Shared genetic aetiology of puberty timing between sexes and with health-related outcomes

Felix R. Day¹, Brendan Bulik-Sullivan²,³,⁴, David A. Hinds⁵, Hilary K. Finucane⁶,⁷, Joanne M. Murabito⁸,⁹, Joyce Y. Tung⁵, Ken K. Ong¹,¹⁰,* & John R.B. Perry¹,*

Understanding of the genetic regulation of puberty timing has come largely from studies of rare disorders and population-based studies in women. Here, we report the largest genomic analysis for puberty timing in 55,871 men, based on recalled age at voice breaking. Analysis across all genomic variants reveals strong genetic correlation (0.74, \(P = 2.7 \times 10^{-70}\)) between male and female puberty timing. However, some loci show sex-divergent effects, including directionally opposite effects between sexes at the SIM1/MCHR2 locus (\(P_{\text{heterogeneity}} = 1.6 \times 10^{-12}\)). We find five novel loci for puberty timing (\(P < 5 \times 10^{-8}\)), in addition to nine signals in men that were previously reported in women. Newly implicated genes include two retinoic acid-related receptors, RORB and RXRA, and two genes reportedly disrupted in rare disorders of puberty, LEPR and KAL1. Finally, we identify genetic correlations that indicate shared aetiologies in both sexes between puberty timing and body mass index, fasting insulin levels, lipid levels, type 2 diabetes and cardiovascular disease.

¹MRC Epidemiology Unit, University of Cambridge School of Clinical Medicine, Box 285 Institute of Metabolic Science, Cambridge Biomedical Campus, Cambridge CB2 0QQ, UK. ²Stanley Center for Psychiatric Research, Broad Institute of MIT and Harvard, Cambridge, Massachusetts 02142, USA. ³Analytic and Translational Genetics Unit, Department of Medicine, Massachusetts General Hospital, Boston, Massachusetts 02114, USA. ⁴Medical and Population Genetics, Broad Institute, Cambridge, Massachusetts 02142, USA. ⁵23andMe Inc., 899 W. Evelyn Avenue, Mountain View, California 94041, USA. ⁶Department of Epidemiology, Harvard School of Public Health, Boston, Massachusetts 02115, USA. ⁷Department of Mathematics, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA. ⁸NHBLI’s and Boston University’s Framingham Heart Study, Framingham, Massachusetts 01702-5827, USA. ⁹Boston University School of Medicine, Department of Medicine, Section of General Internal Medicine, Boston, Massachusetts 02118, USA. ¹⁰Department of Paediatrics, University of Cambridge, Cambridge CB2 0QQ, UK. * These authors contributed equally to this work. Correspondence and requests for materials should be addressed to J.R.B.P. (email: john.perry@mrc-epid.cam.ac.uk).
Voice breaking describes the drop in resonant frequency due to elongation of the larynx in response to androgen exposure. It is a distinct developmental milestone that occurs during late puberty in males (typically between Tanner stages 3 to 4), and age at voice breaking therefore represents a non-invasive marker for the study of puberty timing in men.

Age at menarche, the onset of the first menstrual bleed, is a similar marker of pubertal timing in females and has been more widely studied. Age at menarche in women has been associated with a wide range of disease risks, and previous genome-wide association studies (GWASs) have reported over 100 common loci and five low-frequency coding variants, implicating several previously unsuspected mechanisms involved in the regulation of puberty timing, including post-transcriptional microRNA repression, gamma-aminobutyric acid-B (GABA-B) receptor signalling and nuclear hormone signalling. Those menarche loci were reportedly enriched for variants also associated with puberty timing in boys, but those analyses were limited by the small number of boys with assessment of physical characteristics of puberty.

Consequently, our understanding of the regulation of puberty timing in males is derived in large part from studies of rare disorders of puberty. To date, ~20 genes have been implicated in abnormally delayed or absent puberty, including normosmic or anosmic hypogonadotropic hypogonadism (Kallmann syndrome), while only three genes have been implicated in precocious puberty. Here, we report the first large-scale GWAS of puberty timing in males, based on recalled age at voice breaking in men in the 23andMe study. By combination with data on recalled age at menarche in women, we show that the genetic architecture of puberty timing has a substantial shared component between males and females, which also overlaps the genetic basis of several health-related traits and diseases.

