Understanding the genetic complexity of puberty timing across the allele frequency spectrum

Katherine A Kentistou[1*], Lena R Kaisinger[1*], Stasa Stankovic[1], Marc Vaudel[2,3], Edson M de Oliveira[4], Andrea Messina[5], Robin G Walters[6,7], Xiaoxi Liu[8], Alexander S Busch[9,10], Hannes Helgason[11,12], Deborah J Thompson[13], Federico Santon[5], Konstantin M Petricek[14], Yassine Zouaghi[5], Isabel Huang-Doran[4], Daniel F Gudbjartsson[11,12], Eirik Bratland[2,15], Kuang Lin[6], Eugene J Gardner[1], Yajie Zhao[1], Raina Jia[1], Chikashi Terao[8,16,17], Margie Riggan[18], Manjeet K Bolla[13], Mojgan Yazdanpanah[19], Nahid Yazdanpanah[19], Jonath P Bradfield[20,21], Linda Broer[22,23], Archie Campbell[24], Daniel I Chasman[25], Diana L Cousminer[26,27,28], Nora Franceschini[29], Lude H Franke[30], Giorgia Girotto[31,32], Chunyan He[33,34], Marjo-Riitta Järvelin[35,36,37,38,39], Peter K Joshi[40], Yoichiro Kamatani[41], Robert Karlsson[42], Jian'an Luan[1], Kathryn L Lunetta[43,44], Reedik Mägi[45], Massimo Mangino[46,47], Sarah E Medland[48,49,50], Christa Meisinger[51], Raymond Noordam[52], Teresa Nutile[53], Maria Pina Concas[31], Ozren Polašek[54,55], Eleonora Porcu[56,57], Susan M Ring[58,59], Cinzia Sala[60], Albert V Smith[61,62], Toshiko Tanaka[63], Peter J van der Most[64], Veronique Vitart[65], Carol A Wang[66,67], Gonneke Willemsen[68], Marek Zygmunt[69], Thomas U Ahearn[70], Irene L Andrulis[71,72], Hoda Anton-Culver[73], Antonis C Antoniou[13], Paul L Auer[74], Catriona LK Barnes[40], Matthias W Beckmann[75], Amy Berrington[76], Natalia V Bogdanova[77,78,79], Stig E Bojesen[80,81], Hermann Brenner[82,83,84], Julie E Buring[25], Federico Canzian[85], Jenny Chang-Claude[86,87], Fergus J Couch[88], Angela Cox[89], Laura Crisponi[56], Kamila Czene[42], Mary B Daly[90], Ellen W Demerath[91], Joe Dennis[13], Peter Devilee[92,93], Immaculata De Vivo[94,95], Thilo Dörk[78], Alison M Dunning[96], Miriam Dwek[97], Johan G Eriksson[98], Peter A Fasching[99], Lindsay Fernandez-Rhodes[100], Liana Ferreli[56], Olivia Fletcher[101], Manuela Gago-Dominguez[102], Montserrat García-Closas[70], José A García-Sáenz[103], Anna González-Neira[104], Harald Grallert[105,106], Pascal Guénel[107], Christopher A Haiman[108], Per Hall[42,109], Ute Hamann[110], Hakon Hakonarson[21,26,111,112], Roger J Hart[113], Martha Hickey[114], Maartje J Hooning[115], Reiner Hoppe[116,117], John L Hopper[118], Jouke-Jan Hottenga[68], Frank B Hu[119,95], Hanna Hübner[99], David J Hunter[94,6], ABCTB Investigators[120], Helena Jernström[121], Esther M John[122,123], David Karasik[124,125], Elza K Khusnutdinova[126,127], Vessela N Kristensen[128,129], James V Lacey[130,131], Diether Lambrechts[132,133], Lenore J Launer[134], Penelope A Lind[48,135,50], Annika Lindblom[136,137], Patrik KE Magnusson[42], Arto Mannermaa[138,139], Mark I McCarthy[140,141,142], Thomas Meitinger[143], Cristina Menni[46], Kyriaki Michailidou[13,144], Iona Y Millwood[6,7], Roger L Milne[145,118], Grant W Montgomery[146], Heli Nevanlinna[147], Ilja M Nolte[148], Dale R Nyholt[149], Nadia Obi[150], Katie M O'Brien[151], Kenneth Offit[152,153], Albertine J Oldehinkel[154], Sisse R Ostrowski[155,156], Aarno Palotie[157,158,159,160,161,162], Ole B Pedersen[163,156], Annette Peters[106,164], Giulia Pianigiani[31], Dijana Plaseska-Karanfilska[165], Anneli Pouta[166], Alfred Pozarickij[6], Paolo Radice[167], Gad Rennert[168], Frits R Rosendaal[169], Daniela Ruggiero[53,170], Emmanouil Saloustros[171], Dale P Sandler[151], Sabine Schipf[172], Carsten O Schmidt[172], Marjanka K Schmidt[173,174], Kerrin Small[46], Beatrice Spedicati[32], Meir Stampfer[94,95], Jennifer Stone[175,118], Rulla M Tamimi[176,94], Lauren R Teras[177], Emmi Tikkanen[178,162], Constance Turman[94,179], Celine M Vachon[180], Qin Wang[13], Robert Winqvist[181,182], Alicja Wolk[183], Babette S Zemel[111,184], Wei Zheng[185], Ko W van Dijk[93,186], Behrooz Z Alizadeh[148], Stefania Bandinelli[187], Eric Boerwinkle[188], Dorret I Boomsma[68,189], Marina Ciullo[53,170], Georgia Chenevix-Trench[48], Francesco Cucca[56,57], Tõnu Esko[45], Christian Gieger[105,106,190], Struan FA Grant[26,27,28,111,191], Vilmundur

Gudnason[61,62], Caroline Hayward[65], Ivana Kolčić[54,55], Peter Kraft[94,179], Deborah A Lawlor[58,59], Nicholas G Martin[48], Ellen A Nøhr[192], Nancy L Pedersen[42], Craig E Pennell[66,67,193], Paul M Ridker[25], Antonietta Robino[31], Harold Snieder[148], Ulla Sovio[35,194], Tim D Spector[46], Doris Stöckl[195], Cathie Sudlow[24,196], Nic J Timpson[58,59], Daniela Toniolo[60], André Uitterlinden[22,23], Sheila Ulivi[31], Henry Völzke[172], Nicholas J Wareham[1], Elisabeth Widen[162], James F Wilson[40], The Lifelines Cohort Study†, The Danish Blood Donor study†, The Ovarian Cancer Association Consortium†, The Breast Cancer Association Consortium†, The Biobank Japan Project, The China Kadoorie Biobank Collaborative Group†, Paul DP Pharoah[13,96], Liming Li[197,198], Douglas F Easton[13,96], Pål Njølstad[2,199], Patrick Sulem[11], Joanne M Murabito[44,200], Anna Murray[201], Despoina Manousaki[19,202,203], Anders Juul[204,205,206], Christian Erikstrup[207,208], Kari Stefansson[11,62], Momoko Horikoshi[209], Zhengming Chen[6,7], I Sadaf Farooqi[4], Nelly Pitteloud[5,210], Stefan Johansson[2,15], Felix R Day[1]*, John RB Perry[1,211]*‡, Ken K Ong[1,212]*

1 MRC Epidemiology Unit, University of Cambridge School of Clinical Medicine, Box 285 Institute of Metabolic Science, Cambridge Biomedical Campus, Cambridge CB2 0QQ, UK
2 Mohn Center for Diabetes Precision Medicine, Department of Clinical Science, University of Bergen, NO-5020, Bergen, Norway
3 Department of Genetics and Bioinformatics, Health Data and Digitalization, Norwegian Institute of Public Health, NO-0213, Oslo, Norway
4 University of Cambridge Metabolic Research Laboratories and NIHR Cambridge Biomedical Research Centre, Wellcome-MRC Institute of Metabolic Science, Addenbrooke's Hospital, Cambridge, UK
5 Division of Endocrinology, Diabetology, and Metabolism, Lausanne University Hospital, 1011 Lausanne, Switzerland
6 Nuffield Department of Population Health, University of Oxford, Oxford OX3 7LF, UK
7 MRC Population Health Research Unit, University of Oxford, Oxford OX3 7LF, UK
8 Laboratory for Statistical and Translational Genetics, RIKEN Center for Integrative Medical Sciences, Yokohama, Japan
9 Department of General Pediatrics, University of Münster, Münster, Germany
10 Deptartment of Growth and Reproduction, Copenhagen University Hospital - Rigshospitalet, Copenhagen, Denmark
11 deCODE Genetics/Amgen, Inc., Reykjavik, Iceland
12 School of Engineering and Natural Sciences, University of Iceland, Reykjavik, Iceland
13 Centre for Cancer Genetic Epidemiology, Department of Public Health and Primary Care, University of Cambridge, Cambridge CB1 8RN, UK
14 Charité Universitätsmedizin Berlin, corporate member of Freie Universität Berlin and Humboldt-Universität zu Berlin, Institute of Pharmacology, Berlin, Germany
15 Department of Medical Genetics, Haukeland University Hospital, NO-5021, Bergen, Norway
16 Clinical Research Center, Shizuoka General Hospital, Shizuoka, Japan
17 The Department of Applied Genetics, The School of Pharmaceutical Sciences, University of Shizuoka, Shizuoka, Japan
18 Department of Gynecology, Duke University Medical Center, Durham, North

Carolina, USA
19 Research Center of the Sainte-Justine University Hospital, University of Montreal, Montreal, Quebec, Canada
20 Quantinuum Research, Wayne, PA, USA
21 Center for Applied Genomics, Children's Hospital of Philadelphia, Philadelphia, PA, USA
22 Department of Internal Medicine, Erasmus MC, Rotterdam, The Netherlands
23 Department of Epidemiology, Erasmus MC, Rotterdam, The Netherlands
24 Centre for Genomic and Experimental Medicine, Institute of Genetics & Cancer, University of Edinburgh, Edinburgh, UK
25 Division of Preventive Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA 02215, USA
26 Division of Human Genetics, Children's Hospital of Philadelphia, Philadelphia, PA, USA
27 Department of Genetics, University of Pennsylvania, Philadelphia, PA, USA
28 Center for Spatial and Functional Genomics, Children's Hospital of Philadelphia, Philadelphia, PA, USA
29 Department of Epidemiology, University of North Carolina, Chapel Hill, NC, USA
30 Department of Genetics, University of Groningen, University Medical Center Groningen, Groningen, The Netherlands
31 Institute for Maternal and Child Health – IRCCS ''Burlo Garofolo", Trieste, Italy
32 Department of Medicine, Surgery and Health Sciences, University of Trieste, Trieste, Italy
33 Department of Epidemiology and Biostatistics, Department of Big Data in Health Science, School of Public Health, Zhejiang University School of Medicine, Hangzhou 310058, China
34 Departments of Medical Oncology and Hematology, Sir Runrun Shaw Hospital, School of Medicine, Zhejiang University, Hangzhou 310016, China
35 Department of Epidemiology and Biostatistics, MRC Health Protection Agency (HPA) Centre for Environment and Health, School of Public Health, Imperial College London, UK
36 Institute of Health Sciences, P.O.Box 5000, FI-90014 University of Oulu, Finland
37 Biocenter Oulu, P.O.Box 5000, Aapistie 5A, FI-90014 University of Oulu, Finland
38 Unit of Primary Care, Oulu University Hospital, Kajaanintie 50, P.O.Box 20, FI-90220 Oulu, 90029 OYS, Finland
39 Department of Children and Young People and Families, National Institute for Health and Welfare, Aapistie 1, Box 310, FI-90101 Oulu, Finland
40 Centre for Global Health Research, Usher Institute, University of Edinburgh, Teviot Place, Edinburgh EH8 9AG, Scotland
41 Laboratory of Complex Trait Genomics, Department of Computational Biology and Medical Sciences, Graduate School of Frontier Sciences, The University of Tokyo, Tokyo, Japan
42 Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden
43 Boston University School of Public Health, Department of Biostatistics. Boston, Massachusetts 02118, USA

44 NHLBI's and Boston University's Framingham Heart Study, Framingham, Massachusetts 01702-5827, USA
45 Institute of Genomics, University of Tartu, Tartu, Estonia
46 Department of Twin Research and Genetic Epidemiology, King's College London, London, UK
47 NIHR Biomedical Research Centre at Guy's and St. Thomas' Foundation Trust, London, UK
48 QIMR Berghofer Medical Research Institute, Brisbane, Queensland, Australia
49 School of Psychology, University of Queensland, Brisbane, Queensland, Australia
50 Faculty of Medicine, University of Queensland, Brisbane, Queensland, Australia
51 Epidemiology, Medical Faculty, University of Augsburg, University Hospital of Augsburg, Augsburg, Germany
52 Department of Internal Medicine, Section of Gerontology and Geriatrics, Leiden University Medical Center, Leiden, The Netherlands
53 Institute of Genetics and Biophysics "A. Buzzati-Traverso", CNR, Naples, Italy
54 University of Split School of Medicine, Split, Croatia
55 Algebra University College, Zagreb, Croatia
56 Institute of Genetics and Biomedical Research, National Research Council, Cagliari, Sardinia 09042, Italy
57 University of Sassari, Department of Biomedical Sciences, Sassari, Sassari 07100, Italy
58 MRC Integrative Epidemiology Unit at the University of Bristol, UK
59 Population Health Science, Bristol Medical School, University of Bristol, UK
60 Division of Genetics and Cell Biology, San Raffele Hospital, Milano, Italy
61 Icelandic Heart Association, 201 Kopavogur, Iceland
62 Faculty of Medicine, University of Iceland, 101 Reykjavik, Iceland
63 National Institute on Aging, National Institutes of Health, Baltimore, Maryland, USA
64 Department of Epidemiology, University of Groningen, University Medical Center Groningen, The Netherlands
65 MRC Human Genetics Unit, Institute of Genetics and Cancer, University of Edinburgh, Edinburgh EH4 2XU, UK
66 School of Medicine and Public Health, University of Newcastle, Newcastle, New South Wales 2308, Australia
67 Hunter Medical Research Institute, Newcastle, New South Wales 2305, Australia
68 Dept of Biological Psychology, Vrije Universiteit, Amsterdam; Amsterdam Public Health (APH) research institute, The Netherlands
69 Clinic of Gynaecology and Obstetrics, University Medicine Greifswald, Germany
70 Division of Cancer Epidemiology and Genetics National Cancer Institute, National Institutes of Health, Department of Health and Human Services Bethesda, MD, USA
71 Fred A. Litwin Center for Cancer Genetics Lunenfeld-Tanenbaum Research Institute of Mount Sinai Hospital Toronto, Ontario, Canada
72 Department of Molecular Genetics University of Toronto Toronto, Ontario, Canada
73 Department of Medicine, Genetic Epidemiology Research Institute University of California Irvine Irvine, CA, USA

74 Division of Biostatistics, Institute for Health and Equity, and Cancer Center Medical College of Wisconsin Milwaukee, WI, USA
75 Department of Gynecology and Obstetrics, Comprehensive Cancer Center Erlangen-EMN, Friedrich-Alexander University Erlangen-Nuremberg , University Hospital Erlangen, Erlangen, Germany
76 Division of Genetics and Epidemiology The Institute of Cancer Research, London, UK
77 Department of Radiation Oncology, Hannover Medical School, Hannover, Germany
78 Gynaecology Research Unit, Hannover Medical School, Hannover, Germany
79 N.N. Alexandrov Research Institute of Oncology and Medical Radiology, Minsk, Belarus
80 Copenhagen General Population Study, Herlev and Gentofte Hospital Copenhagen University Hospital, Herlev, Denmark
81 Department of Clinical Biochemistry, Herlev and Gentofte Hospital Copenhagen University Hospital, Herlev, Denmark
82 Division of Clinical Epidemiology and Aging Research German Cancer Research Center (DKFZ), Heidelberg, Germany
83 Division of Preventive Oncology German Cancer Research Center (DKFZ) and National Center for Tumor Diseases (NCT), Heidelberg, Germany
84 German Cancer Consortium (DKTK) German Cancer Research Center (DKFZ), Heidelberg, Germany
85 Genomic Epidemiology Group German Cancer Research Center (DKFZ), Heidelberg, Germany
86 Division of Cancer Epidemiology German Cancer Research Center (DKFZ), Heidelberg, Germany
87 Cancer Epidemiology Group, University Cancer Center Hamburg (UCCH) University Medical Center Hamburg-Eppendorf, Hamburg, Germany
88 Department of Laboratory Medicine and Pathology Mayo Clinic Rochester, MN, USA
89 Sheffield Institute for Nucleic Acids (SInFoNiA), Department of Oncology and Metabolism, University of Sheffield, Sheffield, UK
90 Department of Clinical Genetics Fox Chase Cancer Center Philadelphia, PA, USA
91 Division of Epidemiology and Community Health, School of Public Health, University of Minnesota, USA
92 Department of Pathology, Leiden University Medical Center, Leiden, The Netherlands
93 Department of Human Genetics, Leiden University Medical Center, Leiden, The Netherlands
94 Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, Massachusetts 02115, USA
95 Channing Division of Network Medicine, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, Massachusetts, USA
96 Centre for Cancer Genetic Epidemiology, Department of Oncology, University of Cambridge, Cambridge CB1 8RN, UK
97 School of Life Sciences, University of Westminster, London, UK

