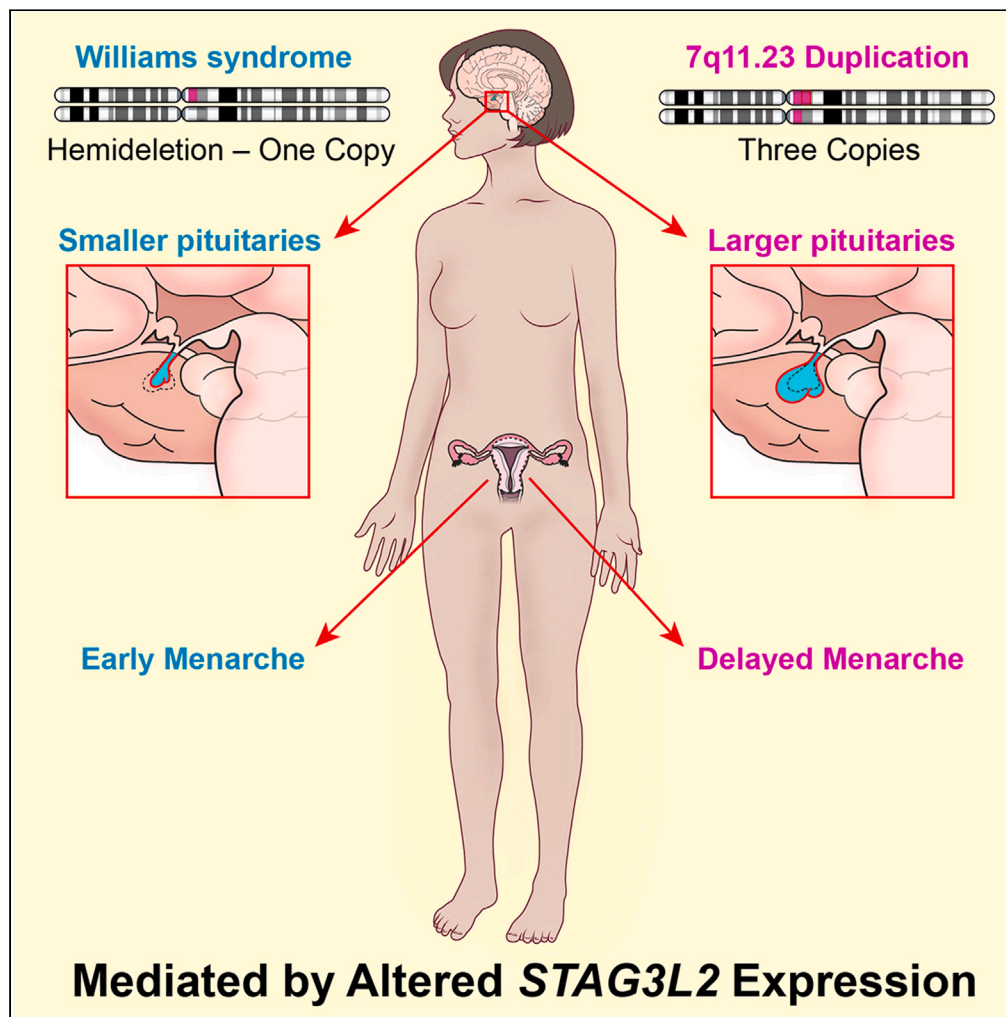


Article

Altered pubertal timing in 7q11.23 copy number variations and associated genetic mechanisms



Shau-Ming Wei,
Michael D.
Gregory, Tiffany
Nash, ..., Bhaskar
Kolachana, Peter J.
Schmidt, Karen F.
Berman

bermank@mail.nih.gov

Highlights

Polygenic scores for age at menarche (AAM) correlate with AAM in developing girls

7q11.23 copy number variation significantly relates to AAM and to pituitary volume

Differential transcription of *STAG3L2* underlies the 7q11.23 CNV-AAM association

Article

Altered pubertal timing in 7q11.23 copy number variations and associated genetic mechanisms

Shau-Ming Wei,^{1,2,5} Michael D. Gregory,^{2,5} Tiffany Nash,² Andrea de Abreu e Gouvêa,² Carolyn B. Mervis,³ Katherine M. Cole,^{1,2} Madeline H. Garvey,² J. Shane Kippenhan,² Daniel P. Eisenberg,² Bhaskar Kolachana,⁴ Peter J. Schmidt,¹ and Karen F. Berman^{2,6,*}

SUMMARY

Pubertal timing, including age at menarche (AAM), is a heritable trait linked to lifetime health outcomes. Here, we investigate genetic mechanisms underlying AAM by combining genome-wide association study (GWAS) data with investigations of two rare genetic conditions clinically associated with altered AAM: Williams syndrome (WS), a 7q11.23 hemideletion characterized by early puberty; and duplication of the same genes (7q11.23 Duplication syndrome [Dup7]) characterized by delayed puberty. First, we confirm that AAM-derived polygenic scores in typically developing children (TD) explain a modest amount of variance in AAM ($R^2 = 0.09$; $p = 0.04$). Next, we demonstrate that 7q11.23 copy number impacts AAM (WS < TD < Dup7; $p = 1.2 \times 10^{-8}$, $\eta^2 = 0.45$) and pituitary volume (WS < TD < Dup7; $p = 3 \times 10^{-5}$, $\eta_p^2 = 0.2$) with greater effect sizes. Finally, we relate an AAM-GWAS signal in 7q11.23 to altered expression in post-mortem brains of *STAG3L2* ($p = 1.7 \times 10^{-17}$), a gene we also find differentially expressed with 7q11.23 copy number ($p = 0.03$). Collectively, these data explicate the role of 7q11.23 in pubertal onset, with *STAG3L2* and pituitary development as potential mediators.

INTRODUCTION

Menarche, when girls begin menstruation, is a pivotal physical developmental milestone that not only holds cultural and social importance in many communities, signifying the onset of reproductive ability, but is also medically meaningful with a range of concurrent and future health implications. Menarche is also a highly important but under-recognized public health consideration. For example, evidence from high-income countries suggests that girls who reach menarche and puberty early without adequate emotional support are likely to engage in earlier sexual relations and substance abuse, posing risks for adolescent pregnancy and other negative health outcomes.^{1–4}

Pubertal maturation of the hypothalamic-pituitary-ovarian axis underlying menarche is complex, and age at menarche (AAM) varies widely across individuals. This variability is associated with subsequent risk for a number of illnesses later in life, including various oncologic, cardio-metabolic, gynecological/obstetric, gastrointestinal, musculoskeletal, and neuropsychiatric diseases.⁴ For instance, early menarche has been associated with increased breast cancer risk,^{5,6} likely due, at least in part, to longer cumulative estrogen exposure throughout reproductive life.^{7,8} Additionally, both early and late AAM have been correlated with several metabolic markers that contribute to increased risk of cardiovascular disease; these include high cholesterol, hypertension, and obesity.^{9–13} Some evidence suggests that, even many decades past menarche, Alzheimer's disease risk may be elevated in those with later AAM,^{14,15} though confirmatory work is needed.^{16,17}

Although the timing of puberty may be influenced by a number of environmental and secondary factors, including nutritional status, systemic health, and central nervous system pathology, it is clear that AAM is also genetically driven, showing 50–80% heritability in prior work.^{18–22} This genetic component raises the possibility that some of the epidemiologic associations between AAM and subsequent disease risk could be mediated not just by the direct sequelae of exposure to a certain cumulative number of lifetime menstrual cycles, but also by underlying pleiotropic genetic variation, though mechanistic evidence from biological experimentation is needed to elaborate this hypothesis. A recent genome-wide association study (GWAS) of ~370,000 women identified 389 significant independent signals that contribute to AAM,²³ indicating a highly polygenic architecture guiding this important developmental trait. This complexity is consistent with studies of puberty-related disorders in both human and animal models, which point to a multifaceted network of signaling molecules and pathways regulating hypothalamic-pituitary-gonadal (HPG) maturation.^{24,25} In addition to the abundance of common, GWAS-defined variants that have been statistically related to normal variability in AAM, investigations into rare highly penetrant genetic mutations, including copy number

¹Behavioral Endocrinology Branch, University of Louisville, Louisville, KY, USA

²Section on Integrative Neuroimaging, Clinical and Translational Neuroscience Branch, University of Louisville, Louisville, KY, USA

³Neurodevelopmental Sciences Laboratory, Department of Psychological and Brain Sciences, University of Louisville, Louisville, KY, USA

⁴Human Brain Collection Core, National Institute of Mental Health, Intramural Research Program, National Institutes of Health, Bethesda, MD, USA

⁵These authors contributed equally

⁶Lead contact

*Correspondence: bermank@mail.nih.gov

<https://doi.org/10.1016/j.isci.2024.109113>



variants (CNVs), can provide unique opportunities to understand specific molecular drivers of clinical phenotypes,^{26,27} including pubertal timing abnormalities.

One example of a chromosomal locus where CNVs affect pubertal timing is the Williams syndrome (WS) critical region in the 7q11.23 chromosomal band. Hemideletion of ~25 genes in this ~1.5 Mb region causes WS, a rare neurodevelopmental disorder,²⁸ compared to typically developing children (TD) who have two copies of these genes and typical age of pubertal onset, it has been clinically noted that children with WS (having one copy of affected genes) may enter puberty early,^{29–32} or even have precocious puberty.³³ More recently and in contrast, individuals with duplications of this same chromosomal locus have been identified (7q11.23 duplication syndrome [Dup7], with three copies of these same genes),³⁴ and clinical observations by ourselves and others have been suggestive of delayed puberty in these individuals.³⁵ If empirically confirmed, these genetically mediated, dose-dependent differences in pubertal onset in 7q11.23 CNVs would provide a unique model for investigating the genetic etiology of variation in the timing of puberty.

