

**Serveur Académique Lausannois SERVAL [serval.unil.ch](http://serval.unil.ch)**

## **Author Manuscript**

### **Faculty of Biology and Medicine Publication**

**This paper has been peer-reviewed but does not include the final publisher proof-corrections or journal pagination.**

Published in final edited form as:

**Title:** Genetic variation in LIN28B is associated with the timing of puberty.

**Authors:** Ong KK, Elks CE, Li S, Zhao JH, Luan J, Andersen LB, Bingham SA, Brage S, Smith GD, Ekelund U, Gillson CJ, Glaser B, Golding J, Hardy R, Khaw KT, Kuh D, Luben R, Marcus M, McGeehin MA, Ness AR, Northstone K, Ring SM, Rubin C, Sims MA, Song K, Strachan DP, Vollenweider P, Waeber G, Waterworth DM, Wong A, Deloukas P, Barroso I, Mooser V, Loos RJ, Wareham NJ

**Journal:** Nature genetics

**Year:** 2009 Jun

**Volume:** 41

**Issue:** 6

**Pages:** 729-33

**DOI:** 10.1038/ng.382

In the absence of a copyright statement, users should assume that standard copyright protection applies, unless the article contains an explicit statement to the contrary. In case of doubt, contact the journal publisher to verify the copyright status of an article.

Published in final edited form as:

*Nat Genet.* 2009 June ; 41(6): 729–733. doi:10.1038/ng.382.

## Genetic variation in *LIN28B* is associated with the timing of puberty

Ken K Ong<sup>1,2,3</sup>, Cathy E Elks<sup>1,2</sup>, Shengxu Li<sup>1,2</sup>, Jing Hua Zhao<sup>1,2</sup>, Jian'an Luan<sup>1,2</sup>, Lars B Andersen<sup>4</sup>, Sheila A Bingham<sup>5,6</sup>, Soren Brage<sup>1,2</sup>, George Davey Smith<sup>7</sup>, Ulf Ekelund<sup>1,2,8</sup>, Christopher J Gillson<sup>1,2</sup>, Beate Glaser<sup>7</sup>, Jean Golding<sup>9</sup>, Rebecca Hardy<sup>10</sup>, Kay-Tee Khaw<sup>11</sup>, Diana Kuh<sup>10</sup>, Robert Luben<sup>11</sup>, Michele Marcus<sup>12,13,14</sup>, Michael A McGeehin<sup>12</sup>, Andrew R Ness<sup>15</sup>, Kate Northstone<sup>16</sup>, Susan M Ring<sup>16</sup>, Carol Rubin<sup>12</sup>, Matthew A Sims<sup>1,2</sup>, Kijoung Song<sup>17</sup>, David P Strachan<sup>18</sup>, Peter Vollenweider<sup>19</sup>, Gerard Waeber<sup>19</sup>, Dawn M Waterworth<sup>17</sup>, Andrew Wong<sup>10</sup>, Panagiotis Deloukas<sup>20</sup>, Inês Barroso<sup>20</sup>, Vincent Mooser<sup>17</sup>, Ruth J Loos<sup>1,2</sup>, and Nicholas J Wareham<sup>1,2</sup>

<sup>1</sup>Medical Research Council (MRC) Epidemiology Unit, Addenbrooke's Hospital, Cambridge, UK. <sup>2</sup>Institute of Metabolic Science, Addenbrooke's Hospital, Cambridge, UK. <sup>3</sup>Department of Paediatrics, University of Cambridge, Cambridge, UK. <sup>4</sup>Institute of Sport Science and Clinical Biomechanics, University of Southern Denmark, Odense, Denmark. <sup>5</sup>MRC Dunn Human Nutrition Unit, Wellcome Trust/MRC Building, Cambridge, UK. <sup>6</sup>MRC Centre for Nutritional Epidemiology in Cancer Prevention and Survival, Cambridge, UK. <sup>7</sup>MRC Centre for Causal Analyses in Translational Epidemiology, Department of Social Medicine, University of Bristol, UK. <sup>8</sup>School of Health and Medical Sciences, Örebro University, Örebro Sweden. <sup>9</sup>Avon Longitudinal Study of Parents and Children (ALSPAC), Department of Community Based Medicine, University of Bristol, Bristol, UK. <sup>10</sup>MRC Unit for Lifelong Health and Ageing, London, UK. <sup>11</sup>Department of Public Health and Primary Care, Institute of Public Health, University of Cambridge, Cambridge, UK. <sup>12</sup>National Center for Environmental Health, Centers for Disease Control and Prevention, Atlanta, Georgia, USA. <sup>13</sup>Department of Epidemiology, Rollins School of Public Health, Emory University, Atlanta, Georgia, USA. <sup>14</sup>Departments of Environmental and Occupational Health, Rollins School of Public Health, Emory University, Atlanta, Georgia, USA. <sup>15</sup>Department of Oral and Dental Science, University of Bristol, Bristol, UK. <sup>16</sup>ALSPAC, Department of Social Medicine, University of Bristol, Bristol, UK. <sup>17</sup>Genetics Division, GlaxoSmithKline, King of Prussia, Pennsylvania, USA. <sup>18</sup>Division of Community Health Sciences, St. George's, University of London, London, UK. <sup>19</sup>Department of Internal Medicine, BH-10 Centre Hospitalier Universitaire Vaudois (CHUV), Lausanne, Switzerland. <sup>20</sup>Wellcome Trust Sanger Institute, Hinxton, Cambridge, UK.