98 Department of General Practice and Primary Healthcare , University of Helsinki, Helsinki University Hospital, Helsinki, Finland
99 Department of Gynecology and Obstetrics, Comprehensive Cancer Center Erlangen-EMN, Friedrich-Alexander University Erlangen-Nuremberg, University Hospital Erlangen, Erlangen, Germany
100 Department of Biobehavioral Health, Pennsylvania State University, University Park, PA, USA
101 The Breast Cancer Now Toby Robins Research Centre, The Institute of Cancer Research, London, UK
102 Genomic Medicine Group, International Cancer Genetics and Epidemiology Group 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
103 Medical Oncology Department, Hospital Clínico San Carlos Instituto de Investigación Sanitaria San Carlos (IdISSC), Centro Investigación Biomédica en Red de Cáncer (CIBERONC), Madrid, Spain
104 Human Genotyping Unit-CeGen, Spanish National Cancer Research Centre (CNIO), Madrid, Spain
105 Research Unit of Molecular Epidemiology, Helmholtz Zentrum München– German Research Center for Environmental Health, Neuherberg, Germany
106 Institute of Epidemiology, Helmholtz Zentrum München–German Research Center for Environmental Health, Neuherberg, Germany
107 Team "Exposome and Heredity", CESP, Gustave Roussy INSERM, University Paris-Saclay, UVSQ Villejuif, France
108 Department of Preventive Medicine, Keck School of Medicine, University of Southern California, Los Angeles, CA, USA
109 Department of Oncology, Södersjukhuset, Stockholm, Sweden
110 Molecular Genetics of Breast Cancer, German Cancer Research Center (DKFZ), Heidelberg, Germany
111 Department of Pediatrics, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA, USA
112 Division of Pulmonary Medicine, Children’s Hospital of Philadelphia, Philadelphia, PA, USA
113 Division of Obstetrics and Gynaecology, University of Western Australia, Western Australia, Australia
114 Department of Obstetrics and Gynaecology at the University of Melbourne and The Royal Women's Hospital, Victoria, Australia
115 Department of Medical Oncology, Erasmus MC Cancer Institute, Rotterdam, The Netherlands
116 Dr. Margarete Fischer-Bosch-Institute of Clinical Pharmacology, Stuttgart, Germany
117 University of Tübingen, Tübingen, Germany
118 Centre for Epidemiology and Biostatistics, Melbourne School of Population and Global Health, The University of Melbourne Melbourne, Victoria, Australia
119 Department of Nutrition, Harvard T.H. Chan School of Public Health School of Public Health, Boston, Massachusetts 02115, USA

120 Australian Breast Cancer Tissue Bank, Westmead Institute for Medical Research, University of Sydney, Sydney, New South Wales, Australia
121 Oncology, Department of Clinical Sciences in Lund, Lund University, Lund, Sweden
122 Department of Epidemiology and Population Health, Stanford University School of Medicine Stanford, CA, USA
123 Department of Medicine, Division of Oncology Stanford Cancer Institute, Stanford University School of Medicine Stanford, CA, USA
124 Hebrew SeniorLife Institute for Aging Research, Boston, Massachusetts, USA
125 Harvard Medical School, Boston, Massachusetts, USA
126 Institute of Biochemistry and Genetics of the Ufa Federal Research Centre of the Russian Academy of Sciences, Ufa, Russia
127 Department of Genetics and Fundamental Medicine, Bashkir State University, Ufa, Russia
128 Department of Medical Genetics, Oslo University Hospital and University of Oslo, Oslo, Norway
129 Institute of Clinical Medicine, Faculty of Medicine, University of Oslo, Oslo, Norway
130 Department of Computational and Quantitative Medicine, City of Hope Duarte, CA, USA
131 City of Hope Comprehensive Cancer Center, City of Hope Duarte, CA, USA
132 Laboratory for Translational Genetics, Department of Human Genetics, KU Leuven, Leuven, Belgium
133 VIB Center for Cancer Biology, VIB, Leuven, Belgium
134 Laboratory of Epidemiology and Population Sciences, National Institute on Aging, Intramural Research Program, National Institutes of Health, Bethesda, Maryland, 20892, USA
135 School of Biomedical Sciences, Queensland University of Technology, Brisbane, Queensland, Australia
136 Department of Molecular Medicine and Surgery, Karolinska Institutet, Stockholm, Sweden
137 Department of Clinical Genetics, Karolinska University Hospital, Stockholm, Sweden
138 Translational Cancer Research Area, University of Eastern Finland, Kuopio, Finland
139 Institute of Clinical Medicine, Pathology and Forensic Medicine, University of Eastern Finland, Kuopio, Finland
140 Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, UK
141 Oxford Centre for Diabetes, Endocrinology, & Metabolism, University of Oxford, Churchill Hospital, Oxford OX3 7LJ, UK
142 NIHR Oxford Biomedical Research Centre, Churchill Hospital, OX3 7LE Oxford, UK
143 Institute of Human Genetics, Klinikum rechts der Isar, Technical University of Munich, School of Medicine, Munich, Germany
144 Biostatistics Unit, The Cyprus Institute of Neurology & Genetics, Nicosia, Cyprus
145 Cancer Epidemiology Division, Cancer Council Victoria, Melbourne, Victoria,

Australia
146 Institute for Molecular Bioscience, The University of Queensland, Brisbane, Queensland, Australia
147 Department of Obstetrics and Gynecology, Helsinki University Hospital, University of Helsinki, Helsinki, Finland
148 Department of Epidemiology, University of Groningen, University Medical Center Groningen, Groningen, The Netherlands
149 School of Biomedical Sciences, Faculty of Health, Centre for Genomics and Personalised Health, Queensland University of Technology, Brisbane, Queensland, Australia
150 Institute for Medical Biometry and Epidemiology, University Medical Center Hamburg-Eppendorf, Hamburg, Germany
151 Epidemiology Branch National Institute of Environmental Health Sciences, NIH Research Triangle Park, NC, USA
152 Clinical Genetics Research Lab, Department of Cancer Biology and Genetics Memorial Sloan Kettering Cancer Center New York, NY, USA
153 Clinical Genetics Service, Department of Medicine Memorial Sloan Kettering Cancer Center New York, NY, USA
154 Interdisciplinary Center Psychopathology and Emotion Regulation, University Medical Center Groningen, University of Groningen, The Netherlands
155 Department of Clinical Immunology, Rigshospitalet - University of Copenhagen, Copenhagen, Denmark
156 Department of Clinical Medicine, Faculty of health and medical sciences, University of Copenhagen, Denmark
157 Psychiatric and Neurodevelopmental Genetics Unit, Massachusetts General Hospital and Harvard Medical School, Boston, Massachusetts, USA
158 Medical and Population Genetics Program, Broad Institute of MIT and Harvard, Cambridge, Massachusetts, USA
159 Stanley Center for Psychiatric Research, Broad Institute of MIT and Harvard, Cambridge, Massachusetts, USA
160 Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, UK
161 Analytic and Translational Genetics Unit, Massachusetts General Hospital and Harvard Medical School, Boston, Massachusetts, USA
162 Institute for Molecular Medicine Finland (FIMM), University of Helsinki, Helsinki, Finland
163 Department of Clinical Immunology, Zealand University Hospital, Køge, Denmark
164 Institute for Medical Information Processing, Biometry and Epidemiology - IBE, Ludwig-Maximilians-Universität München, Munich, Germany
165 Research Centre for Genetic Engineering and Biotechnology "Georgi D. Efremov" MASA Skopje Republic of North Macedonia
166 National Institute for Health and Welfare, Finland
167 Unit of Molecular Bases of Genetic Risk and Genetic Testing, Department of Research Fondazione IRCCS, Istituto Nazionale dei Tumori (INT), Milan, Italy
168 Clalit National Cancer Control Center, Carmel Medical Center and Technion,

Faculty of Medicine, Haifa, Israel
169 Department of Clinical Epidemiology, Leiden University Medical Center, Leiden, The Netherlands
170 IRCCS Neuromed, Pozzilli, Isernia, Italy
171 Department of Oncology, University Hospital of Larissa, Larissa, Greece
172 Institute for Community Medicine, University Medicine Greifswald, Germany
173 Division of Molecular Pathology, The Netherlands Cancer Institute, Amsterdam, The Netherlands
174 Division of Psychosocial Research and Epidemiology, The Netherlands Cancer Institute - Antoni van Leeuwenhoek hospital, Amsterdam, The Netherlands
175 Genetic Epidemiology Group, School of Population and Global Health, University of Western Australia Perth, Western Australia, Australia
176 Department of Population Health Sciences Weill Cornell Medicine New York, NY, USA
177 Department of Population Science American Cancer Society Atlanta, GA, USA
178 Public Health Genomics Unit, Department of Chronic Disease Prevention, National Institute for Health and Welfare, Helsinki, Finland
179 Program in Genetic Epidemiology and Statistical Genetics, Harvard T.H. Chan School of Public Health, Boston, MA, 02115, USA
180 Department of Quantitative Health Sciences, Division of Epidemiology Mayo Clinic Rochester, MN, USA
181 Laboratory of Cancer Genetics and Tumor Biology, Translational Medicine Research Unit, Biocenter Oulu, University of Oulu, Oulu, Finland
182 Laboratory of Cancer Genetics and Tumor Biology, Northern Finland Laboratory Centre Oulu, Oulu, Finland
183 Institute of Environmental Medicine, Karolinska Institutet, Stockholm, Sweden
184 Division of Gastroenterology, Hepatology and Nutrition, Children's Hospital of Philadelphia, Philadelphia, PA, USA
185 Division of Epidemiology, Department of Medicine, Vanderbilt Epidemiology Center, Vanderbilt-Ingram Cancer Center Vanderbilt University School of Medicine Nashville, TN, USA
186 Department of Internal Medicine, Division of Endocrinology, Leiden University Medical Center, Leiden, The Netherlands
187 Geriatric Unit, Local Health Toscana Centro, Florence, Italy
188 Human Genetics Center, University of Texas Health Science Center at Houston, Houston, TX, USA
189 Amsterdam Reproduction & Development research institute, Amsterdam, The Netherlands
190 German Center for Diabetes Research (DZD), Neuherberg, Germany
191 Division of Endocrinology and Diabetes, Children's Hospital of Philadelphia, Philadelphia, PA, USA
192 Institute of Clinical Research, University of Southern Denmark, Department of Obstetrics & Gynecology, Odense University Hospital, Denmark
193 Department of Maternity and Gynaecology, John Hunter Hospital, Newcastle, New South Wales 2305, Australia
194 Department of Obstetrics and Gynaecology, University of Cambridge,

Cambridge, UK
195 Gesundheitsamt Fürstenfeldbruck, Regierung von Oberbayern, Fürstenfeldbruck, Germany
196 Centre for Medical Informatics, Usher Institute, University of Edinburgh
197 Department of Epidemiology and Biostatistics, School of Public Health, Peking University, Beijing, China
198 Center for Public Health and Epidemic Preparedness and Response, Peking University, Beijing, China
199 Department of Pediatrics and Adolescents, Haukeland University Hospital, NO-5021, Bergen, Norway
200 Boston University Chobanian & Avedisian School of Medicine, Department of Medicine, Section of General Internal Medicine, Boston, MA 02118, USA
201 Genetics of Complex Traits, University of Exeter Medical School, University of Exeter, RILD Level 3, Royal Devon & Exeter Hospital, Barrack Road, Exeter, EX2 5DW, UK
202 Department of Pediatrics, University of Montreal, Montreal, Canada
203 Department of Biochemistry and Molecular Medicine, University of Montreal, Montreal, Canada
204 Department of Growth and Reproduction, Copenhagen University Hospital - Rigshospitalet, Copenhagen, Denmark
205 International Center for Research and Research Training in Endocrine Disruption of Male Reproduction and Child Health (EDMaRC), Copenhagen University Hospital - Rigshospitalet, Copenhagen, Denmark
206 Department of Clinical Medicine, University of Copenhagen, Copenhagen, Denmark
207 Department of Clinical Immunology, Aarhus University Hospital, Aarhus, Denmark
208 Department of Clinical Medicine, Aarhus University, Aarhus, Denmark
209 Laboratory for Genomics of Diabetes and Metabolism, RIKEN Center for Integrative Medical Sciences, Yokohama, Japan
210 Faculty of Biology and Medicine, University of Lausanne, Lausanne 1005, Switzerland
211 Metabolic Research Laboratory, Wellcome-MRC Institute of Metabolic Science, University of Cambridge School of Clinical Medicine, Cambridge CB2 0QQ, UK
212 Department of Paediatrics, University of Cambridge, Cambridge CB2 0QQ, UK

* These authors contributed equally to this work.

† A full list is provided in the supplement.

‡ Correspondence should be addressed to: john.perry@mrc-epid.cam.ac.uk

Abstract

Pubertal timing varies considerably and has been associated with a range of health outcomes in later life. To elucidate the underlying biological mechanisms, we performed multi-ancestry genetic analyses in ~800,000 women, identifying 1,080 independent signals associated with age at menarche. Collectively these loci explained 11% of the trait variance in an independent sample, with women at the top and bottom 1% of polygenic risk exhibiting a ~11 and ~14-fold higher risk of delayed and precocious pubertal development, respectively. These common variant analyses were supported by exome sequence analysis of ~220,000 women, identifying several genes, including rare loss of function variants in *ZNF483* which abolished the impact of polygenic risk. Next, we implicated 660 genes in pubertal development using a combination of *in silico* variant-to-gene mapping approaches and integration with dynamic gene expression data from mouse embryonic GnRH neurons. This included an uncharacterized G-protein coupled receptor *GPR83*, which we demonstrate amplifies signaling of *MC3R*, a key sensor of nutritional status. Finally, we identified several genes, including ovary-expressed genes involved in DNA damage response that co-localize with signals associated with menopause timing, leading us to hypothesize that the ovarian reserve might signal centrally to trigger puberty. Collectively these findings extend our understanding of the biological complexity of puberty timing and highlight body size dependent and independent mechanisms that potentially link reproductive timing to later life disease.

Introduction

Age at menarche (AAM), the onset of menses in females, represents the start of reproductive maturity and is a widely reported marker of pubertal timing. Menarche normally occurs between ages 10 to 15 years[1], and its variation is associated with risks of several health outcomes, including obesity, type 2 diabetes (T2D), cardiovascular disease, and hormone-sensitive cancers[2,3,4,5,6]. Thus, widespread secular trends towards earlier puberty timing may have an important impact on public health[7]. AAM is a highly polygenic trait[8] and previous genome-wide association studies (GWAS) have identified ~400 common genetic loci[9,10,11,12,13], the vast majority of which were discovered in samples of European ancestry. AAM has a strong genetic correlation with male puberty timing (R_g=0.68)[14], as well as with adiposity (BMI, R_g=-0.35)[9] and specific pathways have been identified that link nutrient sensing to reproductive hormone axis activation. For example, we recently reported that *MC3R* is the key hypothalamic sensor linking nutritional status to puberty timing[15].

