A Polygenic Risk Score Suggests Shared Genetic Architecture of Voice Break With Early Markers of Pubertal Onset in Boys

María C. Lardone,[1] Alexander S. Busch,[2] José L. Santos,[3] Patricio Miranda,[3] Susana Eyheramendy,[4] Ana Pereira,[5] Anders Juul,[2] Kristian Almstrup,[2] and Verónica Mericq[1]

[1]Institute of Maternal and Child Research, School of Medicine, University of Chile, Santiago 8360160, Chile; [2]Department of Growth and Reproduction, Rigshospitalet, University of Copenhagen 2100, Copenhagen, Denmark; [3]Department of Nutrition, Diabetes and Metabolism, School of Medicine, Pontificia Universidad Católica de Chile, Santiago 8331150, Chile; [4]Faculty of Engineering and Sciences, Universidad Adolfo Ibañez, Santiago 7941169, Chile; and [5]Institute of Nutrition and Food Technology, University of Chile, Santiago 7830490, Chile

ORCiD numbers: 0000-0001-7139-2775 (M. C. Lardone); 0000-0003-4417-569X (A. S. Busch); 0000-0002-0534-4350 (A. Juul); 0000-0002-1832-0307 (K. Almstrup); 0000-0003-2287-0181 (V. Mericq).

Context: Voice break, as a landmark of advanced male puberty in genome-wide association studies (GWAS), has revealed that pubertal timing is a highly polygenic trait. Although voice break is easily recorded in large cohorts, it holds quite low precision as a marker of puberty. In contrast, gonadarche and pubarche are early and clinically well-defined measures of puberty onset.

Objective: To determine whether a polygenic risk score (PRS) of alleles that confer risk for voice break associates with age at gonadarche (AAG) and age at pubarche (AAP) in Chilean boys.

Experimental Design: Longitudinal study.

Subjects and Methods: 401 boys from the Growth and Obesity Chilean Cohort Study (n = 1194; 49.2% boys).

Main Outcome Measures: Biannual clinical pubertal staging including orchidometry. AAG and AAP were estimated by censoring methods. Genotyping was performed using the Multi-Ethnic Global Array (Illumina). Using GWAS summary statistics from the UK-Biobank, 29 significant and independent single nucleotide polymorphisms associated with age at voice break were extracted. Individual PRS were computed as the sum of risk alleles weighted by the effect size.

Results: The PRS was associated with AAG (β=0.01, $P = 0.04$) and AAP (β=0.185, $P = 0.0004$). In addition, boys within the 20% highest PRS experienced gonadarche and pubarche 0.55 and 0.67 years later than those in the lowest 20%, respectively ($P = 0.013$ and $P = 0.007$).

Conclusions: Genetic variants identified in large GWAS on age at VB significantly associate with age at testicular growth and pubic hair development, suggesting that these events share a genetic architecture across ethnically distinct populations. **(*J Clin Endocrinol Metab* 105: e346–e354, 2020)**

Key Words: gonadarche, pubarche, polygenic risk score, GWAS, male puberty

Puberty is a transition period between childhood and adulthood. It represents the process by which sexual

ISSN Print 0021-972X ISSN Online 1945-7197
Printed in USA

Received 24 October 2019. Accepted 7 January 2020.
First Published Online 9 January 2020.
Corrected and Typeset 6 February 2020.

Abbreviations: AAG, age at gonadarche;AAM, age at menarche;AAP, age at pubarche;BMI, body mass index; GnRH, gonadotropin-releasing hormone; GOCS, Growth and Obesity Chilean Cohort Study;GWAS, genome-wide association study;LD, linkage disequilibrium;PRS, polygenic risk score;SNP, single nucleotide polymorphism.

maturation takes place and reproductive capacity is acquired. Puberty depends on the reactivation of hypothalamic gonadotropin-releasing hormone (GnRH) secretion and subsequent gonadal maturation. In males, the normal timing of puberty ranges from 9 to 14 years of age and is clinically marked by an increase of testicular volume to equal or greater than 4 mL, representing a Tanner Stage G2 (gonadarche).

Studies on twins, in different ethnic groups, and family studies suggest that 50% to 80% of the variance in pubertal timing may be genetically controlled in girls (1, 2) as well as in boys (3, 4).

With the emergence of microarray techniques, genome-wide association studies (GWAS) have revealed that one of the major milestones of female pubertal timing, age at menarche (AAM), represents a highly polygenic trait with 389 associated genetic loci, accounting for 7.4% of the trait variance (5). For most, the AAM can be quite precisely recalled, which makes it a very suitable marker for large GWAS.

In boys, GWAS of pubertal timing have used mostly self-reported proxies, such as age at voice break (6, 7) and pubertal height growth (8, 9). A GWAS on the initial clinical milestone of male puberty, ie, testicular enlargement, is not available since testicular size is not regularly assessed in most large-scale epidemiological studies. Consequently, fewer and smaller studies on male pubertal timing exist compared with studies on female pubertal timing. In addition, both voice break and menarche represent late events in the pubertal transition and are physiologically very different from early events, such as gonadarche and, in girls, the appearance of breast buds (thelarche).

Two large GWAS resources, 23andMe and the UK Biobank (5, 6, 10), hold information on the time at voice break and facial hair as male milestones for puberty. Initially, 11 independent genome-wide significant ($P < 5 \times 10^{-8}$) signals located at 9 genomic loci were identified for age at voice break (6). When combined, these 2 GWAS identified 78 independent significant loci associated with a synthetic variable for male puberty based on timing of voice break and facial hair (11). Remarkably, a moderate to strong genome-wide genetic correlation between males and females based on continuous data on voice break and AAM was reported. These findings reinforce the notion that the onset of puberty is a polygenic trait and indicate a shared genetic architecture between sexes.