© 2009 Nature America, Inc. All rights reserved.

Correspondence should be addressed to K.O. (ken.ong@mrc-epid.cam.ac.uk) or R.L. (ruth.loos@mrc-epid.cam.ac.uk).

### AUTHOR CONTRIBUTIONS

**Writing group:** C.E.E., R.J.L., S.L., K.K.O., N.J.W.

**Genome-wide association analyses and meta-analyses:** I.B., P.D., C.E.E., S.L., R.J.L., V.M., K.K.O., N.J.W., J.H.Z.

**Oversight of contributing cohorts:** L.B.A., S.B., U.E., J.G., D.K., K.-T.K., S.A.B., V.M., A.R.N., D.P.S., G.D.S., P.V., N.J.W.

**Phenotype preparation and analyses in contributing cohorts:** S.B., U.E., B.G., J.G., R.H., S.L., J.L., R.L., R.J.L., M.M., M.A.M., A.R.N., K.N., K.K.O., C.R., D.P.S., K.S., G.W., D.M.W., J.H.Z.

**Generating new genotype data:** C.J.G., S.M.R., M.A.S., A.W.

Note: Supplementary information is available on the Nature Genetics website.

### COMPETING INTERESTS STATEMENT

The authors declare competing financial interests: details accompany the full-text HTML version of the paper at <http://www.nature.com/naturegenetics/>.

Reprints and permissions information is available online at <http://npg.nature.com/reprintsandpermissions/>

## Abstract

The timing of puberty is highly variable<sup>1</sup>. We carried out a genome-wide association study for age at menarche in 4,714 women and report an association in *LIN28B* on chromosome 6 (rs314276, minor allele frequency (MAF) = 0.33,  $P = 1.5 \times 10^{-8}$ ). In independent replication studies in 16,373 women, each major allele was associated with 0.12 years earlier menarche (95% CI = 0.08–0.16;  $P = 2.8 \times 10^{-10}$ ; combined  $P = 3.6 \times 10^{-16}$ ). This allele was also associated with earlier breast development in girls ( $P = 0.001$ ;  $N = 4,271$ ); earlier voice breaking ( $P = 0.006$ ,  $N = 1,026$ ) and more advanced pubic hair development in boys ( $P = 0.01$ ;  $N = 4,588$ ); a faster tempo of height growth in girls ( $P = 0.00008$ ;  $N = 4,271$ ) and boys ( $P = 0.03$ ;  $N = 4,588$ ); and shorter adult height in women ( $P = 3.6 \times 10^{-7}$ ;  $N = 17,274$ ) and men ( $P = 0.006$ ;  $N = 9,840$ ) in keeping with earlier growth cessation. These studies identify variation in *LIN28B*, a potent and specific regulator of microRNA processing<sup>2</sup>, as the first genetic determinant regulating the timing of human pubertal growth and development.

Puberty, the transition from childhood to adult body size and sexual maturity, is a complex multistaged process involving growth acceleration, weight gain and the appearance of secondary sexual physical features over a 2- to 3-year period<sup>1</sup>. Early onset and progression of puberty is seen in some overweight and obese children. In addition to its psycho-social and public health implications, early puberty is associated with increased long-term risk for diseases including obesity, diabetes and cancer<sup>3</sup>. The stages of puberty and their transitions are difficult to measure accurately<sup>1</sup>. Epidemiological studies often use age at menarche, the onset of the first menstrual period in girls, to indicate the timing of puberty, as this distinct event is often well recalled many years later<sup>1</sup>. Girls with earlier menarche are heavier and taller than other girls during childhood; they remain heavier but are shorter as adults<sup>3</sup>, reflecting their earlier cessation of growth. Twins studies estimate that 44–95% of the variance in age at menarche may be heritable<sup>1</sup>. However, specific common genetic variants that influence the timing of puberty have not yet been convincingly demonstrated<sup>4</sup>.

To identify common variants associated with the timing of puberty, we conducted a genome-wide association (GWA) study for age at menarche in 4,714 women from two general population studies and one obese case study using the same Affymetrix GeneChip 500K array (Supplementary Note online). Only one SNP, rs314276 in intron 2 of *LIN28B* on chromosome 6 (MAF 0.33), reached genome-wide statistical significance ( $P = 1.5 \times 10^{-8}$ , Supplementary Fig. 1 online). In these GWA studies overall each major C allele at rs314276 was associated with a mean 0.22 years (95% CI = 0.14–0.29) earlier age at menarche.

