, Sassari, Italy, [66]Karolinska Institutet, Department of Medical Epidemiology and Biostatistics, Box 281, 171 77 Stockholm, Sweden, [67]MRC Integrative Epidemiology Unit at the University of Bristol, Oakfield House, Oakfield Grove, Bristol, BS8 2BN, [68]Population Health Science, Bristol Medical School, University of Bristol, Bristol, UK, [69]Department of Clinical Epidemiology, Leiden University Medical Center, Leiden, the Netherlands, [70]Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, MA 02115, USA, [71]Channing Division of Network Medicine, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA 02115, USA, [72]Division of Epidemiology & Community Health, University of Minnesotta, Minneapolis MN 55455, [73]Centre for Cancer Genetic Epidemiology, Department of Oncology, University of Cambridge, Cambridge, UK., [74]School of Life Sciences, University of Westminster, London, UK., [75]Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden., [76]Population and Medical Genetics, Broad Institute, Cambridge, 02141, US, [77]Estonian Genome Center, Institute of Genomics, University of Tartu, Riia 23b, 51010, Tartu, Estonia, [78]David Geffen School of Medicine, Department of Medicine Division of Hematology and Oncology, University of California at Los Angeles, Los Angeles, CA, USA., [79]Survey Research Center, Institute for Social Research, Ann Arbor, MI 48104, [80]Translational Gerontology Branch, National Institute on Aging, Baltimore MD, USA, [81]Department of

Epidemiology, Gillings School of Global Public Health, University of North Carolina, Chapel Hill, NC 27514, [82]Fundación Pública Galega de Medicina Xenómica, Instituto de Investigación Sanitaria de Santiago de Compostela (IDIS), Complejo Hospitalario Universitario de Santiago, SERGAS, Santiago de Compostela, Spain., [83]Moores Cancer Center, University of California San Diego, La Jolla, CA, USA., [84]Department of Medical Sciences, University of Trieste, Italy, [85]Research Unit of Molecular Epidemiology, Helmholtz Zentrum München–German Research Center for Environmental Health, Neuherberg, Germany., [86]Institute of Epidemiology II, Helmholtz Zentrum München–German Research Center for Environmental Health, Neuherberg, Germany., [87]German Center for Diabetes Research (DZD), Neuherberg, Germany., [88]Cancer Epidemiology Division, Cancer Council Victoria, Melbourne, Victoria, Australia., [89]Centre for Epidemiology and Biostatistics, Melbourne School of Population and Global Health, The University of Melbourne, Melbourne, Victoria, Australia., [90]Precision Medicine, School of Clinical Sciences at Monash Health, Monash University, Clayton, Victoria, Australia., [91]Icelandic Heart Association, 201 Kopavogur, Iceland, [92]Faculty of Medicine, University of Iceland, 101 Reykjavik, Iceland, [93]Cancer & Environment Group, Center for Research in Epidemiology and Population Health (CESP), INSERM, University Paris-Sud, University Paris-Saclay, Villejuif, France., [94]Department of Preventive Medicine, Keck School of Medicine, University of Southern California, Los Angeles, CA, USA., [95]Institute of Environmental Medicine, Karolinska Institutet, Stockholm, Sweden., [96]Division of Medical Oncology, Department of Internal Medicine, University of Kentucky College of Medicine, Lexington, KY 40536, USA, [97]The Cancer Prevention and Control Research Program, University of Kentucky Markey Cancer Center, Lexington, KY 40536, USA, [98]Department of Nutrition, Harvard T.H. Chan School of Public Health, Boston, MA 02115, USA, [99]Departments of Epidemiology and Nutrition, Harvard T.H. Chan School of Public Health, Boston, MA 02115, USA, [100]Broad Institute of Harvard and MIT, Cambridge, MA, USA, [101]Department of Epidemiology, Erasmus MC, Rotterdam, The Netherlands, [102]Department of Internal Medicine, The Ohio State University, Columbus, Ohio, [103]Department of Epidemiology & Population Health, Stanford University School of Medicine, Stanford, CA, USA., [104]Department of Medicine, Division of Oncology, Stanford Cancer Institute, Stanford University School of Medicine, Stanford, CA, USA., [105]Hebrew SeniorLife Institute for Aging Research, Boston, Massachusetts, USA, [106]Department of Epidemiology, School of Public Health, University of Michigan, Ann Arbor, MI 48109, [107]Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden, [108]Radiation Epidemiology Branch, Division of Cancer Epidemiology and Genetics, National Cancer Institute, Bethesda, MD, USA., [109]Faculty of Medicine, University of Split, Split, Croatia, [110]Division of Public Health Sciences, Fred Hutchinson Cancer Research Center, Seattle WA 98109, [111]Department of Biostatistics, Harvard T.H. Chan School of Public Health, Boston, MA 02115, USA, [112]Department of Twin Research and Genetic Epidemiology, King's College London, London, UK, [113]Laboratory of Epidemiology and Population Sciences, National Institute on Aging, Intramural Research Program, National Institutes of Health, Bethesda, Maryland, 20892, USA, [114]Div. of Reproductive Endocrinology & Infertility, Dept . of Obstetrics and Gynecology, Erasmus University Medical Center, Rotterdam, The Netherlands, [115]Epidemiology Program, University of Hawaii Cancer Center, Honolulu, HI, USA., [116]Department of Molecular Medicine and Surgery, Karolinska Institutet, Stockholm, Sweden., [117]Department of Clinical Genetics, Karolinska University Hospital, Stockholm, Sweden., [118]Department of Epidemiology, University of Washington, Seattle, WA, 98195, USA, [119]Department of Health Sciences Research, Mayo Clinic, Rochester, MN, USA., [120]Center for Human Genetics, Division of Public Health Sciences, Wake Forest School of Medicine, Wake Forest, NC, USA, [121]Departments of Epidemiology and Medicine Brown University, Brown University, Providence, RI 02912, [122]NIHR Biomedical Research Centre at Guy's and St. Thomas' Foundation Trust, London, UK, [123]Translational Cancer Research Area, University of Eastern Finland, Kuopio, Finland., [124]Institute of Clinical Medicine, Pathology and Forensic Medicine, University of Eastern Finland, Kuopio, Finland.,

[125]Biobank of Eastern Finland, Kuopio University Hospital, Kuopio, Finland., [126]QIMR Berghofer Medical Research Insititute, [127]Central Hospital of Augsburg, MONICA/KORA Myocardial Infarction Registry, Augsburg, Germany, [128]Institute of Human Genetics, Helmholtz Zentrum München, German Research Center for Environmental Health, Neuherberg, Germany, [129]Institute for Molecular Bioscience, The University of Queensland, Brisbane, Australia, [130]Department of Public Health and Primary Care, Leiden University Medical Center, Leiden, the Netherlands, [131]Department of Laboratory Medicine and Pathobiology, University of Toronto, Toronto, ON, Canada., [132]Laboratory Medicine Program, University Health Network, Toronto, ON, Canada., [133]The Institute of Medical Sciences, Aberdeen Biomedical Imaging Centre, University of Aberdeen, Aberdeen, UK, [134]Laboratory of Neurogneetics, National Institute on Aging, National Institutes of Health, Bethesda, MD, USA, [135]Departments of Epidemiology and Medicine, University of Pittsburgh, Pittsburgh, Pennsylvania, [136]Department of Internal Medicine, Section Gerontology and Geriatrics, Leiden University Medical Center, Leiden, the Netherlands, [137]Queensland University of Technology, Faculty of Health, School of Biomedical Sciences, Centre for Genomics and Personalised Health, Brisbane, Queensland, Australia, [138]Department of Epidemiology, Gillings School of Global Public Health and UNC Lineberger Comprehensive Cancer Center, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA., [139]Department of Population Science, American Cancer Society, 250 Williams St., Atlanta, GA 30303, USA, [140]Faculty of Medicine, University of Split, Split, Croatia., [141]Gen-Info Ltd, Zagreb, Croatia., [142]Quantify Research, Stockholm, Sweden, [143]Clalit National Cancer Control Center, Carmel Medical Center and Technion Faculty of Medicine, Haifa, Israel., [144]Women's Health Initiative Branch, National Heart, Lung, and Blood Institute, Bethesda, MD, 20892, USA., [145]Department of Oncology, University Hospital of Larissa, Larissa, Greece., [146]Epidemiology Branch, National Institute of Environmental Health Sciences, NIH, Research Triangle Park, NC, USA., [147]School of Cancer & Pharmaceutical Sciences, Comprehensive Cancer Centre, Guy's Campus, King's College London, London, UK., [148]Boston University School of Public Health, Department of Biostatistics. Boston, Massachusetts 02118, USA, [149]National Institute on Aging, Intramural Research Program, Baltimore, MD, USA, [150]Division of Molecular Pathology, The Netherlands Cancer Institute - Antoni van Leeuwenhoek Hospital, Amsterdam, The Netherlands., [151]Division of Psychosocial Research and Epidemiology, The Netherlands Cancer Institute - Antoni van Leeuwenhoek hospital, Amsterdam, The Netherlands., [152]Division of Genetics and Epidemiology, The Institute of Cancer Research, London, UK., [153]Centre for Cardiovascular Sciences, Queen's Medical Research Institute, University of Edinburgh, Royal Infirmary of Edinburgh, Little France Crescent, Edinburgh, EH16 4TJ, Scotland, [154]Division of Population and Health Genomics, University of Dundee, Dundee, UK, [155]Department of Clinical Pathology, The University of Melbourne, Melbourne, Victoria, Australia., [156]Population Oncology, BC Cancer, Vancouver, BC, Canada., [157]School of Population and Public Health, University of British Columbia, Vancouver, BC, Canada., [158]Department of Obstetrics and Gynaecology, Campus Grosshadern, Ludwig-Maximilians-Universität, Munich, Germany, [159]Institute of Genetic Epidemiology, Helmholtz Zentrum München - German Research Center for Environmental Health, Neuherberg, Germany, [160]Chair of Genetic Epidemiology, IBE, Faculty of Medicine, LMU Munich, Munich, Germany, [161]Institute of Medical Biostatistics, Epidemiology and Informatics (IMBEI), University Medical Center, Johannes Gutenberg University, Mainz, Germany, [162]Division of Breast Cancer Research, The Institute of Cancer Research, London, UK., [163]Institute for Community Medicine, University Medicine Greifswald, 17475 Greifswald, Germany, [164]Faculty of Medicine, School of Health Sciences, University of Iceland, Reykjavik, Iceland, [165]Department of Health Science Research, Division of Epidemiology, Mayo Clinic, Rochester, MN, USA., [166]Interfaculty Institute for Genetics and Functional Genomics, University Medicine Greifswald, 17475 Greifswald, Germany, [167]Department of Medicine, Internal Medicine, Lausanne University Hospital and University of Lausanne, Lausanne, Switzerland, [168]Biostatistics and Computational Biology Branch, National Institute of Environmental Health

Sciences, NIH, Research Triangle Park, NC, USA., [169]Department of Internal Medicine, Division of Endocrinology, Leiden University Medical Center, Leiden, the Netherlands, [170]Einthoven Laboratory for Experimental Vascular Medicine, Leiden University Medical Center, Leiden, the Netherlands, [171]Department of Human Genetics, Leiden University Medical Center, Leiden, the Netherlands, [172]University of Groningen, University Medical Center Groningen, Department of Endocrinology, Groningen, The Netherlands, [173]Department of Surgical Sciences, Uppsala University, Uppsala, Sweden., [174]Department of Obstetrics and Gynecology, University Medicine Greifswald, 17475 Greifswald, Germany, [175]School of Public Health, Peking University Health Science Center, Beijing, 100191, P.R. China, [176]Peking University Center for Public Health and Epidemic Preparedness & Response, Beijing, 100191, P.R. China, [177]Oncode Institute, Utrecht, The Netherlands, [178]Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK, [179]MRC Biostatistics Unit, University of Cambridge, Cambridge, UK, [180]Department of Genetics, University Medical Centre Utrecht, P.O. Box 85500, 3508 GA, Utrecht, The Netherlands, [181]Laboratory of Reproductive Biology, The Juliane Marie Centre for Women, Children and Reproduction, Copenhagen University Hospital and Faculty of Health and Medical Sciences, University of Copenhagen, Denmark, [182]Division of Cancer Epidemiology, German Cancer Research Center (DKFZ), Heidelberg, Germany., [183]Cancer Epidemiology Group, University Cancer Center Hamburg (UCCH), University Medical Center Hamburg-Eppendorf, Hamburg, Germany., [184]NHLBI's and Boston University's Framingham Heart Study, Framingham, Massachusetts 01702-5827, USA., [185]University of Cambridge Metabolic Research Laboratories and MRC Metabolic Diseases Unit, Wellcome - MRC Institute of Metabolic Science, Addenbrooke's Hospital, Cambridge, CB2 0QQ, [186]Boston University School of Medicine, Department of Medicine, Section of General Internal Medicine, Boston, MA 02118, USA., [187]Department of Paediatrics, University of Cambridge, Cambridge CB2 0QQ, UK

Abstract

Reproductive longevity is critical for fertility and impacts healthy ageing in women[1,2], yet insights into the underlying biological mechanisms and treatments to preserve it are limited. Here, we identify 290 genetic determinants of ovarian ageing, assessed using normal variation in age at natural menopause (ANM) in ~200,000 women of European ancestry. These common alleles were associated with clinical extremes of ANM; women in the top 1% of genetic susceptibility have an equivalent risk of premature ovarian insufficiency to those carrying monogenic *FMR1* premutations[3]. Identified loci implicate a broad range of DNA damage response (DDR) processes and include loss-of-function variants in key DDR genes. Integration with experimental models demonstrates that these DDR processes act across the life-course to shape the ovarian reserve and its rate of depletion. Furthermore, we demonstrate that experimental manipulation of DDR pathways highlighted by human genetics increase fertility and extend reproductive life in mice. Causal inference analyses using the identified genetic variants indicates that extending reproductive life in women improves bone health and reduces risk of type 2 diabetes, but increases risks of hormone-sensitive cancers. These findings provide insight into the mechanisms governing ovarian ageing, when they act across the life-course, and how they might be targeted by therapeutic approaches to extend fertility and prevent disease.

Introduction

Over the last 150 years life expectancy has increased from 45 to 85 years[4], but the timing of reproductive senescence (age at natural menopause (ANM)) has remained relatively constant (50-52 years)[5]. The genetic integrity of oocytes decreases with advancing age[6] and natural fertility ceases ~10 years before menopause[1]. More women are choosing to delay childbearing to older ages, resulting in increased use of assisted conception techniques[7,8]. Oocyte and ovarian tissue preservation can prolong fertility but is invasive and there is only a ~6.5% chance of achieving pregnancy with each mature oocyte thawed, which decreases with age[9].

ANM is determined by the non-renewable ovarian reserve, which is established during fetal development and continuously depleted until reproductive senescence (**Extended Data Fig. 1**). DNA damage response (DDR) is the primary biological pathway that regulates reproductive senescence, highlighted by genome-wide association studies (GWAS)[10], rare single gene disorders that cause Premature Ovarian Insufficiency (POI)[11] and animal models[12]. Better understanding of how and when molecular processes influence the establishment and decline of ovarian reserve will inform future therapeutic strategies for infertility treatment and fertility preservation. To address this, our current study increases the number of ANM-associated genetic loci six-fold[13] from 56 to 290. We integrate these data with experiments in mice to characterize the specific DDR processes that contribute to reproductive ageing, providing insights into when they act across the life-course, how they might be modified to preserve fertility and the potential consequences for broader health.

Results

Genome-wide array data, imputed to ~13.1 million genetic variants with minor allele frequency ≥0.1%, were available in 201,323 women of European ancestry (**Extended Data Fig. 2, Supplementary Table 1**). We identified 290 statistically independent signals associated with ANM ($P<5x10^{-8}$), including six on the X-chromosome which was previously untested in large-scale studies (**Figure 1, Supplementary Table 2**). Effect estimates for the 290 signals were consistent between linear and Cox proportional hazard models and across strata of the meta-analysis (**Extended Data Fig. 3**). There was no evidence of test statistic inflation due to population structure (LD score intercept=1.02, s.e. 0.03). All previously reported signals[13] retained genome-wide significance (**Figure 1**).

Additive, per-allele effect sizes for the 290 signals ranged from ~3.5 weeks to ~74 weeks (**Figure 1, Extended Data Fig. 2 and Supplementary Table 2**). Three of these variants exhibited non-additive effects (**Extended Data Fig. 4 a-d, Supplementary Table 3 and Supplementary Results).** We sought to replicate our 290 signals using independent samples from 23andMe, Inc (N=294,828 women). We observed high concordance in effect estimates between the datasets (**Supplementary Table 2 and Extended Data Fig. 3 g),** with nearly all variants at least nominally associated with ANM in 23andMe. Eight variants fell below genome-wide significance in a meta-analysis of our discovery with 23andMe ($P_{max}=2.6x10^{-5}$), half the

number of expected false-positive associations (290*0.05=14.5). We next evaluated these loci in 78,317 women of East Asian ancestry. There was broad replication, consistent with previous observations[14], but substantial heterogeneity of effect sizes and allele frequencies (**Supplementary Table 2**). This was exemplified at the *ENTPD1* locus, where one signal had an effect size ~3 times larger in East Asians (rs1889921), whilst a second independent signal ~20kb away had an effect estimate half the size in East Asians (rs7087644).

Using additional independent samples from the deCODE study (N=16,556 women), we estimated our identified signals cumulatively explained 10.1% of the variance in ANM. This compared to an estimate of 12.3% in UK Biobank (UKBB) using weights for the 290 variants derived from our non-UKBB samples (**Supplementary Table 2**). The identified signals therefore account for 31-38% of the overall genotype-array estimated heritability in UKBB (h^2_g=32.4%, s.e. 0.8%), compared to 15.7-19.8% for the 56 previously reported signals (**Extended Data Fig. 4 e**).

Common variants act on extremes of ANM

It is unclear where in the population distribution of ANM the influence of common genetic variants begins and ends. Our GWAS was restricted to the 99% of women with ANM between 40-60 years. ANM before 40 years (POI) is considered a Mendelian disorder, but may have a polygenic component. To test which parts of the ANM distribution are influenced by common genetic variation, we calculated a polygenic score (PGS) in 108,840 women in UKBB with the full range of ANM using genetic weights derived from the independent non-UKBB component of the meta-analysis (**Supplementary Table 2**). This was coded such that a higher PGS indicates increased susceptibility to later ANM. ANM from 34 to 61 years had a significant polygenic influence (**Figure 2 a**). For example, women with ANM at 34 years had an average -0.5 SD (95% CI 0.26-0.69, P=1.5x10^{-5}) lower PGS than the population mean. We had limited sample size to test outside of these age ranges, however there was some evidence for a depletion of a polygenic influence at ages younger than 34 years (**Figure 2 a**). These data suggest that common genetic variants act on clinically relevant extremes of ANM, although it remains unclear what fraction of POI cases may be polygenic vs monogenic.

Secondly, we evaluated the predictive ability of the PGS. Genetic risk alone proved to be a weak predictor (ROC-AUC 0.65 and 0.64 for early menopause (age <45 years) and POI respectively) (**Figure 2 b** and **c**), however the PGS performed significantly better than smoking status which is the most robust epidemiologically associated risk factor (ROC-AUC 0.58). Adding smoking status to the PGS did not appreciably improve prediction of early menopause (ROC-AUC 0.66). Despite low overall discriminative ability, the PGS was able to identify individuals at high risk of POI (**Figure 2 c**). Women at the top 1% of the PGS (rescaled such that high PGS indicates increased susceptibility to earlier menopause) had equivalent POI risk (PGS OR 4.71 [3.15-7.04] vs 50th centile, P=4.4x10^{-14}) to that reported for women with *FMR1* premutations, the leading tested monogenic cause of POI (OR~5)[3]. It is however notable that the top 1% of genetic risk is more prevalent than the *FMR1* premutation carrier rate (1:250).

Functional genes and pathways implicated

We used a combination of *in silico* fine-mapping and expression quantitative trait (eQTL) data to identify putatively functional genes implicated by our genetic association signals (**Supplementary Table 2**). Firstly, 81 of the 290 independent ANM signals were highly correlated (minimum r^2=0.8) with one or more variants predicted deleterious for gene function, implicating 91 genes (**Supplementary Table 4**). Twelve of these genes harboured predicted loss-of-function variants and seven genes (*MCM8, EXO1, HELB, C1orf112, C19orf57, FANCM* and *FANCA*) contained multiple statistically independent predicted-deleterious variants (**Supplementary Table 4**). We extended this analysis using exome sequence data from 45,351 women in UKBB. Loss-of-function variants near two highlighted genes were associated with ANM (**Supplementary Table 5**). In aggregate, women carrying loss-of-function variants in *BRCA2* (N=143) and *CHEK2* (N=68) reported ANM 1.54 years earlier (95%CI 0.73-2.34, $P=6.8x10^{-5}$) and 3.49 years later (95%CI 2.36-4.63, $P=1x10^{-13}$) respectively. *BRCA1* loss-of-function was the next most significantly associated GWAS-highlighted gene in these analyses (N=32 LOF carriers, 2.63 years earlier ANM, 95%CI 1.00-4.26, $p=1.1x10^{-4}$). Homozygous loss of function variants in *BRCA2* were recently described as a rare cause of POI[15], but we did not identify any such homozygotes for either *BRCA2*, *CHEK2* or *BRCA1*. Notably, identified GWAS signals mapped within 300kb of 20/74 genes that when disrupted cause primary amenorrhea and/or POI (**Supplementary Table 6**), highlighting the common biological processes shared between normal variation in reproductive ageing and clinical extremes.

Next, we integrated publicly available gene expression data across 44 tissue types with our GWAS results (**Supplementary Table 5**). This highlighted expression-linked genes at 116 of the 290 loci (**Supplementary Tables 2 and 5**). Using three computational approaches we observed enrichment in hematopoietic stem cells and their progenitors (**Supplementary Tables 7–12**). Biological pathway enrichment analyses using a range of approaches, highlighted the importance of DDR processes as the key regulator of ANM (**Supplementary Tables 13–16**). We hypothesise that the shared expression profile in both haematopoietic stem cells and oocytes reflects the relative importance of DDR in both cell types[16]. In contrast to puberty timing[17], which represents the beginning of reproductive life, we observed no enrichment of hypothalamic and pituitary expressed genes, but enrichment of genes expressed in the ovary and other reproductive tissues (**Supplementary Table 9**).

Finally, we attempted to leverage data from multi-tissue co-expression networks to identify genes which sit in the centre of these networks and interact with many other genes near ANM-associated variants. Such genes are analogous to the “core” genes proposed in the omnigenic model of genetic architecture[18]. This approach identified 250 genes, 47 of which were within 300kb of one of the identified 290 loci (**Supplementary Tables 17 and 18**). A notable example is *MCM8*, implicated directly by two missense variants and co-expressed with many genes highlighted by our GWAS (**Extended Data Fig. 5**).

ANM genes act across the life-course

Previous analyses highlighted the involvement of DNA repair in the regulation of ovarian ageing. This study supports a much broader DDR involvement as well as metabolic signaling networks such as PI3K[19] with increased resolution of these pathways and when in the life-course they

might act (**Extended Data Fig. 1, Supplementary Results**). We identify DDR pathways associated with replication stress, Fanconi Anemia pathway, DNA-protein crosslink repair, R loops (**Extended Data Fig. 6**), meiotic recombination and 58 genes implicated in regulation of apoptosis (**Supplementary Table 19**) providing evidence that variation in cell death following DDR is an important mechanism for ANM. This includes components and interactors of the central, conserved DDR checkpoint kinases ATR-CHEK1 (single stranded DNA) and ATM-CHEK2 (double strand breaks) that integrate and determine repair and cellular response from a broad variety of DNA repair pathways (**Extended Data Fig. 6**). The expression patterns across developmental stages in human follicles further supports distinct activities across fetal and follicular stages (**Extended Data Fig. 7**, **Supplementary Table 20**), including *TP63*, which was predominantly expressed during follicular stages, consistent with apoptotic inducing activity in response to DNA damage observed in growing oocytes in mouse[20–23]. These observations are consistent with the DDR regulating both the establishment of the ovarian reserve during fetal life and its depletion until ANM.

In utero effects and maternal diet

Previous work in mice demonstrated that a maternal obesogenic diet during pregnancy decreased ovarian reserve in offspring[24]. We extend this observation by demonstrating that two of our highlighted genes (*Dmc1* and *Brsk1*) are differentially expressed in the offspring ovary due to maternal obesity (**Supplementary Table 5, Extended Data Fig. 8**). *Dmc1* is a meiosis-specific DNA recombinase that assembles at the site of DSBs and is essential for meiotic recombination and gamete formation[25]. Expression levels of *Brsk1* were decreased in ovarian tissue of the offspring of obese mice, an effect which appeared to be enhanced further when the offspring were additionally exposed to an obesogenic diet from weaning (**Extended Data Fig. 8**). *Brsk1* acts as a DNA damage sensor and targets *Wee1* and *Mapt1* for phosphorylation, both of which were also up-regulated in our model. *Wee1* is highly expressed in fetal germ cells, inhibits mitosis and is specifically down-regulated late in oogenesis[26]. The mechanisms linking maternal diet-induced altered expression of these genes to reduced ovarian reserve in the offspring remain unclear. However, our findings, in addition to observations that low birthweight is associated with menopause[27], support the hypothesis that DDR mechanisms acting *in utero* to influence reproductive lifespan may be modifiable by maternal exposures.

Extending reproductive life in animals

Our GWAS highlighted loss of function alleles in *CHEK2* associated with later ANM. Whilst previous work has shown genetic manipulation of DDR genes in animal models limits reproductive lifespan, it remains to be tested whether it can also extend it. *CHEK2* plays a crucial role in culling oocytes in mouse mutants defective in meiotic recombination or after artificial induction of double-strand breaks[22,28,29]. In young females, *Chek2* inactivation can partially rescue oocyte loss and in some mutants, fertility, with high levels of non-physiologically induced endogenous and exogenous DNA damage[23,28,30,31]. To better understand the function of the checkpoint kinase pathways in physiological reproductive ageing, we used genetically modified *Chek1* and *Chek2* mice (**Figure 3**, **Extended Data Fig. 9-11**). Follicular atresia was reduced in *Chek2*$^{-/-}$ females around reproductive senescence (13.5 months). This occurred

without a concomitant increase in the ovarian reserve in young mice (1.5 months) (**Figure 3 a**, **Extended Data Fig. 9 a-e**)[28]. The aged *Chek2*$^{-/-}$ females showed elevated anti-Müllerian hormone levels (**Extended Data Fig. 9 f**) and an increased follicular response to gonadotrophin stimulation (**Figure 3 c, Extended Data Fig. 9 g**) consistent with a larger ovarian reserve at 13.5 months. Fertilization, blastocyst formation and litter sizes in naturally-mated aged *Chek2*$^{-/-}$ females were similar to littermate controls (**Extended Data Fig. 9 h-j**), suggesting that the endogenous damage that *Chek2* responds to does not compromise the health of offspring or mothers in later reproductive life (**Extended Data Fig. 9 j, k**). Thus, depletion of the ovarian reserve is slowed in *Chek2*$^{-/-}$ females, resulting in improved ovarian function around the time of reproductive senescence and suggests a potential therapeutic target for enhancing IVF stimulation through short-term apoptotic inhibition.

In contrast to *Chek2*$^{-/-}$, *Chek1*$^{-/-}$ mice are embryonic lethal due to its essential function when DNA replication is perturbed as well as during mitosis[32]. We found that two different maternal, germline-specific conditional knockouts of *Chek1* (*Chek1* cko), one of which also leads to defects in prospermatogonia in males[33], results in infertility in females due to failure during preimplantation embryo development **(Extended Data Fig. 10)**. *Chek1* is required for prophase I arrest and functions in G2/M checkpoint regulation in murine oocytes[23,34] and its activator, *ATR*, is important for meiotic recombination as well as follicle formation[35,36]. An extra copy, ie. three alleles of murine *Chek1* (*SuperChek1* or *sChek1*), is reported to partially rescue lifespan in *ATR*^{Seckel} mice, suggesting that *CHEK1* becomes rate-limiting when cells are under replication stress[37]. We found that *sChek1* on its own increased the ovarian reserve from birth as well as later in life (**Figure 3 b, Extended Data Fig. 11 b-f**). Large antral follicle counts were also elevated in the aged *sChek1* females, compared to litter-mate controls, indicating that follicular activity was also increased. Immediately prior to the typical age at reproductive senescence, *sChek1* females ovulated an increased number of mature MII oocytes (11-13 months) (**Figure 3 c**, **Extended Data Fig. 11 g**). These exhibited increased mRNA expression of *Chek1* (**Extended Data Fig. 11 a**) and had similar capacity for forming blastocyst embryos as wild type (**Extended Data Fig. 11 i, j**). When transferred, these embryos gave rise to healthy, fertile pups over two generations (**Extended Data Fig. 11 k-n**). Thus, *sChek1* causes a larger ovarian reserve to be established at birth and the oocytes appear to maintain their genomic integrity, as confirmed by aneuploidy analysis and efficiency of embryogenesis and fertility of pups (**Extended Data Fig. 11 g-n**), resulting in enhanced follicular activity and delayed reproductive senescence. We speculate that this is due to upregulation of replication-associated DNA repair processes during mitosis and meiosis and that repair might be limiting for establishing and maintaining the ovarian reserve. Taken together, our data show that modulating key DDR genes can extend reproductive lifespan *in vivo*, generating healthy pups that are fertile over several generations. This can occur either by abolishing DDR checkpoints (*Chek2* deletion) or by upregulating repair processes (*sChek1*).

Health consequences of later ANM

We used our identified genetic variants to infer causal relationships, using a Mendelian Randomization (MR) framework, between ANM and several health outcomes (**Supplementary Tables 21–23**). Consistent with previous studies[2,13], each 1-year genetically-mediated later

ANM increased the relative risks of several hormone-sensitive cancers by up to 5% (**Supplementary Table 21**). In contrast, we observed beneficial effects of genetically-mediated later ANM on bone mineral density, fracture risk and type 2 diabetes**.** Our findings are consistent with evidence from randomised controlled trials that oestrogen therapy maintains bone health and protects from type 2 diabetes[38,39]. Furthermore, recent MR studies demonstrate causal associations between sex hormone levels and type 2 diabetes[40]. Trial data in younger women taking HRT suggested no increased risk of cardiovascular disease, stroke or all-cause mortality[39]. In agreement with this we found no evidence to support causal associations for ANM with cardiovascular disease, lipid levels, Alzheimer's disease, body mass or longevity (**Supplementary Table 21**), all of which have been reported in observational studies[41–47]. Finally, we evaluated putative modifiable determinants of ANM reported by observational studies[27]**.** We found that genetically instrumented increased alcohol consumption and tobacco smoking were associated with earlier ANM (**Supplementary Tables 24 and 25**). Each additional cigarette smoked per day decreased ANM by ~2.5 weeks, whilst women who drank alcohol at the maximum recommended limit experienced ~1 year earlier menopause compared to those who drank little. Furthermore, genetically instrumented age at menarche was associated with ~8 weeks earlier ANM per-year earlier menarche.

Collectively our analyses have provided novel insights into the biological processes underpinning reproductive ageing in women, how they can be manipulated to extend reproductive life, and what the consequence of this might be at a population level. We anticipate these findings will greatly inform experimental studies seeking to identify new therapies for enhancement of reproductive function and fertility preservation in women.

References

1. Lambalk, C. B., van Disseldorp, J., de Koning, C. H. & Broekmans, F. J. Testing ovarian reserve to predict age at menopause. *Maturitas* **63**, 280–91 (2009).

2. Collaborative Group on Hormonal Factors in Breast Cancer. Type and timing of menopausal hormone therapy and breast cancer risk: individual participant meta-analysis of the worldwide epidemiological evidence. *Lancet (London, England)* **394**, 1159–1168 (2019).

3. Murray, A. *et al.* Population-based estimates of the prevalence of FMR1 expansion mutations in women with early menopause and primary ovarian insufficiency. *Genet. Med.* **16**, 19–24 (2014).

4. Christensen, K., Doblhammer, G., Rau, R. & Vaupel, J. W. Ageing populations: the challenges ahead. *Lancet (London, England)* **374**, 1196–208 (2009).

5. InterLACE Study Team. Variations in reproductive events across life: a pooled analysis of data from 505 147 women across 10 countries. *Hum. Reprod.* **34**, 881–893 (2019).

6. Gruhn, J. R. *et al.* Chromosome errors in human eggs shape natural fertility over reproductive life span. *Science* **365**, 1466–1469 (2019).

7. Donnez, J. & Dolmans, M.-M. Fertility Preservation in Women. *N. Engl. J. Med.* **377**, 1657–1665 (2017).

8. Yding Andersen, C., Mamsen, L. S. & Kristensen, S. G. FERTILITY PRESERVATION: Freezing of ovarian tissue and clinical opportunities. *Reproduction* **158**, F27–F34 (2019).

