This is the published version of a paper published in *Nature Communications*.

Citation for the original published paper (version of record):

Rare coding variants and X-linked loci associated with age at menarche.
*Nature Communications*, 6
http://dx.doi.org/10.1038/ncomms8756

Access to the published version may require subscription.

N.B. When citing this work, cite the original published paper.

Permanent link to this version:
http://urn.kb.se/resolve?urn=urn:nbn:se:umu:diva-109458
Rare coding variants and X-linked loci associated with age at menarche

Kathryn L. Lunetta et al.

More than 100 loci have been identified for age at menarche by genome-wide association studies; however, collectively these explain only ~3% of the trait variance. Here we test two overlooked sources of variation in 192,974 European ancestry women: low-frequency protein-coding variants and X-chromosome variants. Five missense/nonsense variants (in ALMS1/LAMB2/TNRC6A/TACR3/PRKAG1) are associated with age at menarche (minor allele frequencies 0.08–4.6%; effect sizes 0.08–1.25 years per allele; P < 5 × 10^{-8}). In addition, we identify common X-chromosome loci at IGSF1 (rs762080, P = 9.4 × 10^{-13}) and FAAH2 (rs5914101, P = 4.9 × 10^{-15}). Highlighted genes implicate cellular energy homeostasis, post-transcriptional gene silencing and fatty-acid amide signalling. A frequently reported mutation in TACR3 for idiopathic hypogonatrophic hypogonadism (p.W275X) is associated with 1.25-year-later menarche (P = 2.8 × 10^{-11}), illustrating the utility of population studies to estimate the penetrance of reportedly pathogenic mutations. Collectively, these novel variants explain ~0.5% variance, indicating that these overlooked sources of variation do not substantially explain the ‘missing heritability’ of this complex trait.
A
g at menarche, the onset of first menstruation in females, indicates the start of reproductive maturity and is a
commonly reported marker of pubertal timing. One hundred and six genomic loci for this highly heritable trait have
been mapped by genome-wide association studies (GWAS), implicating many previously unsuspected mechanisms.1
However, to date that approach has been limited to consideration of only those genetic variants captured by
autosomal HapMap2 reference panels. In particular, like most GWAS for other complex traits, previous GWAS for age at
menarche provided poor coverage for low-frequency variants and omitted sex chromosome data.

Here we report a dual strategy for assessing genetic variation overlooked by those prior efforts: low-frequency protein-coding
variants genotyped by large-scale exome-locussed arrays and high-density X-chromosome single-nucleotide polymorphism
(SNP) genotyping and imputation. We identify several new associations between rare protein-coding and X-linked variants
with age at menarche in women of European ancestry. The findings implicate new mechanisms that regulate puberty
timing, but collectively these novel variants explained only ~0.5% of the variance, indicating that these often overlooked
sources of variation do not substantially explain the ‘missing heritability’ of this complex trait.

Results
From the exome array studies, 61,734 low-frequency (minor allele frequency (MAF) <5%) variants passed quality-control (QC)
criteria in a combined sample of up to 76,657 women of European ancestry from 19 studies (Supplementary Table 1).
Gene-based burden and SKAT tests that aggregate the effects of variants with MAF<1% yielded no significant associations with
age at menarche. A linear regression test was used to derive all P values obtained in this study. Meta-analysis of individual
variant associations with questionnaire-reported variation in age at menarche (restricted to the ages of 9–17 years) in this discovery
phase identified one signal at genome-wide statistical significance (P = 5 × 10−8); this was a rare missense variant in the Alström’s
gene (ALMS1, rs45501594, p.T3544S, MAF 1%; P = 4.6 × 10−10). For follow-up testing in up to 116,317
independent women of European ancestry from the deCODE (Diabetes Epidemiology: Collaborative analysis of Diagnostic
criteria in Europe) and 23andMe studies, we selected rs45501594 and 23 other variants that met the following criteria: protein
coding, present in over half of the exome array studies, and with association P < 5 × 10−4. In the follow-up samples, 7 of the 20
variants that passed QC showed directionally concordant confirmatory associations with P < 0.05, of which five reached genome-wide significance in a combined meta-analysis of discovery phase and follow-up data (Table 1, Fig. 1, Supplementary Fig. 1). No significant heterogeneity between studies was observed at any of these loci (Supplementary Fig. 2).