Previously reported GWAS signals in ~370,000 women of European ancestry explained ~7.4% of the population variance in AAM, corresponding to ~25% of the estimated heritability[9]. Here, through an expanded GWAS in up to 799,845 women, including 166,890 of East Asian ancestry, we identify 1080 independent signals for AAM. Female participants who carry an excess of these alleles have equivalent risk of precocious or delayed puberty compared to those carrying clinically relevant monogenic alleles. We complement these common variant analyses by undertaking the first large-scale assessment of rare variation in puberty timing in 222,283 women with exome sequence data. Through subsequent variant to gene mapping approaches, we implicate 660 genes, which collectively shed further light on the biological determinants of puberty timing and the mechanisms linking it to disease risks.

Results

We performed a GWAS meta-analysis for AAM, in up to 799,845 women, by combining data from five strata: i) 38 ReproGen consortium cohorts (N=180,269), ii) UK Biobank (N=238,040), iii) the Breast Cancer Association Consortium and the Ovarian Cancer Association Consortium (N=137,815), iv) 23andMe, Inc. (N=76,831), and v) three East Asian biobanks: the China Kadoorie Biobank, the Biobank Japan, and the Korean Genome and Epidemiology Study (N=166,890). All studies provided GWAS data imputed to at least 1000 Genomes reference panel density (Supplementary Table 1), yielding a total of ~12.7 million genetic variants in the final meta-analysis. We did not find evidence of test statistic inflation due to population structure (LDSC intercept=1.07, SE=0.03).

To maximise the discovery of genomic signals for AAM, we used a combination of distance-based clumping and approximate conditional analysis (see Methods) in the European-strata meta-analysis and in the all-ancestry meta-analysis, to identify signals that are homogenous across the two ancestry groups. European-strata identified signals (N=935) were supplemented with additional signals from the all-ancestry analysis (N=145), resulting in a total of 1080 statistically independent signals for AAM

at a genome-wide significance (P<5x10^{-8}, Figure 1, Supplementary Table 2). Effect sizes ranged from 3.5 months/allele for rarer alleles (MAF=0.9%) to ~5 days/allele for more common variants (Supplementary Figure 1). Across the 145 additional signals, we observed a median 1.16-fold increase in χ^2 for their association with AAM in the all-ancestry analysis compared to European-only, which is proportionate to the added number of East-Asian samples (~21% of the total).

Independent replication data from the Danish Blood Donors study (N=35,467) (Supplementary Table 3) was available for 969/1080 signals[16]. Of these, 862 showed directionally concordant associations (89%, $P_{Binomial}$=2.9×10^{-147}). In this independent sample, the variance explained in AAM doubled from 5.6% for 355 available previously reported signals[9] to 11% for the 969 signals with available data. We also sought indirect confirmation of AAM signals by association with age at voice breaking (AVB) in men from the UK Biobank study (N=191,235) and 23andMe (N=55,871) (Supplementary Table 4)[17,18,14]. Of the 1080 AAM signals, 909/1080 (84%, $P_{Binomial}$=2.6×10^{-122}) showed directionally concordant associations with AVB in UK Biobank (including 354 at P<0.05). Similarly, 852/1067 (79%, $P_{Binomial}$=1.8×10^{-90}) AAM signals available in 23andMe showed directionally concordant associations with AVB (217 at P<0.05).

Exome sequence analyses identify novel rare variants of large effect

Previous genetic studies for AAM have largely been restricted to assessing the role of common, largely non-coding, genetic variation. We sought to address this, by performing an exome-wide association study (ExWAS) in 222,283 European ancestry women in UK Biobank. Gene burden tests were performed by collapsing rare variants (MAF<0.1%) in each gene according to two overlapping predicted functional categories: i) high-confidence protein truncating variants (PTVs) and ii) PTVs plus missense variants with CADD score[19] ≥25 (termed 'damaging variants', DMG). Six genes were associated with AAM at exome-wide significance (P<1.54×10^{-6}, 0.05/32,434 tests, Figure 2, Supplementary Figure 2, Supplementary Figure 3, Supplementary Figure 4, Supplementary Table 5). This included two genes previously reported in rare monogenic disorders of puberty: *TACR3* (beta=0.62 years, P=3.2×10^{-19}, N=489 DMG carriers) previously implicated in normosmic idiopathic hypogonadotropic hypogonadism (IHH)[20], and *MKRN3* (beta=-0.59 years, P=1.4×10^{-7}, N=187 DMG carriers) previously implicated in familial central precocious puberty (CPP)[21]. Furthermore, *MC3R* (beta=0.33 years, P=1.6×10^{-9}, N=796 DMG carriers) was recently reported to link nutrient sensing to key hypothalamic neurons[15].

Of the three novel genes, *KDM4C* (beta=-0.33 years, P=2.5×10^{-7}, N=582 DMG carriers) encodes a lysine-specific histone demethylase likely involved in epigenetically regulating hypothalamic-pituitary-gonadal (HPG) axis genes[22]. A second gene in this small family (*KDM5B*) showed near exome-wide significant association with AAM (P=2.6×10^{-6}). In addition, *PDE10A* (beta=0.58 years, P=1.2×10^{-7}, N=196 DMG carriers) encodes phosphodiesterase 10A, which regulates the intracellular concentration of cyclic nucleotides and hence signal transduction[23].

Finally, *ZNF483* (beta=1.31 years, P=4.9×10^{-11}, N=59 DMG carriers), encodes a zinc finger protein transcription factor involved in neuronal differentiation[24] and self-renewal of pluripotent stem cells[25].

We were able to confirm four of these seven genes (*KDM4C*, *MC3R*, *TACR3* and *ZNF483*) using voice breaking data in 178,625 men with exome sequence data in UK Biobank (P<0.05, Figure 2, Supplementary Table 5). Lack of association with AVB at *MKRN3* is consistent with previous reports that rare *MKRN3* mutations have greater clinical impact in girls than boys[21,26]. None of the seven genes showed an association with childhood or adult adiposity (Supplementary Table 6).

In addition, we specifically examined rare variant associations with AAM or VB for *ANOS1*, *CHD7*, *FGF8* and *WDR11*, which are clinically tested in hypogonadotropic hypogonadism ('high evidence genes' on the Genomics England IHH panel[27]) and show a dominant or X-linked mode of inheritance (Supplementary Table 7). Normal puberty timing (AAM: 10-15 years[1] or VB: "about average") was reported by all carriers of PTVs in *ANOS1* (N=5 male) and *CHD7* (N=5 female, N=1 male). PTVs in *WDR11* showed no association with delayed puberty with only 7/81 female and 5/68 male carriers reporting delayed puberty. Female carriers of PTVs in *FGF8* showed some evidence of later puberty (beta=1.4 years, P=3.6×10^{-3}, N=5/10 reported delayed puberty) but with no effect in males (N=1/8 reported delayed puberty) (Supplementary Figure 5). These observations highlight the lower penetrance of rare deleterious variants in large population-based studies compared to in patient cohorts[28, 29,30].

Common genetic variants influence risk of phenotypic extremes

Rare pathogenic variants such as the above are described to cause disorders of puberty. However, it remains unclear whether common genetic variants also contribute to abnormal puberty timing. To assess this, we generated a polygenic score (PGS) of AAM in a penalised regression framework using lassosum[31] and data from our meta-analysis of European ancestry cohorts but excluding UK Biobank. This PGS explained ~12% of the phenotypic variance in UK Biobank. The PGS was informative in individuals experiencing menarche as early as 8 years old and later than 20, well beyond the normal AAM range (10 to 15 years, Supplementary Figure 6, Supplementary Tables 8 and 9).

We next sought to understand how the risks of early (<10 years) and delayed (>15 years) AAM were influenced by the PGS. Women in the lowest 1% PGS centile reported AAM at mean (SE) 11.49 (0.03) years, compared to 14.46 (0.04) years in the top 1% PGS centile (Supplementary Figure 6, Supplementary Table 10). Compared to women in the 50th PGS centile, those in the top 1% PGS were 10.7 times more likely to report late AAM (OR [8.20-13.96], P=2.6x10^{-68}), while women in the lowest 1% PGS were 14.2 times more likely to report early AAM (OR [7.13-28.39], P=5.1x10^{-14}). Collectively these findings suggest that common genetic variants contribute to the risk of rare clinical disorders of extremely early (precocious) and delayed puberty.

To evaluate the predictive performance of our AAM signals, we compared these to phenotypic predictors in 3,140 female children from the ALSPAC study. The AAM signals in combination explained more variance in AAM than childhood BMI, parental BMI or mother's AAM (Supplementary Table 11). Furthermore, they had a similar ability to predict extremes of AAM (beyond 2 SD) than a multi-phenotype predictor (Supplementary Figure 7), and a combined genotype and phenotype model showed high predictive ability for early AAM (AUROC = 0.75 [95% CI 0.68-0.82]) and late AAM (AUROC = 0.85 [95% CI 0.81-0.92]).

We next tested whether carrying rare variants in the AAM ExWAS genes modifies the common polygenic influence on AAM. We saw that the effect of the common variant PGS on AAM was attenuated in the 49 unrelated carriers of DMG variants in *ZNF483* ($beta_{non\text{-}carriers}$=0.564 years/SD, SE=0.003; $beta_{carriers}$=0.084, SE=0.214; $P_{interaction}$=0.025, Figure 3, Supplementary Table 12). To confirm that this was not a reflection of reduced power due to the low number of carriers, we estimated the expected relationship for non-carriers in 10,000 random subsamples of 49 participants and found that the observed carrier effect was unlikely by chance (P=0.015, Supplementary Figure 8).

Using ENCODE ChIP-seq data[32], we found that the transcriptional targets of *ZNF483* are enriched for in the AAM GWAS (fGWAS[32]; $P=2.6\times10^{-7}$), and that greater *ZNF483* binding confers earlier AAM (SLDP[33]; Z=-4.9, $P=4.8\times10^{-7}$), which is directionally-concordant with our observed effect of rare DMG variants on later AAM. This was further corroborated by functional-domain-specific gene burden analyses, which showed a larger effect on AAM of *ZNF483* DMG variants located within zinc finger domains (beta=1.615 years, SE=0.293, $P=3.59\times10^{-8}$), rather than DMG variants outwith these domains (beta=0.816 years, SE=0.298, $P=6.2\times10^{-3}$, Figure 3). This data suggests that rare DMG variants in *ZNF483* confer later AAM by disrupting the protein's ability to bind to its multiple DNA targets.

Implicating AAM genes through variant to gene mapping approaches

To implicate putatively causal genes that underlie our 1080 common variant signals for AAM, we developed the framework 'GWAS to Genes' (G2G) that integrates genomic and functional data across six sources (Methods, Figure 1, Supplementary Tables 13-16). We identified proximal genes (within 500kb up or downstream) to the 1080 AAM signals and scored genes based on the degree of evidence linking our lead index variants to the function of these genes. To achieve this, we implicated genes by identifying signals that co-localised with a) known enhancers and regulatory elements[34], b) non-synonymous variants, c) expression quantitative trait loci (eQTL) specifically in tissues enriched for AAM associations (Supplementary Figure 9, Supplementary Table 17), and d) circulating protein QTL (pQTL) from whole blood (see Methods). In addition, we integrated gene-level associations for aggregated non-synonymous common variants using MAGMA[35] and gene scores from PoPs[36], which uses bulk human and mouse data with information on scRNA, gene pathways and

protein interactions to link genes to GWAS signals. Individual genes were further upweighted if they were the nearest gene to the signal[37,38].

Using this approach, our 1080 signals were found to be proximal to 10,341 genes, of which 660 'high-confidence AAM genes' were identified as the highest-scoring gene at a locus and with at least two lines of evidence (Supplementary Figure 10 & Supplementary Table 18; top-scoring genes at each of the 1080 loci are also listed in Supplementary Table 19). High-confidence AAM genes include established components of the HPG axis that are disrupted in rare monogenic disorders of puberty (*CADM1*, *CHD4*, *CHD7*, *FEZF1*, *GNRH1*, *KISS1*, *SPRY4*, *TAC3*, *TACR3*, *TYRO3*)[39], and other recently reported candidate genes (*PLEKHA5*, *TBX3*, *ZNF462*)[40,41]. Other AAM genes have recognised roles in sex hormone secretion and gametogenesis (*ACVR2A*, *CYP19A1*, *HSD17B7*, *INHBA*, *INHBB*, *MC3R*, *PCSK2*)[42], are disrupted in rare monogenic disorders of multiple pituitary hormone deficiency (*OTX2*, *SOX2*, *SOX3*, *SST*)[43], monogenic obesity (*BDNF*, *LEPR*, *MC4R*, *NTRK2*, *PCSK1, SH2B1*)[44] or syndromes characterised by hypogonadism (Noonan Syndrome: *BRAF*, *SOS1*; Bardet-Biedl Syndrome: *BBS4*; Prader-Willi/Angelman Syndrome: *NDN*, *SNRPN*, *UBE3A*)[45,46,47,48]. Other mechanisms implicated by high-confidence AAM genes include: insulin and insulin-like growth factor (IGF) signalling (*CALCR*, *GHR*, *IGF1R*, *INSR*, *NEUROD1*, *NSMCE2*, *PAPPA2*, *SOCS2*)[49]; thyroid hormone signalling (*THRB*)[50]; and the Polycomb silencing complex (*CBX4*, *CTBP2*, *FBRSL1*, *JARID2*, *PHC2*, *SCMH1*, *TNRC6A*)[51]. We also found strong supportive evidence for all genes identified by the exome variant associations, except *KDM5B* (Supplementary Figure 11).

Weight gain related and unrelated puberty signals

Phenotypic, genetic and mechanistic links between higher BMI and earlier AAM are well described[15], but it is challenging to distinguish whether individual AAM signals have a primary effect on puberty or weight status[9]. Here, 83 out of the 1080 AAM signals colocalised (at PP>=0.5) and also showed genome-wide significant association with adult BMI (Supplementary Table 20), and 53 further AAM signals colocalised with adult BMI and showed association with BMI at $P<4.6\times10^{-5}$ (based on 1080 tests). Of these 136 colocalising signals, at 126 the AAM-reducing allele was associated with higher adult BMI (Supplementary Table 20).

To identify AAM signals with or without a primary effect on early weight gain, we clustered the 1080 AAM signals by their associations with body weight from birth to age 8 years (before the normal age at puberty onset) in the Norwegian MoBa Cohort (N=26,681 children)[52]. We identified three trajectories: 464 AAM signals (44%) formed a 'moderate early weight gain' trajectory, and 15 (1%) formed a 'high early weight gain' trajectory; both trajectories were characterised by effects of AAM-reducing alleles on higher weight gain across early childhood. The remaining 586 (55%) AAM signals formed a 'no early weight gain' trajectory; yet in combination AAM-reducing alleles in this trajectory increased adult BMI (beta=0.487 kg/m^2/year; $P=1.6\times10^{-20}$) (Supplementary Tables 21 and 22, Figure 4, Supplementary Figure 12). This data indicates a bidirectional causal relationship between AAM and body size, with greater

early weight gain leading to earlier AAM, and also earlier AAM leading to higher adult BMI. This approach provides a clear distinction between AAM signals that have primary effects on puberty timing or early weight gain.

Pathway, tissue and cell-type enrichment of menarche-implicated genes

Genome-wide common variant AAM associations were enriched for genes expressed in several brain regions, and enrichment was highest in the hypothalamus. Outside the brain, we also observed enrichment for genes expressed in the adrenal gland (Supplementary Figure 9 & Supplementary Table 17).

We next performed gene-based pathway analyses on the 660 high-confidence AAM genes using g:Profiler[53] and identified 83 enriched biological pathways (Supplementary Table 23), which grouped into 24 clusters (Figure 4, Supplementary Figure 13, Supplementary Table 24). These included a number of neuro-endocrine, sexual development, protein and chromatin regulation pathways. To explore distinct biological pathways by early weight trajectories, we repeated the gene-based pathways analysis after stratifying the 660 high-confidence AAM genes into 'early weight gain' AAM genes (N=341) or 'no early weight gain' AAM genes (N=315) (Supplementary Table 25). Early weight gain AAM genes specifically highlighted hormone regulation, feeding behaviour, rhythmical process, AKT phosphorylation targets and peptidyl-serine modification (Supplementary Figure 14, Supplementary Table 26). Conversely, the no early weight gain AAM genes highlighted female sex differentiation, histone modification, negative regulation of transcription by RNA polymerase II, synapse organisation and DNA repair (Supplementary Figure 15, Supplementary Table 26). Head development and cellular response to stress pathways were enriched among both early weight AAM gene groups (Figure 4, Supplementary Figures 13-15, Supplementary Table 26).