9. Argyle, C. E., Harper, J. C. & Davies, M. C. Oocyte cryopreservation: where are we now? *Hum. Reprod. Update* **22**, 440–9 (2016).

10. Stolk, L. *et al.* Meta-analyses identify 13 loci associated with age at menopause and highlight DNA repair and immune pathways. *Nat. Genet.* **44**, 260–8 (2012).

11. Venturella, R. *et al.* The Genetics of Non-Syndromic Primary Ovarian Insufficiency: A Systematic Review. *Int. J. Fertil. Steril.* **13**, 161–168 (2019).

12. Titus, S. *et al.* Impairment of BRCA1-related DNA double-strand break repair leads to ovarian aging in mice and humans. *Sci. Transl. Med.* **5**, 172ra21 (2013).

13. Day, F. R. *et al.* Large-scale genomic analyses link reproductive aging to hypothalamic signaling, breast cancer susceptibility and BRCA1-mediated DNA repair. *Nat. Genet.* **47**, 1294–303 (2015).

14. Horikoshi, M. *et al.* Elucidating the genetic architecture of reproductive ageing in the Japanese population. *Nat. Commun.* **9**, 1977 (2018).

15. Caburet, S. *et al.* Homozygous hypomorphic BRCA2 variant in primary ovarian insufficiency without cancer or Fanconi anaemia trait. *J. Med. Genet.* (2020) doi:10.1136/jmedgenet-2019-106672.

16. Thompson, D. J. *et al.* Genetic predisposition to mosaic Y chromosome loss in blood. *Nature* **575**, 652–657 (2019).

17. Day, F. R. *et al.* Genomic analyses identify hundreds of variants associated with age at menarche and support a role for puberty timing in cancer risk. *Nat. Genet.* **10**, 1–19 (2017).

18. Boyle, E. A., Li, Y. I. & Pritchard, J. K. An Expanded View of Complex Traits: From Polygenic to Omnigenic. *Cell* **169**, 1177–1186 (2017).

19. Reddy, P. *et al.* Oocyte-specific deletion of Pten causes premature activation of the primordial follicle pool. *Science* **319**, 611–3 (2008).

20. Suh, E.-K. *et al.* p63 protects the female germ line during meiotic arrest. *Nature* **444**, 624–8 (2006).

21. Deutsch, G. B. *et al.* DNA damage in oocytes induces a switch of the quality control factor TAp63α from dimer to tetramer. *Cell* **144**, 566–76 (2011).

22. Tuppi, M. *et al.* Oocyte DNA damage quality control requires consecutive interplay of CHK2 and CK1 to activate p63. *Nat. Struct. Mol. Biol.* **25**, 261–269 (2018).

23. Rinaldi, V. D., Bloom, J. C. & Schimenti, J. C. Oocyte Elimination Through DNA Damage Signaling from CHK1/CHK2 to p53 and p63. *Genetics* **215**, 373–378 (2020).

24. Aiken, C. E., Tarry-Adkins, J. L., Penfold, N. C., Dearden, L. & Ozanne, S. E. Decreased ovarian reserve, dysregulation of mitochondrial biogenesis, and increased lipid peroxidation in female mouse offspring exposed to an obesogenic maternal diet. *FASEB J.* **30**, 1548–56 (2016).

25. Pittman, D. L. *et al.* Meiotic prophase arrest with failure of chromosome synapsis in mice deficient for Dmc1, a germline-specific RecA homolog. *Mol. Cell* **1**, 697–705 (1998).

26. Nakajo, N. *et al.* Absence of Wee1 ensures the meiotic cell cycle in Xenopus oocytes. *Genes Dev.* **14**, 328–38 (2000).

27. Ruth, K. S. *et al.* Events in Early Life are Associated with Female Reproductive Ageing: A UK Biobank Study. *Sci. Rep.* **6**, 24710 (2016).

28. Bolcun-Filas, E., Rinaldi, V. D., White, M. E. & Schimenti, J. C. Reversal of female infertility by Chk2 ablation reveals the oocyte DNA damage checkpoint pathway. *Science* **343**, 533–536 (2014).

29. Adhikari, D. *et al.* Inhibitory phosphorylation of Cdk1 mediates prolonged prophase I arrest in female germ cells and is essential for female reproductive lifespan. *Cell Res.* **26**, 1212–1225 (2016).

30. Rinaldi, V. D., Bolcun-Filas, E., Kogo, H., Kurahashi, H. & Schimenti, J. C. The DNA Damage Checkpoint Eliminates Mouse Oocytes with Chromosome Synapsis Failure. *Mol. Cell* **67**, 1026-1036.e2 (2017).

31. Tharp, M. E., Malki, S. & Bortvin, A. Maximizing the ovarian reserve in mice by evading LINE-1 genotoxicity. *Nat. Commun.* **11**, 330 (2020).

32. Liu, Q. *et al.* Chk1 is an essential kinase that is regulated by Atr and required for the G(2)/M DNA damage checkpoint. *Genes Dev.* **14**, 1448–59 (2000).

33. Abe, H. *et al.* CHEK1 coordinates DNA damage signaling and meiotic progression in the male germline of mice. *Hum. Mol. Genet.* **27**, 1136–1149 (2018).

34. Chen, L. *et al.* Checkpoint kinase 1 is essential for meiotic cell cycle regulation in mouse oocytes. *Cell Cycle* **11**, 1948–55 (2012).

35. Pacheco, S. *et al.* ATR is required to complete meiotic recombination in mice. *Nat. Commun.* **9**, 2622 (2018).

36. Pacheco, S., Maldonado-Linares, A., Garcia-Caldés, M. & Roig, I. ATR function is indispensable to allow proper mammalian follicle development. *Chromosoma* **128**, 489–500 (2019).

37. López-Contreras, A. J., Gutierrez-Martinez, P., Specks, J., Rodrigo-Perez, S. & Fernandez-Capetillo, O. An extra allele of Chk1 limits oncogene-induced replicative stress and promotes transformation. *J. Exp. Med.* **209**, 455–61 (2012).

38. Salpeter, S. R. *et al.* Meta-analysis: effect of hormone-replacement therapy on components of the metabolic syndrome in postmenopausal women. *Diabetes. Obes. Metab.* **8**, 538–54 (2006).

39. Manson, J. E. *et al.* Menopausal hormone therapy and health outcomes during the intervention and extended poststopping phases of the Women's Health Initiative randomized trials. *JAMA* **310**, 1353–68 (2013).

40. Ruth, K. S. *et al.* Using human genetics to understand the disease impacts of testosterone in men and women. *Nat. Med.* **26**, 252–258 (2020).

41. Dam, V. *et al.* Association of menopausal characteristics and risk of coronary heart disease: a pan-European case-cohort analysis. *Int. J. Epidemiol.* **48**, 1275–1285 (2019).

42. de Kat, A. C. *et al.* Unraveling the associations of age and menopause with cardiovascular risk factors in a large population-based study. *BMC Med.* **15**, 2 (2017).

43. Atsma, F., Bartelink, M.-L. E. L., Grobbee, D. E. & van der Schouw, Y. T. Postmenopausal status and early menopause as independent risk factors for cardiovascular disease: a meta-analysis. *Menopause* **13**, 265–79.

44. Ambikairajah, A., Walsh, E. & Cherbuin, N. Lipid profile differences during menopause: a review with meta-analysis. *Menopause* **26**, 1327–1333 (2019).

45. Pike, C. J. Sex and the development of Alzheimer's disease. *J. Neurosci. Res.* **95**, 671–680 (2017).

46. Zhu, D. *et al.* Body mass index and age at natural menopause: an international pooled

analysis of 11 prospective studies. *Eur. J. Epidemiol.* **33**, 699–710 (2018).

47. Shadyab, A. H. *et al.* Ages at menarche and menopause and reproductive lifespan as predictors of exceptional longevity in women: the Women's Health Initiative. *Menopause* **24**, 35–44 (2017).

Figure legends

Figure 1. Manhattan plot representing GWAS discovery analysis. Previously identified loci in purple, novel loci in blue. Plotted variants have $P<0.01$ with $P<1x10^{-300}$ truncated. **Insert**: Effect sizes and minor allele frequencies of the loci. LOF, loss of function

Figure 2. Polygenic prediction of age at menopause. a, Mean polygenic score (PGS; scaled to have mean=0, SD=1) for a given age at natural menopause (ANM). Higher PGS indicates later ANM. **b, c,** Association of each centile of PGS vs the 50th with, **b,** early menopause and, **c,** premature ovarian insufficiency. Higher PGS indicates earlier ANM.

Figure 3. Genetic manipulation of *Chek1* or *Chek2* extends reproductive lifespan in mouse models. Numbers of follicles in young and aged, **a**, *Chek2*$^{-/-}$ or, **b**, *sChek1* females. Numbers of ovaries analysed in parentheses. **c**, Response to gonadotrophin stimulation of 13.5-month-old *Chek2*$^{-/-}$ and *sChek1* females assessed by the number of MII oocytes retrieved. Numbers of stimulated females in parentheses. Box-and-whisker plots show interquartile range and median. Two-sample t and Fisher's exact tests used for comparisons: *, $P<0.05$; **, $P<0.025$; ***, $P<0.001$.

Figure 1

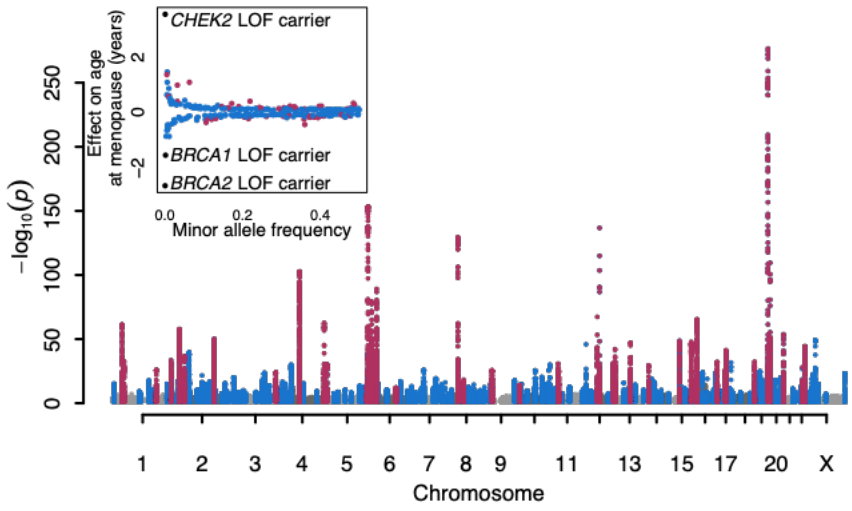


Figure 2

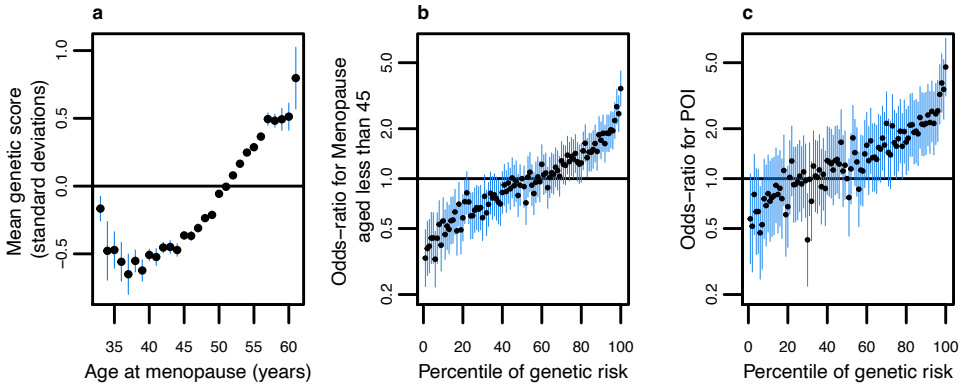
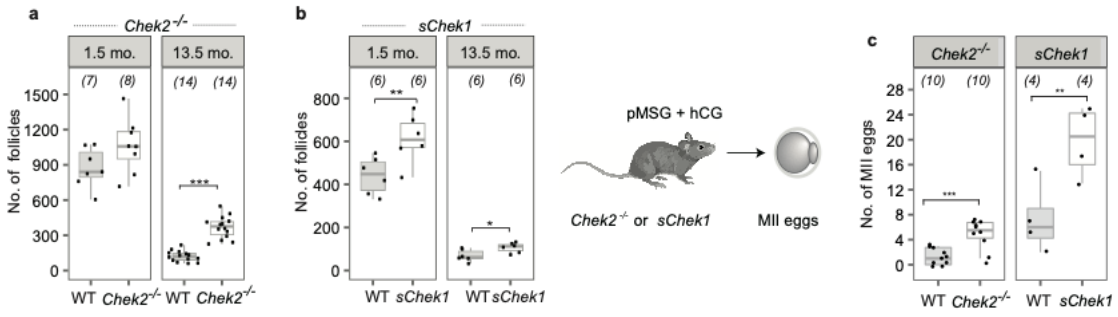


Figure 3



ONLINE METHODS

Information on ethical regulations and approvals for all animal experiments are detailed in the corresponding sections below. Within each of the human population studies included in the genome-wide analyses (all of which have been previously published), each participant provided informed consent and the study protocol was approved by the institutional review board at the parent institution.

Phenotype definition

We included women with age at natural menopause (ANM) from age 40 to 60 inclusive. ANM was derived from self-reported questionnaire data by each study (**Supplementary Table 1**) and was the age at last naturally occurring menstrual period followed by at least 12 consecutive months of amenorrhea. Exclusions were women with menopause caused by hysterectomy, bilateral ovariectomy, radiation or chemotherapy, and those using HRT before menopause. Within each of the studies, each participant provided written informed consent and the study protocol was approved by the institutional review board at the parent institution.

Genome-wide association study meta-analysis

A genome-wide meta-analysis of autosomal and chromosome X variants in women of European ancestry was carried out on summary statistics from analyses in three strata, allowing for the identification of heterogeneity due to different methodology. The three strata were (**Extended Data Fig. 2**): (i) meta-analysis of 1000 Genomes imputed studies; (ii) meta-analysis of samples from the Breast Cancer Association Consortium (BCAC: http://bcac.ccge.medschl.cam.ac.uk); (iii) UK Biobank GWAS. The overall meta-analysis included variants present in at least two of the three strata. All meta-analyses were inverse-variance weighted without GC correction and were carried out in METAL (https://genome.sph.umich.edu/wiki/METAL_Documentation). Analysis was conducted by analysts and two geographically distinct sites independently and the resulting summary statistics were compared for consistency.

The meta-analysis of 1000 Genomes imputed studies included 40 datasets imputed to 1000 Genomes Phase I version 3 for the autosomes and 29 for chromosome X (**Supplementary Table 1, Supplementary Notes**). Each individual study applied quality control to directly genotyped variants and samples prior to imputation (suggested exclusion thresholds for variants were Hardy-Weinberg equilibrium $P<1\times10^{-5}$, call rate <95% and minor allele frequency (MAF) <1%; suggested exclusions for samples were >5% missing genotypes, population outliers, high inbreeding coefficient, heterozygosity outliers, sex mismatches and related samples). Each individual study carried out GWAS using a two-tailed additive linear regression model adjusted for genetic principal components/relationship matrix depending on the software used (**Supplementary Table 1**), without GC correction. Since all samples included were female, chromosome X was analysed as for the autosomes. Once data were submitted, each study underwent quality control centrally according to standard protocols implemented independently by two analysts. Summary statistics for each study were stored centrally. Prior to meta-analysis,

genetic variants ids were converted to “chr:position” format (position in build 37) and alleles for insertion/deletion polymorphisms were coded as “I/D” to ensure consistency across studies. Meta-analysis was carried out including SNPs with imputation quality≥0.4 and MAF≥0.001. Variants in at least half of datasets for either the autosomes or for chromosome X (as appropriate) were taken forward to the overall meta-analysis, resulting in ~10.9 million variants.

GWAS summary statistics for the BCAC data were provided as four datasets, containing breast cancer cases and controls, with each genotyped on the iCOGs and OncoArray genotyping arrays (**Supplementary Table 1**). Quality control was applied to directly genotyped variants prior to imputation and data were imputed to the HRC r1.1 (2016) reference panel. Association analysis and quality control was carried out centrally as for the 1000 Genomes imputed studies. Summary statistics from the four BCAC datasets were meta-analysed, including variants with imputation quality≥0.4 and MAF≥0.001. Variants in two or more of the four datasets were taken forward to the overall meta-analysis, resulting in ~14.5 million variants.

UK Biobank genotyped 488,377 participants on two arrays, 49,950 on the UK BiLEVE Axiom array (807,411 markers) and 438,427 on the UK Biobank Axiom array (825,927 markers), which were then imputed using a combined 1000 Genomes Phase 3 and HRC reference panel. Details of central genotyping, quality control and imputation are described elsewhere[48]. We included 451,454 individuals identified as European in our analysis. Briefly, principal components analyses were used to cluster individuals of White European descent (described more fully elsewhere[49]). We further removed participants who had subsequently withdrawn from the study (n=7) and those where their self-reported sex did not match their genetic sex (n=348) resulting in 451,099 individuals. GWAS was carried out in 106,048 women with ANM by applying a linear mixed model in BOLT-LMM[50] to adjust for population structure and relatedness, also adjusting for study centre and data release. Summary statistics taken forward to the overall meta-analyses were for ~16.6 million variants with imputation quality ≥0.5 and MAF≥0.001. UK Biobank data were analysed by two analysts independently and summary statistics results were compared for consistency.

Genome-wide significance was set at $P<5x10^{-8}$. Statistical independence was determined using a combination of two approaches. Firstly, we used distance-based clumping to select the most significantly associated SNP within a 1Mb window. Secondly, we augmented this list with secondary signals within these 1Mb windows that were identified through approximate conditional analysis implemented in GCTA[51]. We only considered secondary signals that were uncorrelated with other selected signals ($r^2<0.05$) and genome-wide significant in both

univariate and joint models. 10,000 ancestry matched samples from UK Biobank were used in GCTA as an LD reference panel.

Assessing the impact of time to event models on the signals identified

We performed Cox proportional hazards regression for the 290 genome-wide significant ANM signals, allowing inclusion in our analyses of women excluded from the definition of natural menopause. We used UK Biobank imputed genotype data and performed analyses in 379,768 unrelated individuals of European descent (as described previously), of whom 185,293 were included in our Cox analyses (phenotype definition as described previously[27]). Briefly, Cox proportional hazards regression was run using stset and stcox (Breslow method for ties) in Stata v16.0 using age as the time variable, starting at birth (0 years) and ending at last age at risk of natural menopause. Natural menopause was set as the event, with individuals censored at bilateral oophorectomy and/or hysterectomy, or start of HRT use (if ongoing at time of menopause, hysterectomy or oophorectomy). We included the covariates genotyping chip and release of genotype data, recruitment centre and the first five genetic principal components, which were considered to be constant throughout the time at risk. We calculated -1 × natural log(hazard ratio) to allow comparison with effect estimates from linear regression from the full meta-analysis and meta-analysis excluding UK Biobank.

Confirmation of identified signals and variance explained estimates

We sought to confirm our findings by testing the 290 identified loci in an independent sample of 294,828 women from 23andMe. Participants provided informed consent and participated in the research online, under a protocol approved by the external AAHRPP-accredited IRB, Ethical & Independent Review Services (E&I Review). Participants were included in the analysis on the basis of consent status as checked at the time data analyses were initiated.The variant-level data for the 23andMe replication dataset are fully disclosed in the manuscript. Individual-level data are not publicly available due participant confidentiality, and in accordance with the IRB-approved protocol under which the study was conducted. Women's age at menopause was ascertained across multiple surveys using two questions: "About how old were you when you had your last menstrual period? (under_30/30_34/45_49/40_44/55+/50_54/35_39/declined/not_sure)" and "How old were you when you had your last menstrual period?". As menopause age was ascertained in 4-year bins we rescaled the effect estimates appropriately to be on the same 1-year scale as our discovery analyses. Analyses were performed using a linear model (gaussian family), controlling for age (in years), the top 5 genetic principal components and genotyping platform.

To assess the relevance of these loci in women of East Asian ancestry, we meta-analysed data (total N=78,317 women) from the China Kadoorie Biobank study and Biobank Japan (BBJ). A total of 47,140 female participants in BBJ whose age at menopause was available were included in the current study. If different ages at menopause were reported in multiple visits, we took mean of ages at menopause. We excluded individuals 1) with maximum difference more than five years in the reported ages at menopause on multiple visits; 2) whose age at recruitment was younger than reported age at menopause; 3) whose age at menopause was younger than 40 or older than 60 years, or 4) with medical history of hysterectomy, ovariectomy, radiation, chemotherapy and hormone replacement treatment before age at menopause.

Subjects 1) whose DNA microarray data was not available, 2) with low call rate (<0.98), 3) whose genetic data suggested as male, 4) who were genetically identical to other subjects or 5) who were outliers from EAS cluster in PCA plot were excluded from the analyses. We applied the same quality control for variants as the previous literature[52]. After quality control, remaining variants were phased and subsequently imputed onto the reference panel containing the 1000 Genomes Project Phase 3 and around 3,000 Japanese whole-genome sequence data[52]. We restricted subsequent analyses to variants with rsq >0.3. For an association study of age at menopause, we applied a linear mixed model using BOLT-LMMv2.3.4 software correcting for age in years and the top ten genetically determined principal components as covariates.

The China Kadoorie Biobank baseline survey was conducted during 2004-2008 in 10 geographically diverse regions of China (5 rural, 5 urban), with resurveys of approximately 5% of the cohort at 5-yearly intervals. 302632 women aged 35-74 years were enrolled with a mean age at baseline of 51.4 (SD 10.5), of whom 162,929 provided at least one reported age at menopause, in response to the questions “Have you had your menopause? If so, age of completion of menopause?”, with mean (SD) of 48.2 (4.4) years. Genotyping data was available for 31,177 women with values for age at menopause in the range 35-60 years and who had not had prior hysterectomy, oophorectomy, or cancer. Genotyping used custom Affymetrix Axiom® arrays with imputation into the 1000 Genomes Phase 3 reference using SHAPEIT3 and IMPUTE4 (IMPUTE2 for chrX). Age at menopause was adjusted for year-of-birth and year-of-birth-squared, and analyses were carried out separately for each of the 10 recruitment regions using BOLT-LMM v2.3.2 followed by inverse-variance-weighted fixed effect meta-analysis in METAL. Analyses used CKB data release 15.

The variance explained by our identified signals were estimated in a further independent sample of 16,556 women from the Icelandic deCODE study. Of those women, 14,771 were chip-typed and 1,785 are imputed 1st and 2nd degree relatives of chip-typed individuals. We assessed the aggregate significance of the identified loci by testing how many alleles had the same direction of effect using a binomial sign test (null expectation 50%). The proportion of variance explained using replication summary statistics provided by deCODE (n=16,556). We calculated the variance explained by each variant in deCODE (using the formula $2\times\beta^2\times MAF\times(1-MAF)$), dividing the sum of the variance explained in total for the 290 variants by the SE^2 of menopause age in deCODE.

We additionally estimated the proportion of variance in ANM explained by the 290 genome-wide significant signals in UK Biobank by calculating linear regression R^2 in 88,829 unrelated women of European descent (as described previously[49]) who had menopause age recorded. We generated estimates by combining the 290 variants as a genetic risk score with the allelic dosage weighted by the effect size from meta-analysis of the 1KG and BCAC strata only (**Supplementary Table 2**). Genotypes were extracted from imputed data and we included the covariates genotyping chip and release of genotype data, recruitment centre, age and the first five genetic principal components. Genotype-array heritability estimates were calculated using REML implemented in BOLT-LMM to provide a denominator for proportion of heritability explained.

Assessing deviation from an additive genetic model

A dominance deviation test[53] was run for the 290 genome-wide significant ANM signals. Briefly, in this test a dominance deviation term representing the heterozygous group (coded 0, 1 and 0) is fitted jointly with an additive genotype term in the regression model. This test determines whether the average trait value carried by the heterozygous group lies halfway between the two homozygote groups as expected under an additive model. We used best guess genotypes converted from UK Biobank imputed genotype data and performed linear regression analysis in Stata v16.0 in 379,768 unrelated individuals of European descent (identified as described previously[49]. We regressed ANM on genotype including the covariates genotyping chip and release of genotype data, recruitment centre and the first five genetic principal components. We also tested a dominant model, comparing the effect allele heterozygotes/homozygote group with other allele homozygotes, and a recessive model, comparing effect allele homozygotes with heterozygotes and other allele homozygotes. Genetic variants with a P-value for the dominance deviation term that was smaller than Bonferonni corrected P=0.05 (P=0.05/290=0.000172) were considered to show evidence of non-additive effects.

Gene burden analyses of UK Biobank exome sequencing data

We carried out gene burden association testing of rare variants in women identified from ~200K people with exome sequencing data available in the UK Biobank study. We included 45,351 women with ANM between 18–65 years in our analyses to maximise the sample size and ensure inclusion of women with POI who might be expected to be more likely to be carriers of rare variants.

Detailed sequencing methodology is provided by Szustakowski et al[54]. Briefly, exomes were captured with the IDT xGen Exome Research Panel v1.0 which targeted 39Mbp of the human genome with coverage exceeding on average 20x on 95.6% of sites. The OQFE protocol was used for mapping and variant calling to the GRCh38 reference. Variants included in our analyses had individual and variant missingness <10%, Hardy Weinberg Equilibrium p-value $>10^{-15}$, minimum read depth of 7 for SNPs and 10 for indels, and at least one sample per site passed the allele balance threshold > 15% for SNPs and 20% for indels.

Variants in CCDS transcripts were annotated using Variant Effect Predictor[55]. We identified loss-of-function (LoF) variants (stop-gain, frameshift, or abolishing a canonical splice site (-2 or +2 bp from exon, excluding the ones in the last exon)) deemed to be high confidence by LOFTEE (https://github.com/konradjk/loftee). We conducted gene-burden analyses using a SKAT-O test implemented in SAIGE-GENE[56] based on variants with MAF<0.001. SAIGE-GENE implements a generalized mixed-model region-based association test that can account for population stratification and sample relatedness in large-scale analyses. We applied an inverse normal rank transformation to ANM prior to analyses and included recruitment centre as a covariate. For each gene, we present results for the transcript with the smallest SKAT-O p-value. Since the magnitude of effect sizes from SAIGE-GENE are not easily interpretable, we calculated the sum of LoF alleles in *BRCA1*, *BRCA2* and *CHEK2* for each person. We tested each score's association with ANM by performing linear regression in Stata v16.0 in unrelated samples of European descent (identified as described previously [PMID: 30423117]) including recruitment centre and the first five genetic principal components as covariates.

Identifying putatively functional genes

We used two in silico approaches to prioritise putatively functional genes across our highlighted loci. Firstly, To identify variants with functional consequences, we looked up variants in r^2>0.8 with the signals in Variant Effect Predictor (build 38). We identified missense, frameshift, insertion/deletions and stop-gained and splice site disrupting variants, which we then classified according to their VEP, PolyPhen and SIFT impact. We considered 'high impact' variants as those classified as high impact by VEP (stop-gained, frameshift and splice site disrupting). 'Medium impact' variants were missense variants classed as moderate impact by VEP, which were either deleterious in SIFT and were at least possibly damaging in PolyPhen. 'Low impact' variants were missense or inframe insertions/deletions classed as moderate impact by VEP and were tolerated and/or benign in PolyPhen. LD was calculated using PLINK v1.9 from best guess genotypes for 1000 Genomes Phase 3/HRC imputed variants in ~340,000 unrelated UK Biobank participants of white British ancestry. Genetic variant locations were converted from b37 to b38 using UCSC Liftover.

Secondly, we integrated our ANM genome-wide summary statistics with eQTL data using Summary Mendelian Randomization (SMR)[57]. Publicly available expression datasets for 48 tissues in GTEx v7 and 10 brain regions were downloaded from the SMR website (https://cnsgenomics.com/software/smr/#eQTLsummarydata). Whole-blood data in an eQTL meta-analysis of 31,684 samples was available from the eQTLGen consortium [https://www.biorxiv.org/content/10.1101/447367v1] A Bonferroni corrected p-value threshold was used in each expression dataset individually and only associations with HEIDI P > 0.01 were considered to avoid coincidental overlap due to extended patterns of LD. This resulted in a total of 44 (SMR $P<7x10^{-6}$) significant transcriptions in the brain, 96 in whole blood ($P<3x10^{-6}$) and 732 across all GTEx tissues (SMR P<3.6x10-7). We excluded brain and whole blood tissues from the collection of 48 tissues in GTEx as they were better represented by the other expression datasets.

Identifying enriched cell and tissue types

We used three approaches to identify cell and tissue types enriched for ANM associated variants. DEPICT was run using default settings as described previously[58] using GWAS summary statistics including all autosomal variants with P-value $<1x10^{-5}$. The cell-type specific expression matrices used as input to DEPICT were generated from individual single-cell gene expression datasets (see below). Briefly, each data set was processed by first normalizing cells's gene expression to a common transcript count (10,000 transcript per cell) before calculating the average expression of each gene for each cell-type annotation. Averaged data was log-transformed (natural log). We computed cell-type specific gene expression following using a two-step z-score approach - first we calculated gene-wise z-scores (each gene; mean=0, sd=1) to remove the effect of ubiquitous expressed genes, then we calculated cell-type-wise z-scores (each cell-type; mean=0, sd=1) on gene-wise z-scores. For mouse expression datasets we mapped mouse genes to human orthologs using Ensembl (v. 91) keeping only genes with a 1-1 ortholog mapping.

DEPICT analyses were run on two datasets: 1) Tabula Muris (https://tabula-muris.ds.czbiohub.org/)[59], restricted to the fluorescence-activated cell sorting samples. To keep the tissue level information in the dataset, we defined cell-type annotations as 'tissue cell-types' by combining the cell-type label ('cell_ontology_class' column) with the origin tissue of the cell-type ('tissue' column). This allowed us to e.g. distinguish B-cells originating from fat, spleen and marrow tissue. In total we analyzed 115 cell-type annotations from 44,949 cells; 2) Nestorowa et

al. human hematopoietic stem and progenitor cell differentiation dataset[60] was not normalized to a common transcript count because the data was pre-normalized by the authors. We defined cell-type annotations as the 12 distinct hematopoietic stem and progenitor cell (HSPC) phenotypes reported by the authors (shown in their manuscript Figure 3A). The annotations covered 1,483 cells.

Secondly, we additionally performed tissue enrichment analysis using linkage-disequilibrium (LD) score regression to specifically expressed genes (LDSC-SEG)[61]. We used three datasets available on the LDSC-SEG resource page (https://github.com/bulik/ldsc/wiki/Cell-type-specific-analyses), relating to cell and tissue-specific annotations from GTEx[62], Epigenome Roadmap[63] and the "Franke lab"[58,64].

Finally, tissue enrichment analyses were performed using 'Downstreamer', which is described in a separate section below.

Pathway analysis

MAGENTA was used to explore pathway-based associations in the full GWAS data set. MAGENTA implements a gene set enrichment analysis (GSEA)-based approach[65]. We used upstream and downstream limits of 110Kb and 40Kb to assign variants to genes, excluded the HLA region from the analysis and set the number of permutations to 10,000 for GSEA testing, with analysis using 75% and 95% cut-offs. Significance was determined when an individual pathway reached FDR<0.05 in either analysis. In total, 3,222 pathways from Gene Ontology, PANTHER, KEGG and Ingenuity were tested for enrichment of multiple modest associations with ANM.

We additionally performed pathway analyses in 'Downstreamer' (described in section below) and MAGMA[66] v1.08. MAGMA analyses were performed using the full genome-wide summary statistics, but restricted to variants that were predicted deleterious (i.e non-synonymous and loss of function). Gene-sets included in the analyses were obtained from MsigDB v7.2, which included 12,358 curated gene sets from KEGG, Reactome, BioCarta and GO terms consisting of biological processes, cellular components and molecular functions.

Downstreamer methodology

In short, Downstreamer identifies genes connected to genes at GWAS loci (core genes) through expression and identifies enriched pathways. Downstreamer implements a strategy that accounts for LD structure and chromosomal organization, operating in two steps. In the first step, gene-level prioritization scores are calculated for the GWAS trait and a null distribution. In the second step, the gene-level prioritization scores are associated with the co-regulation matrix and pathway annotations. Further details are outlined below.

Downstreamer step 1

Calculation of gene-level prioritization scores (GWAS gene Z-scores)

The primary step is to convert GWAS summary statistics from p-values per variant to an aggregate p-value per gene (gene p-value) while accounting for local LD structure. This aggregate gene level p-value represents the GWAS signal potentially attributable to that gene.