SNP rs314276 lies in a region of high linkage disequilibrium (LD); although no other SNPs in this region were directly genotyped in these arrays, imputation to HapMap build 35 revealed a further 11 SNPs in *LIN28B* that were associated with age at menarche at  $P < 0.5 \times 10^{-7}$  (Fig. 1). The minor allele at a neighboring SNP, rs314277, 337 bp downstream of rs314276, was recently robustly associated with taller adult height<sup>5</sup>. In our GWA studies imputed rs314277 genotypes showed only nominal association with age at menarche ( $P = 0.005$ ). However, as there was low LD between rs314276 and rs314277 ( $r^2 = 0.23$ ;  $D = 1.0$ ) we took both SNPs forward to initial replication.

Study of mothers from the ALSPAC cohort ( $N = 6,456$ ) confirmed the association between rs314276 and age at menarche ( $P = 6.6 \times 10^{-5}$ ). Each common C allele was associated with a mean 0.10 years (95% CI = 0.07–0.13) earlier menarche, the association seemed to be linear (Supplementary Fig. 2 online) and rs314276 explained 0.2% of the variance in age at menarche. In contrast, the adjacent SNP rs314277 was not associated with age at menarche ( $P = 0.08$ ). In multiple regression analysis including both SNPs, rs314276 ( $P = 0.0006$ ), but

not rs314277 ( $P=0.6$ ), was associated with age at menarche. The association between rs314276 and age at menarche was further confirmed in women from the EPIC-Norfolk cohort ( $P=8.7 \times 10^{-6}$ ,  $N=8,411$ ), the MRC National Study of Health and Development (NSHD) ( $P=0.02$ ;  $N=948$ ) and the British 1958 Birth Cohort (B58C) ( $P=0.007$ ;  $N=558$ ; Fig. 2 and Supplementary Fig. 2). In a meta-analysis of all replication studies (total  $N=16,373$  women), each rs314276 C allele was associated with 0.12 years earlier menarche (95% CI = 0.08–0.16;  $P=2.8 \times 10^{-10}$ ; Fig. 2). Notably, the pooled regression coefficient in the replication studies was only around half of that in the GWA studies (heterogeneity between initial and replication studies:  $P=0.02$ ; Fig. 2), likely reflecting the ‘winner's curse’ phenomenon.

Although menarche represents the completion of female sexual maturation and attainment of reproductive capacity, the onset of puberty in girls typically occurs 2–3 years earlier and is manifested by the onset of breast growth and acceleration in height and weight gain. To explore its association with the timing of pubertal onset, we genotyped rs314276 in various childhood studies (see Table 1 for summary).

In ALSPAC girls ( $N=4,271$ ), the common C allele at rs314276 was associated with earlier onset of breast development (Fig. 3, log-rank test:  $P=0.001$ ). At age 10.75 years each C allele was associated with a 20% increased likelihood of breast development (odds ratio (OR) = 1.20, 95% CI = 1.06–1.35;  $P=0.003$ ). In the cross-sectional European Youth Heart Study (EYHS), across all age groups from 9 to 16 years the common C allele at rs314276 was associated with more advanced breast stage in girls (OR = 1.26, 95% CI = 1.05–1.52;  $P=0.006$ ;  $N=1,044$ ) (Table 1 and Supplementary Table 1 online). In ALSPAC girls, the rs314276 C allele was associated with a faster tempo of growth in height between ages 7 and 11 years ( $P=0.00008$ ), and relative acceleration in weight ( $P=0.0003$ ) and body mass index (BMI) ( $P=0.03$ ) (Fig. 4).

In boys, the onset and progression of puberty are manifested by a gradual enlargement of the external genitalia and the spread of pubic hair. By convention these characteristics are categorized into five stages of sexual maturation; however, in epidemiological studies these stages are difficult to assign accurately. Equivalent to menarche in girls, voice breaking in boys represents a distinct event, which typically occurs abruptly in late puberty<sup>6</sup>. In NSHD men ( $N=1,027$ ) at age 15 years the common C allele at rs314276 was associated with more advanced pubic hair stage ( $P=0.05$ ), more advanced voice breaking status ( $P=0.006$ ) and also more advanced tempo of height growth ( $P=0.03$ ), but no apparent difference in genital size (Table 1 and Supplementary Table 2 online). In ALSPAC boys (total  $N=4,588$ ), the C allele at rs314276 was associated with more advanced pubic hair at age 13 years (OR for a one-stage advance in pubic hair stage per C allele = 1.19, 95% CI = 1.04–1.35;  $P=0.01$ ), and with a faster tempo of growth in height between ages 7 and 11 years ( $P=0.03$ ; Supplementary Fig. 3 online). In the smaller sample of EYHS boys (total  $N=910$ ), no association was found with genital or pubic hair stages at ages 9–11 years or 14–16 years (Supplementary Table 1).