In searching for a neural substrate of these genetic findings, we focused on the pituitary gland, a key region in the HPG axis and a central hub for regulating endocrine function in the brain.³⁶ Interestingly, data from the GTEx consortium, a large-scale collaborative effort in which DNA and RNA from multiple tissues were sequenced from almost 1,000 deceased individuals,³⁷ have shown that the pituitary is enriched with genes underlying the AAM-related variants,²³ suggesting that it plays a crucial role in the timing of puberty. In TD children, with average pubertal onset and normal endocrine function, pituitary gland volumes in both sexes increase gradually with age, show a growth spurt during early puberty, and are larger in girls than boys.^{38–40} Studies of pituitary volumes in either precocious or delayed puberty are sparse, with some reports suggesting larger pituitary volumes in children with precocious puberty compared with age-matched controls.^{41,42} Since 7q11.23 CNVs display differential timing of pubertal onset, we hypothesized that pituitary volumes may also vary across these CNVs, though studies examining this potential neural correlate are lacking.

Here, to better characterize the genetic foundations of AAM, we integrated large-scale, general population statistical findings with in-depth investigations of individuals who have rare 7q11.23 CNVs and are enrolled in our long-term, longitudinal studies. First, we sought to confirm and quantify the contributions of common genetic variation to pubertal timing by relating polygenic scores for AAM to clinician-ascertained AAM in a reference cohort of TD girls. Second, to confirm clinical observations, we directly tested whether and to what degree 7q11.23 CNVs are related to AAM. Next, if pubertal onset is indeed altered by 7q11.23 CNVs, and because there is evidence that pituitary morphology in children with early puberty differs from controls, we also tested for an association between 7q11.23 copy number and pituitary volume. To further pinpoint candidate genes that may be responsible for these associations in WS/Dup7, we then examined the 7q11.23 locus for GWAS-defined AAM-associated single nucleotide polymorphisms (SNPs),²³ and we further characterized these signals by identification of relevant quantitative trait-loci (eQTLs) and by testing for differential gene expression in an RNASeq analysis of individuals with 7q11.23 CNVs.

RESULTS

Polygenic contributions to AAM

The degree to which common genetic variation cumulatively influences a phenotype can be estimated as a single value with polygenic scoring methods.⁴³ We used these methods, in conjunction with previously reported summary statistics from a GWAS of retrospectively self-reported AAM that included ~370,000 adult women²³ in order to confirm and quantify the genetic contribution to pubertal timing in a reference sample of TD girls participating in our longitudinal studies. Based on the GWAS summary statistics for AAM previously reported,²³ we calculated AAM polygenic scores (PGS-AAM) for 47 TD girls (whose date of first menstruation was recorded soon after it occurred; Table 1). PGS-AAM was positively related to clinician-ascertained AAM such that individuals with a polygenic proclivity toward later AAM were observed to have later actual AAM ($r = 0.3$, $p = 0.04$, Figure 1). Common genetic variation accounted for approximately 9% of the variance in AAM within this TD reference group, an effect size consistent with prior reports.²³

Rare variant (7q11.23 CNVs) contributions to AAM

Clinical observations have suggested that individuals with WS (hemideletions yielding one copy of ~25 genes at chromosomal locus 7q11.23) have early puberty, whereas those with Dup7 (having three copies of these same genes) have delayed puberty. We sought to confirm these observations empirically and ascertained AAM through clinical interviews of eight girls with WS, 47 TD girls, and nine girls with Dup7 (Table 1). In our rare cohorts of people with 7q11.23 CNVs, we found that AAM significantly differed across copy number groups such that greater gene dosage was associated with older AAM (Average AAM: WS = 10.96 ± 1.1 years, TD = 12.6 ± 1.25 , Dup7 = 15.8 ± 2.7 ; $F = 24.99$, $\eta^2 = 0.45$, $p = 1.2 \times 10^{-8}$, Figure 2), with 7q11.23 copy number accounting for 45% of the variance in AAM. Post-hoc Tukey's honest significant difference tests showed that AAM in the WS group was significantly earlier than was AAM in both TDs ($p = 0.01$) and Dup7 ($p = 5.7 \times 10^{-7}$) and AAM in the Dup7 group was significantly later than in TDs ($p = 2.3 \times 10^{-8}$).

Because common genetic and environmental factors could influence the timing of puberty, we also collected AAM data on typically developing full sibling girls of our participants with WS (8 siblings, AAM = 12.7 ± 1.1) and Dup7 (8 siblings, AAM = 12.2 ± 0.8). These sibling participants (who have the typical two copies of 7q11.23 genes) live and have grown up in the same households as the CNV participants. Therefore, their development could have been affected not only by similar genetic factors (outside of the 7q11.23 region), but also by similar environmental factors that may affect the timing of puberty, including nutrition and socioeconomic status. We found no AAM differences between the unrelated TD group and the 16 siblings of our patients (unrelated TD AAM 12.6 years, sibling AAM 12.45 years, $F = 0.6$, $p = 0.5$). Additionally, we repeated our analysis of AAM as a function of copy number status, substituting these siblings for unrelated TDs to carry out comparison with the two 7q11.23 CNV groups and confirmed the stepwise copy number effect on AAM controlling for these shared genetic and environmental factors ($F = 17.8$, $\eta^2 = 0.5$, $p = 8 \times 10^{-5}$).

Table 1. Demographics

Study	7q11.23 CNVs	Total N	Girls N	Age (mean \pm SD) ^a
AAM Polygenic scores in TD girls	WS	N/A	N/A	N/A
	TD	47	47	12.6 \pm 1.3
	Dup7	N/A	N/A	N/A
Association of AAM and 7q11.23 CNVs	WS	8	8	10.96 \pm 1.1
	Dup7	9	9	15.8 \pm 2.7
	TD	47	47	12.6 \pm 1.3
	Siblings	16	16	12.45 \pm 1.4
Differential Expression Analysis of STAG3L2 in 7q11.23 CNVs	WS	23	15	9.94 \pm 4.0
	TD	40	19	12.86 \pm 3.6
	Dup7	13	8	12.17 \pm 3.1
Pituitary Volume Size Differences with 7q11.23 CNVs	WS	30	20	11.87 \pm 3.9
	TD	50	24	11.88 \pm 3.4
	Dup7	16	7	12.51 \pm 2.8

^afor "AAM polygenic scores in TD girls" and "Association of AAM and 7q11.23 CNVs", age refers to AAM; for "Differential Expression Analysis of STAG3L2 in 7q11.23 CNVs", age refers to age at blood draw; for "Pituitary Volume Size Differences with 7q11.23 CNVs", age refers to age at scan.

Because one individual with Dup7 included in this analysis was treated with growth hormone during childhood, we repeated the analysis testing for association of AAM with 7q11.23 copy number excluding this individual, and the results were largely unchanged and remained highly significant ($F = 20.55$, $\eta^2 = 0.41$, $p = 1.6 \times 10^{-7}$).

Pituitary volume in individuals with 7q11.23 CNVs

The pituitary gland is a key brain region in regulating endocrine function and is central to the HPG axis, producing a surge of hormones at the onset of puberty. Additionally, this region is also enriched with genes underlying AAM-related genetic loci,²³ suggesting that the epidemiologic association between AAM and diseases later in life may be attributed, at least in part, to variability in exposures to sex hormones related to menarche. Prior evidence points to pituitary size increasing in puberty with potentially larger glands in individuals with early puberty.^{41,42} To test whether individuals with rare 7q11.23 CNVs, who (as aforementioned) showed differential timing of AAM, also have differential sizes of pituitary glands, raters, who were blinded to participants' copy number status, manually segmented the pituitary from structural brain MRIs of 30 individuals with WS, 50 TD children and 16 individuals with Dup7, matched for age (Table 1). Total volume of the segmented region was calculated and compared across groups. Pituitary size was significantly associated with 7q11.23 copy number, such that individuals with WS had the smallest pituitary glands, while individuals with Dup7 had the largest ($F = 11.7$, $\eta_p^2 = 0.2$, $p = 3 \times 10^{-5}$, Figure 3). Post-hoc pairwise comparisons showed that the pituitary size of participants with WS was significantly smaller than both TDs ($p = 1.1 \times 10^{-4}$) and those with Dup7 ($p = 4.5 \times 10^{-5}$); TDs were nominally smaller than Dup7 ($p = 0.17$). The directionality of this across-group relationship was opposite to that expected, since as aforementioned, individuals with precocious puberty have larger pituitary volumes,^{41,42} but here, individuals with Dup7,

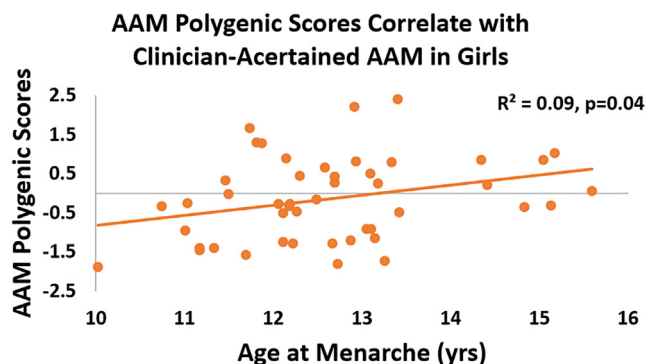


Figure 1. Correlation between PGS-AAM and AAM in typically developing girls

PGS-AAM was positively related to clinician-ascertained AAM such that individuals with a polygenic proclivity toward later actual AAM were observed to have a later AAM ($R^2 = 0.09$, $p = 0.04$). Abbreviations: PGS: polygenic scores; AAM: age at menarche.