First, we applied genomic control to correct for inflation in the GWAS signal. We then integrated the procedure from the PASCAL[67] method into Downstreamer to aggregate variant p-values into a gene p-value while accounting for the LD structure. We aggregated all variants within a 25kb window around the start and end of a gene using the non-Finnish European samples of the 1000 Genomes (1000G) project, Phase 3 to calculate LD [26432245]. We calculated these GWAS gene p-values for all 20,327 protein-coding genes (Ensembl release v75). The gene p-values were then converted to Z-scores for use in subsequent analysis. These are referred to as GWAS gene Z-scores.

Calculation of gene Z-scores for null GWASs to account for chromosomal organization of genes and to calculate empirical p-values.

To account for long range effects of haplotype structure which results in genes getting a similar GWAS gene Z-score, we use a generalized least squares (GLS) regression model for all regressions done in Downstreamer. The GLS model takes a correlation matrix that models this gene-gene correlation.

To calculate this correlation matrix we first simulated 10,000 random phenotypes by drawing phenotypes from a normal distribution and then associating them to the genotypes of the 1000G Phase 3 non-Finnish European samples. We used only overlapping variants between the real traits and the permuted GWASs to avoid biases introduced by genotyping platforms or imputation. We then calculated the GWAS gene Z-scores for each of the 10,000 simulated GWAS signals as described above. Next, we calculated the Pearson correlations between the GWAS gene Z-scores. As simulated GWAS signals are random and independent of each other, any remaining correlation between GWAS gene Z-scores reflects the underlying LD patterns and chromosomal organization of genes.

We simulated an additional 10,000 GWASs as described above to empirically determine enrichment p-values and, finally, we used an additional 100 simulations to estimate the false discovery rate (FDR) of Downstreamer associations.

Downstreamer step 2

Calculation of Z-scores for co-regulation matrix

To calculate core scores, we used a previously generated co-regulation matrix that is based on a large multi-tissue gene network[68]. In short, publicly available RNA-seq samples were downloaded from the European Nucleotide Archive (https://www.ebi.ac.uk/ena). After QC, 56,435 genes and 31,499 samples covering a wide range of human cell-types and tissues remained. We performed a PCA on this dataset and selected 165 components representing 50% of the variation that offered the best prediction of gene function. We then selected the protein coding genes and centred and scaled the eigenvectors for these 165 components (mean = 0, standard deviation = 1) such that each component was given equal weight. The first components mostly describe tissue differences[68], so this normalization ensures that tissue-specific-patterns do not disproportionately drive the co-regulation matrix. The co-regulation matrix is defined as the Pearson correlation between the genes from the scaled eigenvector matrix. The diagonal of the co-regulation matrix was set to zero to avoid the correlation with

itself having a disproportionate effect on the association to the GWAS gene Z-scores. Finally, we converted the Pearson r to Z-scores.

Calculation of Z-scores for pathways and gene sets

To identify pathway and disease enrichments, we used the following databases: Human Phenotype Ontology (HPO), Kyoto Encyclopaedia of Genes and Genomes (KEGG), Reactome and Gene Ontology (GO) Biological Process, Cellular Component and Molecular Function. We have previously predicted how much each gene contributes to these gene sets, resulting in a Z-score per pathway or term per gene[68]. We collapsed genes into meta-genes in parallel with the GWAS step, to ensure compatibility with the GWAS gene Z-scores following the same procedure as in the GWAS pre-processing. Meta-gene Z-scores were calculated as the Z-score sum divided by the square root of the number of genes. Finally, all pathway Z-scores were scaled (mean = 0, standard deviation = 1).

Pre-processing of GWAS gene Z-scores and pruning of highly correlated genes

For each GWAS, both real and simulated, we carried out rank-based inverse normal transformation of GWAS Z-scores to ensure that outliers would not have disproportionate weights. Due to limitations in the PASCAL methodology that result in ties at a minimum significance level of $1x10^{-12}$ for highly significant genes, we used the minimum SNP P-value from the GWAS to identify the most significant gene and resolve the tie. We then used a linear model to correct for gene length, as longer genes will typically harbour more SNPs.

Sometimes, two (or more) genes will be so close to one another that their GWAS gene Z-scores are highly correlated, violating the assumptions of the linear model. Thus, genes with a Pearson correlation $r \geq 0.8$ in the 10,000 GWAS permutations were collapsed into 'meta-genes' and treated as one gene. Meta-gene Z-scores were averaged across the input Z-scores. Lastly, the GWAS Z-scores of the meta genes were scaled (mean = 0, standard deviation = 1).

GLS model to calculate pathway enrichment and core gene scores

We used a GLS regression to associate the GWAS gene Z-scores to the pathway Z-scores and co-regulation Z-scores (described below). These two analyses result in the pathway enrichments and core gene prioritisations, respectively. We used the gene-gene correlation matrix derived from the 10,000 permutations as a measure of conditional covariance of the error term ($\boldsymbol{\Omega}$) in the GLS to account for the relationships between genes due to LD and proximity. The pseudo-inverse of $\boldsymbol{\Omega}$ is used as a substitute for $\boldsymbol{\Omega}^{-1}$

The formula of the GLS is as follows:

$$\beta=(\mathbf{X}^T\boldsymbol{\Omega}^{-1}\mathbf{X})^{-1}\mathbf{X}^T\boldsymbol{\Omega}^{-1}\mathbf{y}$$

Where β is the estimated effect size of pathway, term or gene from the co-regulation matrix, $\boldsymbol{\Omega}$ is the gene-gene correlation matrix, $\mathbf{X}$ is the design matrix of real GWAS Z-scores and $\mathbf{y}$ is the

vector of gene Z-scores per pathway, term or gene from the co-regulation matrix. As we standardized the predictors, we did not include an intercept in the design matrix and **X** only contains one column with the real GWAS gene Z-scores. We estimated the beta's for the 10,000 random GWASs in the same way and subsequently used them to estimate the empirical p-value for β.

Definition of POI and DDR genes

We combined genes implicated in the DDR from a number of sources yielding a total of 778 genes (**Supplementary Table 19**)[69–71]. To identify genes associated with premature ovarian insufficiency/primary ovarian insufficiency (ICD-11 GA30.6), we carried out a search in PubMed for premature ovarian insufficiency, primary ovarian insufficiency, premature ovarian failure and ovarian dysfunction in humans and reviewed all primary studies published in English until 22nd of July, 2020. We included syndromic, non-syndromic, sporadic as well as familial single nucleotide variants, insertion/deletions and copy number variants (CNVs) and included 114 genetic variants from 139 studies. We did not attempt to review the clinical significance of the variants, which ranged from classical POI genes to newly identified CNVs in whole-exome sequencing studies. We expanded our search to review articles and ClinVar. We uncovered another four genes implicated in Perrault Syndrome for which our search terms were not included in the original articles. This gave a total of 118 genes. Our search detected all genetic variants entered in ClinVar as pathogenic, likely pathogenic or with conflicting interpretations of pathogenicity. We excluded genes with variants when no assertion criteria were provided and no published data were available for assessment in ClinVar. Two studies of large chromosomal rearrangements as well as quantitative trait loci consisting of more than a single genetic variant from GWAS in POI populations were excluded resulting in 74 genes (**Supplementary Table 6**). Gene lists were curated independently of the current meta-analysis and genes were only included if there was convincing evidence independent of any GWAS study.

Polygenic prediction of early menopause

To evaluate the impact of common variants on clinical extremes of ANM, we first performed a GWAS meta-analysis excluding the UK Biobank study (N=95,275). Effect estimates from this analysis (**Supplementary Table 2**) were then used for subsequent polygenic score (PGS) construction of ~6.97 million autosomal variants across the genome using LDPRED[72]. The PGS was calculated using PLINK[73] v1.90b4.4 in an independent sample of 108,840 women with the full phenotypic range of ANM ages from the UK Biobank study, rescaled to have a mean of 0 and standard deviation of 1. We then estimated the centile distribution of the genetic risk score for all women with a valid ANM (with no lower or upper phenotype boundary). Two outcomes were defined: early menopause (EM) defined as ANM < 45 (N=11,268) vs all other women (N=97,572); and premature ovarian insufficiency (POI), defined as ANM < 40 years (N=2,407) vs all other women (N=106,433). Logistic regression analyses, adjusting for age, genotype array and 10 genetic principal components, were then performed with either EM or POI as the outcome. This was performed 99 times for each centile of genetic risk (coded 1) vs the 50th centile of genetic risk (coded 0). To assess the relevance of this score to each ANM age group, we estimated the average PGS value by year of ANM. For example, we grouped all women with ANM = 47 and estimated the mean and standard error of the PGS in this group of women. Our

intuition was that any ANM range not influenced by common genetic variants would have the population mean PGS (i.e mean = 0 and SD = 1). Receiver operating characteristics (ROC) models were performed in Stata v14 using the *roctab*, *rocgold* and *rocreg* commands.

Mendelian Randomization analyses

In order to infer causal relationships between ANM and other health related outcomes, we performed Mendelian Randomization (MR). The 290 independent ANM signals were used as a genetic instrument for later ANM. Where a signal was not present in the outcome GWAS, we identified the best HapMap2 proxy with r2>0.5 within 250 kb either side of the signal and its relevant weight was included in our genetic instrument (**Supplementary Table 23**). The genetic variants were identified in publicly available GWAS datasets for a range of outcomes of interest (**Supplementary Table 22)**. These were used in three methods of MR - inverse variance weighted[74], MR-EGGER[75] and weighted median[76]. As a sensitivity analysis we additionally removed signals that appeared to be outliers. This was achieved using the Radial method considering the IVW model[77]. We also performed MR considering the effect of a range of putative modifiable risk factors on ANM as the outcome using the same MR models. Genetic instruments were created for the risk factors using independent genetic variants with effects estimated in published GWAS (**Supplementary Table 25**). For the risk factors of cigarette exposure and alcohol consumption, the MR was performed with a single genetic variant by calculating a Wald ratio for the effect of the variants on ANM divided by the effect on the risk factor using mrrobust in Stata v16.0. The effect of the genetic variant for alcohol consumption was measured in log(drinks per week) (note that drink is a US measure of alcohol consumption equal to 14g pure alcohol, equivalent to 1.75 UK units). Hence a change from 1 drink to 7 drinks (US maximum recommended per week) would be the equivalent of a 1.95 increase in log(drinks per week), which when applied to the Wald estimate, gives the respective change in age at menopause.

Expression of candidate genes identified by human GWAS in a mouse model of environmentally-induced low ovarian reserve

<u>Generation of mouse model</u>

All animal experiments underwent ethical review by the University of Cambridge Animal Welfare and Ethical Review Board and were carried out under the UK Home Office Animals (Scientific Procedures) Act (1986, United Kingdom). Female C57BL/6J mice were randomized to be fed *ad libitum* either a standard laboratory chow diet (7% simple sugars/3% fat; Special Dietary Services, Witham, UK) or an obesogenic diet (10% simple sugars/20% animal lard; Special Diets Services, Witham, UK). The obesogenic diet was supplemented with a separate pot of sweetened condensed milk (55% simple sugars/8% fat; Nestle UK, Gatwick, UK) available to the animals within the cage. A detailed description of the dietary regimen has been published previously[78]. Female mice were placed on the allocated diet six weeks prior to first mating with wild-type males on standard chow diet. The first litter was discarded after weaning, and only proven-breeder females were used for the experimental protocols. Second matings occurred once females on the obesogenic diet had reached at least 10g absolute fat mass, as assessed by time domain nuclear resonance imaging (TDNMR) (Minispec Time Domain Nuclear Resonance, Bruker Optics). The female mice remained on their allocated diets throughout the breeding, pregnancy, and lactation phases. After delivery, each litter was culled to six pups at random to standardize their plane of nutrition from postnatal day 3 in all litters. There was no

significant difference in the pre-culling litter size between obesogenic and control litters. Equal sex ratios within the litters were maintained as far as possible. After weaning at day 21, female offspring were randomly allocated to either the control or the obesogenic diets (identical to those used for the dams) and remained on these diets for the duration of the study. Bodyweight and food intake were measured weekly. At 12 weeks of age, offspring total and fat mass were assessed by weighing and by TDNMR (Minispec Time Domain Nuclear Resonance, Bruker Optics) respectively. Following an overnight fast, the female offspring were weighed and then culled by CO_2 asphyxiation and cervical dislocation. Ovaries were dissected and weighed immediately. One ovary from each animal was snap-frozen in liquid nitrogen or dry ice, and stored at -80^{o}C, the other was fixed in formalin/paraldehyde. The fixed ovary was sectioned and subjected to haematoxylin and eosin (H&E) staining to ensure equal distribution of estrous stages in each experimental group (data not shown). Detailed reproductive and metabolic phenotyping of the female pups has previously been published[24].

Gene expression analysis

A screen of 35 DNA damage response genes highlighted by our previous GWAS on ANM were selected for investigation[13] - *Brca1, Bre, Brsk1, Chd7, Chek2, Dido1, Fbxo18, Helb, Helq, Mcm8, Mlf1ip, Msh5, Msh6, Mycbp, Polg, Prim1, Rad51, Rad54l, Rev3l, Uimc1, Apex, Aptx1, Cdk2ap1, Dmc1, Exo1, Fam175a, Fanci, Ino80, Kntc1, Papd7, Parl, Parp2, Polr2e, Polr2h* and *Tlk1*. Expression levels were measured in whole snap-frozen ovaries. RNA was extracted using a miRNeasy mini Kit (Qiagen, Hilden, Germany). The kit was used according to the manufacturer's instructions, with the addition of DNaseI digestion to ensure that the samples were free from genomic DNA contamination. The extracted RNA was quantified using a Nanodrop spectrophotometer (Nanodrop Technologies, Wilmington, DE, US). cDNA was synthesized from 1µg RNA using oligo-dT primers and M-MLV reverse transcriptase. Gene expression was quantified via RT-PCR (StepOne Plus machine; Applied Biosystems, Warrington, UK) using custom-designed primers (Sigma, Poole, UK) and SYBR green reagents (Applied Biosystems, Warrington, UK). Equal efficiency of reverse transcription between all groups was confirmed using the housekeeper gene *ppia*, and absence of gDNA contamination was confirmed by quantifying *myh6*, which was absent in all samples.

Statistical analysis

All data were initially analyzed using a 2-way ANOVA with maternal diet and offspring diet as the independent variables. In order to correct for multiple hypothesis testing of gene expression levels, p values were transformed to q values to take account of the false discovery rates using the p.adjust function in R stats package (R Foundation for Statistical Computing, Vienna, Austria). Data are represented as means ± SEM. Where p values are reported, an alpha level <0.05 was considered statistically significant. All data analysis was conducted using the R statistical software package version 2.14.1 (R Foundation for Statistical Computing, Vienna, Austria). In all cases, n refers to the number of litters, and n=8 for all groups. Study power was determined based on effect sizes for gene expression differences observed in our previous studies of this model[24].

Human oocytes mRNA screen

Research on RNA expression in human eggs was carried out according to the Helsinki II declaration and was conducted in accordance with national regulation on research on human subjects and material. The research was approved by the Scientific Ethical Committee of the Capital Region of Denmark (Videnskabsetisk Komite) in accordance with Danish National regulation (H-2-2011-044; extension license amm. Nr. 51307; license holder: Claus Yding Andersen and H-1604473; license holder: Eva R. Hoffmann; H-16027088 granted to Marie Louise Grøndahl). The full protocols contained permission to conduct mRNA sequencing on human eggs. GDPR approval was obtained from the national data agency (SUND-2016-60, Eva R Hoffmann and HGH-2016_086 to Marie Louise Grøndahl). All participants provided informed consent according to Danish ethical regulation after receiving written information and oral clarification about participation. Participants could withdraw from the study at any time. Participants did not receive monetary compensation and their participation was fully voluntary and did not affect their fertility treatment.

Single human MII oocytes were collected as described previously[79], lysed in-tube and the cDNA was amplified according to the manufacturer's instructions (Takara Bio; mRNA-Seq, SMART-Seq v4 ultra low input RNA kit, cat. no. 634894). The quality of individual cDNA libraries was verified on an Agilent 2100 Bioanalyzer instrument using a high sensitivity DNA kit (Agilent, 5067-4626). The libraries were prepared with 100 pg input using the Nextera XT DNA library preparation kit (Illumina, FC-131-1024) and the Nextera XT index kit v2 (FC-131-2002) and quantified on a Qubit 3.0 fluorimeter (Thermo Fisher Scientific, Q32854). The quality of the final library was verified on the Agilent 2100 Bioanalyzer high sensitivity DNA chip and pooled to 4 nM. The 4 nM library pools were denatured and loaded according to the recommended NextSeq500 guidelines (Illumina Inc.).

Expression analysis of GWAS genes in human oocytes and granulosa cells at various stages of development

We used processed RNA-seq data of Fetal Primordial Germ Cells from Li *et al* (2017, Accession code: GSE86146)[80] from 17 human female embryos ranging from 5-26 weeks post-fertilisation, and from Zhang *et al* (2018, Accession code GSE107746)[81] studies, follicles at 5 different stages of development from fresh ovarian tissue from 7 adult donors, separated into oocytes and granulosa cell fractions; in addition to our MII Oocytes single-cell RNA-seq dataset (described below).

We transformed the per-cycle base call (BCL) file output from the sequencing run of 11 human MII oocytes into per-read FASTQ files using the bcl2fastq2 Conversion Software v2.19 from Illumina. The samples libraries were multiplexed across four sequencing lanes and the FastQ files from each of the four lanes were concatenated to generate one set of paired fastq files per sample. We performed sample QC and filtering of reads to remove low quality reads, adaptor sequences and low quality bases with trimmomatic[82] version 0.36 in two steps using ILLUMINACLIP:/ /Trimmomatic-0.36/adapters/NexteraPE-PE.fa:2:30:10 (SLIDINGWINDOW:4:20 CROP:72 HEADCROP:10 MINLEN:40 followed by and extra trim of headbases with HEADCROP:10.) Subsequent to filtering, we used the remaining paired reads for alignment by hisat2[83] to the human genome GeneCode v.27 release with the paired GenCode v.27 gtf file containing gene annotations using: ($HISAT2 -p 22 --dta -x .gencode.v27 -1 R1.fastq -2 R2.fastq -S sample.sam) (Pertea *et al.* 2016). The resulting sam files were sorted, indexed and transformed to bam files using samtools[84]. QC measures of aligned reads

was generated using picard metrics (https://slowkow.github.io/picardmetrics) and the CollectRnaSeqMetrics tool from picard tools (http://broadinstitute.github.io/picard). We filtered the bam files for mitochondrial reads and Stringtie was applied to merge and assemble reference guided transcripts for gene level quantifications of raw counts, and transcripts per million (TPM)[85]. Of the 283 consensus genes highlighted by the GWAS (**Supplementary Table 5**), 258 passed QC and were available in the expression dataset. Gene expression levels in TPM were used for further analyses as this unit allows efficient comparison of gene expression levels between samples from different studies. A pseudo-count of 1 was added to all TPM values and converted to log2 scale before the heatmaps were plotted. Hierarchical clustering by euclidean distance, z-score calculation and plotting the heatmap was done using the R package 'pheatmap' (Kolde R, 2019, v1.0.12). Z-scores are calculated by subtracting the mean of TPM values in all samples for a gene and dividing by the standard deviation. Samples with only TPM>5 were considered for heatmap showing the GWAS genes.

sChek1, Chek1 cKO, and Chek2 mice

Mouse work at the University of Copenhagen (*sChek1*) was licensed under 2016-15-0202-00043 by the Danish Animal Experiments Inspectorate (Dyreforsøgstilsynet, Denmark). Mouse work at UAB (Chek2) was approved by the UAB and the Catalan Ethics Committee for Animal Experimentation (CEEAAH 1091; DAAM6395). Mouse work at CCHMC (*Chek1* cKO, *Ddx4-Cre*) was performed according to the guidelines of the Institutional Animal Care and Use Committee (protocol no. IACUC2018-0040) approved by CCHMC. The *Chek1 cKO, Zp3-Cre* embryology was conducted at the Institute of Animal Physiology and Genetics CAS in Libechov (Czech Republic), abiding by the policies of the Expert Committee for the Approval of Projects of Experiments on Animals of the Academy of Sciences of the Czech Republic (# 43-2015).

Chek1 cKO (Ddx4-Cre), sChek1, and *Chek2* mutant mice were generated previously[33,37,86]. The lines were maintained in C57BL/6-129Sv and inbred C57BL/6-129Sv (*sChek1 and Chek2*) backgrounds respectively. The *chek2* mouse is available under accession number BRC03481 at the RIKEN Bioresource Centre. The *Chek1* cKO *Zp3-Cre* embryos were generated by crossing mice with *Zp3*-Cre transgene[87] to mice with *Chek1* allele containing *LoxP* sites[88] resulting in mice expressing Cre-recombinase under the control of the oocyte specific *zona pellucida 3* promotor (*Zp3::Cre*) to produce *Chek1* cKO (*Zp3-Cre*). All experiments were carried out using litter mate controls or with animals of closely related parents as controls. The four mutant strains were kept at the University of Copenhagen (*sChek1*), Autonomous University of Barcelona (*Chek2*), Cincinnati Children's Hospital Medical Center (*Chek1 cKO* - Ddx4-Cre) and Institute of Animal Physiology and Genetics CAS in Libechov, Czech Republic (Chek1 cKO Zp3-Cre). Breeding cages were set in a conventional way with strict specific pathogen-free barrier and mice used for experiments were kept in individual ventilated cages (IVC). 12h light exposure was provided. Temperature, relative humidity and air changes per hour were 22 °C (+/-2 °C), 55% +/-10 %, and 17 respectively. Food and water were provided *ad libitum*. Animals were genotyped two times, initially upon weaning and again before experimental procedures were carried out. Mouse genotyping was performed by PCR analysis using the following primers for the *Chek1 cKO (Ddx4)*: F1 (5′-ACC TGC CCG CAA CTC CCT TTC-3′) and R2 (5′-TGC AAC AGC TTC AGT TAT TC-3′); for the cKO Chek1(Zp3-Cre): Cre_low (5'-TAT TCG GAT CAT CAG CTA-3'), Cre_up (5'-GGT GGG AGA ATG TTA ATC-3'), CHK1F1 (5'-ACC TGC CCG CAA CTC

CCT TTC-3'), CHK1R1(5'-CCA TGA CTC CAA GCA CAG CGA-3'). The sizes of products were 318 bp for wild type and 380 bp for *loxP*/*loxP* transgene. The size of the *Zp3*-Cre transgene was 139 bp. For *sChek1* the primers were: gsChek1_left “TGT CTT CCC TTC CCT GCT TA”, gsChek1_right1 “TCC CAA GGG TCA GAG ATC AT” and g*s*Chek1_5’PCR2 “GTA AGC CAG TAT ACA CTC CGC TA”. The wild type gene yields a size of 400 bp whereas the transgene is 270 bp. For *Chek2*, the primers WT1F (5’–GTGTGCGCCACCACTATCCTG–3’), WT2R (5’–CCCTTGGCCATGTTTCATCTG–3’) and NeoMutR (5’–TCCTCGTGCTTTACGGTATC–3’) were used to detect the wild type (450 bp) and the mutant (625 bp) alleles in one PCR reaction. The Qiagen Taq polymerase PCR kit was used for genotyping (Cat No 201203 / 201205).

Mouse ovarian histology and follicle count

Ovaries were dissected and placed in 4% formaldehyde (*Chek1 cKO (Ddx4)*) & Bouin´s fixative solution (70% saturated picric acid solution (Applichem, A2520, 1000), 25% formaldehyde, 5% glacial acetic acid (Merck, 1.00063.2500)) or 4% formaldehyde for *Chek1 cKO (Ddx4-Cre)* overnight at 4 °C. The ovaries were washed two times with cold PBS for 30 minutes followed by dehydration with an increasing concentration of ethanol. Subsequently, the samples were submerged in Histo-Clear II (Cat. # HS-202, National Diagnostics) for 30 min. at room temperature. This was repeated another two times (three times in total) with fresh Histo-Clear II. Ovaries were embedded in paraffin blocks and cut to a thickness of 7 µm (*sCHEK1* and *Chek2)* and 6 µm (*Chek1 cKO (Ddx4-Cre)*) and mounted on poly-L-lysine coated slides. After de-paraffinization and rehydration, the slides were stained with PAS-hematoxylin. The tissue was imaged using a Zeiss Axio scanner Z.1 and follicles with a visible nucleus were counted using the Zen Blue lite software from Zeiss. Primordial follicles contain one layer of flat granulosa cells surrounding the oocytes, primary follicles have one layer of cuboid granulosa cells. Secondary follicles contain two or more layers of granulosa cells and antral follicles are those with one or several cavities (the antrum).

Mouse ovulation induction and oocyte collection

Ovulation was induced by injection of 5 IU of PMSG (Prospec; ref HOR-272) followed by 5 IU of hCG (Chorulon Vet; ref 422741) after 47 hours. For 11-13, 16 and 24 months old mice, 7.5 IU of each hormones were used. 12 hours post-hCG injection, the mice were sacrificed and oviducts were dissected under a stereo-microscope to release the cumulus masses into 90 µl drop of fertilization medium covered with mineral oil (NordilCell; ref 90142). Oocytes were recovered from oviducts by gently tearing swollen ampulla of oviducts to release cumulus masses into medium. Recipe of fertilization medium was previously published elsewhere[89].

RT-qPCR on mice oocytes

Total RNA from oocytes was isolated with the *Arcturus PicoPure RNA Isolation Kit* from *Applied Biosystems* following the manufacturer’s instructions. Reverse transcription reactions were done with twenty eight nanograms of RNA using the *Maxima First Strand cDNA Synthesis Kit for RT-qPCR with dsDNase* (Thermo Fisher Scientific). cDNA was quantified by qPCR with the *Applied Biosystems 7500 FAST Real-Time PCR System* using *Power SYBR green PCR Master Mix* from *Thermo Fisher Scientific*. The sequences of the primers used are the following: *Chek1*-For: 5’- AAGCCACGAGAATGTAGTGAAA-3’, *Chek1*-Rev: 5’- AGCATCTTGTTCAGGCATCC-3’,

Actb-For: 5'-CCAACCGTGAAAAGATGACC-3', *Actb*-Rev: 5'-ACCAGAGGCATACAGGGACA-3'. Values were normalized to the expression of *Actb* housekeeping gene.

Mouse embryo development in vitro

Fresh pre-thawed frozen sperm from a proven fertile C57BL/6 wild-type male was used for *in vitro* fertilization and poured into a dish containing mature MII eggs in fertilization medium. Disappearance of germinal vesicle (GV) and polar body extrusion confirmed fertilization. Zygotes were incubated at 5% CO_2 and 37 °C. After incubating zygotes in fertilisation medium for overnight, We transferred zygotes to a 60 mm petri dish containing 50 µl KSOM (Chemicon, cat MR-106-D) covered by mineral oil(NordilCell; ref 90142). Two separate dishes were prepared for embryos from each genotype. The embryos were again incubated at 5% CO_2 and 37 °C. The developmental stage of embryos was assessed using a stereomicroscope at the equivalent of 0.5, 1.5, 2.5, 3.5, 4.4 and 5.5 days post-coitum (dpc). For *chek2*, where the wild type frequency of fertilization was lower than in the the *Chek1-cko* and *sChek1* strains, we used young C57BL/6J.Ola.Hsd females to control for the efficiency of IVF (85%).

Mouse ovulation and embryo development (Chek1 cKO, Zp3-Cre)

Chek1 ctrl and cKO females were stimulated with 5 IU of PMSG (HOR-272, Prospec) followed by 5 IU of hCG (Ovitrelle, Merck) after 44 hours. After 18 hours, the females were sacrificed using cervical dislocation according to the protocols authorized by the ethics committee, and ovulated MII oocytes and zygotes were collected in M2 media (M7167-50ML, Sigma-Aldrich) by tearing ampulla from oviduct. The oocytes and zygotes in cumulus mass were placed into a drop of M2 media supplied with 300 µg/ml hyaluronidase (H4272, Sigma-Aldrich) to release the cumulus cells. The MII oocytes and zygotes were cultured at 5% CO_2 and 37°C in EmbryoMax® KSOM media (MR-106-D, Sigma-Aldrich) and after 10 hours were scored using Leica DMI 6000 microscope. Only zygotes with visible pronuclei were left for subsequent culture.

Immunofluorescence analysis of mouse preimplantation embryos (Chek1 cKO, Zp3-Cre)

The embryos were 3x briefly washed in PBS supplied with 1mg/ml poly(vinyl alcohol) and fixed in 3.7% formaldehyde for 45 min. They were permeabilized thereafter by 0.5% Triton X-100 in PBS for 45 min. To block unspecific antibody binding, the embryos were incubated in 2% normal donkey serum (NDS) for 2 hours. The embryos were incubated overnight at 4°C at a dilution 1:200 in primary antibody against gH2AX (9718, Cell Signaling Technology). The next day, they were incubated for 90 min at a dilution 1:100 in Rhodamine (TRITC)-AffiniPure Donkey Anti-Rabbit IgG (711-025-152, Jackson Immuno Research). Then they were mounted in ProLong™ Gold Antifade Mountant with DAPI (P36941, Invitrogen) with a spacer to uphold the embryonic 3D structure. The embryos were washed 5x for 8 min in PBS supplied with 1mg/ml bovine serum albumin or 0.2% NDS between each steps. The embryos were scanned using a confocal microscope (Leica TCS SP5) and Fiii software[90] was employed for image analysis.

Mouse embryo transfer

Wild-type female recipient mice (surrogate) were prepared to receive embryos by mating them with an infertile male one night before the transfer of embryos. Successful preparation of recipient mice for embryo transfer was confirmed by checking for the presence of a plug. Two cell-stage (1.5dpc) embryos were transferred into a single horn of recipient mice and anaesthesia were maintained during this procedure. Pups were born after 19 days of embryo transfer.

Natural breeding, assessment of health of offspring and fertility in mouse

To test the natural breeding efficiency, we set cages with one or two adult (2-months or 12-month-old) control or females with a male of proven fertility. We registered litter sizes and dates of delivery for all litters obtained during a period for up to one year.

Mice Serum AMH analysis

Mice of various ages were anesthetized. Blood was collected in a plain tube, allowed to clot for one hour at room temperature and then centrifuged at 3000 rpm (1500g) for 15 minutes at 4 °C. After centrifugation, supernatant (serum) was collected in a 1.5 ml tube and stored at -80 °C. Serum AMH levels were determined by using AMH ELISA kit (cat. # AL-113) from Ansh Labs, Webster, TX.

Assessment of the health of the offspring from control and mutant breeding was performed on a weekly basis by the personnel of the respective animal facilities following the standard health monitoring protocols approved by the Copenhagen or Catalan Ethics Committee for Animal Experimentation.

Data availability

Full genome-wide association summary statistics for the discovery meta-analysis are available from the ReproGen website (www.reprogen.org).

MII Oocyte dataset EGAS00001004947. Access to EGAS00001004947 is granted in accordance with the ethics permission under which the data were collected from participants and under appropriate GDPR compliant data processor agreements.

SMR https://cnsgenomics.com/software/smr/#eQTLsummarydata

Tabula Muris https://tabula-muris.ds.czbiohub.org/

LDSC-SEG https://github.com/bulik/ldsc/wiki/Cell-type-specific-analyses

RNA-seq samples https://www.ebi.ac.uk/ena

Human oocyte expression analyses: GSE107746, GSE107746

Methods references

48. Bycroft, C. *et al.* The UK Biobank resource with deep phenotyping and genomic data. *Nature* **562**, 203–209 (2018).

49. Tyrrell, J. *et al.* Using genetics to understand the causal influence of higher BMI on depression. *Int. J. Epidemiol.* **48**, 834–848 (2019).

50. Loh, P.-R. *et al.* Efficient Bayesian mixed-model analysis increases association power in large cohorts. *Nat. Genet.* **47**, 284–290 (2015).

51. Yang, J., Lee, S. H., Goddard, M. E. & Visscher, P. M. GCTA: a tool for genome-wide complex trait analysis. *Am. J. Hum. Genet.* **88**, 76–82 (2011).

52. Terao, C. *et al.* Chromosomal alterations among age-related haematopoietic clones in Japan. *Nature* **584**, 130–135 (2020).

53. Wood, A. R. *et al.* Variants in the FTO and CDKAL1 loci have recessive effects on risk of obesity and type 2 diabetes, respectively. *Diabetologia* **59**, 1214–21 (2016).

54. Szustakowski, J. D. *et al.* Advancing Human Genetics Research and Drug Discovery through Exome Sequencing of the UK Biobank. *medRxiv* 2020.11.02.20222232 (2020) doi:10.1101/2020.11.02.20222232.

55. McLaren, W. *et al.* The Ensembl Variant Effect Predictor. *Genome Biol.* **17**, 122 (2016).

56. Zhou, W. *et al.* Scalable generalized linear mixed model for region-based association tests in large biobanks and cohorts. *Nat. Genet.* **52**, 634–639 (2020).

