Genome-wide association study of sexual maturation in males and females highlights a role for body mass and menarche loci in male puberty

Diana L. Cousminer[1], Evangelia Stergiakouli[3], Diane J. Berry[5], Wei Ang[6], Maria M. Groen-Blokhuis[7,8], Antje Körner[10], Niina Siitonen[11], Ioanna Ntalla[14,15], Marcella Marinelli[16,17,18], John R.B. Perry[19], Johannes Kettunen[1,20], Rick Jansen[8,9,21], Ida Surakka[1,20], Nicholas J. Timpson[3], Susan Ring[4], George Mcmahon[4], Chris Power[5], Carol Wang[6], Mika Kähönen[22], Jorma Viikari[23], Terho Lehtimäki[24], Christel M. Middeldorp[7,9,21], Hilleke E. Hulshoff Pol[25], Madlen Neef[10], Sebastian Weise[10], Katja Pahkala[11,12], Harri Niinikoski[11,13], Eleftheria Zeggini[26], Kalliope Panoutsopoulou[26], Mariona Bustamante[16,18,33], Brenda W.J.H. Penninx[8,9,21], the ReproGen Consortium[†], Joanne Murabito[27], Maties Torrent[18,34], George V. Dedoussis[14], Wieland Kiess[10], Dorret I. Boomsma[7,8,9], Craig E. Pennell[6], Olli T. Raitakari[11,28], Elina Hyppönen[5,29,35], George Davey Smith[3], Samuli Ripatti[1,2,26], Mark I. McCarthy[30,31,32] and Elisabeth Widén[1,*] for the Early Growth Genetics (EGG) Consortium[†]

[1]Institute for Molecular Medicine Finland (FIMM), [2]Hjelt Institute, University of Helsinki, Helsinki, Finland, [3]MRC Integrative Epidemiology Unit, [4]Avon Longitudinal Study of Parents and Children (ALSPAC), School of Social and Community Medicine, University of Bristol, Bristol, UK, [5]Centre for Paediatric Epidemiology and Biostatistics, UCL Institute of Child Health, London, UK, [6]School of Women's and Infants' Health, The University of Western Australia, Perth, WA, Australia, [7]Department of Biological Psychology, [8]EMGO+ Institute for Health and Care Research, [9]Neuroscience Campus Amsterdam, VU University, Amsterdam, the Netherlands, [10]Center of Pediatric Research, Department of Women's & Child Health, University of Leipzig, Leipzig, Germany, [11]Research Centre of Applied and Preventive Cardiovascular Medicine, [12]Paavo Nurmi Centre, Sports & Exercise Medicine Unit, Department of Physical Activity and Health, [13]Department of Pediatrics, University of Turku, Turku, Finland, [14]Department of Nutrition and Dietetics, Harokopio University of Athens, Athens, Greece, [15]Genetic Epidemiology Group, Department of Health Sciences, University of Leicester, Leicester, UK, [16]Center for Research in Environmental Epidemiology (CREAL), Barcelona, Catelonia, Spain, [17]Hospital del Mar Research Institute (IMIM), Barcelona, Catalonia, Spain, [18]Spanish Consortium for Research on Epidemiology and Public Health (CIBERESP), Barcelona, Catalonia, Spain, [19]MRC Epidemiology Unit, Institute of Metabolic Science, Addenbrooke's Hospital, Cambridge, UK, [20]Department of Chronic Disease Prevention, National Institute for Health and Welfare, Helsinki, Finland, [21]Department of Psychiatry, VU University Medical Center/GGZ inGeest, Amsterdam, The Netherlands, [22]Department of Clinical Physiology, Tampere University Hospital and University of Tampere School of Medicine, Tampere 33521, Finland, [23]Department of Medicine, University of Turku and Turku University Hospital, Turku 20521, Finland, [24]Department of Clinical Chemistry, Fimlab Laboratories and University of Tampere School of Medicine, Tampere 33520, Finland, [25]Department of Psychiatry, Brain Center Rudolf Magnus, University Medical Center Utrecht, Utrecht, the Netherlands, [26]Wellcome Trust Sanger Institute, Hinxton, UK, [27]Department of Medicine, Section of General Internal Medicine, Boston University School of Medicine, Boston, MA, USA,

---

*To whom correspondence should be addressed at: Institute for Molecular Medicine Finland (FIMM), University of Helsinki, Tukholmankatu 8, PO Box 20, Helsinki 00014, Finland. Tel: +358 503812738; Fax: +358 919125737; Email: elisabeth.widen@helsinki.fi
[†]See Supplementary Material for a full list of members.

[28]Department of Clinical Physiology and Nuclear Medicine, Turku University Hospital, Turku, Finland, [29]School of Population Health and Sansom Institute for Health Research, University of South Australia, North Terrace, Adelaide, Australia, [30]Oxford Centre for Diabetes, Endocrinology and Metabolism, University of Oxford, Churchill Hospital, Old Road, Headington, Oxford, UK, [31]Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, UK, [32]Oxford NIHR Biomedical Research Centre, Churchill Hospital, Oxford, UK, [33]Genes and genomes, Center for Genomic Regulation (CRG), Barcelona, Catalonia, Spain and [34]ib-salut, Area de Salut de Menorca, Balearic Islands, Spain [35]South Australian Health and Medical Research Institute, North Terrace, Adelaide, Australia

Little is known about genes regulating male puberty. Further, while many identified pubertal timing variants associate with age at menarche, a late manifestation of puberty, and body mass, little is known about these variants' relationship to pubertal initiation or tempo. To address these questions, we performed genome-wide association meta-analysis in over 11 000 European samples with data on early pubertal traits, male genital and female breast development, measured by the Tanner scale. We report the first genome-wide significant locus for male sexual development upstream of *myocardin-like 2* (*MKL2*) ($P = 8.9 \times 10^{-9}$), a menarche locus tagging a developmental pathway linking earlier puberty with reduced pubertal growth ($P = 4.6 \times 10^{-5}$) and short adult stature ($p = 7.5 \times 10^{-6}$) in both males and females. Furthermore, our results indicate that a proportion of menarche loci are important for pubertal initiation in both sexes. Consistent with epidemiological correlations between increased prepubertal body mass and earlier pubertal timing in girls, body mass index (BMI)-increasing alleles correlated with earlier breast development. In boys, some BMI-increasing alleles associated with earlier, and others with delayed, sexual development; these genetic results mimic the controversy in epidemiological studies, some of which show opposing correlations between prepubertal BMI and male puberty. Our results contribute to our understanding of the pubertal initiation program in both sexes and indicate that although mechanisms regulating pubertal onset in males and females may largely be shared, the relationship between body mass and pubertal timing in boys may be complex and requires further genetic studies.

INTRODUCTION

Puberty is an important milestone in life that marks the passage from childhood to reproductive maturity. It is a complex stage characterized by a defined sequence of developmental events, for example breast and genital development, menarche and the pubertal growth spurt, many of which are challenging to assess in an epidemiological setting. The timing of pubertal events is highly variable, with 4–6 years mean variation in onset within each sex and ∼2 years variation between the sexes (1–3). Pubertal traits are also highly heritable, with heritability estimates of up to 0.8–0.9 (4,5). Nonetheless, while altered pubertal timing is correlated with risk for the development of metabolic syndrome-related disorders such as obesity, diabetes and cardiovascular disease (6,7) as well as hormone-dependent cancer (8–10) later in life, the molecular machinery underlying pubertal variation still remains poorly understood.

Genome-wide association (GWA) studies of pubertal timing have mainly targeted females. The largest studies have assessed age at menarche (AAM), a late manifestation of puberty that is accurately recalled (11) and thus easy to assess. These studies revealed over 30 associated loci in European women (12–16), a proportion of which also associated with menarcheal timing in African Americans (17). In addition, two GWA studies of the pubertal growth spurt, utilizing longitudinal height measurements in both girls and boys, have pinpointed altogether five pubertal timing loci (18,19), four of which were associated in both

sexes. Since most pubertal timing loci have been identified in analyses of females, the genetic underpinnings of male pubertal timing are still underexplored. It also remains unknown to what extent menarche-associated loci may act on pubertal initiation, especially in males.

To identify genetic loci associated with the timing of pubertal onset, particularly in males, we chose to assess the earliest physical signs closely following the central activation of the hypothalamic–pituitary–gonadal (HPG) axis which triggers pubertal onset, namely genital enlargement in boys and breast development in girls, using visually assessed Tanner scale measurements ranging from pre-adolescent (Stage 1) to fully mature morphology (Stage 5) (2,3). Within the Early Growth Genetics (EGG) Consortium (20), we included European-based cohorts with Tanner genital and breast stage data in slightly >11 000 adolescents in a GWA study of these early pubertal traits, as well as an investigation into the overlap among variants that impact pubertal onset in both sexes.