In contrast to their relatively taller stature during childhood, men and women with early puberty are shorter as adults, owing to their earlier cessation of growth, but they remain heavier and more overweight<sup>3</sup>. Consistent with its associations with earlier puberty, the rs314276 C allele was also associated with shorter adult height in ALSPAC mothers (mean  $\pm$  s.e.m. difference:  $-0.37 \pm 0.12$  cm per C allele,  $P=0.002$ ), EPIC-Norfolk women ( $-0.36 \pm 0.10$  cm,  $P=0.0002$ ) and EPIC-Norfolk men ( $-0.30 \pm 0.11$  cm,  $P=0.006$ ). In multiple regression among ALSPAC mothers including both rs314276 and rs314277, rs314276 ( $P=0.003$ ) but not rs314277 ( $P=0.3$ ) was associated with adult height. In contrast to the

findings with adult height, rs314276 showed no association with adult weight, BMI, waist circumference or percentage body fat (all  $P > 0.10$ , see Supplementary Table 3 online)

Several candidate genes, encoding regulators of sex steroid secretion and action, have been proposed to regulate the timing of puberty; however, no common variants have yet been established<sup>4</sup>. In our GWA studies, SNPs in *FGFR1*, *GOLT1A*, *KISS1*, *LEPR* and *SHBG* showed modest associations with age at menarche ( $0.01 < P < 0.05$ ; Supplementary Table 4 online). Our study therefore reinforces the merits of the genome-wide compared to the candidate gene approach, although much larger studies will be needed to identify the likely large number of SNPs associated with timing of puberty.

A recent GWA study for adult height identified rs314277 as one of ten loci with robust associations ( $P = 5.9 \times 10^{-9}$  in 15,821 individuals), whereas rs314276 showed far weaker association ( $P \sim 1 \times 10^{-4}$ ); however, rs314276 was not directly genotyped in many of those samples<sup>5</sup>. In our study, where both SNPs were directly genotyped, the effects of rs314277 on height were explained by rs314276. rs314276 lies in a large LD block of around 200 kb that covers the 5' region and first three exons of *LIN28B*. Imputation revealed a further 11 SNPs that were in complete LD with rs314276 and that also reached genome-wide association with age at menarche. Fine mapping studies in similar populations of European ancestry are therefore unlikely to be able to distinguish the causal variant(s) from other linked variants. *LIN28B* has two isoforms<sup>7</sup> distinguished by the presence or truncation of a highly conserved cold-shock domain (CSD), which is crucial for protein function<sup>2</sup>. Different *LIN28B* isoforms could therefore contribute to the heritability of the timing of growth and development.

*LIN28B* shows high sequence, structural and functional homology with *LIN28* on chromosome 1 (ref. 7); however, we found no association with 12 directly genotyped or imputed SNPs in *LIN28* (all  $P > 0.2$ ; see Supplementary Table 4). Both *LIN28B* and *LIN28* show similar sequence homology to the heterochronic gene *lin-28* in *Caenorhabditis elegans*<sup>7</sup>. Deleterious mutations in *lin-28* produce an abnormal rapid tempo of development through larval stages to adult cuticle formation<sup>8</sup>. Conversely, enhancement of *lin-28* expression by deletion of regulatory elements delays larval stage progression<sup>9</sup>. Both *LIN28B* and *LIN28* encode potent and specific regulators of pre-processing of the let7 family of microRNAs<sup>2</sup> and regulate cell pluripotency<sup>10</sup> and cancer growth<sup>7</sup>.

In conclusion, we have identified SNP rs314276, or another related variant within *LIN28B*, as the first genetic marker associated with the timing of pubertal growth and development and in both girls and boys. Our findings suggest the conservation of a fundamental cell regulatory system that controls the tempo of somatic development and also suggest a physiological role for microRNA processing in the timing of human growth and development.

## METHODS

### Genome-wide association samples

**EPIC-Norfolk Obesity Case-Cohort**—The control cohort comprised 2,566 individuals (1,364 women) randomly selected from the EPIC-Norfolk study of 25,663 men and women of European descent. The case cohort comprised 1,685 obese individuals (718 women), defined as BMI index  $>30 \text{ kg/m}^2$ , randomly selected from the obese individuals within EPIC-Norfolk. Participants were aged 39–79 years and were recruited in Norfolk, UK between 1993 and 1997 (ref. 11). Following exclusions due to quality control criteria and missing data, on 625 obese and 1,215 control women were available for genome-wide analysis.

**Cohorte LAUSannoise (CoLaus)**—CoLaus is a cross-sectional study of a random sample of 6,188 European adults (including 2,976 women), aged 35–75 years, living in Lausanne, Switzerland<sup>12</sup>. Following exclusions due to quality control criteria and missing data, 2,874 women were included in the genome-wide analyses.