57. Zhu, Z. *et al.* Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets. *Nat. Genet.* **48**, 481–487 (2016).

58. Pers, T. H. *et al.* Biological interpretation of genome-wide association studies using predicted gene functions. *Nat. Commun.* **6**, 5890 (2015).

59. Tabula Muris Consortium *et al.* Single-cell transcriptomics of 20 mouse organs creates a Tabula Muris. *Nature* **562**, 367–372 (2018).

60. Nestorowa, S. *et al.* A single-cell resolution map of mouse hematopoietic stem and progenitor cell differentiation. *Blood* **128**, e20-31 (2016).

61. Finucane, H. K. *et al.* Heritability enrichment of specifically expressed genes identifies disease-relevant tissues and cell types. *Nat. Genet.* **50**, 621–629 (2018).

62. GTEx Consortium. Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. *Science* **348**, 648–60 (2015).

63. Roadmap Epigenomics Consortium *et al.* Integrative analysis of 111 reference human epigenomes. *Nature* **518**, 317–30 (2015).

64. Fehrmann, R. S. N. *et al.* Gene expression analysis identifies global gene dosage

sensitivity in cancer. *Nat. Genet.* **47**, 115–25 (2015).

65. Segrè, A. V *et al.* Common inherited variation in mitochondrial genes is not enriched for associations with type 2 diabetes or related glycemic traits. *PLoS Genet.* **6**, (2010).

66. de Leeuw, C. A., Mooij, J. M., Heskes, T. & Posthuma, D. MAGMA: Generalized Gene-Set Analysis of GWAS Data. *PLoS Comput. Biol.* **11**, (2015).

67. Lamparter, D., Marbach, D., Rueedi, R., Kutalik, Z. & Bergmann, S. Fast and Rigorous Computation of Gene and Pathway Scores from SNP-Based Summary Statistics. *PLoS Comput. Biol.* **12**, e1004714 (2016).

68. Deelen, P. *et al.* Improving the diagnostic yield of exome- sequencing by predicting gene-phenotype associations using large-scale gene expression analysis. *Nat. Commun.* **10**, 2837 (2019).

69. Knijnenburg, T. A. *et al.* Genomic and Molecular Landscape of DNA Damage Repair Deficiency across The Cancer Genome Atlas. *Cell Rep.* **23**, 239-254.e6 (2018).

70. Pearl, L. H., Schierz, A. C., Ward, S. E., Al-Lazikani, B. & Pearl, F. M. G. Therapeutic opportunities within the DNA damage response. *Nat. Rev. Cancer* **15**, 166–80 (2015).

71. Álvarez-Quilón, A. *et al.* Endogenous DNA 3' Blocks Are Vulnerabilities for BRCA1 and BRCA2 Deficiency and Are Reversed by the APE2 Nuclease. *Mol. Cell* **78**, 1152-1165.e8 (2020).

72. Vilhjálmsson, B. J. *et al.* Modeling Linkage Disequilibrium Increases Accuracy of Polygenic Risk Scores. *Am. J. Hum. Genet.* **97**, 576–92 (2015).

73. Purcell, S. *et al.* PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**, 559–75 (2007).

74. Burgess, S., Butterworth, A. & Thompson, S. G. Mendelian randomization analysis with multiple genetic variants using summarized data. *Genet. Epidemiol.* **37**, 658–65 (2013).

75. Burgess, S. & Thompson, S. G. Interpreting findings from Mendelian randomization using the MR-Egger method. *Eur. J. Epidemiol.* **32**, 377–389 (2017).

76. Bowden, J., Davey Smith, G., Haycock, P. C. & Burgess, S. Consistent Estimation in Mendelian Randomization with Some Invalid Instruments Using a Weighted Median Estimator. *Genet. Epidemiol.* **40**, 304–14 (2016).

77. Bowden, J. *et al.* Improving the visualization, interpretation and analysis of two-sample summary data Mendelian randomization via the Radial plot and Radial regression. *Int. J. Epidemiol.* **47**, 2100 (2018).

78. Samuelsson, A.-M. *et al.* Diet-induced obesity in female mice leads to offspring hyperphagia, adiposity, hypertension, and insulin resistance: a novel murine model of developmental programming. *Hypertens. (Dallas, Tex. 1979)* **51**, 383–92 (2008).

79. Sankar, A. *et al.* KDM4A regulates the maternal-to-zygotic transition by protecting broad

H3K4me3 domains from H3K9me3 invasion in oocytes. *Nat. Cell Biol.* **22**, 380–388 (2020).

80. Li, L. *et al.* Single-Cell RNA-Seq Analysis Maps Development of Human Germline Cells and Gonadal Niche Interactions. *Cell Stem Cell* **20**, 858-873.e4 (2017).

81. Zhang, Y. *et al.* Transcriptome Landscape of Human Folliculogenesis Reveals Oocyte and Granulosa Cell Interactions. *Mol. Cell* **72**, 1021-1034.e4 (2018).

82. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).

83. Chen, S. *et al.* AfterQC: automatic filtering, trimming, error removing and quality control for fastq data. *BMC Bioinformatics* **18**, 80 (2017).

84. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).

85. Pertea, M., Kim, D., Pertea, G. M., Leek, J. T. & Salzberg, S. L. Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown. *Nat. Protoc.* **11**, 1650–67 (2016).

86. Takai, H. *et al.* Chk2-deficient mice exhibit radioresistance and defective p53-mediated transcription. *EMBO J.* **21**, 5195–205 (2002).

87. Lewandoski, M., Wassarman, K. M. & Martin, G. R. Zp3-cre, a transgenic mouse line for the activation or inactivation of loxP-flanked target genes specifically in the female germ line. *Curr. Biol.* **7**, 148–51 (1997).

88. Lam, M. H., Liu, Q., Elledge, S. J. & Rosen, J. M. Chk1 is haploinsufficient for multiple functions critical to tumor suppression. *Cancer Cell* **6**, 45–59 (2004).

89. Takeo, T. & Nakagata, N. Superovulation using the combined administration of inhibin antiserum and equine chorionic gonadotropin increases the number of ovulated oocytes in C57BL/6 female mice. *PLoS One* **10**, e0128330 (2015).

90. Schindelin, J. *et al.* Fiji: an open-source platform for biological-image analysis. *Nat. Methods* **9**, 676–82 (2012).

Acknowledgements

This research has been conducted using the UK Biobank resource under application numbers 871 (Exeter) and 9797 (Cambridge). Full individual study acknowledgements can be found in the **Supplementary Information**. The authors wish to dedicate this work to the memory of Professor Petr Solc.

Author contributions

All authors reviewed the original and revised manuscripts. Leads on manuscript writing: K.S.R, F.R.D, E.R.H, A.Murray, I.Roig, J.RB.P. Central statistical genetics analysis team: K.S.R, F.R.D, A.Murray, J.RB.P. Animal model working group: J.H, A.M-M, C.E.A, L.K, H.A, J.L.T, J.MG, S.T, E.PTH, M.F, Y.H, A.S, A.Puj, A.J.L, J.A.D, S.E.O, S.H.N, P.Solc, E.R.H, I.Roig. Human oocyte expression working group: A.Azad, V.S, R.B, K.W.O, M.K.H, M.LG, C.Y, E.R.H. Sample collection, genotyping, phenotyping and individual study analysis: K.S.R, F.R.D, D.J.T, P.F, A.Clar, O.B.B, P.Sul, R.G.W, C.T, M.H, K.L, N.O, P.N.T, P.A, S.Stan, P.RHJ.T, T.U.A, B.Z.A, E.N, I.L.A, A.M.A, K.J.A, A.Aug, S.Band, C.M.B, R.N.B, H.B, M.W.B, S.Beno, S.Berg, M.B, E.B, S.E.B, M.K.B, D.I.B, N.B, J.A.B, L.B, J.E.B, A.Camp, H.C, J.E.C, E.C, S.J.C, G.C, M.C, T.C, F.J.C, A.Cox, L.C, S.S.C, F.C, K.C, G.D, E.JCN.d, R.d, I.D, E.W.D, J.D, A.M.D, M.D, M.E, T.E, P.A.F, J.D.F, L.Fer, N.F, T.M.F, M.G-D, M.Mezz, M.G-C, C.G, G.G.G, H.G, D.F.G, V.G, P.G, C.A.H, N.H, P.H, C.Ha, C.He, W.H, G.H, J.L.H, J.J.H, F.H, D.H, M.A.I, R.D.J, M.DR.J, E.M.J, P.K.J, D.K, S.LR.K, C.Kart, R.K, C.M.K, I.K, C.Koop, P.K, A.W.K, Z.K, M.LaBi, G.L, C.L, L.J.L, J.SE.L, D.A.L, L.LM, J.Li, A.L, S.Lind, T.L, M.Lin, Y.L, S.Liu, J.Lu, R.M, P.KE.M, M.Mang, A.Mann, B.Mar, J.Mar, N.G.M, H.M, B.McK, S.E.M, C.Meis, T.M, C.Men, A.Mets, L.M, R.L.M, G.W.M, D.O.M, A.Mulas, A.M.M, Alison.M, M.A.N, A.N, R.N, T.N, D.R.N, A.F.O, H.O, J.N.P, A.V.P, N.L.P, N.P, A.Pet, U.P, P.DP.P, O.P, E.Por, B.M.P, I.Rah, G.R, H.S.R, P.M.R, S.M.R, A.R, L.M.R, F.R.R, J.R, I.Rud, R.R, D.R, C.F.S, E.S, D.P.S, S.San, E.J.S, C.Sar, D.Schl, M.K.S, M.J.S, K.E.S, C.Sco, S.Shek, A.V.S, B.H.S, J.A.S, R.S, M.C.S, T.D.S, J.J.S, M.S, D.Sto, J.BJ.v, K.Str, U.S, A.J.S, T.Tan, L.R.T, A.T, U.Þ, N.J.T, D.T, M.T, M.A.T, T.Tru, J.T, A.G.U, S.U, C.M.V, V.V, U.V, P.V, H.V, Q.W, N.J.W, C.R.W, D.R.W, A.N.W, K.W, G.W, J.F.W, B.HR.W, A.W, A.R.W, W.Z, M.Z, Z.C, L.Li, L.Fra, S.Burg, P.D, T.H.P, K.Stef, J.C, Y.T.v, K.L.L, D.I.C, D.F.E, J.A.V, J.M.M, K.K.O, A.Murray, J.RB.P

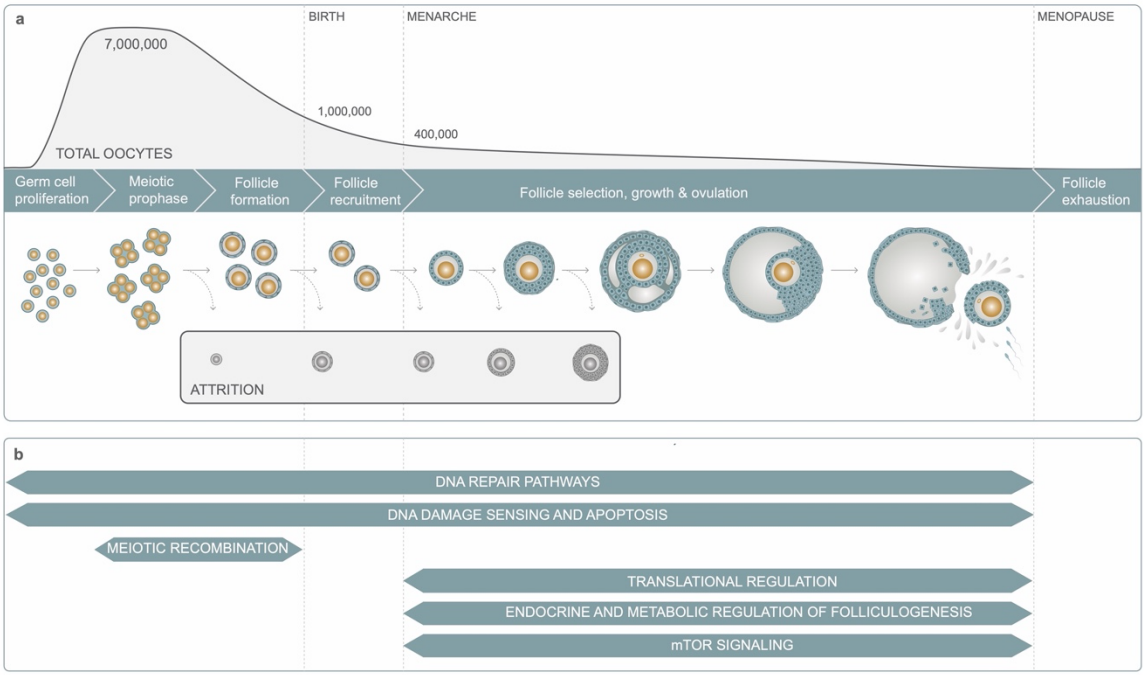
Competing interests

Full individual study and author disclosures can be found in the **Supplementary Information**.

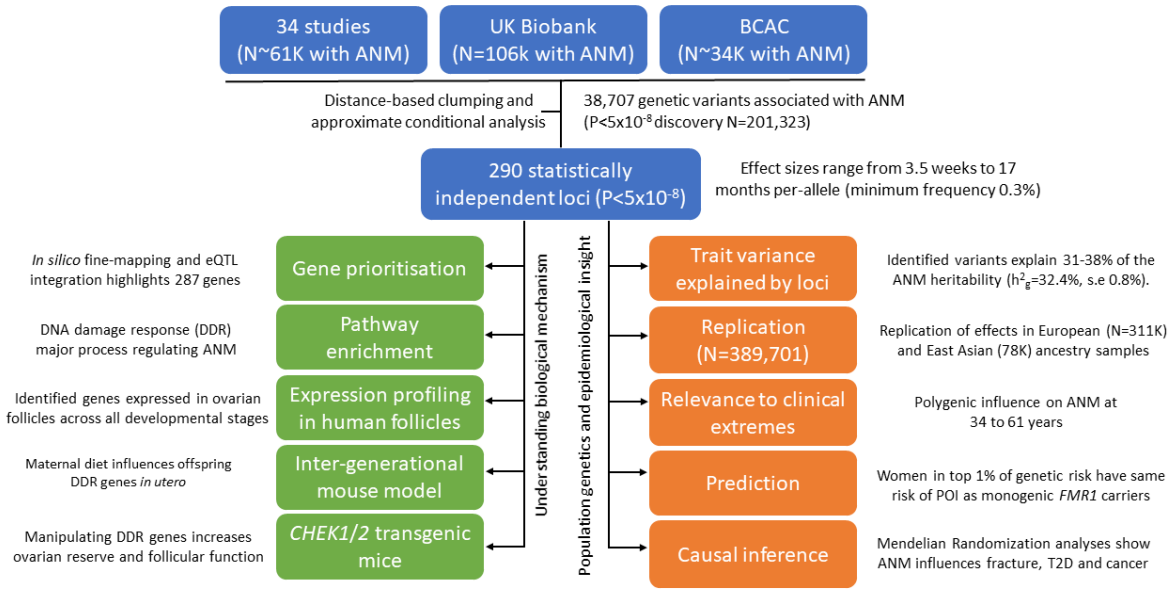
Supplementary Information is available for this paper.

Extended Data Figures

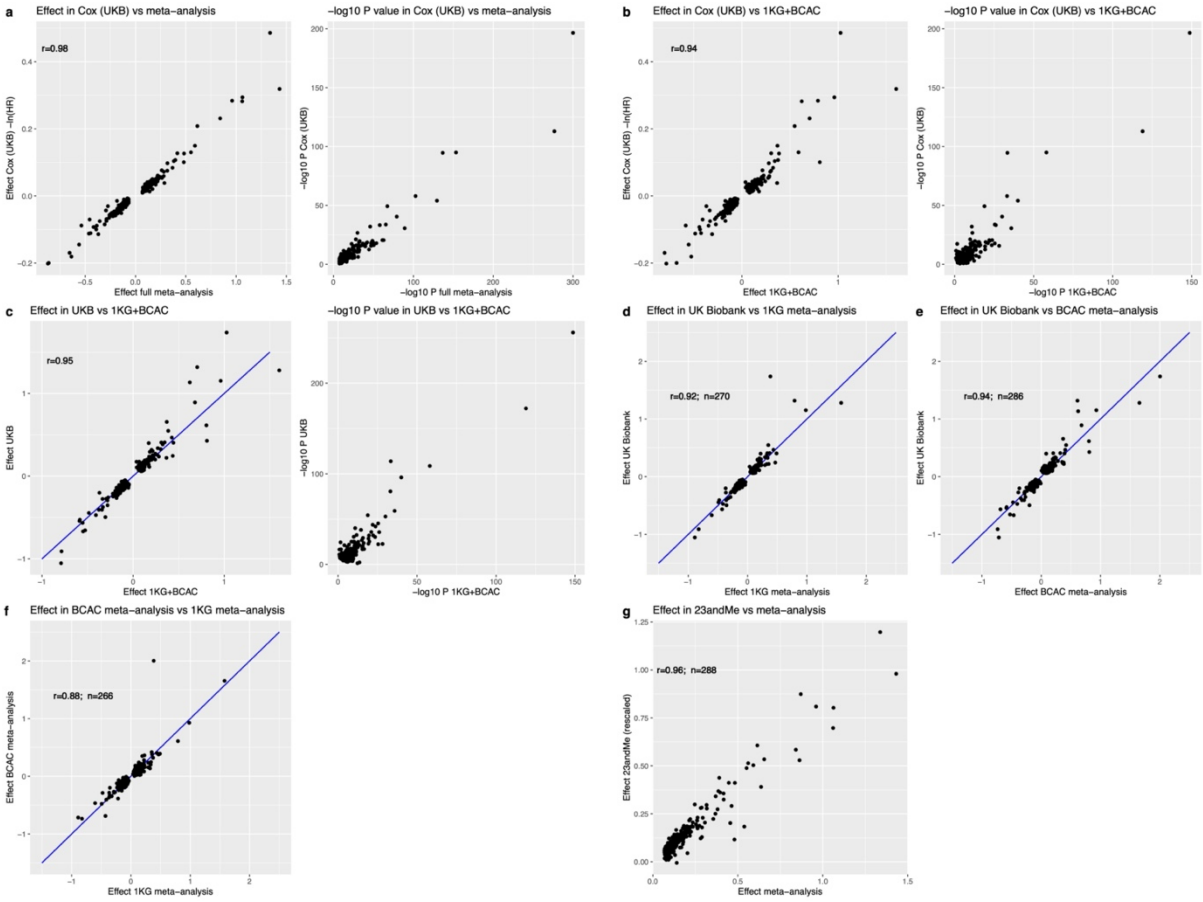
Extended Data Figure 1. Overview of ovarian reserve and follicular activity across reproductive life. **a,** Key processes involved in follicular activity from fetal development to menopause showing the numbers of oocytes at each stage; **b,** Summary of key biological pathways involved in follicular activity and their relationship to stage of reproductive life. Follicles, consisting of oocytes and surrounding granulosa cells are formed *in utero* and maintained as resting primordial follicles in the cortex constituting the ovarian reserve. Follicles are sequentially recruited from the ovarian reserve at a rate of several hundred per month in childhood, peaking at around 900 per month at ~15 years of age. Following recruitment, follicles grow by mitotic division of granulosa cells and expansion of oocyte volume for almost 6 months until meiosis is reinitiated at ovulation and the mature oocyte is released into the oviduct. Waves of atresia (follicle death) accompany developmental transitions and growing follicles are continuously induced to undergo cell death such that, typically, only a single follicle matures to ovulate each month. As ovarian reserve declines the rate of follicle recruitment decreases, but the preovulatory follicles continue to produce substantial amounts of oestrogen, while other important hormones such as anti-Müllerian hormone and inhibin-B decline, leading to upregulation of the hypothalamus-pituitary gonadal axis.



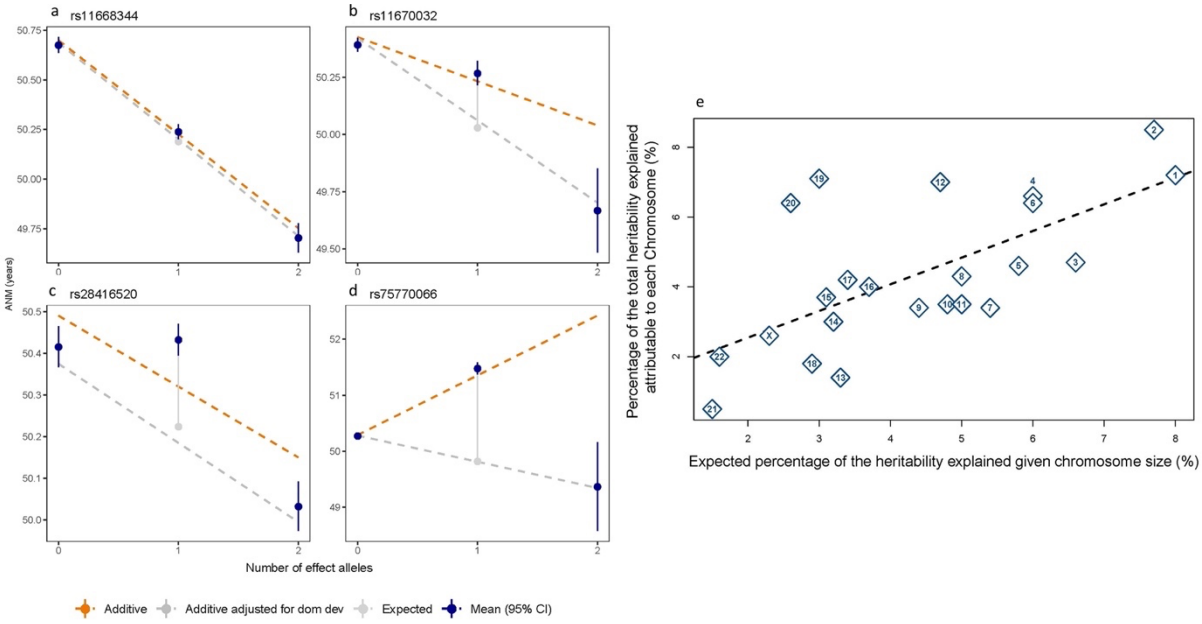
Extended Data Figure 2. Overview of performed analyses.



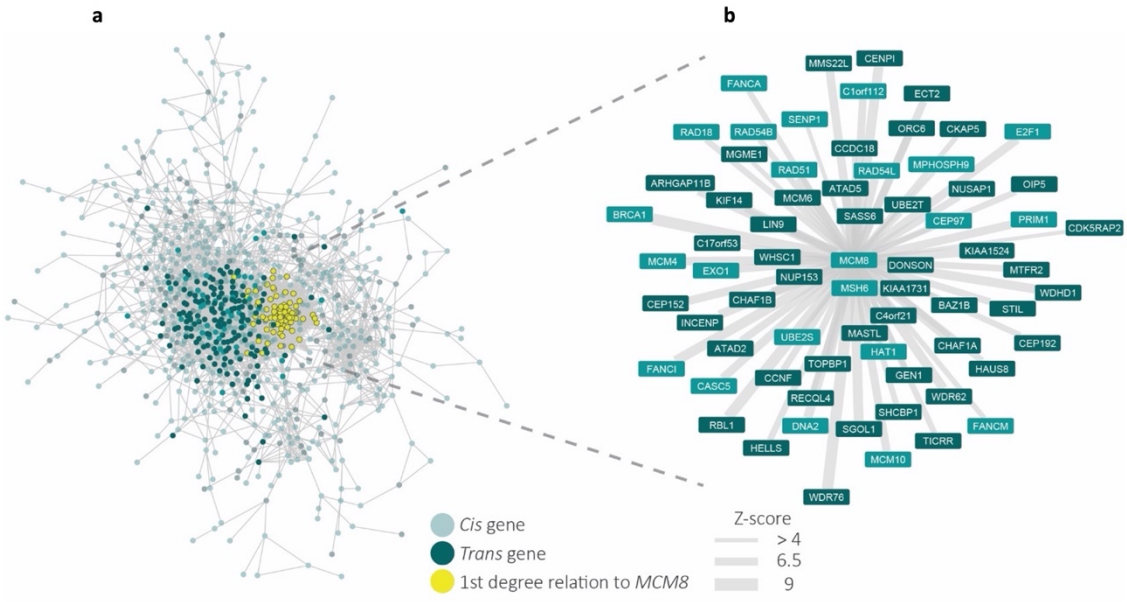
Extended Data Figure 3. Consistency of effect estimates across analyses methods and strata. Comparison of effect estimates from: **a,** Cox proportional hazards regression in UK Biobank with linear regression effect estimates from the overall meta-analysis ("Effect full meta-analysis"); **b,** Cox proportional hazards regression in UK Biobank with linear regression effect estimates from the meta-analysis excluding UK Biobank ("Effect 1KG+BCAC"); **c,** linear regression in UK Biobank with linear regression effect estimates from the meta-analysis excluding UK Biobank ("Effect 1KG+BCAC"). Comparison of linear regression effect estimates from: **d**, UK Biobank GWAS vs. the meta-analysis of 1000 Genomes imputed studies; **e**, UK Biobank GWAS vs. meta-analysis of samples from the Breast Cancer Association Consortium (BCAC); **f**, meta-analysis of BCAC samples vs. the meta-analysis of 1000 Genomes imputed studies; **g**, 23andMe replication analysis (rescaled) vs. overall meta-analysis. HR, hazard ratio from Cox proportional hazards model; r, Pearson correlation coefficient; blue line is y=x for reference. Note: *P* values < 1×10^{-300} are shown as 1×10^{-300}.



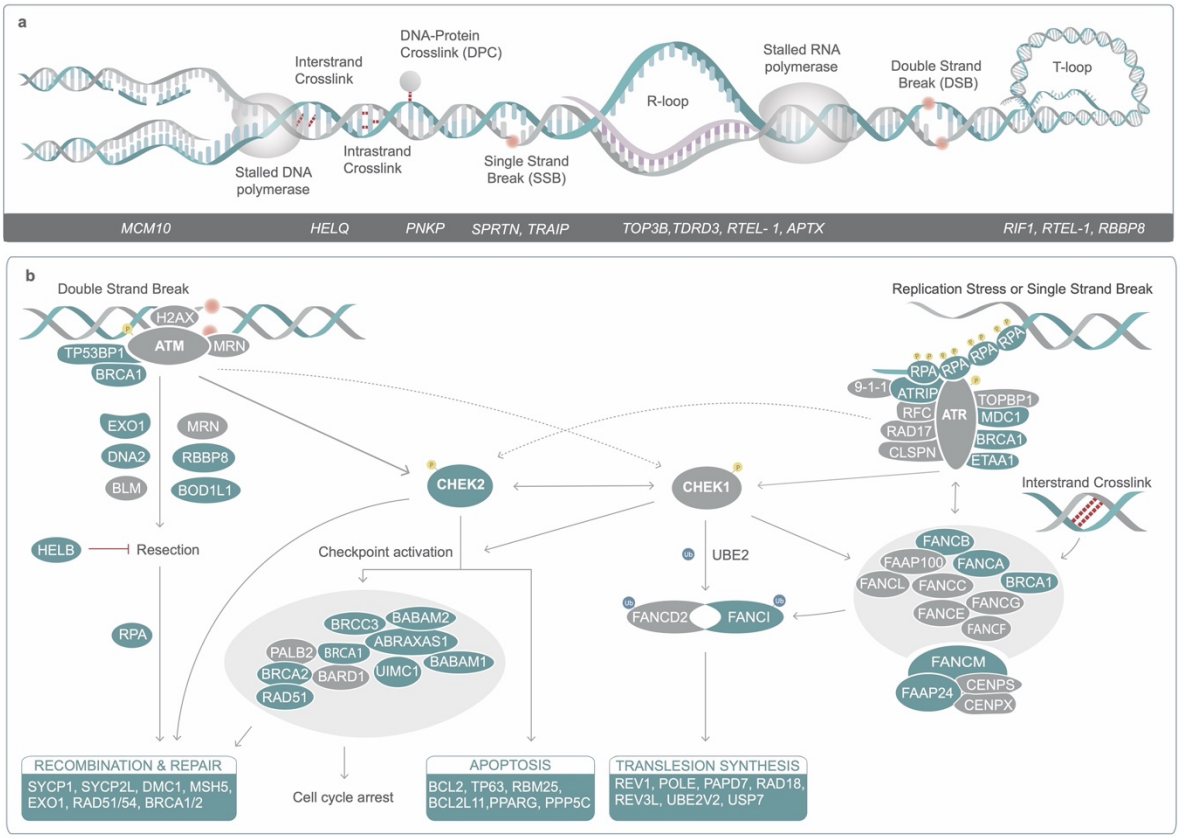
Extended Data Figure 4. Deviation from additive effects and distribution of estimated heritability across chromosomes. a-d, Genome-wide significant signals showing departure from an additive model. We tested the identified signals for departure from an additive allelic model. **a,** rs11668344 shows no deviation from an additive allelic model; **b,** rs11670032 and **c,** rs28416520 show deviation from the additive allelic model and a recessive effect; and **d,** rs75770066 shows a heterozygote effect. The mean and 95% confidence interval around the mean estimate are shown for each genotype. The expected mean ANM for the heterozygotes is the average of the mean ANM in the homozygote groups. The dashed orange line shows the effect estimate by genotype from linear regression based on an additive allelic model. Estimated ANM for each genotype was calculated as constant from regression model + number alleles × effect estimate from regression model. The dashed grey line indicates the additive effect estimate by genotype from a model adjusting for the dominance deviation effect of the heterozygote group (solid grey line). All regression models were adjusted for centre, genotyping chip and genetic principal components. ANM, age at natural menopause; dom dev, dominance deviation. **e,** The percentage of the total heritability explained that was attributable to each chromosome (observed heritability) is compared with the expected proportion calculated on the basis of chromosome size. The heritability of ANM was not uniformly distributed across chromosomes in proportion to their size. The X-chromosome did not explain more heritability than expected given its size, however chromosome 19 explained 2.36% [1.98-2.75] of the trait variance – greater than the individual contributions of nearly all larger chromosomes (weighted average for chromosomes 1-18: 1.7%, s.e 0.2%) and ~2.5x more than expected given its size. This was partially attributable to a single locus at 19q13 which explained ~0.75% trait variance and where we mapped 6 independent signals (**Supplementary Table 2**). The dashed line shows the mean ratio of expected to observed heritability across all chromosomes. Chromosome size was estimated based on the number of genetic variants.