To understand how AAM-associated genes may exert effects on the HPG axis, we explored their expressional dynamics in mouse embryonic GnRH neurons [manuscript in preparation]. RNAseq in GnRH neurons previously identified 2182 genes that showed differential expression between embryonic migration stages (early, intermediate or late, Figure 5), and were categorised into 23 spatio-temporal expression trajectories [manuscript in preparation]. At the genome-wide level, we observed enrichment for GWAS AAM associations among genes that become upregulated in the late (Trajectory01, $P_{adj.}=3.8\times10^{-5}$) and mid- to late-stages of GnRH neuron development (Trajectory03, $P_{adj.}=0.032$, Supplementary Table 27), i.e. when GnRH neurons have completed their migration process and start to make synaptic connections. Of the 660 high-confidence AAM genes, 28 assign to Trajectory01 ($P_{Exact}=2.3\times10^{-6}$), including *NEGR1* and *TNRC6A*, and 31 assign to Trajectory03 ($P_{Exact}=5.2\times10^{-3}$), including *KDM4C*, *PDE10A* and *TP53BP1*. Both of these GnRH expressional trajectories remained enriched when considering only the subset of non-early weight high-confidence AAM genes (Trajectory01 $P_{Exact}=5.4\times10^{-4}$; Trajectory03 $P_{Exact}=1.4\times10^{-2}$), while Trajectory01 was also enriched when considering only AAM

724 genes that influence early weight gain (Trajectory01 $P_{Exact}=9.3\times10^{-4}$; Trajectory03
725 $P_{Exact}=0.08$).

726 **GPCRs and puberty timing**

727 G protein-coupled receptors (GPCRs) regulate several endocrine processes and
728 diseases, including puberty timing[54] and are therapeutic targets. Here, 24 of the 161
729 brain-expressed GPCRs (Methods) were implicated in AAM by at least one G2G
730 predictor (Figure 6, Supplementary Table 28). These include *MC3R*, where we
731 recently reported that rare LOF variants, which impair signalling, were associated with
732 delayed puberty[15], and *GPR83* which encodes a $G\alpha_{q11}$- and Gαi-coupled GPCR widely
733 expressed in several brain regions[55,56] and is implicated in energy metabolism[57]. In
734 mice, *Gpr83* and *Mc3r* are reportedly co-expressed in key hypothalamic neurons that
735 control reproduction (KNDy neurons) and growth (GHRH neurons)[15].

736 Since dimerisation between GPCRs may affect their signalling[58], we tested for physical
737 and functional interactions between *MC3R* and *GPR83 in vitro*. Using a
738 Bioluminescence Resonance Energy Transfer (BRET)-based assay in HEK293 cells,
739 we observed a physical and specific interaction between *GPR83* and *MC3R*
740 (Supplementary Figure 16, Supplementary Table 29). We then tested whether *GPR83*
741 modifies canonical *MC3R* signalling, by measuring NDP-α-melanocyte-stimulating
742 hormone (NDP-αMSH)-stimulated cyclic AMP production in HEK293 cells following
743 transfection with plasmids encoding wild-type *GPR83* and *MC3R* separately or
744 together (1:1 ratio). *GPR83* and *MC3R* co-transfection increased cAMP production by
745 43% compared to *MC3R* alone (P=0.03, Figure 6, Supplementary Figure 16,
746 Supplementary Tables 30 and 31).

747 Consistent with this *in vitro* interaction, we observed statistical genetic epistasis
748 between the common AAM signals at *MC3R* rs3746619; a 5'UTR SNP highly
749 correlated with predicted deleterious coding variants, and *GPR83* rs592068; which
750 colocalises with eQTLs for *GPR83* in brain[59] and across tissues[60]. Among white
751 European unrelated UK Biobank female participants, *MC3R* function-increasing
752 alleles conferred increasingly earlier AAM in the presence of *GPR83* expression-
753 increasing alleles ($beta_{interaction}$=-0.034 ± 0.015 years, $P_{interaction}$=0.02, Figure 6). These
754 findings extend our previous observation that *MC3R* loss of function causes delayed
755 puberty[15], by indicating that increased *MC3R* function through enhanced *GPR83*
756 expression leads to earlier puberty timing.

757 **Joint regulation of ages at menarche and menopause**

758 Previous GWASs have estimated a modest shared genetic aetiology between AAM
759 and age at natural menopause (ANM) (genome-wide genetic correlation: r_g=0.14;
760 P=0.003)[61]. ANM gene candidates are mainly expressed in the ovary and implicate
761 DNA damage sensing and repair (DDR) processes that maintain genome stability and
762 hence preserve the ovarian primordial follicle pool[62].

763 Of the 1080 AAM signals, nine colocalised (at PP>=0.5) and showed genome-wide
764 significant ($P<5\times10^{-8}$) association with ANM, and a further 11 AAM signals colocalised

and showed association with ANM at $P<4.6\times10^{-5}$ (=0.05/1080, Supplementary Table 32). We also considered if ANM signals influence AAM. Of the 290 previously reported ANM signals[62], 21 colocalised and showed association with AAM at $P<1.7\times10^{-4}$ (=0.05/290), 13 of which were additional to the above AAM signals (Supplementary Table 33). Consistent with the phenotypic association between AAM and ANM[63], most of the shared common signals (25/33) showed directionally-concordant effects on AAM and ANM (shifting reproductive lifespan earlier or later). Several of these shared signals map to components of the HPG axis, including *GNRH1*, *INHBB* and *FSHB* (lead SNP rs11031006), which has previously reported associations with related reproductive phenotypes[8,64,65].

Several other shared AAM and ANM signals map to genes that encode components of DDR processes (*CHD4*, *CHEK2*, *DEPTOR*, *E2F1*, *MSH6*, *MSI2*, *PPARG*, *RAD18*, *RAD51*, *RAD52*, *SCAI*, *SPRY4*, *SUMO1*, *TP53BP1*, *TRIP12*, and *WWOX,* see Supplementary Table 34)**,** central to the establishment and maintenance of ovarian oocyte numbers[62], and not previously implicated in puberty timing. A notable example is rs746979919 ($P_{AAM}=1.5\times10^{-20}$; $P_{ANM}=1.5\times10^{-34}$), which is intronic in *MSH6*, a DNA mismatch repair gene that is primarily expressed in peripheral reproductive tissues, such as ovary and uterus (Supplementary Table 35). Furthermore, the co-localised ANM signal at *CHEK2* captures the previously described frameshift variant 1100delC[66]. This association was further supported by the exome data, with the 347 women carrying rare *CHEK2* protein truncating variants (excluding 1100delC) reporting on average 2 months later AAM (SE=0.99, P=0.04). *CHEK2* encodes a cell cycle checkpoint inhibitor that plays a crucial role in culling oocytes with unrepaired DNA damage[62].

Three of the shared AAM and ANM signals that map to DDR genes were assigned to the 'moderate early weight gain' trajectory and further colocalised with adult BMI (*RAD52*: $P_{BMI}=4.8\times10^{-22}$, *TP53BP1*: $P_{BMI}=1.2\times10^{-5}$ and *TRIP12*: $P_{BMI}=2.9\times10^{-13}$), suggesting that some DDR genes might influence AAM via early weight gain (Supplementary Table 20). Other shared AAM and ANM signals that map to DDR genes were assigned to the 'no early weight gain' trajectory (*CHD4*, *MSH6*, *SCAI* and *SUMO1*) and/or showed no association (P>0.05) with adult BMI (*CHEK2*, *MSI2*, *PPARG*, and *WWOX*).

Summary and conclusions

The GWAS signals identified by this expanded multi-ancestry GWAS double the variance explained in AAM compared to previous findings[9]. Furthermore, the common variant PGS contributes substantially to risks of extremely early and late puberty timing. Future studies should explore the potential of this PGS to predict extreme disorders of puberty timing, in contrast to the effects of known monogenic causes.

We describe the first systematic characterisation of common genetic determinants of both ends of reproductive lifespan, AAM and ANM. The 33 identified shared signals highlight the concordant effects of HPG axis genes on both AAM and ANM, and also the influence of ovary-expressed genes involved in DNA damage response. DDR processes have been well described to regulate ovarian oocyte numbers throughout life[62], but have not previously been implicated in puberty timing. When stratified by their effects on early childhood weight, DDR pathways were enriched among AAM genes that do not show a primary effect on early weight gain. Our findings suggest that the ovarian reserve, established during early fetal development, might signal centrally to influence the timing of puberty.

We address the considerable challenge of deriving biological insights from common variant signals[67] by developing G2G (GWAS to Genes), an analytical pipeline that integrates a variety of data sources to enable gene prioritisation. While comprehensive experimental validation of G2G is infeasible, its utility is supported by the prioritisation from GWAS AAM data of many genes with known involvement in sex hormone regulation and rare monogenic or syndromic disorders of puberty, obesity and hormone function. The validity of G2G prioritised genes is also supported by evidence for their enrichment for dynamic expression in GnRH neurons during their late stage of embryonic migration, when they begin their integration into the hypothalamic neural network controlling puberty[68]. Furthermore, we provide experimental support for one novel high-scoring AAM gene, *GPR83*, which is co-expressed with, interacts with, and enhances *MC3R* function. Future studies should explore further the emerging role of brain expressed GPCRs in linking central nutritional sensing to reproductive function.

Finally, we provide one of the few examples to date of epistatic interaction between common and rare genetic variants. Linked to puberty timing by both common and rare variants, the transcription factor *ZNF483* has diverse binding sites across the genome. We infer that greater *ZNF483* binding promotes earlier AAM, whereas rare deleterious variants in *ZNF483* appear to abolish the influence of common genetic influence on puberty timing.

Together, these insights shed light on mechanisms, including early weight gain and adiposity, hormone secretion and response, and cellular susceptibility to DNA damage, that potentially explain the widely reported relationships between earlier puberty timing and higher risks of later life mortality, metabolic disease, and cancer.

Acknowledgements

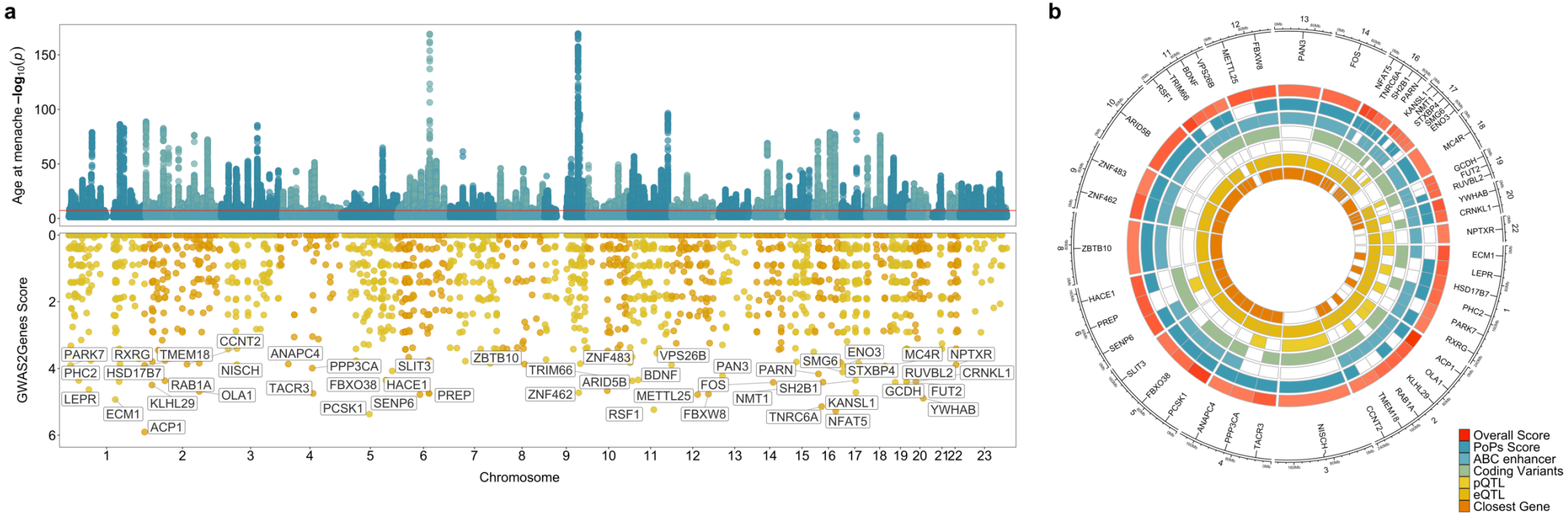
This research has been conducted using the UK Biobank Resource under Application Number 9905. Other study specific acknowledgements can be found in the supplementary note.

Data availability

Cohorts should be contacted individually for access to their raw data. UK Biobank data are available on application (https://www.ukbiobank.ac.uk/enable-your-research/register). Summary statistics from the meta-analysis excluding 23andMe will be made available upon publication. Access to the full summary statistics including 23andMe results, can be obtained from 23andMe after completion of a Data Transfer Agreement (https://research.23andme.com/dataset-access/).

Competing interests

J.R.B.P. and E.J.G. are employed by Adrestia Therapeutics. D.J.T. is employed by Genomics PLC. D.L.C. and J.P.B. are employed by GSK. E.T. is employed by Pfizer. D.A.L. has received support from Roche Disgnostics and Medtroic Ltd for work unrelated to the research in this paper. T.D.S. is co-founder and stakeholder of Zoe Global Ltd. P.A.F. conducts research funded by Amgen, Novartis and Pfizer, he received Honoraria from Roche, Novartis and Pfizer. M.W.B. conducts research funded by Amgen, Novartis and Pfizer.



856

857 **Figure 1 | Age at menarche GWAS and gene prioritisation.** (a) Miami plot showing signals from the European meta-analysis for
858 age at menarche (upper panel) and genome-wide G2G scores with names of the top 50 genes annotated (lower panel). The upper
859 panel Y-axis is capped at $-\log_{10}(1\times10^{-150})$ for visibility. (b) The 50 top scoring genes implicated by G2G, annotated by their sources
860 of evidence. Relevant data are included in Supplementary Tables 2 and 13-16.

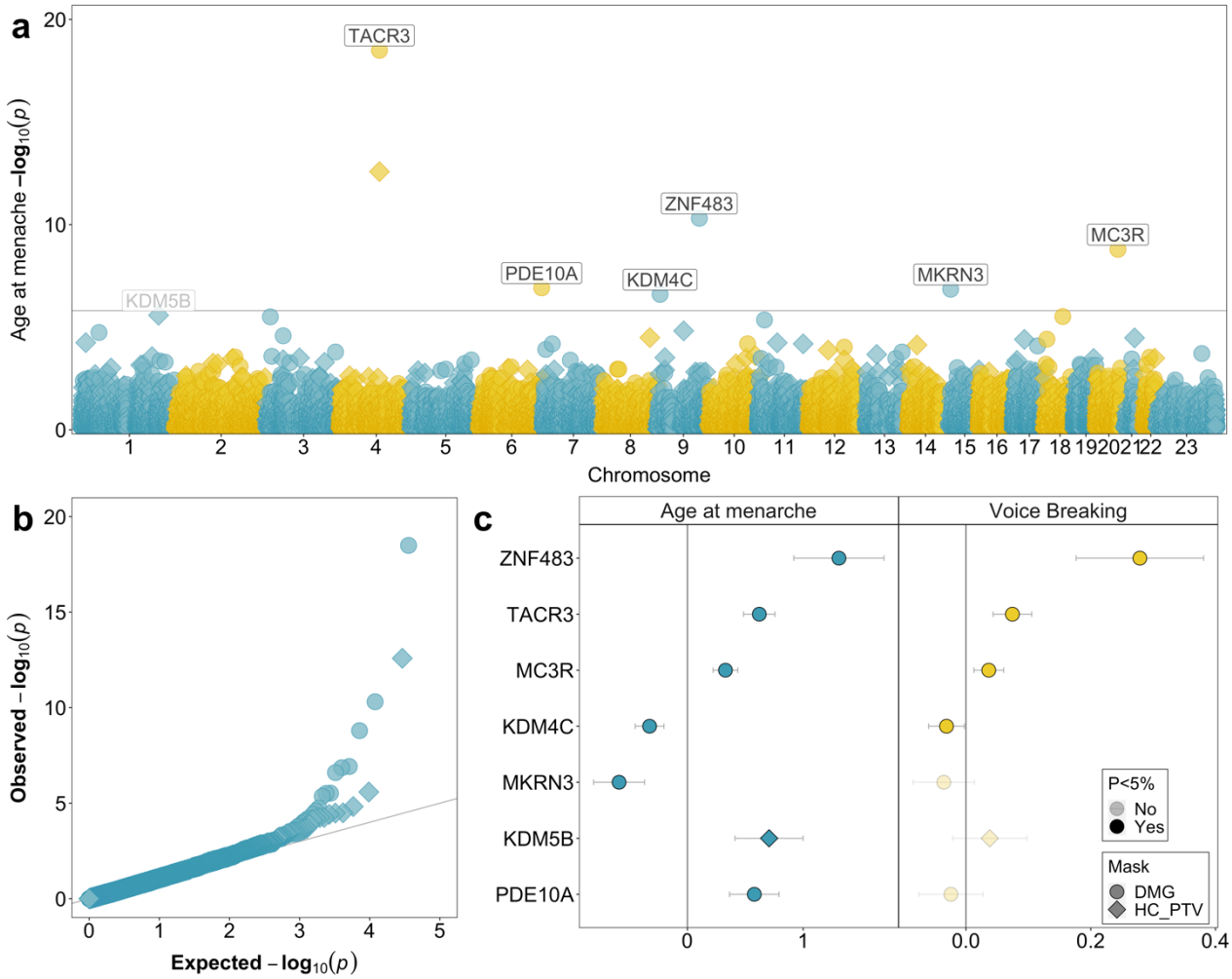


Figure 2 | Exome-wide rare (MAF <0.1%) variant associations with age at menarche. (a) Manhattan plot showing gene burden test results for age at menarche. Genes passing exome-wide significance ($P<1.54\times10^{-6}$) are highlighted; in addition, *KDM5B* shows a sub-threshold association ($P=2.6\times10^{-6}$). Point shapes indicate variant predicted functional class (DMG, damaging; HC PTV, high confidence protein truncating). (b) QQ plot for gene burden tests. (c) Comparison of gene burden associations for age at menarche (female participants, years) and age at voice breaking (men, 3 categories). Relevant data are included in Supplementary Table 5.