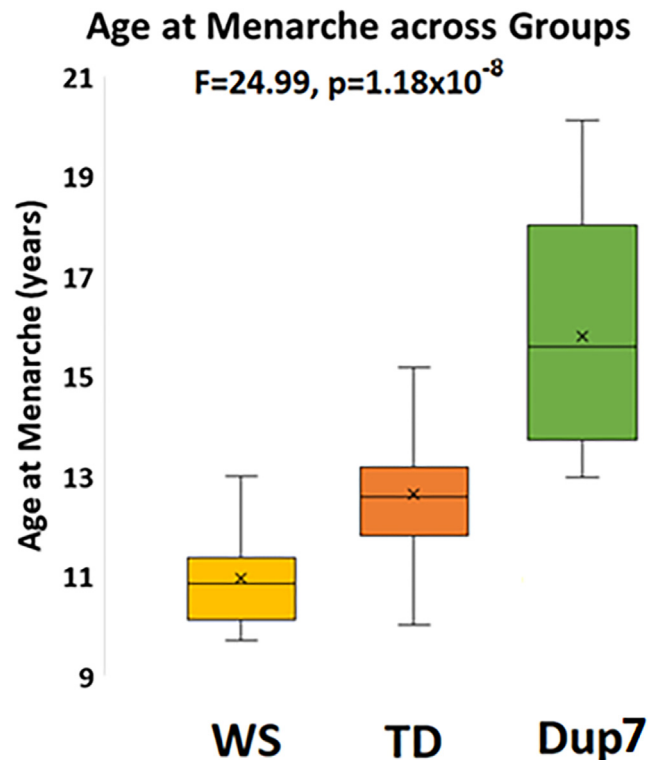


Figure 2. 7q11.23 CNVs and differential pubertal timing

AAM significantly differed across 7q11.23 CNV groups such that greater gene dosage was associated with older AAM (Average AAM: WS=10.96±1.1 years, TD=12.6±1.25, Dup7=15.8±2.7; F=24.99, $\eta^2=0.45$, $p=1.2\times 10^{-8}$). Post-hoc tukey's HSD tests showed AAM in the WS group significantly differed from TDs ($p=0.01$) as did AAM of the Dup7 group compared to TDs ($p=5.7\times 10^{-7}$). X's represent mean of each group, lines represent median, boxes represent the range of the central 50% of data, and whiskers represent maximum and minimum values. Abbreviations: WS: Williams syndrome; Dup7: 7q11.23 Duplication syndrome; TD: typically-developing children.

who have later AAM, had larger pituitary volumes. In reviewing the images after raters blinded to copy number status had manually performed pituitary segmentation, it was noted that glands of several individuals with Dup7 had a more cystic appearance, potentially suggesting a buildup of unreleased hormone.^{44,45}

Next, to examine developmental patterns of pituitary volume longitudinally within 7q11.23 copy number groups, investigators, again blind to diagnosis, segmented pituitaries from 226 longitudinal visits of these same participants, many of whom return approximately every two years. This analysis included scans from 87 visits of 30 participants with WS, 102 visits of 50 TDs, and 37 visits of 16 participants with Dup7. There were significant age ($p = 2\times 10^{-16}$) and group ($p = 2\times 10^{-4}$) effects and an age-by-group interaction ($F = 10.02$, $p = 2.5\times 10^{-6}$), wherein pituitary volumes of the three groups were similar after the start of puberty, but then begin to diverge, and the reduced volumes in WS and the increased volumes in Dup7 become apparent (Figure 4).

Additionally, to ensure that these analyses were not driven by treatment with any hormonal medications, we performed sensitivity analyses for both the cross-sectional and longitudinal assessment of pituitary volumes after excluding participants who had any history of receiving treatment with hormonal medications. The results of these sensitivity analyses were largely unchanged from those of the primary analyses: for the cross-sectional analysis (22 participants with WS, 50 TDs, and 15 participants with Dup7; $F = 8.22$, $\eta_p^2 = 0.17$, $p = 5.5\times 10^{-4}$; and for the longitudinal age-by-group interaction $F = 6.02$, $p = 4.5\times 10^{-4}$, based on 186 total visits.

7q11.23 genetic signals impacting AAM

To determine which, if any, of the 389 genetic signals previously reported to be significantly associated with AAM in a large-scale GWAS²³ are located at the 7q11.23 locus, we referenced the reported summary statistics from that study. One genome-wide significant signal spanning the telomeric end of the WS critical region was identified, extending into the low-copy repeat region flanking the WS 7q11.23 locus (peak SNP rs2267812, $p = 1.7\times 10^{-17}$, Figure 5). To further characterize this significant locus, we searched for eQTLs in data derived from postmortem brains of healthy individuals from the BrainSeq database (<http://eqtl.brainseq.org/phase1/eqtl/>). Though this peak SNP, rs2267812, is located within the *GTF2I* gene, we found that its variation was most significantly associated with expression of two transcripts of the *STAG3L2* gene (ENST00000448772 and ENST00000380775, $p_{\text{bonf}} = 0.002$ and $p_{\text{bonf}} = 0.01$, respectively), suggesting that the latter gene may be responsible for the association of this locus with AAM. Importantly, the ENST00000380775 transcript of *STAG3L2* spans 194kb, nearly the entire locus seen

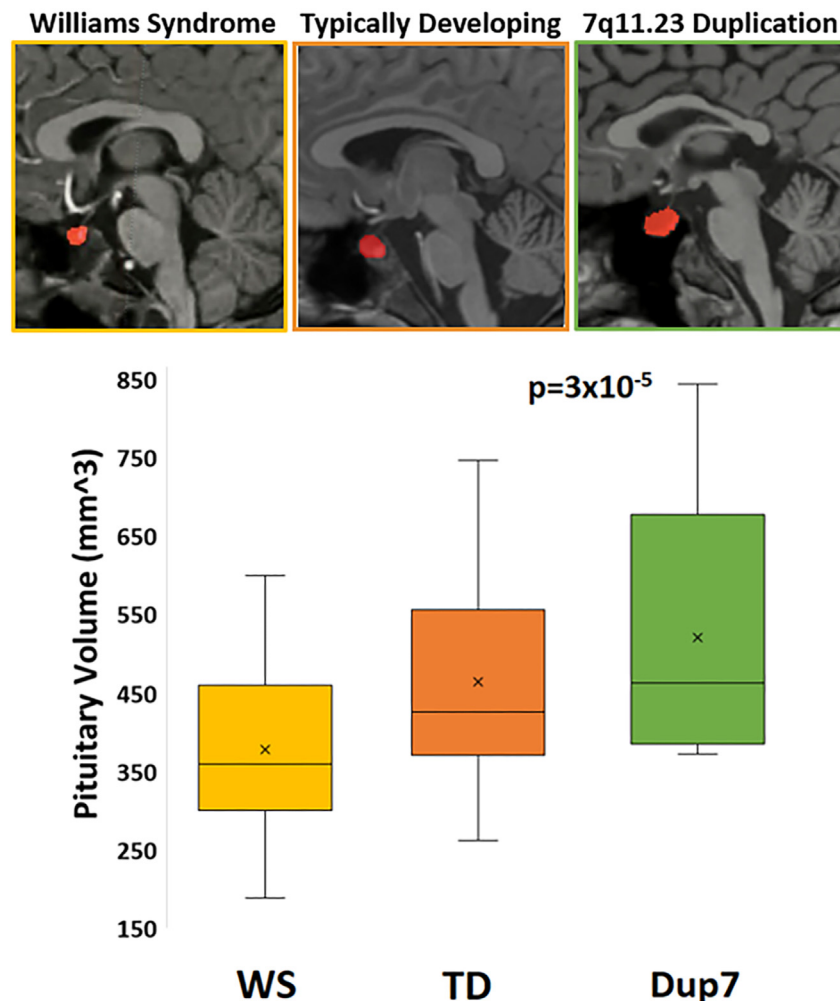


Figure 3. 7q11.23 CNVs and differential pituitary volumes

Pituitary size was significantly associated with 7q11.23 copy number, such that individuals with WS had the smallest pituitary glands while individuals with Dup7 had the largest ($p=3 \times 10^{-5}$). Top: Sample pituitary segmentations of a representative individual with WS (left), a TD participant (middle), and an individual with Dup7 (right). Bottom: Box plots of pituitary volumes as a function of CNV group (WS N=30, age=11.87 ± 3.9 years; TD N=50, age=11.88 ± 3.4 years; Dup7 N=16, age=12.51 ± 2.8 years). X's represent mean of each group, lines represent median, boxes represent the range of the central 50% of data, and whiskers represent maximum and minimum values. Abbreviations: WS: Williams syndrome; Dup7: 7q11.23 Duplication syndrome; TD: typically-developing children.

in Figure 5 and includes an exon underlying the *GTF2I* gene in the telomeric portion of the 7q11.23 region hemideleted in WS (Figure 6). In contrast, the ENST00000448772 transcript lies exclusively in the flanking region that is not typically involved in the classic WS hemideletion (Figure 6), further emphasizing the importance of ENST00000380775 for this work.

Though one exon of the ENST00000380775 transcript of *STAG3L2* extends into the classic WS critical region, the typical reporting of this gene lies outside of the WS locus, in the flanking low-copy repeat region (Figure 6, bottom). Therefore, to test whether individuals with WS and Dup7 have altered expression of *STAG3L2*, we analyzed RNAseq data from blood lymphocytes of 23 children with WS, 40 TD children, and 13 children with Dup7 (Table 1), and found that expression of *STAG3L2* was related to CNV dosage ($p = 0.03$; Figure 7). Post-hoc t-tests showed that individuals with WS had greater *STAG3L2* expression than both TDs ($p = 0.03$) and individuals with Dup7 ($p = 0.04$) but the greater expression in TDs compared to Dup7 was not significant ($p = 0.2$). Surprisingly, the directionality of this finding was such that expression of *STAG3L2* was highest in individuals with WS (who have hemideletions) and lowest in individuals with Dup7 (who have an extra copy of the ~25 genes in the WS 7q11.23 locus), suggesting that hemideletion of this region may include the removal of regulatory element(s) that inhibit *STAG3L2* transcription. We also examined RNAseq data regarding the expression of the other *STAG3*-like genes (which are not located in the 7q11.23 CNV region) and found no associations between 7q11.23 copy number status and expression of these genes. These include *STAG3* ($p = 0.98$), *STAG3L1* ($p = 0.2$), *STAG3L3* ($p = 0.96$), *STAG3L4* ($p = 0.42$), *STAG3L5P* ($p = 0.99$), and *STAG3L5P-PVRIG2P-PILRB* ($p = 0.98$).