RESULTS

GWA meta-analysis on male genital stage and follow-up genotyping reveals a novel male puberty locus at chr 16p13.12

Within the EGG Consortium, we performed GWA meta-analyses for Tanner genital stage in boys at age 12.6–15 years,

Table 1. GWA meta-analysis results for Tanner genital stage (males), Tanner breast stage (females) and males and females combined. All SNPs associating with $P < 1 \times 10^{-6}$ are reported

SNP	Chr	Pos[a]	Nearby gene	Effect/other allele	Effect allele freq	β	SE	P[b]	N	P gender het[c]
Males										
rs246185	16	14 302 933	*MKL2*	T/C	0.68	0.13	0.026	5.01×10^{-7}	3769	4.2×10^{-4}
Females										
rs9391253	6	105 474 309	*LIN28B*	T/A	0.33	-0.088	0.017	2.5×10^{-7}	6146	0.29
rs2095812	6	105 490 671	*LIN28B*	G/C	0.33	-0.088	0.017	2.6×10^{-7}	6147	0.28
rs395962	6	105 504 111	*LIN28B*	T/G	0.33	-0.088	0.017	3.0×10^{-7}	6146	0.28
rs314263	6	105 499 438	*LIN28B*	T/C	0.67	0.088	0.017	3.1×10^{-7}	6147	0.29
rs314276	6	105 514 692	*LIN28B*	C/A	0.66	0.087	0.017	3.7×10^{-7}	6146	0.37
rs167539	6	105 516 741	*LIN28B*	C/A	0.34	-0.086	0.017	4.3×10^{-7}	6147	0.38
rs7759938	6	105 485 647	*LIN28B*	T/C	0.68	0.088	0.018	5.6×10^{-7}	6145	0.34
rs314273	6	105 568 575	*LIN28B*	T/G	0.34	-0.085	0.017	6.8×10^{-7}	6146	0.43
rs1149336	1	7 454 877	*CAMTA1*	T/C	0.27	0.09	0.018	8.7×10^{-7}	6145	0.34
rs314268	6	105 524 671	*LIN28B*	G/A	0.34	-0.084	0.017	9.6×10^{-7}	6143	0.43
Combined										
rs314276	**6**	**105 514 692**	***LIN28B***	**C/A**	**0.66**	**0.08**	**0.014**	**1.6×10^{-8}**	**9915**	**0.37**
rs167539	6	105 516 741	*LIN28B*	C/A	0.34	-0.079	0.014	1.8×10^{-8}	9916	0.38
rs9391253	6	105 474 309	*LIN28B*	T/A	0.33	-0.079	0.014	2.3×10^{-8}	9915	0.29
rs314273	6	105 568 575	*LIN28B*	T/G	0.34	-0.079	0.014	2.3×10^{-8}	9915	0.43
rs2095812	6	105 490 671	*LIN28B*	G/C	0.33	-0.079	0.014	2.5×10^{-8}	9916	0.28
rs395962	6	105 504 111	*LIN28B*	T/G	0.33	-0.079	0.014	2.9×10^{-8}	9915	0.28
rs314263	6	105 499 438	*LIN28B*	T/C	0.67	0.078	0.014	3.0×10^{-8}	9916	0.29
rs7759938	6	105 485 647	*LIN28B*	T/C	0.68	0.08	0.014	3.4×10^{-8}	9914	0.34
rs314268	6	105 524 671	*LIN28B*	G/A	0.34	-0.078	0.014	3.7×10^{-8}	9912	0.43
rs369065	6	105 550 751	*LIN28B*	T/C	0.66	0.077	0.014	5.1×10^{-8}	9916	0.5
rs1149336	1	7 454 877	*CAMTA1*	T/C	0.27	0.081	0.015	7.2×10^{-8}	9914	0.34
rs1149332	1	7 451 433	*CAMTA1*	T/C	0.27	0.08	0.015	9.7×10^{-8}	9916	0.37

Markers passing a genome-wide significance threshold of 1.67×10^{-8} (adjusting for three GWA analyses) are marked in bold.
[a]Marker position reported according to genome build 36 and allele coding based on the positive strand.
[b]P-value adjusted for genomic control by GWAMA ([67],[68]) during meta-analysis.
[c]Test for heterogeneity among male and female effects performed by GWAMA during meta-analysis.

including up to 3769 samples from four cohorts. While there were no genome-wide significant signals after the primary meta-analysis, one marker at chromosome 16p13.12 (rs246185) suggestively associated at $P = 5.0 \times 10^{-7}$ (Table 1). When we included a further 208 samples from an additional cohort that had DNA available for follow-up genotyping, rs246185 surpassed genome-wide significance ($P = 8.9 \times 10^{-9}$) (Table 2b). Although our follow-up sample was limited, we found no evidence for heterogeneity of effects between the discovery and follow-up studies (Cochran's statistic $Q = 0.23$, $P = 0.99$; Supplementary Material, Fig. S1) and sample characteristics were similar between the discovery and follow-up study subjects (Supplementary Material, Table S1). Thus, we report the first genetic locus associated with male sexual development.

GWA meta-analysis of female breast stage shows overlap with leading menarche signals

With data available on Tanner breast development at the age of 10.5–12.5 years in 6147 samples from six cohorts, we also ran a female-specific GWA analysis. Furthermore, since girls enter puberty an average of 2 years prior to boys (1) these data provide an estimate of pubertal timing at a similar stage of development, and because the phenotypes were measured on the same scale in boys and girls, these measures of early puberty are directly comparable between the sexes. Thus, we also performed a combined-sex meta-analysis. The female-specific and combined-gender analyses both displayed deviation

from the expected P-value distribution under the null hypothesis (Supplementary Material, Fig. S2). However, only rs314276 ($P = 1.6 \times 10^{-8}$; $n = 9915$) near *LIN28B*, a locus previously implicated in the timing of menarche (16), breast development (12), the timing of the pubertal height growth spurt (18,19) and adult stature (21), reached genome-wide significance in the combined male and female analysis, corrected for analyzing three models ($P < 1.7 \times 10^{-8}$) (Table 1). Similar to previous findings, the association was more significant in females (β (SE) $= 0.087$ (0.017), $P = 3.7 \times 10^{-7}$) than in males (β (SE) $= 0.064$ (0.024), $P = 0.0085$). Two signals at a suggestive locus for males and females combined on chromosome 1 falling within the *CAMTA1* gene, rs1149336 and rs1149332, failed to reach genome-wide significance despite the addition of 1376 and 1305 additional samples, respectively (Table 2a).

Rs246185 is associated with advanced menarche and diminished pubertal growth

Further assessment of rs246185 with puberty and growth phenotypes showed strong association with traits that support a role for this locus in central GnRH activation in both males and females. Central activation of puberty by gonadotropin-releasing hormone (GnRH) release from the hypothalamus triggers a cascade of downstream events that culminate in manifestations of pubertal development such as the appearance of secondary sex characteristics, the onset of menstruation in girls and the pubertal growth spurt in both sexes (22). To assess the association

Table 2. Follow-up genotyping of suggestively associated signals. Novel signals with a *P*-value of $<1 \times 10^{-6}$ were genotyped or analyzed in additional cohorts. The *P*-value threshold for genome-wide significance is 1.67×10^{-8}

	Effect allele/other allele	Effect allele frequency	β	SE	P	N
a. Inclusion of further samples (*de novo* genotyped and GWA data) for two signals at the chromosome 1p36.23 locus at *CAMTA1*						
rs1149336						
Primary analysis	T/C	0.27	0.081	0.015	7.22×10^{-8}	9914
Additional samples	T/C	0.25	-0.025	0.032	0.44	1376
Combined	T/C	0.26	0.062	0.014	5.79×10^{-6}	11 290
rs1149332						
Primary analysis	T/C	0.27	0.08	0.015	9.66×10^{-8}	9916
Additional samples	T/C	0.25	-0.023	0.033	0.49	1305
Combined	T/C	0.26	0.062	0.014	5.09×10^{-6}	11 221
b. Inclusion of *de novo* genotyped samples for the chromosome 16p13.12 locus near *MKL2*						
rs246185						
Primary analysis	T/C	0.68	0.131	0.026	5.01×10^{-7}	3769
Additional samples	T/C	0.67	0.231	0.076	0.0025	208
Combined	**T/C**	**0.68**	**0.141**	**0.025**	$\mathbf{8.88 \times 10^{-9}}$	**3977**

Markers surpassing the genome-wide significant threshold are marked in bold.