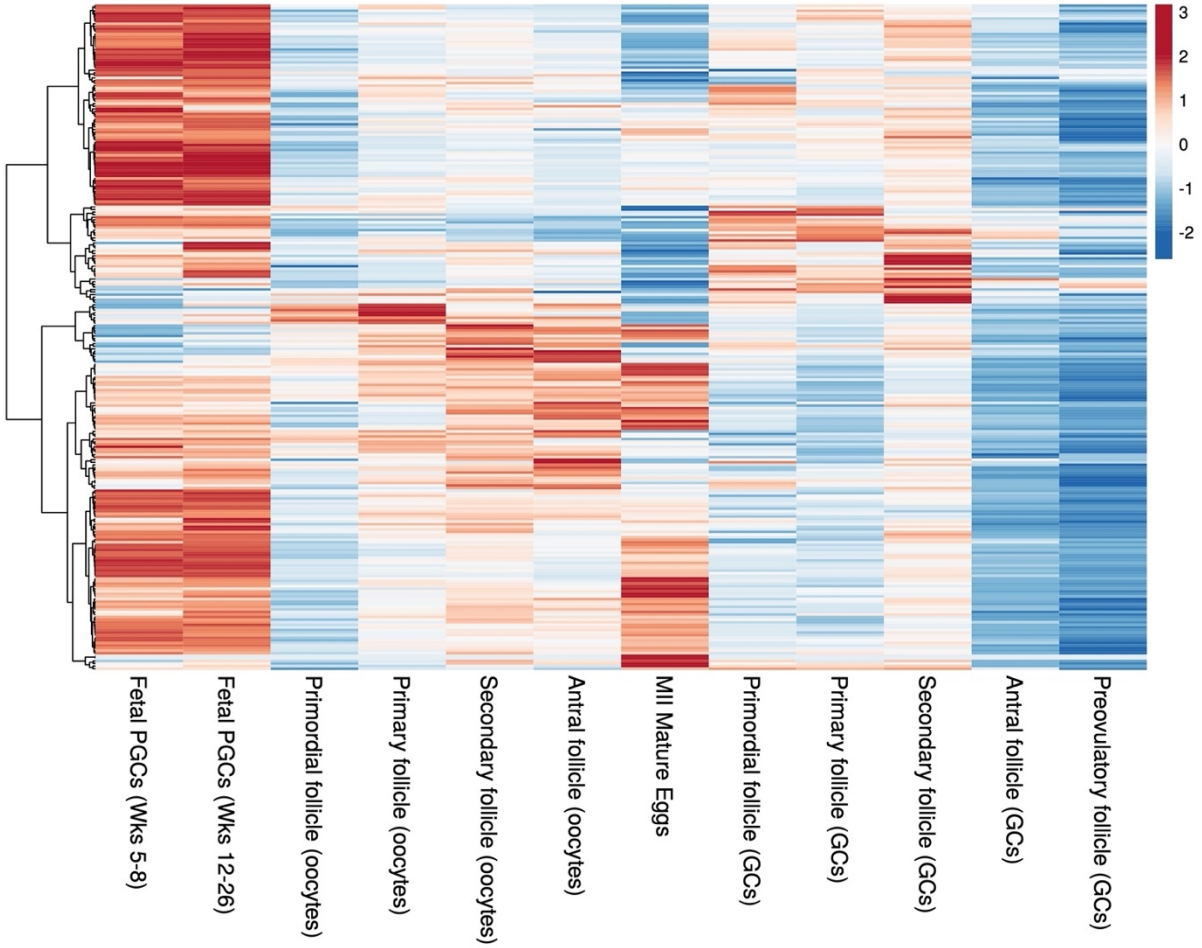
Extended Data Figure 5. Gene co-regulation networks for age at menopause genes with those co-regulated with *MCM8* highlighted. **a,** Gene co-regulation network for genes relating to age at menopause. Nodes indicate genes that either in a *cis* region from the GWAS or have been prioritized by Downstreamer, edges indicate a co-regulation relationship with a Z-score >4. Co-regulation is defined as the Pearson correlation between genes in a scaled eigenvector matrix derived from a multi-tissue gene network[68]. *Cis* genes are defined as genes that are within +/-300kb of a GWAS top hit for age at menopause. *Trans* genes are defined as having been prioritized by Downstreamer's co-regulation analysis and are not within +/-300kb of a GWAS top hit. Downstreamer prioritizes genes by associating the gene p-value profile of the GWAS (calculated using PASCAL[67]) to the co-regulation profile of each protein coding gene. Only genes where this association passes Bonferroni significance are shown as trans genes. Colours of nodes indicate the following: Teal indicates *Cis* genes, Dark Teal indicates *Trans* genes and Yellow indicates genes with a 1st degree relation to *MCM8*. **b,** Gene co-regulation network showing the genes that have a first degree relationship with *MCM8* with a Z-score >4. Width of the edge indicates the Z-score of the co-regulation relationship. Colours indicate the same as in **a,** with the exception of Yellow, as all genes indicated have a 1st degree relation to *MCM8.*



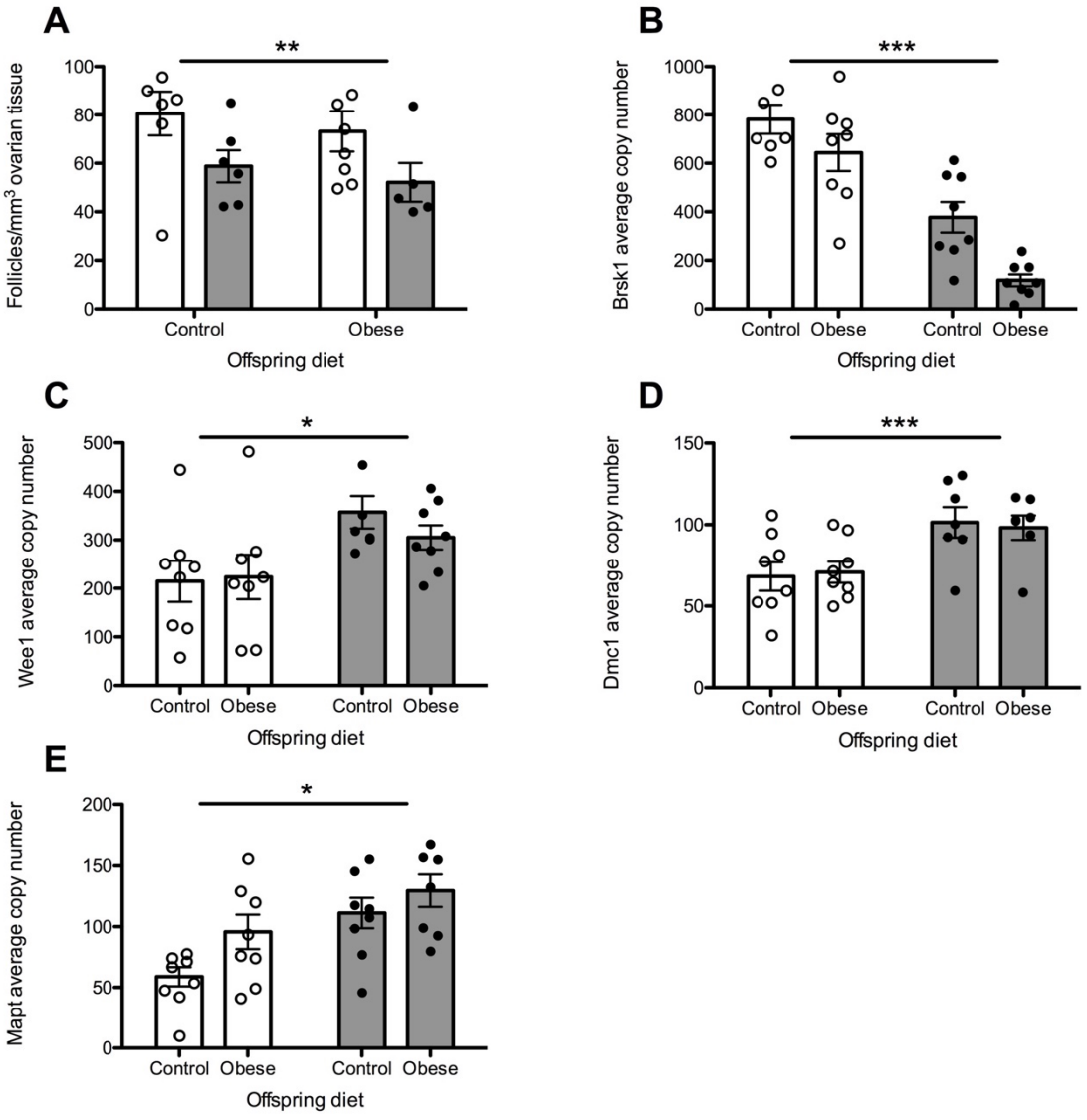
Extended Data Figure 6. DNA damage response and repair pathways implicated in reproductive ageing in humans. **a**, Consequences of replication stress annotated with genes involved that were within 300kb of the age at natural menopause (ANM) signals; **b**, Genes involved in downstream DNA damage response and repair pathways with those within 300kb of an ANM signal shown in blue. A full list of genes involved in DNA damage response and apoptosis annotated with genome-wide signals for ANM is provided in **Supplementary Table 19**. MRN, *MRN- MRE11-RAD50-NBS1* complex; RPA, Replication Protein A including a subunit encoded by *RPA1;* RFC, Replication Factor C including a subunit encoded by *RFC1*; 9-1-1, *RAD9-HUS1-RAD1* complex.



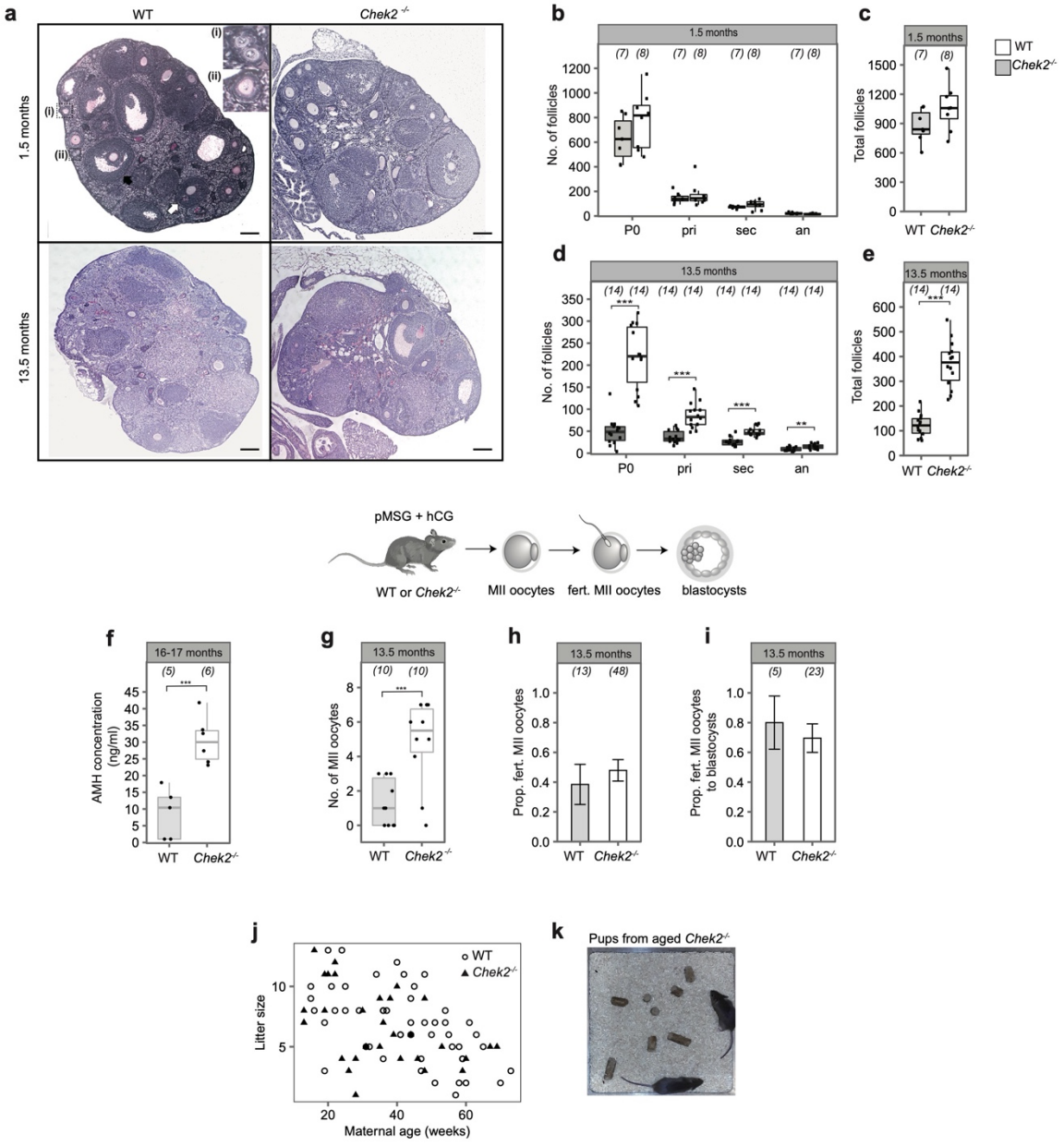
Extended Data Figure 7. Cluster plot of expression of consensus genes identified from the genome-wide analyses in germ cells across different developmental stages. Genes were selected from the GWAS signals, based on *in silico* prioritisation (**Supplementary Table 5**)**.** Of the 283 consensus genes highlighted by the GWAS, 258 passed QC and were available in the expression dataset. Gene expression was measured in human fetal primordial germ cells[80,81], and oocytes and granulosa cells in adult follicles (dataset generated in this study). Plot shows Z-scores, calculated by subtracting the mean transcripts per million (TPM) in all samples for a gene and dividing by the standard deviation. GC, granulosa cell; MII, meiosis II; PGC, primordial germ cell; Wks, weeks.



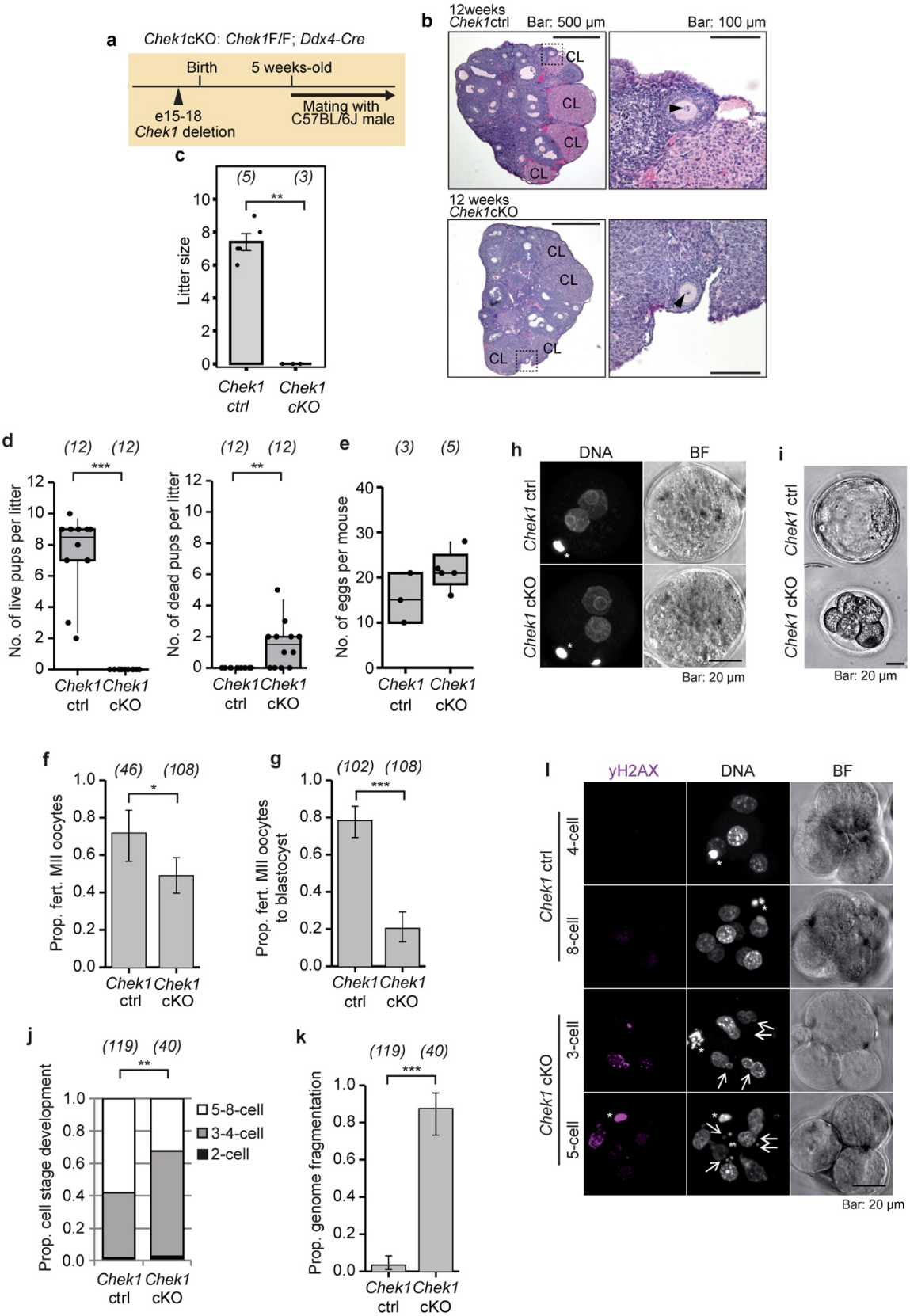
Extended Data Figure 8. Relationship between decreased ovarian reserve and gene expression. Open bar/dot groups – control maternal diet, normal ovarian reserve. Grey bar/dot groups: obesogenic maternal diet, reduced ovarian reserve. **a**, Ovarian follicular reserve in young adulthood in wild-type mice. Total follicles/mm^3 ovarian tissue at 12 weeks. Dots: individual observations. Bar heights and error bars: mean± SEM. n= 8 biologically independent animals from different litters in each group. *P*=0.0091 derived from 2-way ANOVA after correction for multiple hypothesis testing. **b,** *Brsk1* expression in the same animals, measured using qrtPCR and expressed as average copy number. Dots: individual observations. Bar heights and error bars: mean± SEM. n= 8 biologically independent animals from different litters in each group. *P*=0.0001 derived from 2-way ANOVA after correction for multiple hypothesis testing. **c,** *Wee1* expression in the same animals, measured using qrtPCR and expressed as average copy number. Dots: individual observations. Bar heights and error bars: mean± SEM. n= 8 biologically independent animals from different litters in each group. *P*=0.0256 derived from 2-way ANOVA after correction for multiple hypothesis testing. **d,** *Dmc1* expression in the same animals, measured using qrtPCR and expressed as average copy number. Dots: individual observations. Bar heights and error bars: mean± SEM. n= 8 biologically independent animals from different litters in each group. *P*=0.00001 derived from 2-way ANOVA after correction for multiple hypothesis testing. **e**, *Mapt* expression in the same animals, measured using qrtPCR and expressed as average copy number. Dots: individual observations. Bar heights and error bars: mean± SEM. n= 8 biologically independent animals from different litters in each group. *P*=0.0378 derived from 2-way ANOVA after correction for multiple hypothesis testing. qrtPCR, quantitative reverse transcription polymerase chain reaction; SEM, standard error of mean. *, *P*<0.05; **, *P*<0.01; ***, *P*<0.001.


A
Follicles/mm3 ovarian tissue
100
80
60
40
20
0
**
Control
Obese
Offspring diet
B
Brsk1 average copy number
1000
800
600
400
200
0
***
Control
Obese
Control
Obese
Offspring diet
C
Wee1 average copy number
500
400
300
200
100
0
*
Control
Obese
Control
Obese
Offspring diet
D
Dmc1 average copy number
150
100
50
0
***
Control
Obese
Control
Obese
Offspring diet
E
Mapt average copy number
200
150
100
50
0
*
Control
Obese
Control
Obese
Offspring diet

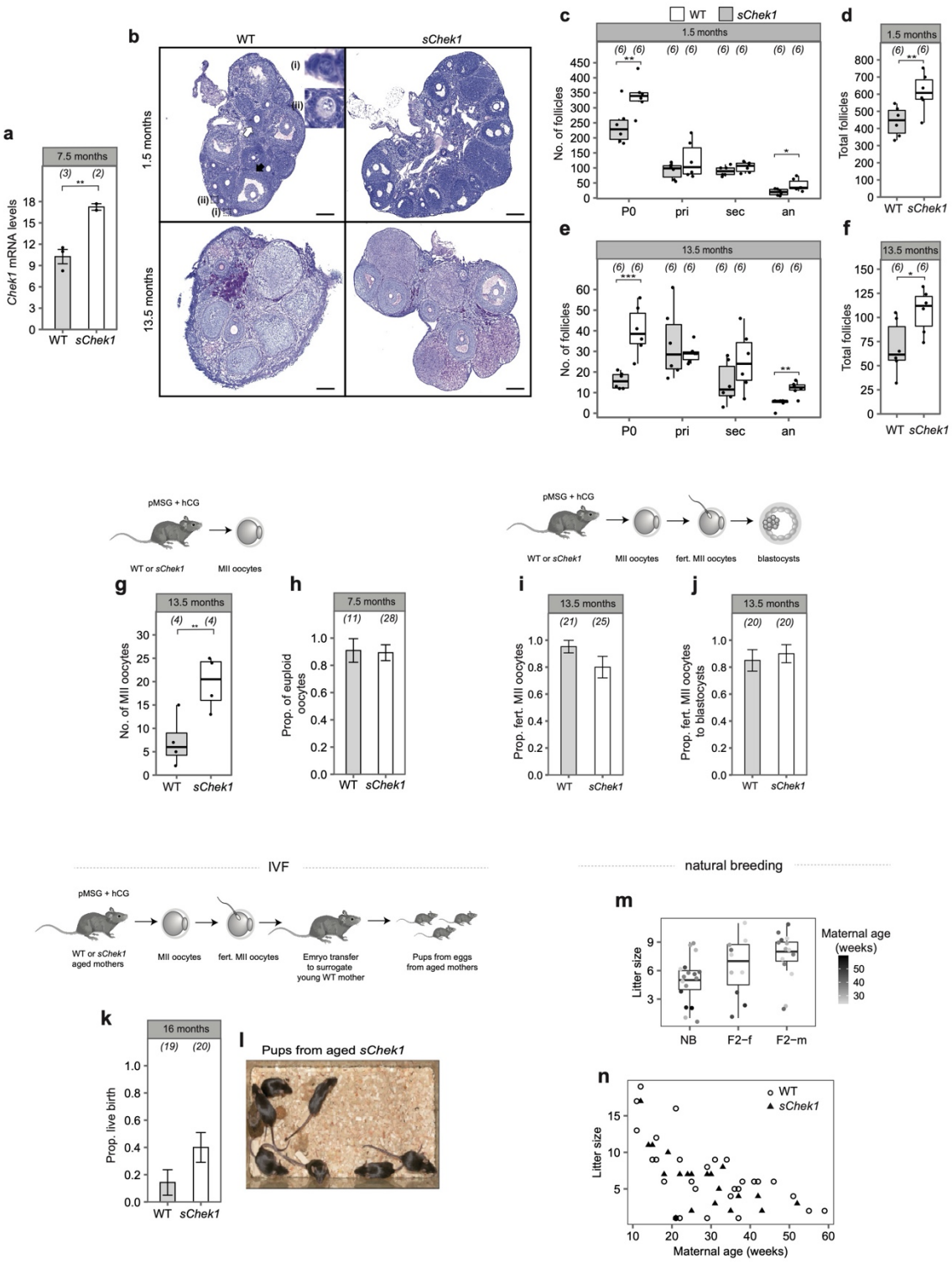
Extended Data Figure 9. *Chek2* deletion increases reproductive lifespan in mouse. a, Representative images of ovarian sections of 1.5-and 13.5-month-old wild type (WT) and *Chek2*$^{-/-}$ mice stained with PAS-Hematoxylin. Primordial follicles (inset (i)), primary follicles (inset (ii)), secondary follicle (white arrow) and antral follicle (black arrow) are shown. Scale bar represents 200 μm. **b-e,** Quantification of the number of follicles (by class and total) present in WT and *Chek2*$^{-/-}$ mice ovaries: **b, c,** 1.5-month-old; **d, e,** 13.5-month-old. The numbers in parentheses correspond to the total number of ovaries analysed. **f,** Serum AMH (ng/ml) in 16-17 months old *Chek2*$^{-/-}$ mice. The numbers in parentheses correspond to the number of mice assessed. **g-i,** Diagram illustrates the gonadotrophin stimulation of 13.5-month old females. Numbers in parentheses show: **g,** the number of MII oocytes retrieved per female; **h,** the number of MII oocytes fertilized; and **i,** the number of fertilized oocytes assessed for blastocyst formation. **j,** Litter size of WT and *Chek2*$^{-/-}$ females throughout the reproductive life span. Litter sizes from 9 WT and 5 *Chek2*$^{-/-}$ females are shown. Breeding cages contained one male and one female. Generalized linear model analysis showed maternal age effect, but no effect on genotype on litter sizes. **k,** Image of healthy pups born to 13 month-old *Chek2*$^{-/-}$ females. **b-i**, Two sample *t* and Fisher's exact tests were used to compare WT and *Chek2*$^{-/-}$ for statistical significance: *, $P<0.05$; **, $P<0.025$; ***, $P<0.001$. All *P*-values are two sided. Error bars indicate standard error of mean. Box-and-whisker plots show interquartile range and median (**b-g**). an, antral follicle; hCG, human chorionic gonadotrophin; pMSG, pregnant mare serum gonadotrophin; pri, primary follicle; P0, primordial follicle; sec, secondary follicle; WT=wildtype. Mouse strain: maintained on a mixed background, C57BL/6 129Sv, accession number BRC03481 at the RIKEN Bioresource Centre.


a
WT
Chek2 -/-
1.5 months
13.5 months
(i)
(ii)
b
1.5 months
No. of follicles
P0
pri
sec
an
c
1.5 months
Total follicles
WT
Chek2-/-
d
13.5 months
No. of follicles
e
13.5 months
Total follicles
pMSG + hCG
WT or Chek2-/-
MII oocytes
fert. MII oocytes
blastocysts
f
16-17 months
AMH concentration (ng/ml)
g
13.5 months
No. of MII oocytes
h
13.5 months
Prop. fert. MII oocytes
i
13.5 months
Prop. fert. MII oocytes to blastocysts
j
Litter size
Maternal age (weeks)
k
Pups from aged Chek2-/-

Extended Data Figure 10. Conditional knockout *Chek1* females are infertile due to requirement for *Chek1* during preimplantation embryo development. **a,** Schematic of the conditional-knockout mouse model of *Chek1* (*Chek1* cKO) in the female germline using the *Ddx4-Cre*. A similar approach was used for *Zp3-Cre*. **b,** In the ovarian sections stained with haematoxylin and eosin, we found follicles, corpora lutea (CL) and oocytes which contain nuclear structures (indicated with arrowheads in the magnified right hand panel). These findings suggest that estrus cycles and ovulation followed by corpus luteum formation are independent from *Chek1* disruption in oocytes *in vivo*. **c,** Litter size of *Chek1* cKO females. Three females older than 5 weeks age were mated with C57BL/6J males. Five independent littermate females (F/+, Tg-/Tg-; F/F, Tg-/Tg-; or F/+, Tg+/Tg-) were used as *Chek1* controls (ctrl). While *Chek1* ctrl females delivered normally, *Chek1* cKO females delivered no litters (**, Mann Whitney test, *P*=0.0179). Thus, these results indicate that *CHEK1* is essential in the female germline. **d,** Litter size of *Chek1*-cKO and controls using the *Zp3-Cre* during follicular growth. 3 months old control (*Chek1 F/F*; C*hek1* ctrl, n=4) and conditional knockout (*Chek1 F/F; Chek1* cKO with *Zp3*-Cre, n=4) were three-times consecutively mated with wild-type (*Chek1*$^{+/+}$) males, and the number of live (left) and dead (right) pups was monitored. While C*hek1* ctrl delivered a normal amount of live pups, *Chek1* cKO had only a reduced amount of perinatally dead pups (Mann-Whitney U Test: ***, *P*<0.001; **, *P* <0.01). Numbers in parentheses show the number of litters. **e,** The mean number of all ovulated eggs (the sum of MII oocytes and fertilized MII oocytes) per mouse with SEM (Mann Whitney U Test, *P* =0.126). Each data point presents the no. of eggs per mouse. 3-5 months old C*hek1* ctrl (n=3) and *Chek1* cKO (n=5) females were mated with wild-type (*Chek1*$^{+/+}$) males after pMSG + hCG stimulation. The number of ovulated eggs isolated 18 h post hCG stimulation and additional 10 h cultured *in vitro* was scored. The number of mice is shown in brackets. **f,** The proportion of fertilized MII oocytes to all ovulated eggs with a binomial confidence interval (*, Fisher's Exact Test, *P* =0.012; 95% CI 1.9–6.0; OR: 2.62). Numbers in parentheses show the total number of analysed eggs. **g,** The proportion of embryos that developed to blastocysts with binomial confidence interval (***, Fisher's Exact Test, *P*<0.0001). Fertilized MII oocytes (zygotes) were isolated from pMSG + hCG stimulated females 18h post hCG administration and cultured *in vitro* for 96 hours (~ E3.5) when development to blastocyst was scored. Data are pooled from four independent experiments. The number of embryos is shown in brackets. **h,** Fertilized eggs from C*hek1* ctrl (n=18) and *Chek1* cKO (n=13) females were fixed and stained for DNA (DAPI). All fertilized eggs from both genotypes showed normal pronuclei formation. The data were pooled from two independent experiments. Asterisks mark polar bodies. **i,** The majority of C*hek1* ctrl embryos formed blastocyst (g), but *Chek1* cKO embryos were arrested mainly in 3-8 cell stages. Representative bright-field images are shown. **j,** Proportion of developmental stages 2 cell, 3-4 cell and 5-8 cell (**, Cochran-Armitage Trend Test, *P*=0.0073). C*hek1* ctrl and *Chek1* cKO zygotes were isolated from 13 C*hek1* ctrl and 6 *Chek1* cKO pMSG + hCG stimulated females 18h post hCG administration and cultured *in vitro* for 49 hours. Embryos were fixed and stained for ɣH2AX by immunofluorescence. DNA was visualized by DAPI (**l**). **k,** Proportion of embryos with genome fragmentation with binomial confidence interval (***, Fisher's Exact Test, *P*<0.0001). Data are pooled from two independent experiments. The number of embryos is shown in brackets. **l,** C*hek1* ctrl and *Chek1* cKO zygotes (**j,k**) were fixed and stained for ɣH2AX (magenta) by immunofluorescence. DNA (gray) was visualized by DAPI. Arrows indicate genome fragments. Asterisks indicate polar bodies. These findings suggest that maternally expressed *Chek1* is critical for genome integrity protection during first divisions of preimplantation embryos in mice. All *P*-values are two sided. Box-and-whisker plots show interquartile range and median. Strains: C57BL/6-FVB mixed background for **a-c** (*Chek1* cKO*, Ddx4-Cre*); C57BL6-CD1 mixed background (*Chek1* cKO*, Zp3-Cre*) for panels **d-l**.


a
Chek1cKO: Chek1F/F; Ddx4-Cre
Birth
5 weeks-old
e15-18
Chek1 deletion
Mating with
C57BL/6J male
b
12weeks
Chek1ctrl
Bar: 500 µm
Bar: 100 µm
CL
12 weeks
Chek1cKO
c
(5) (3)
Litter size
Chek1 ctrl
Chek1 cKO
d
(12) (12)
No. of live pups per litter
No. of dead pups per litter
Chek1 ctrl
Chek1 cKO
e
(3) (5)
No. of eggs per mouse
h
DNA
BF
Chek1 ctrl
Chek1 cKO
Bar: 20 µm
i
f
(46) (108)
Prop. fert. MII oocytes
g
(102) (108)
Prop. fert. MII oocytes to blastocyst
j
(119) (40)
Prop. cell stage development
5-8-cell
3-4-cell
2-cell
k
(119) (40)
Prop. genome fragmentation
l
yH2AX
DNA
BF
4-cell
8-cell
3-cell
5-cell
Bar: 20 µm

Extended Data Figure 11. Extended reproductive lifespan in females carrying an extra copy of *Chek1 (sChek1).* **a,** mRNA expression levels of *Chek1* in oocytes, numbers in parentheses show the number of mice stimulated for retrieving the oocytes. **b,** Representative images of ovarian sections of 1.5 and 13.5-month-old wild type (WT) and *sChek1* mice stained with PAS-hematoxylin. Primordial follicles (inset (i)), primary follicles (inset (ii)), secondary follicle (white arrow) and antral follicle (black arrow) are shown. Scale bar: 200 μm. **c-f,** Quantification of the number of follicles (by class and total) present in WT and *sChek1* littermates: **c, d,** 1.5-month-old; **e, f,** 13.5-month-old. The numbers in parentheses correspond to the total number of ovaries analysed. **g-j,** MII oocytes retrieved in response to pMSG and hCG, proportion of euploid oocytes, proportion fertilized and proportion developed to blastocysts at different ages of WT and *sChek1* mice. Numbers in parentheses show: **g,** the number of MII oocytes retrieved per female; **h,** the number of oocytes assessed for aneuploidy; **i,** the number of MII oocytes fertilized; and **j,** the number of fertilized oocytes assessed for blastocyst development. **k,** Proportion of live births relative to transferred embryos from *in vitro* fertilized oocytes from aged mice (16 months), the numbers in parenthesis show the embryos transferred. **l,** Photo of healthy pups born to 16-month old *sChek1* females after IVF. **m,** Litter sizes from F2 females or males from aged *sChek1* females after IVF treatment in **k,** compared to females of equivalent ages that were naturally breeding. Note that for natural breeding there were two females and one male per breeding cage, whereas F2 cages contained a single male and one female. Therefore, litter sizes are an underestimate for the IVF-conceived pups. **n,** Litter sizes of WT and *sChek1* females throughout their reproductive life span. Data are from six breeding cages, three for each genotype. Each breeding cage contained one WT male and two females that were either WT or *sChek1.* Generalized linear model analysis showed maternal age effect, but no effect on genotype on litter sizes. **a-k,** Two sample *t* and Fisher's exact tests were used to compare WT and s*Chek1* for statistical significance: *, $P<0.05$; **, $P<0.025$; ***, $P<0.001$. All *P*-values are two sided. Error bars indicate standard error of mean. Box-and-whisker plots show interquartile range and median (**c-g, m**). an=antral follicle; hCG= human chorionic gonadotrophin; IVF=*in vitro* fertilization; NB=natural breeding; F2-f= F2 female; F2-m= F2 male; pMSG=pregnant mare serum gonadotrophin; pri=primary follicle; P0=primordial follicle; sec=secondary follicle; WT=wild type. Mouse strain: inbred from mixed background C57BL/6 129Sv.


a
7.5 months
(3) (2)
**
Chek1 mRNA levels
WT sChek1
b
WT
sChek1
1.5 months
13.5 months
(i)
(ii)
c
WT sChek1
1.5 months
(6) (6) (6) (6) (6) (6) (6) (6)
**
*
No. of follicles
P0 pri sec an
d
1.5 months
(6) (6)
**
Total follicles
WT sChek1
e
13.5 months
(6) (6) (6) (6) (6) (6) (6) (6)
***
**
No. of follicles
P0 pri sec an
f
13.5 months
(6) (6)
*
Total follicles
WT sChek1
pMSG + hCG
WT or sChek1
MII oocytes
fert. MII oocytes
blastocysts
g
13.5 months
(4) (4)
**
No. of MII oocytes
WT sChek1
h
7.5 months
(11) (28)
Prop. of euploid oocytes
WT sChek1
i
13.5 months
(21) (25)
Prop. fert. MII oocytes
WT sChek1
j
13.5 months
(20) (20)
Prop. fert. MII oocytes to blastocysts
WT sChek1
IVF
pMSG + hCG
WT or sChek1 aged mothers
MII oocytes
fert. MII oocytes
Emryo transfer to surrogate young WT mother
Pups from eggs from aged mothers
k
16 months
(19) (20)
Prop. live birth
WT sChek1
l
Pups from aged sChek1
natural breeding
m
Litter size
NB F2−f F2−m
Maternal age (weeks)
50
40
30
n
WT
sChek1
Litter size
Maternal age (weeks)

Supplementary Information

Contents:

Supplementary Results

Non-additive effects

Given that additive models can identify variants that exhibit stronger dominant or recessive effects[1], we tested all the identified signals for departure from an additive model. We identified three variants exhibiting non-additive effects (**Supplementary Table 3 and Extended Data Fig. 4 a-d**). For a common variant in *PIWIL1* (rs28416520, MAF=46%, $P=2\times10^{-14}$) a recessive model was the best fit (**Extended Data Fig. 4 c**). Deletion of *Piwil1* in mice results in sterility in males, but not females, and its role in human oogenesis is uncertain. It is however expressed as a dense paranuclear granule in human primordial follicle oocytes[2]. A low-frequency missense variant in *HELB* (rs75770066, MAF=3%, $P=7\times10^{-16}$) appeared to exhibit a heterozygous advantage effect (**Extended Data Fig. 4 d**), with higher mean ANM in the heterozygous group (95% CI 51.37-51.58 years) than the common (50.24-50.30) and rare homozygote (48.58-50.16) groups. Further fine-mapping and experimental work will be required to understand the complex biological mechanism(s) at this locus.