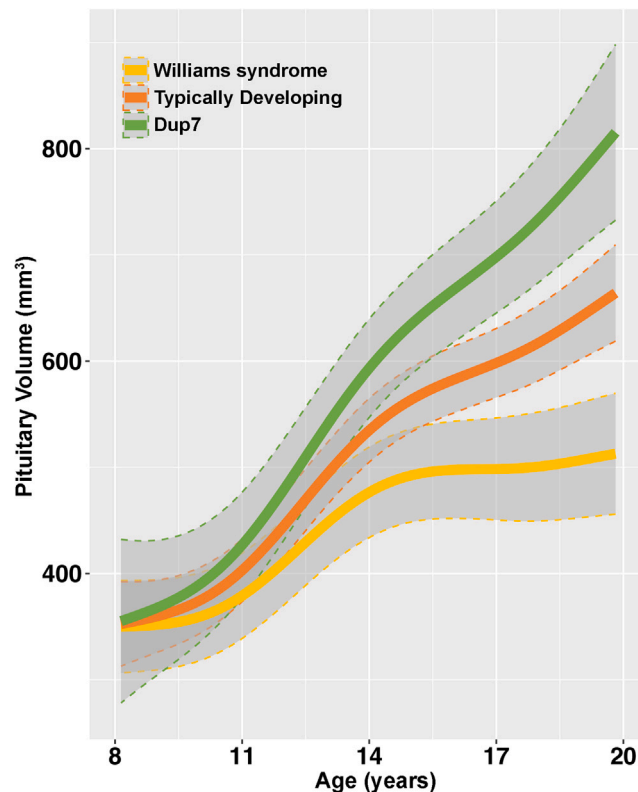


Figure 4. Developmental trajectories of 226 longitudinally acquired pituitary volumes (in cubic millimeters) graphed by 7q11.23 copy number status Pituitary volumes from T1-weighted MRIs of 30 participants with Williams syndrome (WS), 87 visits), 50 typically developing children (TD), 102 visits), and 16 participants with 7q11.23 Duplication (Dup7), 37 visits) over 1–5 visits for each participant across ages 8–20 years. The three developmental trajectory curves (WS, yellow; TD, orange; and Dup7, green) were longitudinally calculated with mixed-effects spline modeling; gray ribbons depict 95% confidence intervals for each group's curve. There were significant age ($p = 2 \times 10^{-16}$) and group ($p = 2 \times 10^{-4}$) effects and an age-by-group interaction ($p = 2 \times 10^{-6}$): pituitary volumes of the three groups were similar until after the start of puberty, but then began to diverge, with largest pituitary volumes in individuals with Dup7 and smallest volumes in WS, a pattern that continued throughout the studied age range; volume increased at a faster rate in the Dup7 group.

DISCUSSION

Consistent with prior clinical observations, we empirically confirmed that individuals with 7q11.23 CNVs have differential pubertal timing, such that greater gene dosage (i.e., in participants with Dup7) was associated with older AAM, whereas hemideletion (i.e., in WS) was associated with earlier pubertal timing. In support of our focus on these rare neurodevelopmental conditions in order to better understand genetic sources of AAM variability, we found that CNV status explained approximately 45% of our sample's total variance in AAM, reflecting both the wide range of AAM in our CNV cohorts and the strength of the highly penetrant CNV effect. Additionally, our observation is likely an underestimation of the actual CNV effect, as the girls with WS who were treated with hormonal medication to delay their early onset of puberty were not able to be included in this analysis as their AAM was pharmacologically altered. In contrast to this robust copy number effect, the cumulative effects of GWAS-identified common genetic variation, while present, accounted for ~9% of the variance in AAM in our reference sample of TD children, which importantly, is similar to the previously reported ~7.4% of the population variance estimated from prior large-scale GWAS data.²³

Though there is a significant genetic component driving AAM, environmental factors, including nutrition and socioeconomic status, could also affect development. Our analysis of AAM comparing patients with 7q11.23 CNVs to full siblings (who have the typical two copies of 7q11.23 genes and live and have grown up in the same households as the patients) was highly significant. In fact, the effect size estimate actually increased (from $\eta^2 = 0.45$ to 0.5) despite a substantially smaller cohort of siblings than there were unrelated TD individuals with 2 copies of 7q11.23 ($N_s = 47$ TDs versus 16 sibling girls), suggesting that taking the shared genetic (outside of the 7q11.23 region) and environmental factors into consideration brought the observed copy number effect into even clearer focus.

In addition to AAM, pituitary size was also associated with 7q11.23 copy number, such that individuals with WS had the smallest pituitary glands while individuals with Dup7 had the largest. This result indicates that pituitary development is likely subject to genetic regulation originating, at least in part, from within the 7q11.23 locus. The pituitary gland, a central endocrine hub governing sex hormone homeostasis and playing a crucial role in the pubertal transition, shows volumetric changes associated with sexual development

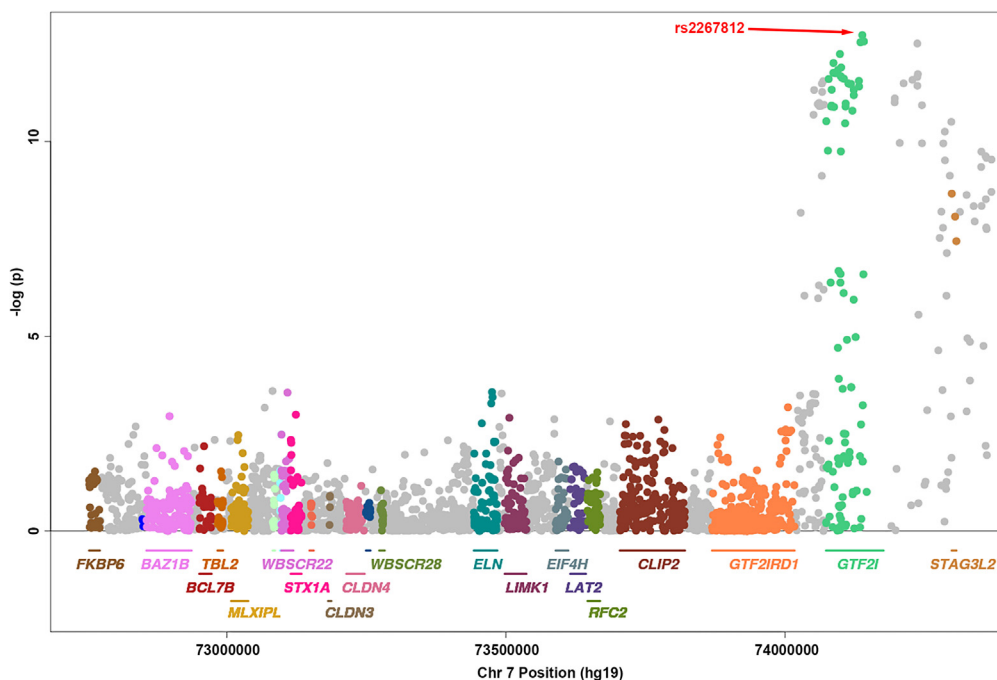


Figure 5. SNP associations with AAM in the expanded 7q11.23 region

A “mini-Manhattan” plot of SNP associations with AAM across the expanded WS critical region. One genome-wide significant signal spanning the telomeric end of the WS critical region was found, extending into the low-copy repeat region that telomerically flanks the WS 7q11.23 locus (peak SNP rs2267812, $p = 1.7 \times 10^{-17}$). Abbreviations: SNP: single nucleotide polymorphism; AAM: age at menarche; WS: Williams syndrome.

and sex hormone milieu^{46,47} and is enriched for AAM-associated gene transcripts.²³ Moreover, SNPs near genes involved in hormone synthesis, bioactivity (e.g., *ESR1*, *PGR*), and pituitary function (e.g., *TACR3*, *LGPR4*) have been found to be associated with menarche timing.⁴⁸ However, the directionality of the association between pituitary size and 7q11.23 CNVs—individuals with WS having smaller pituitary volumes (and earlier puberty) and individuals with Dup7 having larger pituitaries (and later puberty)—suggests that the molecular mechanisms driving earlier puberty in WS may be different from those for clinical disorders associated with formal diagnoses of central precocious puberty, wherein pituitary volumes have been found to be enlarged.^{41,42} Additionally, our qualitative observation of more cystic-appearing pituitary glands in individuals with Dup7 will require further investigation, including determining whether this observation might indicate a buildup of unreleased hormone. Regardless, because of the pituitary’s important role in pubertal development and concomitant morphological changes mirroring pubertal timing abnormalities, it is unlikely that the parallel 7q11.23 CNV associations with both AAM and pituitary size are independent phenomena. Future work will be needed to delineate potential biological pathways leading from specific WS critical region gene hemideletion/duplication to pituitary structural and functional change resulting in AAM alterations.

Given the results we obtained from combining large-scale GWAS-level information with incisive studies of individuals with rare 7q11.23 CNVs, one candidate gene deserving of such investigation is *STAG3L2*. We identified one genome-wide significant signal from the Day et al., findings²³ that spans the telomeric end of the 7q11.23 CNV region, extending into the low-copy repeat region flanking the WS locus (peak SNP rs2267812). RNASeq data from postmortem brains suggest that, though the peak SNP is physically located in an intron of the *GTF2I* gene, variation at rs2267812 results in differential expression of a different gene, *STAG3L2*. This gene is ubiquitously expressed, particularly in testes, ovaries and pituitary,³⁷ and is similar in structure to at least five other “STAG3-like” genes, all of which lie outside the 7q11.23 WS critical region in other low copy repeat regions on chromosome 7.