Table 3. Association between rs246185 (T) and related puberty and growth traits

Trait	β	SE	P	N	Reference
Tanner breast stage at age 11–12	0.019	0.019	0.31	6144	NA
Tanner breast stage in 80th percentile of BMI	0.06	0.03	0.043	2637	NA
AAM[a]	-0.047	0.008	$\mathbf{1.03 \times 10^{-8}}$	up to 87 802	Elks 2010
10F[b]	-0.028	0.020	0.16	6974	Cousminer 2013
12M[b]	0.035	0.020	0.072	6983	Cousminer 2013
10F + 12M[b]	0.004	0.014	0.77	13 957	Cousminer 2013
PGF[b]	-0.062	0.021	**0.0027**	5756	Cousminer 2013
PGM[b]	-0.042	0.023	0.065	5043	Cousminer 2013
PGF + PGM[b]	-0.053	0.015	**0.00054**	10 799	Cousminer 2013
PTF[b]	-0.069	0.023	**0.003**	4946	Cousminer 2013
PTM[b]	-0.071	0.025	0.0052	4282	Cousminer 2013
PTF + PTM[b]	-0.070	0.017	$\mathbf{4.61 \times 10^{-5}}$	9228	Cousminer 2013
Adult stature[a]	-0.029	NA	$\mathbf{7.5 \times 10^{-6}}$	132 029	Lango Allen 2010

We queried the female breast meta-analysis results and previously published datasets for association of puberty and growth traits with rs246185. 10F, height SDS at age 10 years in females. 12M, height SDS at age 12 years in males. PGF, total pubertal growth in females (height SDS change between age 8 years and adult). PGM, total pubertal growth in males (height SDS change between age 8 years and adult). PTF, pubertal timing in females (height SDS change between age 14 years and adult). PTM, pubertal timing in males (height SDS change between age 14 years and adult). *P*-value threshold for significance is 0.004 (Bonferroni correction for 12 traits).
[a]Reported here for rs246185, although rs1659127, which is in high LD with rs246185 ($r^2 = 0.84$), is more strongly associated with AAM (16) and adult stature (21).
[b]For a detailed description of these models, see (18). Markers surpassing the genome-wide significant threshold are marked in bold.

between this locus and the timing of pubertal activation, we queried our GWA data on breast development as well as previously published data on menarcheal timing, the pubertal growth spurt, and adult stature (Table 3) for association with rs246185. Although no significant effect could be detected in our GWA data on girls' breast development at age 10.5–12.5 years despite a consistent direction of effect, rs246185 is in high LD ($r^2 = 0.84$) with a previously published menarche SNP (rs1659127) (16). Looking up our best SNP in the complete ReproGen dataset (16) showed that the T allele at rs246185 associated with 2.1 weeks earlier menarche ($P = 1.0 \times 10^{-8}$), consistent with our data showing association between the same allele and advanced genital development (higher Tanner score) in boys.

Furthermore, we found that the puberty-advancing allele also associated with decreased total pubertal growth (standardized height difference between age 8 years and adult, relative to the population mean), and decreased late pubertal growth (standardized height difference between age 14 years and adult, relative to the population mean) (18). The early cessation of growth during puberty results in shorter adult stature, evidenced by the significant association between rs1659127 and reduced final adult height ($P = 1.1 \times 10^{-11}$) in the large meta-analysis of adult stature performed by the GIANT Consortium (21). Our leading SNP also associated with adult stature, although to a lesser degree ($P = 7.5 \times 10^{-6}$). Thus, this locus tags a pathway which impacts adult stature predominantly through earlier pubertal timing. While adult height SNPs affect growth during multiple growth periods (18,19,23,24), rs246185 primarily affects growth during puberty. These findings are consistent with epidemiological observations which link early puberty to reduced pubertal growth and shorter adult stature (25,26).

Functional exploration of the genetic landscape around rs246185

We were unable to link rs246185 at the 16p13.12 locus to a gene in the region, although this variant falls in a predicted promoter region containing a cluster of transcription factor-binding sites. Expression quantitative trait locus (eQTL) analyses from whole blood (27,28) ($N > 5000$ (see Supplementary Material) (29), lymphoblastoid cell lines (30,31) and skin and adipose tissue (30) did not reveal any links between genotype at rs246185 and nearby gene expression levels (data not shown). However, studies of genes nearby height-associated SNPs indicate that the most likely candidate gene is the closest gene to a given variant (21,32). The nearest gene to rs246185, *MKL2*, falls 38 kb upstream from the SNP and is a transcriptional activator involved in cell differentiation and development. Still, other possible candidate genes lie in the region (Supplementary Material, Fig. S3 and Table S2).

To further investigate the region surrounding rs246185, we fine-mapped this region in the four cohorts which contributed to the primary male analysis [Avon Longitudinal Study of Parents and Children (ALSPAC), Western Australia Pregnancy Study (RAINE), Young Finns Study (YFS) and TEENs of Attica: genes and Environment (TEENAGE)] by imputing against the 1000 Genomes reference set (33). Only a handful of newly imputed markers showed evidence for association of similar strength as the leading GWA a SNP (Supplementary Material, Fig. S3). None of these variants affect nearby gene structure. However, rs193536, an SNP in high LD with rs246185 ($r^2 = 0.85$), had a slightly stronger association with male genital stage and is predicted to be likely to affect transcription factor binding by the RegulomeDB web tool (34) based on multiple lines of evidence including transcription factor-binding motifs, DNase footprints and DNase peaks. Furthermore, when we queried additional SNPs in high LD (here defined as $r^2 > 0.6$) with rs246185 but not present in the 1000 Genomes imputed data, we detected evidence for another nearby marker, rs74755650 ($r^2 = 0.75$), as potentially impacting transcription factor binding (Supplementary Material, Table S3). For each of these polymorphic sites, two transcription factor-binding motifs were predicted to be affected by variation in their target sequences based on position weight matrices and footprinting, which combines genome sequence information with experimental data to map bound TF-binding sites (35,36). Variation at rs193536 ($r^2 = 0.85$) is predicted to affect binding of PAX-3 and ER, and variation at rs74755650 ($r^2 = 0.75$) is predicted to affect binding of PATZ1 (also known as MAZR) and WT1, each supported by evidence from multiple cell types (35,36). However, examination of the consensus-binding motifs (Supplementary Material, Fig. S4) shows that the polymorphisms at the PATZ1 and PAX-3 target sequences would have greater impact on the ability of these factors to recognize their target sites and bind to DNA than the polymorphisms at the WT1- and ER-binding sites. For the PATZ1-binding motif, a G is strongly preferred to a T or A variant at rs74755650, while the PAX-3 transcription factor strongly prefers a C allele to an A allele at rs193536. Variation at these TF-binding sites may thus influence variability in nearby gene expression.

Overlap of menarche loci with Tanner association results

If a marker associates with both early and late manifestations of puberty, it can be assumed that the marker is tagging mechanisms that regulate the timing of pubertal onset. With GWA data on early markers of pubertal initiation, we evaluated what proportion of menarche signals tag the onset of puberty. With data also on male development, we were able to assess the overlap between regulators of male and female puberty. We extracted all significantly associated ($n = 32$; 12–16) and possible ($n = 12$; 16,18) menarche loci from each of the three Tanner meta-analysis results (Supplementary Material, Table S5).

While our data provided limited power (Supplementary Material, Table S4) to detect significant associations with much smaller sample sizes than were available to the ReproGen Consortium, we did observe deviation from the line of null association for menarche variants in both the male and female Tanner stage GWA meta-analysis results (Fig. 1). Furthermore, when we performed regression analyses comparing the association effect on pubertal timing between the menarche and Tanner associations, we found them to be highly consistent for the majority of loci (Fig. 1), with the menarche-advancing allele also associating with higher Tanner score and thus earlier development (females $P = 0.02$, males $P = 0.006$, combined $P = 0.0009$). For the combined-gender analysis, in addition to *LIN28B* and *MKL2*, the AAM locus at *TMEM38B* was significantly associated with Tanner pubertal stage ($P = 1.8 \times 10^{-5}$; significance threshold accounting for 44 tests = 0.001). Furthermore, rs3743266 at the gene *RORA*, previously listed as a possible menarche locus that did not surpass a stringent genome-wide significance threshold (16) but was associated with menarche in African-American women (17), was associated with Tanner stage in our data ($P = 0.0008$), providing further evidence for this signal as a true puberty locus (Supplementary Material, Table S5).

Overlap of adiposity loci with Tanner association results

There is a clear correlation between higher body mass index (BMI) and advanced pubertal timing in girls. While the relationship between body mass and pubertal timing remains controversial in boys, the majority of studies show obesity to be associated with earlier puberty, with a subset of overweight boys exhibiting a delay (37). Furthermore, it has been shown that pubertal development and childhood adiposity partially share a genetic basis (38), and loci which associate with both BMI and AAM have been explored (39).

