

# Joint models reveal genetic architecture of transitions between pubertal stages and their association with BMI in a Latino population

Lucas Vicuña<sup>1,2,3</sup>, Esteban Barrientos<sup>3</sup>, Valeria Leiva-Yamaguchi<sup>4</sup>, Danilo Alvarez<sup>4</sup>, Veronica Mericq<sup>5</sup>, Ana Pereira<sup>6</sup>, Susana Eyheramendy<sup>2,3,7\*</sup>

**1** Department of Medicine, Genetics Section, University of Chicago, Chicago, USA

**2** Faculty of Engineering and Sciences, Universidad Adolfo Ibáñez, Santiago, Chile.

**3** Instituto Milenio Fundamentos de los Datos (IMFD), Santiago, Chile.

**4** MRC Biostatistics Unit, University of Cambridge, Cambridge, UK.

**5** Institute of Maternal and Child Research, Faculty of Medicine, University of Chile, Santiago, Chile.

**6** Institute of Nutrition and Food Technology, University of Chile, Santiago, Chile.

**7** Data Observatory Foundation, ANID Technology Center No. DO210001.

\* Corresponding author: [susana.eyheramendy@uai.cl](mailto:susana.eyheramendy@uai.cl)

## Abstract

Early or late pubertal onset can lead to disease in adulthood, including cancer, obesity, type 2 diabetes, metabolic disorders, bone fractures and psychopathologies. Thus, knowing the age at which puberty is attained is crucial as it can serve as a risk factor for future diseases. Pubertal development is divided into five stages of sexual maturation in boys and girls according to the standardized Tanner scale. We performed genome-wide association studies (GWAS) on the GOCS cohort composed of admixed children with European and Native American ancestry. Using joint models that integrate time-to-event survival parameters and longitudinal trajectories of body-mass index (BMI), we identified genetic variants associated with phenotypic transitions between pairs of Tanner stages. We identified 43 novel significant associations, most of them in boys. The GWAS on Tanner 3 → 4 transition in boys captured an association peak around the growth-related genes *LARS2* and *LIMD1* genes, the former of which causes ovarian dysfunction when mutated. The associated variants are expression- and splicing Quantitative Trait Loci regulating gene expression and alternative splicing in multiple tissues. Further, higher individual Native American genetic ancestry proportions predicted a significantly earlier arrival to Tanner 2 stage in boys but not in girls. Finally, the joint models identified longitudinal BMI parameters significantly associated in several Tanner stages' transitions, confirming the association of BMI on pubertal timing.

## Introduction

Puberty is a complex process influenced by various factors such as genetics, nutrition, and environmental factors. Early or delayed onset of puberty can have significant consequences on long-term health outcomes and disease risk [1]. For example, early onset of puberty in girls has been linked to an increased risk of breast cancer, endometrial cancer, type 2 diabetes, cardiovascular diseases and psychopathologies later in life [1] [2] [3]. This is explained at least in part by a longer exposure to estrogen and insulin growth factor (IGF-1), which can affect cell proliferation and differentiation, as

well as apoptosis [4]. Similarly, early puberty onset in boys has been associated with increased risk to obesity and high blood pressure in adulthood [1]. On the other hand, delayed puberty can lead to reduced adult height and decreased bone density, which can increase the risk of fractures and osteoporosis later in life [5] [6].

Pubertal maturation involves the appearance of a combination of physical changes that define secondary sexual characteristics in girls and boys. In girls, hallmark events are thelarche (the appearance of breasts), pubarche (the first appearance of pubic hair) and menarche (the first occurrence of menstruation), whereas in boys, hallmark events are changes in the scrotum and penis, pubarche, and increases in testicular volume. These sequences of events were described in seminal studies by Marshall and Tanner [7,8], who defined 5 stages of maturation from prepubertal (Tanner stage 1) to fully mature (Tanner stage 5), with Tanner stage 2 representing the onset of puberty.

Pubertal maturation is influenced by diverse factors. One such factor is adiposity. Higher adiposity (i.e. BMI) correlates with an earlier age of onset of puberty, especially in girls [9] [10]. Indeed, BMI is synergistically related to pubertal timing as well as to other anthropometric traits like height [11,12]. Thus, BMI is an important indicator of normal versus pathological sexual maturation [9,13–16]. Ultimately, these processes are intertwined and regulated by shared endocrine mechanisms controlled by the hypothalamic-pituitary-gonadal system [17].

Ethnicity can also affect the variability of pubertal phenotypes [9]. For instance, the mean age of onset of breast development is higher in White American girls than in African American girls (10.0 and 8.9 years, respectively) [17]. Also, while menarcheal age is > 13 years old in northeastern European girls, in Latin American populations like Chile [18] and Venezuela is ~ 12.5 years old. A study performed among Mapuche Indigenous and non-Indigenous girls from the Araucanía region in Southern Chile showed that the age at menarche is 12.6 years in Mapuche girls, whereas it is 12.2 years in non-Indigenous girls (after correcting for socio-economic status) [19]. Further, Mapuche origin in Chile is an independent risk factor for precocious gonadarche (i.e. the earliest gonadal changes) and pubarche in boys but not in girls [20]. However, whether these differences are due to genetic ancestry or other factors is currently not well-understood [17].

Genetic factors explain 70-80% of the variance in pubertal timing, as revealed by monozygotic twin studies [17]. For instance, a large-scale GWAS on a European population identified 389 genetic variants associated with menarche, explaining ~ 7.4% of the population variance [2]. In European boys, 76 variants were associated with the onset of facial hair (a marker of male puberty timing) [21].

The vast majority of genetic studies on pubertal variability have focused on cross-sectional phenotypes, namely, those that occur at specific time points (e.g. mean height at 8 years old), or the time point where a pubertal event takes place (e.g. age of menarche in girls). However, cross-sectional studies do not capture phenotypic changes over time. Instead, longitudinal models are suitable for analyzing phenotypic trajectories, such as BMI as a function of time. For instance, longitudinal GWAS have identified genetic variants associated with infant, child and/or adult BMI [22–24].

In this study, we investigated how genetic variability affects the transitions between sexual maturation stages. In particular, we were interested in estimating the length of time until the occurrence of a well-defined pubertal characteristic. We implemented survival models, which are best suited to analyse time-to-event observations. These models assume censored data as input [25–27]. There are different types of censorship, but we focused on interval and right censoring. The event is said to be interval censored when the event occurs within an interval of time, but the exact time of the event is unknown [28]. For example, in the context of this study, there was interval censorship in the data that records the stage number of the Tanner of a youth on each visit to the

practitioner. A youth could be on Tanner 2 on one visit and in Tanner 3 on the following visit. Between the two visits, i.e. in that interval of time, he or she has changed from Tanner 2 to 3. On the other hand, the (very common) right censoring is when the event of interest is not observed in the last registration time or the patient has dropout of the study. This situation occurs, for example, when the youth in his/her last visit was not yet in Stage Tanner 5.

In many studies, longitudinal measurements are recorded along time-to-event data, but usually these two sources of data are analyzed separately. In certain settings, a joint modeling approach can benefit the analysis. For instance, when the association between longitudinal and survival outcomes is of interest [29] [30].

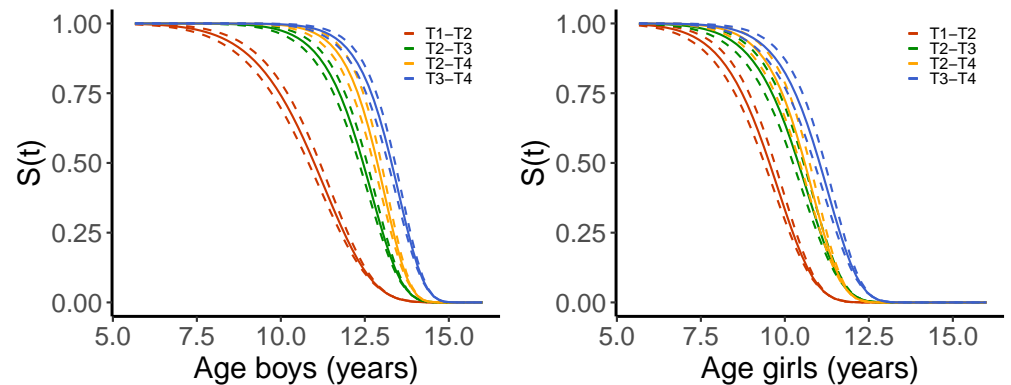
In this study, we implemented survival as well as joint models on data from admixed boys and girls with Native American and European ancestries. We were interested to assess whether some specific parameters from the BMI trajectories are associated with the transitions between Tanner stages. This can be studied in a joint model that combines the longitudinal modelling of BMI trajectories with time-to-event data fitting the transitions between Tanner stages.

In summary, the specific aims of this study are the following. First, to statistically model the transitions between Tanner stages. Second, to quantify how Native American genetic ancestry affects the timing of such transitions, when compared with European ancestry. Third, to identify genetic variants associated with the transitions. Forth, to assess the influence of BMI trajectories on the transition between Tanner Stages.

## Results

### Statistical modeling of transitions between Tanner stages.

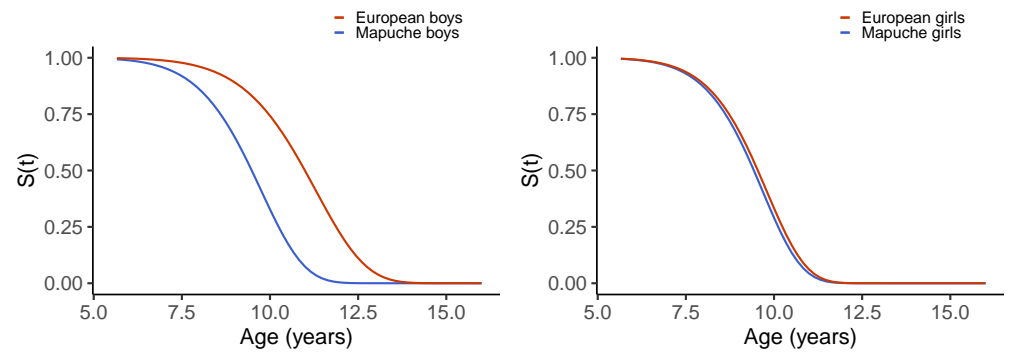
We studied the transitions between pairs of Tanner stages in boys and girls from the GOCS cohort. This cohort has longitudinal growth measurements from admixed Chilean children, collected for over 16 years [31]. We considered transitions from the consecutive Tanner stages pairs  $1 \rightarrow 2$ ,  $2 \rightarrow 3$  and  $3 \rightarrow 4$ . We also included the Tanner  $2 \rightarrow 4$  transition because it encompasses the total duration of puberty, thus enabling us to analyse shared genetic factors across puberty (next sections). We implemented survival models, which estimate the probability  $S(t)$  at each age that an event of interest has not yet occurred. In our case the event of interest was the time point at which the Tanner stage changed, which occurred within an interval delimited by two consecutive visits to the practitioner. In the case of Tanner  $1 \rightarrow 2$ , the lower limit of the interval was set at the age of adiposity rebound (Age-AR; see Methods). **Fig 1** shows the survival curves obtained for Tanner  $1 \rightarrow 2$ ,  $2 \rightarrow 3$ ,  $3 \rightarrow 4$  and  $2 \rightarrow 4$  transitions in boys and girls. The parameter values of the model are shown in **S1\_Table**. As expected, we observed that puberty in girls started and finished earlier than in boys (see **S2\_Table** for the estimated ages at  $S(t) = 0.25$ ,  $0.5$  and  $0.75$  in boys and girls). The progression from Tanner  $2 \rightarrow 3$  was significantly slower in males than in females at most ages (**Fig 1** and **S3\_Table**). For instance, at  $S(0.25)$ , the Tanner  $2 \rightarrow 3$  transition lasted 1.13 years in boys and 0.81 years in girls. **S3\_Table** shows the time intervals between Tanner stages for  $S(t) = 0.25$ ,  $0.5$  and  $0.75$ .



**Fig 1. Survival curves for girls and boys.** Probability  $S(t)$  of not moving to the next Tanner at a particular time for boys (left panel) and girls (right panel). Shown are the curves for Tanner 1  $\rightarrow$  2, Tanner 2  $\rightarrow$  3, Tanner 2  $\rightarrow$  4, and Tanner 3  $\rightarrow$  4 transitions.