Interestingly, the existence of these multiple “STAG3-like” genes is believed to be due to segmental duplications that first occurred in a common ancestor to all hominids after diverging from macaques 12–19 Mya, though the emergence of the *STAG3L2* gene of interest here likely occurred during a much more recent segmental duplication in evolution, estimated to have occurred 2.55–2.89 Mya.⁴⁹ Though little is known about these “STAG3-like” genes, including the *STAG3L2* gene of focus here, the ancestral *STAG3* gene, which lies telomeric to the WS critical region, at chromosomal locus 7q22, is a cohesion complex component known to be involved in meiotic chromosome pairing^{50,51} and is particularly integral to the sistering of chromatid arms during meiosis I.⁵² Further, mutations in this ancestral *STAG3* gene have been implicated in disorders of reproductive functions in both men (spermatogenic impairment)^{53–55} and women (primary ovarian insufficiency).^{51,53,56–60} Indeed, the findings reported here suggest that the structurally similar *STAG3L2* gene is also an important mediator in endocrine function and has the potential to also influence the onset of puberty.

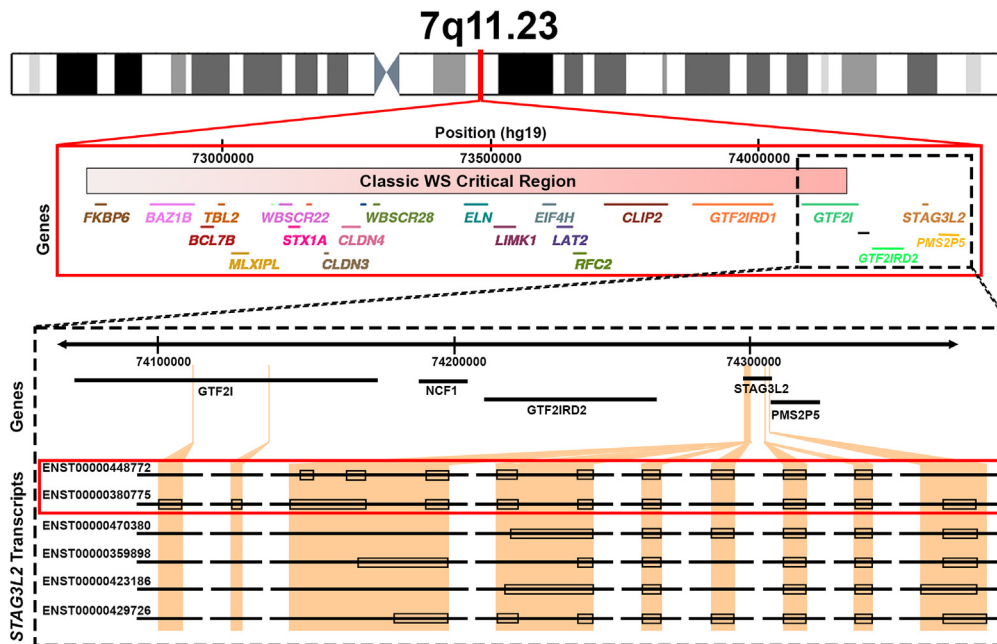


Figure 6. STAG3L2 alternate transcripts

Figure highlights the genes in the WS region and STAG3L2 transcripts significantly related to AAM-associated SNP rs2267812. Top shows WS critical region at the 7q11.23 locus; colored lines and gene names indicate affected individual genes, and the red gradient bar represents the classic WS hemideletion/Dup7 duplication. Bottom shows locations of introns for STAG3L2 transcripts as reported in Ensembl (<http://www.ensembl.org>). The first two transcripts, highlighted by the red box, represent those showing significant eQTLs with rs2267812 in data derived from postmortem brains of healthy individuals from the brainseq database (<http://eqtl.brainseq.org/phase1/eqtl/>; ENST00000448772 and ENST00000380775, $p_{\text{bonf}} = 0.002$ and $p_{\text{bonf}} = 0.01$, respectively). Importantly, the ENST00000380775 transcript spans 194kb, nearly the entire identified significant region, and includes the telomeric portion of the 7q11.23 region hemideleted in WS. In contrast, the ENST00000448772 transcript lies exclusively in the flanking region and is not typically involved in the classical WS deletion.

The observation from our blood lymphocyte RNASeq analyses of individuals with 7q11.23 CNVs that an increase in STAG3L2 RNA expression is linked to WS is surprising, as one might expect hemideletion to result in a decrease in expression. This unexpected result suggests that hemideletion of the WS critical region may lead to the removal of regulatory elements that normally inhibit transcription. These findings are particularly intriguing in light of the reported links between STAG3 loss-of-function mutations and primary ovarian insufficiency in women, which is also characterized by delayed puberty. This association the ancestral STAG3 gene with pubertal timing is consistent with our observations in individuals with Dup7, in whom STAG3L2 expression is decreased and puberty is delayed, versus in WS, where STAG3L2 expression is increased and puberty is early. Thus, it seems plausible that the altered STAG3L2 expression observed in WS and Dup7 and altered STAG3L2 expression associated with common variation linked AAM (i.e., rs2267812) could reflect a unitary genetic mechanism impacting the timing of puberty. Understanding the precise mechanisms underlying these associations will require further research, but these findings may ultimately shed light on the complex interplay between genetics and pubertal timing, and the role of STAG3L2 in these processes.

While much of our focus on the GWAS signal studied here has been on the STAG3L2 gene, the physical location of the peak SNP in the GTF2I gene offers a possible alternative candidate for the gene of interest underlying this locus. However, given the associations between STAG3 and ovarian insufficiency, the eQTL linking rs2267812 with STAG3L2 expression in the postmortem brain, and the alternative STAG3L2 transcript that spans the entire significant GWAS locus, STAG3L2 appears to be the more likely candidate responsible for the differential impacts on pubertal timing observed in individuals with WS and Dup7. Nonetheless, GTF2I codes for a general transcription factor (TFII-I) that regulates methylation of genes outside of the WS critical region⁶¹ and its targets include regulators of phosphorylation and WNT signaling.⁶² Additionally, variation in this gene has been associated with the hypersocial phenotype seen in WS⁶³ and recent evidence has linked this gene to oligodendrocyte function and myelination in the human and murine brain.⁶⁴ Therefore, this important gene warrants further study, but may be the less likely candidate for the pubertal effects reported here.

Overall, the findings presented here provide valuable insights into the genetic mechanisms underlying the timing of puberty and the role of the 7q11.23 locus, particularly the STAG3L2 gene, in this pivotal developmental process. Specifically, the study confirmed that individuals harboring AAM-associated common polymorphisms and those with 7q11.23 CNVs have differential pubertal timing, with relatively modest versus large effect sizes, respectively. Furthermore, alongside AAM, the size of the pituitary may be modulated by 7q11.23 CNVs and merits investigation as a potential neural substrate for relevant pubertal events. Additionally, we suggest that STAG3L2 may yoke both rare CNV and common-SNP findings in this arena, potentially representing a key mediator of altered pubertal onsets in 7q11.23 CNV syndromes that may

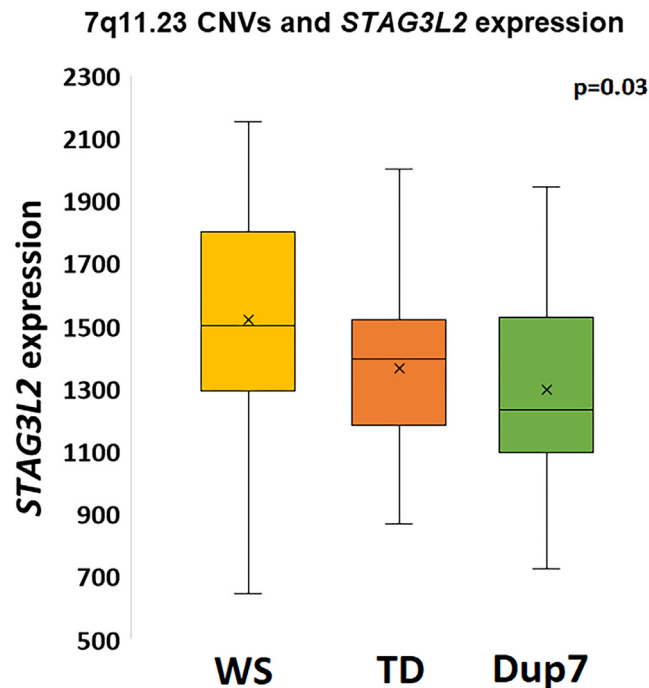


Figure 7. 7q11.23 CNVs and *STAG3L2* expression

RNASeq analysis showed that expression of *STAG3L2* was highest in individuals with WS (who have hemideletions) and lowest in individuals with Dup7 (who have an extra copy at 7q11.23, $p = 0.03$). Y axis shows normalized *STAG3L2* counts from DESEQ2. X's represent mean of each group, lines represent median, boxes represent the range of the central 50% of data, and whiskers represent maximum and minimum values. Abbreviations: WS: Williams syndrome; Dup7: 7q11.23 Duplication syndrome; TD: typically developing children.

also generalize to normative variation in the timing of puberty. Further research will be necessary in order to understand the precise mechanisms underlying these associations, but collectively, the present data provide valuable insights into, and pathways for further discovery of, the complex interplay between genetics and puberty timing. They also present a potential avenue to investigate the adverse health consequences associated with the timing of puberty, which present substantial public health concerns.