To assess whether adiposity loci also associate with earlier markers of pubertal onset in both sexes, we performed a similar analysis, extracting known adiposity loci ($n = 31$ (40,41)) from the Tanner meta-analysis results. We again observed deviation from the null association in both sexes (Fig. 2) for BMI SNPs in the Tanner staging results. Regression analyses of BMI loci that are also associated with AAM (Fig. 2) showed that the majority of BMI-increasing alleles associated with earlier pubertal development, especially in girls (female $P = 0.019$), consistent with studies of BMI loci and menarcheal timing (16,39). Furthermore, rs3817334 (T) at *MTCH2*, a locus
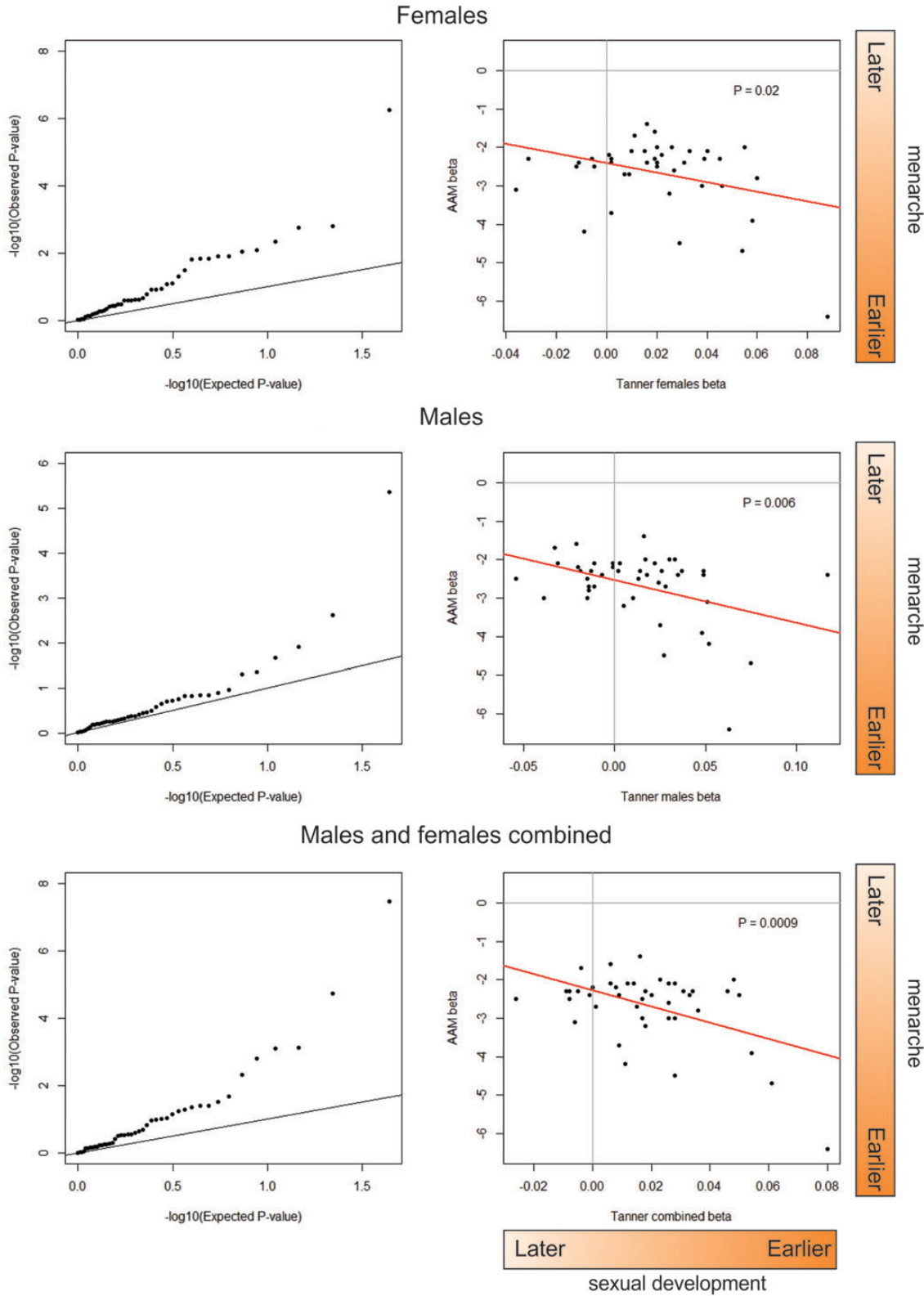
Figure 1. Menarche-advancing alleles in Tanner stage meta-analysis results. QQ plots for the *P*-values of AAM loci (16,18) in the female Tanner breast stage meta-analysis, male Tanner genital stage meta-analysis and male and female combined meta-analysis can be seen on the left. On the right, the Tanner analysis effect size is plotted for each menarche-advancing allele against the effect size on the timing of menarche [taken from Elks *et al.* (16) or Cousminer *et al* (18)]. *β*, SE and *P*-value for each marker are presented in more detail in the Supplementary Material, Table S5.
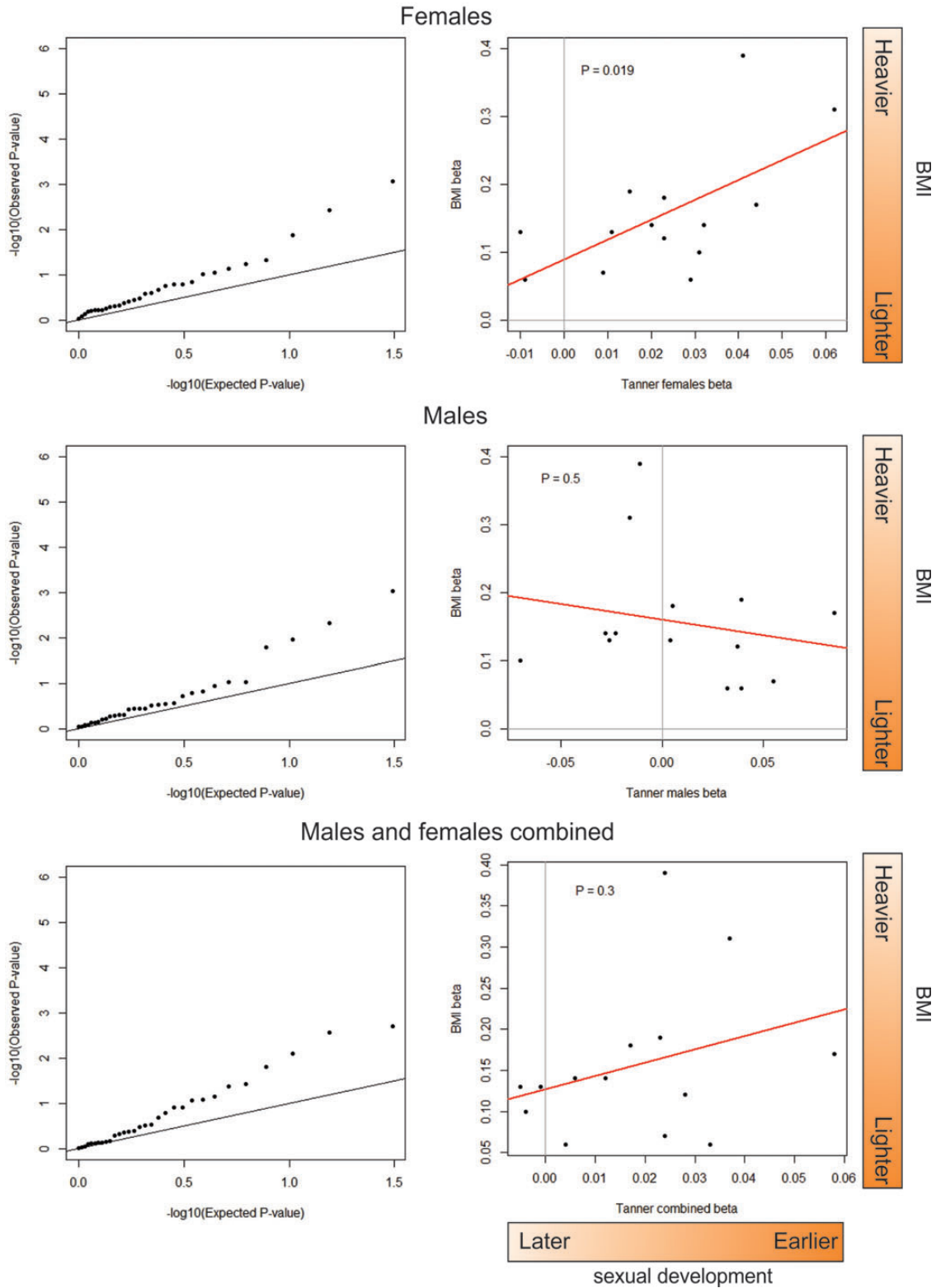
Figure 2. BMI-increasing alleles in Tanner stage meta-analysis results. QQ plots for alleles which are associated with BMI (39,40) are presented on the left for the female Tanner breast stage meta-analysis, male Tanner genital stage meta-analysis, and the combined male and female meta-analysis. On the right, scatter plots show the Tanner analysis effect size for BMI-increasing alleles that have also been associated with AAM (16,38). β, SE and P-value for each marker are presented in more detail in Supplementary Material, Table S6.

suggestively associated with decreased growth across puberty (18) but not the timing of menarche, here associated with advanced breast development ($P = 0.0008$; threshold accounting for testing 31 loci $= 0.0016$; sex difference $P = 0.003$) (Supplementary Material, Table S6).

Although the QQ plot for BMI loci in the male genital analysis showed a similar deviation from the line of null association as in females, the correlation between previously reported BMI-increasing effects and pubertal timing effects was different between boys and girls (Fig. 2; Supplementary Material, Table S6 and Fig. S5). While some body mass increasing loci showed the same trend towards association with earlier pubertal development, several of these alleles robustly displayed the opposite association. For example, the BMI-increasing allele (A) at rs571312 (*MC4R*), a locus not previously associated with puberty, was in our data associated with delayed male genital development ($P = 0.0009$; sex difference $P = 0.003$). Of the menarche-associated BMI alleles, the T allele at rs887912 (*FANCL*) also associated with delayed sexual development in boys, although the association was non-significant after correction for multiple testing ($P = 0.005$; sex difference $P = 0.004$).