In all studies, age at menarche to the nearest completed whole year was ascertained at baseline by questionnaire. Values ranging from 8 to 18 years were included as these were deemed to be physiological. All participants gave written informed consent, and project protocols were approved by the appropriate research ethics committees (Supplementary Note).

### Genome-wide SNP genotyping and quality control

All GWA studies were genotyped using the Affymetrix GeneChip 500K array set, and genotypes were called using the BRLMM algorithm. Each study applied genotyping quality control criteria and conducted tests for population stratification: 352,700 SNPs passed quality control in EPIC-Norfolk and 390,631 in the CoLaus cohort. In CoLaus correction for population substructure was done by principal components. The genomic control inflation factors ( $I_{GC}$ ) for each GWA study were  $<1$  for EPIC-Norfolk control cohort and obese cases and 1.016 in CoLaus, indicating that the extent of residual population substructure is modest, and no further genomic control corrections were applied.

### Genome-wide analyses and meta-analyses

Age at menarche showed a normal distribution. In each GWA study, linear regression was carried out (assuming an additive model) to test the association between each SNP and age at menarche using PLINK (CoLaus) or SAS/Genetics 9.1 (EPIC-Norfolk). Subsequently, summary statistics of the SNP-menarche age associations of each GWA study were combined in meta-analyses using the inverse variance-weighted method with a fixed-effects model using SAS 9.1. The overall results of the meta-analysis were visualized using HAPLOVIEW.

### Imputation of genotypes

To increase genomic coverage, polymorphic HapMap CEU SNPs were imputed using IMPUTE. Associations with age at menarche were subsequently tested with these imputed SNPs using linear regression under additive genetic assumptions.

### Replication samples

Initial replication of the top hit rs314276 and adjacent rs314277 was done in mothers from the Avon Longitudinal Study of Parents and Children (ALSPAC). This population-based study recruited pregnant women with expected delivery dates between April 1991 and December 1992 from Bristol, UK<sup>13</sup>. Genotypes for rs314276 and rs314277 and recalled menarche data were available for 6,456 mothers. Replication samples were also available from a further 10,824 women in the EPIC-Norfolk Study, excluding women who had been analyzed in the GWA study. rs314276 was directly genotyped in 10,453 women, of whom 8,411 had valid data on age at menarche. SNP rs314276 was genotyped in a further 948 women with a valid age at menarche from the MRC National Survey of Health and Development (NSHD), a prospective birth cohort study comprising a stratified sample of all births in England, Scotland and Wales in one week in March 1947 (ref. 14). A further 558 women from the 1958 British Birth Cohort (B58C), a prospective birth cohort originally consisting of all births in England, Wales and Scotland during one week in 1958, had rs314276 genotypes and age at menarche data available.

## Replication genotyping and quality control

Genotyping of rs314276 and rs314277 was conducted using TaqMan (Applied Biosystems) (EPIC and NSHD) or by KBiosciences (ALSPAC). Genotype data from B58C was obtained using an Affymetrix GeneChip Human Mapping 500K Array. Genotype frequencies were in Hardy-Weinberg equilibrium (all  $P > 0.2$ ), and call rates were all  $>90\%$ .

## Statistical analyses in replication studies

Associations between SNPs and age at menarche (women only), height, weight, BMI, waist circumference and percentage body fat were conducted in men and women separately using linear regression assuming an additive genetic model and adjusting for age. In ALSPAC mothers, multiple regression models were performed including both rs314276 and rs314277 to identify the independent effects of each SNP on age at menarche and height. Summary statistics were meta-analyzed with the initial GWA study sets using the inverse variance-weighted method. Analyses were performed using Stata/s.e.m. 9.2 for Windows (StataCorp).

## Childhood studies

We analyzed 8,859 ALSPAC children (4,588 boys and 4,271 girls) to examine the association between rs312726 genotype and timing of adolescent growth (between ages 7 to 11 years), pubic hair stage in boys and breast development in girls. Data on pubertal development (voice breaking status and pubic hair stage) at age 15 years were available for NSHD men ( $N = 1,026$ ). We analyzed 1,964 individuals from the Danish and Estonian cohorts of the cross-sectional European Youth Heart Study (EYHS) to determine the association between rs312726 and pubertal stage.

## Childhood studies genotyping and quality control

rs314276 was genotyped using TaqMan SNP genotyping assay (Applied Biosystems) (EYHS) or by KBiosciences (ALSPAC). Genotype frequencies were in Hardy-Weinberg equilibrium (all  $P > 0.1$ ), and call rates were all  $>96\%$ .