### Effects of ancestry on transitions between Tanner stages.

We estimated the ancestry proportions of the individuals from the admixed GOCS cohort (**S1\_Fig**). We used the ADMIXTURE software,  $K=4$  ancestral populations, and European, Native American and African reference populations (see Methods). These proportions were on average 52.1% European, 43.8% Mapuche Native American, 2.6% Aymara Native American, and 1.5% African. As explained above, we fitted a survival model to estimate the probability  $S(t)$  that at a given age a youth has not yet attained a particular Tanner stage given that he/she is in the previous Tanner stage. In this model, we adjusted for ancestry proportion. To highlight the effect of global ancestry, we considered survival curves from hypothetical individuals with 100% Mapuche ancestry and with 100% European ancestry, referred to as “Mapuche” and “Europeans” from now on. Global ancestry did have a significant effect in Tanner 1  $\rightarrow$  2 transition in boys, with an effect size of 1.34 and a  $P$ -value = 0.026. However, global ancestry did not have a significant effect on the other transitions (**S1\_Table**). **Fig 2** shows the survival curves of the Tanner 1  $\rightarrow$  2 transition in Mapuche and European boys and girls. In quantitative terms, if we consider a 0.5 probability that boys have (not) attained Tanner 2 stage given that they are in Tanner 1, our model predicts that Mapuche and European boys will be 9.5 and 11.2 years old, respectively. **S4\_Table** shows the predicted ages for probabilities  $S(t)$  of 0.25, 0.5 and 0.75, which follow a similar trend.



**Fig 2. Effect of global ancestry on transitions between Tanner stages.** Survival curves, which represent the probability  $S(t)$  of not moving to the next Tanner at a particular age. Shown are the curves for the Tanner 1  $\rightarrow$  2 transition in Mapuche and European boys (left panel) and girls (right panel).

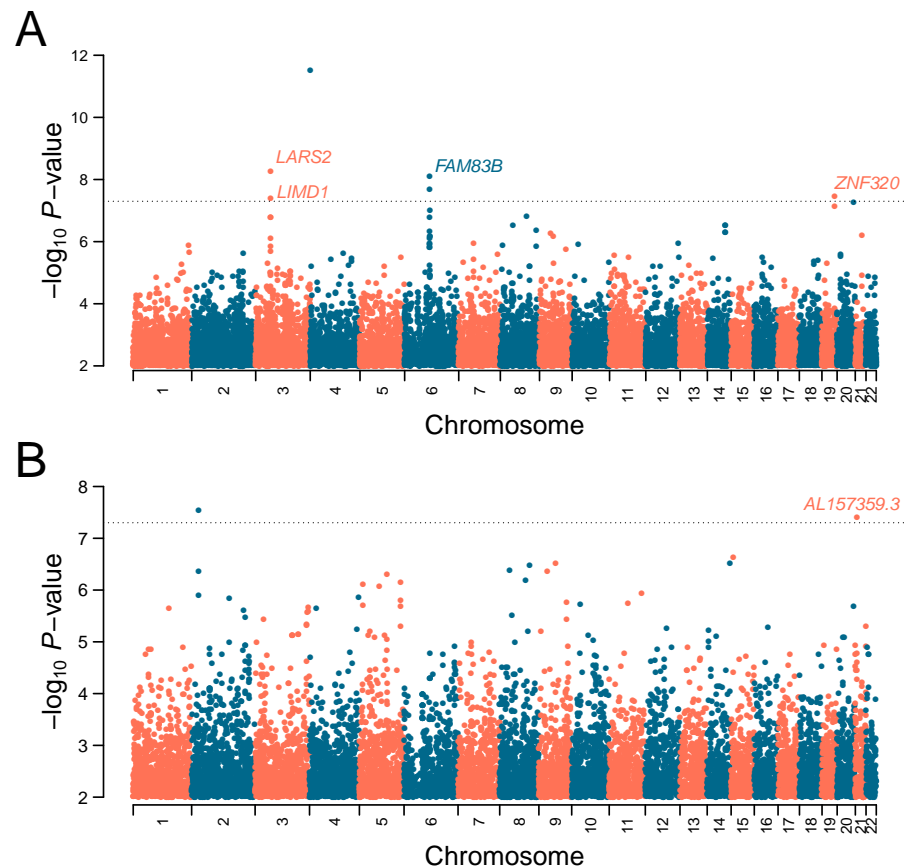
### GWAS on transitions between Tanner stages.

We sought to identify single nucleotide polymorphism (SNP) alleles associated genome-wide with transitions between pairs of Tanner stages in boys and girls separately. We implemented survival models (see details in Methods) according to the following strategy. First, we ran GWAS with survival models for the interval censoring data consisting on Tanner stages between visits of the adolescents to the practitioner. In these models, the covariates used were the genetic variant, local ancestry for the SNP, and the global ancestry for the adolescent. Second, we implemented joint models that incorporate as a covariate a parameter from the longitudinal modeling of the log-BMI trajectories into the survival model (see details in Methods). The later step is crucial because, as mentioned before, BMI can affect the onset of puberty. Third, since running the GWAS based on the joint models on the whole array of SNPs is computationally expensive, we selected the SNPs from the survival analyses with association  $P$ -values lower than 0.1, obtaining between 49,000 and 93,000 SNPs (depending on the analysis). The joint model's captured 43 significant associations ( $P$ -value  $< 5 \times 10^{-8}$ ) in total. **Table 1** shows these associations, together with the corresponding effect sizes and association  $P$ -values of the survival model, as well as functional annotations.

The GWAS for Tanner 1  $\rightarrow$  2, Tanner 2  $\rightarrow$  3, Tanner 3  $\rightarrow$  4 and Tanner 2  $\rightarrow$  4 transitions captured three, two, 20 and 18 significant associations, respectively. Out of them, 35 associations occurred in boys and 8 in girls. The 43 associations relate to 42 unique loci and 30 unique genes. The only SNP allele identified in two analyses was *rs464034-T*, which maps the long non-coding RNA *AL157359.3*, and was captured by the Tanner 2  $\rightarrow$  4 and Tanner 3  $\rightarrow$  4 GWAS in girls. No single SNP was significantly associated in both sexes in any analysis. Interestingly, the Tanner 3  $\rightarrow$  4 GWAS in boys captured several variants forming peaks around genes *LARS2*, *LIMD1* (chromosome 3) and *FAM83B* (chromosome 6). The Manhattan plot in **Fig 3** shows the GWAS associations of the survival analysis, which display the association peaks. We do not show the Manhattans of the joint model's GWAS because they were based on a reduced subset of SNPs. Significant associations surrounding less clear peaks also appear at chromosome 2 in boys and at chromosome 17 in girls in the Tanner 2  $\rightarrow$  4 GWAS (**S4 Figure**). Most of the significantly associated SNPs from the remaining GWAS seem to drive the association alone (i.e. without linked variants). **S2 Figure**, **S3 Figure** and **S4 Figure** show the Manhattan plots for Tanner 1  $\rightarrow$  2, Tanner 2  $\rightarrow$  3 and Tanner 2  $\rightarrow$  4 transitions from the survival models' GWAS.

SNP ID-Allele	Location	Consequence	Biotype	Gene	$\beta$ -GT (J)	P-val (J)	$\beta$ -GT (S)	P-val (S)	Sex	Stages
rs73094504-C	12:33134883	upstream	proc. pseud.	RP11-267D19.2	1.2	2.1E-08	1.0	3.4E-06	M	T1-T2
rs11830066-T	12:4470170	downstream	protein	C12orf5	-3.4	4.9E-08	-3.3	1.5E-07	M	T1-T2
rs17077311-G	6:148085737	intron/NCT	lincRNA	RP11-307P5.1	2.5	6.0E-08	2.2	1.6E-08	F	T1-T2
rs10061608-T	5:89527520	downstream	lincRNA	RP11-61G23.1	1.7	4.3E-09	1.5	5.0E-07	F	T2-T3
rs114974272-G	3:181497942	intron/NCT	sense ov.	SOX2-OT	2.3	3.6E-08	2.3	3.7E-08	M	T2-T3
rs10796166-A	10:14208652	intron	protein	FRMD4A	-5.9	3.9E-11	-5.6	1.2E-06	M	T3-T4
rs4812078-C	20:57957997	intergenic	-	-	-6.1	9.0E-11	-5.8	5.6E-08	M	T3-T4
rs1325549-C	6:40234603	downstream	lincRNA	RP11-552E20.1	-2.6	9.6E-11	-1.6	6.5E-03	M	T3-T4
rs34336178-T	3:45596658	downstream	protein	LARS2	-0.7	9.7E-10	-0.7	5.3E-09	M	T3-T4
rs17259299-T	14:87377092	intron	protein	RP11-322L20.1	0.8	2.0E-09	0.7	2.9E-07	M	T3-T4
rs73315357-C	14:87381550	intron	protein	RP11-322L20.1	0.8	2.0E-09	0.7	2.9E-07	M	T3-T4
rs13075111-G	3:45604131	intron/NCT	proc. tran.	LIMD1	-0.7	2.8E-09	-0.6	4.0E-08	M	T3-T4
rs464034-T	21:20341133	upstream	lincRNA	AL157359.3	0.8	6.2E-09	0.7	3.9E-08	F	T3-T4
rs7121055-A	11:23722786	intergenic	-	-	-3.0	6.4E-09	-2.0	7.2E-03	M	T3-T4
rs1448835-A	2:168531930	intergenic	-	-	-3.0	6.8E-09	-2.0	6.4E-03	M	T3-T4
rs147231055-A	16:47918000	intron/NCT	lincRNA	RP11-523L20.2	1.8	1.7E-08	1.6	4.6E-06	M	T3-T4
rs2704627-T	2:18648838	intergenic	-	-	-0.7	2.0E-08	-0.7	2.8E-08	F	T3-T4
rs962946-T	12:21797594	intron	protein	LDHB	-3.1	2.0E-08	-1.9	5.7E-02	M	T3-T4
rs1677078-T	12:21807432	intron	protein	LDHB	-3.1	2.0E-08	-1.9	5.7E-02	M	T3-T4
rs8100166-C	19:53396485	intron	protein	ZNF320	0.6	2.5E-08	0.6	3.4E-08	M	T3-T4
rs61761075-A	12:25397063	intron	protein	KRAS	1.3	2.9E-08	1.2	6.3E-06	M	T3-T4
rs12974965-T	19:53377733	downstream	protein	ZNF320	-0.6	3.2E-08	-0.5	7.4E-08	M	T3-T4
rs72939007-T	2:214226271	intron	protein	SPAG16	0.8	4.0E-08	0.6	3.3E-06	F	T3-T4
rs140141991-A	18:13613578	intron	protein	LDLRAD4	2.2	4.4E-08	1.8	5.5E-05	M	T3-T4
rs973205-A	6:54729949	intron	protein	FAM83B	-0.6	4.4E-08	-0.6	2.0E-08	M	T3-T4
rs13117815-A	4:566177	downstream	lincRNA	RP11-1263C18.1	-6.2	2.8E-10	-5.7	7.8E-09	M	T2-T4
rs7561821-G	2:33715742	intron	protein	RASGRP3	-0.7	4.2E-10	-0.7	1.8E-09	M	T2-T4
rs6716035-C	2:33711967	intron	protein	RASGRP3	-0.8	1.4E-09	-0.8	1.2E-08	M	T2-T4
rs28817094-G	1:218753327	intergenic	-	-	-0.9	4.2E-09	-0.8	1.4E-07	M	T2-T4
rs17676059-G	16:7574482	intron	protein	RBFOX1	-0.9	6.9E-09	-0.9	1.3E-08	M	T2-T4
rs464034-T	21:20341133	upstream	lincRNA	AL157359.3	0.8	8.1E-09	0.7	3.2E-07	F	T2-T4
rs2193600-T	7:53786903	intron/NCT	lincRNA	GS1-179L18.1	0.9	8.2E-09	0.8	5.4E-06	M	T2-T4
rs4496215-G	17:6038074	intergenic	-	-	0.8	9.3E-09	0.8	5.9E-09	F	T2-T4
rs3125761-C	13:55688006	intergenic	-	-	-0.6	1.1E-08	-0.5	6.2E-06	M	T2-T4
rs2052815-G	2:126440151	intergenic	-	-	-0.7	1.4E-08	-0.6	7.5E-07	M	T2-T4
rs13240918-A	7:47441254	intron	protein	TNS3	-2.0	2.0E-08	-1.8	1.4E-07	M	T2-T4
rs12220579-A	10:6313880	upstream	lincRNA	RP11-563J2.2	-1.0	2.8E-08	-0.9	4.4E-08	M	T2-T4
rs935710-A	8:93638037	intron/NCT	lincRNA	RP11-587H10.2	0.6	2.8E-08	0.6	1.5E-06	M	T2-T4
rs7602289-G	2:103445960	intron/NCT	proc. tran.	TMEM182	-0.7	3.8E-08	-0.7	4.0E-08	M	T2-T4
rs2741127-A	8:6731487	intron	protein	DEFB1	-0.7	3.9E-08	-0.6	2.3E-07	M	T2-T4
rs72858350-A	11:10891054	intron/NCT	antisense	ZBED5-AS1	-2.5	3.9E-08	-1.7	1.5E-06	M	T2-T4
rs7255633-G	19:54937469	intron	protein	TTYH1	-0.8	4.0E-08	-0.6	4.9E-06	F	T2-T4
rs10796166-A	10:14208652	intron	protein	FRMD4A	-6.3	4.8E-08	-5.3	3.2E-05	M	T2-T4

**Table 1. Significant GWAS associations for Tanner transition.** Shown are the significantly associated variants identified by the joint models. SNP rs ID with the associated allele, physical location, sequence ontology (SO) consequence, biotype, gene, effect size of the genotype ( $\beta_{GT}$ ) in the joint (J) model, association  $P$ -value of the genotype ( $P_{GT}$ ) in the joint model,  $\beta_{GT}$  in the survival (S) model,  $P_{GT}$  in the survival model, sex and Tanner transition stages. For each Tanner transition, variants were sorted in ascending order according to the  $P_{GT}$  in the joint model. M: male; F: female; NCT: noncoding transcript variant; proc.pseu.: processed pseudogene.