In order to replicate and validate the results from GWAS data, the polygenic risk score method has the advantage of aggregating the effects of variants identified in large GWAS to predict genetic overlap between traits and to estimate a genetic propensity to a trait and heritability at the individual level in a small sample size group (12). In this sense, the replication analysis in a smaller population of men from the UK (ALSPAC) found that the polygenic risk score based on 78 voice break signals was also associated with self-reported advanced Tanner stage of genital development, age at appearance of facial hair, age at peak height velocity, and appearance of armpit hair (11).

Our aim was to test whether a polygenic puberty timing risk score based on genetic variants associated with age at voice break was also associated with early pubertal markers, such as age at gonadarche (AAG) and age at pubarche (AAP) in a cohort of clinically deeply phenotyped healthy Chilean boys who were followed longitudinally during puberty.

Methods

Subjects

Healthy Chilean boys participating in the Growth and Obesity Chilean Cohort Study (GOCS) with clinical pubertal staging were included in this study. GOCS is a longitudinal study that recruited children aged 2.6 to 4.9 years (n = 1195; 49.2% boys) from public nursery schools of 6 counties in Santiago during 2006. The inclusion criteria were (1) single birth during 2002-2003 (2) birth weight from 2500 to 4500 g; and (3) absence of physical (eg, skin burn), medical (eg, brain tumor), or endocrine diseases (eg, hyperthyroidism, hyperprolactinemia) that could alter the growth and/or onset of puberty. A complete description of the cohort has been previously reported (13). A sample was selected on the basis of having clinical staging of genital development (Tanner stages) as well as available high-quality DNA for genotyping. The GOCS protocol was approved by the institutional review board of the Institute of Nutrition and Food Technology (INTA) of the University of Chile. Informed consent was obtained from parents or guardians and children assented to participate when they turned 7 years of age.

Clinical assessment

Clinical examinations were conducted annually by dietitians until 2009. Thereafter, bi-annual evaluations with Tanner staging data collection were included. Secondary sex characteristics were evaluated by a single male dietitian trained specially for this purpose, with permanent supervision of a single pediatric endocrinologist (V.M.). The assessment included measurement of genital volume by palpation using Prader orchidometer and classification according to Tanner stages (14). In addition, we evaluated the presence of pubic hair and classified according to the different Tanner stages. Concordance between the dietitian and pediatric endocrinologist featured a kappa statistic = 0.8 for genitalia measurement (unpublished report). Testicular volume ≥ 4 mL (at least one testis) was considered to be a marker of onset of puberty (15). Body mass index (BMI) was calculated as weight (kg) divided by height squared (m^2) and BMI-for-age z score (z-BMI) was estimated based on the World Health Organization 2007 growth standards (16).

Genotyping

Genomic DNA was isolated from peripheral blood using QIAamp DNA Mini kit (Qiagen, Germany). Genotyping was performed using the Multi-Ethnic Global Array (MEGA; Illumina) on 914 participants of the GOCS cohort and 8 HGDP-CEPH Panel duplicate controls. Quality control was implemented in Genome Studio v2.0.3 (Illumina. Inc.). Participants with call rate < 0.98 and variants with heterozygous genotype on the X chromosome for males and genotypes on Y chromosome for females were removed. The concordance with the HGDP-CEPH Panel control was 99.99 ± 0.02% and the total genotyping rate was 0.9993. Further filtering was performed in PLINK 1.9 (17) based on gender mismatch, relatedness, heterozygosity rate and ancestry outliers. Single nucleotide polymorphisms (SNPs) with minor allele frequency less than 1%, missing genotype data > 5%, duplicated physical positions and deviations from the Hardy-Weinberg Equilibrium ($P < 1 \times 10^{-4}$) were excluded.

Calculation of the polygenic puberty timing risk score

GWAS summary statistics for relative timing of voice break (n = 154 459) based on the UK Biobank cohort study (http://biobank.ctsu.ox.ac.uk/crystal/field.cgi?id = 2385), were obtained from an online resource (http://www.nealelab.is/uk-biobank/). The summary statistics were analyzed and annotated using FUMA SNP2GENE platform (FUMA, http://fuma.ctglab.nl/), where 31 independent SNPs were significantly associated with relative timing of voice break.

UK Biobank genotyping data and their relative age of voice break ("younger than average", "about average" or "older than average") were available in 154 459 male individuals of white British ancestry. Effect estimates were obtained based on logistic regression models in 3 categories: "younger than average", when the subject answered to have had voice break younger than his peers; "about average age", if the subject answered to have had voice break at the same age like his peers; and "older than average", when the subject answered to have had voice break older than his peers (18). Significant and independent SNPs were identified as those with a genome-wide P value $<5 \times 10^{-8}$ and not in linkage disequilibrium (LD, $r^2 < 0.6$). The region comprising SNPs with $r^2 \geq 0.6$ compared to one of the significant and independent SNPs is denoted as the genomic risk locus and contains the candidate SNPs.

The polygenic puberty timing risk score was calculated using the following formula:

$$PRS = \sum_{i=1}^{k} \beta * SNP_i$$

Where, SNP_i is a genotype coded 0/1/2 such that the subject has 0, 1, or 2 risk alleles, ie, alleles associated with a delay in voice break, and β is the effect size reported for that SNP_i in the UK Biobank GWAS. The effect sizes represent log(OR), where a positive value means the effect size for an association with later voice break and a negative value means the effect size for an association with early voice break.