The rare missense variant in ALMS1 (rs45501594, Supplementary Fig. 3) remained the strongest signal identified
using exome array studies (combined: P = 6.8 × 10−20). In the follow-up samples, each rare allele was associated with 0.23-year-
later age at menarche, an effect size more than double that of any genetic variant previously reported for puberty timing in the
general population. This strong signal was not detected by the previous HapMap2-based GWAS as it is poorly tagged by
common SNPs in that reference panel (maximum proxy SNP, r² = 0.24). Deleterious mutations in this gene cause Alström’s
syndrome (OMIM no. 203800), a rare, autosomal-recessive disorder characterized by cone–rod dystrophy, sensorineural
hearing loss, dilated cardiomyopathy, childhood obesity, insulin resistance, diabetes mellitus, hypogonadotropic hypogonadism in
males, menstrual irregularities and early puberty in females, and short stature in adulthood2. Hypogonadism was also invariably
observed in an ALMS1 gene-trapped mouse model.3

The variant with largest effect was a rare stop-gain mutation in the tachykinin receptor 3 gene (TACR3; rs144292455, MAF = 0.08%, combined P = 2.8 × 10−11, Supplementary Fig. 3); in follow-up samples each rare allele was associated with 1.25-
year-later age at menarche. Common HapMap2 SNPs at the TACR3 locus were previously associated with age at menarche;4
however, the rare variant rs144292455 is not tagged by the HapMap2 or conventional 1000G imputation (it was directly genotyped in 23andMe and was imputed in deCODE). Statistical independence was confirmed by observing significant association with the common TACR3 SNP in a sensitivity analysis within a
participating study (Women’s Genome Health Study (WGHS), Supplementary Table 1) that excluded rare allele carriers. The rare allele causes a premature stop codon (p.W275X) in the fifth transmembrane segment of the 465 amino-acid receptor for the
neuropeptide neurokinin B, and is the most frequently reported TACR3 mutation in the rare reproductive disorder idiopathic
hypogonadotropic hypogonadism (idiopathic hypogonadotropic hypogonadism (IHH), OMIM no. 614840).4 Both homozygous
and heterozygous p.W275X variants have been reported in male IHH cases with features of ‘early androgen deficiency’; however, notably the heterozygous cases showed evidence of spontaneous
neuroendocrine recovery. Our findings suggest that heterozygous p.W275X variants contribute to the normal variation in puberty
timing, whereas homozygous inheritance or possibly compound heterozygosity is required for IHH.

A low-frequency missense variant in the LAMB2 gene was associated with 0.08-year-later age at menarche (rs35713889, p.G914R, MAF 4%; P = 1.1 × 10−11, Supplementary Fig. 3). In the same region (3p21.31) we previously reported a HapMap2 GWAS locus for age at menarche (locus 19a and 19b in ref. 1); however, the low-frequency variant rs35713889 is poorly tagged by common HapMap2 SNPs (the best proxy rs1134043, r² = 0.24, was reportedly not associated with age at menarche: P = 0.35 (ref. 1)). The strongest reported1 HapMap2 signal at this locus is only weakly correlated with rs35713889 (rs3870341, MAF = 26%; r² = 0.07, distance 422 kb), and both signals remained significant when jointly tested in a follow-up sample of 76,831 women from the
23andMe study (in separate models: rs35713889: β = 0.08 years per allele, P = 0.0001 and rs3870341: β = 0.04, P = 4.5 × 10−5; in the joint model: rs35713889: β = 0.06, P = 0.004 and rs3870341: β = 0.03, P = 0.001). LAMB2 encodes one of 15 subunits of Laminin, an extracellular matrix glycoprotein with a key role in the attachment, migration and organization of cells into tissues during embryonic development. Rare recessive mutations in LAMB2 cause Pierson’s syndrome (OMIM no. 609049), a disorder characterized by congenital nephrotic syndrome and ocular anomalies, typically with microcoria;5 neurological abnormalities are also described likely because of cortical laminar disorganization6. Common variants in/near other Laminin genes have been reported for a broad range of complex traits, including type 2 diabetes,7 refractive error8, colorectal cancer,9 IgG glycosylation,10 ulcerative colitis11 and coffee consumption12.