**Fig 3. Survival GWAS on Tanner 3 → 4 transition.** The Manhattan plots represent genotype-phenotype associations along the 22 autosomes in boys (A) and girls (B). Highlighted are genes harboring variants whose associations were significant in both the joint and survival models' GWAS. We show the plot made from the survival data, because the joint analysis was performed on a small subset of SNPs. The dotted black line represents the genome-wide significance threshold of  $P = 5 \times 10^{-8}$ . Variants with  $-\log_{10}(P\text{-value})$  between 0 and 2 are not shown.

### Effect of longitudinal BMI on Tanner transitions

We assessed the effect of longitudinal BMI on Tanner transitions through a joint model (described in the Method Section). The joint model is a survival model that incorporates as a covariate a random effect from the modeling of the log-BMI trajectories through a mixed model. We assessed four different covariates to share from the random effects of the longitudinal mixed model. **S5\_Table** shows the  $P$ -values for the association parameter when each of the four possible random effects were considered into the survival model. The four possible random effects considered were: the intercept, the slope, intercept plus the slope, and the mean (see details in the Methods Section). In all the joint models that we considered, the slope random effect was incorporated as a covariate. The effect of this covariate, called the association effect (or factor), was significant in boys in the Tanner 1 → 2 transition, with an effect size of 0.74 ( $P = 0.01$ ) (**S1\_Table**). In girls, the association factor was significant in Tanner 1 → 2 (effect size=0.9;  $P = 0.008$ ), Tanner 3 → 4 (effect size=1.23;  $P = 0.002$ ) and Tanner 2 → 4 (effect size=1.7;  $P = 1.6 \times 10^{-5}$ ) (**S1\_Table**).

## Functional annotations of associated genes.

We queried the NHGRI GWAS Catalog [32] to identify reported GWAS associations for the genes detected in our analyses. We only considered reported genes with genome-wide association  $P$ -values lower than  $5 \times 10^{-8}$ . Even though several genes have been previously GWAS-associated with diverse traits, we focused on those genes and traits more related to puberty. We defined three groups of trait categories: (i) Puberty, growth and endocrine function; (ii) obesity and anthropometric traits; and (iii) brain function (see Discussion). Each category was divided into several subcategories, and many genes belong to more than one (sub)category, reflecting pleiotropic effects. **Table 2** shows the identified genes, as well as their assigned categories and subcategories.

An important proportion of disease heritability is explained by loci that regulate the expression of genes (“expression Quantitative Trait Loci” or “eQTL”) and/or alternative splicing of pre-mRNA (“splicing Quantitative Trait Loci” or “sQTL”) [33]. Thus, we mined the Genotype-Tissue Expression (GTEx) [34] database to identify SNPs captured by our GWAS that are known eQTLs and/or sQTLs in single tissues. In total, we found 11 SNPs corresponding to eQTL and/or sQTL achieving significant associations in the Genotype-Tissue Expression (GTEx) project [34]. Among them, six are eQTLs only, one is a sQTL only, and four are both eQTL and sQTL. The 11 eQTLs were significantly associated with the expression of 17 unique genes (“eGenes”) in at least one tissue (154 associations in total) (**S1 File**). The five sQTLs were significantly associated with the splicing of pre-mRNAs of 6 genes (“sGenes”; 83 associations in total) (**S2 File**). Next, we queried the NHGRI GWAS Catalog to identify previous puberty-related GWAS associations for the eGenes and sGenes. These genes, as well as genes harboring an eQTL or sQTL captured by our GWAS were assigned to the same categories and subcategories as before (**Table 2**).



Trait category	Trait subcategory	Genes
Puberty, growth and endocrine function	Testosterone levels	<i>TNS3</i> <sup>§</sup> , <i>KRAS</i> <sup>§</sup> , <i>MFSD9</i> <sup>eE</sup> , <i>FGFRL1</i> <sup>s</sup> [35]
	Sex hormone-binding globulin levels	<i>KRAS</i> <sup>§</sup> [36], <i>FGFRL1</i> <sup>s</sup> [35]
	Age at first birth	<i>LARS2</i> <sup>§esES</sup> [37]
	Age at onset of menarche	<i>FAM83B</i> <sup>§e</sup> , <i>RBFOX1</i> <sup>§</sup> [38], <i>WSCD1</i> <sup>eE</sup> [38]
	Several male pubertal phenotypes	<i>SOX2-OT</i> <sup>§</sup> [21] [39] [40] [41] [42]
	Adolescent idiopathic scoliosis	<i>SPAG16</i> <sup>§es</sup> , <i>TNS3</i> <sup>§</sup> , <i>TMEM182</i> <sup>§</sup> , <i>RBFOX1</i> <sup>§</sup> , <i>FRMD4A</i> <sup>§</sup> [43]
	Heel bone mineral density	<i>FGFRL1</i> <sup>s</sup> [44]
Obesity and anthropometric traits	Type 2 diabetes	<i>LDLRAD4</i> <sup>§</sup> [45], <i>FRMD4A</i> <sup>§</sup> [46]
	BMI	<i>LDLRAD4</i> <sup>§</sup> , <i>SOX2-OT</i> <sup>§</sup> [47], <i>LIMD1</i> <sup>§esES</sup> [48], <i>RBFOX1</i> <sup>§</sup> [49], <i>LARS2</i> <sup>§esES</sup> [50]
	Waist-hip ratio	<i>ZBED5-AS1</i> <sup>§esE</sup> [38]
	Cholesterol levels	<i>KRAS</i> <sup>§</sup> [51]
	Body weight	<i>RBFOX1</i> <sup>§</sup> , <i>LDLRAD4</i> <sup>§</sup> [52]
	Birth weight	<i>TNS3</i> <sup>§</sup> [53]
	Height	<i>FRMD4A</i> <sup>§</sup> , <i>LIMD1</i> <sup>§esES</sup> , <i>LDHB</i> <sup>§e</sup> [38], <i>WSCD1</i> <sup>eE</sup> [54], <i>MFSD9</i> <sup>eE</sup> , <i>ZNF468</i> <sup>eE</sup> , <i>FGFRL1</i> <sup>s</sup> [55]
	Body size or surface area	<i>LARS2</i> <sup>§esES</sup> [56], <i>FGFRL1</i> <sup>s</sup> [57]
	Biological sex	<i>TMEM182</i> <sup>§</sup> [58]
Brain function	Educational attainment	<i>SPAG16</i> <sup>§es</sup> , <i>RBFOX1</i> <sup>§</sup> [59], <i>TMEM182</i> <sup>§</sup> [60], <i>TTYH1</i> <sup>§sS</sup> [61], <i>CDC42EP5</i> <sup>esE</sup> [60], <i>LENG8-AS1</i> <sup>eE</sup> [61], <i>LENG8</i> <sup>esES</sup> [60]
	Cognitive ability	<i>SPAG16</i> <sup>§es</sup> [59], <i>RBFOX1</i> <sup>§</sup> [62]
	Mathematical ability	<i>TMEM182</i> <sup>§</sup> [59]
	Neuroticism	<i>RBFOX1</i> <sup>§</sup> [63]
	Depression or bipolar disorder	<i>SPAG16</i> <sup>§es</sup> [64], <i>RBFOX1</i> <sup>§</sup> [65]
	Anxiety	<i>RBFOX1</i> <sup>§</sup> [63]
	Risk-taking behavior	<i>TMEM182</i> <sup>§</sup> [66]
	Cannabis or cigarette smoking	<i>TMEM182</i> <sup>§</sup> [67]
	Smoking initiation	<i>RBFOX1</i> <sup>§</sup> [68]
	Decaffeinated coffee consumption	<i>RBFOX1</i> <sup>§</sup> [69]
	Alcohol consumption	<i>FRMD4A</i> <sup>§</sup> [70], <i>WSCD1</i> <sup>eE</sup> [54]
	Diet habits	<i>TMEM182</i> <sup>§</sup> [71]
	Chronotype or morning person	<i>FAM83B</i> <sup>§e</sup> [72]

**Table 2. Functional annotations of genes.** Genes related to our GWAS were classified according to three trait categories and 29 subcategories related to puberty, based on published GWAS. According to existing variant and gene annotations, we classified the same genes into five categories. <sup>§</sup>: gene associated to variant captured by our GWAS; <sup>e</sup>: gene harboring eQTL captured by our GWAS; <sup>s</sup>: gene harboring sQTL captured by our GWAS; <sup>E</sup>: eGene of eQTL captured by our GWAS; <sup>S</sup>: sGene of eQTL captured by our GWAS. Several genes were assigned to more than one category.

## Discussion

In this study, we performed GWAS to identify genetic variants associated with the transitions between stages of pubertal maturation. Our methodological approach proved useful for capturing multiple genome-wide associations from a longitudinal cohort, even considering the small sample size of the GOCS cohort (**S6-Table**). To our knowledge, we are the first ones to implement this approach to dissect the genetic architecture of body growth, in particular of pubertal changes. The survival model's GWAS alone detected more than 70 significant associations. However, after including a parameter from the BMI trajectories into the joint model, the total number of significant associations was reduced to 43 (see association *P*-values in **Table 1**). This indicates that some of the associations were mediated by BMI. Noteworthy, no gene or variant was shared between boys and girls, indicating a different genetic architecture between sexes. The only locus detected in more than one Tanner transition was the *AL157359.3* variant allele *rs464034-T*, which was captured by the Tanner 3 → 4 and Tanner 2 → 4 transitions in girls (**Table 1**). This suggests that this locus plays a role during the whole duration of female puberty, but that its effect sizes in Tanner 1 → 2 and Tanner 2 → 3 might be smaller than in the Tanner 3 → 4 transition. Future studies on a bigger cohort will be needed to test this hypothesis.

Among the 43 novel associations captured in our analyses, we identified genes and variants that had not been previously related to puberty. Perhaps the most relevant are *LARS2* and *LIMD1*. The associated variant alleles *rs34336178-T* and *rs13075111-G* top the peak at chromosome 3 in the Manhattan plot from the Tanner 3 → 4 transition in boys (**Fig 3**). *rs34336178-T* is located 1 bp downstream of *LARS2* and 228 bp of *LIMD1*. *LARS2* encodes the mitochondrial leucyl-tRNA synthetase, and enzyme that catalyzes the aminoacylation of tRNAs to bind their cognate amino acid during translation. Mutations in this gene cause Perrault syndrome, a rare autosomal recessive disorder which causes ovarian dysfunction in females, and bilateral sensorineural hearing loss in females and males [73]. *LIMD1* is a tumor-suppressor gene involved in a variety of functions, including gene expression regulation. Of note, *LARS2* and *LIMD1* have been GWAS-associated with obesity-related and anthropomorphic traits, such as BMI and height (**Table 2**). Importantly, *rs34336178* is a sQTL that regulates the alternative splicing of its own gene in six tissues, and it is also an eQTL that regulates its own expression, as well as *LIMD1*'s expression across 27 diverse tissues (**SFile1** and **SFile2**). The intronic/non-coding transcript variant *rs13075111* is an eQTL and sQTL regulating the expression of *LARS2*, *LIMD1* and/or *SACM1L* across 45 and 21 tissues, respectively (**SFile1** and **SFile2**). The observation that there are loci that work as eQTLs and sQTLs is expected, since alternative splicing can alter gene expression [33].