Results
Genome-wide association signals for age at voice breaking. We identified 11 independent genome-wide significant (P < 5 × 10⁻⁸) signals for age at voice breaking in men located at nine genomic loci (Table 1). Of these, nine signals (mapping to seven loci: in/near LIN28B, MKL2, BSX, TMEM38B, NR4A2, IGSF1 and ALMS1) are correlated (r² > 0.05) with reported loci for age at menarche in women. The two novel signals for puberty timing are located in/near LEPR and KAL1, both of which are disrupted in rare disorders of puberty. The strongest common signal (minor allele frequency (MAF) > 5%) for voice breaking was at the LIN28B locus (rs991253, P = 8 × 10⁻²⁵) where three independent signals were identified (Table 1), which is consistent with the allelic heterogeneity at this locus reported for age at menarche. All signals are common variants, except for a rare (MAF = 1%) intergenic variant near ALMS1, which is associated with 0.32 year per allele later age at voice breaking, and is highly correlated with a rare non-synonymous variant in ALMS1 (rs45501594, T3542S, r² = 0.83) that is reportedly associated with age at menarche in women.

Genetic overlap between puberty timing in men and women. To estimate the shared genetic aetiology between timing of puberty in men and women, we used LD Score Regression to calculate the genome-wide genetic correlation (r_g) between age at voice breaking in men and age at menarche in women. The observed strong positive correlation (r_g = 0.74, P = 2.7 × 10⁻⁷⁰) indicates that many variants have similar influences on puberty timing in males and females. This sex concordance is illustrated by comparison of the effect sizes of all genome-wide significant loci for age at menarche/voice breaking (Fig. 1). Only six of the 123 reported age at menarche loci show significant sex-discordant effects on puberty timing (P_heterogeneity < 4.0 × 10⁻⁸); these were rs889122-OLFM2, rs466639-RXRG, rs2688325-CSMD1, rs1254337-SIX6, rs6555855-SLIT3 and rs9321659-SIM1/MCHR2. Notably, at the SIM1/MCHR2 locus the reported menarche-age raising allele is associated with younger age at voice breaking in men (P = 7.9 × 10⁻⁵, P_heterogeneity = 1.6 × 10⁻¹²).

Confirmation of novel puberty timing signals. Given the strong overall genetic correlation between puberty timing loci in men and women, and the paucity of male puberty timing data available in similarly sized studies, we sought confirmation of novel puberty timing loci identified in men using GWAS data on age at menarche in women (from a combination of publicly available HapMap2-imputed data from 182,416 women in the ReproGen consortium and 1,000-Genomes-imputed data from 76,831 additional women in the 23andMe study). Both of the novel genome-wide significant signals for age at voice breaking in men (in/near LEPR and KAL1) show directionally concordant associations with age at menarche in women (Table 1, LEPR P = 1.85 × 10⁻⁵; KAL1 2.4 × 10⁻⁴). However, the LEPR signal (rs140410685 a 3 bp indel) shows a threefold larger effect on puberty timing in men than in women (P_heterogeneity = 3.7 × 10⁻⁸).

We next attempted to confirm sub-genome-wide significant signals for age at voice breaking (5 × 10⁻⁸ < P < 1 × 10⁻⁶). Of the five signals at this threshold, three were strongly correlated with reported signals for age at menarche, showing directionally concordant associations with puberty timing in both sexes (rs10980922-ZNF483, rs2282752-WDR6 and rs6681737-NR5A2). The other two signals (rs9350100-RNF144B/ID4 and rs6506352-RORB) showed directionally discordant associations with age at menarche (P < 0.01), but with significant heterogeneity between sexes (Table 1). Given this heterogeneity we meta-analysed the two estimates in a random-effects model. Variants at the novel RNF144B-ID4 locus reached genome-wide significance for puberty timing in men and women combined. Although rs6506352 did not reach significance in this meta-analysis (P = 2.1 × 10⁻⁷), an uncorrelated variant (rs4237264, r² ~ 0) at this RORB locus was previously reported as a possible signal for age at menarche (P = 9 × 10⁻⁶). In a combined meta-analysis, we robustly confirmed rs4237264 as a novel signal for puberty timing in men and women (P = 2 × 10⁻¹¹).