MAGENTA pathway analysis

Finally, to get a glimpse of the biological pathways that are enriched for genes with markers associated with Tanner stage at lower than expected *P*-values, we entered the results from the GWA meta-analyses into MAGENTA to perform gene-set enrichment (GSE) analysis (42) (Supplementary Material, Table S7). For the combined male and female GWA results, pathways that were enriched for genes near or containing associated SNPs ($P < 0.001$, false discovery rate, FDR < 0.05) included aminopeptidase activity, hormone-sensitive lipase (HSL) mediated triacylglycerol hydrolysis and apoptosis. For the female analysis, enriched pathways included thyrotropin-releasing hormone (TRH) receptor signaling, cardiomyopathy, cancer (acute myeloid leukemia, endometrial cancer, prostate cancer and the PI3 pathway), as well as apoptosis and the regulation of the actin cytoskeleton. For males, only the protease inhibitor pathway surpassed an FDR of <0.05.

DISCUSSION

Male puberty is challenging to address in epidemiological and genetic studies due to the lack of an easy-to-assess non-invasive marker of pubertal development (43). While the ReproGen Consortium reported association between menarcheal timing and over 30 common genetic variants (16), male pubertal development has not been addressed in a similar setting. In this study, we pooled known European-based cohorts with Tanner staging data and genotype information within the EGG Consortium, amassing up to 11 290 study subjects in total for GWA and follow-up genotyping. With the ability to assess male genital staging, our study successfully uncovered the first robust genome-wide significant locus for male sexual development on chromosome 16 upstream of *MKL2*. This locus associates both with measurements of pubertal timing and height growth during puberty and can thus be added to a growing number of loci which tag a pathway linking earlier pubertal timing to decreased pubertal growth and reduced adult stature (18).

Although *MKL2* could not be definitively implicated as the causative gene as no eQTLs were found that would provide a direct link between polymorphisms at this locus and nearby gene expression, this may be partly due to the lack of eQTL-gene expression studies in tissues relevant for sexual maturation. While 1000 Genomes imputation and fine-mapping of the region surrounding our leading GWA SNP did not enrich the association signals at this locus, two transcription factor-binding sites, for PATZ1 and PAX-3, were predicted to be affected by polymorphisms in LD with rs246185, the leading associated SNP. While both PATZ1 and PAX-3 are compelling candidates as mediators of variation in the timing of sexual development as they are both important for morphological development during embryogenesis (44,45), PATZ1 is both a transcriptional repressor (46) and activator (47) that has been shown to have a critical role in spermatogenesis (48), embryonic and postnatal growth (44), and acts as a corepressor of androgen receptor-dependent transcription (49), which is important for normal puberty in both males and females (50). Furthermore, both male and female knock-out mice for the *Patz1* gene were infertile (48) (Mouse Genome Informatics database; http://www.informatics.jax.org/), indicating a potentially broad role for this factor in the sexual development of both sexes. Still, further study is needed to investigate whether rs74755650, the polymorphism predicted to affect PATZ1 binding, is the underlying functional variant at this locus, and through which nearby gene it may act.

While our study was unable to detect novel genetic associations for female breast development, our ability to assess a marker of early puberty in both genders revealed that menarche loci appear to be important for pubertal timing in both sexes and are thus tagging biological effects upstream of sex-specific mechanisms, as evidenced by the high correlation between menarche-advancing effects and earlier sexual development in both boys and girls. While the effects of some specific alleles may vary between the sexes (18) (Supplementary Material, Table S5), it appears that the overall genetic architecture regulating pubertal initiation is similar in both boys and girls. Furthermore, these results show a high degree of overlap of genetic variants on early manifestations of puberty as analyzed in the current study, and late pubertal manifestations, e.g. age of menarche, which is consistent with a recent study which found strong genetic overlap between genes involved in the first pubertal processes, such as increased hormone secretion, and the later development of secondary sex characteristics (51).

Additionally, with data on early puberty in both genders, we were able to perform GSE pathway analyses which revealed both pathways expected to be involved in pubertal onset and development and those which may lead to a greater understanding of the biology underlying puberty. For example, apoptosis, a hallmark of tissue remodeling, appeared multiple times as an enriched pathway (in KEGG, REACTOME, GOterm and Panther data sets). These analyses also picked up hypothalamus−pituitary pathways, such as the TRH receptor signaling pathway. Similar to GnRH, TRH is released from the hypothalamus into the pituitary, where it stimulates the release of thyroid-stimulating hormone and prolactin. Thyroid hormones are essential for normal sexual development (52). The presence of other pathways known to be important in puberty, such as

steroid hormone biosynthesis and the hormone biosynthetic process, in the secondary tier of results shows that multiple real signals beneath the significance threshold, and other, less expected pathways, may lead to new insights into the mechanisms regulating the initiation of puberty.

While recent studies found no evidence for sex-specific BMI effects (53), our data show that some BMI alleles display sex-specific associations to pubertal initiation. BMI-increasing loci correlated with advanced female breast development, as expected based on previous studies of BMI alleles and menarche (16,39) as well as epidemiological observations. In boys, despite controversy surrounding the possible correlation between overweight/obesity and pubertal timing, we found that BMI variants had lower *P*-values than expected based on a null hypothesis (Fig. 2) in male genital development. While the majority of BMI-increasing alleles also tended to associate with earlier sexual development as in girls (Supplementary Material, Fig. S5), specific alleles strongly showed the opposite association, providing the first suggestion of the genetic architecture underlying the epidemiological observation that a subset of overweight boys experience pubertal delay. Moreover, energy metabolism was picked up in the GSE analysis results for both sexes. For example, HSL-mediated triacylglycerol hydrolysis was enriched in the combined-gender analysis, although it is unclear whether this is a female-driven result. Triacylglycerol is an important energy store and provides cholesterol for steroid biosynthesis (54,55). Additionally, targeted disruption of HSL in mice resulted in male sterility (56), further evidence for this pathway as important in sexual development. Taken together, our results suggest a contrasting picture of the association between body mass and puberty between boys and girls and highlight the need for further in-depth genetic studies of the sex-specific correlations between energy metabolism and pubertal timing.

Only SNPs near *LIN28B* were associated with female breast development. Although high BMI (57) and self-assessment rather than clinician-assessment (58,59) both may interfere with an accurate evaluation of breast development, our limited ability to detect genome-wide significant associations is likely due to low power rather than an incorrectly assessed phenotype (Supplementary Material, Table S4). For example, when we excluded the top 20th percentile of BMI in contemporary cohorts and re-ran the Tanner breast stage association analysis, we observed only a slight attenuation of the association signal at the *LIN28B* locus (full dataset, rs314276 β (SE) = 0.09 (0.02), 20th percentile of BMI cutoff β (SE) = 0.06 (0.03)). Finally, we observe high consistency between menarche-advancing alleles (16) and association to advanced breast development (Fig. 1; Supplementary Material, Table S5), showing that our data are robust.

In conclusion, we report the first locus for male pubertal development near the gene *MKL2* on chromosome 16 and show that this locus associates with both the timing of sexual development and the pubertal height growth spurt. This study sheds light on the genetic overlap between male and female pubertal maturation and contributes to our understanding of the pubertal initiation program. Furthermore, we provide new insights into the genetic architecture of male puberty and show that although the leading association signals vary between the sexes, there is likely a substantial overlap of the molecular mechanisms that regulate pubertal initiation in males and females.

MATERIALS AND METHODS

Phenotypes and study subjects

This study was carried out with the cooperation of the EGG Consortium (20) (http://egg-consortium.org/). Tanner breast stage in girls or genital stage in boys was treated as a quantitative trait on a scale of 1–5. Girls enter puberty on average 2 years prior to boys (1), so we chose separate age ranges for boys and girls at which the Tanner scale data were approximately normally distributed.

Female study subjects with a Tanner breast stage assessment within 10.5–12.5 years of age were included in the female discovery set. Cohorts contributing to the female discovery analysis were the ALSPAC, 1958 British Birth Cohort (BC58-T1DGC and BC58-WTCCC), Cardiovascular Risk in YFS, Netherlands Twin Registry (NTR) and TEENAGE study. In total, 6147 girls were included in the primary analysis.

Male subjects with a Tanner genital stage assessment between 12.6 and 15 years of age were included in the male discovery analysis. Cohort studies that contributed to the male discovery analysis were the ALSPAC, Western Australia Pregnancy Study (RAINE), Cardiovascular Risk in YFS and TEENAGE study. A total of 3769 boys were included.