We also confirmed genes that had been GWAS-associated with pubertal phenotypes. One example is *SOX2-OT* (**S3-Figure**), which has been associated with pubertal and endocrine-related traits mostly in males, such as late vs. average onset facial hair, age at voice breaking [21], male-pattern baldness [39], excessive hairiness [40], monobrow [41] and eyebrow thickness [42]. The *SOX2-OT* variant allele *rs114974272-G* was detected in the Tanner 2 → 3 transition in boys. Interestingly, it is Tanner 3 stage where the spurt of pubic hair occur in boys [8]. *SOX2-OT* is a transcriptional regulator of *SOX2*, which is one of the major regulators of pluripotency [74]. *SOX2* is located within the sex-determining region of the Y chromosome (SRY) [75], which is necessary for male sex determination in mammals. Also, it is a key gene for the normal development and function of the hypothalamo-pituitary and reproductive axes, as revealed by clinical phenotypes of patients harboring heterozygous sequence variations in this gene [75]. We found three genes –*FAM83B*, *RBFox1* and *WSCD1*– in boys that have been directly GWAS-associated with female puberty, in particular with the age at onset of menarche (**Table 2**). However, we did not find associations between these

genes' variants and Tanner transitions in girls. This result makes sense, since a proportion of menarche loci are important for pubertal initiation in boys as well [10]. The *FAM83B* intronic variant *rs973205*, which is part of the peak at chromosome 6 in the Manhattan plot of the Tanner 3 → 4 transition in boys (**Fig 3**), is an eQTL associated with the regulation of the undescribed *RP11-524H19.2* gene in the suprapubic skin, lower leg skin and esophageal mucosa (**SFile1** and **SFile2**).

Another interesting finding is the *TTYH1* intronic variant *rs7255633*, which was captured by the GWAS on Tanner 2 → 4 transition in girls. This variant is an eQTL and sQTL in 59 and 50 tissues, respectively. Its associated sGenes are *TTYH1*, *LENG8*, whereas its eGenes are *LENG8*, *CDC42EP5*, *LENG8-AS1*, *AC008746.12* and *CTD-2587H19.2*. Most of these genes have been GWAS-associated with educational attainment (**Table 2**), suggesting that *rs7255633* might play a role in brain functions related to sexual maturation across female puberty. The observation that many of the genes identified by our GWAS have known associations with brain function is in agreement with the critical bilateral effects between puberty and the brain. For example, our results show multiple associations with psychiatric conditions such as depression, bipolar disorder, anxiety, neuroticism or substance abuse (**Table 2**), and it is known that incidence rates of psychiatric disorders peak in adolescence [76].

We analysed and quantified how pubertal transitions are affected by Native American and European genetic ancestry. In all our survival and joint model we incorporated the global ancestry of each individual and local ancestry at each SNPs to adjust for population structure. Our results show that increased Mapuche proportions are significantly associated with an earlier arrival to Tanner 2 in boys. This result supports previous epidemiological findings in the GOCS cohort which show significant associations between Mapuche origin (based on the number of Mapuche last names) and precocious gonadarche and pubarche in boys [20]. While gonadarche refers to the earliest gonadal changes of puberty, pubarche refers to the first appearance of pubic hair at puberty. Moreover, a previous study by our group on the GOCS cohort found that the age at peak height velocity, namely, the age where the maximum rate of growth occurs during puberty (~ 12.7 years old among admixed Chileans), is 0.73 years earlier in Mapuche compared with European adolescents [77]. These observations provide further evidence that genetic ancestry is a relevant factor affecting the early physiological changes during puberty, particularly the first sexual maturation changes triggered in Tanner 2 stage.

A limitation of our study was that we could not replicate our findings in an independent cohort due to the lack of longitudinal pediatric growth cohorts with comprehensive data on Tanner stages, BMI trajectory and having a Native American ancestry component.

The results of this study are important to understanding the genetic architecture of pubertal changes, which hitherto has barely been analysed. In addition, they hold medical relevance, since they will serve to estimate with higher accuracy the genetic risk to potential associated diseases in adulthood. Finally, our results highlight the importance of implementing joint models of survival and longitudinal variables in GWA studies of growth processes, and hold great potential to get more insights into their complex genetic architecture.

## Methods

### Determination of Tanner stages

The “Growth and Obesity Chilean Cohort Study” (GOCS) [31] includes children with singleton births only, gestational age between 37 – 42 weeks, birth weight of 2,500 to

4,500 g, and no physical or psychological conditions that could severely affect growth. Among the 1196 participants of the cohort 943 (489 girls and 454 boys) met all inclusion criteria. To measure the sexual maturation of adolescents from GOCS, early clinical anthropometric evaluations were performed until 2009. Thereafter, at age 6.7 years, a single pediatric endocrinologist assessed breast, pubarche as well as genital development by palpation and classified breast, pubic hair and testes according to Tanner [7] [8]. Afterward, every 6 months, secondary sex characteristics were evaluated by a single dietitian trained for this purpose, with permanent supervision from the pediatric endocrinologist. Concordance between the dietitian and pediatric endocrinologist was 0.9 for breast and genitalia evaluation [31]. Age at menarche was self-reported. Since several boys did not want their testicles' to be examined, Tanner 4 in males was labeled as the time of break voice. On the subset of boys that agreed to be examined, there was always an agreement between the ascertainment of Tanner 4 stage when labeled as break voice as well as testicular enlargement.

## Genotyping

We used SNP array data from GOCS adolescents obtained in a previous study [77]. Individuals were genotyped with the Infinium<sup>®</sup> Multi-Ethnic Global BeadChip (Illumina). We used Plink 1.9 [78] to exclude 18 samples with call rate <0.98 (18 samples), 10 samples with gender mismatch, and one sample from each pair of highly related individuals (IBS >0.2). We excluded variants with a minor allele frequency (MAF) < 0.01, and variants following at least one of the following conditions: have heterozygous genotypes on male X chromosome, call genotypes on the Y chromosome in females, have high heterozygosity ( $\pm 3$  SD from the mean), have >5% missing genotype data, have duplicated physical positions (one variant was kept from each duplicate pair) and show significant deviations from Hardy-Weinberg equilibrium ( $P = 1 \times 10^{-6}$ ). We removed A-T and C-G transversions to avoid inconsistencies with the reference human genome. We also excluded 25 boys whose last BMI measurement was taken before they were 12 years old. After quality control filtering, we obtained the final data set of 904 individuals and 521,788 autosomal SNPs.

## Local Ancestry Inference.

We used RFmix.v.2.0 [79] to infer the local ancestry of each SNP allele from our Chilean sample. As reference populations we used Yoruba (YRI, n=108) for the African ancestry, Iberian Populations in Spain (IBS, n=107) for the European ancestry, Peruvian in Lima Peru (PEL) individuals with > 95% Native American ancestry (n=29). All of these samples are from the 1000 Genomes Project [80]. We excluded individuals with > 5% SNP missing rate. We estimated the haplotype phase from genotype data with Shapeit2 [81], using the HapMap37 human genome build 37 recombination map as a reference. We used the `forward-backward` parameter to run the software. RFmix identified three ancestries: Native American, European, and African.

## Global Ancestry Inference.

Global ancestry proportions of Chilean children were estimated with Admixture 1.3.0 [82] in unsupervised mode. We used Yoruba (YRI, n=108) as the reference population for the African ancestry and Iberian Populations in Spain (IBS, n=107) for the European ancestry [80]. We used a merged dataset of 11 Mapuche [83] and 73 Aymara [84] [85] as the reference Native American population panel. Using K=4 the software clearly identified 4 ancestral groups corresponding to European, African, Aymara and Mapuche. The latter two groups represent the main Native American

sub-ancestries of admixed Chileans. We used the Mapuche ancestry proportion as a covariate in the GWAS regressions. To distinguish Peruvian (PEL) individuals with > 95% Native American ancestry, we used K=3 ancestral groups and the aforementioned reference populations.

## Variant and gene annotations.

SNP annotations were retrieved with the web server Variant Effect Predictor (VEP) from Ensembl [86], using the GRCh37 (hg19) human genome assembly. Upstream and downstream variants were defined as those located 10 Kb upstream or downstream of the gene, respectively. Intergenic variants were defined as those located > 100 Kb upstream or downstream of the closest gene. Reported GWAS associations were retrieved from the NHGRI GWAS Catalog [32]. We only considered genome-wide significant associations ( $P < 5 \times 10^{-8}$ ). When more than one variant in the same gene has been associated with the same phenotype, we reported the strongest association. eQTLs and sQTLs data were retrieved using the Genotype-Tissue Expression (GTEx) Project [34] portal and variants achieving the significance threshold of  $P = 2.5 \times 10^{-7}$  for eQTL mapping were considered significant [34].

## Derivation of the age of adiposity rebound

The derivation of Age-AR was described previously [87]. Before the derivation, 51 individuals were excluded because they had not enough measurements between 2 and 10 years old. Also, 156 individuals were excluded because all their measurements were taken after 4.5 years old. The final dataset had 696 individuals with measurements between 2 and 10 years old. The sample sizes for each analysis are displayed in **S6\_Table**. To estimate Age-AR, the following longitudinal statistical model of BMI was implemented:

$$\text{LBMI}_i(t) = (\beta_1 S_i + \beta_0) + (\beta_2 S_i + \beta_3)t + (\beta_4 S_i + \beta_5)t^2 + (\beta_6 S_i + \beta_7)t^3 + b_{0i} + b_{1i}t + b_{2i}t^2 + b_{3i}t^3 + \epsilon_i(t), \quad i = 1, \dots, 696, \quad (1)$$

where  $\text{LBMI}_i(t)$  represents the log-BMI of individual  $i$  at time/age  $t$  and  $S_i$  represents the individual's sex (0 female, 1 male). The parameters  $\beta_k$  for  $k = 1, \dots, 7$  are fixed effects whereas  $b_{ki}$  for  $k = 0, \dots, 3$  are random effects for each individual  $i$  where  $b_{ki} \sim \text{Normal}(0, \sigma_k^2)$ . The error terms  $\epsilon_i(t)$  are assumed independent across the different individuals  $i$  but dependent between observations for the same individual and are normally distributed with mean at zero and variance  $\sigma^2$ .

The predicted trajectory for individual  $i$  can then be written as

$$\widehat{\text{LBMI}}_i(t) = (\hat{\beta}_1 S_i + \hat{\beta}_0) + (\hat{\beta}_2 S_i + \hat{\beta}_3)t + (\hat{\beta}_4 S_i + \hat{\beta}_5)t^2 + (\hat{\beta}_6 S_i + \hat{\beta}_7)t^3 + \hat{b}_{0i} + \hat{b}_{1i}t + \hat{b}_{2i}t^2 + \hat{b}_{3i}t^3, \quad (2)$$

where  $\hat{\cdot}$  (hat) is the estimator of the parameters. Age-AR is obtained by finding the time/age  $t$  at which the minimum BMI is found, by solving the equation:

$$3(\hat{\beta}_6 S_i + \hat{\beta}_7 + \hat{b}_{3i})x_i^2 + 2(\hat{\beta}_4 S_i + \hat{\beta}_5 + \hat{b}_{2i})x_i + (\hat{\beta}_2 S_i + \hat{\beta}_3 + \hat{b}_{1i}) = 0, \quad (3)$$

which has as its solution:

$$\text{Age-}\hat{\text{AR}}_i = \frac{1}{3(\hat{\beta}_6 S_i + \hat{\beta}_7 + \hat{b}_{3i})} \left( -(\hat{\beta}_4 S_i + \hat{\beta}_5 + \hat{b}_{2i}) \pm 2\sqrt{(\hat{\beta}_4 S_i + \hat{\beta}_5 + \hat{b}_{2i})^2 - 3(\hat{\beta}_6 S_i + \hat{\beta}_7 + \hat{b}_{3i})(\hat{\beta}_2 S_i + \hat{\beta}_3 + \hat{b}_{1i})} \right), \quad (4)$$

where the minimum obtained above needs to satisfy:

$$6(\hat{\beta}_6 S_i + \hat{\beta}_7 + \hat{b}_{3i}) \text{Age-AR}_i + 2(\hat{\beta}_4 S_i + \hat{\beta}_5 + \hat{b}_{2i}) > 0. \quad (5)$$

## GWAS for transitions between Tanner stages using a survival model

A survival model estimates the probability that a certain event has not occurred yet, which depends on some covariates. The hazard function represents the probability that the event occurred at time  $t$ .