Limitations of the study

Several caveats warrant mention regarding the present work. First, because of the relatively small sample sizes of these rare CNVs, we are unable to directly assess the effects of race or early stress, which, in addition to genetics, might influence AAM. Also, there is relatively limited data assessing age at voice breaking in boys, a corresponding pubertal milestone in males that shares many similar genetic signals with AAM in girls.²³ Though there may be multiple reasons for these limited data, including the fact that our cohort of individuals with WS has a sampling bias toward girls, the timing voice breaking in boys has proved to be more difficult to ascertain than AAM. In our experience, individuals and parents tend to remember the exact date of the discrete event of menarche in girls but are less likely to remember the approximate date of the relatively slower process of voice breaking in boys. Lastly, mechanistic interpretation of these findings related to sex steroids is limited by the lack of information regarding circulating hormone levels in these participants. Therefore, the presence of any such association with these data will require further study.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- [KEY RESOURCES TABLE](#)
- [RESOURCE AVAILABILITY](#)
 - Lead contact
 - Materials availability
 - Data and code availability
- [EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS](#)
- [METHOD DETAILS](#)

- Analysis of AAM polygenic scores in TD girls
- Association of AAM and 7q11.23 CNVs
- Pituitary Volume Size Differences with 7q11.23 CNVs
- Differential expression analysis of STAG3L2 in 7q11.23 CNVs
- **QUANTIFICATION AND STATISTICAL ANALYSIS**
- **ADDITIONAL RESOURCES**

ACKNOWLEDGMENTS

We thank the participants who volunteered for our research. This work was supported by two ongoing protocols. Protocol 11M0251 (NCT01434368) is supported by the NIMH Intramural Research Programs (ZIAMH002717 for K.F.B.; ZIAMH002933for P.J.S.) of the NIH/HHS, USA and a NIH Director's Challenge Grant (K.F.B., and P.J.S.). Protocol 10M0112 (NCT01132885) was supported by the NIMH Intramural Research Program (ZIAMH002863), as well as a 2010 NIH Bench-to-Bedside award and a 2014 Brain and Behavior Research Foundation Distinguished Investigator Award to K.F.B. Data were collected under clinical protocol 10-M-0112/NCT01132885. CBM's participation in this project was supported by NICHD (R37 HD29957), NINDS (R01 NS35102), the Simons Foundation (SFARI 238896), and the Williams Syndrome Association (WSA 0104, WSA 0111).

AUTHOR CONTRIBUTIONS

M.D.G., S.M.W., and K.F.B. conceived of the project; C.B.M., T.N., M.H.G., M.D.G., S.M.W., A.D., K.M.C., and P.J.S. managed participants; C.B.M., D.P.E., and M.D.G. evaluated WS and Dup7 participants; M.D.G., M.H.G., and B.K. performed RNA analyses; M.D.G., S.M.W., and B.K. performed genetic analyses; M.D.G., S.M.W., J.S.K., T.N., A.D., M.H.G., and K.F.B. performed pituitary volume analyses. S.M.W. and M.D.G. wrote the first draft of the paper. All authors read and approved the final manuscript.

DECLARATION OF INTERESTS

The authors declare no conflicts of interests.

Received: September 5, 2023

Revised: November 20, 2023

Accepted: January 31, 2024

Published: February 6, 2024

REFERENCES

1. Deardorff, J., Gonzales, N.A., Christopher, F.S., Roosa, M.W., and Millsap, R.E. (2005). Early puberty and adolescent pregnancy: The influence of alcohol use. *Pediatrics* 116, 1451–1456. <https://doi.org/10.1542/peds.2005-0542>.
2. Patton, G.C., and Viner, R. (2007). Adolescent health 1 - Pubertal transitions in health. *Lancet* 369, 1130–1139. [https://doi.org/10.1016/S0140-6736\(07\)60366-3](https://doi.org/10.1016/S0140-6736(07)60366-3).
3. Mendle, J., Turkheimer, E., and Emery, R.E. (2007). Detrimental psychological outcomes associated with early pubertal timing in adolescent girls. *Dev. Rev.* 27, 151–171. <https://doi.org/10.1016/j.dr.2006.11.001>.
4. Day, F.R., Ruth, K.S., Thompson, D.J., Lunetta, K.L., Pervjakova, N., Chasman, D.I., Stolk, L., Finucane, H.K., Sulem, P., Bulik-Sullivan, B., et al. (2015). Large-scale genomic analyses link reproductive aging to hypothalamic signaling, breast cancer susceptibility and BRCA1-mediated DNA repair. *Nat. Genet.* 47, 1294–1303. <https://doi.org/10.1038/ng.3412>.
5. Mullooly, M., Khodr, Z.G., Dallal, C.M., Nyante, S.J., Sherman, M.E., Falk, R., Liao, L.M., Love, J., Brinton, L.A., and Gierach, G.L. (2017). Epidemiologic Risk Factors for In Situ and Invasive Breast Cancers Among Postmenopausal Women in the National Institutes of Health-AARP Diet and Health Study. *Am. J. Epidemiol.* 186, 1329–1340. <https://doi.org/10.1093/aje/kwx206>.
6. Collaborative Group on Hormonal Factors in Breast Cancer (2012). Menarche, menopause, and breast cancer risk: individual participant meta-analysis, including 118 964 women with breast cancer from 117 epidemiological studies. *Lancet Oncol* 13, 1141–1151. [https://doi.org/10.1016/S1470-2045\(12\)70425-4](https://doi.org/10.1016/S1470-2045(12)70425-4).
7. Endogenous Hormones and Breast Cancer Collaborative Group, Key, T.J., Appleby, P.N., Reeves, G.K., Roddam, A.W., Helzlsouer, K.J., Alberg, A.J., Rollison, D.E., Dorgan, J.F., Brinton, L.A., et al. (2011). Circulating sex hormones and breast cancer risk factors in postmenopausal women: reanalysis of 13 studies. *Br. J. Cancer* 105, 709–722. <https://doi.org/10.1038/bjc.2011.254>.
8. Day, F.R., Helgason, H., Chasman, D.I., Rose, L.M., Loh, P.R., Scott, R.A., Helgason, A., Kong, A., Masson, G., Magnusson, O.T., et al. (2016). Physical and neurobehavioral determinants of reproductive onset and success. *Nat. Genet.* 48, 617–623. <https://doi.org/10.1038/ng.3551>.
9. Prentice, P., and Viner, R.M. (2013). Pubertal timing and adult obesity and cardiometabolic risk in women and men: a systematic review and meta-analysis. *Int. J. Obes.* 37, 1036–1043. <https://doi.org/10.1038/ijo.2012.177>.
10. Lakshman, R., Forouhi, N.G., Sharp, S.J., Luben, R., Bingham, S.A., Khaw, K.T., Wareham, N.J., and Ong, K.K. (2009). Early age at menarche associated with cardiovascular disease and mortality. *J. Clin. Endocrinol. Metab.* 94, 4953–4960. <https://doi.org/10.1210/jc.2009-1789>.
11. Day, F.R., Elks, C.E., Murray, A., Ong, K.K., and Perry, J.R.B. (2015). Puberty timing associated with diabetes, cardiovascular disease and also diverse health outcomes in men and women: the UK Biobank study. *Sci. Rep.* 5, 11208. <https://doi.org/10.1038/srep11208>.
12. Luijken, J., van der Schouw, Y.T., Mensink, D., and Onland-Moret, N.C. (2017). Association between age at menarche and cardiovascular disease: A systematic review on risk and potential mechanisms. *Maturitas* 104, 96–116. <https://doi.org/10.1016/j.maturitas.2017.07.009>.
13. Feng, Y., Hong, X., Wilker, E., Li, Z., Zhang, W., Jin, D., Liu, X., Zang, T., Xu, X., and Xu, X. (2008). Effects of age at menarche, reproductive years, and menopause on metabolic risk factors for cardiovascular diseases. *Atherosclerosis* 196, 590–597. <https://doi.org/10.1016/j.atherosclerosis.2007.06.016>.
14. Yoo, J.E., Shin, D.W., Han, K., Kim, D., Won, H.S., Lee, J., Kim, S.Y., Nam, G.E., and Park, H.S. (2020). Female reproductive factors and the risk of dementia: a nationwide cohort study. *Eur. J. Neurol.* 27, 1448–1458. <https://doi.org/10.1111/ene.14315>.