Of the 31 significant and independent SNPs associated with voice break extracted from the UK Biobank GWAS, only 5 SNPs were included in the Infinium Multi-Ethnic Global-8 kit (Illumina) which was used to genotype GOCS participants. For the remaining SNPs, the adjacent candidate SNP with the most significant P value in each genomic locus was selected. Two genomic regions (lead SNPs rs115260227 and rs2108321) were excluded since none of the candidate SNPs reach the threshold P value $<5 \times 10^{-5}$. Finally, 29 SNPs were included for the present analysis. The polygenic puberty timing risk score was computed as the summation of risk alleles across the 29 SNPs weighted by the effect size of each SNP.

Annotation and functional mapping of genes

Gene annotation was extracted using the GENE2FUNC function of FUMA platform (https://fuma.ctglab.nl/) applying the positional approach and a maximum distance of 10 kilobases. When the SNPs were not mapped to a protein coding gene using the proximity approach, we looked up these SNPs in the genetic portal Open Targets Genetics (https://genetics.opentargets.org/) which link a variant to potentially functional implicated genes integrating functional and biological data. Functional mapping of annotated genes was performed using the GENE2FUNC function of FUMA with default parameters.

Statistical analysis

To assess the association of the polygenic puberty timing risk score with AAG and AAP, we used linear regression model using standardized estimates obtained from Lifereg procedure in SAS (SAS Institute, Cary, NC). To estimate AAG and AAP, probit analyses (lifereg procedure) integrating left-, right-, and interval-censored observations were performed. Briefly, longitudinal data for boys who had the age at last gonadal Tanner stage 1 and the age at the first testicular volume ≥ 4 mL visit (Tanner stage 2), and the age at the last visit with absence of pubic hair and the visit immediately after with the presence of pubic hair, respectively, were included in the probit analyses as interval censored data. On the other hand, if a boy did not have testicular enlargement or presence of pubic hair at his last examination, the age at examination was used as right-censored data, and if a boy had any of these signs at his first examination, the age at examination was used as left-censored data.

To assess the association of individual SNPs with AAG and AAP, we used the "*association*" function in R Studio v3.5.2 under de codominant model. P values < 0.05 were considered statistical significant. Spearman correlation coefficient was used to assess correlation between AAG or AAP and BMI z-score.

Results

Clinical characteristics

Table 1 shows the AAG, AAP, and BMI z scores in the group of 401 boys with genotyping out of the 505 of the complete cohort who had the clinical evaluation. There were no differences in AAG, AAP, and the BMI z-score at these ages between the groups ($P > 0.05$). In the genotyped group, a significant inverse correlation was observed between AAG or AAP and BMI z-scores ($r = -0.209$; $P < 0.001$ and $r = -0.156$; $P < 0.001$, respectively)

Observation of pubarche as the first sign of puberty was higher (38.9 %) than the observation of gonadarche as the first manifestation (28.3%) or the synchronous appearance of these traits (32.8 %) in the cohort. On the other hand, precocious gonadarche or pubarche before the age of 9 years was observed in 8.9% and 8.7% of boys, respectively. Before the age of 7 years, only 0.3% and 1.2% of the boys presented gonadarche and pubarche, respectively. In the group of genotyped boys, 1/401 (0.2%) and 5/401 (1.2%) had gonadarche and pubarche before 7 years of age, respectively, and none presented both pubertal events before the age of 7 years.

Polygenic risk score and regression analysis with pubertal traits

The selected SNPs for the construction of the polygenic puberty timing risk score, their effect sizes and *P* values are shown in Table 2. SNPs in genomic locus 2, 5, 11, 14, and 21 were the lead SNPs of the region and the rest of the 24 SNPs were candidate SNPs. The candidate SNPs were also in high LD (D′: 0.751 to D′: 1.0) with the corresponding lead SNPs when analyzing the American population data of the 1000 Genomes Project Phase 3 in the LDlink application. The variant rs314268 located in the exonic region of *LIN28B* (6:105417978) reached the most significant *P* value ($P = 4.59 \times 10^{-38}$) in the GWAS (UK Biobank). This variant has been associated with Tanner genital stage in males and Tanner breast stage in females (19).

When the list of SNPs were analyzed independently in the GOCS cohort, 3/29 SNPs reached a *P* value at the nominal level ($P < 0.05$) for association (codominant model) with AAG (rs3126259, $P = 0.0043$ and rs1516882, $P = 0.0243$) or AAP (rs72901035, $P = 0.0292$). There were no notable differences in allele frequencies between our population and the signal-identified European population. The mean absolute difference in allele frequency was 15%, with the largest difference at rs2964296 (position chr5:166509450).

The polygenic puberty timing risk score was calculated in 401 boys from the GOCS cohort and it was normally distributed (Fig. 1). In order to determine the cumulative effect of significant risk loci for voice break on the time of pubertal onset we performed a linear regression analysis using censored data of Tanner genital and pubic hair stages (boys n = 21, n = 0, n = 380 and n = 26, n = 6, n = 369 with right censored values, left censored values and interval censored values for age at testis volume ≥4 ml and Tanner stage 2 of pubic hair, respectively). The regression analysis revealed a significant association of AAG ($\beta = 0.01$, $P = 0.041$, r^2 0.01) and AAP ($\beta = 0.185$, $P = 0.0004$, $r^2 = 0.034$) with the polygenic puberty timing risk score, respectively (Fig. 1A and 1B, respectively).