A low-frequency missense variant in the TNRCC6A gene was associated with later age at menarche (rs113388806; p.Q1112H; MAF 4.7%; β = 0.08 years per allele; P = 1.1 × 10−11, Supplementary Fig. 3). This signal was only moderately well tagged by common HapMap2 SNPs (best proxy: rs12447003, r² = 0.36, reported association with age at menarche: P = 0.0005 (ref. 1)). TNRCC6A encodes an Argonaute-navigator protein, responsible for post-transcriptional gene silencing through RNA interference and microRNA pathways.13 This finding further
Table 1 | Association statistics for the novel low-frequency and X-chromosome variants.

<table>
<thead>
<tr>
<th>Gene</th>
<th>SNP</th>
<th>Location</th>
<th>Alleles*</th>
<th>Discovery</th>
<th>Follow-up</th>
<th>Combined</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Effect (s.e.)</td>
<td>P</td>
<td>N</td>
</tr>
<tr>
<td>Exome array</td>
<td>rs45501594</td>
<td>2p13.1</td>
<td>G/C/1.1%</td>
<td>0.26 (0.04)</td>
<td>4.6E-10</td>
<td>57,867</td>
</tr>
<tr>
<td>LAMB2</td>
<td>rs3573889</td>
<td>3p21.31</td>
<td>T/C/4.4%</td>
<td>0.11 (0.02)</td>
<td>5.0E-07</td>
<td>58,695</td>
</tr>
<tr>
<td>TNRC6A</td>
<td>rs11338806</td>
<td>1q21.1</td>
<td>T/A/4.7%</td>
<td>0.09 (0.02)</td>
<td>1.7E-05</td>
<td>76,657</td>
</tr>
<tr>
<td>TACR3</td>
<td>rs144292545</td>
<td>4q24</td>
<td>T/C/0.8%</td>
<td>0.71 (0.15)</td>
<td>1.3E-06</td>
<td>68,487</td>
</tr>
<tr>
<td>PRKAG1</td>
<td>rs1026930</td>
<td>12q13.2</td>
<td>C/G/3.4%</td>
<td>0.11 (0.02)</td>
<td>4.4E-07</td>
<td>76,657</td>
</tr>
</tbody>
</table>

1000G X-chromosome

<table>
<thead>
<tr>
<th>Gene</th>
<th>SNP</th>
<th>Location</th>
<th>Alleles*</th>
<th>Effect (s.e.)</th>
<th>P</th>
<th>N</th>
<th>Effect (s.e.)</th>
<th>P</th>
<th>N</th>
<th>Effect (s.e.)</th>
<th>P</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>IGSF1</td>
<td>rs762080</td>
<td>Xq26.2</td>
<td>A/C/24%</td>
<td>-0.07 (0.01)</td>
<td>4.1E-12</td>
<td>76,831</td>
<td>-0.04 (0.01)</td>
<td>0.04%</td>
<td>6.7E-03</td>
<td>39,486</td>
<td>-0.06 (0.008)</td>
<td>9.4E-13</td>
</tr>
<tr>
<td>FAAH2</td>
<td>rs5914101</td>
<td>Xp11.21</td>
<td>A/G/24%</td>
<td>-0.07 (0.01)</td>
<td>1.1E-09</td>
<td>76,831</td>
<td>-0.03 (0.01)</td>
<td>0.03%</td>
<td>2.0E-02</td>
<td>39,486</td>
<td>-0.05 (0.009)</td>
<td>4.9E-10</td>
</tr>
</tbody>
</table>

Figure 1 | A ‘Manhattan plot’ of menarche association statistics for the genotyped low-frequency exome array variants. Test statistics are shown from the exome-chip discovery-phase samples, with the exception of the five labelled loci that indicate results from the combined discovery and replication set.