Regardless of the type of censorship present in the data (interval or right censoring), the hazard function is usually estimated using a specification either parametric (e.g., Weibull baseline hazard) or semi-parametric (e.g., Cox proportional hazard) [88]. The hazard function for both approaches can be written, for an individual  $i$  at age  $t$ , as follows:

$$h_i(t) = h_0(t) \exp \{ \gamma_1 \text{GT}_{ij} + \gamma_2 \text{GA}_i + \gamma_3 \text{LA}_{ij} \}, \quad (6)$$

where  $h_0(t) = \lambda \phi t^{\phi-1}$  specifies a Weibull baseline hazard while Cox's proposal  $h_0(t)$  is left unspecified. The following are the covariates in our models:  $\text{GT}_{ij}$  is the additive representation of the genotype of SNP  $j$ , which takes values in  $\{0, 1, 2\}$ ;  $\text{GA}_i$  represents the global Mapuche Native American ancestry proportion; and  $\text{LA}_i$  corresponds to the local ancestry that takes the values 0, 1 or 2 depending on the number of Native American alleles for SNP  $j$  and individual  $i$ . The effect of each variable is measured through its respective coefficient,  $\gamma = (\gamma_1, \gamma_2, \gamma_3)$ .

## Transitions between Tanner stages using a joint model of longitudinal and survival variables

Our joint modeling involves specifying two submodels, one longitudinal and one survival, that share information. The longitudinal model fits the trajectories of the logarithm of the BMI for each individual as a function of age using a mixed model. The survival model estimates the probability that individual  $i$  at age  $t$  has not yet attained Tanner stage  $k$ , for some  $k \in \{2, 3, 4\}$ , as a function of some covariates and some random effects estimated in the longitudinal models. By bringing random effects from the longitudinal model as covariates in the survival model we share information from the BMI trajectory of each individual. Different models can be specified by selecting a different set of random effects from the longitudinal model. After assessing several models we selected the model with the best performance according to the AIC criteria.

Specifically, the longitudinal submodel is a mixed model defined as follows [89]:

$$\begin{aligned} \text{LBMI}_i(t) &= \beta_0 + b_{0i} + (\beta_1 + b_{1i})t + \beta_2 t^2 + \beta_3 t^3 + \epsilon_i(t), \\ \mathbf{b}_i &\sim \text{Normal}(\mathbf{0}, \mathbf{\Sigma}) \quad \perp \quad \epsilon_i(t) \sim \text{Normal}(0, \sigma^2), \end{aligned} \quad (7)$$

where  $\text{LBMI}_i(t)$  represents the log-BMI of individual  $i$  at age  $t$ ;  $(\beta_0, \beta_1, \beta_2, \beta_3)$  are intercept, linear, quadratic, and cubic fixed effects, respectively;  $\mathbf{b}_i = (b_{0i}, b_{1i})$  are intercept and slope random effects, respectively; and  $\epsilon_i(t)$  is an error term.  $\mathbf{\Sigma}$  and  $\sigma^2$  are the variance-covariance matrix of random effects and the variance of error, respectively, and are parameters to be estimated.

The survival submodel (6) is defined as follows:

$$h_i(t) = h_0(t) \exp \{ \gamma_1 \text{GT}_{ij} + \gamma_2 \text{GA}_i + \gamma_3 \text{LA}_{ij} + \alpha b_{1i} t \}, \quad (8)$$

where  $\text{GT}_{ij}$  is the additive representation of the genotype of SNP  $j$  at individual  $i$  taking values in  $\{0, 1, 2\}$ ,  $\text{GA}_i$  represents the global Mapuche Native American ancestry

proportion,  $LA_{ij}$  corresponds to the local ancestry and takes the values 0, 1 or 2 depending on the number of Native American alleles for SNP  $j$  and individual  $i$ .  $\alpha$  quantifies the strength of the association between the longitudinal and survival processes, the so-called association effect. Note that we are only sharing the time-dependent random effect (i.e., the slope). In our context, we are interested in the association of this effect (BMI trend) with time to puberty.

Estimating the parameters and random effects simultaneously is a computationally demanding task. In particular, marginal likelihood is required (i.e., integrating out the random effects), making the inference even more expensive and impractical for large datasets. For these reasons, we implement the procedure proposed by [90], which consists of two-stage inference: (1) the longitudinal submodel is fitted, and (2) the estimation of  $b_{1i}$  is used as a covariate into the survival submodel.

## Consent to Participate

Informed consent from participants was obtained from parents or guardians. Children agreed to participate when they turned 7 years old. This study was approved by the Scientific Ethics Committees of Instituto de Nutrición y Tecnología en Alimentos (INTA) and Pontificia Universidad Católica de Chile.

## Funding

This work was supported by the Fondo Nacional de Ciencia y Tecnología (FONDECYT) [1200146 to S.E., L.V. and E.B.; 1190346 to V.M.]. S.E., and L.V. were additionally supported by the Instituto Milenio de Investigación Sobre los Fundamentos de los Datos (IMFD). D.A. was supported by the UKRI Medical Research Council, grant number MC.UU.00002/5.

## Author Contributions

S.E. conceived the study. S.E., L.V. and D.A. designed experiments. E.B., L.V. and V.L-Y. performed analyses. V.M. and A.P. performed anthropometric measurements and collected medical data. L.V., S.E. and D.A. wrote the manuscript. All the authors critically reviewed and accepted the final version.

## Conflict of Interests Statement

The authors declare that there is no conflict of interest.

## Data Availability Statement

The genetic data used in this study are from adolescents, many of which are less than 18 years old. Thus, we are not allowed to publish or share their raw data. The summary statistics of this study can be found in the NHGRI-EBI GWAS Catalog server [32], under the code GCP000453.

## References

1. Day FR, Elks CE, Murray A, Ong KK, Perry JRB. Puberty timing associated with diabetes, cardiovascular disease and also diverse health outcomes in men and women: the UK Biobank study. *Sci Rep*. 2015;5:11208. doi:10.1038/srep11208.
2. Day FR, , Thompson DJ, Helgason H, Chasman DI, Finucane H, et al. Genomic analyses identify hundreds of variants associated with age at menarche and support a role for puberty timing in cancer risk. *Nature Genetics*. 2017;49(6):834–841. doi:10.1038/ng.3841.
3. Graber JA. Pubertal timing and the development of psychopathology in adolescence and beyond. *Hormones and Behavior*. 2013;64(2):262–269. doi:<https://doi.org/10.1016/j.yhbeh.2013.04.003>.
4. Biro FM, Huang B, Wasserman H, Gordon CM, Pinney SM. Pubertal Growth, IGF-1, and Windows of Susceptibility: Puberty and Future Breast Cancer Risk. *Journal of Adolescent Health*. 2021;68(3):517–522. doi:10.1016/j.jadohealth.2020.07.016.
5. Vandendput L, Kindblom JM, Bygdell M, Nethander M, Ohlsson C. Pubertal timing and adult fracture risk in men: A population-based cohort study. *PLOS Medicine*. 2019;16(12):e1002986. doi:10.1371/journal.pmed.1002986.
6. Zhu J, Chan YM. Adult Consequences of Self-Limited Delayed Puberty. *Pediatrics*. 2017;139(6). doi:10.1542/peds.2016-3177.
7. Marshall WA, Tanner JM. Variations in pattern of pubertal changes in girls. *Archives of Disease in Childhood*. 1969;44(235):291–303. doi:10.1136/adc.44.235.291.
8. Marshall WA, Tanner JM. Variations in the Pattern of Pubertal Changes in Boys. *Archives of Disease in Childhood*. 1970;45(239):13–23. doi:10.1136/adc.45.239.13.
9. Biro FM, Greenspan LC, Galvez MP, Pinney SM, Teitelbaum S, Windham GC, et al. Onset of breast development in a longitudinal cohort. *Pediatrics*. 2013;132(6):1019–27. doi:10.1542/peds.2012-3773.
10. Cousminer DL, Stergiakouli E, Berry DJ, Ang W, Groen-Blokhuis MM, Körner A, et al. Genome-wide association study of sexual maturation in males and females highlights a role for body mass and menarche loci in male puberty. *Human Molecular Genetics*. 2014;23(16):4452–4464. doi:10.1093/hmg/ddu150.
11. Silventoinen K, Jelenkovic A, Palviainen T, Dunkel L, Kaprio J. The Association Between Puberty Timing and Body Mass Index in a Longitudinal Setting: The Contribution of Genetic Factors. *Behavior Genetics*. 2022;52(3):186–194. doi:10.1007/s10519-022-10100-3.
12. Upners EN, Busch AS, Almstrup K, Petersen JH, Assens M, Main KM, et al. Does height and IGF-I determine pubertal timing in girls? *Pediatric Research*. 2020;90(1):176–183. doi:10.1038/s41390-020-01215-6.
13. Aksglaede L, Juul A, Olsen LW, Sørensen TIA. Age at Puberty and the Emerging Obesity Epidemic. *PLoS ONE*. 2009;4(12):e8450. doi:10.1371/journal.pone.0008450.



14. Rosenfield RL, Lipton RB, Drum ML. Thelarche, Pubarche, and Menarche Attainment in Children With Normal and Elevated Body Mass Index. *Pediatrics*. 2009;123(1):84–88. doi:10.1542/peds.2008-0146.
15. Fan HY, Lee YL, Hsieh RH, Yang C, Chen YC. Body mass index growth trajectories, early pubertal maturation, and short stature. *Pediatric Research*. 2019;88(1):117–124. doi:10.1038/s41390-019-0690-3.
16. Ella SSAE, Barseem NF, Tawfik MA, Ahmed AF. BMI relationship to the onset of puberty: assessment of growth parameters and sexual maturity changes in Egyptian children and adolescents of both sexes. *Journal of Pediatric Endocrinology and Metabolism*. 2020;33(1):121–128. doi:10.1515/jpem-2019-0119.
17. Parent AS, Teilmann G, Juul A, Skakkebaek NE, Toppari J, Bourguignon JP. 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(5):668–93. doi:10.1210/er.2002-0019.
18. Pereira A, Corvalan C, Merino PM, Leiva V, Mericq V. Age at Pubertal Development in a Hispanic-Latina Female Population: Should the Definitions Be Revisited? *Journal of Pediatric and Adolescent Gynecology*. 2019;32(6):579–583. doi:10.1016/j.jpag.2019.08.008.
19. Amigo H, Lara M, Bustos P, Muñoz S. Postmenarche growth: cohort study among indigenous and non-indigenous Chilean adolescents. *BMC Public Health*. 2015;15:51. doi:10.1186/s12889-015-1389-y.
20. Fernández M, Pereira A, Corvalán C, Mericq V. Precocious pubertal events in Chilean children: ethnic disparities. *Journal of Endocrinological Investigation*. 2018;42(4):385–395. doi:10.1007/s40618-018-0927-8.
21. Hollis B, Day FR, Busch AS, Thompson DJ, Soares ALG, et al. Genomic analysis of male puberty timing highlights shared genetic basis with hair colour and lifespan. *Nature Communications*. 2020;11(1). doi:10.1038/s41467-020-14451-5.
22. Couto Alves A, De Silva NMG, Karhunen V, Sovio U, Das S, Taal HR, et al. GWAS on longitudinal growth traits reveals different genetic factors influencing infant, child, and adult BMI. *Sci Adv*. 2019;5(9):eaaw3095. doi:10.1126/sciadv.aaw3095.
23. Warrington NM, Howe LD, Paternoster L, Kaakinen M, Herrala S, Huikari V, et al. A genome-wide association study of body mass index across early life and childhood. *Int J Epidemiol*. 2015;44(2):700–12. doi:10.1093/ije/dyv077.
24. Sovio U, Mook-Kanamori DO, Warrington NM, Lawrence R, Briollais L, Palmer CNA, et al. Association between common variation at the FTO locus and changes in body mass index from infancy to late childhood: the complex nature of genetic association through growth and development. *PLoS Genet*. 2011;7(2):e1001307. doi:10.1371/journal.pgen.1001307.
25. Van Der Net JB, Janssens ACJ, Eijkemans MJ, Kastelein JJ, Sijbrands EJ, Steyerberg EW. Cox proportional hazards models have more statistical power than logistic regression models in cross-sectional genetic association studies. *European Journal of Human Genetics*. 2008;16(9):1111.