In some cohorts, Tanner stage was assessed by a clinician or trained researcher, and in others, Tanner stage was based on self-reports using pictures or schematic drawings. Tanner stage was measured by self-assessment in ALSPAC, RAINE and TEENAGE. Self-assessment or maternal assessment was performed in Infancia y Medio Ambiente. Studies that had Tanner stage assessed by a medical professional were B58C-WTCCC, B58C-T1DGC, YFS, the Leipzig Childhood Cohort (LEIPZIG) and the Special Turku Coronary Risk Factor Intervention Project (STRIP). In NTR, data collection were performed in several subprojects (see Supplementary Material, Table S1 on cohort details for more information on the specific questions asked and collection methods).

For the BMI-restricted Tanner breast stage analysis, females in the lower 80th percentile of BMI were included from each respective cohort. Contemporary studies that contributed to this analysis were the ALSPAC, Cardiovascular Risk in YFS and NTR. Leading SNPs from the *LIN28B* locus were then extracted and compared with the original female meta-analysis results. 2637 girls were included in this analysis.

Genotyping and quality control

Directly genotyped SNPs from high-density genotyping platforms (Illumina and Affymetrix) and imputed SNPs were analyzed for association (see Supplementary Material, Table S1). Prior to imputation, SNPs with a minor allele frequency (MAF) <1%, a call rate <95% and Hardy–Weinberg frequency of $P < 1 \times 10^{-6}$ were excluded. Further exclusions included duplicates, excess heterozygosity, non-European ancestry or ambiguous sex. Imputation was performed using MACH (60), IMPUTE (61,62) or BEAGLE (63) for ~2.5 million SNPs against HapMap Phase II (release 21/22). Prior to meta-analysis, the imputed SNPs were filtered to exclude

poorly imputed SNPs (IMPUTE filter PROPER INFO <0.4, MACH filter $r^2 < 0.3$ or Plink INFO < 0.3).

GWA analyses

Each cohort individually performed GWA using an additive model accounting for genotype imputation uncertainty for all autosomes by linear regression, adjusting for population substructure (the first principal components) if necessary and age at measurement (in years, to the nearest month, if available). Programs used to perform the GWA analyses were MACH2QTL (60,64), QUICKTEST (65), SnpTest (61), Plink (66) and Prob-ABEL (67).

Meta-analyses

Prior to meta-analysis, cohort-specific results were further filtered to exclude SNPs with MAF < 0.03. Meta-analysis of individual cohort results was performed using GWAMA (68,69) (http:// www.well.ox.ac.uk/gwama/index.shtml) version 2.0.5 for quantitative traits with two genomic control corrections enabled (on individual study results and on the meta-analysis summary output). Gender-differentiated and gender-heterogeneity analysis options were also enabled (see the GWAMA website for more details on these options).

Follow-up genotyping

Markers with a P-value of $< 5 \times 10^{-7}$ in the discovery analyses were selected for follow-up. Genotyped markers included rs246185 (*MKL2*), rs1149336 and rs1129332 (*CAMTA1*). GWA was performed in Infancia y Medio Ambiente (INMA), with the relevant markers extracted for follow-up analysis in females. *De novo* follow-up genotyping was performed in both male and female samples in STRIP and the LEIPZIG. Genotyping of STRIP samples was performed with TaqMan Pre-Designed SNP Genotyping Assays on the LightCycler 480 Real-Time PCR System (Roche) according to the manufacturer's instructions at the Finnish Genome Center to the Technology Center at FIMM (Helsinki, Finland) with a success rate of 0.94, 0.87 and 0.84. LEIPZIG sample genotyping was performed using TaqMan Pre-Designed SNP Genotyping Assay run on an Applied Biosystem 7500 Real Time PCR System and applying the autocaller algorithm of 7500 Software V.2.0.3 with calling rates of 0.98, 0.98 and 0.93, respectively.

Association of genotype with Tanner stage was performed by the contributing cohorts. The follow-up association results were then meta-analyzed together with the primary discovery summary statistics using linear regression in PLINK (66) v 1.07.

Expression quantitative trait loci

eQTLs were queried from existing datasets. Whole blood eQTLs were queried from the DILGOM study (27) and NTR and Netherlands Study of Depression and Anxiety (28) [NESDA (29)]. GENEVAR (70) (GENe Expression VARiation; http:// www.sanger.ac.uk/resources/software/genevar/) was used to look up data on rs246185 from lymphoblastoid cell lines (30,31) and skin and adipose tissue (31).

1000 Genomes imputation and meta-analysis of the Chr 16 locus

ALSPAC, YFS, RAINE and TEENAGE all had 1000 Genomes imputed data for chromosome 16 available. Imputation was performed with IMPUTE (61), MaCH (60,64) and Minimac (71). The association analysis was performed in each cohort using SNPtest V2 (61–63), mach2qtl (64,71) and QUICKTEST (65). After the cohort-specific association was performed, results were filtered based on imputation quality (PROPER INFO <0.4, $r^2 < 0.3$ or INFO < 0.3). Meta-analysis was then performed using GWAMA (68,69), and the 1 Mb region surrounding rs246185 was extracted.

Functional investigation of variants in LD with rs246185

SNPs nearby rs246185 were queried from the 1000 Genomes Pilot I dataset in the online SNAP proxy search v2.2 web tool (http:// www.broadinstitute.org/mpg/snap/ldsearch.php) with an r^2 threshold of 0.6. The resulting 12 SNPs (plus rs246185) were then entered into the RegulomeDB (33) webtool (http://www. regulome.stanford.edu/index), which annotates variants with known and predicted regulatory elements in intergenic genomic regions. Known and predicted regulatory elements include regions of DNAase hypersensitivity, transcription factor-binding sites and biochemically characterized promoter regions that affect transcription. These data are sourced from public datasets from GEO, the ENCODE project and other published literature.

Follow-up of rs246185

This marker was extracted from the female Tanner GWA analysis results as well as from previously published GWA studies on AAM by the ReproGen Consortium (16), three pubertal height growth phenotypes (18) and adult stature (21).

Overlap of published menarche and BMI SNPs with Tanner staging

32 previously published menarche (16) loci, 10 possible menarche loci, 2 pubertal growth loci with evidence for menarche-association (18) and 31 SNPs associated with adult (40) and/or childhood (41) BMI were extracted from the Tanner GWA discovery analysis results sets.

GSE analysis

GSE analysis was performed using MAGENTA (41) on the discovery-set results for males and females separately and combined. Default values were used for the number of gene-set permutations, for the upstream and lower boundaries surrounding each gene, and for the gene-set size limits. The GSE analysis was run for 75th and 95th percentiles. The genes in the human leukocyte antigen (HLA) region were removed prior to analysis. Pathways are reported that had a FDR of ≤ 0.05 and a nominal GSE $P < 1 \times 10^{-3}$.

SUPPLEMENTARY MATERIAL

Supplementary Material is available at *HMG* online.