RORB encodes retinoic acid receptor (RAR)-related orphan receptor beta; notably RORA and RXRG (encoding retinoid X receptor gamma) were previously implicated in age at menarche. We therefore tested single nucleotide polymorphisms (SNPs) within 500 kb of the remaining six RAR, RAR-related and retinoid X receptor (RXR) encoding genes for associations with puberty timing in our pooled sample of men and women, identifying one additional novel signal ~350 kb downstream of RXRA (rs416390, P = 2 × 10⁻⁶) and a further suggestive signal ~330 kb upstream of RXRB (rs241438, P = 5 × 10⁻⁶). In aggregate, SNPs in or near a reported list of nuclear hormone receptor genes that contain these RAR-related and RXR genes are significantly enriched for associations with age at voice breaking in men (P = 7 × 10⁻³), as reported for age at menarche in women.

Genetic correlation between puberty timing and other traits. To inform the likely aetiological relevance of puberty timing in men and women to other health-related outcomes, we used LD Score Regression to test their genetic correlations (r_g) with 27 other traits or complex diseases. There is no significant heterogeneity observed in genetic correlations between men and women.
women. In men and women combined, significant genetic correlations (conservatively adjusted for multiple testing: $P < 1.85 \times 10^{-3}$ (≈ 0.05/27)) are observed between puberty timing and nine other traits or disease outcomes (Table 2). The strongest genetic correlation is with body mass index (BMI; $r_g = -0.34$, $P = 4.6 \times 10^{-10}$). Further inverse genetic correlations are observed with polycystic ovary syndrome, fasting insulin levels, type 2 diabetes, triglyceride levels, cardiovascular disease and bone mineral density; and positive genetic correlations are observed with high-density lipoprotein cholesterol levels and adult height.

**Discussion**

We report a large genetic study of puberty timing in males and females. Our findings are the first to quantify the strongly shared genetic basis for puberty timing between sexes, and this is consistent with the largely sex-concordant effects of disruptive mutations in rare disorders of puberty.8,9. Accordingly, our findings support the validity of recalled age at voice breaking in men as a marker of puberty timing in epidemiological studies, consistent with the informative prospective assessment of this phenotype.8

Independent signals at seven loci previously identified for age at menarche in women5 (including three independent signals at the LIN28B locus) passed the genome-wide statistical significance threshold for age at voice breaking in men (Table 1). These included the strongest reported common and low-frequency signals for age at menarche at LIN28B and ALMS1, respectively, and other signals with relatively large effects. This observation is consistent with the largely shared genetic architecture for puberty timing between sexes.

The overlapping genetic architecture for puberty timing in men and women provided the rationale for a pooled meta-analysis across the sexes. Notably, we identified two novel signals, near RORB and RXRA, which add to the reported signals near to other retinoic acid receptor-related genes, RORA and RXRG5, and a fifth signal (near RXRB) showed sub-genome-wide significant association with puberty timing. The retinoic acid receptor-related and retinoid X receptors function as transcription factors that dimerize and regulate nuclear receptors to influence cell differentiation, development, circadian rhythm and metabolism13. Their receptor partners include the canonical receptors for oestrogen and androgens among other hormones and metabolites14, and their conformational changes may alter receptor sensitivity15. Collectively these findings strengthen the evidence for an aetiological role of retinoic acid and retinoid receptors in the regulation of puberty timing, although their relative importance to male versus female puberty remains to be established. Further studies are also needed to identify functional links between these allelic signals and specific gene and protein functions.