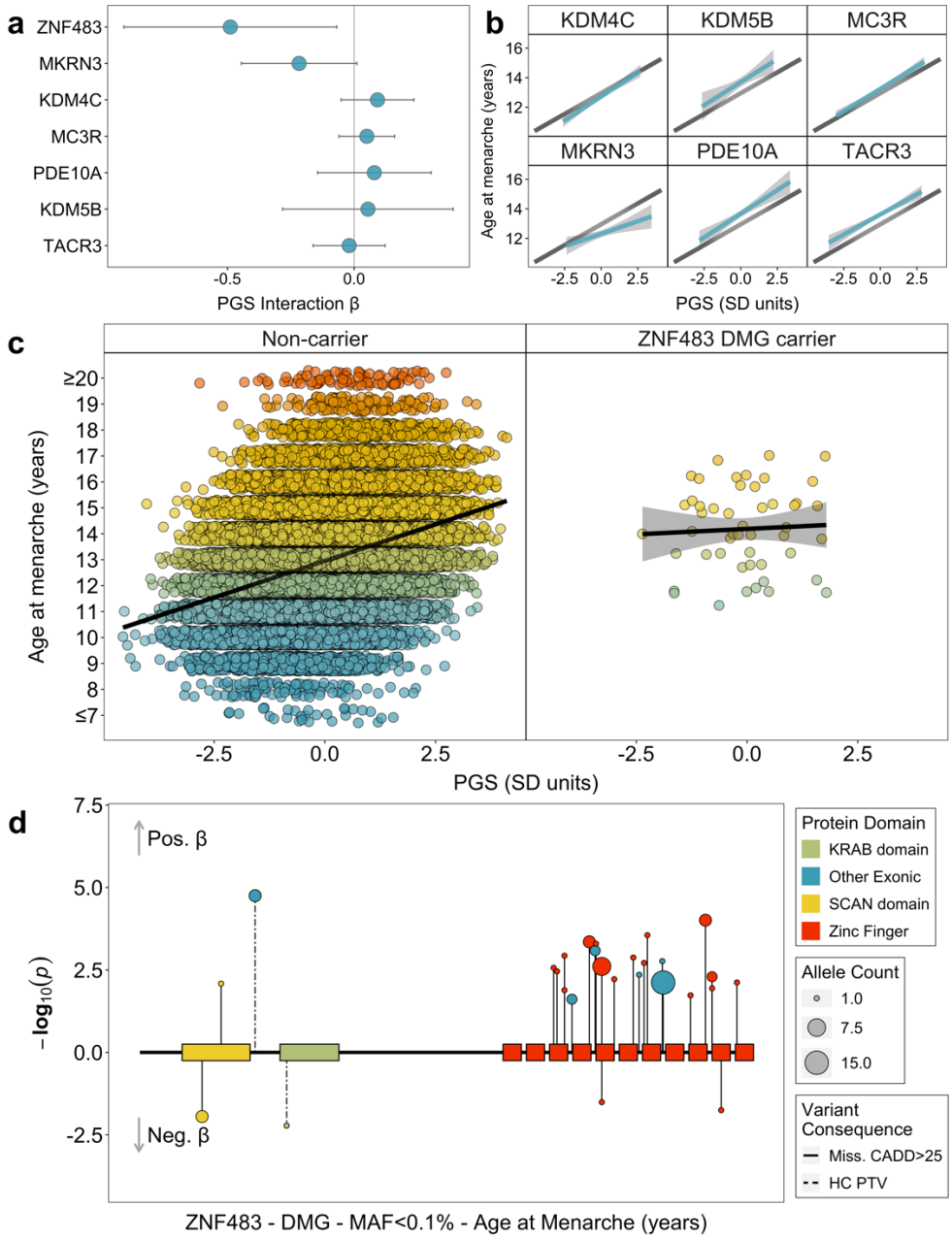


Figure 3 | Epistatic interactions between rare coding variants and common genetic susceptibility on age at menarche in UK Biobank. (**a**) Interaction effects (95% CI) on age at menarche between a GWAS polygenic score (PGS) and carriage of qualifying rare variants in seven exome-highlighted genes. Predicted mean (95% CI) age at menarche in (**b**) non-carriers (black) and carriers (light blue) of rare variants in six genes without significant interaction effects and (**c**) in non-carriers (left panel) and carriers (right) of rare variants in *ZNF483* which shows significant interaction. In (c) points show individual age at menarche values. (**d**) Plot of individual rare damaging (DMG) variant associations with age at menarche by *ZNF483* functional domains. The coding part of *ZNF483* is depicted by the horizontal black line. Included damaging variants had a minor allele frequency (MAF) <0.1% and were annotated to either be high-confidence protein truncating variants or missense variants with CADD score >=25. Relevant data are included in Supplementary Table 12.

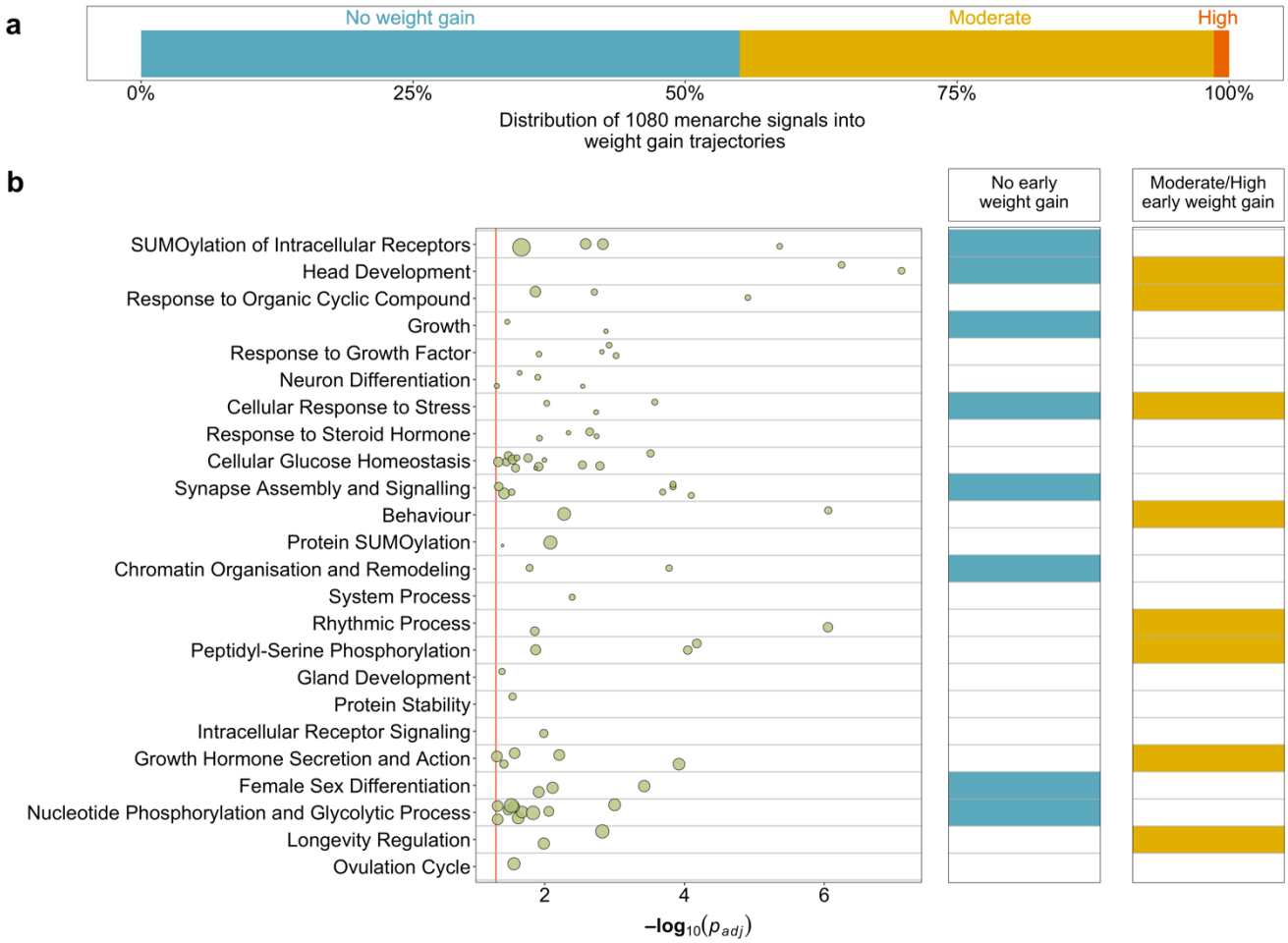


Figure 4 | Stratification of age at menarche signals and biological pathway enrichments by their influence on early childhood weight. (a) Proportion of GWAS signals for age at menarche by early childhood weight trajectory. (b) Biological pathways enriched for high confidence age at menarche genes, plus enrichment within early childhood weight trajectories. Row names describe pathway clusters. Strength of associations with individual pathways are indicated by circles. Circle size reflects the proportion of pathway genes that are high confidence age at menarche genes. The right-hand panel indicates whether each pathway cluster remains enriched for age at menarche genes when stratified by early childhood weight trajectory. Extended data are included in Supplementary Tables 21 and 23-26.

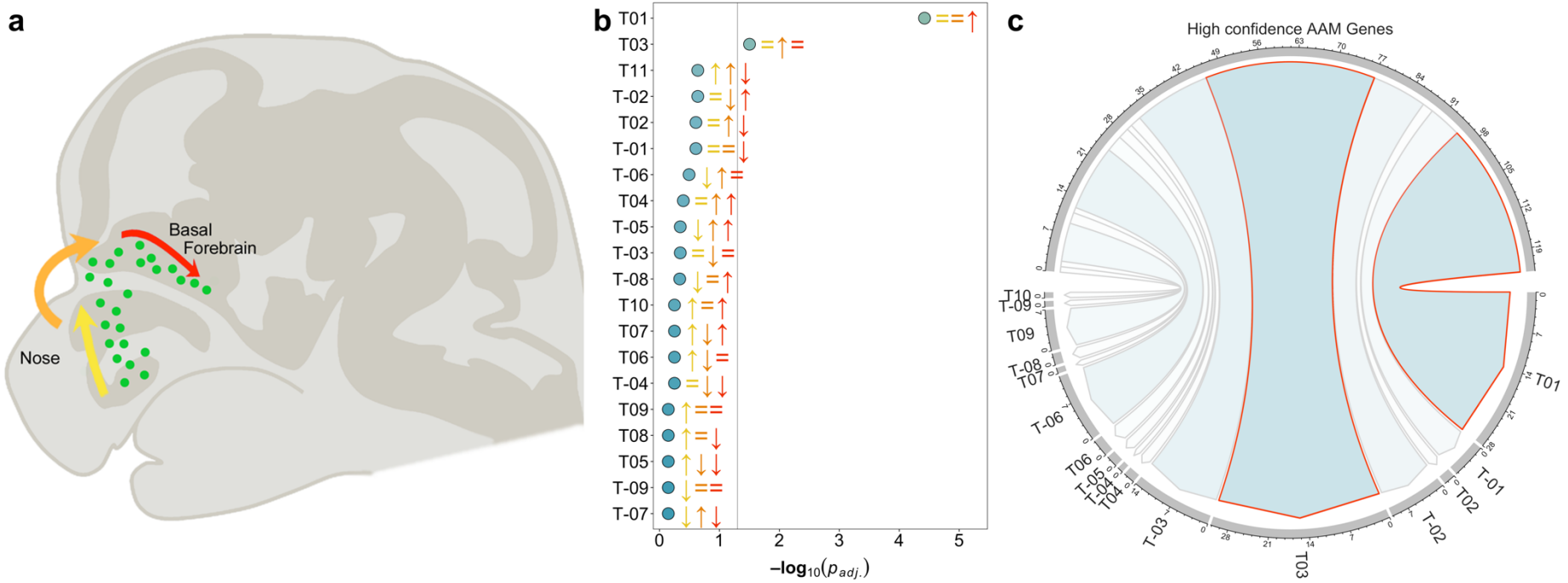


Figure 5 | Enrichment of gene drivers of GnRH migration and maturation in the age at menarche GWAS. (a) Schematic representation of the stages of GnRH neuron migration during embryonic development. Using RNAseq data, Pitteloud and colleagues [manuscript in preparation] grouped differentially-expressed genes into 23 expressional trajectories based on their comparative level of expression during the Early (yellow), Intermediate (amber) and Late (red) stages of GnRH migration. (b) Genome-wide MAGMA enrichment for age at menarche associations within each expression trajectory. (c) Trajectories significantly enriched at the genome-wide level in (b) show significant overlap with the 660 high-confidence age at menarche genes. Extended data are included in Supplementary Tables 18 and 27.