26. Lin D, Wu C, Li D, Jia W, Hu Z, Zhou Y, et al. Genome-wide association study identifies common variants in SLC39A6 associated with length of survival in esophageal squamous-cell carcinoma. *Nature Genetics*. 2013;45(6):632–638. doi:10.1038/ng.2638.
27. Joshi PK, Fischer K, Schraut KE, Campbell H, Esko T, Wilson JF. Variants near CHRNA3/5 and APOE have age-and sex-related effects on human lifespan. *Nature Communications*. 2016;7. doi:10.1038/ncomms11174.
28. Sparling YH, Younes N, Lachin JM, Bautista OM. Parametric survival models for interval-censored data with time-dependent covariates. *Biostatistics*. 2006;7(4):599–614.
29. Rizopoulos D, Verbeke G, Molenberghs G. Multiple-imputation-based residuals and diagnostic plots for joint models of longitudinal and survival outcomes. *Biometrics*. 2010;66(1):20–29.
30. Wu L, Wei L, Yi GY, Huang Y. Analysis of longitudinal and survival data: joint modeling, inference methods, and issues. *Journal of Probability and Statistics*. 2011;2012:1–17.
31. Pereira A, Garmendia ML, González D, Kain J, Mericq V, Uauy R, et al. Breast bud detection: a validation study in the Chilean growth obesity cohort study. *BMC Womens Health*. 2014;14:96. doi:10.1186/1472-6874-14-96.
32. Welter D, MacArthur J, Morales J, Burdett T, Hall P, Junkins H, et al. The NHGRI GWAS Catalog, a curated resource of SNP-trait associations. *Nucleic Acids Res*. 2014;42(Database issue):D1001–6. doi:10.1093/nar/gkt1229.
33. Yamaguchi K, Ishigaki K, Suzuki A, Tsuchida Y, Tsuchiya H, Sumitomo S, et al. Splicing QTL analysis focusing on coding sequences reveals mechanisms for disease susceptibility loci. *Nature Communications*. 2022;13(1). doi:10.1038/s41467-022-32358-1.
34. Lonsdale J, Thomas J, Salvatore M, Phillips R, Lo E, Shad S, et al. The Genotype-Tissue Expression (GTEx) project. *Nature Genetics*. 2013;45(6):580–585. doi:10.1038/ng.2653.
35. Ruth KS, Day FR, Tyrrell J, Thompson DJ, Wood AR, et al. Using human genetics to understand the disease impacts of testosterone in men and women. *Nature Medicine*. 2020;26(2):252–258. doi:10.1038/s41591-020-0751-5.
36. Sinnott-Armstrong N, Tanigawa Y, Amar D, Mars N, Benner C, Aguirre M, et al. Genetics of 35 blood and urine biomarkers in the UK Biobank. *Nature Genetics*. 2021;53(2):185–194. doi:10.1038/s41588-020-00757-z.
37. Mills MC, Tropf FC, Brazel DM, van Zuydam N, Vaez A, Agbessi M, et al. Identification of 371 genetic variants for age at first sex and birth linked to externalising behaviour. *Nature Human Behaviour*. 2021;5(12):1717–1730. doi:10.1038/s41562-021-01135-3.
38. Kichaev G, Bhatia G, Loh PR, Gazal S, Burch K, Freund MK, et al. Leveraging Polygenic Functional Enrichment to Improve GWAS Power. *The American Journal of Human Genetics*. 2019;104(1):65–75. doi:10.1016/j.ajhg.2018.11.008.
39. Yap CX, Sidorenko J, Wu Y, Kemper KE, Yang J, Wray NR, et al. Dissection of genetic variation and evidence for pleiotropy in male pattern baldness. *Nature Communications*. 2018;9(1). doi:10.1038/s41467-018-07862-y.

40. Endo C, Johnson TA, Morino R, Nakazono K, Kamitsuji S, Akita M, et al. Genome-wide association study in Japanese females identifies fifteen novel skin-related trait associations. *Scientific Reports*. 2018;8(1). doi:10.1038/s41598-018-27145-2.
41. Pickrell JK, Berisa T, Liu JZ, Ségurel L, Tung JY, Hinds DA. Detection and interpretation of shared genetic influences on 42 human traits. *Nature Genetics*. 2016;48(7):709–717. doi:10.1038/ng.3570.
42. Wu S, Zhang M, Yang X, Peng F, Zhang J, Tan J, et al. Genome-wide association studies and CRISPR/Cas9-mediated gene editing identify regulatory variants influencing eyebrow thickness in humans. *PLOS Genetics*. 2018;14(9):e1007640. doi:10.1371/journal.pgen.1007640.
43. Liu J, Zhou Y, Liu S, Song X, Yang XZ, et al. The coexistence of copy number variations (CNVs) and single nucleotide polymorphisms (SNPs) at a locus can result in distorted calculations of the significance in associating SNPs to disease. *Human Genetics*. 2018;137(6-7):553–567. doi:10.1007/s00439-018-1910-3.
44. Kim SK. Identification of 613 new loci associated with heel bone mineral density and a polygenic risk score for bone mineral density, osteoporosis and fracture. *PLOS ONE*. 2018;13(7):e0200785. doi:10.1371/journal.pone.0200785.
45. Vujkovic M, Keaton JM, Lynch JA, Miller DR, Zhou J, Tcheandjieu C, et al. Discovery of 318 new risk loci for type 2 diabetes and related vascular outcomes among 1.4 million participants in a multi-ancestry meta-analysis. *Nature Genetics*. 2020;52(7):680–691. doi:10.1038/s41588-020-0637-y.
46. Pipal KV, Mamtani M, Patel AA, Jaiswal SG, Jaisinghani MT, Kulkarni H. Susceptibility Loci for Type 2 Diabetes in the Ethnically Endogamous Indian Sindhi Population: A Pooled Blood Genome-Wide Association Study. *Genes*. 2022;13(8):1298. doi:10.3390/genes13081298.
47. Pulit SL, Stoneman C, Morris AP, Wood AR, Glastonbury CA, Tyrrell J, et al. Meta-analysis of genome-wide association studies for body fat distribution in 694 649 individuals of European ancestry. *Human Molecular Genetics*. 2018;28(1):166–174. doi:10.1093/hmg/ddy327.
48. Zhu Z, Guo Y, Shi H, Liu CL, Panganiban RA, Chung W, et al. Shared genetic and experimental links between obesity-related traits and asthma subtypes in UK Biobank. *Journal of Allergy and Clinical Immunology*. 2020;145(2):537–549. doi:10.1016/j.jaci.2019.09.035.
49. Wang SH, Su MH, Chen CY, Lin YF, Feng YCA, Hsiao PC, et al. Causality of abdominal obesity on cognition: a trans-ethnic Mendelian randomization study. *International Journal of Obesity*. 2022;46(8):1487–1492. doi:10.1038/s41366-022-01138-8.
50. Kichaev G, Bhatia G, Loh PR, Gazal S, Burch K, Freund MK, et al. Leveraging Polygenic Functional Enrichment to Improve GWAS Power. *Am J Hum Genet*. 2019;104(1):65–75. doi:10.1016/j.ajhg.2018.11.008.
51. Klarin D, Damrauer SM, Cho K, Sun YV, Teslovich TM, et al. Genetics of blood lipids among ~300, 000 multi-ethnic participants of the Million Veteran Program. *Nature Genetics*. 2018;50(11):1514–1523. doi:10.1038/s41588-018-0222-9.

52. Sakaue S, Kanai M, Tanigawa Y, Karjalainen J, Kurki M, Koshihara S, et al. A cross-population atlas of genetic associations for 220 human phenotypes. *Nature Genetics*. 2021;53(10):1415–1424. doi:10.1038/s41588-021-00931-x.
53. Warrington NM, , Beaumont RN, Horikoshi M, Day FR, Helgeland Ø, et al. Maternal and fetal genetic effects on birth weight and their relevance to cardio-metabolic risk factors. *Nature Genetics*. 2019;51(5):804–814. doi:10.1038/s41588-019-0403-1.
54. Cheng S, Wen Y, Liu L, Cheng B, Liang C, Ye J, et al. Traumatic events during childhood and its risks to substance use in adulthood: an observational and genome-wide by environment interaction study in UK Biobank. *Translational Psychiatry*. 2021;11(1). doi:10.1038/s41398-021-01557-7.
55. Yengo L, Vedantam S, Marouli E, Sidorenko J, Bartell E, Sakaue S, et al. A saturated map of common genetic variants associated with human height. *Nature*. 2022;610(7933):704–712. doi:10.1038/s41586-022-05275-y.
56. Richardson TG, Sanderson E, Palmer TM, Ala-Korpela M, Ference BA, Smith GD, et al. Evaluating the relationship between circulating lipoprotein lipids and apolipoproteins with risk of coronary heart disease: A multivariable Mendelian randomisation analysis. *PLOS Medicine*. 2020;17(3):e1003062. doi:10.1371/journal.pmed.1003062.
57. Yu XH, Cao RR, Yang YQ, Deng FY, Bo L, Lei SF. Body surface area is a potential obesity index: Its genetic determination and its causality for later-life diseases. *Obesity*. 2023;31(1):256–266. doi:10.1002/oby.23590.
58. Pirastu N, Cordioli M, Nandakumar P, Mignogna G, Abdellaoui A, Hollis B, et al. Genetic analyses identify widespread sex-differential participation bias. *Nature Genetics*. 2021;53(5):663–671. doi:10.1038/s41588-021-00846-7.
59. Lee JJ, , Wedow R, Okbay A, Kong E, Maghzian O, et al. Gene discovery and polygenic prediction from a genome-wide association study of educational attainment in 1.1 million individuals. *Nature Genetics*. 2018;50(8):1112–1121. doi:10.1038/s41588-018-0147-3.
60. Okbay A, Wu Y, Wang N, Jayashankar H, Bennett M, Nehzati SM, et al. Polygenic prediction of educational attainment within and between families from genome-wide association analyses in 3 million individuals. *Nature Genetics*. 2022;54(4):437–449. doi:10.1038/s41588-022-01016-z.
61. Pisman JA, Demange PA, Guloksuz S, Willemsen AHM, Abdellaoui A, ten Have M, et al. Genetic Risk for Smoking: Disentangling Interplay Between Genes and Socioeconomic Status. *Behavior Genetics*. 2021;52(2):92–107. doi:10.1007/s10519-021-10094-4.
62. Davies G, Lam M, Harris SE, Trampush JW, Luciano M, Hill WD, et al. Study of 300, 486 individuals identifies 148 independent genetic loci influencing general cognitive function. *Nature Communications*. 2018;9(1). doi:10.1038/s41467-018-04362-x.
63. Nagel M, , Jansen PR, Stringer S, Watanabe K, de Leeuw CA, et al. Meta-analysis of genome-wide association studies for neuroticism in 449, 484 individuals identifies novel genetic loci and pathways. *Nature Genetics*. 2018;50(7):920–927. doi:10.1038/s41588-018-0151-7.