Three additional novel signals reached genome-wide significance for puberty timing, represented by variants in/near RNF144B-ID4, LEPR and KAL1. All three were associated with puberty timing in both men and women, two of which had stronger effects in males (LEPR and RNF144B-ID4). The signal at 6p22.3 resides in a gene desert with the two nearest genes, type 2 diabetes, triglyceride levels, cardiovascular disease and bone mineral density; and positive genetic correlations are observed with high-density lipoprotein cholesterol levels and adult height. Notably, we identified two novel signals, near RORB and RXRA, which add to the reported signals near to other retinoic acid receptor-related genes, RORA and RXRG5, and a fifth signal (near RXRB) showed sub-genome-wide significant association with puberty timing. The retinoic acid receptor-related and retinoid X receptors function as transcription factors that dimerize and regulate nuclear receptors to influence cell differentiation, development, circadian rhythm and metabolism13. Their receptor partners include the canonical receptors for oestrogen and androgens among other hormones and metabolites14, and their conformational changes may alter receptor sensitivity15. Collectively these findings strengthen the evidence for an aetiological role of retinoic acid and retinoid receptors in the regulation of puberty timing, although their relative importance to male versus female puberty remains to be established. Further studies are also needed to identify functional links between these allelic signals and specific gene and protein functions.

Three additional novel signals reached genome-wide significance for puberty timing, represented by variants in near RNF144B-ID4, LEPR and KAL1. All three were associated with puberty timing in both men and women, two of which had stronger effects in males (LEPR and RNF144B-ID4). The signal at 6p22.3 resides in a gene desert with the two nearest genes, type 2 diabetes, triglyceride levels, cardiovascular disease and bone mineral density; and positive genetic correlations are observed with high-density lipoprotein cholesterol levels and adult height. Notably, we identified two novel signals, near RORB and RXRA, which add to the reported signals near to other retinoic acid receptor-related genes, RORA and RXRG5, and a fifth signal (near RXRB) showed sub-genome-wide significant association with puberty timing. The retinoic acid receptor-related and retinoid X receptors function as transcription factors that dimerize and regulate nuclear receptors to influence cell differentiation, development, circadian rhythm and metabolism13. Their receptor partners include the canonical receptors for oestrogen and androgens among other hormones and metabolites14, and their conformational changes may alter receptor sensitivity15. Collectively these findings strengthen the evidence for an aetiological role of retinoic acid and retinoid receptors in the regulation of puberty timing, although their relative importance to male versus female puberty remains to be established. Further studies are also needed to identify functional links between these allelic signals and specific gene and protein functions.

Three additional novel signals reached genome-wide significance for puberty timing, represented by variants in near RNF144B-ID4, LEPR and KAL1. All three were associated with puberty timing in both men and women, two of which had stronger effects in males (LEPR and RNF144B-ID4). The signal at 6p22.3 resides in a gene desert with the two nearest genes, type 2 diabetes, triglyceride levels, cardiovascular disease and bone mineral density; and positive genetic correlations are observed with high-density lipoprotein cholesterol levels and adult height.

Notably, we identified two novel signals, near RORB and RXRA, which add to the reported signals near to other retinoic acid receptor-related genes, RORA and RXRG5, and a fifth signal (near RXRB) showed sub-genome-wide significant association with puberty timing. The retinoic acid receptor-related and retinoid X receptors function as transcription factors that dimerize and regulate nuclear receptors to influence cell differentiation, development, circadian rhythm and metabolism13. Their receptor partners include the canonical receptors for oestrogen and androgens among other hormones and metabolites14, and their conformational changes may alter receptor sensitivity15. Collectively these findings strengthen the evidence for an aetiological role of retinoic acid and retinoid receptors in the regulation of puberty timing, although their relative importance to male versus female puberty remains to be established. Further studies are also needed to identify functional links between these allelic signals and specific gene and protein functions.