REFERENCES

1. Palmert, M.R. and Boepple, P.A. (2001) Variation in the timing of puberty: clinical spectrum and genetic investigation. *J. Clin. Endocrinol. Metab.*, **86**, 2364–2368.
2. Marshall, W.A. and Tanner, J.M. (1969) Variations in pattern of pubertal changes in girls. *Arch. Dis. Child.*, **44**, 291–303.
3. Marshall, W.A. and Tanner, J.M. (1970) Variations in the pattern of pubertal changes in boys. *Arch. Dis. Child.*, **45**, 13–23.
4. Morris, D.H., Jones, M.E., Schoemaker, M.J., Ashworth, A. and Swerdlow, A.J. (2011) Familial concordance for age at menarche: analyses from the Breakthrough Generations Study. *Paediatr. Perinat. Epidemiol.*, **25**, 306–311.
5. Silventoinen, K., Haukka, J., Dunkel, L., Tynelius, P. and Rasmussen, F. (2008) Genetics of pubertal timing and its associations with relative weight in childhood and adult height: the Swedish Young Male Twins Study. *Pediatrics*, **121**, e885–e891.
6. Prentice, P. and Viner, R.M. (2013) 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.
7. Widén, E., Silventoinen, K., Sovio, U., Ripatti, S., Cousminer, D.L., Hartikainen, A.-L., Laitinen, J., Pouta, A., Kaprio, J., Järvelin, M.-R. *et al.* (2012) Pubertal timing and growth influences cardiometabolic risk factors in adult males and females. *Diabetes Care*, **35**, 850–856.
8. Golub, M.S., Collman, G.W., Foster, P.M.D., Kimmel, C.A., Rajpert-De Meyts, E., Reiter, E.O., Sharpe, R.M., Skakkebaek, N.E. and Toppari, J. (2008) Public health implications of altered puberty timing. *Pediatrics*, **121**, S218–S230.
9. Biro, F.M. and Deardorff, J. (2013) Identifying opportunities for cancer prevention during preadolescence and adolescence: puberty as a window of susceptibility. *J. Adolesc. Heal. Off. Publ. Soc. Adolesc. Med.*, **52**, S15–S20.
10. Mantovani, A. and Fucic, A. (2013) Puberty dysregulation and increased risk of disease in adult life: Possible modes of action. *Reprod. Toxicol. (Elmsford NY)*, doi: 10.1016/j.reprotox.2013.06.002.
11. Must, A., Phillips, S.M., Naumova, E.N., Blum, M., Harris, S., Dawson-Hughes, B. and Rand, W.M. (2002) Recall of early menstrual history and menarcheal body size: after 30 years, how well do women remember? *Am. J. Epidemiol.*, **155**, 672–679.
12. Ong, K.K., Elks, C.E., Li, S., Zhao, J.H., Luan, J., Andersen, L.B., Bingham, S.A., Brage, S., Smith, G.D., Ekelund, U. *et al.* (2009) Genetic variation in LIN28B is associated with the timing of puberty. *Nat. Genet.*, **41**, 729–733.
13. Sulem, P., Gudbjartsson, D.F., Rafnar, T., Holm, H., Olafsdottir, E.J., Olafsdottir, G.H., Jonsson, T., Alexandersen, P., Feenstra, B., Boyd, H.A. *et al.* (2009) Genome-wide association study identifies sequence variants on 6q21 associated with age at menarche. *Nat. Genet.*, **41**, 734–738.
14. He, C., Kraft, P., Chen, C., Buring, J.E., Paré, G., Hankinson, S.E., Chanock, S.J., Ridker, P.M., Hunter, D.J. and Chasman, D.I. (2009) Genome-wide association studies identify loci associated with age at menarche and age at natural menopause. *Nat. Genet.*, **41**, 724–728.
15. Perry, J.R.B., Stolk, L., Franceschini, N., Lunetta, K.L., Zhai, G., McArdle, P.F., Smith, A.V., Aspelund, T., Bandinelli, S., Boerwinkle, E. *et al.* (2009) Meta-analysis of genome-wide association data identifies two loci influencing age at menarche. *Nat. Genet.*, **41**, 648–650.
16. Elks, C.E., Perry, J.R.B., Sulem, P., Chasman, D.I., Franceschini, N., He, C., Lunetta, K.L., Visser, J.A., Byrne, E.M., Cousminer, D.L. *et al.* (2010) Thirty new loci for age at menarche identified by a meta-analysis of genome-wide association studies. *Nat. Genet.*, **42**, 1077–1085.
17. Demerath, E.W., Liu, C.-T., Franceschini, N., Chen, G., Palmer, J.R., Smith, E.N., Chen, C.T.L., Ambrosone, C.B., Arnold, A.M., Bandera, E.V. *et al.* (2013) Genome-wide association study of age at menarche in African-American women. *Hum. Mol. Genet.*, **22**, 3329–3346.
18. Cousminer, D.L., Berry, D.J., Timpson, N.J., Ang, W., Thiering, E., Byrne, E.M., Taal, H.R., Huikari, V., Bradfield, J.P., Kerkhof, M. *et al.* (2013) Genome-wide association and longitudinal analyses reveal genetic loci linking pubertal height growth, pubertal timing and childhood adiposity. *Hum. Mol. Genet.*, **22**, 2735–2747.
19. Widén, E., Ripatti, S., Cousminer, D.L., Surakka, I., Lappalainen, T., Järvelin, M.-R., Eriksson, J.G., Raitakari, O., Salomaa, V., Sovio, U. *et al.* (2010) Distinct variants at LIN28B influence growth in height from birth to adulthood. *Am. J. Hum. Genet.*, **86**, 773–782.
20. Freathy, R.M., Mook-Kanamori, D.O., Sovio, U., Prokopenko, I., Timpson, N.J., Berry, D.J., Warrington, N.M., Widen, E., Hottenga, J.J., Kaakinen, M. *et al.* (2010) Variants in ADCY5 and near CCNL1 are associated with fetal growth and birth weight. *Nat. Genet.*, **42**, 430–435.
21. Lango Allen, H., Estrada, K., Lettre, G., Berndt, S.I., Weedon, M.N., Rivadeneira, F., Willer, C.J., Jackson, A.U., Vedantam, S., Raychaudhuri, S. *et al.* (2010) Hundreds of variants clustered in genomic loci and biological pathways affect human height. *Nature*, **467**, 832–838.
22. Sisk, C.L. and Foster, D.L. (2004) The neural basis of puberty and adolescence. *Nat. Neurosci.*, **7**, 1040–1047.
23. Sovio, U., Bennett, A.J., Millwood, I.Y., Molitor, J., O'Reilly, P.F., Timpson, N.J., Kaakinen, M., Laitinen, J., Haukka, J., Pillas, D. *et al.* (2009) Genetic determinants of height growth assessed longitudinally from infancy to adulthood in the northern Finland birth cohort 1966. *PLoS Genet.*, **5**, e1000409.
24. Paternoster, L., Howe, L.D., Tilling, K., Weedon, M.N., Freathy, R.M., Frayling, T.M., Kemp, J.P., Smith, G.D., Timpson, N.J., Ring, S.M. *et al.* (2011) Adult height variants affect birth length and growth rate in children. *Hum. Mol. Genet.*, **20**, 4069–4075.
25. McIntyre, M.H. (2011) Adult stature, body proportions and age at menarche in the United States National Health and Nutrition Survey (NHANES) III. *Ann. Hum. Biol.*, **38**, 716–720.
26. McIntyre, M.H. and Kacerosky, P.M. (2011) Age and size at maturity in women: a norm of reaction? *Am. J. Hum. Biol. Off. J. Hum. Biol. Counc.*, **23**, 305–312.
27. Inouye, M., Kettunen, J., Soininen, P., Silander, K., Ripatti, S., Kumpula, L.S., Hämäläinen, E., Jousilahti, P., Kangas, A.J., Männistö, S. *et al.* (2010) Metabonomic, transcriptomic, and genomic variation of a population cohort. *Mol. Syst. Biol.*, **6**, 441.
28. Jansen, R., Batista, S., Brooks, A.I., Tischfield, J.A., Willemsen, G., van Grootheest, G., Hottenga, J.-J., Milaneschi, Y., Mbarek, H., Madar, V. *et al.* (2014) Sex differences in the human peripheral blood transcriptome. *BMC Genomics*, **15**, 33.
29. Wright, F., Sullivan, P., Brooks, A., Zou, F., Sun, W., Xia, K., Madar, V., Jansen, R., Chung, W., Zhou, Y. *et al.* (2014) Heritability and genomics of gene expression in peripheral blood. *Nature Genetics*, in press.
30. Stranger, B.E., Montgomery, S.B., Dimas, A.S., Parts, L., Stegle, O., Ingle, C.E., Sekowska, M., Smith, G.D., Evans, D., Gutierrez-Arcelus, M. *et al.*

(2012) Patterns of cis regulatory variation in diverse human populations. *PLoS Genet.*, **8**, e1002639.

31. Nica, A.C., Parts, L., Glass, D., Nisbet, J., Barrett, A., Sekowska, M., Travers, M., Potter, S., Grundberg, E., Small, K. *et al.* (2011) The architecture of gene regulatory variation across multiple human tissues: the MuTHER study. *PLoS Genet.*, **7**, e1002003.

32. Lui, J.C., Nilsson, O., Chan, Y., Palmer, C.D., Andrade, A.C., Hirschhorn, J.N. and Baron, J. (2012) Synthesizing genome-wide association studies and expression microarray reveals novel genes that act in the human growth plate to modulate height. *Hum. Mol. Genet.*, **21**, 5193–5201.

33. Consortium, T.E.P. (2012) An integrated encyclopedia of DNA elements in the human genome. *Nature*, **489**, 57–74.

34. Boyle, A.P., Hong, E.L., Hariharan, M., Cheng, Y., Schaub, M.A., Kasowski, M., Karczewski, K.J., Park, J., Hitz, B.C., Weng, S. *et al.* (2012) Annotation of functional variation in personal genomes using RegulomeDB. *Genome Res.*, **22**, 1790–1797.

35. Pique-Regi, R., Degner, J.F., Pai, A.A., Gaffney, D.J., Gilad, Y. and Pritchard, J.K. (2011) Accurate inference of transcription factor binding from DNA sequence and chromatin accessibility data. *Genome Res.*, **21**, 447–455.

36. Matys, V., Kel-Margoulis, O.V., Fricke, E., Liebich, I., Land, S., Barre-Dirrie, A., Reuter, I., Chekmenev, D., Krull, M., Hornischer, K. *et al.* (2006) TRANSFAC and its module TRANSCompel: transcriptional gene regulation in eukaryotes. *Nucleic Acids Res.*, **34**, D108–D110.

37. Wagner, I.V., Sabin, M.A., Pfäffle, R.W., Hiemisch, A., Sergeyev, E., Körner, A. and Kiess, W. (2012) Effects of obesity on human sexual development. *Nat. Rev. Endocrinol.*, **8**, 246–254.

38. Johnson, W., Choh, A.C., Curran, J.E., Czerwinski, S.A., Bellis, C., Dyer, T.D., Blangero, J., Towne, B. and Demerath, E.W. (2013) Genetic risk for earlier menarche also influences peripubertal body mass index. *Am. J. Phys. Anthropol.*, **150**, 10–20.