Menopause associated genes act across the life-course

Previous large-scale genetic analyses highlighted a clear involvement of homologous recombination and the *BRCA1-A* complex in the regulation of ovarian ageing. Our current study supports much broader DDR involvement, providing increased resolution of these pathways and informing when in the life-course they might act.

Our identified genes and pathway analyses strongly implicate repair pathways associated with replication stress, in particular removal of interstrand crosslinks, which covalently join both strands of the DNA helix, as well as DNA-protein crosslinks and R loops (DNA:RNA hybrids). All of these lesions stall DNA replication and prevent transcription (**Extended Data Fig. 6)**. This observation is supported by recent work demonstrating the role of the interstrand crosslink pathway *in utero* for resolving DNA damage in pre-meiotic, primordial germ cells[3]. This process begins with replication fork remodelling at interstrand crosslinks by *FANCM*[4–6], where we identify two independent ANM-associated missense variants (**Supplementary Table 4**). This subsequently leads to recruitment of the core Fanconi Anaemia (FA) complex to signal DNA damage, where we map missense variants in two of the eight genes – *FANCA* and *FANCB*. Furthermore we identify variants mapping key genes in the downstream repair systems coordinated by the FA pathway, including homologous recombination (e.g *RAD51*, *BRCA1*, *BRCA2*) as well as translesion synthesis (e.g *REV1*, *REV3L* and *RAD18*)[7].

Several DDR genes highlighted by our study have critical meiotic functions in fetal oocytes where at least 500 programmed double-strand breaks (DSBs) initiate recombination[8]. We implicate key recombination and synaptonemal complex genes with functions in meiotic prophase (*STAG3*, *SMC1β, EXO1*, *RAD51, DMC1, HELQ, RAD52, MSH5*). Mouse models of these genes show defective repair of meiotic recombination and subsequent apoptosis of fetal oocytes resulting in decreased primordial follicles from birth and infertility[9–17]. We note that several of our ANM-associated variants overlap those recently reported for recombination rate[18], however, despite more nominally significant associations than expected by chance, there was no clear relationship between the direction of effect on menopause and recombination rate across the 290 ANM loci (**Supplementary Table 26**).

A range of factors likely contribute to the rate at which follicles are recruited and the follicular reserve depleted. Our data implicate key genes in the mTOR complex 1 (mTORC1) in ANM, including *STK11* and *DEPTOR.* The mTOR protein kinase that controls cell growth by regulating protein and

nucleotide synthesis and is activated by the PI3K pathway. Oocyte-specific deletion of *Pten* in mice removes the inhibiting effect of the PI3K pathway on primordial follicle activation, leading to premature recruitment and exhaustion of the entire primordial follicle pool[19]. Other ANM-implicated genes include *FSHB, NOBOX, INHBB, INHBC, LHCGR, IGF1, IGFBP1, PPARG* and *BMPR1B,* highlighting broader endocrine and metabolic mechanisms governing ANM. We also identified common variants in *FTO* associated with ANM (**Supplementary Table 2)** which are distinct from the well-established body weight association in this region (r^2 with lead BMI variant rs1558902 = 0.0002).

Finally, the majority of known genes causing POI implicate aberrant DNA damage or the inability to repair it, with limited evidence in humans that defects in the downstream cell-death signaling pathways impact variation in reproductive ageing. In contrast, our study identifies more than 58 genes implicated in regulation of apoptosis associated with ANM (**Supplementary Table 19**), providing evidence that variation in cell death following DDR is an important mechanism. This includes components and interactors of the central, conserved DDR checkpoint kinases ATR-CHEK1 (single stranded DNA) and ATM-CHEK2 (double strand breaks), that integrate and determine repair and cellular response from a broad variety of DNA repair pathways (**Extended Data Fig. 6**).

Whilst the breadth of DDR pathways identified suggests our identified loci may exert their effect at different stages across the life-course, we sought to evaluate this by assessing patterns of germ cell gene expression across different developmental stages. The individual expression profiles of our 283 consensus genes (**Supplementary Table 2**) were assessed in human fetal primordial germ cells from 5 to 26 weeks gestation, in addition to oocyte and granulosa expression in adult follicles at different stages of growth (**Extended Data Fig. 7** and **Supplementary Table 20**). Collectively these data identified distinct clusters of genes that were active at different stages of life and follicle growth. The majority of our identified genes appeared most active in fetal primordial germ cells and fetal oocytes, however distinct expression profiles were evident across all developmental stages and between oocytes and granulosa cells (**Extended Data Fig. 7**). In many cases the pattern of expression was consistent with the known biological roles of those genes, for example Fanconi anemia genes were predominantly expressed in the fetal germ cells as well as the oocytes of the growing follicles, with less pronounced expression in granulosa cells (**Supplementary Table 20**). In contrast, genes such as *POLG* and *TP63* were predominantly expressed during follicular stages, consistent with apoptotic inducing activity in response to DNA damage observed in growing oocytes in mouse[20–23]. Further studies will be required to build on our observations and confirm the mechanism underlying the genetic associations.

Supplementary References

1. Wood, A. R. *et al.* Variants in the FTO and CDKAL1 loci have recessive effects on risk of obesity and type 2 diabetes, respectively. *Diabetologia* **59**, 1214–21 (2016).

2. Gomes Fernandes, M. *et al.* Human-specific subcellular compartmentalization of P-element induced wimpy testis-like (PIWIL) granules during germ cell development and spermatogenesis. *Hum. Reprod.* **33**, 258–269 (2018).

3. Hill, R. J. & Crossan, G. P. DNA cross-link repair safeguards genomic stability during premeiotic germ cell development. *Nat. Genet.* **51**, 1283–1294 (2019).

4. Meetei, A. R. *et al.* A human ortholog of archaeal DNA repair protein Hef is defective in Fanconi anemia complementation group M. *Nat. Genet.* **37**, 958–63 (2005).

5. Gari, K., Décaillet, C., Stasiak, A. Z., Stasiak, A. & Constantinou, A. The Fanconi anemia protein FANCM can promote branch migration of Holliday junctions and replication forks. *Mol. Cell* **29**, 141–8 (2008).

6. Gari, K., Décaillet, C., Delannoy, M., Wu, L. & Constantinou, A. Remodeling of DNA replication structures by the branch point translocase FANCM. *Proc. Natl. Acad. Sci. U. S. A.* **105**, 16107–12 (2008).

7. Sale, J. E., Lehmann, A. R. & Woodgate, R. Y-family DNA polymerases and their role in tolerance of cellular DNA damage. *Nat. Rev. Mol. Cell Biol.* **13**, 141–52 (2012).

8. Lenzi, M. L. *et al.* Extreme heterogeneity in the molecular events leading to the establishment of chiasmata during meiosis i in human oocytes. *Am. J. Hum. Genet.* **76**, 112–27 (2005).

9. Caburet, S. *et al.* Mutant cohesin in premature ovarian failure. *N. Engl. J. Med.* **370**, 943–949 (2014).

10. Titus, S. *et al.* Impairment of BRCA1-related DNA double-strand break repair leads to ovarian aging in mice and humans. *Sci. Transl. Med.* **5**, 172ra21 (2013).

11. Wei, K. *et al.* Inactivation of Exonuclease 1 in mice results in DNA mismatch repair defects, increased cancer susceptibility, and male and female sterility. *Genes Dev.* **17**, 603–14 (2003).

12. Abreu, C. M. *et al.* Shu complex SWS1-SWSAP1 promotes early steps in mouse meiotic recombination. *Nat. Commun.* **9**, 3961 (2018).

13. Yoshida, K. *et al.* The mouse RecA-like gene Dmc1 is required for homologous chromosome synapsis during meiosis. *Mol. Cell* **1**, 707–18 (1998).

14. Pittman, D. L. *et al.* Meiotic prophase arrest with failure of chromosome synapsis in mice deficient for Dmc1, a germline-specific RecA homolog. *Mol. Cell* **1**, 697–705 (1998).

15. Ward, J. D. *et al.* Overlapping mechanisms promote postsynaptic RAD-51 filament disassembly during meiotic double-strand break repair. *Mol. Cell* **37**, 259–72 (2010).

16. Revenkova, E. *et al.* Cohesin SMC1 beta is required for meiotic chromosome dynamics, sister chromatid cohesion and DNA recombination. *Nat. Cell Biol.* **6**, 555–62 (2004).

17. de Vries, S. S. *et al.* Mouse MutS-like protein Msh5 is required for proper chromosome synapsis in male and female meiosis. *Genes Dev.* **13**, 523–31 (1999).

18. Halldorsson, B. V *et al.* Characterizing mutagenic effects of recombination through a

sequence-level genetic map. *Science* **363**, (2019).

19. Reddy, P. *et al.* Oocyte-specific deletion of Pten causes premature activation of the primordial follicle pool. *Science* **319**, 611–3 (2008).

20. Suh, E.-K. *et al.* p63 protects the female germ line during meiotic arrest. *Nature* **444**, 624–8 (2006).

21. Deutsch, G. B. *et al.* DNA damage in oocytes induces a switch of the quality control factor TAp63α from dimer to tetramer. *Cell* **144**, 566–76 (2011).

22. Tuppi, M. *et al.* Oocyte DNA damage quality control requires consecutive interplay of CHK2 and CK1 to activate p63. *Nat. Struct. Mol. Biol.* **25**, 261–269 (2018).

23. Rinaldi, V. D., Bloom, J. C. & Schimenti, J. C. Oocyte Elimination Through DNA Damage Signaling from CHK1/CHK2 to p53 and p63. *Genetics* **215**, 373–378 (2020).

Supplementary Notes

1) Acknowledgments and sources of funding for studies included in the genome-wide association analyses

Study acronym	Full study name	Acknowledgments and sources of funding
23andMe	23andMe	We would like to thank the research participants and employees of 23andMe for making this work possible. The following members of the 23andMe Research Team contributed to this study: Michelle Agee, Stella Aslibekyan, Adam Auton, Elizabeth Babalola, Robert K. Bell, Jessica Bielenberg, Katarzyna Bryc, Emily Bullis, Briana Cameron, Daniella Coker, Gabriel Cuellar Partida, Devika Dhamija, Sayantan Das, Sarah L. Elson, Teresa Filshtein, Kipper Fletez-Brant, Will Freyman, Pooja M. Gandhi, Karl Heilbron, Barry Hicks, David A. Hinds, Karen E. Huber, Ethan M. Jewett, Yunxuan Jiang, Aaron Kleinman, Katelyn Kukar, Vanessa Lane, Keng-Han Lin, Maya Lowe, Marie K. Luff, Jennifer C. McCreight, Matthew H. McIntyre, Kimberly F. McManus, Steven J. Micheletti, Meghan E. Moreno, Joanna L. Mountain, Sahar V. Mozaffari, Priyanka Nandakumar, Elizabeth S. Noblin, Jared O'Connell, Aaron A. Petrakovitz, G. David Poznik, Morgan Schumacher, Anjali J. Shastri, Janie F. Shelton, Jingchunzi Shi, Suyash Shringarpure, Chao Tian, Vinh Tran, Joyce Y. Tung, Xin Wang, Wei Wang, Catherine H. Weldon, Peter Wilton
AGES	Age, Gene/Environment Susceptibility-Reykjavik Study	This study has been funded by NIH contracts N01-AG-1-2100 and 271201200022C, the NIA Intramural Research Program, Hjartavernd (the Icelandic Heart Association), and the Althingi (the Icelandic Parliament). The study is approved by the Icelandic National Bioethics Committee, VSN: 00-063. The researchers are indebted to the participants for their willingness to participate in the study.
ALSPAC	Avon Longitudinal Study of Parents and Children	We are extremely grateful to all the families who took part in this study, the midwives for their help in recruiting them, and the whole ALSPAC team, which includes interviewers, computer and laboratory technicians, clerical workers, research scientists, volunteers, managers, receptionists and nurses. The UK Medical Research Council and the Wellcome Trust (Grant ref: 217065/Z/19/Z) and the University of Bristol provide core support for ALSPAC. GWAS data for ALSPAC mothers was funded by the Wellcome Trust (WT088806) and phenotype data by the British Heart Foundation (SP/07/008/24066), Wellcome Trust (WT092830M) and MRC (G1001357). GWAS data for ALSPAC offspring was generated by Sample Logistics and Genotyping Facilities at the Wellcome Trust Sanger Institute and LabCorp (Laboratory Corporation of America) using support from 23andMe. DAL, NJT, SMR and GDS work in a Unit that receives support from the University of Bristol and MRC (MC_UU_00011/1 and MC_UU_00011/6).
ARIC	Atherosclerosis Risk in Communities Study	The Atherosclerosis Risk in Communities study has been funded in whole or in part with Federal funds from the National Heart, Lung, and Blood Institute, National Institutes of Health, Department of Health and Human Services, under Contract nos. (HHSN268201700001I, HHSN268201700002I, HHSN268201700003I, HHSN268201700005I, HHSN268201700004I). The authors thank the staff and participants of the ARIC study for their important contributions. The Atherosclerosis Risk in Communities study has been funded in whole or in part with Federal funds from the National Heart, Lung, and Blood Institute, National Institutes of Health, Department of Health and Human Services (contract numbers HHSN268201700001I, HHSN268201700002I, HHSN268201700003I, HHSN268201700004I and HHSN268201700005I), R01HL087641, R01HL086694; National Human Genome Research Institute contract U01HG004402; and National Institutes of Health contract HHSN268200625226C. The authors thank the staff and participants of the ARIC study for their important contributions. Infrastructure was partly supported by Grant Number UL1RR025005, a component of the National Institutes of Health and NIH Roadmap for Medical Research.

BCAC/ iCOGs	Breast Cancer Association Consortium/ iCOGs	See separate section below for details of individual studies.
CARL	INGI-CARLANTINO	We would like to thank all the participants in the study for their contribution and support The study was supported through the Italian Ministry of Health
CHS	Cardiovascular Health Study	This CHS research was supported by NHLBI contracts HHSN268201200036C, HHSN268200800007C, HHSN268201800001C, N01HC55222, N01HC85079, N01HC85080, N01HC85081, N01HC85082, N01HC85083, N01HC85086; and NHLBI grants U01HL080295, R01HL087652, R01HL105756, R01HL103612, R01HL120393, U01HL130114 and 75N92021D00006 with additional contribution from the National Institute of Neurological Disorders and Stroke (NINDS). Additional support was provided through R01AG023629 from the National Institute on Aging (NIA). A full list of principal CHS investigators and institutions can be found at CHS-NHLBI.org. The provision of genotyping data was supported in part by the National Center for Advancing Translational Sciences, CTSI grant UL1TR001881, and the National Institute of Diabetes and Digestive and Kidney Disease Diabetes Research Center (DRC) grant DK063491 to the Southern California Diabetes Endocrinology Research Center. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.
CILENTO	CILENTO	We thank the populations of Cilento for their participation in the study. This work was supported by grants from the Italian Ministry of Universities (FIRB-RBNE08NKH7, INTEROMICS Flagship Project), the Assessorato Ricerca Regione Campania, the Fondazione con il SUD (2011-PDR-13) and the Istituto Banco di Napoli - Fondazione to MC.
COLAUS	CoLaus (Etude Cohorte Lausannoise)	The CoLaus study was and is supported by research grants from GlaxoSmithKline, the Faculty of Biology and Medicine of Lausanne, and the Swiss National Science Foundation (grants 33CSCO-122661, 33CS30-139468, 33CS30-148401 and 33CS30_177535/1).
CROATIA Korcula	CROATIA Korcula	We would like to acknowledge the contributions of the recruitment team in Korcula, the administrative teams in Croatia and Edinburgh and the people of Korcula. The SNP genotyping for the KORCULA cohort was performed in Helmholtz Zentrum München, Neuherberg, Germany. Medical Research Council UK and the Ministry of Science, Education and Sport in the Republic of Croatia (number 108-1080315-0302).
CROATIA Vis	CROATIA Vis	We would like to acknowledge the staff of several institutions in Croatia that supported the field work, including but not limited to The University of Split and Zagreb Medical Schools, Institute for Anthropological Research in Zagreb and Croatian Institute for Public Health. Medical Research Council UK and the Ministry of Science, Education and Sport in the Republic of Croatia (number 108-1080315-0302).
EGCUT-370 and OmniX	Estonian Genome Center, University of Tartu	This study was supported by EU H2020 grants 692145, 676550, 654248, Estonian Research Council Grant IUT20-60, NIASC, and EU through the European Regional Development Fund (Project No. 2014-2020.4.01.15-0012 GENTRANSMED).
EPIC-Norfolk	The EPIC-Norfolk Study	The EPIC-Norfolk study (DOI 10.22025/2019.10.105.00004) has received funding from the Medical Research Council (MR/N003284/1 MC-UU_12015/1 and MC_UU_00006/1) and Cancer Research UK (C864/A14136). The genetics work in the EPIC-Norfolk study was funded by the Medical Research Council (MC_PC_13048). We are grateful to all the participants who have been part of the project and to the many members of the study teams at the University of

		Cambridge who have enabled this research.
FHS	Framingham Heart Study	The authors thank the Framingham Heart Study participants and staff. The Framingham Heart Study phenotype-genotype analyses were supported by NIA R01AG29451 JMM, KLL). The Framingham Heart Study of the National Heart Lung and Blood Institute of the National Institutes of Health and Boston University School of Medicine was supported by the National Heart, Lung and Blood Institute's Framingham Heart Study Contract No. N01-HC-25195 and its contract with Affymetrix, Inc for genotyping services (Contract No. N02-HL-6-4278). Analyses reflect intellectual input and resource development from the Framingham Heart Study investigators participating in the SNP Health Association Resource (SHARe) project. A portion of this research was conducted using the Linux Cluster for Genetic Analysis (LinGA-II) funded by the Robert Dawson Evans Endowment of the Department of Medicine at Boston University School of Medicine and Boston Medical Center. Framingham Heart Study Contract No. N01-HC-25195, HHSN268201500001 The authors are pleased to acknowledge that the computational work reported on in this paper was performed on the Shared Computing Cluster which is administered by Boston University's Research Computing Services. URL: www.bu.edu/tech/support/research/. The authors thank the participants for their dedication to the study.
FVG	INGI- FRIULI VENEZIA GIULIA	We would like to thank all the participants in the study for their contribution and support The study was supported by Regione FVG (L.26.2008) and Italian Ministry of Health
GS	Generation Scotland: Scottish Family Health Study	We are grateful to all the families who took part, the general practitioners and the Scottish School of Primary Care for their help in recruiting them, and the whole Generation Scotland team, which includes interviewers, computer and laboratory technicians, clerical workers, research scientists, volunteers, managers, receptionists, healthcare assistants and nurses. Generation Scotland received core support from the Chief Scientist Office of the Scottish Government Health Directorates [CZD/16/6] and the Scottish Funding Council [HR03006] and is currently supported by the Wellcome Trust [216767/Z/19/Z]. Genotyping of the GS:SFHS samples was carried out by the Genetics Core Laboratory at the Edinburgh Clinical Research Facility, University of Edinburgh, Scotland and was funded by the Medical Research Council UK and the Wellcome Trust (Wellcome Trust Strategic Award "STratifying Resilience and Depression Longitudinally" (STRADL) Reference 104036/Z/14/Z).
HANDLS		
Health ABC	The Health, Aging, and Body Composition Study	This study utilized the high-performance computational capabilities of the Biowulf Linux cluster at the National Institutes of Health, Bethesda, Md. (http://biowulf.nih.gov). The Health ABC Study was supported by NIA contracts N01AG62101, N01AG62103, and N01AG62106 and, in part, by the NIA Intramural Research Program. The genome-wide association study was funded by NIA grant 1R01AG032098-01A1 to Wake Forest University Health Sciences and genotyping services were provided by the Center for Inherited Disease Research (CIDR). CIDR is fully funded through a federal contract from the National Institutes of Health to The Johns Hopkins University, contract number HHSN268200782096C.

HRS	Health and Retirement Study	HRS is supported by the National Institute on Aging (NIA U01AG009740). The genotyping was funded separately by the National Institute on Aging (RC2 AG036495, RC4 AG039029). Our genotyping was conducted by the NIH Center for Inherited Disease Research (CIDR) at Johns Hopkins University. Genotyping quality control and final preparation of the data were performed by the Genetics Coordinating Center at the University of Washington.
INCHIANTI	Invecchiare in Chianti	The InCHIANTI study baseline (1998-2000) was supported as a "targeted project" (ICS110.1/RF97.71) by the Italian Ministry of Health and in part by the U.S. National Institute on Aging (Contracts: 263 MD 9164 and 263 MD 821336).
InterAct cases and cohort	European Prospective Investigation into Cancer & Nutrition - InterAct	We thank all EPIC participants and staff and the InterAct Consortium members for their contributions to the study. The InterAct project received funding from the European Union (Integrated Project LSHM-CT-2006-037197 in the Framework Programme 6 of the European Community). We thank staff from the technical, field epidemiology and data teams of the Medical Research Council Epidemiology Unit in Cambridge, UK, for carrying out sample preparation, DNA provision and quality control, genotyping and data handling work.
KORA F3 / KORA F4	KORA F3 and F4: Cooperative Health Research in the Augsburg Region, Follow up for Study S3 and S4	The KORA study was initiated and financed by the Helmholtz Zentrum München – German Research Center for Environmental Health, which is funded by the German Federal Ministry of Education and Research (BMBF) and by the State of Bavaria. The KORA-Study Group consists of A. Peters (speaker), H. Schulz, R. Holle, R. Leidl, C. Meisinger, K. Strauch, and their co-workers, who are responsible for the design and conduct of the KORA studies. Furthermore, KORA research was supported within the Munich Center of Health Sciences (MC-Health), Ludwig-Maximilians-Universität, as part of LMUinnovativ. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. We thank all study participants and the study staff.
LifeLines	The LifeLines Cohort Study	The authors wish to acknowledge Behrooz Z. Alizadeh, Annemieke Boesjes, Marcel Bruinenberg, Noortje Festen, Pim van der Harst, Ilja Nolte, Lude Franke, Mitra Valimohammadi for their help in creating the GWAS database, and Rob Bieringa, Joost Keers, René Oostergo, Rosalie Visser, Judith Vonk for their work related to data-collection and validation. The authors are grateful to the study participants, the staff from the LifeLines Cohort Study and the contributing research centers delivering data to LifeLines and the participating general practitioners and pharmacists. The Lifelines Biobank initiative has been made possible by subsidy from the Dutch Ministry of Health, Welfare and Sport, the Dutch Ministry of Economic Affairs, the University Medical Center Groningen (UMCG the Netherlands), University Groningen and the Northern Provinces of the Netherlands. The Lifelines Cohort Study was supported by the Netherlands Organization for Scientific Research (NWO) [grant 175.010.2007.006]; the Economic Structure Enhancing Fund (FES) of the Dutch government; the Ministry of Economic Affairs; the Ministry of Education, Culture and Science; the Ministry for Health, Welfare and Sports; the Northern Netherlands Collaboration of Provinces (SNN); the Province of Groningen; University Medical Center Groningen; the University of Groningen; the Dutch Kidney Foundation; and the Dutch Diabetes Research Foundation. The Lifelines Cohort Study is supported by the National Consortium for Healthy Ageing and the BioSHaRE-EU consortium (KP7, project reference 261433). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

NEO	Netherlands Epidemiology of Obesity	The authors of the NEO study thank all individuals who participated in the Netherlands Epidemiology in Obesity study, all participating general practitioners for inviting eligible participants and all research nurses for collection of the data. We thank the NEO study group, Pat van Beelen, Petra Noordijk and Ingeborg de Jonge for the coordination, lab and data management of the NEO study. The genotyping in the NEO study was supported by the Centre National de Génotypage (Paris, France), headed by Jean-Francois Deleuze. The NEO study is supported by the participating Departments, the Division and the Board of Directors of the Leiden University Medical Center, and by the Leiden University, Research Profile Area Vascular and Regenerative Medicine. Dennis Mook-Kanamori is supported by Dutch Science Organization (ZonMW-VENI Grant 916.14.023).
NHS Affymetrix / NHS Illumina / NHS Omni-Express	The Nurses' Health Study (NHS)	The NHS GWAS were supported by grants from the National Institutes of Health [NCI (CA40356, CA087969, CA055075, CA98233, U01 CA137088, R01 CA059045, R01 CA137178, R01 CA082838, R01 CA131332), NIDDK (DK058845, DK070756), NHGRI (HG004399, HG004728), NHLBI (HL35464), NIAMS (R01 AR056291)]. We would like to thank the participants and staff of the NHS and NHSII for their valuable contributions as well as the following state cancer registries for their help: AL, AZ, AR, CA, CO, CT, DE, FL, GA, ID, IL, IN, IA, KY, LA, ME, MD, MA, MI, NE, NH, NJ, NY, NC, ND, OH, OK, OR, PA, RI, SC, TN, TX, VA, WA, WY. The authors assume full responsibility for analyses and interpretation of these data.
NTR	The Netherlands Twin Register	The Netherland Twin Register: would like to thank all study participants for their contributions to our scientific efforts, the SURF SARA institute for computational resources and the Avera institute of Human Genetics for genotyping of samples. Funding was obtained from the Netherlands Organization for Scientific Research (NWO) and The Netherlands Organization for Health Research and Development (ZonMW) grants 904-61-090, 985-10-002, 912-10-020, 904-61-193,480-04-004, 463-06-001, 451-04-034, 400-05-717, Addiction-31160008, 016-115-035, 481-08-011, 400-07-080, 056-32-010, Middelgroot-911-09-032, OCW_NWO Gravity program –024.001.003, NWO-Groot 480-15-001/674, Center for Medical Systems Biology (CSMB, NWO Genomics), NBIC/BioAssist/RK(2008.024), Biobanking and Biomolecular Resources Research Infrastructure (BBMRI –NL, 184.021.007 and 184.033.111), X-Omics 184-034-019; Spinozapremie (NWO- 56-464-14192), KNAW Academy Professor Award (PAH/6635) and University Research Fellow grant (URF) to DIB; Amsterdam Public Health research institute (former EMGO+) , Neuroscience Amsterdam research institute (former NCA); Amsterdam Research & Development (AR&D) research institute; the European Community's Fifth and Seventh Framework Program (FP5- LIFE QUALITY-CT-2002-2006, FP7- HEALTH-F4-2007-2013, grant 01254: GenomEUtwin, grant 01413: ENGAGE and grant 602768: ACTION); the European Research Council (ERC Starting 284167, ERC Consolidator 771057, ERC Advanced 230374), Rutgers University Cell and DNA Repository (NIMH U24 MH068457-06), the National Institutes of Health (NIH, R01D0042157-01A1, R01MH58799-03, MH081802, DA018673, R01 DK092127-04, Grand Opportunity grants 1RC2 MH089951, and 1RC2 MH089995); the Avera Institute for Human Genetics, Sioux Falls, South Dakota (USA). Part of the genotyping and analyses were funded by the Genetic Association Information Network (GAIN) of the Foundation for the National Institutes of Health. Computing was supported by NWO through grant 2018/EW/00408559, BiG Grid, the Dutch e-Science Grid and SURFSARA.
ORCADES	Orkney Complex Disease Study	DNA extractions were performed at the Genetics Core Laboratory at the Edinburgh Clinical Research Facility, University of Edinburgh, Scotland. We would like to acknowledge the invaluable contributions of the research nurses in Orkney, the administrative team in Edinburgh and the people of Orkney. ORCADES was supported by the Chief Scientist Office of the Scottish Government (CZB/4/276, CZB/4/710), the Royal Society, the MRC Human Genetics Unit, Arthritis Research UK and the European Union framework program 6 EUROSPAN project (contract no. LSHG-CT-2006-018947).

QIMR	QIMR Adult Cohort	We thank the participants and their families for contributing to this research. We also thank A Henders, B Usher, E Souzeau, A Kuot, A McMellon, MJ Wright, MJ Campbell, A Caracella, L Bowdler, S Smith, B Haddon, A Conciatore, D Smyth, H Beeby, O Zheng and B Chapman for their input into project management, databases, phenotype collection, and sample collection, processing and genotyping. The QIMR cohort was supported by National Institutes of Health (NIH) Grants AA07535, AA07728, AA13320, AA13321, AA14041, AA11998, AA17688, DA012854, DA019951, AA010249, AA013320, AA013321, AA011998, AA017688, and DA027995; by Grants from the Australian National Health and Medical Research Council (NHMRC) (241944, 339462, 389927, 389875, 389891, 389892, 389938, 442915, 442981, 496739, 552485, 552498 and 1075175); by Grants from the Australian Research Council (ARC) (A7960034, A79906588, A79801419, DP0770096, DP0212016, and DP0343921); DRN (FT0991022, 613674) SEM (1103623) and GWM (619667) were supported by the ARC Future Fellowship and NHMRC Fellowship Schemes.
RSI / RSII / RSIII	Rotterdam Study I, 2 and 3	The Rotterdam Study (PMID: 32367290) is funded by Erasmus Medical Center and Erasmus University, Rotterdam, Netherlands Organization for the Health Research and Development (ZonMw), the Research Institute for Diseases in the Elderly (RIDE), the Ministry of Education, Culture and Science, the Ministry for Health, Welfare and Sports, the European Commission (DG XII), and the Municipality of Rotterdam. The authors are grateful to the study participants, the staff from the Rotterdam Study and the participating general practitioners and pharmacists. The generation and management of GWAS genotype data for the Rotterdam Study (RS I, RS II, RS III) was executed by the Human Genotyping Facility of the Genetic Laboratory of the Department of Internal Medicine, Erasmus MC, Rotterdam, The Netherlands. The GWAS datasets are supported by the Netherlands Organisation of Scientific Research NWO Investments (nr. 175.010.2005.011, 911-03-012), the Genetic Laboratory of the Department of Internal Medicine, Erasmus MC, the Research Institute for Diseases in the Elderly (014-93-015; RIDE2), the Netherlands Genomics Initiative (NGI)/Netherlands Organisation for Scientific Research (NWO) Netherlands Consortium for Healthy Aging (NCHA), project nr. 050-060-810. We thank Pascal Arp, Mila Jhamai, Marijn Verkerk, Lizbeth Herrera and Marjolein Peters, MSc, and Carolina Medina-Gomez, MSc, for their help in creating the GWAS database, and Karol Estrada, PhD, Yurii Aulchenko, PhD, and Carolina Medina-Gomez, MSc, for the creation and analysis of imputed data. The Rotterdam Study has been approved by the Medical Ethics Committee of the Erasmus MC (registration number MEC 02.1015) and by the Dutch Ministry of Health, Welfare and Sport (Population Screening Act WBO, license number 1071272-159521-PG). The Rotterdam Study Personal Registration Data collection is filed with the Erasmus MC Data Protection Officer under registration number EMC1712001. The Rotterdam Study has been entered into the Netherlands National Trial Register (NTR; www.trialregister.nl) and into the WHO International Clinical Trials Registry Platform (ICTRP; www.who.int/ictrp/network/primary/en/) under shared catalogue number NTR6831. All participants provided written informed consent to participate in the study and to have their information obtained from treating physicians.
SARDINIA	SardiNIA	We thank all the volunteers who generously participated in this study and made this research possible. This research was supported by the Intramural Research Program of the NIH, National Institute on Aging, with contracts N01-AG-1-2109 and HHSN271201100005C; by PB05 InterOmics MIUR Flagship Project and by grant FaReBio2011 “Farmaci e Reti Biotecnologiche di Qualità".
SASBAC cases / controls		This work was supported by grants from NIH (RO1-CA58427) and the Agency for Science, Technology and Research (A *STAR; Singapore).