Three additional novel signals reached genome-wide significance for puberty timing, represented by variants in near RNF144B-ID4, LEPR and KAL1. All three were associated with puberty timing in both men and women, two of which had stronger effects in males (LEPR and RNF144B-ID4). The signal at 6p22.3 resides in a gene desert with the two nearest genes, type 2 diabetes, triglyceride levels, cardiovascular disease and bone mineral density; and positive genetic correlations are observed with high-density lipoprotein cholesterol levels and adult height.

Notably, we identified two novel signals, near RORB and RXRA, which add to the reported signals near to other retinoic acid receptor-related genes, RORA and RXRG5, and a fifth signal (near RXRB) showed sub-genome-wide significant association with puberty timing. The retinoic acid receptor-related and retinoid X receptors function as transcription factors that dimerize and regulate nuclear receptors to influence cell differentiation, development, circadian rhythm and metabolism13. Their receptor partners include the canonical receptors for oestrogen and androgens among other hormones and metabolites14, and their conformational changes may alter receptor sensitivity15. Collectively these findings strengthen the evidence for an aetiological role of retinoic acid and retinoid receptors in the regulation of puberty timing, although their relative importance to male versus female puberty remains to be established. Further studies are also needed to identify functional links between these allelic signals and specific gene and protein functions.

Three additional novel signals reached genome-wide significance for puberty timing, represented by variants in near RNF144B-ID4, LEPR and KAL1. All three were associated with puberty timing in both men and women, two of which had stronger effects in males (LEPR and RNF144B-ID4). The signal at 6p22.3 resides in a gene desert with the two nearest genes, type 2 diabetes, triglyceride levels, cardiovascular disease and bone mineral density; and positive genetic correlations are observed with high-density lipoprotein cholesterol levels and adult height. Notably, we identified two novel signals, near RORB and RXRA, which add to the reported signals near to other retinoic acid receptor-related genes, RORA and RXRG5, and a fifth signal (near RXRB) showed sub-genome-wide significant association with puberty timing. The retinoic acid receptor-related and retinoid X receptors function as transcription factors that dimerize and regulate nuclear receptors to influence cell differentiation, development, circadian rhythm and metabolism13. Their receptor partners include the canonical receptors for oestrogen and androgens among other hormones and metabolites14, and their conformational changes may alter receptor sensitivity15. Collectively these findings strengthen the evidence for an aetiological role of retinoic acid and retinoid receptors in the regulation of puberty timing, although their relative importance to male versus female puberty remains to be established. Further studies are also needed to identify functional links between these allelic signals and specific gene and protein functions.
Figure 1 | Scatterplot comparing effect sizes of the same genetic variants on age at menarche in women and age at voice breaking in men.
Dots indicate newly identified voice breaking or previously reported age at menarche variants. NS, not significant.

Table 2 | Genetic correlations between puberty timing and 27 health-related outcomes.