Next, we analyzed whether the AAG and the AAP changed in proportion to the polygenic puberty timing risk score by separating the population of scores in quintiles. This analysis showed that boys within the 20% highest scores entered puberty 0.55 ± 0.22 years later than boys within the 20% lowest scores and this difference is statistically significant ($P = 0.013$; 95% CI, 0.11–0.98, analysis of maximum likelihood parameter estimates) (Fig. 2A). Similarly, boys within the 20% highest scores experienced pubarche 0.67 ± 0.24 years later than boys within the 20% lowest scores ($P = 0.007$; 95% CI, 0.18–11.5) (Fig. 2A).

Additionally, we analyzed whether the time between the transition of Tanner stage G2 and Tanner stage G4 (pubertal tempo), and the timing of pubertal events defined as the difference between the AAG and the AAP associates with the polygenic puberty timing risk score. The regression analysis showed no association of these variables with the score (Fig. 2B). Furthermore, no association was observed between the polygenic puberty timing risk score and BMI *z* score at AAG or AAP.

Genomic characterization of voice break risk loci and related genes

Thirty-one loci were associated with voice break in the UK Biobank data, including 35 lead SNPs and 6197 candidate SNPs (18). Using the GENE2FUNC function of FUMA we aimed to annotate the mapped genes in their biological context. Within the 29 genomic locus selected for the construction of the polygenic puberty

Table 1. Age at Gonadarche and Age at Pubarche in Boys From the Growth and Obesity Chilean Cohort Study

	All boys (n = 505)	Boys with genotyping (n = 401)
Age at gonadarche	10.85 ± 1.42 (8.5–12.9)	10.77 ± 1.42 (8.51–12.91)
BMI z score at age of gonadarche	1.16 ± 1.23 (–0.88 to 2.9)	1.14 ± 1.24 (–0.93 to 2.91)
Age at pubarche	10.87 ± 1.59 (8.69–13.57)	10.83 ± 1.6 (8.73–13.58)
BMI z-score at age of pubarche	1.09 ± 1.18 (–0.93 to 2.81)	1.07 ± 1.2 (–0.94 to 2.79)

Values are mean ± SD (5th percentile—95th percentile).

Table 2. Twenty-Nine Selected SNPs Included in the Calculation of the Polygenic Risk Score

Genomic Locus	rsID	Chr	Position	Minor Allele	MAF	Effect Allele	EAF	EAF (GOCS)	Effect Size	SE	*P* value	Nearest Gene
1	rs4655723	1	66080784	T	0.199	T	0.181	0.291	-0.009	0.001	*6.56E-09*	*LEPR*
2	rs1514177*	1	74991402	C	0.428	G	0.576	0.611	0.007	0.001	*4.98E-10*	*FPGT-TNNI3K:TNNI3K*
3	rs12759783	1	199881715	A	0.476	A	0.490	0.402	0.007	0.001	*4.93E-10*	RP11-567B20.2/*NR5A2*
4	rs72901035[&]	2	156650258	T	0.121	T	0.129	0.075	0.011	0.002	*8.56E-11*	AC093375.1/*NR4A2*
5	rs1345417*	3	181511951	C	0.375	G	0.599	0.458	-0.009	0.001	*3.00E-15*	SOX2-OT/*SOX2*
7	rs13179411	5	133900513	T	0.159	T	0.151	0.288	0.009	0.002	*2.62E-08*	*JADE2*
8	rs55972276	5	135653737	A	0.126	A	0.139	0.070	0.009	0.002	*2.99E-08*	*TRPC7*
9	rs2964296	5	165936455	G	0.464	G	0.474	0.223	-0.007	0.001	*9.23E-09*	CTB-63M22.1/*TENM2*
10	rs60875622	6	100115317	G	0.292	G	0.308	0.291	-0.009	0.001	*1.11E-13*	Y_RNA/*PRDM13*
11	rs314268*	6	105417978	G	0.353	A	0.661	0.737	-0.016	0.001	*4.59E-38*	*LIN28B*
13	rs4728218	7	130989335	T	0.457	G	0.543	0.691	0.006	0.001	*5.19E-08*	*MKLN1*
14	rs7032296*	9	86713990	G	0.453	G	0.465	0.419	-0.006	0.001	*4.87E-08*	RP11-158D2.2/*HNRNPK; RMI1*
15	rs1516882[#]	9	108912911	G	0.322	G	0.314	0.323	-0.012	0.001	*2.40E-20*	RP11-6F6.1/*TMEM38B*
16	rs12003641	9	114279497	T	0.081	T	0.082	0.169	0.012	0.002	*2.81E-09*	*ZNF483*
17	rs6265	11	27679916	T	0.197	T	0.189	0.129	0.008	0.001	*3.81E-08*	*BDNF*
18	rs11037611	11	43788330	C	0.401	C	0.409	0.414	0.006	0.001	*2.75E-08*	*HSD17B12*
19	rs3824915	11	44331509	G	0.468	G	0.499	0.364	0.007	0.001	*3.20E-10*	*ALX4*
20	rs7107502	11	122830214	G	0.429	A	0.568	0.640	0.008	0.001	*3.47E-12*	*BSX*
21	rs12913832*	15	28365618	A	0.364	G	0.786	0.760	0.011	0.001	*1.06E-14*	*HERC2*
22	rs3743266	15	60781513	C	0.345	C	0.333	0.211	-0.007	0.001	*3.01E-09*	RP11-219B17.1/*RORA*
23	rs246185	16	14395432	C	0.328	C	0.322	0.246	0.011	0.001	*3.86E-18*	RP11-65J21.3/*MKL2*
24	rs11079719	17	43840006	G	0.244	G	0.224	0.168	0.010	0.001	*2.02E-13*	*ARL17B*
25	rs8081915	17	55227670	A	0.394	A	0.393	0.458	-0.006	0.001	*5.59E-07*	*AKAP1*
26	rs913001	20	54835406	A	0.345	C	0.706	0.817	0.007	0.001	*1.58E-07*	RP11-380D15.2/*MC3R*
27	rs3126259[#]	23	55574773	G	0.272	G	0.272	0.392	-0.004	0.001	*6.46E-06*	*PAGE2*
28	rs10482156	23	62318885	G	0.219	G	0.201	0.089	0.006	0.001	*1.45E-09*	*CBX1P1*
29	rs78050478	23	65100647	A	0.116	A	0.109	0.064	0.006	0.001	*4.02E-06*	RP11-314A15.2/*VSIG4*
30	rs111381836	23	100895354	A	0.173	A	0.150	0.103	-0.006	0.001	*5.30E-08*	RNU6-30P/*ARMCX6*
31	rs5932882	23	130426766	A	0.180	A	0.184	0.084	-0.009	0.001	*5.43E-19*	*IGSF1*