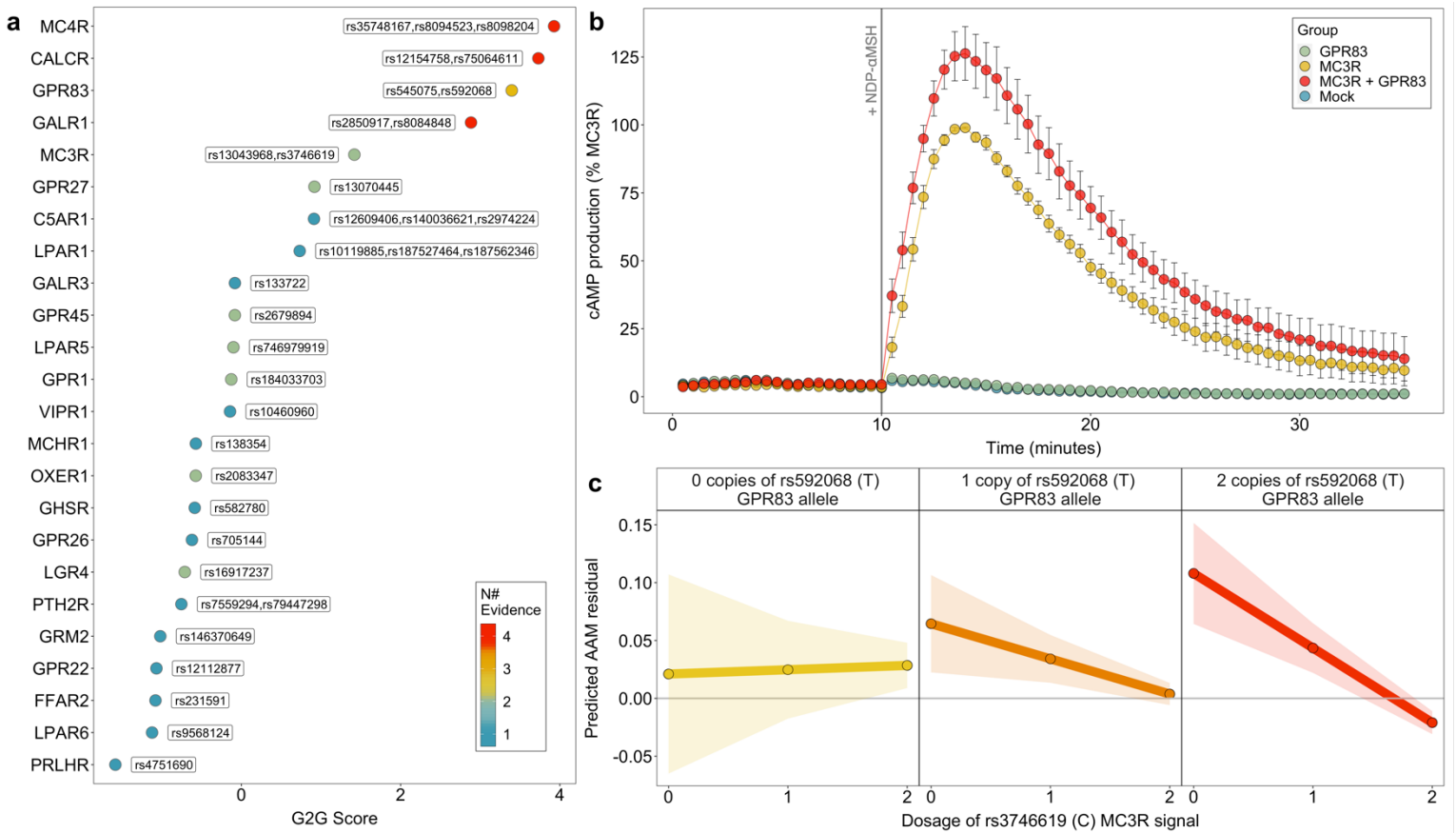


Figure 6 | Interactions between G protein-coupled receptors (GPCRs) on age at menarche. (a) 24 brain-expressed GPCRs implicated in age at menarche by G2G analysis of white European GWAS data (b) Time-resolved NDP-αMSH-stimulated cAMP production in HEK293 cells expressing *MC3R*-alone or with both *MC3R* and *GPR83*. Data are mean (standard error) % of the maximal *MC3R*-alone response (from 6 independent experiments). (c) Predicted mean (95% CI) age at menarche according to dosage of *MC3R* function-increasing C alleles at rs3746619 (X-axis in each panel) and *GPR83* expression-increasing T alleles at rs592068 (panels). $Beta_{interaction}$= -0.034 ± 0.015 years, $P_{interaction}$=0.02. Extended data are included in Supplementary Tables 28 and 30.

Methods

GWAS meta-analysis for age at menarche

Association summary statistics were collated from studies on age at menarche (AAM, predominantly recalled in adulthood) and genome-wide SNP arrays imputed to the 1000 Genomes reference panel or more recent (Supplementary Table 1). Genetic variants and individuals were filtered based on study-specific quality control metrics. In each study, genetic variants were tested for association with AAM in additive linear regression models, including as covariates: age and any study-specific variables, such as genetic principal components. Insertion and deletion polymorphisms were coded as "I" and "D" to allow harmonization across all studies. Association statistics for each SNP were then processed centrally using a standardized quality control pipeline[69]. Each variant was meta-analysed using a fixed-effects inverse-variance-weighted model using METAL[70]. This was done in two stages. First, summary statistics from studies within each stratum (i. ReproGen consortium studies, ii. reproductive cancer consortium studies, iii. East Asian studies) were meta-analysed and then filtered to include only variants present in more than half of the studies within each stratum. Second, strata-level results were meta-analysed with data from UK Biobank[71], using 'first instance' data for AAM (field 2714), and 23andMe. Initially we performed a European-only analysis (N=632,955). This combined file was filtered to include only variants present in the UK Biobank and at least one other stratum. Variants were also filtered to include minor allele frequency (MAF) >= 0.1%. We then performed a second analysis, by adding the data from the East-Asian studies, and followed the same sample filtering steps and identification of independent signals (described below).

Replication and explained variance

Independent replication of identified signals was performed in data from the Danish Blood Donors study[16] (DBDS). The DBDS includes questionnaire recalled AAM data on 35,472 European women ("Age when menstruation started?"). Mean age at recall was 38.4 years (SD=12.9 years) and mean AAM was 13.1 years (SD=1.4 years). Indirect confirmation of AAM signals was sought by association with age at voice breaking (AVB) in men in UK Biobank[17] (N=191,235 European men - data field 2385) and the 23andMe study[18] (N=55,781 European men). For signals with missing data for either AVB dataset, we identified proxies using the UK Biobank White European dataset (within 1Mb of the reported signal and R^2>0.6), choosing the variant with the highest R^2 value. Given the smaller sample sizes of these cohorts, we performed a Binomial sign test for global replication. The variance explained by each lead AAM signal in the DBDS was calculated using the formula 2×f(1−f)β2a, where f denotes the variant MAF and βa is the effect estimate in additive models. Overall variance explained was calculated as the sum of individual variants.

UK Biobank phenotype preparation

For downstream analyses in UK Biobank, we derived a AAM variable, using data from field 2714. To maximise sample size, individuals with missing or implausibly early or late 'first instance' AAM (<8 years or >19 years old) were imputed using data from the next available instance (if plausible). We also derived two binary traits to represent abnormally early (precocious) and delayed puberty. Early puberty was defined as AAM <10 years old (N=1,321). Delayed puberty was defined as AAM >15 years old (N=10,530). For comparison, women reporting AAM at 12 or 13 years were controls (N=81,950). All data analysis and visualisation were conducted in R (version 4.2.1, 2022-06-23), unless otherwise stated.

Rare variant associations with age at menarche

To identify gene-level rare variant associations with AAM, we performed an exome-wide association study analysis (ExWAS) using whole-exome sequencing (WES) data on 222,283 UK Biobank women of European genetic-ancestry[72]. WES data processing and quality control was performed as described by Gardner *et al*.[30]. Individual gene burden tests were performed by collapsing variants with MAF <0.1% per gene according to their predicted functional consequence. We defined two functional categories of rare variants i) high confidence protein truncating variants (HC PTV) annotated using VEP[73] and LOFTEE[74] and ii) damaging variants (DMG) including HC PTVs plus missense variants with CADD score[19] ≥25. We analysed a maximum of 17,885 protein-coding genes, each with at least 10 rare allele carriers in either of the two variant categories ($P<1.54\times10^{-6}$, 0.05/32,434 tests). Gene burden association tests were performed using BOLT-LMM[75]. Validity of the ExWAS analysis was indicated by the absence of significant association with the synonymous variant mask (Supplementary Figure 1) and low exome-wide inflation scores (λ-PTV =1.047 and λ-DMG =1.047). Where applicable, protein domains were annotated using information from UniProt[76] and domain-level burden tests were then performed using linear models.

Rare variant associations with other traits

We assessed the associations of any ExWAS AAM-associated genes in UK Biobank with a range of related phenotypes: age at natural menopause (based on field 3581), body mass index (BMI, field 21001), comparative body size age 10 (field 1687), adult height (field 50), comparative height age 10 (field 1697), and circulating IGF-1 concentrations (field 30770). We considered only the top AAM-associated variant mask for each gene. We also performed a similar look-up of these genes across a broader range of phenotypes using the AstraZeneca Portal[77].

Rare variants in IHH panel app genes

We selected high evidence ("green") genes with an established monoallelic/X-linked mode of inheritance from the routine clinical investigation Genomics England gene panel for idiopathic hypogonadotropic hypogonadism (IHH). At the time of the study, this included four genes, *ANOS1*, *CHD7*, *FGF8* and *WDR11*. We performed a look-up of these genes in the UK Biobank WES data for AAM (N=222,283) and VB (N=178,625) considering only HC PTVs with MAF <0.1%. We also extracted the phenotype of individual carriers. As in the ExWAS analysis, normal pubertal timing was defined in women as AAM between 10-15 years of age[1] and in men as AVB at an "about average age" (UK Biobank data field 2385).

Polygenic score calculation

We calculated a genome-wide polygenic score (PGS) for AAM using lassosum[31]. To keep PGS generation independent of PGS testing, we generated the PGS using our European-ancestry GWAS data excluding UK Biobank. We randomly selected 25,000 unrelated Europeans in UK Biobank to generate the LD (Linkage Disequilibrium) reference. The resulting PGS was standardised, by subtracting the mean and dividing by the standard deviation.

We divided the PGS into 100 centiles, and calculated the mean AAM for each PGS centile, as well as PGS centile-specific odds ratios for precocious or delayed AAM (as defined above) compared to individuals in the 50th centile of the PGS. We also calculated the mean PGS for each completed whole year of AAM.

We next tested whether carriage of ExWAS AAM-associated rare variants modifies the influence of the PGS on AAM. We performed linear models that included interaction terms [PGS × rare variant carrier status] in the subsample of unrelated white-European UK Biobank women with WES, PGS and AAM data (N=187,941). To test for chance effects due to low sample size, we randomly subsampled non-carriers to a sample size equivalent to that of carriers and compared this distribution of AAM to that observed in carriers.

A PGS comprising the 882 available lead AAM SNPs or their proxies (out of the 935 independent AAM signals from the EUR-MA) was computed in 3,140 girls with available imputed GWAS data from the ALSPAC study. Linear or logistic regression models for continuous AAM, early AAM (-2 SDs, corresponding to <10.38 years), and delayed AAM (+2 SDs, corresponding to >14.95 years) were tested, controlling for the first 20 genetic PCs. Other models assessed the predictive performance of BMI at age 8 years, and mother's AAM; finally a model including all predictors as covariables was calculated. The predictive performance of each model was evaluated by the R^2 metric for continuous AAM and by the area under the receiver operating characteristic curve (AUROC) for binary AAM outcomes.

GWAS to Genes (G2G) pipeline

Mapping GWAS signals to genes

To perform signal selection, GWAS AAM summary statistics were filtered to remove variants with MAF <0.1%. The remaining variants were merged with allele information from UK Biobank to provide the genomic sequence for any missing alleles. Genome-wide significant signals (P<5x10^{-8}) were selected initially based on proximity (in 1Mb windows). Secondary signals within these windows were then identified using approximate conditional analysis (GCTA[78]), using an LD reference panel derived from 25,000 randomly selected UK Biobank participants. Secondary signals were defined as uncorrelated (R^2<0.05) with another signal and without an overt change in their AAM association between baseline and conditional models (change in beta <20% or change in P-value by less than four orders of magnitude). Primary and secondary AAM signals were further checked for pairwise LD within 10 Mb windows using plink (v1.90b6.18)[79] and only independent signals (R^2<0.05) were retained, prioritising distance-based signals in the case of linkage. Signal selection was performed first using the European-ancestry GWAS meta-analysis, and then supplemented by any signals identified by the All-ancestry GWAS meta-analysis that were uncorrelated (R^2<0.05) with any European-ancestry signal.

Independent GWAS AAM signals were examined for proximal genes, defined as those within 500kb up- or downstream of the genes start or end sites, using the NCBI RefSeq gene map for GRCh37 (via http://hgdownload.soe.ucsc.edu/goldenPath/hg19/database/).

Colocalization with expression or protein QTL data

Tissue enrichment for GWAS associations was performed using LD score regression applied to tissue-specific expression (LDSC-SEG)[80] and tissue-specific annotations from GTEx, accessed via https://github.com/bulik/ldsc/wiki/Cell-type-specific-analyses. Significantly enriched tissues (P<0.05) were then included in colocalization analyses with the tissue-specific and cross-tissue meta-analysed GTEx eQTL data (V7[60], available via https://gtexportal.org and using the fixed-effects summary statistics for the latter), in addition to data from the eQTLGen[81] and Brain-eMeta[59] studies.

Including genomic variants with at least suggestive association with AAM (GWAS P<5x10^{-5}), we applied Summary data-based Mendelian Randomization and Heterogeneity in Independent Instruments (SMR-HEIDI, version 0.68[82]) and the Approximate Bayes Factor (ABF) method in the R package “coloc” (version 5.1.0[83]). For the former, we considered gene expression to be influenced by the same GWAS AAM variant if the FDR-corrected SMR test P<0.05 and HEIDI test P>0.001. For the latter, genomic regions were defined as ±500 kb around each gene and loci exhibiting a H4 posterior probability >0.75 were considered to show evidence of colocalization.

We also tested for colocalization between GWAS AAM variants and pQTLs using data from the Fenland study[84] and using the same procedure as above. It is important to note that colocalization analysis cannot determine causal relationships or the direction of causality between the two phenotypes.

Mapping GWAS signals to enhancers and coding variants

For genes proximal to (within 500 kb) GWAS AAM signals, we calculated genomic windows of high LD (R^2>0.80) around each signal and mapped these to the locations of known enhancers for the genes, using activity-by-contact (ABC) enhancer maps[34]. This was done across the 131 available cell/tissue types and genes were matched to enhancers only in the tissues/cells where they were actively expressed.

We also checked whether GWAS AAM signals were in LD (R^2>0.80) with any coding variants within the paired genes and what the predicted consequence of those coding variants, using SIFT[85] and POLYPHEN[86].

Gene-level GWAS associations with AAM

We performed a gene-level MAGMA analysis[35], which collapses common GWAS variants within each gene and calculates aggregate gene-level associations with the outcome trait, as described by de Leeuw et al.[35]. To enhance validity of this approach, we restricted the analysis to include only coding variants. Genes with FDR-corrected MAGMA P<0.05 were considered associated with AAM.

Finally, we used the Polygenic Priority Score (PoPS[36]), which is a similarity-based gene prioritisation method and uses cell-type specific gene expression, biological pathways, and protein-protein interactions to prioritise likely causal genes from GWAS data. At each locus, the gene with the numerically highest PoPS score was determine to be the PoPS-prioritised gene.

Calculation of G2G scores

From the above analyses, gene-level results were scored for each of the six sources as follows:

1. **Closest gene.** Gene proximity to a GWAS signal is a good predictor of causality[37]. The genes closest to each AAM signal (if also within 500 kb) were assigned. All genes with an intragenic signal were assigned as closest. Closest genes were scored 1.5 points.
2. **eQTL colocalisation.** Genes with evidence of eQTL colocalisation via both SMR-HEIDI and coloc were scored 1.5 points. Genes with evidence of colocalisation via only one of these received 1.0 point. A further (1.0) point was assigned to genes if

the most likely shared causal variant between eQTL and GWAS AAM was independent of the proximal GWAS signal (R^2<0.05).

3. **pQTL colocalisation.** The same scoring as in (2) was applied to pQTL analyses.
4. **Coding variants.** As the evidence was overlapping for coding variant gene-level MAGMA analysis and signals correlated with coding variants, these analyses were scored concomitantly. Genes with an FDR-corrected MAGMA *P*<0.05 were scored 0.5 points. Genes containing coding variants of deleterious or damaging predicted consequence in LD with GWAS AAM signals were scored 1.0 point, or only 0.5 points if the coding variants were predicted to be benign or tolerated.
5. **ABC enhancers.** Genes targeted by enhancers which overlapped with or were correlated with GWAS AAM signals were scored 1.0 point.
6. **PoPS.** PoPs prioritised genes at each locus were scored 1.5 points.

G2G scores for each gene-signal pair were calculated as the sum of scores from these six sources. Genes that scored >0 points and were located within 500 kb of a GWAS AAM signal were considered further. To account for confounding due to large LD blocks, G2G scores were adjusted for signal LD window size (defined as the genomic distance containing variants with pairwise R^2>0.50 with the lead SNP) using linear regression models.