Table 1 extends the range of epigenetic mechanisms implicated in the regulation of puberty. A low-frequency missense variant in PRKAG1 was associated with earlier age at menarche (rs1126930; p.T98S; MAF 3.4%; β = −0.09 years per allele, P = 9.6 × 10−11; Supplementary Fig. 3). This low-frequency variant is only moderately well tagged by common HapMap2 SNPs (max r2 = 0.36), which reported showed subgenome-wide significant association with age at menarche (rs11837234, P = 3.1 × 10−6). This PRKAG1 missense variant is in the same locus as, but not correlated to, a reported1 common signal for age at menarche (rs7138803, 848 kb apart, r2 = 0.02). PRKAG1 encodes the gamma-1 regulatory subunit of AMP-activated protein kinase, which senses and maintains cellular energy homeostasis by promoting fatty-acid oxidation and inhibiting fatty-acid synthesis; PRKAG1 is overexpressed in ovarian carcinomas15 and is somatically mutated in colorectal cancers16.

Our second genotyping approach considered X-chromosome GWAS SNPs in up to 76,831 women of European ancestry from the 23andMe study17. Imputation was performed against the 1000 Genomes reference, yielding genotype data for ~266,000 X-chromosome variants (MAF > 1%). Two signals, in/near IGSF1 and FAAH2, reached genome-wide significance for association with age at menarche and both associations were confirmed in 39,486 independent women of European ancestry from the deCODE study.

Common variants in and near IGSF1 were robustly associated with age at menarche (lead SNP: rs762080, MAF = 24%; β = 0.06 years per allele, P = 9.4 × 10−13; Supplementary Fig. 4). IGSF1 encodes the immunoglobulin superfamily member 1, which is a plasma membrane glycoprotein highly expressed in the pituitary gland and tests. Rare X-linked mutations in IGSF1 were recently described to cause central hypothyroidism, hypoprolactinemia, delayed puberty and macro-orchidism in males (OMIM no. 300888)18,19. Heterozygous female carriers reportedly had normal age at menarche; however, 6/18 had central hypothyroidism and 4/18 underwent oophorectomy for ovarian cysts19.

The second X-chromosome locus, in Xp11.21 (lead SNP rs5914101 is intronic in FAAH2, MAF 24%, β = 0.05 years per allele, P = 1.9 × 10−10; Supplementary Fig. 4), lies within the critical region for Turner’s syndrome, which is the most common cause of primary ovarian insufficiency20. FAAH2 encodes fatty-acid amide hydrolase 2. This enzyme catalyses the hydrolysis and degradation of bioactive fatty-acid amides, a large class of endogenous signalling lipids including the endocannabinoids, which modulate several physiological processes, including feeding, inflammation, pain, sleep and various reproductive processes, including hypothalamic gonadotropin-releasing hormone secretion21,22.

We sought to further functionally characterize the seven genes implicated by these analyses using expression data on 53 tissue types from the Genotype-Tissue Expression consortium23. All seven genes showed high relative tissue expression in the ovary and/or brain (specifically the hypothalamus; Supplementary Fig. 5); however, none of the lead SNPs showed a significant association with primary ovarian insufficiency20. FAAH2 encodes fatty-acid amide hydrolase 2. This enzyme catalyses the hydrolysis and degradation of bioactive fatty-acid amides, a large class of endogenous signalling lipids including the endocannabinoids, which modulate several physiological processes, including feeding, inflammation, pain, sleep and various reproductive processes, including hypothalamic gonadotropin-releasing hormone secretion21,22.

Discussion

In summary, by large-scale analysis of genetic variation not captured by previous GWAS for age at menarche, we identified several low-frequency exonic variants of relatively large effect and two common X-chromosome signals. The implicated genes...
provide new insights into the mechanisms that link energy homeostasis to puberty timing, indicate possible roles of RNA-mediated gene silencing and fatty-acid amide signalling, and link genes behind rare autosomal, X-linked and syndromic disorders of puberty to normal variation in reproductive timing. Our findings using dense exome arrays in large unselected populations are informative for the clinical interpretation of heterogeneous TACR3 variants in patients with rare disorders. In the deCODE study these novel variants collectively explained only 0.5% of the variance in age at menarche, suggesting that these often overlooked sources of genetic variation do not contribute disproportionately to the missing heritability of this complex trait. While variants with MAF below 1% are likely not well represented here, our findings indicate that, similar to other complex traits, the genetic architecture of puberty timing is likely dominated by the additive effects of hundreds or even thousands of variants, each with relatively small effect.