39. Fernández-Rhodes, L., Demerath, E.W., Cousminer, D.L., Tao, R., Dreyfus, J.G., Esko, T., Smith, A.V., Gudnason, V., Harris, T.B., Launer, L. *et al.* (2013) Association of Adiposity Genetic Variants With Menarche Timing in 92,105 Women of European Descent. *Am. J. Epidemiol.*, doi: 10.1093/aje/kws473.

40. Speliotes, E.K., Willer, C.J., Berndt, S.I., Monda, K.L., Thorleifsson, G., Jackson, A.U., Lango Allen, H., Lindgren, C.M., Luan, J., Mägi, R. *et al.* (2010) Association analyses of 249,796 individuals reveal 18 new loci associated with body mass index. *Nat. Genet.*, **42**, 937–948.

41. Bradfield, J.P., Taal, H.R., Timpson, N.J., Scherag, A., Lecoeur, C., Warrington, N.M., Hypponen, E., Holst, C., Valcarcel, B., Thiering, E. *et al.* (2012) A genome-wide association meta-analysis identifies new childhood obesity loci. *Nat. Genet.*, **44**, 526–531.

42. Segrè, A.V., DIAGRAM Consortium, MAGIC investigators, Groop, L., Mootha, V.K., Daly, M.J. and Altshuler, D. (2010) Common inherited variation in mitochondrial genes is not enriched for associations with type 2 diabetes or related glycemic traits. *PLoS Genet.*, **6**, doi: 10.1371/journal.pgen.1001058.

43. Tinggaard, J., Mieritz, M.G., Sørensen, K., Mouritsen, A., Hagen, C.P., Aksglaede, L., Wohlfahrt-Veje, C. and Juul, A. (2012) The physiology and timing of male puberty. *Curr. Opin. Endocrinol. Diabetes Obes.*, **19**, 197–203.

44. Valentino, T., Palmieri, D., Vitiello, M., Simeone, A., Palma, G., Arra, C., Chieffi, P., Chiariotti, L., Fusco, A. and Fedele, M. (2013) Embryonic defects and growth alteration in mice with homozygous disruption of the Patz1 gene. *J. Cell. Physiol.*, **228**, 646–653.

45. Buckingham, M. and Relaix, F. (2007) The role of Pax genes in the development of tissues and organs: Pax3 and Pax7 regulate muscle progenitor cell functions. *Annu. Rev. Cell Dev. Biol.*, **23**, 645–673.

46. Fedele, M., Benvenuto, G., Pero, R., Majello, B., Battista, S., Lembo, F., Vollono, E., Day, P.M., Santoro, M., Lania, L. *et al.* (2000) A novel member of the BTB/POZ family, PATZ, associates with the RNF4 RING finger protein and acts as a transcriptional repressor. *J. Biol. Chem.*, **275**, 7894–7901.

47. Kobayashi, A., Yamagiwa, H., Hoshino, H., Muto, A., Sato, K., Morita, M., Hayashi, N., Yamamoto, M. and Igarashi, K. (2000) A combinatorial code for gene expression generated by transcription factor Bach2 and MAZR (MAZ-related factor) through the BTB/POZ domain. *Mol. Cell. Biol.*, **20**, 1733–1746.

48. Fedele, M., Franco, R., Salvatore, G., Paronetto, M.P., Barbagallo, F., Pero, R., Chiariotti, L., Sette, C., Tramontano, D., Chieffi, G. *et al.* (2008) PATZ1 gene has a critical role in the spermatogenesis and testicular tumours. *J. Pathol.*, **215**, 39–47.

49. Pero, R., Lembo, F., Palmieri, E.A., Vitiello, C., Fedele, M., Fusco, A., Bruni, C.B. and Chiariotti, L. (2002) PATZ attenuates the RNF4-mediated enhancement of androgen receptor-dependent transcription. *J. Biol. Chem.*, **277**, 3280–3285.

50. Walters, K.A., Simanainen, U. and Handelsman, D.J. (2010) Molecular insights into androgen actions in male and female reproductive function from androgen receptor knockout models. *Hum. Reprod. Update*, **16**, 543–558.

51. Koenis, M.M.G., Brouwer, R.M., van Baal, G.C.M., van Soelen, I.L.C., Peper, J.S., van Leeuwen, M., Delemarre-van de Waal, H.A., Boomsma, D.I. and Hulshoff Pol, H.E. (2013) Longitudinal study of hormonal and physical development in young twins. *J. Clin. Endocrinol. Metab.*, **98**, E518–E527.

52. Weber, G., Vigone, M.C., Stroppa, L. and Chiumello, G. (2003) Thyroid function and puberty. *J. Pediatr. Endocrinol. Metab. JPEM*, **16**(Suppl. 2), 253–257.

53. Randall, J.C., Winkler, T.W., Kutalik, Z., Berndt, S.I., Jackson, A.U., Monda, K.L., Kilpeläinen, T.O., Esko, T., Mägi, R., Li, S. *et al.* (2013) Sex-stratified genome-wide association studies including 270,000 individuals show sexual dimorphism in genetic loci for anthropometric traits. *PLoS Genet.*, **9**, e1003500.

54. Holm, C., Osterlund, T., Laurell, H. and Contreras, J.A. (2000) Molecular mechanisms regulating hormone-sensitive lipase and lipolysis. *Annu. Rev. Nutr.*, **20**, 365–393.

55. Kraemer, F.B. and Shen, W.-J. (2006) Hormone-sensitive lipase knockouts. *Nutr. Metab.*, **3**, 12.

56. Osuga, J., Ishibashi, S., Oka, T., Yagyu, H., Tozawa, R., Fujimoto, A., Shionoiri, F., Yahagi, N., Kraemer, F.B., Tsutsumi, O. *et al.* (2000) Targeted disruption of hormone-sensitive lipase results in male sterility and adipocyte hypertrophy, but not in obesity. *Proc. Natl. Acad. Sci. USA*, **97**, 787–792.

57. Bonat, S., Pathomvanich, A., Keil, M.F., Field, A.E. and Yanovski, J.A. (2002) Self-assessment of pubertal stage in overweight children. *Pediatrics*, **110**, 743–747.

58. Rollof, L. and Elfving, M. (2012) Evaluation of self-assessment of pubertal maturation in boys and girls using drawings and orchidometer. *J. Pediatr. Endocrinol. Metab.*, **25**, 125–129.

59. Desmangles, J.-C., Lappe, J.M., Lipaczewski, G. and Haynatzki, G. (2006) Accuracy of pubertal Tanner staging self-reporting. *J. Pediatr. Endocrinol. Metab.*, **19**, 213–221.

60. Li, Y., Willer, C., Sanna, S. and Abecasis, G. (2009) Genotype imputation. *Annu. Rev. Genomics Hum. Genet.*, **10**, 387–406.

61. Marchini, J., Howie, B., Myers, S., McVean, G. and Donnelly, P. (2007) A new multipoint method for genome-wide association studies by imputation of genotypes. *Nat. Genet.*, **39**, 906–913.

62. Howie, B.N., Donnelly, P. and Marchini, J. (2009) A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLoS Genet.*, **5**, e1000529.

63. Browning, S.R. and Browning, B.L. (2007) Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering. *Am. J. Hum. Genet.*, **81**, 1084–1097.

64. Li, Y., Willer, C.J., Ding, J., Scheet, P. and Abecasis, G.R. (2010) MaCH: using sequence and genotype data to estimate haplotypes and unobserved genotypes. *Genet. Epidemiol.*, **34**, 816–834.

65. Kutalik, Z., Johnson, T., Bochud, M., Mooser, V., Vollenweider, P., Waeber, G., Waterworth, D., Beckmann, J.S. and Bergmann, S. (2011) Methods for testing association between uncertain genotypes and quantitative traits. *Biostatistics*, **12**, 1–17.

66. Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A.R., Bender, D., Maller, J., Sklar, P., de Bakker, P.I.W., Daly, M.J. *et al.* (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.*, **81**, 559–575.

67. Aulchenko, Y.S., Struchalin, M.V. and van Duijn, C.M. (2010) ProbABEL package for genome-wide association analysis of imputed data. *BMC Bioinformatics*, **11**, 134.

68. Magi, R., Lindgren, C.M. and Morris, A.P. (2010) Meta-analysis of sex-specific genome-wide association studies. *Genet. Epidemiol.*, **34**, 846–853.

69. Mägi, R. and Morris, A.P. (2010) GWAMA: software for genome-wide association meta-analysis. *BMC Bioinformatics*, **11**, 288.

70. Yang, T.-P., Beazley, C., Montgomery, S.B., Dimas, A.S., Gutierrez-Arcelus, M., Stranger, B.E., Deloukas, P. and Dermitzakis, E.T. (2010) Genevar: a database and Java application for the analysis and visualization of SNP-gene associations in eQTL studies. *Bioinforma. Oxf. Engl.*, **26**, 2474–2476.

71. Chen, W.-M. and Abecasis, G.R. (2007) Family-based association tests for genomewide association scans. *Am. J. Hum. Genet.*, **81**, 913–926.