Nearest protein coding gene is denoted in italic. Effect size in the GWAS (logOR).
Abbreviations: EAF, effect allele frequency; MAF, Minor allele frequency; SE, standard error.
*Lead SNP of the genomic region. [#]Significant association with AAG in the codominant model. [&]Significant association with AAP in the codominant model

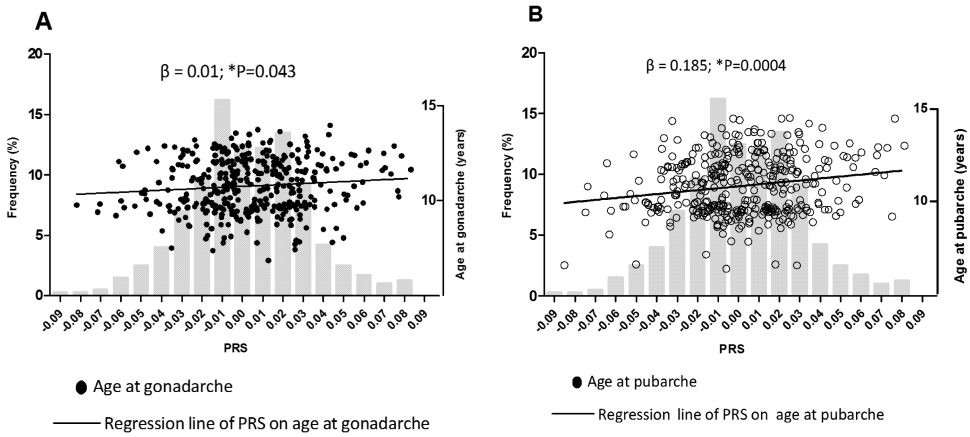

Figure 1. Polygenic puberty timing risk scores (PRS) frequency distribution and association with age at gonadarche (A) and pubarche (B). *Analysis of Maximum Likelihood Parameter Estimates (SAS).

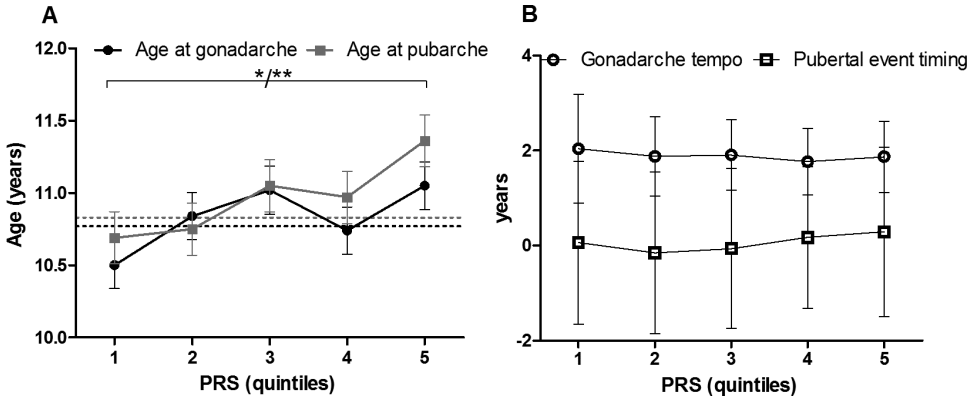

Figure 2. Comparison of pubertal traits in categorized data of polygenic puberty timing risk score (PRS). (A) Age at gonadarche (testicular volume ≥ 4) and age at pubarche expressed as mean ± standard error (SE) within PRS quintiles. The dotted line denotes the mean age at gonadarche and pubarche in the cohort. */** Significant difference between the first and the fifth quintile on age at gonadarche and pubarche respectively ($P = 0,.013$ and $P = 0.007$). (B) Pubertal tempo (time difference between the transition of Tanner stage G2 to Tanner stage G4) and pubertal event timing (time difference between age at gonadarche and age at pubarche) expressed as mean ± SE within PRS quintiles.

timing risk score, 59 protein coding genes were mapped. The GWAS catalog showed that these genes have been associated with a wide range of phenotypic traits; however, they are principally represented in GWAS associated to AAM, male pattern baldness, BMI and obesity, and age at voice break (Supplementary Fig. 1A)(20). In agreement with this data, tissue expression analysis using the result of annotated genes showed that most significant genes are expressed in the pituitary and the uterus (Supplementary Fig. 1B)(20). Moreover, tissue expression analysis of prioritized genes showed that these genes are upregulated in adrenal gland compared with other tissues, although no statistical significance is reached (Supplementary Fig. 1C)(20).