Methods

Exome array discovery analysis. Exome array genotype data were generated across discovery studies in up to 76,857 women of genetically determined European ancestry with questionnaire-reported age at menarche between ages 9 and 17 years (Supplementary Table 1). Exome array genotype calling for three studies (Framingham Heart Study (FHS), The Atherosclerosis Risk in Communities (ARIC) and Rotterdam Study (RS); totaling ~9,000 women) was performed jointly as part of the CHARGE joint calling protocol25, which included over 62,000 individuals. Four additional studies (Cambridge Cancer, KORA, Kocalla, Generation Scotland, total N = 9,700) used the cluster file made available by CHARGE to call genotypes. Other studies followed standard calling and QC protocols for the Exome array (Supplementary Table 2). Each contributing study ran a linear regression model on age at menarche, adjusted for birth year and principal components derived from genotypes, using the skatMeta/sqMeta package in R. Studies with family data included a random effect to account for relationships. Alleles were aligned to a common reference file before association testing (SNPInfo_HumanExome-12v1, rev5.2.txt, available at http://www.chargeconsortium.com/main/exomechip/) and variants shared across the meta-analysis were excluded. We excluded X-chromosome gene-based testing (within sqMeta) for low-frequency variants using fixed effect burden tests, which assume that all rare variants have the same effect direction and size (scaled by a weight determined by allele frequency), and SKAT tests, which assume that rare variant effects are random and can contain a mixture of null, protective and risk rare alleles. These tests were run using three variant filters, all of which included only variants with MAF <1%: (1) all non-synonymous; (2) non-synonymous annotated as ‘damaging’ (conserved and predicted damaging, see http://www.chargeconsortium.com/main/exomechip/); and (3) only loss of function. The multiple testing adjustment included two tests × three filters × number of genes, resulting in a type I error threshold P<1.14 × 10−6. For individual variants, a fixed-effects inverse variance-weighted meta-analysis was performed across all studies using METAL (http://www.sph.umich.edu/csg/abecasis/Metal/), with associations considered significant at a conservative P<5 × 10−8.

Exome array follow-up studies. We performed follow-up testing of selected exome array variants in the 23andMe study (as described below) and also in 39,486 independent women of European ancestry from the deCODE study, Iceland, who had genotypes on over 34 million variants by imputation of whole-genome sequencing-identified SNPs and indels on Illumina SNP chip data (Supplementary Table 1)26. Variants from both studies were required to either pass genotyping QC (23andMe only, described below) or have imputation quality score >0.4. X-chromosome follow-up was performed in the deCODE study alone. All participants in all published studies provided informed consent, and the research protocol of each study was approved by their local research ethics committee.

1000G X-chromosome discovery meta-analysis. X-chromosome SNP data were generated in up to 76,831 women of European ancestry from the 23andMe study27,28, with questionnaire-reported age at menarche between the ages of 8 and 16 years, and who were genotyped on one or more of three GWAS arrays that also included customized content on human pathogenic variants (Supplementary Table 1)28. 23andMe participants provided informed consent to take part in this research under a protocol approved by the AAHRPP-accredited institutional review board, Ethical and Independent Review Services. Before imputation, we excluded SNPs with Hardy–Weinberg equilibrium P<10−5; call rate <95% or with large allele frequency discrepancies compared with European 1000 Genomes reference data. Frequency discrepancies were identified by computing a 2 × 2 table of allele counts for European 1000 Genomes samples and 2,000 randomly sampled 23andMe customers with European ancestry, and identifying SNPs with a χ2

References


Acknowledgements

Sources of funding for the contributing studies are listed in Supplementary Table 3.

Author contributions


Additional information

Supplementary Information accompanies this paper at http://www.nature.com/naturecommunications

Competing financial interests: The authors declare no competing financial interests.

Reprints and permission information is available online at http://npg.nature.com/reprintsandpermissions/

How to cite this article: Lunetta, K. L. et al. Rare coding variants and X-linked loci associated with age at menarche. Nat. Commun. 6:7756 doi: 10.1038/ncomms8756 (2015).