SHIP	Study of Health in Pomerania	SHIP is part of the Community Medicine Research net of the University of Greifswald, Germany, which is funded by the Federal Ministry of Education and Research (grants no. 01ZZ9603, 01ZZ0103, and 01ZZ0403), the Ministry of Cultural Affairs as well as the Social Ministry of the Federal State of Mecklenburg-West Pomerania. Genome-wide data have been supported by the Federal Ministry of Education and Research (grant no. 03ZIK012) and a joint grant from Siemens Healthcare, Erlangen, Germany and the Federal State of Mecklenburg- West Pomerania. The University of Greifswald is a member of the Caché Campus program of the InterSystems GmbH.
SHIP-Trend	Study of Health in Pomerania - Trend	SHIP is part of the Community Medicine Research net of the University of Greifswald, Germany, which is funded by the Federal Ministry of Education and Research (grants no. 01ZZ9603, 01ZZ0103, and 01ZZ0403), the Ministry of Cultural Affairs as well as the Social Ministry of the Federal State of Mecklenburg-West Pomerania. The University of Greifswald is a member of the Caché Campus program of the InterSystems GmbH.
TWINGENE	TwinGene	TwinGene is part of the Swedish Twin Registry which is managed by Karolinska Institutet and receives funding through the Swedish Research Council under the grant no 2017-00641. The Ministry for Higher Education; The Swedish Research Council (M-2005-1112); GenomEUtwin (EU/QLRT-2001-01254; QLG2-CT-2002-01254); NIH DK U01-066134; The Swedish Foundation for Strategic Research (SSF); Heart and Lung foundation no. 20070481
TWINSUK	TwinsUK	TwinsUK is funded by the Wellcome Trust, Medical Research Council, European Union, Chronic Disease Research Foundation (CDRF), Zoe Global Ltd and the National Institute for Health Research (NIHR)-funded BioResource, Clinical Research Facility and Biomedical Research Centre based at Guy's and St Thomas' NHS Foundation Trust in partnership with King's College London.
UK Biobank	UK Biobank	This research has been conducted using the UK Biobank resource under application numbers 871 (Exeter) and 9797 (Cambridge).
VB	Val Borbera	We thank all the participants to the project, the San Raffaele Hospital MDs who contributed to clinical data collection, prof. Clara Camaschella who coordinated the data collection, Corrado Masciullo and Massimiliano Cocca for the database informatics. The research was supported by funds from Compagnia di San Paolo, Torino, Italy; Fondazione Cariplo, Italy; Telethon Italy; Ministry of Health, Ricerca Finalizzata 2008 and 2011-2012 and Public Health Genomics Project 2010.
WGHS	Women's Genome Health Study	The WGHS is supported by the National Heart, Lung, and Blood Institute (HL043851 and HL080467) and the National Cancer Institute (CA047988 and UM1CA182913), with funding for genotyping provided by Amgen. NHLBI/NCI (Buring, Lee, PIs); Amgen (Chasman, Ridker, PIs).
WHI GARNET / WHIMS	Women's Health Initiative	The WHI program is funded by the National Heart, Lung, and Blood Institute, National Institutes of Health, U.S. Department of Health and Human Services through contracts HHSN268201600018C, HHSN268201600001C, HHSN268201600002C, HHSN268201600003C, and HHSN268201600004C. The authors thank the WHI investigators and staff for their dedication, and the study participants for making the program possible. A full listing of WHI investigators can be found at: https://s3-us-west-2.amazonaws.com/www-whi-org/wp-content/uploads/WHI-Investigator-Long-List.pdf
China Kadoorie Biobank	China Kadoorie Biobank	The chief acknowledgment is to the participants, the project staff, and the China National Centre for Disease Control and Prevention (CDC) and its regional offices for assisting with the fieldwork. China's National Health Insurance provides electronic linkage to all hospital treatments. We thank Judith Mackay in Hong Kong; Yu Wang, Gonghuan Yang, Zhengfu Qiang, Lin Feng, Maigeng Zhou, Wenhua Zhao, Yan Zhang and Zheng Bian in China CDC; Lingzhi Kong, Xiucheng Yu, and Kun Li in the Chinese Ministry of Health; and Garry Lancaster, Sarah Clark, Martin Radley, Mike Hill, Hongchao Pan, and Jill Boreham in the CTSU, Oxford, for assisting with the design,

		planning, organisation, and conduct of the study. The CKB baseline survey and the first re-survey were supported by the Kadoorie Charitable Foundation, Hong Kong. Long-term follow-up was supported by UK Wellcome Trust(212946/Z/18/Z, 202922/Z/16/Z, 104085/Z/14/Z, 088158/Z/09/Z), National Key Research and Development Program of China (2016YFC0900500, 2016YFC0900501, 2016YFC0900504, 2016YFC1303904) and National Natural Science Foundation of China (91843302). DNA extraction and genotyping was funded by GlaxoSmithKline, and the UK Medical Research Council (MC-PC-13049, MC-PC-14135). The project is supported by core funding from UK Medical Research Council (MC_UU_00017/1,MC_UU_12026/2 MC_U137686851), Cancer Research UK (C16077/A29186; C500/A16896), and British Heart Foundation (CH/1996001/9454) to the Clinical Trial Service Unit and Epidemiological Studies Unit at Oxford University.

2) Acknowledgments and sources of funding for studies included in Breast Cancer Association Consortium

Study acronym	Full study name	Acknowledgments and sources of funding
BCAC	Breast Cancer Association Consortium	We thank all the individuals who took part in these studies and all the researchers, clinicians, technicians and administrative staff who have enabled this work to be carried out. BCAC is funded by the European Union's Horizon 2020 Research and Innovation Programme (grant numbers 634935 and 633784 for BRIDGES and B-CAST respectively), and PERSPECTIVE I&I, funded by the Government of Canada through Genome Canada and the Canadian Institutes of Health Research, the Ministère de l'Économie et de l'Innovation du Québec through Genome Québec, the Quebec Breast Cancer Foundation. The EU Horizon 2020 Research and Innovation Programme funding source had no role in study design, data collection, data analysis, data interpretation or writing of the report. Additional funding for BCAC is provided via the Confluence project which is funded with intramural funds from the National Cancer Institute Intramural Research Program, National Institutes of Health. The breast cancer genome-wide association analyses were supported by the Government of Canada through Genome Canada and the Canadian Institutes of Health Research, the 'Ministère de l'Économie, de la Science et de l'Innovation du Québec' through Genome Québec and grant PSR-SIIRI-701, The National Institutes of Health (U19 CA148065, X01HG007492), Cancer Research UK (C1287/A10118, C1287/A16563, C1287/A10710) and the European Union (HEALTH-F2-2009-223175 and H2020 633784 and 634935). All studies and funders are listed in Michailidou et al (2017).
ABCFS	Australian Breast Cancer Family Study	ABCFS thank Maggie Angelakos, Judi Maskiell, Gillian Dite. The Australian Breast Cancer Family Study (ABCFS) was supported by grant UM1 CA164920 from the National Cancer Institute (USA). The content of this manuscript does not necessarily reflect the views or policies of the National Cancer Institute or any of the collaborating centers in the Breast Cancer Family Registry (BCFR), nor does mention of trade names, commercial products, or organizations imply endorsement by the USA Government or the BCFR. The ABCFS was also supported by the National Health and Medical Research Council of Australia, the New South Wales Cancer Council, the Victorian Health Promotion Foundation (Australia) and the Victorian Breast Cancer Research Consortium. J.L.H. is a National Health and Medical Research Council (NHMRC) Senior Principal Research Fellow. M.C.S. is a NHMRC Senior Research Fellow.
ABCS		ABCS thanks the Blood bank Sanquin, The Netherlands. The ABCS study was supported by the Dutch Cancer Society [grants NKI 2007-3839; 2009 4363].

BBCC		The work of the BBCC was partly funded by ELAN-Fond of the University Hospital of Erlangen.
BCINIS		The BCINIS study would not have been possible without the contributions of Dr. K. Landsman, Dr. N. Gronich, Dr. A. Flugelman, Dr. W. Saliba, Dr. E. Liani, Dr. I. Cohen, Dr. S. Kalet, Dr. V. Friedman, Dr. O. Barnet of the NICCC in Haifa, and all the contributing family medicine, surgery, pathology and oncology teams in all medical institutes in Northern Israel. BIGGS thanks Niall McInerney, Gabrielle Colleran, Andrew Rowan, Angela Jones.
BREOGAN	BREast Oncology GAlician Network	The BREOGAN study would not have been possible without the contributions of the following: Manuela Gago-Dominguez, Jose Esteban Castelao, Angel Carracedo, Victor Muñoz Garzón, Alejandro Novo Domínguez, Maria Elena Martinez, Sara Miranda Ponte, Carmen Redondo Marey, Maite Peña Fernández, Manuel Enguix Castelo, Maria Torres, Manuel Calaza (BREOGAN), José Antúnez, Máximo Fraga and the staff of the Department of Pathology and Biobank of the University Hospital Complex of Santiago-CHUS, Instituto de Investigación Sanitaria de Santiago, IDIS, Xerencia de Xestion Integrada de Santiago-SERGAS; Joaquín González-Carreró and the staff of the Department of Pathology and Biobank of University Hospital Complex of Vigo, Instituto de Investigacion Biomedica Galicia Sur, SERGAS, Vigo, Spain. The BREast Oncology GAlician Network (BREOGAN) is funded by Acción Estratégica de Salud del Instituto de Salud Carlos III FIS PI12/02125/Cofinanciado FEDER; Acción Estratégica de Salud del Instituto de Salud Carlos III FIS Intrasalud (PI13/01136); Programa Grupos Emergentes, Cancer Genetics Unit, Instituto de Investigacion Biomedica Galicia Sur. Xerencia de Xestion Integrada de Vigo-SERGAS, Instituto de Salud Carlos III, Spain; Grant 10CSA012E, Consellería de Industria Programa Sectorial de Investigación Aplicada, PEME I + D e I + D Suma del Plan Gallego de Investigación, Desarrollo e Innovación Tecnológica de la Consellería de Industria de la Xunta de Galicia, Spain; Grant EC11-192. Fomento de la Investigación Clínica Independiente, Ministerio de Sanidad, Servicios Sociales e Igualdad, Spain; and Grant FEDER-Innterconecta. Ministerio de Economia y Competitividad, Xunta de Galicia, Spain.
CBCS	Canadian Breast Cancer Study	CBCS thanks study participants, co-investigators, collaborators and staff of the Canadian Breast Cancer Study, and project coordinators Agnes Lai and Celine Morissette. CBCS is funded by the Canadian Cancer Society (grant # 313404) and the Canadian Institutes of Health Research.
CCGP		CCGP thanks Styliani Apostolaki, Anna Margiolaki, Georgios Nintos, Maria Perraki, Georgia Saloustrou, Georgia Sevastaki, Konstantinos Pompodakis. CCGP is supported by funding from the University of Crete.
CECILE		The CECILE study was supported by Fondation de France, Institut National du Cancer (INCa), Ligue Nationale contre le Cancer, Agence Nationale de Sécurité Sanitaire, de l'Alimentation, de l'Environnement et du Travail (ANSES), Agence Nationale de la Recherche (ANR).
CGPS		CGPS thanks staff and participants of the Copenhagen General Population Study. For the excellent technical assistance: Dorthe Uldall Andersen, Maria Birna Arnadottir, Anne Bank, Dorthe Kjeldgård Hansen. The Danish Cancer Biobank is acknowledged for providing infrastructure for the collection of blood samples for the cases. The CGPS was supported by the Chief Physician Johan Boserup and Lise Boserup Fund, the Danish Medical Research Council, and Herlev and Gentofte Hospital.
CPSII		Investigators from the CPS-II cohort thank the participants and Study Management Group for their invaluable contributions to this research. They also acknowledge the contribution to this study from central cancer registries supported through the Centers for Disease Control and Prevention National Program of Cancer Registries, as well as cancer registries supported by the National Cancer Institute Surveillance Epidemiology and End Results program.

		The American Cancer Society funds the creation, maintenance, and updating of the CPS-II cohort.
DIETCOMPLYF		The DietCompLyf study was funded by the charity Against Breast Cancer (Registered Charity Number 1121258) and the NCRN. The University of Westminster curates the DietCompLyf database funded by Against Breast Cancer Registered Charity No. 1121258 and the NCRN.
ICICLE		ICICLE thanks Kelly Kohut, Michele Caneppele, Maria Troy. ICICLE was supported by Breast Cancer Now, CRUK and Biomedical Research Centre at Guy's and St Thomas' NHS Foundation Trust and King's College London.
KARBAC		Financial support for KARBAC was provided through the regional agreement on medical training and clinical research (ALF) between Stockholm County Council and Karolinska Institutet, the Swedish Cancer Society, The Gustav V Jubilee foundation and Bert von Kantzows foundation.
KARMA		KARMA and SASBAC thank the Swedish Medical Research Counsel. The KARMA study was supported by Märit and Hans Rausings Initiative Against Breast Cancer.
KBCP		KBCP thanks Eija Myöhänen, Helena Kemiläinen. The KBCP was financially supported by the special Government Funding (EVO) of Kuopio University Hospital grants, Cancer Fund of North Savo, the Finnish Cancer Organizations, and by the strategic funding of the University of Eastern Finland.
KCONFAB/ AOCS		kConFab/AOCS wish to thank Heather Thorne, Eveline Niedermayr, all the kConFab research nurses and staff, the heads and staff of the Family Cancer Clinics, and the Clinical Follow Up Study (which has received funding from the NHMRC, the National Breast Cancer Foundation, Cancer Australia, and the National Institute of Health (USA)) for their contributions to this resource, and the many families who contribute to kConFab. kConFab is supported by a grant from the National Breast Cancer Foundation, and previously by the National Health and Medical Research Council (NHMRC), the Queensland Cancer Fund, the Cancer Councils of New South Wales, Victoria, Tasmania and South Australia, and the Cancer Foundation of Western Australia. Financial support for the AOCS was provided by the United States Army Medical Research and Materiel Command [DAMD17-01-1-0729], Cancer Council Victoria, Queensland Cancer Fund, Cancer Council New South Wales, Cancer Council South Australia, The Cancer Foundation of Western Australia, Cancer Council Tasmania and the National Health and Medical Research Council of Australia (NHMRC; 400413, 400281, 199600). G.C.T. and P.W. are supported by the NHMRC. RB was a Cancer Institute NSW Clinical Research Fellow.
MARIE		MARIE thanks Petra Seibold, Dieter Flesch-Janys, Judith Heinz, Nadia Obi, Alina Vrieling, Sabine Behrens, Ursula Eilber, Muhabbet Celik, Til Olchers and Stefan Nickels. The MARIE study was supported by the Deutsche Krebshilfe e.V. [70-2892-BR I, 106332, 108253, 108419, 110826, 110828], the Hamburg Cancer Society, the German Cancer Research Center (DKFZ) and the Federal Ministry of Education and Research (BMBF) Germany [01KH0402].
MCBCS		The MCBCS was supported by the NIH grants CA192393, CA116167, CA176785 an NIH Specialized Program of Research Excellence (SPORE) in Breast Cancer [CA116201], and the Breast Cancer Research Foundation and a generous gift from the David F. and Margaret T. Grohne Family Foundation.
MCCS	Melbourne Collaborative Cohort Study	The MCCS was made possible by the contribution of many people, including the original investigators, the teams that recruited the participants and continue working on follow-up, and the many thousands of Melbourne residents who continue to participate in the study. The Melbourne Collaborative Cohort Study (MCCS) cohort recruitment was funded by VicHealth and Cancer Council Victoria. The MCCS was further augmented by Australian National Health and Medical Research Council grants 209057, 396414 and

		1074383 and by infrastructure provided by Cancer Council Victoria. Cases and their vital status were ascertained through the Victorian Cancer Registry and the Australian Institute of Health and Welfare, including the National Death Index and the Australian Cancer Database.
MEC		The MEC was supported by NIH grants CA63464, CA54281, CA098758, CA132839 and CA164973.
MISS		The MISS study is supported by funding from ERC-2011-294576 Advanced grant, Swedish Cancer Society, Swedish Research Council, Local hospital funds, Berta Kamprad Foundation, Gunnar Nilsson.
MMHS		We thank the coordinators, the research staff and especially the MMHS participants for their continued collaboration on research studies in breast cancer. The MMHS study was supported by NIH grants CA97396, CA128931, CA116201, CA140286 and CA177150. MSKCC is supported by grants from the Breast Cancer Research Foundation and Robert and Kate Niehaus Clinical Cancer Genetics Initiative.
NC-BCFR & OFBCR	Northern California Breast Cancer Family Registry (NC-BCFR) and Ontario Familial Breast Cancer Registry (OFBCR)	The Northern California Breast Cancer Family Registry (NC-BCFR) and Ontario Familial Breast Cancer Registry (OFBCR) were supported by grants U01CA164920 and U01CA167551 from the USA National Cancer Institute of the National Institutes of Health. The content of this manuscript does not necessarily reflect the views or policies of the National Cancer Institute or any of the collaborating centers in the Breast Cancer Family Registry (BCFR) or the Colon Cancer Family Registry (CCFR), nor does mention of trade names, commercial products, or organizations imply endorsement by the USA Government or the BCFR or CCFR.
NCBCS	Carolina Breast Cancer Study (NCBCS)	The Carolina Breast Cancer Study (NCBCS) was funded by Komen Foundation, the National Cancer Institute (P50 CA058223, U54 CA156733, U01 CA179715), and the North Carolina University Cancer Research Fund.
OFBCR		The OFBCR thanks Teresa Selander, Nayana Weerasooriya and Steve Gallinger. ORIGO thanks E. Krol-Warmerdam, and J. Blom for patient accrual, administering questionnaires, and managing clinical information.
PBCS		PBCS thanks Louise Brinton, Mark Sherman, Neonila Szeszenia-Dabrowska, Beata Peplonska, Witold Zatonski, Pei Chao, Michael Stagner. The PBCS was funded by Intramural Research Funds of the National Cancer Institute, Department of Health and Human Services, USA.
PKARMA		
PLCO		Genotyping for PLCO was supported by the Intramural Research Program of the National Institutes of Health, NCI, Division of Cancer Epidemiology and Genetics. The PLCO is supported by the Intramural Research Program of the Division of Cancer Epidemiology and Genetics and supported by contracts from the Division of Cancer Prevention, National Cancer Institute, National Institutes of Health.
SASBAC		The SASBAC study was supported by funding from the Agency for Science, Technology and Research of Singapore (A*STAR), the US National Institute of Health (NIH) and the Susan G. Komen Breast Cancer Foundation.
SBCS		SBCS thanks Sue Higham, Helen Cramp, Dan Connley, Ian Brock, Sabapathy Balasubramanian and Malcolm W.R. Reed. The SBCS was supported by Sheffield Experimental Cancer Medicine Centre and Breast Cancer Now Tissue Bank.
SEARCH		We thank the SEARCH and EPIC teams. SEARCH is funded by Cancer Research UK [C490/A10124, C490/A16561] and supported by the UK National Institute for Health Research Biomedical Research Centre at the University of Cambridge. The University of Cambridge has received salary support for PDPP from the NHS in the East of England through the Clinical Academic Reserve.
SISTER	The Sister Study	The Sister Study (SISTER) is supported by the Intramural Research Program of the NIH, National Institute of Environmental Health Sciences (Z01-ES044005 and Z01-

		ES049033).
SMC		The SMC is funded by the Swedish Cancer Foundation and the Swedish Research Council (VR 2017-00644) grant for the Swedish Infrastructure for Medical Population-based Life-course Environmental Research (SIMPLER).
UKBGS		UKBGS thanks Breast Cancer Now and the Institute of Cancer Research for support and funding of the Breakthrough Generations Study, and the study participants, study staff, and the doctors, nurses and other health care providers and health information sources who have contributed to the study. We acknowledge NHS funding to the Royal Marsden/ICR NIHR Biomedical Research Centre. The UKBGS is funded by Breast Cancer Now and the Institute of Cancer Research (ICR), London. ICR acknowledges NHS funding to the NIHR Biomedical Research Centre.
USRT		The USRT Study was funded by Intramural Research Funds of the National Cancer Institute, Department of Health and Human Services, USA.

3) Further acknowledgments and sources of funding

Genome-wide analyses

Exeter:

Robin N. Beaumont is funded by the Wellcome Trust and Royal Society grant: 104150/Z/14/Z. Anna Murray is supported by the Wellcome Trust Institutional Strategic Support Award (WT097835MF). Katherine S. Ruth was supported by funding from the Gillings Family Foundation, MRC P2D and is supported by Cancer Research UK [grant number C18281/A29019]. Jessica Tyrrell is supported by an Academy of Medical Sciences (AMS) Springboard award, which is supported by the AMS, the Wellcome Trust, GCRF, the Government Department of Business, Energy and Industrial strategy, the British Heart Foundation and Diabetes UK [SBF004\1079]. Andrew R. Wood is supported by the European Research Council grants: SZ-245 50371-GLUCOSEGENES-FP7-IDEAS-ERC and 323195. The authors acknowledge the use of the University of Exeter High-Performance Computing facility in carrying out this work.

Edinburgh:

Caroline Hayward, Veronique Vitart and James F. Wilson are supported by an MRC University Unit Programme Grant MC_UU_00007/10 (QTL in Health and Disease).

Cambridge:

Claudia Langenberg and Nicholas J. Wareham are funded by the Medical Research Council (MC_UU_12015/1 and MC_UU_00006/1). Nicholas J Wareham is a NIHR Senior Investigator.

Ken Ong, John Perry, Stasa Stankovic and Felix Day are supported by the Medical Research Council (Unit programmes: MC_UU_12015/2 and MC_UU_00006/2).

Mouse models of environmentally-induced low ovarian reserve:

Catherine Aiken and Susan Ozanne are funded by grants from the Academy of Medical Sciences, the Addenbrooke's Charitable Trust, an Isaac Newton Trust/Wellcome Trust ISSF/ University of Cambridge Joint Research Grant and the MRC (MRC_MC_UU_12012/4 and MRC_MC_UU_00014/4) and supported by the NIHR Cambridge Biomedical Research Centre. The views expressed are those of the author(s) and not necessarily those of the NIHR or the Department of Health and Social Care.

Mouse models of DNA damage response and ovarian ageing:

Jazib Hussain is the recipient of an international PhD fellowship from the Punjab Educational Endowment Fund, Pakistan. Members of the Eva R Hoffmann were funded by an ERC Consolidator Grant (724718-ReCAP), a Novo Nordisk Foundation Young Investigator Award (NNF15OC0016662), and IRFD-FSS (0134-00299B). They acknowledge the support with histology from Pernille Sjølin Froh and Thi Cam Ha Nguyen at their Department. Andres Lopez-Contreras was funded by an ERC Starter grant (ERC-2015-STG-679068) and Nordea Fonden (02-2017-1749). Paula Aguilera was supported by a Juan de la Cierva postdoctoral contract from the Spanish Ministry of Science and Innovation. Eva R Hoffmann and Andres Lopez-Contreras acknowledge core funding from the Danish National Research Foundation Center Grant (6110-00344B). Jeremy A. Daniel was supported by the Novo Nordisk Foundation (grant agreement NFF14CC0001). Satoshi H. Namekawa was funded by the National Institutes of Health (R01 GM098605). Petr Solc and Lucie Knoblochova were funded by The Institutional Research Concept RVO67985904 and grant #20-27742S from the Grant Agency of the

Czech Republic. Lucie Knoblochova was supported by the Charles University, project GA UK No. 1402217. Claus Yding Andersen, Marie Louise Grøndahl, Kristina W. Olsen, Eva R. Hoffmann acknowledge funding from Ferring Pharmaceuticals and ReproUnion (www.reprounion.eu). Kristina W. Olsen was a recipient of a fellowship from Helsefonden (20-B-0334). Yan Huang is a recipient of a fellowship of the China Scholarship Council (201607040048). Ignasi Roig was supported by the Spanish Ministerio de Ciencia e Innovación (BFU2013-43965-P, BFU2016-80370-P and PID2019-107082RB-I00) and a fellowship of the Fundación Dexeus Mujer.

4) Individual study disclosures

Joanne M. Murabito discloses a consulting relationship with Merck as a guest lecturer.

Deborah J. Thompson is now employed by Genomics plc.

Pierre Fontanillas is employed by and hold stock or stock options in 23andMe, Inc.

All other authors and studies declared no conflict of interest.

5) Consortium membership

Biobank-based Integrative Omics Study (BIOS) Consortium

Management Team Bastiaan T. Heijmans (chair)[1], Peter A.C. 't Hoen[2], Joyce van Meurs[3], Aaron Isaacs[4], Rick Jansen[5], Lude Franke[6].
Cohort collection Dorret I. Boomsma[7], René Pool[7], Jenny van Dongen[7], Jouke J. Hottenga[7] (Netherlands Twin Register); Marleen MJ van Greevenbroek[8], Coen D.A. Stehouwer[8], Carla J.H. van der Kallen[8], Casper G. Schalkwijk[8] (Cohort study on Diabetes and Atherosclerosis Maastricht); Cisca Wijmenga[6], Lude Franke[6], Sasha Zhernakova[6], Ettje F. Tigchelaar[6] (LifeLines Deep); P. Eline Slagboom[1], Marian Beekman[1], Joris Deelen[1], Diana van Heemst[9] (Leiden Longevity Study); Jan H. Veldink[10], Leonard H. van den Berg[10] (Prospective ALS Study Netherlands); Cornelia M. van Duijn[4], Bert A. Hofman[11], Aaron Isaacs[4], André G. Uitterlinden[3] (Rotterdam Study).

Data Generation Joyce van Meurs (Chair)[3], P. Mila Jhamai[3], Michael Verbiest[3], H. Eka D. Suchiman[1], Marijn Verkerk[3], Ruud van der Breggen[1], Jeroen van Rooij[3], Nico Lakenberg[1].
Data management and computational infrastructure Hailiang Mei (Chair)[12], Maarten van Iterson[1], Michiel van Galen[2], Jan Bot[13], Dasha V. Zhernakova[6], Rick Jansen[5], Peter van 't Hof[12], Patrick Deelen[6], Irene Nooren[13], Peter A.C. 't Hoen[2], Bastiaan T. Heijmans[1], Matthijs Moed[1].

Data Analysis Group Lude Franke (Co-Chair)[6], Martijn Vermaat[2], Dasha V. Zhernakova[6], René Luijk[1], Marc Jan Bonder[6], Maarten van Iterson[1], Patrick Deelen[6], Freerk van Dijk[14], Michiel van Galen[2], Wibowo Arindrarto[12], Szymon M. Kielbasa[15], Morris A. Swertz[14], Erik. W van Zwet[15], Rick Jansen[5], Peter-Bram 't Hoen (Co-Chair)[2], Bastiaan T. Heijmans (Co-Chair)[1].

1. Molecular Epidemiology Section, Department of Medical Statistics and Bioinformatics, Leiden University Medical Center, Leiden, The Netherlands
2. Department of Human Genetics, Leiden University Medical Center, Leiden, The Netherlands
3. Department of Internal Medicine, ErasmusMC, Rotterdam, The Netherlands
4. Department of Genetic Epidemiology, ErasmusMC, Rotterdam, The Netherlands
5. Department of Psychiatry, VU University Medical Center, Neuroscience Campus Amsterdam, Amsterdam, The Netherlands
6. Department of Genetics, University of Groningen, University Medical Centre Groningen, Groningen, The Netherlands
7. Department of Biological Psychology, VU University Amsterdam, Neuroscience Campus Amsterdam, Amsterdam, The Netherlands
8. Department of Internal Medicine and School for Cardiovascular Diseases (CARIM), Maastricht University Medical Center, Maastricht, The Netherlands
9. Department of Gerontology and Geriatrics, Leiden University Medical Center, Leiden, The Netherlands
10. Department of Neurology, Brain Center Rudolf Magnus, University Medical Center Utrecht, Utrecht, The Netherlands
11. Department of Epidemiology, ErasmusMC, Rotterdam, The Netherlands
12. Sequence Analysis Support Core, Leiden University Medical Center, Leiden, The Netherlands
13. SURFsara, Amsterdam, the Netherlands
14. Genomics Coordination Center, University Medical Center Groningen, University of Groningen, Groningen, the Netherlands
15. Medical Statistics Section, Department of Medical Statistics and Bioinformatics, Leiden University Medical Center, Leiden, The Netherlands

eQTLGen Consortium

Mawussé Agbessi[1], Habibul Ahsan[2], Isabel Alves[1], Anand Andiappan[3], Wibowo Arindrarto[4], Philip Awadalla[1], Alexis Battle[5,6], Frank Beutner[7], Marc Jan Bonder[8,9,10], Dorret Boomsma[11], Mark Christiansen[12], Annique Claringbould[8], Patrick Deelen[8,13], Tõnu Esko[14], Marie-Julie Favé[1], Lude Franke[8], Timothy Frayling[15], Sina A. Gharib[16,12], Gregory Gibson[17], Bastiaan T. Heijmans[4], Gibran Hemani[18], Rick Jansen[19], Mika Kähönen[20,21], Anette Kalnapenkis[14], Silva Kasela[14], Johannes Kettunen[22], Yungil Kim[6,23], Holger Kirsten[24], Peter Kovacs[25], Knut Krohn[26], Jaanika Kronberg-Guzman[14], Viktorija Kukushkina[14], Zoltan Kutalik[27], Bernett Lee[3], Terho Lehtimäki[28,29], Markus Loeffler[24], Urko M. Marigorta[17], Hailang Mei[4], Lili Milani[14], Grant W. Montgomery[30], Martina Müller-Nurasyid[31,32,33], Matthias Nauck[34], Michel Nivard[11], Brenda Penninx[19], Markus Perola[35], Natalia Pervjakova[14], Brandon L. Pierce[2], Joseph Powell[36], Holger Prokisch[37,38], Bruce M. Psaty[12,39,40], Olli T. Raitakari[41,42], Samuli Ripatti[43], Olaf Rotzschke[3], Sina Rüeger[27], Ashis Saha[6], Markus Scholz[24], Katharina Schramm[31,32], Ilkka Seppälä[28,29], Eline P. Slagboom[4], Coen D.A. Stehouwer[44], Michael Stumvoll[45], Patrick Sullivan[46], Peter-Bram 't Hoen[47], Alexander Teumer[48], Joachim Thiery[49], Lin Tong[2], Anke Tönjes[45], Jenny van Dongen[11], Maarten van Iterson[4], Joyce van Meurs[50], Jan H. Veldink[51], Joost Verlouw[50], Peter M. Visscher[30], Uwe Völker[52], Urmo Võsa[8,14], Harm-Jan Westra[8], Cisca Wijmenga[8], Hanieh Yaghootkar[15], Jian Yang[30,53], Biao Zeng[17], Futao Zhang[30]

Author list is ordered alphabetically.