Discussion

In the present study, we calculated the genetic puberty timing risk score based on 29 independent loci associated with voice break from the UK biobank GWAS to test this combined effect on early pubertal events in Chilean boys. Our replication study comprises 401 boys who were longitudinally evaluated from the age of 3 years throughout puberty. Therefore, the AAG, defined as testicular enlargement, and the AAP were estimated by the mid-point visit or censored approaches.

Our analysis showed a positive association between the polygenic puberty timing risk score and AAG, suggesting that voice break and gonadarche share an overlapping genetic architecture. This finding are supported by a replication study in 1,554 boys belonging to the Avon Longitudinal Study of Parents and Children (ALSPAC) from United Kingdom, where a 78-SNPs genetic risk score based on independent signals for 2 male pubertal milestones, ie, voice break and facial hair, was associated with Tanner stage of genital development, including self-reported genitalia and pubic hair characteristics (11). Moreover, we found that boys who have the 20% higher polygenic puberty timing risk score entered puberty 0.55 years later compared with those with

the 20% lowest scores. In this sense, the UK biobank GWAS study showed that effect size estimates tend to be higher for relatively late voice break, suggesting estimated genome-wide heritability is higher for late onset of puberty in boys (5).

Furthermore, we found that AAP was strongly associated with the polygenic puberty timing risk score in the linear regression analysis, explaining 3.4% of the phenotypic variance. Although this cumulative effect size is minimal in terms of age, the significance is important to stablish a genetic association between these phenotypic traits. In addition, we showed that the burden of genetic signals with an advancing or delaying effect on age at voice break have the same direction in the appearance of pubic hair, denoting a common pubertal timing.

The appearance of pubertal hair or pubarche is rather a phenotypic trait of adrenarche than gonadarche (21), demonstrating the increasing levels of adrenal androgens and their action on peripheral tissues including the skin. Taking into consideration that androgens are largely responsible for changes in the larynx that lead to voice drop, it is tempting to think that much of the genetic variability encompassed in the polygenic puberty timing risk score is due to pathways involving not only gonadal but also adrenal androgens. Supporting this hypothesis, there is enrichment of genes with up-regulated expression in adrenal glands among the set of genes associated with voice break (GTEX analysis).

Gonadarche and pubarche are 2 early events in the progression of puberty. Although development of axillary and pubic hair requires adrenarcheal concentrations of androgen, clinically evident pubarche is usually preceded by physical evidence of gonadarche within 6 months between Tanner genital stage 3 (G3) and 4 (G4) in men of different ethnic groups (22–24). The Chilean population shows admixture of 52% European and 44% Amerindian origin determined by a set of ancestral identity markers (SNPs) (25, 26). Recent data from our cohort revealed that pubertal timing was not different in girls (age at thelarche) of indigenous or nonindigenous origin. On the contrary, in boys, indigenous origin was an independent risk factor for precocious gonadarche and pubarche (15).

On the other extreme of the chronological spectrum of pubertal development, peak high velocity and voice break are proxies for the degree of pubertal maturation occurring at genital stage 3 to 4 and during genital stage 4, respectively (27, 28). The progression through puberty or pubertal tempo usually lasts 3 to 4 years (21); however, it has been suggested that nutritional history, adiposity, and epigenetics may predict pubertal development in a sex-dependent manner (29). In our study, we could assess the development between genital stage 2 and 4 since most of our boys have not completed sexual maturity yet. This partial pubertal tempo represents 2.3 ± 1.17 years and it was similar throughout the genetic scoring range. This result suggests that the genetic variance that affects pubertal timing does not control the pubertal progression, which shows a steady velocity.

A common genetic basis between obesity and puberty has been postulated after known obesity-related SNPs (BMI, waist-hip ratio, and obesity) were associated with earlier AAM in girls (30). On the other hand, a genetic risk score constructed with known variants for advanced AAM demonstrated a positive association with BMI z-score prior to, during, and after puberty in girls and boys (31), suggesting that puberty and body composition in childhood and adolescence share a common genetic etiology (31). In this sense, genome-wide genetic correlation using LD score regression showed strong genetic correlation between male puberty timing, assessed by time of voice break, and BMI and adiposity traits such as body fat mass, hip circumference, and leg/arm fat percentage (6, 11, 32). Further, based on a Mendelian randomization approach, a recent study identified a causal effect of BMI on timing of voice break in boys (32). Although our voice break-based polygenic risk score does not reflect a shared genetic architecture between these 2 traits, in the present study, 3 of the selected loci, MC3R (33, 34), BDNF (35) and TNNI3K (36), were identified to be associated with obesity susceptibility in different populations. This observation suggests that some variants that have an effect on timing of pubertal onset may also control body composition.

In recent years, GWAS on AAM and age at voice break have been used to explain the genetic control of puberty onset. Nevertheless, in more comprehensive study, only 7.4% of the trait variability of AAM has been explained by genetic polymorphisms (5), and a similar result could be suggested for puberty in men, since many of these variants are associated with age at voice break. However, it is interesting to note that genes implicated in puberty disorders are also represented among these common signals, especially genes regulating the GnRH neuroendocrine network and the energy metabolism (37). In this sense, our polygenic risk score include significant signals mapping *LEPR*, *SOX2*, *IGSF1*, *JADE2* genes involved in the above-mentioned pathways (38, 39), as well as genes involved in sex steroid hormone metabolism and action such as *HSD17B12* and *NR5A2* (40, 41).