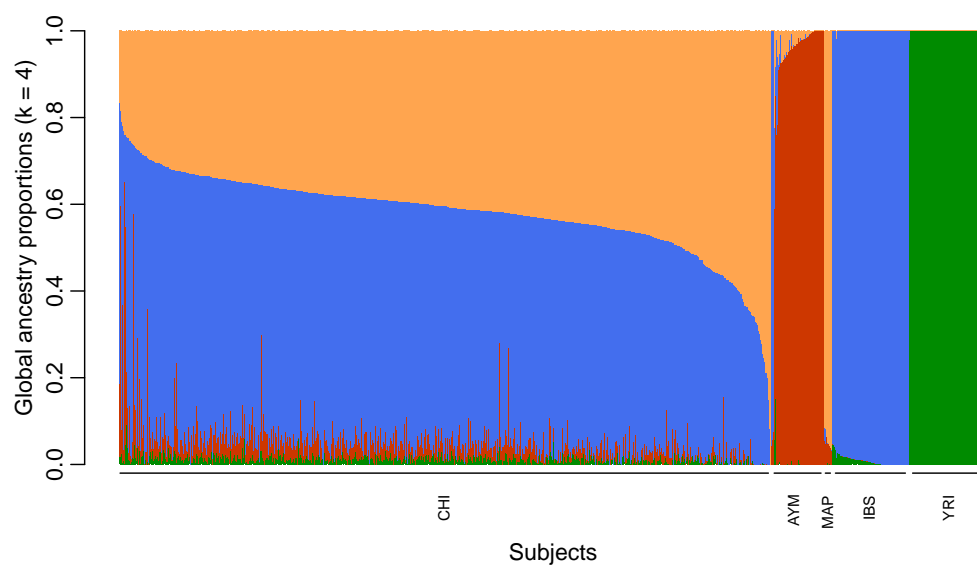
64. Lee PH, Anttila V, Won H, Feng YCA, Rosenthal J, Zhu Z, et al. Genomic Relationships, Novel Loci, and Pleiotropic Mechanisms across Eight Psychiatric Disorders. *Cell*. 2019;179(7):1469–1482.e11. doi:10.1016/j.cell.2019.11.020.
65. Howard DM, Adams MJ, Clarke TK, Hafferty JD, Gibson J, Shirali M, et al. Genome-wide meta-analysis of depression identifies 102 independent variants and highlights the importance of the prefrontal brain regions. *Nature Neuroscience*. 2019;22(3):343–352. doi:10.1038/s41593-018-0326-7.
66. Baselmans B, Hammerschlag AR, Noordijk S, Ip H, van der Zee M, de Geus E, et al. The Genetic and Neural Substrates of Externalizing Behavior. *Biological Psychiatry Global Open Science*. 2022;2(4):389–399. doi:10.1016/j.bpsgos.2021.09.007.
67. Song W, Lin GN, Yu S, Zhao M. Genome-wide identification of the shared genetic basis of cannabis and cigarette smoking and schizophrenia implicates NCAM1 and neuronal abnormality. *Psychiatry Research*. 2022;310:114453. doi:10.1016/j.psychres.2022.114453.
68. Liu M, , Jiang Y, Wedow R, Li Y, Brazel DM, et al. Association studies of up to 1.2 million individuals yield new insights into the genetic etiology of tobacco and alcohol use. *Nature Genetics*. 2019;51(2):237–244. doi:10.1038/s41588-018-0307-5.
69. Yin B, Wang X, Huang T, Jia J. Shared Genetics and Causality Between Decaffeinated Coffee Consumption and Neuropsychiatric Diseases: A Large-Scale Genome-Wide Cross-Trait Analysis and Mendelian Randomization Analysis. *Frontiers in Psychiatry*. 2022;13. doi:10.3389/fpsy.2022.910432.
70. Ye J, Cheng S, Chu X, Wen Y, Cheng B, Liu L, et al. Associations between electronic devices use and common mental traits: A gene–environment interaction model using the UK Biobank data. *Addiction Biology*. 2021;27(2). doi:10.1111/adb.13111.
71. Cole JB, Florez JC, Hirschhorn JN. Comprehensive genomic analysis of dietary habits in UK Biobank identifies hundreds of genetic associations. *Nature Communications*. 2020;11(1). doi:10.1038/s41467-020-15193-0.
72. Jones SE, Lane JM, Wood AR, van Hees VT, Tyrrell J, Beaumont RN, et al. Genome-wide association analyses of chronotype in 697, 828 individuals provides insights into circadian rhythms. *Nature Communications*. 2019;10(1). doi:10.1038/s41467-018-08259-7.
73. Soldà G, Caccia S, Robusto M, Chierighin C, Castorina P, Ambrosetti U, et al. First independent replication of the involvement of LARS2 in Perrault syndrome by whole-exome sequencing of an Italian family. *Journal of Human Genetics*. 2015;61(4):295–300. doi:10.1038/jhg.2015.149.
74. Shahryari A, Jazi MS, Samaei NM, Mowla SJ. Long non-coding RNA SOX2OT: expression signature, splicing patterns, and emerging roles in pluripotency and tumorigenesis. *Frontiers in Genetics*. 2015;6. doi:10.3389/fgene.2015.00196.
75. Kelberman D. Mutations within Sox2/SOX2 are associated with abnormalities in the hypothalamo-pituitary-gonadal axis in mice and humans. *Journal of Clinical Investigation*. 2006;doi:10.1172/jci28658.

76. Ernst M, Benson B, Artiges E, Gorka AX, Lemaitre H, Lago T, et al. Pubertal maturation and sex effects on the default-mode network connectivity implicated in mood dysregulation. *Translational Psychiatry*. 2019;9(1). doi:10.1038/s41398-019-0433-6.
77. Vicuña L, Norambuena T, Miranda JP, Pereira A, Mericq V, Ongaro L, et al. Novel loci and Mapuche genetic ancestry are associated with pubertal growth traits in Chilean boys. *Human Genetics*. 2021;140(12):1651–1661. doi:10.1007/s00439-021-02290-3.
78. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet*. 2007;81(3):559–75. doi:10.1086/519795.
79. Maples BK, Gravel S, Kenny EE, Bustamante CD. RFMix: a discriminative modeling approach for rapid and robust local-ancestry inference. *Am J Hum Genet*. 2013;93(2):278–88. doi:10.1016/j.ajhg.2013.06.020.
80. 1000 Genomes Project Consortium, Auton A, Brooks LD, Durbin RM, Garrison EP, Kang HM, et al. A global reference for human genetic variation. *Nature*. 2015;526(7571):68–74. doi:10.1038/nature15393.
81. Delaneau O, Marchini J, Zagury JF. A linear complexity phasing method for thousands of genomes. *Nat Methods*. 2011;9(2):179–81. doi:10.1038/nmeth.1785.
82. Alexander DH, Novembre J, Lange K. Fast model-based estimation of ancestry in unrelated individuals. *Genome Research*. 2009;19(9):1655–1664. doi:10.1101/gr.094052.109.
83. Vidal EA, Moyano TC, Bustos BI, Pérez-Palma E, Moraga C, Riveras E, et al. Whole genome sequence, variant discovery and annotation in Mapuche-Huilliche Native south Americans. *Sci Rep*. 2019;9(1):2132.
84. Lindo J, Haas R, Hofman C, Apata M, Moraga M, Verdugo RA, et al. The genetic prehistory of the Andean highlands 7000 years BP through European contact. *Sci Adv*. 2018;4(11):eaau4921. doi:10.1126/sciadv.aau4921.
85. Crawford JE, Amaru R, Song J, Julian CG, Racimo F, Cheng JY, et al. Natural selection on genes related to cardiovascular health in high-altitude adapted Andeans. *Am J Hum Genet*. 2017;101(5):752–767.
86. McLaren W, Gil L, Hunt SE, Riat HS, Ritchie GRS, Thormann A, et al. The Ensembl Variant Effect Predictor. *Genome Biol*. 2016;17(1):122. doi:10.1186/s13059-016-0974-4.
87. Vicuña L, Barrientos E, Norambuena T, Alvares D, Gana JC, Leiva-Yamaguchi V, et al. New insights from GWAS on BMI-related growth traits in a longitudinal cohort of admixed children with Native American and European ancestry. *iScience*. 2023;26(2):106091. doi:10.1016/j.isci.2023.106091.
88. Kalbfleisch J, Prentice R. *The statistical analysis of failure time data*. John Wiley & Sons; 2002.
89. Pinheiro J, Bates D. *Linear mixed-effects models: basic concepts and examples*. In: *Mixed-effects models in S and S-Plus*. Springer; 2000. p. 3–56.

90. Tsiatis AA, DeGruttola V, Wulfsohn MS. Modeling the relationship of survival to longitudinal data measured with error. Applications to survival and CD4 counts in patients with AIDS. *Journal of the American Statistical Association*. 1995;90(429):27–37.

## Supplementary Material

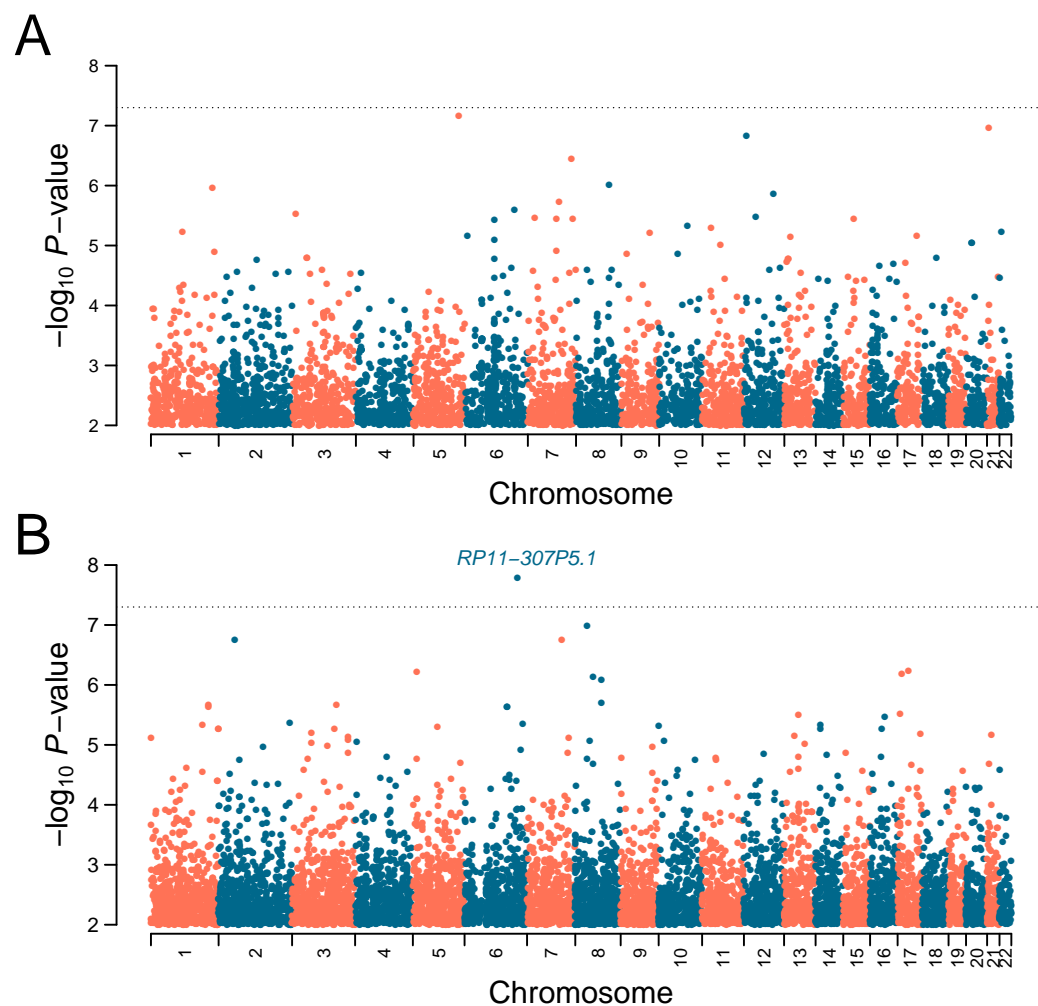
### Supplementary Figures



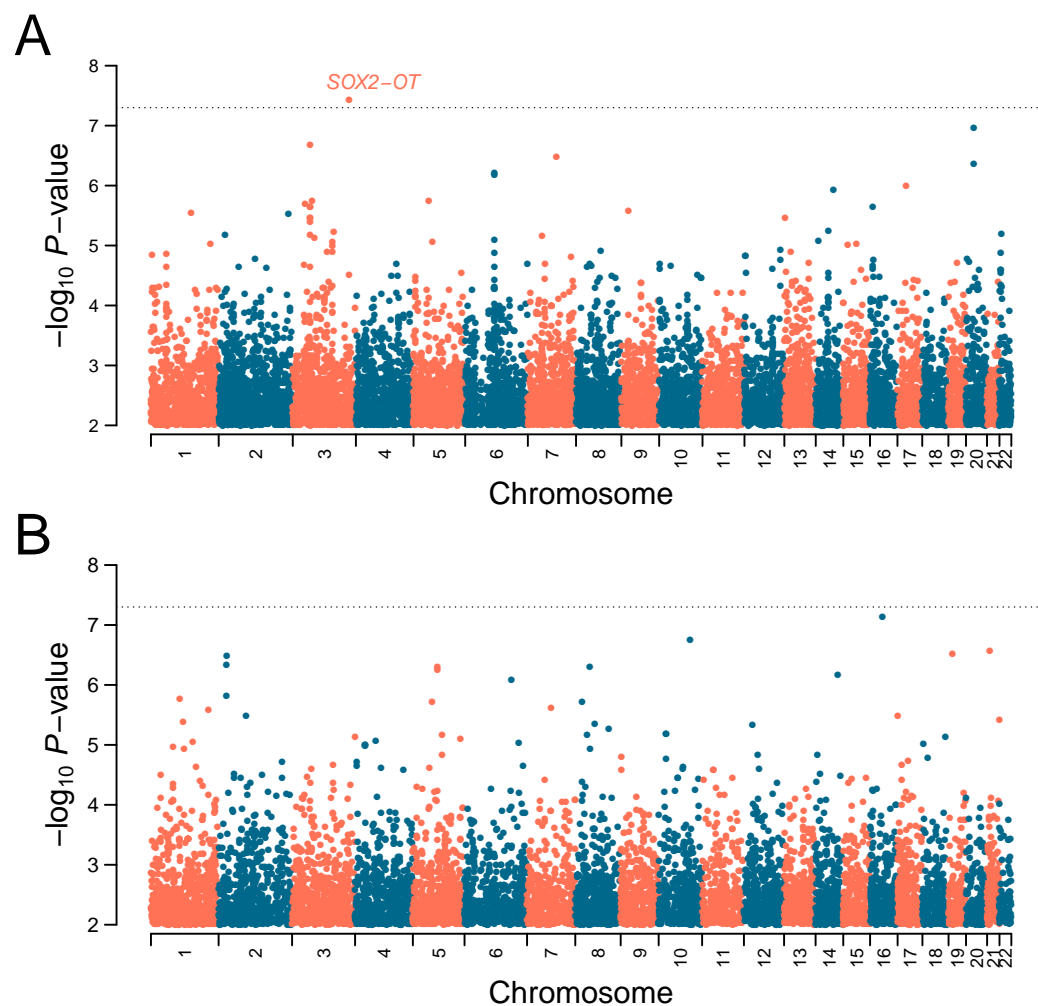
**S1 Figure. Global ancestry proportions of admixed Chilean individuals.**

The number of ancestral populations used was  $K = 4$ . CHI: Admixed Chileans; AYM: Aymara; MAP: Mapuche; IBS: Spaniards; YRI: Yoruba.