<table>
<thead>
<tr>
<th>Outcome†</th>
<th>Age at menarche (n=209,820)</th>
<th>Age at voice breaking (n=55,871)</th>
<th>Age at menarche and voice breaking, combined (N=265,691)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>rs g</td>
<td>s.e.</td>
<td>P value</td>
</tr>
<tr>
<td>Body mass index (~234k)</td>
<td>–0.35</td>
<td>0.02</td>
<td>1.4E – 01</td>
</tr>
<tr>
<td>HDL cholesterol (~100k)</td>
<td>0.17</td>
<td>0.03</td>
<td>3.4E – 09</td>
</tr>
<tr>
<td>Polycystic ovary syndrome (~88k)</td>
<td>–0.25</td>
<td>0.05</td>
<td>7.2E – 07</td>
</tr>
<tr>
<td>Adult height (~253k)</td>
<td>0.13</td>
<td>0.02</td>
<td>2.5E – 10</td>
</tr>
<tr>
<td>Fasting insulin (~52k)</td>
<td>–0.23</td>
<td>0.05</td>
<td>3.6E – 06</td>
</tr>
<tr>
<td>Triglycerides (~97k)</td>
<td>–0.12</td>
<td>0.03</td>
<td>5.1E – 06</td>
</tr>
<tr>
<td>Cardiovascular disease (87k)</td>
<td>–0.12</td>
<td>0.04</td>
<td>5.0E – 03</td>
</tr>
<tr>
<td>Type 2 diabetes (69k)</td>
<td>–0.13</td>
<td>0.04</td>
<td>5.2E – 04</td>
</tr>
<tr>
<td>Lumbar spine BMD (33k)</td>
<td>–0.08</td>
<td>0.03</td>
<td>1.4E – 02</td>
</tr>
<tr>
<td>Femoral neck BMD (33k)</td>
<td>–0.08</td>
<td>0.03</td>
<td>1.3E – 02</td>
</tr>
<tr>
<td>HbA1c (46k)</td>
<td>–0.11</td>
<td>0.05</td>
<td>2.1E – 02</td>
</tr>
<tr>
<td>Fasting glucose (58k)</td>
<td>–0.06</td>
<td>0.04</td>
<td>8.7E – 02</td>
</tr>
<tr>
<td>Bipolar disorder (17k)</td>
<td>0.07</td>
<td>0.04</td>
<td>9.2E – 02</td>
</tr>
<tr>
<td>Autism (10k)</td>
<td>–0.10</td>
<td>0.04</td>
<td>2.3E – 02</td>
</tr>
<tr>
<td>Schizophrenia (70k)</td>
<td>0.03</td>
<td>0.03</td>
<td>3.1E – 01</td>
</tr>
<tr>
<td>Birth weight (27k)</td>
<td>–0.07</td>
<td>0.05</td>
<td>1.5E – 01</td>
</tr>
<tr>
<td>Rheumatoid arthritis (26k)</td>
<td>0.04</td>
<td>0.04</td>
<td>3.8E – 01</td>
</tr>
<tr>
<td>Alzheimer’s disease (54k)</td>
<td>0.04</td>
<td>0.05</td>
<td>4.2E – 01</td>
</tr>
<tr>
<td>Major depressive disorder (19k)</td>
<td>–0.04</td>
<td>0.06</td>
<td>4.7E – 01</td>
</tr>
<tr>
<td>Crohn’s disease (21k)</td>
<td>0.04</td>
<td>0.03</td>
<td>1.8E – 01</td>
</tr>
<tr>
<td>Diabetic blood pressure (69k)</td>
<td>–0.03</td>
<td>0.03</td>
<td>3.8E – 01</td>
</tr>
<tr>
<td>ADHD (5k)</td>
<td>–0.06</td>
<td>0.08</td>
<td>4.2E – 01</td>
</tr>
<tr>
<td>Systolic blood pressure (69k)</td>
<td>–0.03</td>
<td>0.03</td>
<td>4.5E – 01</td>
</tr>
<tr>
<td>Ever smoker (74k)</td>
<td>0.01</td>
<td>0.04</td>
<td>7.3E – 01</td>
</tr>
<tr>
<td>LDL cholesterol (95k)</td>
<td>0.01</td>
<td>0.03</td>
<td>6.4E – 01</td>
</tr>
<tr>
<td>Anorexia nervosa (18k)</td>
<td>0.03</td>
<td>0.04</td>
<td>4.3E – 01</td>
</tr>
<tr>
<td>Ulcerative colitis (27k)</td>
<td>0.04</td>
<td>0.05</td>
<td>4.1E – 01</td>
</tr>
</tbody>
</table>

**ADHD:** attention-deficit hyperactivity disorder; BMD: bone mineral density; HDL: high-density lipoprotein; LDL: low-density lipoprotein.
†Sample size in each GWAS for puberty-related outcomes is shown in parentheses after each outcome name.
‡P values below the multiple test-corrected threshold <1.85 × 10⁻⁵ (≈ 0.05/27) are highlighted in bold.