Some limitations can be recognized in our work. First, the cutoff for the GWAS *P* value to select the SNPs to be included in the polygenic puberty timing risk score was determined a priori as $P < 5 \times 10^{-5}$ and was not

calculated over a range of thresholds. However, with this approach we included most of the loci significantly associated with voice break in the GWAS and few effect estimates were excluded. Secondly, AAG and AAP could be determined in most but not all genotyped subjects, due to participants who did not reach a Tanner stage G2 at the last visit or abandoned the protocol. To settle this issue, censored data of Tanner genital and pubic hair stages was used.

In conclusion, our results indicate that genetic variants identified from large scale studies using late events of male puberty timing show a significant overlap with early pubertal events monitored in deeply phenotyped cohorts of peripubertal children, suggesting that these events share a genetic predisposition. Moreover, our findings point to a good genetic concordance for pubertal events in cohorts of different ethnicity. To our knowledge, this is the first study to associate male puberty with massive genotyping in a Chilean population.

Acknowledgments

Genotyping of GOCS was performed in the Human Genotyping laboratory at the Spanish National Cancer Research Centre, a member of CeGen (PRB2-ISCIII), and supported by grants PT13/ 0001/0005 of PE I + D + i 2013–2016 funded by ISCIII and European Regional Development Fund, and FONDECYT grant #1160833 (ICM and Instituto Milenio Fundamento de los Datos).

Financial Support: National Fund for Scientific and Technological Development of Chile to V.M. (grant number 1140447 and 1190346). Concurso Proyecto de Internalización UCH-1566–2018 to M.C.L.

Additional Information

Correspondence and Reprint Requests: Verónica Mericq, MD, P.O. Box 226–3, Santiago, Chile. E-mail: vmericq@med.uchile.cl.

Disclosure Summary: The authors have no conflicts of interest to disclose.

Data Availability: The datasets generated during and/or analyzed during the current study are not publicly available but are available from the corresponding author on reasonable request.

References

1. Palmert MR, Boepple PA. Variation in the timing of puberty: clinical spectrum and genetic investigation. *J Clin Endocrinol Metab*. 2001;**86**(6):2364–2368.
2. Sørensen K, Juul A, Christensen K, Skytthe A, Scheike T, Kold Jensen T. Birth size and age at menarche: a twin perspective. *Hum Reprod*. 2013;**28**(10):2865–2871.
3. Silventoinen K, Haukka J, Dunkel L, Tynelius P, Rasmussen F. Genetics of pubertal timing and its associations with relative weight in childhood and adult height: the Swedish Young Male Twins Study. *Pediatrics*. 2008;**121**(4):e885–e891.
4. Wehkalampi K, Silventoinen K, Kaprio J, et al. Genetic and environmental influences on pubertal timing assessed by height growth. *Am J Hum Biol*. 2008;**20**(4):417–423.
5. Day FR, Thompson DJ, Helgason H, et al.; LifeLines Cohort Study; InterAct Consortium; kConFab/AOCS Investigators; Endometrial Cancer Association Consortium; Ovarian Cancer Association Consortium; PRACTICAL consortium. Genomic analyses identify hundreds of variants associated with age at menarche and support a role for puberty timing in cancer risk. *Nat Genet*. 2017;**49**(6):834–841.
6. Day FR, Bulik-Sullivan B, Hinds DA, et al. Shared genetic aetiology of puberty timing between sexes and with health-related outcomes. *Nat Commun*. 2015;**6**:8842.
7. Ong KK, Elks CE, Li S, et al. Genetic variation in LIN28B is associated with the timing of puberty. *Nat Genet*. 2009;**41**(6):729–733.
8. Cousminer DL, Berry DJ, Timpson NJ, et al.; ReproGen Consortium; Early Growth Genetics (EGG) Consortium. Genome-wide association and longitudinal analyses reveal genetic loci linking pubertal height growth, pubertal timing and childhood adiposity. *Hum Mol Genet*. 2013;**22**(13):2735–2747.
9. Widén E, Ripatti S, Cousminer DL, et al. Distinct variants at LIN28B influence growth in height from birth to adulthood. *Am J Hum Genet*. 2010;**86**(5):773–782.
10. Sudlow C, Gallacher J, Allen N, et al. UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *Plos Med*. 2015;**12**(3):e1001779.
11. Hollis B, Felix D, Busch AS, et al. Expanded genomic analyses for male voice-breaking highlights a shared phenotypic and genetic basis between puberty timing and hair colour. *bioRxiv preprint server*. Posted November 30, 2018. doi:https://doi.org/10.1101/483933. Accessed 15 January 2019.
12. Choi SW, Mak TSH, O'Reilly P. A guide to performing polygenic risk score analyses. *bioRxiv preprint server*. Posted September 14, 2018. doi:https://doi.org/10.1101/416545. Accessed 15 January 2019.
13. Kain J, Corvalán C, Lera L, Galván M, Uauy R. Accelerated growth in early life and obesity in preschool Chilean children. *Obesity (Silver Spring)*. 2009;**17**(8):1603–1608.
14. Marshall WA, Tanner JM. Variations in the pattern of pubertal changes in boys. *Arch Dis Child*. 1970;**45**(239):13–23.
15. Fernández M, Pereira A, Corvalán C, Mericq V. Precocious pubertal events in Chilean children: ethnic disparities. *J Endocrinol Invest*. 2019;**42**(4):385–395.
16. World Health Organization. https://www.who.int/growthref/en/. Accessed 15 June 2011.
17. Purcell S, Neale B, Todd-Brown K, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet*. 2007;**81**(3):559–575.
18. Neale Lab. http://www.nealelab.is/uk-biobank/. Accessed 14 January 2019.
19. Cousminer DL, Stergiakouli E, Berry DJ, et al.; ReproGen Consortium; Early Growth Genetics Consortium. Genome-wide association study of sexual maturation in males and females highlights a role for body mass and menarche loci in male puberty. *Hum Mol Genet*. 2014;**23**(16):4452–4464.
20. Lardone MC, Busch AS, Santos JL, et al. Data from: A polygenic risk score suggests shared genetic architecture of voice break with early markers of pubertal onset in boys. Repositorio Academico Universidad de Chile. Deposited December 20, 2019. http://repositorio.uchile.cl/handle/2250/172976.
21. Abreu AP, Kaiser UB. Pubertal development and regulation. *Lancet Diabetes Endocrinol*. 2016;**4**(3):254–264.