## Statistical analyses in childhood studies

Height, weight and BMI measurements in each individual were converted to s.d. scores (SDS) adjusted for sex and age. Growth tempo, a measure of child's current height relative to height potential, was calculated as the child's height SDS minus their mean parental height SDS15. Associations between rs314276 and changes in growth tempo, weight and BMI SDS between ages 7 to 11 years were calculated using time series analyses (repeated measures ANOVA). Association between rs314276 and the onset of pubertal breast development in girls was calculated using a log-rank test. Cross-sectional associations between rs314276 and pubertal stage in girls and boys were done by ordinal logistic regression. All analyses assumed an additive genetic model, and were done using SPSS v.14 for Windows.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

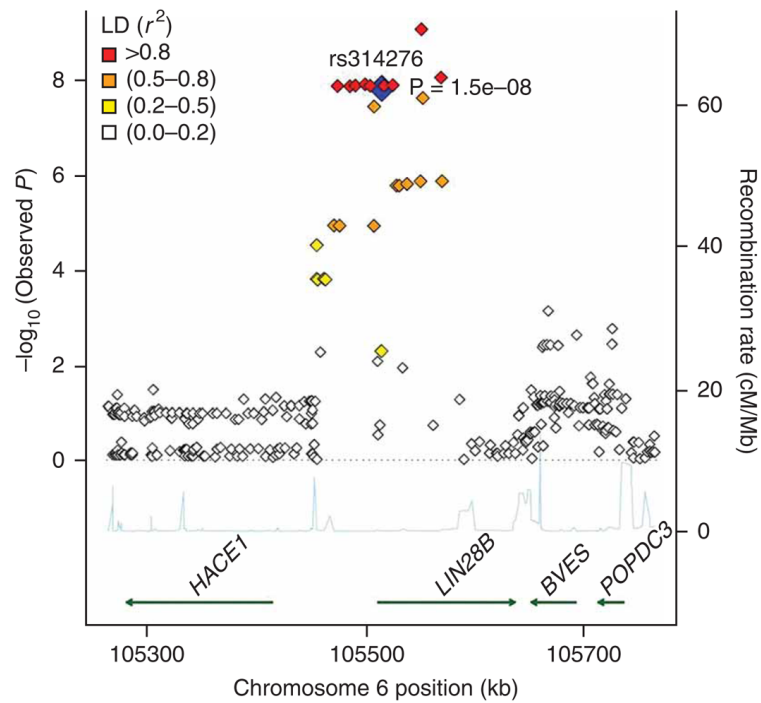
## Acknowledgments

We are grateful to all of the participants in each of the studies contributing to this effort. Full acknowledgments can be found in the Supplementary Note. Support for this research was provided by: the UK Medical Research Council; the Wellcome Trust; University of Bristol; the Faculty of Biology and Medicine of Lausanne, Switzerland; and GlaxoSmithKline.