For genes proximal to more than one GWAS AAM signal (and hence with multiple G2G scores), the signal with the most concordant sources for that gene (highest residual G2G score) was retained and a further (1.0) point was added to reflect evidence from multiple signals. This resulted in a unique summarised G2G score for each included gene. To account for confounding due to gene size, G2G scores were further adjusted for gene length using linear regression models. The resulting residuals were considered to be the final G2G scores.

To prioritise likely causal AAM genes, all G2G scored genes (i.e., highlighted as potentially causal by at least one source) were ranked and also allocated a G2G centile position. In addition, the number of concordant predictors (sources) for each gene were noted (range: 1 to 6 sources). Finally, to reflect uncertainty due to multiple high scoring genes for the same signal, genes were flagged if they were proximal (within 1 Mb) to other genes with a similar G2G score (within 0.5 points or greater, and highlighted by at least the same number of sources).

High-confidence AAM genes

Independent GWAS signals from the All- and the European-ancestry meta-analyses were annotated with their top G2G scoring gene, using corresponding GWAS data (i.e., European analysis signals were annotated with genes from the European G2G, etc.). Genes implicated by at least two concordant sources were considered to be high-confidence AAM genes.

High-confidence AAM genes were functionally annotated using STRING[87]. Links to rare monogenic disorders were annotated from the Online Mendelian Inheritance in Man (OMIM) database (via Online Mendelian Inheritance in Man, OMIM®. McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University (Baltimore, MD), accesed November 2022. World Wide Web URL: https://omim.org/). Finally, we used GTEx, a publicly available resource of tissue-specific gene expression, to lookup the tissue expression of 1080 AAM genes highlighted by G2G[60].

ZNF483 genome-wide binding analysis

We used fGWAS (v.0.3.6[32]), a hierarchical model for joint analysis of GWAS and genomic annotations, to test for enrichment of GWAS AAM signals among ZNF483 transcription factor binding sites. fGWAS models a maximum likelihood parameter estimate for enrichment of a transcription factor (in this case ZNF483). To perform this, we annotated the European-ancestry GWAS AAM summary statistics with the ZNF483 binding sites from the ENCODE ChIP-seq data derived from human HepG2 cell line (ENCSR436PIH).

We also used Signed LD Profile regression (SLDP, https://github.com/yakirr/sldp, Reshef *et al*. [33]) to explore the directional effect of ZNF483 function on AAM. We tested whether alleles that are predicted to increase the binding of ZNF483 have a combined tendency to increase or decrease AAM. SLDP requires signed LD profiles for ZNF483 binding, a signed background model and reference panel in a SLDP compatible format. We used a 1000 Genomes Phase 3 European reference panel containing approximately 10M SNPs and 500 individuals.

Clustering of AAM signals by early childhood body weight

We analysed repeated measurements of early childhood body weight from the MoBa cohort study[52,88] to investigate the relationship between early growth and puberty timing. Childhood body weight values were extracted from the study questionnaires for 12 different time-points from birth to age 8 years using previously reported exclusion criteria[52]. Weight values were standardized and adjusted for sex and gestational age using the generalized additive model for location, scale and shape (GAMLSS; v5.1-7, via www.gamlss.com) in R (v3.6.1) as previously reported[52] with the exception that a Box-Cox t distribution was used to standardize body weight values (instead of the log-normal distribution used for BMI)[52]. GWAS for these traits was performed using BOLT-LMM (v2.3.4) as previously reported[52].

We performed Mendelian randomisation (MR) analyses to assess the likely causal effects of AAM on childhood weight at each time-point[89]. As instrumental variables (IVs), we used all 1080 AAM-associated lead SNPs individually. As outcome data, we used childhood weight at the 12 time-points. For SNPs with missing outcome data, we

identified proxies within 1 Mb and R^2>0.6, choosing the variant with the highest R^2 value, using a random selection of 25,000 unrelated European-ancestry UK Biobank individuals for the LD reference. Genotypes at all variants were aligned to the AAM-increasing allele. We used inverse-variance weighted (IVW) MR models, as this has the greatest statistical power[90].

Next, we stratified the 1080 AAM lead SNPs by their effects on early childhood weight used a k-means clustering approach for longitudinal data[91]. We performed five different models with k-means for k∈{2,3,4,5,6} clusters 20 times each. To find the optimal partition, we used the "nearlyAll" option which uses several different initialisation methods in alternation. As the assumption of homoscedasticity was not met, we used the Carolinski-Harabatz criterion, a non-parametric quality criterion, to derive the optimal number of clusters.

We then performed additional MR analyses, combining AAM signals within each identified cluster as IVs and, as the outcomes, childhood weight at each time-point and also adult BMI (on N=450,706 UK Biobank participants). We grouped together 'high early weight' and 'moderate early weight' AAM SNPs into a single IV to maximise power.

Biological pathway enrichment analysis

We performed gene-centric biological pathway enrichment analysis using g:Profiler (via the R client "gprofiler2", version 0.2.1[53]). We used a filtered set of GO pathways (accessed on the 21/02/2023), focusing on GO:BP, KEGG and REACTOME, and restricted the analysis to those pathways with 1000 genes or fewer, reasoning that these are more biologically specific. Pathway enrichment analyses were performed using the set of 660 high-confidence AAM genes, and repeated when stratified by their effects on early childhood weight (see above). Pathways with Bonferroni corrected P<0.05 were considered to be associated with AAM.

As the pathways derived from overlapping sources, we clustered the AAM-associated pathways to aid interpretation. Clustering was based on shared AAM genes across pathways. We used a "complete" clustering algorithm and a custom distance calculated as [one minus the proportion of the overlap between any two pathways relative to the pathway with the smaller overlap]. Thus, between two pathways a value of 0 indicates that all the shared AAM genes in the pathway with fewer genes are also enriched in the other pathway. To define clusters, we chose an arbitrary overlap value of 0.5, which indicates that pathways in the same cluster share 50% or more of their AAM genes.

Each pathway cluster was annotated by i) the pathway with the most significant enrichment, ii) the pathway with the highest proportion of AAM genes, iii) biological coherence of the pathways, and iv) shared genes common to all included pathways.

We considered that pathways were overlapping between the total AAM gene set and the two early-weight subgroups if there were common pathways across either i) the most significant pathway or ii) the pathway with the highest proportion of AAM-associated genes.

Expression of AAM genes in GnRH neurons

We tested for enrichment of AAM-associated genes in RNAseq data from embryonic GnRH mouse neurons [manuscript in preparation]. All expressed genes were sorted into different expressional trajectories, based on shared dynamic expression profiles across three developmental stages (early, intermediate or late), as described by Pitteloud et al. [manuscript in preparation]. We tested for enrichment of AAM-associated genes (from our European-ancestry GWAS meta-analysis) in any identified trajectory, using MAGMA[35] with custom pathways. As a sensitivity test, we used Fisher's Exact test to confirm over-representation of AAM-associated genes within each trajectory.

Colocalization of AAM signals with BMI and menopause

To explore the shared genetic architecture between AAM, age at natural menopause (ANM) and adult BMI, we performed a colocalization analysis for each of the 1080 AAM signals. ANM GWAS summary statistics were from reported ReproGen data on ~250,000 women of European ancestry[62]. Adult BMI GWAS summary statistics were derived from 450,706 individuals in UK Biobank. For AAM signals with missing outcome GWAS data, we identified proxies within 1 Mb and with an R^2>0.6 using our 25,000 participant UK Biobank LD reference. We applied both Bonferroni correction ($P \leq 0.05/1080 = 4.6 \times 10^{-5}$) for association with the outcome trait, and a posterior probability (PP) of colocalization PP>0.5.

The same approach was applied in the opposite direction, by testing ANM signals identified in the most recent ReproGen GWAS[62] for association with AAM. ANM signals were highlighted if they passed Bonferroni correction ($P \leq 0.05/290 = 1.7 \times 10^{-4}$) for association with AAM. As ANM signals are well-established to be enriched for DNA damage repair genes (DDR), we built a comprehensive list of DDR genes, integrating five different sources: i) an expert curated DDR gene list ("Broad DDR") from the laboratory of Professor Stephen Jackson, this list encompasses a range of related pathways: DNA repair genes, broader DNA damage response genes (such as damage-induced chromatin remodelling, transcription regulation or cell cycle checkpoint induction); and general maintenance of genome stability (such as genes involved in DNA replication); ii) a second expert curated list previously reported[62], assembled by John Perry, Eva Hoffmann and Anna Murray; iii) genes listed in the REACTOME[92] "DNA Repair" pathway (R-HSA-73894); iv) genes listed in the Gene

Ontology "DNA Repair" pathway (GO:0006281); and v) genes listed in the Gene Ontology[93] "Cellular response to DNA damage stimulus" (GO:0006974).

GPR83-MC3R interaction

Brain expressed GPCRs

We tested whether any brain-expressed G-protein coupled receptors (GPCRs) were implicated by GWAS AAM associations (G2G gene scores). We tested a curated list of brain-expressed GPCRs (Stephen O'Rahilly, personal communication): ACKR1, ACKR2, ACKR3, ACKR4, ADRB1, ADRB2, ADRB3, AGTR1, AGTR2, BRS3, C5AR1, C5AR2, CALCR, CASR, CCKAR, CCR1, CCR10, CCR2, CCR3, CCR4, CCR5, CCR6, CCR7, CCR9, CCRL2, CNR1, CNR2, CXCR1, CXCR2, CXCR3, CXCR4, CXCR6, DRD1, DRD2, DRD3, DRD4, DRD5, EDNRA, EDNRB, FFAR1, FFAR2, FFAR3, FFAR4, FPR1, FPR2, FPR3, FSHR, GALR1, GALR2, GALR3, GHRHR, GHSR, GIPR, GLP1R, GLP2R, GNRHR, GPER1, GPR1, GPR12, GPR15, GPR17, GPR18, GPR19, GPR20, GPR22, GPR25, GPR26, GPR27, GPR3, GPR34, GPR35, GPR37, GPR39, GPR4, GPR42, GPR45, GPR52, GPR55, GPR6, GPR61, GPR62, GPR63, GPR75, GPR78, GPR82, GPR83, GPR84, GPR85, GPR87, GPR88, GRM1, GRM2, GRM3, GRM4, GRM5, GRM6, GRM7, GRM8, GRPR, HCAR1, HCAR2, HCAR3, HRH1, HRH2, HRH3, HRH4, LGR4, LGR5, LGR6, LPAR1, LPAR2, LPAR3, LPAR4, LPAR5, LPAR6, MC3R, MC4R, MC5R, MCHR1, MCHR2, NMBR, NMUR1, NMUR2, NPSR1, NPY1R, NPY2R, NPY4R, NPY5R, NPY6R, OXER1, OXGR1, P2RY1, P2RY2, P2RY4, P2RY6, P2RY8, PRLHR, PTAFR, PTH1R, PTH2R, QRFPR, RGR, RXFP1, RXFP2, S1PR1, S1PR2, S1PR3, S1PR4, S1PR5, SCTR, SSR1, SSR2, SSR3, SSR4, TSHR, VIPR1, VIPR2, VN1R1, VN1R2, VN1R5 and XCR1. For any GPCR scored by our G2G AAM pipeline, colocalisation was tested between GWAS signals for AAM and adult BMI (colocalisation methods as described above).

Cell culture and transfection

To investigate the effect of *GPR83* of MC3R signalling, we performed *in vitro* assays in transiently transfected HEK293 cells maintained in Dulbecco's modified eagle medium (high glucose DMEM, GIBCO, 41965) supplemented with 10% fetal bovine serum (GIBCO, 10270), 1% GlutaMAX™ (100X) (GIBCO, 35050), and 100 units/mL penicillin and 100 mg/mL streptomycin (Sigma-Aldrich, P0781). Cells were incubated at 37°C in humidified air containing 5% CO2 and transfections were performed using Lipofectamine 2000 (GIBCO, 11668) in serum-free Opti-MEM I medium (GIBCO, 31985), according to the manufacturer's protocols. The plasmids used encode the C-FLAG-tagged human *GPR83* WT (NM_016540.4) or N-FLAG-tagged human MC3R WT (NM_019888.3) ligated into pcDNA3.1(+) (Invitrogen).

Bioluminescence Resonance Energy Transfer (BRET) to measure dimerization

Heterodimerization between *GPR83* and *MC3R* was quantified using BRET1 in titration configuration. Briefly, 12,000 HEK293 cells seeded in 96-well plates were transfected with a constant dose of MC3R-RlucII plasmid (0.5 ng/well) and increasing doses of *GPR83*-Venus plasmids, or soluble (s) Venus as negative control. All conditions were topped up with empty vector (pcDNA3.1 (+)) to a total of 100 ng plasmid/well. Twenty-four hours post transfection, cells were washed once with Thyrode's buffer and total Venus fluorescence was measured in a Spark 10M Microplate reader (Tecan) using monochromators (excitation 485 ± 20 nm, emission 535 ± 20 nm). BRET was quantified 10 minutes after the addition of coelenterazine H (NanoLight Technology, 2.5 mM). netBRET was calculated as [(absorbance at 533 ± 25 nm/absorbance at 480 ± 40 nm)] – [background (absorbance at 533 ± 25 nm/absorbance at 480 ± 40 nm)], with the background corresponding to the signal in cells expressing the RlucII protomer alone under similar conditions. Data on the X-axis represent the ratio between acceptor (Venus) fluorescence and donor (RlucII) luminescence. Representative data are from four independent experiments.

Time-resolved cAMP assay

Measurement of ligand-induced cAMP generation in HEK293 cells transiently expressing either *MC3R* or both *MC3R* and *GPR83* was performed using GloSensor™ cAMP biosensor (Promega), according to manufacturer's protocol. Briefly, 12,000 cells were seeded in white 96-well poly-D-lysine-coated plates. After 24 hours, cells were transfected with both 100 ng/well of pGloSensorTM-20F cAMP plasmid (Promega, E1171) and 30 ng/well of each plasmid encoding either MC3R or MC3R and GPR83, using Lipofectamine 2000 (GIBCO, 11668). All conditions were topped up with empty vector (pcDNA3.1 (+)) to a total of 160 ng plasmid/well. The day after transfection, cell media were replaced by 90 mL of fresh DMEM with 2% v/v GloSensorTM cAMP Reagent (Promega, E1290) and incubated for 120 min at 37°C. Firefly luciferase activity was measured at 37°C and 5% CO2 using a Spark 10M microplate reader (Tecan). After initial measurement of the baseline signal for 10 min (30 seconds intervals), cells were stimulated with 10 mL of 10x stock solution of the MC3R agonist NDP-aMSH (final concentration 1 mM) and real-time chemiluminescent signals were quantified for 25 minutes (30 seconds intervals). In each experiment, a negative control using mock transfected cells (empty pcDNA3.1(+) plasmid) was assayed. The area under the curve (AUC) was calculated for each cAMP production curve considering total peak area above the baseline calculated as the average signal for mock pcDNA3.1(+)-transfected cells. For data normalization, the AUC from mock transfected cells was set as 0 and the AUC from WT MC3R was set as 100%. Results are from six independent experiments.

Genetic epistasis between GPR83 and MC3R

To corroborate the *in vitro* interaction, we tested for evidence of a specific epistatic interaction between AAM GWAS signals at *GPR83* (rs592068-C) and *MC3R*

(rs3746619-A). We extracted genotypes for these SNPs in white-European unrelated UK Biobank women (N=204,303). After adjusting AAM for standard covariates (GWAS chip, age, sex, PC1-10), we modelled the interaction between genotype dosages at the two signals using a linear model.