15. Hong, X., Zhang, X., and Li, H. (2001). [A case-control study of endogenous estrogen and risk of Alzheimer's disease]. *Zhonghua Liu Xing Bing Xue Za Zhi* 22, 379–382.
16. Geerlings, M.I., Ruitenberg, A., Witterman, J.C., van Swieten, J.C., Hofman, A., van Duijn, C.M., Breteler, M.M., and Launer, L.J. (2001). Reproductive period and risk of dementia in postmenopausal women. *JAMA* 285, 1475–1481. <https://doi.org/10.1001/jama.285.11.1475>.
17. Prince, M.J., Acosta, D., Guerra, M., Huang, Y., Jimenez-Velazquez, I.Z., Llibre Rodriguez, J.J., Salas, A., Sosa, A.L., Chua, K.C., Dewey, M.E., et al. (2018). Reproductive period, endogenous estrogen exposure and dementia incidence among women in Latin America and China; A 10/66 population-based cohort study. *PLoS One* 13, e0192889. <https://doi.org/10.1371/journal.pone.0192889>.
18. Morris, D.H., Jones, M.E., Schoemaker, M.J., Ashworth, A., and Swerdlow, A.J. (2011). Familial concordance for age at menarche: analyses from the Breakthrough Generations Study. *Paediatr. Perinat. Epidemiol.* 25, 306–311. <https://doi.org/10.1111/j.1365-3016.2010.01183.x>.
19. Kaprio, J., Rimpelä, A., Winter, T., Viken, R.J., Rimpelä, M., and Rose, R.J. (1995). Common genetic influences on BMI and age at menarche. *Hum. Biol.* 67, 739–753.
20. Hur, Y.M., Jin, H.J., and Lee, S. (2019). Heritability of Age at Menarche in South Korean Female Twins. *Twin Res. Hum. Genet.* 22, 183–186. <https://doi.org/10.1017/thg.2019.17>.
21. Dvornyk, V., and Waqar ul, H. (2012). Genetics of age at menarche: a systematic review. *Hum. Reprod. Update* 18, 198–210. <https://doi.org/10.1093/humupd/dmr050>.
22. Anderson, C.A., Duffy, D.L., Martin, N.G., and Visscher, P.M. (2007). Estimation of variance components for age at menarche in twin families. *Behav. Genet.* 37, 668–677. <https://doi.org/10.1007/s10519-007-9163-2>.
23. Day, F.R., Thompson, D.J., Helgason, H., Chasman, D.I., Finucane, H., Sulem, P., Ruth, K.S., Whalen, S., Sarkar, A.K., Albrecht, E., et al. (2017). Genomic analyses identify hundreds of variants associated with age at menarche and support a role for puberty timing in cancer risk. *Nat. Genet.* 49, 834–841. <https://doi.org/10.1038/ng.3841>.
24. Abreu, A.P., Dauber, A., Macedo, D.B., Noel, S.D., Brito, V.N., Gill, J.C., Cukier, P., Thompson, I.R., Navarro, V.M., Gagliardi, P.C., et al. (2013). Central precocious puberty caused by mutations in the imprinted gene MKRN3. *N. Engl. J. Med.* 368, 2467–2475. <https://doi.org/10.1056/NEJMoa1302160>.
25. Silveira, L.G., Noel, S.D., Silveira-Neto, A.P., Abreu, A.P., Brito, V.N., Santos, M.G., Bianco, S.D.C., Kuohung, W., Xu, S., Gryngarten, M., et al. (2010). Mutations of the KISS1 gene in disorders of puberty. *J. Clin. Endocrinol. Metab.* 95, 2276–2280. <https://doi.org/10.1210/jc.2009-2421>.
26. Schork, N.J., Murray, S.S., Frazer, K.A., and Topol, E.J. (2009). Common vs. rare allele hypotheses for complex diseases. *Curr. Opin. Genet. Dev.* 19, 212–219. <https://doi.org/10.1016/j.gde.2009.04.010>.
27. Sullivan, P.F., Daly, M.J., and O'Donovan, M. (2012). Genetic architectures of psychiatric disorders: the emerging picture and its implications. *Nat. Rev. Genet.* 13, 537–551. <https://doi.org/10.1038/nrg3240>.
28. Kozel, B.A., Barak, B., Kim, C.A., Mervis, C.B., Osborne, L.R., Porter, M., and Pober, B.R. (2021). Williams syndrome. *Nat. Rev. Dis. Primers* 7, 42. <https://doi.org/10.1038/s41572-021-00276-z>.
29. Morris, C.A. (1993). Williams Syndrome. In *GeneReviews*(R), M.P. Adam, G.M. Mirzaa, R.A. Pagon, S.E. Wallace, L.J.H. Bean, K.W. Gripp, and A. Amemiya, eds.
30. Stanley, T.L., Leong, A., and Pober, B.R. (2021). Growth, body composition, and endocrine issues in Williams syndrome. *Curr. Opin. Endocrinol. Diabetes Obes.* 28, 64–74. <https://doi.org/10.1097/MED.0000000000000588>.
31. Pober, B.R. (2010). Williams-Beuren syndrome. *N. Engl. J. Med.* 362, 239–252. <https://doi.org/10.1056/NEJMr0903074>.
32. Kim, Y.M., Cho, J.H., Kang, E., Kim, G.H., Seo, E.J., Lee, B.H., Choi, J.H., and Yoo, H.W. (2016). Endocrine dysfunctions in children with Williams-Beuren syndrome. *Ann. Pediatr. Endocrinol. Metab.* 21, 15–20. <https://doi.org/10.6065/apem.2016.21.1.15>.
33. Partsch, C.J., Japing, I., Siebert, R., Gosch, A., Wessel, A., Sippell, W.G., and Pankau, R. (2002). Central precocious puberty in girls with Williams syndrome. *J. Pediatr.* 141, 441–444. <https://doi.org/10.1067/mpd.2002.127280>.
34. Somerville, M.J., Mervis, C.B., Young, E.J., Seo, E.J., del Campo, M., Bamforth, S., Peregrine, E., Loo, W., Lilley, M., Pérez-Jurado, L.A., et al. (2005). Severe expressive-language delay related to duplication of the Williams-Beuren locus. *N. Engl. J. Med.* 353, 1694–1701. <https://doi.org/10.1056/NEJMoa051962>.
35. Morris, C.A., Mervis, C.B., Paciorkowski, A.P., Abdul-Rahman, O., Dugan, S.L., Rope, A.F., Bader, P., Hendon, L.G., Velleman, S.L., Klein-Tasman, B.P., and Osborne, L.R. (2015). 7q11.23 Duplication syndrome: Physical characteristics and natural history. *Am. J. Med. Genet. A* 167A, 2916–2935. <https://doi.org/10.1002/ajmg.a.37340>.
36. Spaziani, M., Tarantino, C., Tahani, N., Gianfrilli, D., Sbardella, E., Lenzi, A., and Radicioni, A.F. (2021). Hypothalamo-Pituitary axis and puberty. *Mol. Cell. Endocrinol.* 520, 111094. <https://doi.org/10.1016/j.mce.2020.111094>.
37. GTEx Consortium (2020). The GTEx Consortium atlas of genetic regulatory effects across human tissues. *Science* 369, 1318–1330. <https://doi.org/10.1126/science.aaz1776>.
38. Çolaklar, A., and Fitoz, Ö.S. (2023). Pituitary gland volumes in children with normal endocrine function. *Pediatr. Radiol.* 53, 452–460. <https://doi.org/10.1007/s00247-022-05505-5>.
39. MacMaster, F.P., Keshavan, M., Mirza, Y., Carrey, N., Upadhyaya, A.R., El-Sheikh, R., Buhagiar, C.J., Taormina, S.P., Boyd, C., Lynch, M., et al. (2007). Development and sexual dimorphism of the pituitary gland. *Life Sci.* 80, 940–944. <https://doi.org/10.1016/j.lfs.2006.11.040>.
40. Takano, K., Utsunomiya, H., Ono, H., Ohfu, M., and Okazaki, M. (1999). Normal development of the pituitary gland: assessment with three-dimensional MR volumetry. *AJNR. Am. J. Neuroradiol.* 20, 312–315.
41. Sharafuddin, M.J., Luisiri, A., Garibaldi, L.R., Fulk, D.L., Klein, J.B., Gillespie, K.N., and Graviss, E.R. (1994). MR imaging diagnosis of central precocious puberty: importance of changes in the shape and size of the pituitary gland. *AJR. Am. J. Roentgenol.* 162, 1167–1173. <https://doi.org/10.2214/ajr.162.5.8166005>.
42. Wu, S., Yang, Y., Wang, Y., Liu, Q., Zhu, Z., and Gu, W. (2020). Diagnostic value of pituitary volume in girls with precocious puberty. *BMC Pediatr.* 20, 425. <https://doi.org/10.1186/s12887-020-02283-7>.
43. Schizophrenia Working Group of the Psychiatric Genomics Consortium (2014). Biological insights from 108 schizophrenia-associated genetic loci. *Nature* 511, 421–427. <https://doi.org/10.1038/nature13595>.
44. Greenman, Y., and Melmed, S. (1996). Diagnosis and management of nonfunctioning pituitary tumors. *Annu. Rev. Med.* 47, 95–106. <https://doi.org/10.1146/annurev.med.47.1.95>.
45. Kontogeorgos, G., Kovacs, K., Horvath, E., and Scheithauer, B.W. (1993). Null cell adenomas, oncocyctomas, and gonadotroph adenomas of the human pituitary: An immunocytochemical and ultrastructural analysis of 300 cases. *Endocr. Pathol.* 4, 20–27. <https://doi.org/10.1007/BF02914485>.
46. Peper, J.S., Brouwer, R.M., van Leeuwen, M., Schnack, H.G., Boomsma, D.I., Kahn, R.S., and Hulshoff Pol, H.E. (2010). HPG-axis hormones during puberty: a study on the association with hypothalamic and pituitary volumes. *Psychoneuroendocrinology* 35, 133–140. <https://doi.org/10.1016/j.psyneuen.2009.05.025>.
47. Wong, A.P.Y., Pipitone, J., Park, M.T.M., Dickie, E.W., Leonard, G., Perron, M., Pike, B.G., Richer, L., Veillette, S., Chakravarty, M.M., et al. (2014). Estimating volumes of the pituitary gland from T1-weighted magnetic-resonance images: effects of age, puberty, testosterone, and estradiol. *Neuroimage* 94, 216–221. <https://doi.org/10.1016/j.neuroimage.2014.02.030>.
48. Perry, J.R., Day, F., Elks, C.E., Sulem, P., Thompson, D.J., Ferreira, T., He, C., Chasman, D.I., Esko, T., Thorleifsson, G., et al. (2014). Parent-of-origin-specific allelic associations among 106 genomic loci for age at menarche. *Nature* 514, 92–97. <https://doi.org/10.1038/nature13545>.
49. Antonelli, A., de Luis, O., Domingo-Roura, X., and Pérez-Jurado, L.A. (2005). Evolutionary mechanisms shaping the genomic structure of the Williams-Beuren syndrome chromosomal region at human 7q11.23. *Genome Res.* 15, 1179–1188. <https://doi.org/10.1101/gr.3944605>.
50. Pezzi, N., Prieto, I., Kremer, L., Pérez Jurado, L.A., Valero, C., Del Mazo, J., Martínez-A, C., and Barbero, J.L. (2000). STAG3, a novel gene encoding a protein involved in meiotic chromosome pairing and location of STAG3-related genes flanking the Williams-Beuren syndrome deletion. *FASEB J* 14, 581–592. <https://doi.org/10.1096/fasebj.14.3.581>.
51. Sang, Q., Ray, P.F., and Wang, L. (2023). Understanding the genetics of human infertility. *Science* 380, 158–163. <https://doi.org/10.1126/science.adf7760>.
52. Prieto, I., Suja, J.A., Pezzi, N., Kremer, L., Martínez-A, C., Rufas, J.S., and Barbero, J.L. (2001). Mammalian STAG3 is a cohesin specific to sister chromatid arms in meiosis I. *Nat. Cell Biol.* 3, 761–766. <https://doi.org/10.1038/35087082>.
53. Jaillard, S., McElreavy, K., Robevska, G., Akloul, L., Ghieh, F., Sreenivasan, R., Beaumont, M., Bashamboo, A., Bignon-Topalovic, J., Neyroud, A.S., et al. (2020).