1 Boston University School of Public Health, Department of Biostatistics, Boston, Massachusetts 02118, USA. 2 NHLBI’s and Boston University’s Framingham Heart Study, Framingham, Massachusetts 01702-5827, USA. 3 MRC Epidemiology Unit, University of Cambridge School of Clinical Medicine, Box 285 Institute of Metabolic Science, Cambridge Biomedical Campus, Cambridge CB2 0QQ, UK. 4 deCODE genetics/Amgen, Inc., Reykjavik 101-102, Iceland. 5 Genetics of Complex Traits, University of Exeter Medical School, University of Exeter, Exeter EX1 2LU, UK. 6 23andMe Inc., 1390 Shorebird Way, Mountain View, California 94043, USA. 7 Estonian Genome Center, University of Tartu, Tartu 51010, Estonia. 8 Division of Endocrinology, Boston Children’s Hospital, Boston, MA 02115, USA. 9 Department of Genetics, Harvard Medical School, Boston, MA 02115, USA. 10 Broad Institute of the Massachusetts Institute of Technology and Harvard University, Cambridge, MA 02142, USA. 11 Research Unit of Molecular Epidemiology, Helmholtz Zentrum München–German Research Center for Environmental Health, Neufahrn 85764, Germany. 12 Institute of Genetic Epidemiology, Helmholtz Zentrum München–German Research Center for Environmental Health, Neufahrn 85764, Germany. 13 Department of Epidemiology, Indiana University Richard M. Fairbanks School of Public Health, Indianapolis, IN 46202, USA. 14 Indiana University Melvin and Bren Simon Cancer Center, Indianapolis, IN 46202, USA. 15 Medical Research Council Human Genetics Unit, Institute of Genetics and Molecular Medicine, University of Edinburgh, Edinburgh EH4 2XU, UK. 16 Institute of Genetics and
ARTICLE

NATURE COMMUNICATIONS | DOI: 10.1038/ncomms8756 | www.nature.com/naturecommunications