with rs5934505) and KAL1 encodes the extracellular matrix glycoprotein anosmin-1 implicated in the embryonic migration of gonadotrophin releasing hormone and olfactory neurons. Deleterious mutations in KAL1 cause X-linked Kallmann syndrome, characterized by hypogonadotropic hypogonadism and anosmia. rs140410685 near LEPR, which encodes the leptin receptor, is uncorrelated with a reported neighbouring signal for age at menarche (r²~0 with rs10789181), and shows no reported association with adult BMI in reported GWAS meta-analyses (rs2186245 r²=1). In rare disorders of puberty, disruptive mutations usually have similar effects on puberty timing in both sexes.
exceptions are rare mutations in the genes that encode the pituitary hormones, follicle-stimulating hormone and luteinizing hormone, which disrupt puberty in females but not in males. In normal populations, rapid postnatal weight gain predicts earlier puberty timing in both sexes, but the influence of low birth weight is apparent only in females. We found that only a small minority of signals showed sex-discordant effects, most notably rs9321659 at the SIM1-MCHR2 locus. SIM1 encodes a transcription factor regulator of hypothalamic paraventricular nucleus development and function. Rare deleterious mutations cause hyperphagia and severe early onset obesity affecting both males and females, however, rs9321659 is reportedly not associated with adult BMI (P = 0.56)2,17. Overall, these data suggest that combining male and female data genome wide is likely to yield novel shared puberty loci, but care should be taken to assess for potential heterogeneity in the effects.

Finally, our observed genetic correlations indicate the relevance of puberty timing to later life health outcomes in both men and women. Consistent with traditional epidemiological evidence20–22, earlier puberty timing was genetically related to higher risks of adverse health outcomes, including higher BMI, type 2 diabetes, lipid profiles and cardiovascular disease; it was favourable only for bone mineral density. LD score regression is a powerful tool to identify potential causal relationships between traits; however, limitations include the inability to establish causal directions and to adjust for potential mediating factors. The relationship between puberty timing and obesity risk is complex, with plausible bi-directional mechanisms. Previous studies have reported that genome-wide significant signals for higher BMI, both in combination and individually, are associated with earlier puberty timing in females2,5,23,24. Conversely, in the opposite direction, earlier puberty timing associated with the LIN28B rs314276 C-allele leads to faster adolescent weight gain and higher post-pubertal BMI, without effects on pre-pubertal BMI.25. Further studies are required to explore whether health outcomes related to earlier puberty timing are fully mediated by higher BMI.

In summary, this large-scale assessment of the genetic architecture of puberty timing in males quantifies the extent of shared aetiology between sexes, extends the evidence implicating retinoic acid-related receptors in the regulation of puberty timing, and supports the relevance of puberty timing in both sexes to the aetiologies of various health-related outcomes.

Methods

Genome-wide association study for age at voice breaking. Genome-wide SNP data were generated from one or more of three genotyping arrays in up to 55,671 men aged 18 or older of European ancestry from the 23andMe study10,12,19, who reported their recalled age at voice breaking, by online questionnaire in response to the question ‘How old were you when your voice began to crack/deepen?’ Participants answered into one of the predefined age bins (under 9, 9–10 years old, 11–12 years old, 13–14 years old, 15–16 years old, 17–18 years old, 19 years old or older), scored from 0 to 6. Genetic effect estimates from these 2-year bins were re-calculated to 1-year effect estimates post analysis. We previously validated the accuracy of this approach by comparing re-estimated 2-year estimates for age at menarche (recorded in the same way) to those obtained from studies recording age at menarche by year. No significant heterogeneity was detected across these two approaches for known menarche loci2, 23andMe participants provided informed consent to take part in this research under a protocol approved by Ethical and Independent Review Services, an institutional review board accredited by the Association for the Accreditation of Human Research Protection Programs. Before imputation, we excluded SNPs with Hardy–Weinberg equilibrium P < 10−6, call rate < 95%, or with large allele frequency discrepancies compared with European 1,000 Genomes reference data. Frequency discrepancies were identified by computing a 2 × 2 table of allele counts for European 1,000 Genomes samples and 2,000 randomly sampled 23andMe participants with European ancestry, and identifying SNPs with a χ² P < 10 −12. Genotype data were imputed against the March 2012 release of 1,000 Genomes reference haplotype panel. Genetic association results were obtained from linear regression models assuming additive allelic effects. These models included as covariates—age and the top five genetically determined principal components to account for population structure. The reported SNP association test P values were computed from likelihood ratio tests. Results were further adjusted for a lambda GC value of 1.069 to correct for any residual test statistic inflation due to population stratification. LD score regression analysis26 also confirmed that principal component correction had appropriately controlled for inflation due to population stratification (pre-GC calculated intercept = 1) before the more conservative GC correction.