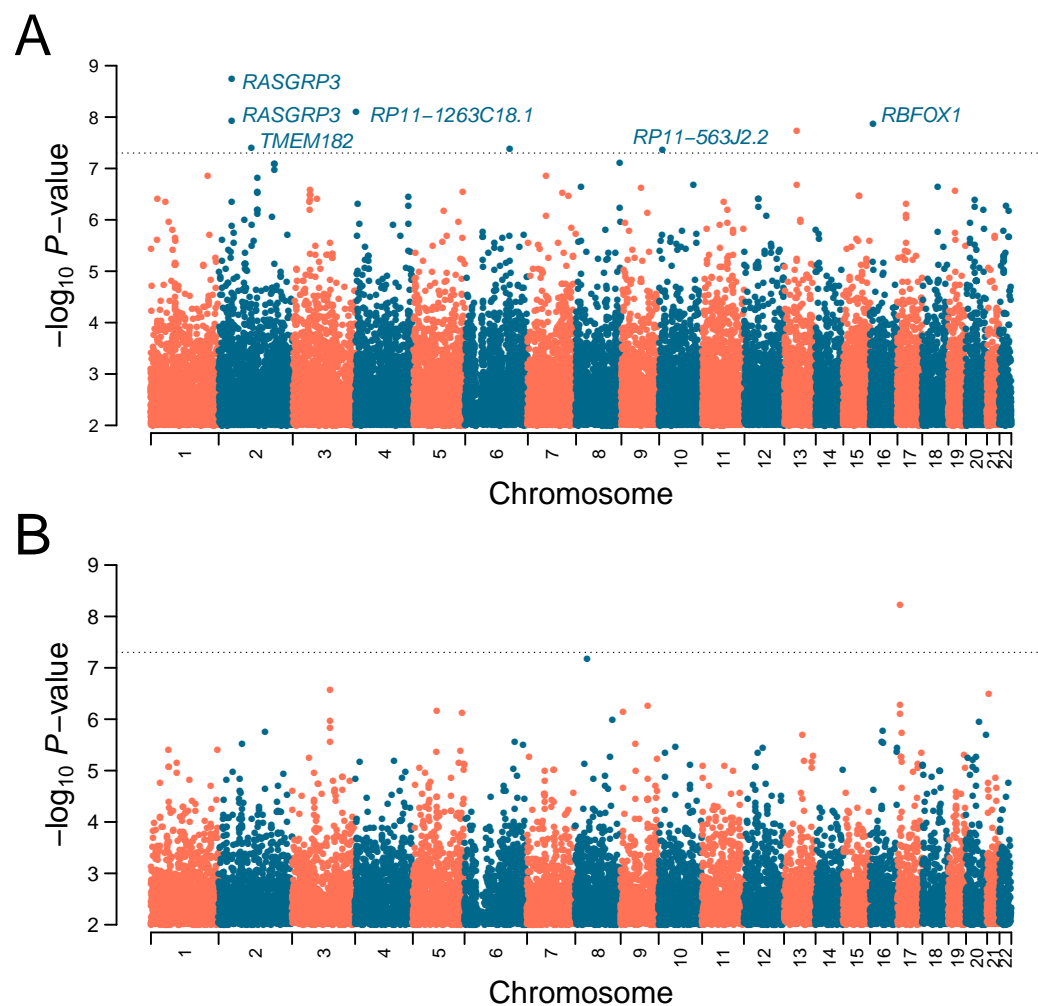




**S2 Figure. GWAS of Tanner 1  $\rightarrow$  2 transition captured by the survival model.** The Manhattan plots represent genotype-phenotype associations along the 22 autosomes in boys (A) and girls (B). Highlighted are genes associated with variants whose associations were significant in the joint and survival models' GWAS. The dotted black line represents the genome-wide significance threshold of  $P = 5 \times 10^{-8}$ . Variants with  $-\log_{10}(P\text{-value})$  between 0 and 2 are not shown.



**S3 Figure. GWAS of Tanner 2 → 3 transition captured by the survival model.** The Manhattan plots represent genotype-phenotype associations along the 22 autosomes in boys (A) and girls (B). Highlighted are genes associated with variants whose associations were significant in the joint and survival models' GWAS. The dotted black line represents the genome-wide significance threshold of  $P = 5 \times 10^{-8}$ . Variants with  $-\log_{10}(P\text{-value})$  between 0 and 2 are not shown.



**S4 Figure. GWAS of Tanner 2 → 4 transition captured by the survival model.** The Manhattan plots represent genotype-phenotype associations along the 22 autosomes in boys (**A**) and girls (**B**). Highlighted are genes associated with variants whose associations were significant in the joint and survival models' GWAS. The dotted black line represents the genome-wide significance threshold of  $P = 5 \times 10^{-8}$ . Variants with  $-\log_{10}(P\text{-value})$  between 0 and 2 are not shown.

## Supplementary Tables

Sex	Tanner Trans.	Variable	Effect Size	SE	P-value
Girls	$T_1 \rightarrow T_2$	Global ancestry	-0.120	0.734	0.870
		Assoc. Factor	0.901	0.341	0.008
	$T_2 \rightarrow T_3$	Global ancestry	-0.199	0.852	0.815
		Assoc. Factor	0.587	0.311	0.059
	$T_3 \rightarrow T_4$	Global ancestry	0.221	0.814	0.786
		Assoc. Factor	1.233	0.401	0.002
	$T_2 \rightarrow T_4$	Global ancestry	0.166	0.851	0.845
		Assoc. Factor	1.709	0.393	<0.001
Boys	$T_1 \rightarrow T_2$	Global ancestry	1.321	0.621	0.033
		Assoc. Factor	0.739	0.287	0.010
	$T_2 \rightarrow T_3$	Global ancestry	0.942	0.61	0.128
		Assoc. Factor	0.429	0.317	0.176
	$T_3 \rightarrow T_4$	Global ancestry	0.732	0.639	0.252
		Assoc. Factor	0.496	0.314	0.114
	$T_2 \rightarrow T_4$	Global ancestry	0.337	0.786	0.668
		Assoc. Factor	0.411	0.407	0.313

**S1 Table. Summary statistics for global ancestry and BMI association of the joint model by Tanner transition.** Shown are the variables and their respective effect size, standard error (SE), and *P*-value at each Tanner stage transition according to sex.

Tanner Trans.	S(t)	Age M (L)	Age M	Age M (U)	Age F (L)	Age F	Age F (U)
$T_1 \rightarrow T_2$	0.25	11.7	12.0	12.2	10.1	10.3	10.6
$T_2 \rightarrow T_3$	0.25	12.9	13.1	13.3	10.9	11.2	11.3
$T_2 \rightarrow T_4$	0.25	13.2	13.4	13.6	11.1	11.3	11.5
$T_3 \rightarrow T_4$	0.25	13.7	13.9	14.1	11.6	11.8	12.0
$T_1 \rightarrow T_2$	0.50	10.7	11.2	11.6	9.2	9.7	10.2
$T_2 \rightarrow T_3$	0.50	12.2	12.6	13.1	10.0	10.5	11.0
$T_2 \rightarrow T_4$	0.50	12.4	12.9	13.4	10.2	10.7	11.1
$T_3 \rightarrow T_4$	0.50	12.9	13.4	13.9	10.7	11.2	11.6
$T_1 \rightarrow T_2$	0.75	9.3	10.0	10.7	8.0	8.7	9.4
$T_2 \rightarrow T_3$	0.75	11.1	11.8	12.5	9.0	9.7	10.4
$T_2 \rightarrow T_4$	0.75	11.7	12.4	13.1	9.3	10.0	10.8
$T_3 \rightarrow T_4$	0.75	12.0	12.8	13.5	9.6	10.3	11.0

**S2 Table. Mean ages attained at Tanner transitions.** Shown are the Tanner pair, the value (probability) of the survival function  $S(t)$ , the median age attained at the Tanner transitions in males (M) and females (F), including the confidence intervals' lower (L) and upper bounds (U).

Interval	S(t)	Males (years)	Females (years)
T12 - T23	0.25	1.13*	0.81*
T12 - T23	0.50	1.45*	0.81
T12 - T23	0.75	1.78*	0.97
T23 - T34	0.25	0.81*	0.65*
T23 - T34	0.50	0.81	0.65
T23 - T34	0.75	0.97	0.65
T23 - T24	0.25	0.32*	0.16
T23 - T24	0.50	0.32	0.16
T23 - T24	0.75	0.65	0.32
T24 - T34	0.25	0.48*	0.48*
T24 - T34	0.50	0.48	0.48
T24 - T34	0.75	0.32	0.32

**S3 Table. Time intervals between survival curves.** Shown are pairs of Tanner stages' pairs representing the curves, the value (probability) of the survival function S(t) and the time interval between the curves in males and females. \* Significant association.

Ancestry	S(t)	Age (L)	Age (median)	Age (U)
Mapuche	0.25	9.37	10.34	11.64
	0.5	8.73	9.54	10.83
	0.75	7.92	8.73	9.70
European	0.25	11.80	11.96	12.12
	0.5	10.83	11.15	11.31
	0.75	9.86	10.02	10.34

**S4 Table. Age attained at Tanner  $T_1 \rightarrow T_2$  transition in Mapuche and European boys.** Shown are the median age and its confidence intervals' lower (L) and upper (U) bounds for S(t) = 0.25, 0.5 and 0.75.

Sex	Tanner Trans.	Intercept	Slope	Intercept + Slope	Mean
Girls	$T_1 \rightarrow T_2$	0.881	0.008	0.004	0.005
	$T_2 \rightarrow T_3$	0.828	0.057	0.020	0.025
	$T_3 \rightarrow T_4$	0.735	0.001	0.001	0.002
	$T_2 \rightarrow T_4$	0.333	<0.001	<0.001	<0.001
Boys	$T_1 \rightarrow T_2$	0.478	0.007	0.016	0.020
	$T_2 \rightarrow T_3$	0.371	0.015	0.063	0.062
	$T_3 \rightarrow T_4$	0.598	0.089	0.066	0.056
	$T_2 \rightarrow T_4$	0.488	0.284	0.558	0.511

**S5 Table. P-value for association parameter of the joint model by Tanner transition.** Shown are the P-values for association parameters of the joint model with different shared component specifications at each Tanner stage transition according to sex.

Sex	Tanner Transition	N Survival	N Joint
Girls	$T_1 \rightarrow T_2$	332	292
	$T_2 \rightarrow T_3$	254	247
	$T_3 \rightarrow T_4$	212	211
	$T_2 \rightarrow T_4$	252	251
Boys	$T_1 \rightarrow T_2$	383	381
	$T_2 \rightarrow T_3$	346	346
	$T_3 \rightarrow T_4$	323	323
	$T_2 \rightarrow T_4$	318	318

**S6 Table. Sample sizes of survival and joint analyses.** Shown are the sample sizes for girls and boys.

### Supplementary Files' Legends

**SFile1. eQTLs associated with the 42 loci captured by our GWAS.** Shown are the SNP rs ID, physical location, association P-value, effect size, gene ID and eGene ID.

**SFile2. sQTLs associated with the 42 loci captured by our GWAS.** Shown are the SNP rs ID, physical location, association P-value, effect size, gene ID and sGene ID.