22. Herman-Giddens ME, Wang L, Koch G. Secondary sexual characteristics in boys: estimates from the national health and nutrition examination survey III, 1988–1994. *Arch Pediatr Adolesc Med*. 2001;**9**(155):1022–1028.
23. Chulani VL, Gordon LP. Adolescent growth and development. *Prim Care*. 2014;**41**(3):465–487.
24. Sørensen K, Aksglaede L, Petersen JH, Juul A. Recent changes in pubertal timing in healthy Danish boys: associations with body mass index. *J Clin Endocrinol Metab*. 2010;**95**(1):263–270.
25. Fuentes M, Pulgar I, Gallo C, et al. [Gene geography of Chile: regional distribution of American, European and African genetic contributions]. *Rev Med Chil*. 2014;**142**(3):281–289.
26. Datos Genómicos – Chilegenómico. http://www.chilegenomico.cl/datos-genomicos/. Accessed 04 March 2016.
27. Juul A, Magnusdottir S, Scheike T, Prytz S, Skakkebaek NE. Age at voice break in Danish boys: effects of pre-pubertal body mass index and secular trend. *Int J Androl*. 2007;**30**(6):537–542.
28. Bordini B, Rosenfield RL. Normal pubertal development: part II: clinical aspects of puberty. *Pediatr Rev*. 2011;**32**(7):281–292.
29. German A, Shmoish M, Hochberg Z. Predicting pubertal development by infantile and childhood height, BMI, and adiposity rebound. *Pediatr Res*. 2015;**78**(4):445–450.
30. Elks CE, Perry JR, Sulem P, et al.; GIANT Consortium. Thirty new loci for age at menarche identified by a meta-analysis of genome-wide association studies. *Nat Genet*. 2010;**42**(12):1077–1085.
31. Johnson W, Choh AC, Curran JE, et al. Genetic risk for earlier menarche also influences peripubertal body mass index. *Am J Phys Anthropol*. 2013;**150**(1):10–20.
32. Busch AS, Hollis B, Day FR, et al. Voice break in boys-temporal relations with other pubertal milestones and likely causal effects of BMI. *Hum Reprod*. 2019;**34**(8):1514–1522.
33. Comuzzie AG, Cole SA, Laston SL, et al. Novel genetic loci identified for the pathophysiology of childhood obesity in the Hispanic population. *Plos One*. 2012;**7**(12):e51954.
34. Gerhard GS, Chu X, Wood GC, et al. Next-generation sequence analysis of genes associated with obesity and nonalcoholic fatty liver disease-related cirrhosis in extreme obesity. *Hum Hered*. 2013;**75**(2-4):144–151.
35. Thorleifsson G, Walters GB, Gudbjartsson DF, et al. Genome-wide association yields new sequence variants at seven loci that associate with measures of obesity. *Nat Genet*. 2009;**41**(1):18–24.
36. Zhao J, Bradfield JP, Zhang H, et al. Role of BMI-associated loci identified in GWAS meta-analyses in the context of common childhood obesity in European Americans. *Obesity (Silver Spring)*. 2011;**19**(12):2436–2439.
37. Howard SR. The genetic basis of delayed puberty. *Front Endocrinol (Lausanne)*. 2019;**10**:423.
38. Howard SR. Genes underlying delayed puberty. *Mol Cell Endocrinol*. 2018;**476**:119–128.
39. Hou H, Uusküla-Reimand L, Makarem M, et al. Gene expression profiling of puberty-associated genes reveals abundant tissue and sex-specific changes across postnatal development. *Hum Mol Genet*. 2017;**26**(18):3585–3599.
40. Hiltunen JK, Kastaniotis AJ, Autio KJ, Jiang G, Chen Z, Glumoff T. 17B-hydroxysteroid dehydrogenases as acyl thioester metabolizing enzymes. *Mol Cell Endocrinol*. 2019;**489**:107–118.
41. Yazawa T, Inanoka Y, Mizutani T, Kuribayashi M, Umezawa A, Miyamoto K. Liver receptor homolog-1 regulates the transcription of steroidogenic enzymes and induces the differentiation of mesenchymal stem cells into steroidogenic cells. *Endocrinology*. 2009;**150**(8):3885–3893.