Independent signals were identified using a combination of distance-based clumping and approximate conditional analysis. Firstly, regions were defined on the basis of physical proximity, with the most strongly associated SNP representing the association signal for that region. We then tested for the presence of multiple statistically independent signals in each region using approximate conditional analysis implemented in GCTA27. Independent signals indicated by SNP P values < 1 × 10−6 were considered in follow-up analyses. A signal was considered to be the same as a previously reported menarche locus if it had a pairwise r² > 0.05.

Combined analyses with other puberty data sets. We followed up these selected SNPs in two additional sources of data: reported publicly available HapMap2 reference panel-imputed GWAS results for age at menarche from 182,416 women in the ReproGen consortium5; and independent 1,000 genomes reference panel-imputed GWAS data for age at menarche in 76,831 women in the 23andMe study10.

These additional samples were considered in three analytical designs. Firstly, all data in women (n = 259,247) were combined to estimate effects of the individual selected SNPs on age at menarche, using fixed-effects inverse variance-weighted meta-analysis with all effect estimates reported on a per year scale. Secondly, for genetic correlation analyses with age at menarche, women in ReproGen consortium cohorts genotyped by the custom ‘iCOG’ array were excluded to obtain a consistent sample size across GWAS SNPs (leaving data for analysis on 209,820 women). Thirdly, all GWAS results for puberty timing in men and women were combined into one analysis, with all effect estimates for individual SNP associations combined using fixed-effects inverse variance fixed-effects meta-analysis. LD Score Regression11,26 showed that combining GWAS data from men and women did not introduce substantial test statistic inflation due to possible relatedness between strata (cross-trait intercept 0.016, s.e. 0.005). Heritability between men and women for individual SNP associations was quantified by the I² statistic generated by METAL software. SNPs with a combination of distance and age heterogeneity were analysed in a random-effects model implemented by Han and Esko28.

Genetic correlations. Genetic correlations (r) were calculated between age at voice breaking in men, age at menarche in women, and 27 other complex traits/diseases in publicly available data sets using LD Score Regression11,26. Data sets used can be downloaded from http://www.med.unc.edu/pgc/downloads. Sample numbers in the studies of each trait are shown in Table 2. The only study with publicly available data set used was a polycystic ovary syndrome GWAS performed on 5,184 self-reported cases and 82,759 controls from the 23andMe study29. A conservative Bonferroni corrected P value threshold of 0.185 × 10−3 (e = 0.05/27) was used to define significant associations.

References


© 2015 Macmillan Publishers Limited. All rights reserved.

**Acknowledgements**
This work was supported by the Medical Research Council [U106179472; MC_U106179472; U106179471; MC_U106179471] and the National Human Genome Research Institute of the National Institutes of Health (grant number R44HG006981 to 23andMe). We thank the research participants and employees of 23andMe for making this work possible. HKF is supported by the Fannie and John Hertz Foundation.

**Author contributions**
All authors read and contributed to each version of the manuscript. F.R.D., B.B.-S., H.K.F. and J.R.B.P. carried out analysis. D.A.H., J.M.M. and J.Y.T. performed genotyping and phenotyping. D.A.H., J.Y.T., K.K.O. and J.R.B.P. contributed to study design and project management.

**Additional information**
Competing financial interests: D.A.H. and J.Y.T. are employees of and own stock or stock options in 23andMe, Inc. The remaining authors declare no conflict of interest.

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


This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article’s Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/