Biomedical Research, National Research Council, Cagliari, Sardinia 09042, Italy. 17 University of Sassari, Department of Biomedical Sciences, Sassari, Sassari 07100, Italy. 18 Center for Statistical Genetics, Ann Arbor, University of Michigan, Michigan 48109-2029, USA. 19 Institute for Maternal and Child Health—IRCCS “Burlo Garofolo”, Trieste 34137, Italy. 20 Division of Preventive Medicine, Brigham and Women’s Hospital, Boston, MA 02215. 21 Fred Hutchinson Cancer Research Center, Public Health Sciences Division, Seattle, WA 98109-1024, USA. 22 Department of Internal Medicine, Erasmus MC, Rotterdam 3015GE, the Netherlands. 23 Institute for Community Medicine, University Medicine Greifswald, Greifswald 17475, Germany. 24 Center for Cancer Genetics Epidemiology, Department of Public Health and Primary Care, University of Cambridge, CB1 8RN, UK. 25 Division of Genetics and Cell Biology, Ran Raffeile Scientific Institute, Milano 20132, Italy. 26 School of Women’s and Infants’ Health, The University of Western Australia, WA-6009, Australia. 27 Program in Personalized Medicine, Division of Endocrinology, Diabetes and Nutrition—University of Maryland School of Medicine, Baltimore, MD 21201, USA. 28 Boston University School of Medicine, Department of Medicine, Sections of Preventive Medicine and Endocrinology, Boston, MA, USA. 29 Division of Epidemiology & Community Health, University of Minnesota, Minneapolis, MN 55455, USA. 30 Centre for Cancer Genetic Epidemiology, Department of Oncology, University of Cambridge, Cambridge CB1 8RN, UK. 31 Department of Clinical Medical Sciences, Surgical and Health, University of Trieste, Trieste 34149, Italy. 32 Human Genetics Center, School of Public Health, The University of Texas Health Science Center at Houston, Houston, TX 77030, USA. 33 School of Engineering and Natural Sciences, University of Iceland, Reykjavik IS-101, Iceland. 34 Musculoskeletal Research Programme, Division of Applied Medicine, University of Aberdeen, Aberdeen AB25 2ZD, UK. 35 Genetic Epidemiology Unit Department of Epidemiology, Erasmus MC, Rotterdam 3015 GE, the Netherlands. 36 State Key Laboratory of Medical Genomics, Shanghai Institute of Hematology, Rui Jin Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai 200025, China. 37 Department of Internal Medicine, The Ohio State University, Columbus, Ohio 43210, USA. 38 Hebrew SeniorLife Institute for Aging Research, Boston, MA 02131, USA. 39 Harvard Medical School, Boston, MA 02115, USA. 40 German Center for Diabetes Research, Neuberg 85764, Germany. 41 Department of Genetics, University of North Carolina, Chapel Hill, NC 27599, USA. 42 Department of Biostatistics, University of North Carolina, Chapel Hill, NC 27599, USA. 43 Department of Epidemiology, Harvard School of Public Health, Boston, MA 02115, USA. 44 British Heart Foundation Glasgow Cardiovascular Research Centre, Institute of Cardiovascular and Medical Sciences, College of Medical, Veterinary and Life Sciences, University of Glasgow, Glasgow G12 8TA, UK. 45 Faculty of Medicine, University of Split, Split, Croatia. 46 Medical Genetics Section, Centre for Genomic and Experimental Medicine, Institute of Genetics and Molecular Medicine, University of Edinburgh, Edinburgh EH4 2XL, UK. 47 Institute for Population Health Sciences and Informatics, University of Edinburgh, Teviot Place, Edinburgh EH8 9AG, Scotland. 48 National Institute on Aging, Intramural Research Program, Baltimore, MD 20892, USA. 49 Institute of Epidemiology II, Helmholtz Zentrum München - German Research Center for Environmental Health, Neuberg 85764, Germany. 50 Interfaculty Institute for Genetics and Functional Genomics, University Medicine Greifswald, Greifswald 17475, Germany. 51 Department of Physiology and Biophysics, University of Mississippi Medical Center, Jackson, MS 39216, USA. 52 Department of Obstetrics and Gynecology, University Medicine Greifswald, Greifswald 17475, Germany. 53 Changing Division of Network Medicine, Department of Medicine, Brigham and Women’s Hospital and Harvard Medical School, Boston, MA 02115, USA. 54 Department of Nutrition, Harvard School of Public Health, Boston, MA 02115, USA. 55 Departments of Epidemiology and Medicine Brown University, Brown University, Providence, RI 02912, USA. 56 Institute of Molecular and Cell Biology, University of Tartu, Tartu 51010, Estonia. 57 Institute of Medical Informatics, Biometry and Epidemiology, Chair of Genetic Epidemiology, Ludwig-Maximilians-Universität, Munich 81377, Germany. 58 Department of Medicine, Vanderbilt University Medical Center, Nashville, Tennessee 37203, USA. 59 Department of Epidemiology, University of North Carolina, Chapel Hill, NC 27599, USA. 60 Faculty of Medicine, University of Iceland, Reykjavik IS-101, Iceland. 61 Boston University School of Medicine, Department of Medicine, Section of General Internal Medicine, Boston, MA 02118, USA. 62 Department of Paediatrics, University of Cambridge, Cambridge CB2 0QQ, UK. * These authors contributed equally to this work. †A full list of consortium members appears at the end of the paper.

**EPIC-InterAct Consortium**

Nita G. Forouhi2, Nicola D. Kerrison2, Stephen J. Sharp2, Matt Sims3, Inês Barroso63,64, Panos Deloukas63, Mark I. McCarthy65,66,67, Larraitz Arriola68,69,70, Beverley Balkau71,72, Aurelio Barricarte70,73, Heiner Boeing74, Paul W. Franks75,76, Carlos Gonzalez77, Sara Grioni78, Rudolf Kaaks79, Timothy J. Key80, Carmen Navarro70,81,82, Peter M. Nilsson79, Kim Overvad83,84, Domenico Palli85, Salvatore Panico86, J. Ramón Quirós87, Olov Rolandsson3,76, Carlotta Sacerdoti88,89, Maria-José Sánchez70,90,91, Nadia Slimani92, Anne Tjonneland93, Rosario Tumino94,95, Daphne L. van der A96, Yvonne T. van der Schouw97 & Elio Riboli98