## References

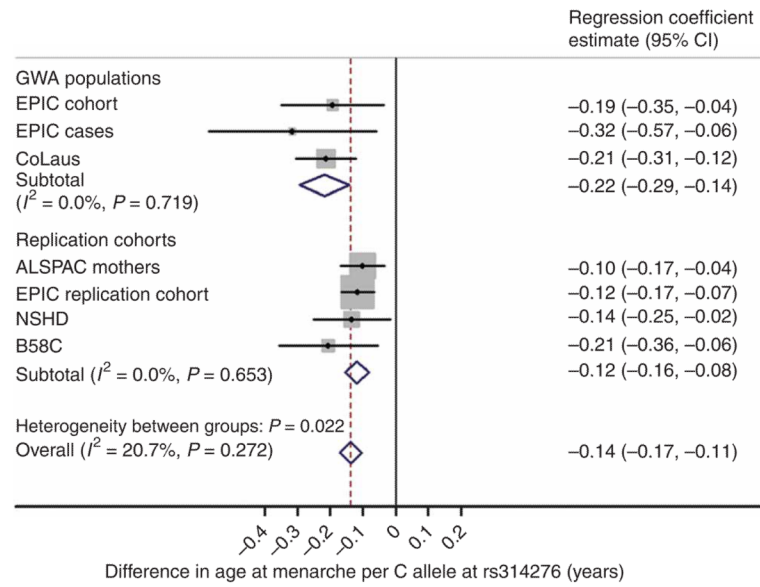
1. Parent AS, et al. The timing of normal puberty and the age limits of sexual precocity: variations around the world, secular trends, and changes after migration. *Endocr. Rev.* 2003; 24:668–693. [PubMed: 14570750]
2. Viswanathan SR, Daley GQ, Gregory RI. Selective blockade of microRNA processing by Lin28. *Science.* 2008; 320:97–100. [PubMed: 18292307]
3. Lakshman R, et al. Association between age at menarche and risk of diabetes in adults: results from the EPIC-Norfolk cohort study. *Diabetologia.* 2008; 51:781–786. [PubMed: 18320165]
4. Gajdos ZK, Hirschhorn JN, Palmert MR. What controls the timing of puberty? An update on progress from genetic investigation. *Curr. Opin. Endocrinol. Diabetes Obes.* 2009; 16:16–24. [PubMed: 19104234]
5. Lettre G, et al. Identification of ten loci associated with height highlights new biological pathways in human growth. *Nat. Genet.* 2008; 40:584–591. [PubMed: 18391950]
6. Harries ML, Walker JM, Williams DM, Hawkins S, Hughes IA. Changes in the male voice at puberty. *Arch. Dis. Child.* 1997; 77:445–447. [PubMed: 9487971]
7. Guo Y, et al. Identification and characterization of lin-28 homolog B (LIN28B) in human hepatocellular carcinoma. *Gene.* 2006; 384:51–61. [PubMed: 16971064]
8. Ambros V, Horvitz HR. Heterochronic mutants of the nematode *Caenorhabditis elegans*. *Science.* 1984; 226:409–416. [PubMed: 6494891]
9. Moss EG, Lee RC, Ambros V. The cold shock domain protein LIN-28 controls developmental timing in *C. elegans* and is regulated by the lin-4 RNA. *Cell.* 1997; 88:637–646. [PubMed: 9054503]
10. Yu J, et al. Induced pluripotent stem cell lines derived from human somatic cells. *Science.* 2007; 318:1917–1920. [PubMed: 18029452]
11. Day N, et al. EPIC-Norfolk: study design and characteristics of the cohort. *European Prospective Investigation of Cancer. Br. J. Cancer.* 1999; 80(Suppl. 1):95–103. [PubMed: 10466767]
12. Firmann M, et al. The CoLaus study: a population-based study to investigate the epidemiology and genetic determinants of cardiovascular risk factors and metabolic syndrome. *BMC Cardiovasc. Disord.* 2008; 8:6. [PubMed: 18366642]
13. Golding J, Pembrey ME, Jones R. ALSPAC — the Avon Longitudinal Study of Parents and Children. I. Study methodology. *Paediatr. Perinat. Epidemiol.* 2001; 15:74–87. [PubMed: 11237119]
14. Wadsworth M, Kuh D, Richards M, Hardy R. Cohort Profile: The 1946 National Birth Cohort (MRC National Survey of Health and Development). *Int. J. Epidemiol.* 2006; 35:49–54. [PubMed: 16204333]
15. Wehkalampi K, et al. Genetic and environmental influences on pubertal timing assessed by height growth. *Am. J. Hum. Biol.* 2008; 20:417–423. [PubMed: 18293372]





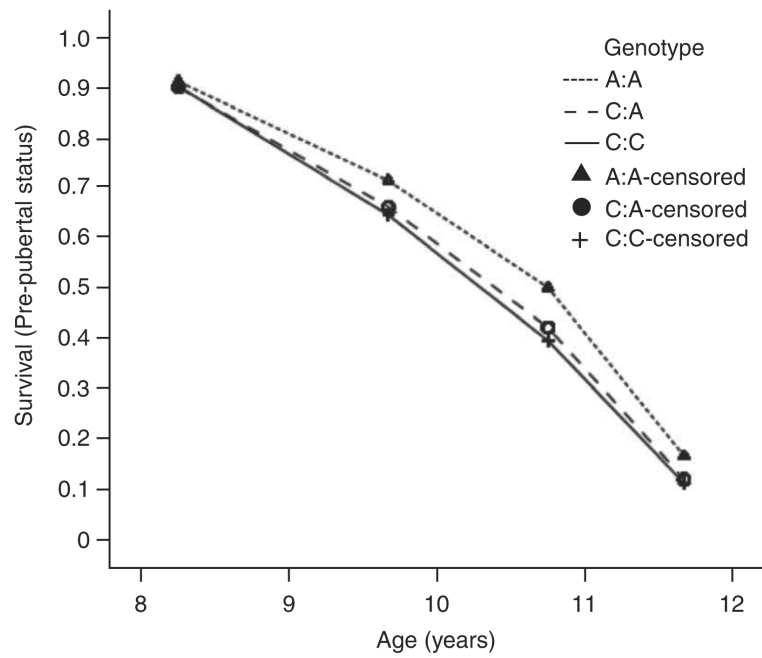
**Figure 1.**

Regional plot of the locus around *LIN28B* associated with age at menarche. SNPs are plotted by position on chromosome 6 against GWAS association with age at menarche ( $-\log_{10} P$  value). SNP rs314276 is shown in blue, labeled with its stage 1  $P$  value. Estimated recombination rates (from HapMap) are plotted in cyan to reflect the local LD structure. The directly genotyped or imputed SNPs surrounding rs314276 are color-coded to reflect their LD with rs314276 as in the inset (taken from pairwise  $r^2$  values from the HapMap CEU database). Genes and their directions of transcription are noted at the bottom of the plot.