References

1. Parent, A.-S. *et al.* The timing of normal puberty and the age limits of sexual precocity: variations around the world, secular trends, and changes after migration. *Endocr. Rev.* **24**, 668–693 (2003).
2. Gajbhiye, R., Fung, J. N. & Montgomery, G. W. Complex genetics of female fertility. *NPJ Genomic Med.* **3**, 29 (2018).
3. McGrath, I. M., Mortlock, S. & Montgomery, G. W. Genetic Regulation of Physiological Reproductive Lifespan and Female Fertility. *Int. J. Mol. Sci.* **22**, 2556 (2021).
4. Elks, C. E. *et al.* Age at menarche and type 2 diabetes risk: the EPIC-InterAct study. *Diabetes Care* **36**, 3526–3534 (2013).
5. Prentice, P. & Viner, R. M. Pubertal timing and adult obesity and cardiometabolic risk in women and men: a systematic review and meta-analysis. *Int. J. Obes. 2005* **37**, 1036–1043 (2013).
6. Bodicoat, D. H. *et al.* Timing of pubertal stages and breast cancer risk: the Breakthrough Generations Study. *Breast Cancer Res. BCR* **16**, R18 (2014).
7. Cheng, T. S., Ong, K. K. & Biro, F. M. Trends Toward Earlier Puberty Timing in Girls and Its Likely Mechanisms. *J. Pediatr. Adolesc. Gynecol.* **35**, 527–531 (2022).
8. Perry, J. R. B., Murray, A., Day, F. R. & Ong, K. K. Molecular insights into the aetiology of female reproductive ageing. *Nat. Rev. Endocrinol.* **11**, 725–734 (2015).
9. 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.* **49**, 834–841 (2017).
10. Lunetta, K. L. *et al.* Rare coding variants and X-linked loci associated with age at menarche. *Nat. Commun.* **6**, 7756 (2015).
11. Elks, C. E. *et al.* Thirty new loci for age at menarche identified by a meta-analysis of genome-wide association studies. *Nat. Genet.* **42**, 1077–1085 (2010).
12. Perry, J. R. *et al.* Parent-of-origin-specific allelic associations among 106 genomic loci for age at menarche. *Nature* **514**, 92–97 (2014).
13. Horikoshi, M. *et al.* Elucidating the genetic architecture of reproductive ageing in the Japanese population. *Nat. Commun.* **9**, 1977 (2018).
14. Hollis, B. *et al.* Genomic analysis of male puberty timing highlights shared genetic basis with hair colour and lifespan. *Nat. Commun.* **11**, 1536 (2020).
15. Lam, B. Y. H. *et al.* MC3R links nutritional state to childhood growth and the timing of puberty. *Nature* **599**, 436–441 (2021).
16. Erikstrup, C. *et al.* Cohort Profile: The Danish Blood Donor Study. *Int. J. Epidemiol.* dyac194 (2022) doi:10.1093/ije/dyac194.
17. Sudlow, C. *et al.* UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS Med.* **12**, e1001779 (2015).
18. Day, F. R. *et al.* Shared genetic aetiology of puberty timing between sexes and with health-related outcomes. *Nat. Commun.* **6**, 8842 (2015).

19. Rentzsch, P., Witten, D., Cooper, G. M., Shendure, J. & Kircher, M. CADD: predicting the deleteriousness of variants throughout the human genome. *Nucleic Acids Res.* **47**, D886–D894 (2019).
20. Young, J. *et al.* TAC3 and TACR3 defects cause hypothalamic congenital hypogonadotropic hypogonadism in humans. *J. Clin. Endocrinol. Metab.* **95**, 2287–2295 (2010).
21. Valadares, L. P. *et al.* MKRN3 Mutations in Central Precocious Puberty: A Systematic Review and Meta-Analysis. *J. Endocr. Soc.* **3**, 979–995 (2019).
22. Manotas, M. C., González, D. M., Céspedes, C., Forero, C. & Rojas Moreno, A. P. Genetic and Epigenetic Control of Puberty. *Sex. Dev. Genet. Mol. Biol. Evol. Endocrinol. Embryol. Pathol. Sex Determ. Differ.* **16**, 1–10 (2022).
23. Russwurm, C., Koesling, D. & Russwurm, M. Phosphodiesterase 10A Is Tethered to a Synaptic Signaling Complex in Striatum. *J. Biol. Chem.* **290**, 11936–11947 (2015).
24. Yasui, G. *et al.* Zinc finger protein 483 (ZNF483) regulates neuronal differentiation and methyl-CpG-binding protein 2 (MeCP2) intracellular localization. *Biochem. Biophys. Res. Commun.* **568**, 68–75 (2021).
25. Oleksiewicz, U. *et al.* TRIM28 and Interacting KRAB-ZNFs Control Self-Renewal of Human Pluripotent Stem Cells through Epigenetic Repression of Pro-differentiation Genes. *Stem Cell Rep.* **9**, 2065–2080 (2017).
26. Simon, D. *et al.* Mutations in the maternally imprinted gene MKRN3 are common in familial central precocious puberty. *Eur. J. Endocrinol.* **174**, 1–8 (2016).
27. Martin, A. R. *et al.* PanelApp crowdsources expert knowledge to establish consensus diagnostic gene panels. *Nat. Genet.* **51**, 1560–1565 (2019).
28. Shekari, S. *et al.* Monogenic causes of Premature Ovarian Insufficiency are rare and mostly recessive. 2022.11.21.22282589 Preprint at https://doi.org/10.1101/2022.11.21.22282589 (2022).
29. Mirshahi, U. L. *et al.* Reduced penetrance of MODY-associated HNF1A/HNF4A variants but not GCK variants in clinically unselected cohorts. *Am. J. Hum. Genet.* **109**, 2018–2028 (2022).
30. Gardner, E. J. *et al.* Damaging missense variants in IGF1R implicate a role for IGF-1 resistance in the etiology of type 2 diabetes. *Cell Genomics* **2**, None (2022).
31. Mak, T. S. H., Porsch, R. M., Choi, S. W., Zhou, X. & Sham, P. C. Polygenic scores via penalized regression on summary statistics. *Genet. Epidemiol.* **41**, 469–480 (2017).
32. Pickrell, J. K. Joint analysis of functional genomic data and genome-wide association studies of 18 human traits. *Am. J. Hum. Genet.* **94**, 559–573 (2014).
33. Reshef, Y. A. *et al.* Detecting genome-wide directional effects of transcription factor binding on polygenic disease risk. *Nat. Genet.* **50**, 1483–1493 (2018).
34. Nasser, J. *et al.* Genome-wide enhancer maps link risk variants to disease genes. *Nat. 2021 5937858* **593**, 238–243 (2021).
35. de Leeuw, C. A., Mooij, J. M., Heskes, T. & Posthuma, D. MAGMA: generalized gene-set analysis of GWAS data. *PLoS Comput. Biol.* **11**, e1004219 (2015).

36. Weeks, E. M. *et al.* Leveraging polygenic enrichments of gene features to predict genes underlying complex traits and diseases. 2020.09.08.20190561 Preprint at https://doi.org/10.1101/2020.09.08.20190561 (2020).
37. Stacey, D. *et al.* ProGeM: a framework for the prioritization of candidate causal genes at molecular quantitative trait loci. *Nucleic Acids Res.* **47**, e3 (2019).
38. Aragam, K. G. *et al.* Discovery and systematic characterization of risk variants and genes for coronary artery disease in over a million participants. *Nat. Genet.* **54**, 1803–1815 (2022).
39. Cangiano, B., Swee, D. S., Quinton, R. & Bonomi, M. Genetics of congenital hypogonadotropic hypogonadism: peculiarities and phenotype of an oligogenic disease. *Hum. Genet.* **140**, 77–111 (2021).
40. Sertedaki, A. *et al.* Whole Exome Sequencing Points towards a Multi-Gene Synergistic Action in the Pathogenesis of Congenital Combined Pituitary Hormone Deficiency. *Cells* **11**, 2088 (2022).
41. Butz, H., Nyírő, G., Kurucz, P. A., Likó, I. & Patócs, A. Molecular genetic diagnostics of hypogonadotropic hypogonadism: from panel design towards result interpretation in clinical practice. *Hum. Genet.* **140**, 113–134 (2021).
42. Herbison, A. E. Control of puberty onset and fertility by gonadotropin-releasing hormone neurons. *Nat. Rev. Endocrinol.* **12**, 452–466 (2016).
43. Parkin, K., Kapoor, R., Bhat, R. & Greenough, A. Genetic causes of hypopituitarism. *Arch. Med. Sci. AMS* **16**, 27–33 (2020).
44. Loos, R. J. F. & Yeo, G. S. H. The genetics of obesity: from discovery to biology. *Nat. Rev. Genet.* **23**, 120–133 (2022).
45. Zenker, M., Edouard, T., Blair, J. C. & Cappa, M. Noonan syndrome: improving recognition and diagnosis. *Arch. Dis. Child.* **107**, 1073–1078 (2022).
46. El Bouchikhi, I. *et al.* Noonan syndrome-causing genes: Molecular update and an assessment of the mutation rate. *Int. J. Pediatr. Adolesc. Med.* **3**, 133–142 (2016).
47. Forsythe, E. & Beales, P. L. Bardet-Biedl syndrome. *Eur. J. Hum. Genet. EJHG* **21**, 8–13 (2013).
48. Buiting, K., Williams, C. & Horsthemke, B. Angelman syndrome - insights into a rare neurogenetic disorder. *Nat. Rev. Neurol.* **12**, 584–593 (2016).
49. Cole, T. J., Ahmed, M. L., Preece, M. A., Hindmarsh, P. & Dunger, D. B. The relationship between Insulin-like Growth Factor 1, sex steroids and timing of the pubertal growth spurt. *Clin. Endocrinol. (Oxf.)* **82**, 862–869 (2015).
50. Mourouzis, I., Lavecchia, A. M. & Xinaris, C. Thyroid Hormone Signalling: From the Dawn of Life to the Bedside. *J. Mol. Evol.* **88**, 88–103 (2020).
51. Vazquez, M. J., Daza-Dueñas, S. & Tena-Sempere, M. Emerging Roles of Epigenetics in the Control of Reproductive Function: Focus on Central Neuroendocrine Mechanisms. *J. Endocr. Soc.* **5**, bvab152 (2021).
52. Helgeland, Ø. *et al.* Characterization of the genetic architecture of infant and early childhood body mass index. *Nat. Metab.* **4**, 344–358 (2022).
53. Kolberg, L., Raudvere, U., Kuzmin, I., Vilo, J. & Peterson, H. gprofiler2 -- an R package for gene list functional enrichment analysis and namespace conversion toolset g:Profiler. *F1000Research* **9**, ELIXIR-709 (2020).

54. Vassart, G. & Costagliola, S. G protein-coupled receptors: mutations and endocrine diseases. *Nat. Rev. Endocrinol.* **7**, 362–372 (2011).
55. Adams, F., Grassie, M., Shahid, M., Hill, D. R. & Henry, B. Acute oral dexamethasone administration reduces levels of orphan GPCR glucocorticoid-induced receptor (GIR) mRNA in rodent brain: potential role in HPA-axis function. *Brain Res. Mol. Brain Res.* **117**, 39–46 (2003).
56. Brézillon, S., Detheux, M., Parmentier, M., Hökfelt, T. & Hurd, Y. L. Distribution of an orphan G-protein coupled receptor (JP05) mRNA in the human brain. *Brain Res.* **921**, 21–30 (2001).
57. Müller, T. D. *et al.* The orphan receptor Gpr83 regulates systemic energy metabolism via ghrelin-dependent and ghrelin-independent mechanisms. *Nat. Commun.* **4**, 1968 (2013).
58. Lohse, M. J. Dimerization in GPCR mobility and signaling. *Curr. Opin. Pharmacol.* **10**, 53–58 (2010).
59. Qi, T. *et al.* Identifying gene targets for brain-related traits using transcriptomic and methylomic data from blood. *Nat. Commun. 2018 91* **9**, 1–12 (2018).
60. GTEx Consortium. Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. *Science* **348**, 648–660 (2015).
61. 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**, 1294–1303 (2015).
62. Ruth, K. S. *et al.* Genetic insights into biological mechanisms governing human ovarian ageing. *Nature* **596**, 393–397 (2021).
63. Ruth, K. S. *et al.* Events in Early Life are Associated with Female Reproductive Ageing: A UK Biobank Study. *Sci. Rep.* **6**, 24710 (2016).
64. Mbarek, H. *et al.* Identification of Common Genetic Variants Influencing Spontaneous Dizygotic Twinning and Female Fertility. *Am. J. Hum. Genet.* **98**, 898–908 (2016).
65. Day, F. *et al.* Large-scale genome-wide meta-analysis of polycystic ovary syndrome suggests shared genetic architecture for different diagnosis criteria. *PLoS Genet.* **14**, e1007813 (2018).
66. Thompson, D. J. *et al.* Genetic predisposition to mosaic Y chromosome loss in blood. *Nature* **575**, 652–657 (2019).
67. Lappalainen, T. & MacArthur, D. G. From variant to function in human disease genetics. *Science* **373**, 1464–1468 (2021).
68. Wen, S., Ai, W., Alim, Z. & Boehm, U. Embryonic gonadotropin-releasing hormone signaling is necessary for maturation of the male reproductive axis. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 16372–16377 (2010).
69. Winkler, T. W. *et al.* Quality control and conduct of genome-wide association meta-analyses. *Nat. Protoc.* **9**, 1192–1212 (2014).
70. Willer, C. J., Li, Y. & Abecasis, G. R. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinforma. Oxf. Engl.* **26**, 2190–2191 (2010).

71. Bycroft, C. *et al.* The UK Biobank resource with deep phenotyping and genomic data. *Nature* **562**, 203–209 (2018).
72. Backman, J. D. *et al.* Exome sequencing and analysis of 454,787 UK Biobank participants. *Nature* **599**, 628–634 (2021).
73. McLaren, W. *et al.* The Ensembl Variant Effect Predictor. *Genome Biol.* **17**, 122 (2016).
74. Karczewski, K. J. *et al.* The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature* **581**, 434–443 (2020).
75. Loh, P.-R. *et al.* Efficient Bayesian mixed-model analysis increases association power in large cohorts. *Nat. Genet.* **47**, 284–290 (2015).
76. UniProt Consortium. UniProt: the Universal Protein Knowledgebase in 2023. *Nucleic Acids Res.* **51**, D523–D531 (2023).
77. Wang, Q. *et al.* Rare variant contribution to human disease in 281,104 UK Biobank exomes. *Nature* **597**, 527–532 (2021).
78. 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).
79. Purcell, S. *et al.* PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. *Am. J. Hum. Genet.* **81**, 559–575 (2007).
80. Finucane, H. K. *et al.* Heritability enrichment of specifically expressed genes identifies disease-relevant tissues and cell types. *Nat. Genet.* **50**, 621–629 (2018).
81. Võsa, U. *et al.* Large-scale cis- and trans-eQTL analyses identify thousands of genetic loci and polygenic scores that regulate blood gene expression. *Nat. Genet. 2021 539* **53**, 1300–1310 (2021).
82. Zhu, Z. *et al.* Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets. *Nat. Genet.* **48**, 481 (2016).
83. Giambartolomei, C. *et al.* Bayesian test for colocalisation between pairs of genetic association studies using summary statistics. *PLoS Genet* **10**, e1004383 (2014).
84. Pietzner, M. *et al.* Mapping the proteo-genomic convergence of human diseases. *Science* **374**, eabj1541 (2021).
85. Vaser, R., Adusumalli, S., Leng, S. N., Sikic, M. & Ng, P. C. SIFT missense predictions for genomes. *Nat. Protoc. 2015 111* **11**, 1–9 (2015).
86. Adzhubei, I., Jordan, D. M. & Sunyaev, S. R. Predicting Functional Effect of Human Missense Mutations Using PolyPhen-2. *Curr. Protoc. Hum. Genet.* **76**, 7.20.1-7.20.41 (2013).
87. Szklarczyk, D. *et al.* The STRING database in 2021: customizable protein-protein networks, and functional characterization of user-uploaded gene/measurement sets. *Nucleic Acids Res.* **49**, D605–D612 (2021).
88. Magnus, P. *et al.* Cohort profile: the Norwegian Mother and Child Cohort Study (MoBa). *Int. J. Epidemiol.* **35**, 1146–1150 (2006).
89. Smith, G. D. & Ebrahim, S. 'Mendelian randomization': can genetic epidemiology contribute to understanding environmental determinants of disease? *Int. J. Epidemiol.* **32**, 1–22 (2003).

90. Slob, E. A. W. & Burgess, S. A comparison of robust Mendelian randomization methods using summary data. *Genet. Epidemiol.* **44**, 313–329 (2020).
91. Genolini, C. & Falissard, B. KmL: a package to cluster longitudinal data. *Comput. Methods Programs Biomed.* **104**, e112-121 (2011).
92. Gillespie, M. *et al.* The reactome pathway knowledgebase 2022. *Nucleic Acids Res.* **50**, D687–D692 (2022).
93. Gene Ontology Consortium. The Gene Ontology resource: enriching a GOld mine. *Nucleic Acids Res.* **49**, D325–D334 (2021).