63The Wellcome Trust Sanger Institute, Cambridge, CB10 1SA, UK. 64University of Cambridge Metabolic Research Laboratories, Cambridge CB2 OQO, UK. 65Oxford Centre for Diabetes, Endocrinology and Metabolism (OCDEM), University of Oxford, OX3 7LU, UK. 66Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford OX3 7BN, UK. 67Oxford NIHR Biomedical Research Centre, Oxford OX3 7LU, UK. 68Public Health Division of Gipuzkoa, 20013 San Sebastian, Spain. 69Instituto BIO-Donostia, Basque Government, 20014 Donostia, San Sebastian, Spain. 70CIBER Epidemiology y Salud Pública (CIBERESP), 28029 Madrid, Spain. 71Inserm, CESP, U1018, 94807 Villejuif cedex, France. 72Univ Paris-Sud, USMR 1018, F-94805, Villejuif, France. 73Navarre Public Health Institute (ISPIN), 31003 Pamplona, Spain. 74German Institute of Human Nutrition Potsdam-Rehbruecke, 14558 Nuthetal, Germany. 75Department of Clinical Sciences, Lund University, S-205 02 Malmo, Sweden. 76Department of Public Health and Clinical Medicine, Umeå University, 90187 Umeå, Sweden. 77Catalan Institute of Oncology (ICO), 08916 Badalona, Barcelona, Spain. 78Epidemiology and Prevention Unit, 20133 Milan, Italy. 79German Cancer Research Centre (DKFZ), 69120 Heidelberg, Germany. 80Cancer Epidemiology Unit, Nuffield Department of Population Health, University of Oxford, OX3 7LF, UK. 81Department of Epidemiology, Murcia Regional Health Council, 30008 Murcia, Spain. 82Unit of Preventive Medicine and Public Health, School of Medicine, University of Murcia, 30100 Espinardo, Spain. 83Department of Public Health, Sects for Epidemiology, Aarhus University, DK-8000 Aarhus C, Denmark. 84Aalborg University Hospital, 9010 Aalborg, Denmark. 85Cancer Research and Prevention Institute (ISPO), 50141 Florence, Italy. 86Dipartimento di Medicina Clinica e Chirurgia, Federico II University, 80131 Naples, Italy. 87Public Health Directorate, 33006-Oviedo-Asturias, Spain. 88Unit of Cancer Epidemiology, Città della Salute e della Scienza University-Hospital of Turin and Center for Cancer Prevention (CPO), 10126 Turin, Italy. 89Human Genetics Foundation (HuGeF), 10126 Torino, Italy. 90Andalusian School of Public Health, 18080 Granada, Spain. 91Instituto de Investigacion Biosanitaria de Granada (Granada.ips), 18012 Granada, Spain. 92International Agency for Research on Cancer, 69372 Lyon Cedex 08, France. 93Danish Cancer Society Research Center, 2100 Copenhagen, Denmark. 94ASP, 97100 Ragusa, Italy. 95Ragusa Cancer Registry, Aire Onlus, 97100 Ragusa, Italy. 96National Institute for Public Health and the Environment (RIVM), 3720 BA Bilthoven, The Netherlands. 97University Medical Center Utrecht, 3508 GA Utrecht, the Netherlands. 98School of Public Health, Imperial College London, W2 1PG, UK.
Generation Scotland
Blair H. Smith\textsuperscript{46,99}, Archie Campbell\textsuperscript{99}, Ian J. Deary\textsuperscript{99,100,101}, Andrew M. McIntosh\textsuperscript{99,102}

\textsuperscript{99}Centre for Genomic and Experimental Medicine, Institute of Genetics and Molecular Medicine, University of Edinburgh, Edinburgh EH4 2XU, UK.
\textsuperscript{100}Centre for Cognitive Ageing and Cognitive Epidemiology, University of Edinburgh, Edinburgh, UK. \textsuperscript{101}Department of Psychology, University of Edinburgh, EH8 9JZ, UK. \textsuperscript{102}Division of Psychiatry, University of Edinburgh, Edinburgh, EH10 5HF, UK.