1. Computational Biology, Ontario Institute for Cancer Research, Toronto, Canada
2. Department of Public Health Sciences, University of Chicago, Chicago, United States of America
3. Singapore Immunology Network, Agency for Science, Technology and Research, Singapore, Singapore
4. Department of Biomedical Data Sciences, Leiden University Medical Center, Leiden, The Netherlands
5. Department of Biomedical Engineering, Johns Hopkins University, Baltimore, United States of America
6. Department of Computer Science, Johns Hopkins University, Baltimore, United States of America
7. Heart Center Leipzig, Universität Leipzig, Leipzig, Germany
8. Department of Genetics, University Medical Centre Groningen, Groningen, The Netherlands
9. European Molecular Biology Laboratory, European Bioinformatics Institute, Wellcome Genome Campus, Hinxton, United Kingdom
10. Genome Biology Unit, European Molecular Biology Laboratory, Heidelberg, Germany
11. Department of Biological Psychology, Vrije Universiteit Amsterdam, Amsterdam, The Netherlands
12. Cardiovascular Health Research Unit, University of Washington, Seattle, United States of America
13. Genomics Coordination Center, University Medical Centre Groningen, Groningen, The Netherlands
14. Estonian Genome Center, Institute of Genomics, University of Tartu, Tartu 51010, Estonia
15. Exeter Medical School, University of Exeter, Exeter, United Kingdom
16. Department of Medicine, University of Washington, Seattle, United States of America
17. School of Biological Sciences, Georgia Tech, Atlanta, United States of America
18. MRC Integrative Epidemiology Unit, University of Bristol, Bristol, United Kingdom
19. Department of Psychiatry and Amsterdam Neuroscience, Amsterdam UMC, Vrije Universiteit, Amsterdam, the Netherlands
20. Department of Clinical Physiology, Tampere University, Tampere, Finland
21. Department of Clinical Physiology, Finnish Cardiovascular Research Center – Tampere, Faculty of Medicine and Life Sciences, University of Tampere, Tampere, Finland
22. Centre for Life Course Health Research, University of Oulu, Oulu, Finland
23. Genetics and Genomic Science Department, Icahn School of Medicine at Mount Sinai, New York, United States of America

24. Institut für Medizinische Informatik, Statistik und Epidemiologie, LIFE – Leipzig Research Center for Civilization Diseases, Universität Leipzig, Leipzig, Germany
25. IFB Adiposity Diseases, Universität Leipzig, Leipzig, Germany
26. Interdisciplinary Center for Clinical Research, Faculty of Medicine, Universität Leipzig, Leipzig, Germany
27. Institute of Social and Preventive Medicine, Lausanne University Hospital, Lausanne, Switzerland
28. Department of Clinical Chemistry, Fimlab Laboratories, Tampere, Finland
29. Department of Clinical Chemistry, Finnish Cardiovascular Research Center - Tampere, Faculty of Medicine and Life Sciences, University of Tampere, Tampere, Finland
30. Institute for Molecular Bioscience, University of Queensland, Brisbane, Australia
31. Institute of Genetic Epidemiology, Helmholtz Zentrum München - German Research Center for Environmental Health, Neuherberg, Germany
32. Department of Medicine I, University Hospital Munich, Ludwig Maximilian's University, München, Germany
33. DZHK (German Centre for Cardiovascular Research), partner site Munich Heart Alliance, Munich, Germany
34. Institute of Clinical Chemistry and Laboratory Medicine, University Medicine Greifswald, Greifswald, Germany
35. National Institute for Health and Welfare, University of Helsinki, Helsinki, Finland
36. Garvan Institute of Medical Research, Garvan-Weizmann Centre for Cellular Genomics, Sydney, Australia
37. Institute of Human Genetics, Helmholtz Zentrum München, Neuherberg, Germany
38. Institute of Human Genetics, Technical University Munich, Munich, Germany
39. Departments of Epidemiology, Medicine, and Health Services, University of Washington, Seattle, United States of America
40. Kaiser Permanente Washington Health Research Institute, Seattle, WA, United States of America
41. Department of Clinical Physiology and Nuclear Medicine, Turku University Hospital, Turku, Finland
42. Research Centre of Applied and Preventive Cardiovascular Medicine, University of Turku, Turku, Finland
43. Statistical and Translational Genetics, University of Helsinki, Helsinki, Finland
44. Department of Internal Medicine, Maastricht University Medical Centre, Maastricht, The Netherlands
45. Department of Medicine, Universität Leipzig, Leipzig, Germany
46. Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden
47. Center for Molecular and Biomolecular Informatics, Radboud Institute for Molecular Life Sciences, Radboud University Medical Center Nijmegen, Nijmegen, The Netherlands
48. Institute for Community Medicine, University Medicine Greifswald, Greifswald, Germany
49. Institute for Laboratory Medicine, LIFE – Leipzig Research Center for Civilization Diseases, Universität Leipzig, Leipzig, Germany
50. Department of Internal Medicine, Erasmus Medical Centre, Rotterdam, The Netherlands
51. Department of Neurology, University Medical Center Utrecht, Utrecht, The Netherlands
52. Interfaculty Institute for Genetics and Functional Genomics, University Medicine Greifswald, Greifswald, Germany
53. Institute for Advanced Research, Wenzhou Medical University, Wenzhou, Zhejiang 325027, China

The Biobank Japan Project

Koichi Matsuda. Laboratory of Genome Technology, Human Genome Center, Institute of Medical Science, The University of Tokyo, Tokyo, Japan. Laboratory of Clinical Genome Sequencing, Graduate School of Frontier Sciences, The University of Tokyo, Tokyo, Japan.

kmatsuda@edu.k.u-tokyo.ac.jp

Yuji Yamanashi. Division of Genetics, The Institute of Medical Science, The University of Tokyo, Tokyo, Japan. yyamanas@ims.u-tokyo.ac.jp

Yoichi Furukawa. Division of Clinical Genome Research, Institute of Medical Science, The University of Tokyo, Tokyo, Japan. furukawa@ims.u-tokyo.ac.jp

Takayuki Morisaki. Division of Molecular Pathology IMSUT Hospital, Department of Internal Medicine Project Division of Genomic Medicine and Disease Prevention The Institute of Medical Science The University of Tokyo, Tokyo, Japan. morisaki@ims.u-tokyo.ac.jp

Yoshinori Murakami. Department of Cancer Biology, Institute of Medical Science, The University of Tokyo, Tokyo, Japan. ymurakam@ims.u-tokyo.ac.jp

Yoichiro Kamatani. Laboratory of Complex Trait Genomics, Graduate School of Frontier Sciences, The University of Tokyo, Tokyo, Japan.

Laboratory of Clinical Genome Sequencing, Graduate School of Frontier Sciences, The University of Tokyo, Tokyo, Japan. kamatani.yoichiro@edu.k.u-tokyo.ac.jp

Kaori Muto. Department of Public Policy, Institute of Medical Science, The University of Tokyo, Tokyo, Japan. krmt@ims.u-tokyo.ac.jp

Akiko Nagai. Department of Public Policy, Institute of Medical Science, The University of Tokyo, Tokyo, Japan. akikongi@ims.u-tokyo.ac.jp

Wataru Obara. Department of Urology, Iwate Medical University, Iwate, Japan. watao@iwate-med.ac.jp

Ken Yamaji. Department of Internal Medicine and Rheumatology, Juntendo University Graduate School of Medicine, Tokyo, Japan. k.yamaji@juntendo.ac.jp

Kazuhisa Takahashi. Department of Respiratory Medicine, Juntendo University Graduate School of Medicine, Tokyo, Japan. kztakaha@juntendo.ac.jp

Satoshi Asai. Division of Pharmacology, Department of Biomedical Science, Nihon University School of Medicine, Tokyo, Japan.

Division of Genomic Epidemiology and Clinical Trials, Clinical Trials Research Center, Nihon University. School of Medicine, Tokyo, Japan. asai.satoshi@nihon-u.ac.jp

Yasuo Takahashi. Division of Genomic Epidemiology and Clinical Trials, Clinical Trials Research Center, Nihon University School of Medicine, Tokyo, Japan. takahashi.yasuo@nihon-u.ac.jp

Takao Suzuki. Tokushukai Group, Tokyo, Japan. takao.suzuki@tokushukai.jp

Nobuaki Sinozaki. Tokushukai Group, Tokyo, Japan. nobuaki.shinozaki@tokushukai.jp

Hiroki Yamaguchi. Departmentof Hematology, Nippon Medical School, Tokyo, Japan. y-hiroki@fd6.so-net.ne.jp

Shiro Minami. Department of Bioregulation, Nippon Medical School, Kawasaki, Japan. shirom@nms.ac.jp

Shigeo Murayama. Tokyo Metropolitan Geriatric Hospital and Institute of Gerontology, Tokyo, Japan. smurayam@bbarjp.net

Kozo Yoshimori. Fukujuji Hospital, Japan Anti-Tuberculosis Association, Tokyo, Japan. yoshimorik@fukujuji.org

Satoshi Nagayama. The Cancer Institute Hospital of the Japanese Foundation for Cancer Research, Tokyo, Japan. snagayama2009@hotmail.co.jp

Daisuke Obata. Center for Clinical Research and Advanced Medicine, Shiga University of Medical Science, Shiga, Japan. dobata@belle.shiga-med.ac.jp

Masahiko Higashiyama. Department of General Thoracic Surgery, Osaka International Cancer Institute, Osaka, Japan. higashiyama-m@higashiosaka-hosp.jp

Akihide Masumoto. IIZUKA HOSPITAL, Fukuoka,Japan. amasumotoh1@aih-net.com

Yukihiro Koretsune. National Hospital Organization Osaka National Hospital, Osaka, Japan. koretune318@hotmail.co.jp

China Kadoorie Biobank Collaborative Group

International Steering Committee: Junshi Chen, Zhengming Chen (PI), Robert Clarke, Rory Collins, Yu Guo, Liming Li (PI), Chen Wang, Jun Lv, Richard Peto, Robin Walters.

International Co-ordinating Centre, Oxford: Daniel Avery, Ruth Boxall, Derrick Bennett, Ka Hung Chan, Yumei Chang, Yiping Chen, Zhengming Chen, Robert Clarke, Huaidong Du, Zammy Fairhurst-Hunter, Wei Gan, Simon Gilbert, Alex Hacker, Parisa Hariri, Mike Hill, Michael Holmes, Pek Kei Im, Andri Iona, Maria Kakkoura, Christiana Kartsonaki, Rene Kerosi, Kuang Lin, Iona Millwood, Qunhua Nie, Alfred Pozarickij, Paul Ryder, Sam Sansome, Dan Schmidt, Paul Sherliker, Rajani Sohoni, Becky Stevens, Iain Turnbull, Robin Walters, Lin Wang, Neil Wright, Ling Yang, Xiaoming Yang, Pang Yao.

National Co-ordinating Centre, Beijing: Yu Guo, Xiao Han, Can Hou, Chun Li, Chao Liu, Jun Lv, Pei Pei, Canqing Yu.

10 Regional Co-ordinating Centres:

Guangxi Provincial CDC: Naying Chen, Duo Liu, Zhenzhu Tang. Liuzhou CDC: Ningyu Chen, Qilian Jiang, Jian Lan, Mingqiang Li, Yun Liu, Fanwen Meng, Jinhuai Meng, Rong Pan, Yulu Qin, Ping Wang, Sisi Wang, Liuping Wei, Liyuan Zhou. **Gansu** Provincial CDC: Caixia Dong, Pengfei Ge, Xiaolan Ren. Maiji CDC: Zhongxiao Li, Enke Mao, Tao Wang, Hui Zhang, Xi Zhang. **Hainan** Provincial CDC: Jinyan Chen, Ximin Hu, Xiaohuan Wang. Meilan CDC: Zhendong Guo, Huimei Li, Yilei Li, Min Weng, Shukuan Wu. **Heilongjiang** Provincial CDC: Shichun Yan, Mingyuan Zou, Xue Zhou. Nangang CDC: Ziyan Guo, Quan Kang, Yanjie Li, Bo Yu, Qinai Xu. **Henan** Provincial CDC: Liang Chang, Lei Fan, Shixian Feng, Ding Zhang, Gang Zhou. Huixian CDC: Yulian Gao, Tianyou He, Pan He, Chen Hu, Huarong Sun, Xukui Zhang. **Hunan** Provincial CDC: Biyun Chen, Zhongxi Fu, Yuelong Huang, Huilin Liu, Qiaohua Xu, Li Yin. Liuyang CDC: Huajun Long, Xin Xu, Hao Zhang, Libo Zhang. **Jiangsu** Provincial CDC: Jian Su, Ran Tao, Ming Wu, Jie Yang, Jinyi Zhou, Yonglin Zhou. Suzhou CDC: Yihe Hu, Yujie Hua, Jianrong Jin Fang Liu, Jingchao Liu, Yan Lu, Liangcai Ma, Aiyu Tang, Jun Zhang. **Qingdao** Qingdao CDC: Liang Cheng, Ranran Du, Ruqin Gao, Feifei Li, Shanpeng Li, Yongmei Liu, Feng Ning, Zengchang Pang, Xiaohui Sun, Xiaocao Tian, Shaojie Wang, Yaoming Zhai, Hua Zhang, Licang CDC: Wei Hou, Silu Lv, Junzheng Wang. **Sichuan** Provincial CDC: Xiaofang Chen, Xianping Wu, Ningmei Zhang, Weiwei Zhou. Pengzhou CDC: Xiaofang Chen, Jianguo Li, Jiaqiu Liu, Guojin Luo, Qiang Sun, Xunfu Zhong. **Zhejiang** Provincial CDC: Weiwei Gong, Ruying Hu, Hao Wang, Meng Wan, Min Yu. Tongxiang CDC: Lingli Chen, Qijun Gu, Dongxia Pan, Chunmei Wang, Kaixu Xie, Xiaoyi Zhang.

kConFab

David Amor, Medical Geneticist, Genetic Health Services, Victoria Royal Children's Hospital, Melbourne VIC 3050

Lesley Andrews, Hereditary Cancer Clinic, Prince of Wales Hospital Randwick, NSW 2031

Yoland Antill, Dept. Haem and Medical Oncology, Peter MacCallum Cancer Centre, St Andrews Place, East Melbourne VIC 3002

Rosemary Balleine, Department of Translational Oncology, C/- Department of Medical Oncology, Westmead Hospital, Westmead NSW 2145

Jonathan Beesley, Research Officer, Queensland Institute of Medical Research, Herston Road Herston, Qld 4002 Australia

Ian Bennett, Silverton Place, 101 Wickham, Terrace Brisbane QLD 4000

Michael Bogwitz, Familial Cancer Centre, The Royal Melbourne Hospital, Grattan Street Parkville, Victoria 3050 Australia

Leon Botes, Clinical Nurse Specialist, Hereditary Cancer Centre, Prince of Wales Hospital, Barker St, Randwick, NSW 2031 Australia

Meagan Brennan, NSW Breast Cancer Institute, PO Box 143, Westmead NSW 2145

Melissa Brown, Department of Biochemistry, University of Queensland, St. Lucia QLD 4072

Michael Buckley, Molecular and Cytogenetics Unit, Prince of Wales Hospital, Randwick NSW 2031

Jo Burke, Royal Hobart Hospital, GPO Box 1061L Hobart TAS 7001

Phyllis Butow, Medical Psychology Unit, Royal Prince Alfred Hospital, Camperdown NSW 2204

Liz Caldon, Replication and Genome Stability Cancer Division, Garvan Institute of Medical Research, 370 Victoria Street, Darlinghurst NSW 2010, Australia

Ian Campbell, Peter MacCallum Cancer Centre, St Andrew's Place, East Melbourne VIC 3002

Deepa Chauhan, School of Psychology, Brennan McCallum (Building A18), University of Sydney 2006

Manisha Chauhan, St Vincents Hospital Cancer Genetics Clinic, The Kinghorn Cancer Centre Sydney NSW

Georgia Chenevix-Trench, Queensland Institute of Medical Research, Royal Brisbane Hospital, Herston QLD 4029

Alice Christian, Genetics Department, Central Region Genetics Service, Wellington Hospital, New Zealand

Paul Cohen, Director of Gynaecological Cancer Research, St John of God Subiaco Hospital, 12 Salvado Road Subiaco WA 6008, Australia

Alison Colley, Department of Clinical Genetics, Liverpool Health Service, PO Box 103, Liverpool NSW 2170

Ashley Crook, Department of Clinical Genetics, Level 3E, Royal North Shore Hospital, St Leonards NSW 2065

James Cui, Epidemiology and Preventive Medicine, Monash University, Prahan Vic 3004, Australia

Margaret Cummings, Department of Pathology, University of Queensland Medical School, Herston NSW 4006

Sarah-Jane Dawson, Molecular Genetics Department, Cambridge University, England

Anna deFazio, Dept. Gynaecological Oncology, Westmead Institute for Cancer Research, Westmead Hospital Westmead NSW 2145

Martin Delatycki, Director, Clinical Genetics, Austin Health, Heidelberg Repatriation Hospital, PO Box 5444, Heidelberg West Vic 3081, Australia

Rebecca Dickson, Associate Genetic Counsellor, Level 2, Block 51 Royal North Shore Hospital, North Shore NSW 2408

Joanne Dixon, Central Regional Genetic Services, Wellington Hospital Private bag 7902, Wellington, New Zealand

Ted Edkins, Clinical Chemistry, Princess Margret Hospital for Children, Box D184, Perth WA 6001

Stacey Edwards, Department of Biochemistry and Molecular Biology, University of Queensland, St Lucia Qld 4072, Australia

Gelareh Farshid, Tissue Pathology, IMVS, Adelaide SA 5000

Andrew Fellows, Molecular Diagnostic Development, Pathology Department, Peter MacCallum Cancer Centre, Melbourne East Melbourne Vic 3002

Georgina Fenton, South West Family Cancer Clinic, Liverpool Hospital Liverpool, BC NSW 1871

Michael Field, Clinical Geneticist, Royal North Shore Hospital, Level 2, Vindin House, St Leonards NSW 2065

James Flanagan, Epigenetics Unit, Department of Surgery and Oncology, Imperial College London, London W12 0NN, England

Peter Fong, Medical Oncology Department, Regional Cancer and Blood, Services, Level 1 Building 7, Auckland City Hospital, 2 Park Rd. Grafton, Auckland 1023, New Zealand

Laura Forrest, Psychosocial Cancer Genetics Research Group, Parkville Familial Cancer Centre, 305 Grattan Street, Melbourne Vic 3000, Australia

Stephen Fox, Pathology Department, Level 1, Peter MacCallum Cancer Centre, St Andrew's Place, East Melbourne Vic 3002

Juliet French, School of Molecular and Microbial Sciences, University of Queensland, St Lucia Qld 4072

Michael Friedlander, Professor of Medicine, Department of Medical Oncology, Prince of Wales Hospital, Randwick NSW 2031

Clara Gaff, Victorian Clinical Genetics Service, Royal Melbourne Hospital, Parkville VIC 3052

Mike Gattas, Queensland Clinical Genetic Service, Royal Children's Hospital, Bramston Terrace, Herston QLD 4020

Peter George, Clinical Biochemistry Unit, Canterbury Health Labs, PO Box 151, Christchurch, New Zealand

Sian Greening, Illawarra Cancer Centre, Wollongong Hospital, Private Mail Bag 8808, South Coast Mail Centre, NSW 2521

Marion Harris, Familial Cancer Clinic, Peter MacCallum Cancer Centre, St Andrews Place East Melbourne VIC 3002

Stewart Hart, Breast and Ovarian Cancer Genetics, Monash Medical Centre, 871 Centre Road, Bentleigh East VIC 3165

Nick Hayward, Queensland Institute for Medical Research, Royal Brisbane Hospital, Post Office, Herston QLD 4029

John Hopper, Centre for M.E.G.A. Epidemiology, University of Melbourne, Level 1, 723 Swanston Street, Carlton VIC 3010

Cass Hoskins, Parkville Familial Cancer Centre, Peter MacCallum Cancer Centre & The Royal Melbourne Hospital, Melbourne, 3000

Clare Hunt, Southern Health Familial Cancer Centre, Monash Medical Centre, Special Medicine Building, 246, Clayton Rd Clayton Victoria 3168, Australia

Paul James, Clinical Geneticist, Genetic Health Sevices, Monash Medical Centre, Clayton Vic

Mark Jenkins, Centre for M.E.G.A. Epidemiology, The University of Melbourne, 723 Swanston Street, Carlton VIC 3053

Alexa Kidd, Clinical Genetics Departments, Central Regional Genetics Service, Wellington Hospital, New Zealand

Judy Kirk, Familial Cancer Service, Department of Medicine, Westmead Hospital, Westmead NSW 2145

Jessica Koehler, Hereditary Cancer Clinic, Prince of Wales Hospital, Randwick NSW 2031

James Kollias, Breast Endocrine and Surgical Unit, Royal Adelaide Hospital, North Terrace SA 5000

Sunil Lakhani, UQ Centre for Clinical Research, Level 6 Building 71/918, University of Queensland The Royal Brisbane &amp, Women's Hospital Herston, 4029

Mitchell Lawrence, Prostate Cancer Research Program, 19 Innovation Walk, Level 3, Monash University, Clayton, 3800

Geoff Lindeman, Breast Cancer Laboratory, Walter and Eliza Hall Institute, PO Royal Melbourne Hospital, Parkville VIC 3050

Lara Lipton, Medical Oncology and Clinical Haematology Unit, Western Hospital, Footscray VIC

Liz Lobb, Medical Psychology Research Unit, Room 332, Brennan MacCallum Building (A18), The University of Sydney, Camperdown, 2006

Graham Mann, Westmead Institute for Cancer Research, Westmead Millennium Institute, Westmead NSW 2145

Deborah Marsh, Kolling Institute of Medical Research, Royal North Shore Hospital, St Leonards NSW 2065

Sue Anne McLachlan, Department of Oncology, St Vincent's Hospital, 41 Victoria Parade, Fitzroy VIC 3065

Bettina Meiser, Hereditary Cancer Clinic, Prince of Wales Hospital, Randwick NSW 2031

Roger Milne, Centro Nacional de Investigaciones Oncologicas, C/ Melchor Fernández Almagro, 3, E-28029 Madrid, Spain

Sophie Nightingale, Western Health and Peter MacCallum Cancer Centre, Consultant, General, Breast and Melanoma Surgeon, St Andrew's Place, East Melbourne Victoria 3002

Shona O'Connell, Southern Health Familial Cancer Centre, Special Medicine Building, 246 Clayton Road, Clayton Vic 3168

Sarah O'Sullivan, Genetic Services of Western, Level 3 Agnes Walsh House, 374 Bagot Road, Subiaco WA 6008, Australia

David Gallego Ortega, Tumour Development Group, Garvan Institute of Medical Research, The Kinghorn Cancer Centre, 370 Victoria St, Darlinghurst NSW 2010, Australia

Nick Pachter, Familial Cancer and Clinical Genetics, Royal Melbourne Hospital, Grattan Street, Parkville VIC 3050, Australia

Briony Patterson, Tas Clinical Genetics Service, Royal Hobart Hospital, GPO Box 1061, Hobart Tasmania 7001, Australia

Amy Pearn, The Gene Council, Perth, Australia, PO Box 510, North Perth WA 6906, Australia

Kelly Phillips, Department of Medical Oncology, Peter MacCallum Cancer Centre, St Andrew's Place, East Melbourne VIC 3002

Ellen Pieper, Associate Genetic Counsellor, Parkville Familial Cancer Centre and Genomic Medicine, VCCC Grattan Street, Melbourne Vic 3000, Australia

Edwina Rickard, Familial Cancer centre, Westmead Hospital, Westmead NSW 2145

Bridget Robinson, Oncology Service, Christchurch Hospital Private Bag 4710, Christchurch, New Zealand

Mona Saleh, Centre for Genetic Education, Prince of Wales Hospital, Randwick NSW 2031

Elizabeth Salisbury, Anatomical Pathology, Conjoint Associate Professor, UNSW, Prince of Wales Hospital Randwick, 2031 NSW

Christobel Saunders, School of Surgery and Pathology, QE11 Medical Centre, M block 2nd Floor, Nedlands WA 6907

Jodi Saunus, Breast Pathology, University of Queensland Centre for Clinical Research, Building 71/918 Royal Brisbane and Women's Hospital, Herston Qld 4029

Rodney Scott, Hunter Area Pathology Service, John Hunter Hospital, Locked Bag 1 Regional Mail Centre, NSW 2310

Clare Scott, Research Department, WEHI C/o, Royal Melbourne Hospital, Parkville, 3050

Adrienne Sexton, Familial Cancer Centre, Royal Melbourne Hospital, Grattan Street, Parkville Vic 3050

Andrew Shelling, Obstetrics and Gynaecology, University of Auckland, New Zealand

Peter Simpson, The University of Queensland, Building 71/918, RBWH Campus, Herston Qld 4029

Melissa Southey, Genetic Epidemiology Laboratory, Departemnt of Pathology, University of Melbourne, VIC 3010

Amanda Spurdle, Cancer Unit, Queensland Institute of Medical Research, Herston QLD 4029

Jessica Taylor, Familial Cancer and Genetics Medicine, Royal Melbourne Hospital, 2nd Floor Grattan Street, Parkville Vic 3050 Australia

Renea Taylor, Deputy Head, Cancer Program. Monash University, Rm 349, Level 3, Building 76, 19 Innovation Walk, Clayton VIC 3800

Heather Thorne, Research Department, Peter MacCallum Cancer Centre, St Andrew's Place, East Melbourne VIC 3002

Alison Trainer, University of NSW, Prince of Wales Hospital, Barker Street, Randwick NSW 2031

Kathy Tucker, Heredity Cancer Clinic, Prince of Wales Hospital, Randwick NSW 2031

Jane Visvader, The Walter and Eliza Hall Institute of Medical Research, Post Office Royal Melbourne Hospital, Parkville VIC 3050

Logan Walker, Molecular Cancer Epidemiology Laboratory, Queensland Institute of Medical Research, P.O. Royal Brisbane Hospital, Herston Qld 4027, Australia

Rachael Williams, Family Cancer Clinic, St Vincent's Hospital, Darlinghurst NSW 2010
Ingrid Winship, Department of Genetics, Royal Melbourne Hospital, Parkville, 3050
Mary Ann Young, Genome.One, 370 Victoria St Darlinghurst, 2010 NSW

The LifeLines Cohort Study

Behrooz Z. Alizadeh[1], H. Marike Boezen[1], Lude Franke[2], Pim van der Harst[3], Gerjan Navis[4], Marianne Rots[5], Harold Snieder[1], Morris Swertz[2], Bruce H. R. Wolffenbuttel[6] & Cisca Wijmenga[2]

[1]Department of Epidemiology. University of Groningen, University Medical Center Groningen, Groningen, The Netherlands.

[2]Department of Genetics, University of Groningen, University Medical Center Groningen, Groningen, The Netherlands.

[3]Department of Cardiology, University of Groningen, University Medical Center Groningen, Groningen, The Netherlands.

[4]Division of Nephrology, Department of Internal Medicine, University of Groningen, University Medical Center Groningen, Groningen, The Netherlands.

[5]Department of Pathology and Medical Biology, University of Groningen, University Medical Center Groningen, Groningen, The Netherlands.

[6]Department of Endocrinology, University of Groningen, University Medical Center Groningen, Groningen, The Netherlands.

The InterAct consortium

Claudia Langenberg(1), Robert A Scott(1), Stephen J Sharp(1), Nita G Forouhi(1), Nicola D Kerrison(1), Matt Sims(1), Debora ME Lucarelli(1), Inês Barroso(1,2), Panos Deloukas(3), Mark I McCarthy(4,5,6), Antonio Agudo(7), Beverley Balkau(8), Aurelio Barricarte(9,10,11), Heiner Boeing(12), Miren Dorronsoro(13,14,11), Paul W Franks(15,16), Sara Grioni(17), Rudolf Kaaks(18), Timothy J Key(19), Carmen Navarro(20,11,21), Peter M Nilsson(22), Kim Overvad(23,24), Domenico Palli(25), Salvatore Panico(26), J. Ramón Quirós(27), Olov Rolandsson(28), Carlotta Sacerdote(29), María-José Sánchez(30,11,31), Nadia Slimani(32), Annemieke MW Spijkerman(33), Anne Tjonneland(34), Rosario Tumino(35,36), Yvonne T van der Schouw(37), Elio Riboli(38), Nicholas J Wareham(1)

Affiliations (de-duplicated by ID Number):

(1) MRC Epidemiology Unit, University of Cambridge School of Clinical Medicine, Box 285 Institute of Metabolic Science, Cambridge Biomedical Campus Cambridge, CB2 0QQ, United Kingdom, (2) University of Cambridge Metabolic Research Laboratories, Institute of Metabolic Science Addenbrooke's Hospital CB2 0QQ Cambridge, (3) Department of Human Genetics, Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge, CB10 1SA, United Kingdom, (4) Oxford Centre for Diabetes, Endocrinology and Metabolism (OCDEM), University of Oxford, Churchill Hospital, Old Road, Headington, Oxford, OX3 7LJ, UK, (5) Wellcome Trust Centre for Human Genetics, University of Oxford, Roosevelt Drive, Oxford, OX3 7BN, UK, (6) Oxford NIHR Biomedical Research Centre, Churchill Hospital, Old Road, Headington, Oxford, OX3 7LJ, UK, (7) Unit of Nutrition, Environment and Cancer, Cancer Epidemiology Research Program, Catalan Institute of Oncology, 08908 Barcelona, Spain, (8) CESP, University Paris-South, Faculty of Medicine, University Versailles-St Quentin, Inserm U1018, University Paris-Saclay, 94800 Villejuif, France, (9) Navarre Public Health Institute, Leyre 15, 31003 Pamplona, Navarra, Spain, (10) Navarra Institute for Health Research (IdiSNA) Pamplona, Spain, (11) Consortium for Biomedical Research in Epidemiology and Public Health (CIBER Epidemiología y Salud Pública), Av. Monforte de Lemos, 3-5. Pabellón 11. Planta 0 28029 Madrid, Spain, (12) Department of Epidemiology, German Institute of Human Nutrition Potsdam-Rehbruecke, Arthur-Scheunert-Allee 114-116, 14558 Nuthetal, Germany, (13) Public Health Division of Gipuzkoa, Basque Government, Av. Navarra 4, 20013 San Sebastian, Spain, (14) Instituto BIO-Donostia, Basque Government, San Sebastian, Spain, (15) Department of Clinical Sciences, Clinical Research Center, Skåne University Hospital, Lund University, 20502 Malmö, Sweden, (16) Department of Public Health and Clinical Medicine, Umeå University, 90187 Umeå, Sweden, (17) Fondazione IRCCS Istituto Nazionale dei Tumori Milan, Via Venezian, 1, 20133 Milan, Italy, (18) Division of Cancer Epidemiology, German Cancer Research Center (DKFZ), Im Neuenheimer Feld 581, 69120 Heidelberg, Germany, (19) Cancer Epidemiology Unit, Nuffield Department of Population Health, University of Oxford, Oxford, United Kingdom, (20) Department of Epidemiology, Murcia Regional Health Council, IMIB-Arrixaca, Ronda de Levante, 11, 30008, Murcia, Spain, (21) Unit of Preventive Medicine and Public Health, School of Medicine, University of Murcia, Spain, (22) Department of Clinical Sciences, Clinical Research Center, Skåne University Hospital, Lund University, 20502 Malmö, Sweden, (23) Department of Public Health, Aarhus University, Bartholins Allé 2, DK-8000 Aarhus C, Denmark, (24) Department of Cardiology, Aalborg University Hospital, Sdr. Skovvej 15, DK-9000 Aalborg, Denmark, (25) Cancer Risk Factors and Life-Style Epidemiology Unit, Cancer Research and Prevention Institute – ISPO, Florence, Italy, (26) Dipartimento di Medicina Clinica e Chirurgia, Federico II University, via Pansini 5-80131 Naples, Italy, (27) Consejería de Sanidad, Public Health Directorate, C/Ciriaco Miguel Vigil 9, 33006- Oviedo-Asturias, Spain, (28) Department of Public Health and Clinical Medicine, Family Medicine, Umeå University 90187 Umeå, Sweden, (29) Unit of Cancer Epidemiology, AO Citta' della Salute e della Scienza Hospital-University of Turin and

Center for Cancer Prevention (CPO), Via Santena 7, 10126 Torino, Italy, (30) Andalusian School of Public Health, Cuesta del Observatorio 4, CP 18080 Granada, Spain, (31) Instituto de Investigación Biosanitaria de Granada (ibs.GRANADA), Universidad de Granada, Granada, Spain, (32) International Agency for Research on Cancer, Dietary Exposure Assessment Group (DEX), 150 Cours Albert Thomas, 69372 Lyon Cedex 08, France, (33) National Institute for Public Health and the Environment (RIVM), PO Box 1, 3720 BA Bilthoven, The Netherlands, (34) Danish Cancer Society Research Center, Strandboulevarden 49, 2100 Copenhagen, Denmark, (35) Cancer Registry and Histopathology Department, Azienda Sanitaria Provinciale No 7, Piazza Igea Nr 1, 97100 Ragusa, Italy, (36) Associazone Iblea per la Ricerca Epidemiologica -Organizazione Non Lucrativa di Utilità Sociale, Piazza Amcione No 2, 97100 Ragusa, Italy, (37) Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Utrecht University, 3584 CG Utrecht, the Netherlands, (38) School of Public Health, Imperial College London, Norfolk Place, London W2 1PG, UK

23andMe Research Team

Michelle Agee, Robert K Bell, Katarzyna Bryc, Sarah L Elson, David A Hinds, Karen E Huber, Aaron Kleinman, Nadia K Litterman, Jennifer C McCreight, Matthew H McIntyre, Joanna L Mountain, Elizabeth S Noblin, Carrie A M Northover, J Fah Sathirapongsasuti, Olga V Sazonova, Janie F Shelton, Suyash Shringarpure, Chao Tian, Joyce Y Tung, Vladimir Vacic

Affiliation: 23andMe Inc., 223N Mathilda Ave, Sunnyvale, CA, 94086, USA.