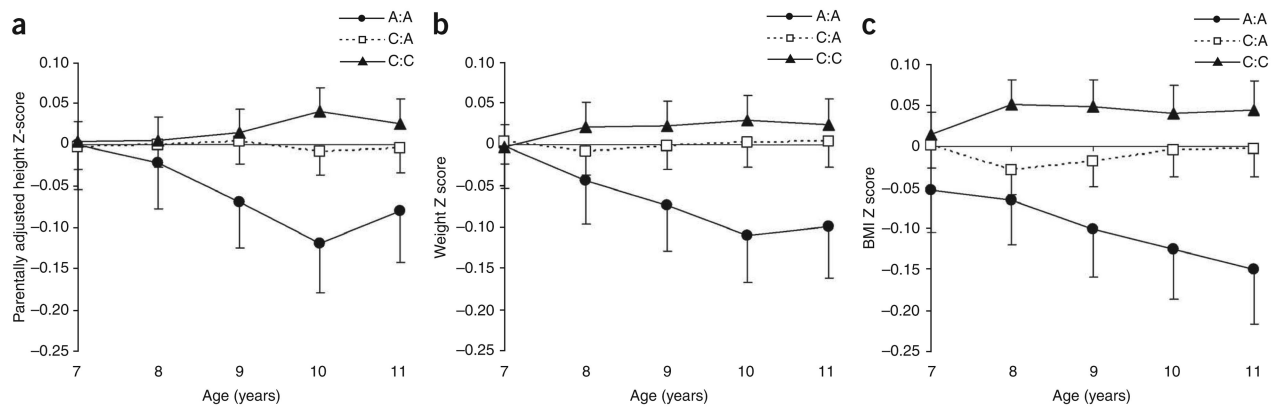


**Figure 2.**

A forest plot showing the meta-analysis of the effect of each C allele at rs314276 in *LIN28B* on earlier age at menarche in the GWA populations, the replication cohorts and in all groups.



**Figure 3.** Kaplan-Maier plot of survival in pre-pubertal status (Tanner breast stage 1) by age and *LIN28B* rs314276 genotype in ALSPAC girls ( $N=3,233$ ).  $P=0.001$ , log-rank test for genotype difference.



**Figure 4.**

Adolescent growth in ALSPAC girls by *LIN28B* rs314276 genotype. (a–c) Mean ( $\pm$  s.e.) standard deviation scores (SDS) for tempo of growth (child's height SDS minus mean parental height SDS) (a), weight (b) and BMI (c) are plotted against age from 7 to 11 years by genotype. *P* values (0.00008, 0.0003 and 0.03 for height, weight and BMI, respectively) are from time-series analyses (repeated measures ANOVA) for the change in SDS over time by genotype (additive genetic models).

**Table 1**Summary of the pubertal and growth phenotype associations with *LIN28B* rs314276 genotype

Variable	Estimated effect per C allele mean/odds ratio (95% CI)	P value	Study population
Girls/Women			
Earlier menarche	-0.12 years (-0.08 to -0.16)	$2.8 \times 10^{-10}$	ALSPAC mothers, EPIC-Norfolk, NSHD, B58C
More likely breast development at 10 years <sup>a</sup>	OR = 1.20 (1.06 to 1.35)	0.003	ALSPAC girls
More advanced breast stage at ages 9 to 16 years <sup>b</sup>	OR = 1.26 (1.05 to 1.52)	0.006	EYHS
Faster height tempo at age 10 years <sup>c</sup>	0.074 SDS (0.042 to 0.106)	0.02	ALSPAC girls
Shorter adult height	-0.37 cm (-0.51 to -0.22)	$3.6 \times 10^{-7}$	ALSPAC mothers, EPIC-Norfolk
Boys/Men			
More advanced voice breaking status at age 15 years <sup>d</sup>	OR = 1.26 (1.07 to 1.50)	0.006	NSHD
More advanced pubic hair stage at age 13 years <sup>e</sup>	OR = 1.19 (1.04 to 1.35)	0.01	ALSPAC boys
More advanced pubic hair stage at age 15 years <sup>f</sup>	OR = 1.18 (0.99 to 1.40)	0.05	NSHD
Faster height tempo at age 10 years <sup>c</sup>	0.085 SDS (0.054 to 0.116)	0.005	ALSPAC boys
Shorter adult height	-0.30 cm (-0.41 to -0.19)	0.006	EPIC-Norfolk

<sup>a</sup>Breast development was defined as Tanner stage 2 or greater.

<sup>b</sup>Breast stage was defined by the five-point Tanner scale.

<sup>c</sup>Height tempo was calculated as height s.d. score minus parental height s.d. score.

<sup>d</sup>Voice breaking status was categorized using the following scale: 1, no; 2, starting to break; 3, completely broken.

<sup>e</sup>Pubic hair stage was defined by the five-point Tanner scale.

<sup>f</sup>Pubic hair stage was categorized using the following scale: 1, no; 2, yes, sparse; 3, yes, profuse.