- STAG3 homozygous missense variant causes primary ovarian insufficiency and male non-obstructive azoospermia. *Mol. Hum. Reprod.* 26, 665–677. <https://doi.org/10.1093/molehr/gaaa050>.
54. Riera-Escamilla, A., Enguita-Marruedo, A., Moreno-Mendoza, D., Chianese, C., Sleddens-Linkels, E., Contini, E., Benelli, M., Natali, A., Colpi, G.M., Ruiz-Castañé, E., et al. (2019). Sequencing of a 'mouse azoospermia' gene panel in azoospermic men: identification of RNF212 and STAG3 mutations as novel genetic causes of meiotic arrest. *Hum. Reprod.* 34, 978–988. <https://doi.org/10.1093/humrep/dez042>.
 55. van der Bijl, N., Röpke, A., Biswas, U., Wöste, M., Jessberger, R., Kliesch, S., Friedrich, C., and Tüttelmann, F. (2019). Mutations in the stromal antigen 3 (STAG3) gene cause male infertility due to meiotic arrest. *Hum. Reprod.* 34, 2112–2119. <https://doi.org/10.1093/humrep/dez204>.
 56. França, M.M., Nishi, M.Y., Funari, M.F.A., Lerario, A.M., Baracat, E.C., Hayashida, S.A.Y., Maciel, G.A.R., Jorge, A.A.L., and Mendonca, B.B. (2019). Two rare loss-of-function variants in the STAG3 gene leading to primary ovarian insufficiency. *Eur. J. Med. Genet.* 62, 186–189. <https://doi.org/10.1016/j.ejmg.2018.07.008>.
 57. Heddar, A., Dessen, P., Flatters, D., and Misrahi, M. (2019). Novel STAG3 mutations in a Caucasian family with primary ovarian insufficiency. *Mol. Genet. Genom.* 294, 1527–1534. <https://doi.org/10.1007/s00438-019-01594-4>.
 58. Le Quesne Stabej, P., Williams, H.J., James, C., Tekman, M., Stanescu, H.C., Kleta, R., Ocaka, L., Lescai, F., Storr, H.L., Bitner-Glindzic, M., et al. (2016). STAG3 truncating variant as the cause of primary ovarian insufficiency. *Eur. J. Hum. Genet.* 24, 135–138. <https://doi.org/10.1038/ejhg.2015.107>.
 59. Mellone, S., Zavattaro, M., Vurchio, D., Ronzani, S., Caputo, M., Leone, I., Prodam, F., and Giordano, M. (2021). A Long Contiguous Stretch of Homozygosity Disclosed a Novel STAG3 Biallelic Pathogenic Variant Causing Primary Ovarian Insufficiency: A Case Report and Review of the Literature. *Genes (Basel)* 12, 1709. <https://doi.org/10.3390/genes12111709>.
 60. Xiao, W.J., He, W.B., Zhang, Y.X., Meng, L.L., Lu, G.X., Lin, G., Tan, Y.Q., and Du, J. (2019). In-Frame Variants in STAG3 Gene Cause Premature Ovarian Insufficiency. *Front. Genet.* 10, 1016. <https://doi.org/10.3389/fgene.2019.01016>.
 61. Strong, E., Butcher, D.T., Singhanian, R., Mervis, C.B., Morris, C.A., De Carvalho, D., Weksberg, R., and Osborne, L.R. (2015). Symmetrical Dose-Dependent DNA-Methylation Profiles in Children with Deletion or Duplication of 7q11.23. *Am. J. Hum. Genet.* 97, 216–227. <https://doi.org/10.1016/j.ajhg.2015.05.019>.
 62. Kopp, N.D., Nygaard, K.R., Liu, Y., McCullough, K.B., Maloney, S.E., Gabel, H.W., and Dougherty, J.D. (2020). Functions of Gtf2i and Gtf2ird1 in the developing brain: transcription, DNA binding and long-term behavioral consequences. *Hum. Mol. Genet.* 29, 1498–1519. <https://doi.org/10.1093/hmg/ddaa070>.
 63. Jabbi, M., Chen, Q., Turner, N., Kohn, P., White, M., Kippenhan, J.S., Dickinson, D., Kolachana, B., Mattay, V., Weinberger, D.R., and Berman, K.F. (2015). Variation in the Williams syndrome GTF2I gene and anxiety proneness interactively affect prefrontal cortical response to aversive stimuli. *Transl. Psychiatry* 5, e622. <https://doi.org/10.1038/tp.2015.98>.
 64. Barak, B., Zhang, Z., Liu, Y., Nir, A., Trangle, S.S., Ennis, M., Levandowski, K.M., Wang, D., Quast, K., Boulting, G.L., et al. (2019). Neuronal deletion of Gtf2i, associated with Williams syndrome, causes behavioral and myelin alterations rescuable by a remyelinating drug. *Nat. Neurosci.* 22, 700–708. <https://doi.org/10.1038/s41593-019-0380-9>.
 65. Gregory, M.D., Kippenhan, J.S., Callicott, J.H., Rubinstein, D.Y., Mattay, V.S., Coppola, R., and Berman, K.F. (2019). Sequence Variation Associated with SLC12A5 Gene Expression Is Linked to Brain Structure and Function in Healthy Adults. *Cereb. Cortex* 29, 4654–4661. <https://doi.org/10.1093/cercor/bhy344>.
 66. Chang, C.C., Chow, C.C., Tellier, L.C., Vattikuti, S., Purcell, S.M., and Lee, J.J. (2015). Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience* 4, 7. <https://doi.org/10.1186/s13742-015-0047-8>.
 67. Bergen, S.E., Ploner, A., Howrigan, D., CNV Analysis Group and the Schizophrenia Working Group of the Psychiatric Genomics, O'Donovan, M.C., Smoller, J.W., Sullivan, P.F., Sebat, J., Neale, B., and Kendler, K.S. (2019). Joint Contributions of Rare Copy Number Variants and Common SNPs to Risk for Schizophrenia. *Am J Psychiatry* 176, 29–35. <https://doi.org/10.1176/appi.ajp.2018.17040467>.
 68. Dickinson, D., Zaidman, S.R., Giangrande, E.J., Eisenberg, D.P., Gregory, M.D., and Berman, K.F. (2020). Distinct Polygenic Score Profiles in Schizophrenia Subgroups With Different Trajectories of Cognitive Development. *Am. J. Psychiatry* 177, 298–307. <https://doi.org/10.1176/appi.ajp.2019.19050527>.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Deposited data		
GWAS summary statistics for AAM from Day et al. 2017	https://www.reprogen.org/	Menarche_1KG_NatGen2017_WebsiteUpload.zip
Software and algorithms		
Plink2	https://www.cog-genomics.org/plink/2.0	NA
IBM SPSS Statistics v29.0.0.0	https://www.ibm.com/spss	NA
Gamm4 R package	https://cran.r-project.org/web/packages/gamm4	NA
STAR version 2.6.1	https://github.com/alexdobin/STAR	NA
DESeq2	https://bioconductor.org/packages/release/bioc/html/DESeq2.html	NA
QoRTS version 1.3.6	https://github.com/hartleys/QoRTs/releases/tag/v1.3.6	NA

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to the lead contact, Karen Berman (bermank@mail.nih.gov).

Materials availability

This study did not generate new unique reagents.

Data and code availability

- Data: GWAS summary statistics used for polygenic scoring were obtained from the supplementary data associated with the cited Day et al. publication.²³ Some data from the NIMH Intramural Research Program study defining the brain phenotype of children with WS and Dup7 are not publicly available due to institutional (IRB) restrictions. Data from the NIMH Intramural Longitudinal Study of the Endocrine and Neurobiological Events Accompanying Puberty will be publicly shared as permitted via IRB approval and participant consents.
- Code: This paper did not report original code. Software packages used are reported in the [key resources table](#).
- All Other Items: Any additional information required to reanalyze the data reported in this paper is available from the lead author upon request.

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

Participants were recruited as part of our long-term, longitudinal NIMH Intramural Research protocol, Defining the Brain Phenotype of Children with WS and Dup7, and were studied at the NIH Clinical Center (Protocol #10-M-0112, NCT01132885, ZIAMH002863). Additionally, some TD girls were recruited from the NIMH Intramural Longitudinal Study of the Endocrine and Neurobiological Events Accompanying Puberty (Protocol #11-M-0251, NCT01434368, ZIAMH002717/ZIAMH002933). Study procedures were approved by the NIH IRB. Parents of minor participants provided written informed consent and children provided assent. All participants were in good physical health, based on physical examination and a review of available medical records, had IQs in the low-normal to normal range based on neuropsychological testing, and had no significant morphological abnormalities on clinical brain MRI as well as intact cerebral vasculature, based on a radiologist's reading of MR images and MR angiography, respectively. Finally, no WS or Dup7 participants reported thyroid issues and no participants with Dup7 were placed on medications to alter pubertal timing. Demographic characteristics, including age and sex, are reported in [Table 1](#), separately for each analysis reported below.

METHOD DETAILS

Analysis of AAM polygenic scores in TD girls

Blood samples were collected from participants and genotyped using Illumina HumanOMNI-5M SNP chips. Data were phased using Shapeit and imputed using Impute2 software, as previously reported.⁶⁵ Polygenic scores (PGSs) for AAM were calculated with PLINK 2's "score" function,⁶⁶ based on previously reported GWAS summary statistics,²³ using 10 p value thresholds (ranging from 5×10^{-8} to 1.0). The polygenic signal was concentrated into a single score through principal component analysis and the first principal component was carried forward for

further analyses, based on previously described methods.^{67,68} The relation between the resulting PGS-AAM scores and clinician-ascertained AAM was determined using Pearson correlation in SPSS.

Association of AAM and 7q11.23 CNVs

AAM was ascertained, as above, through clinical interview of participants and their parents, including 17 girls with 7q11.23 CNVs: eight with WS, nine with Dup7, and 47 typically developing (TD) girls. AAM was also ascertained from 8 siblings of individuals with WS and 8 siblings of individuals with Dup7. Seven additional girls with WS had been excluded from analysis as they were treated with hormonal medications to pharmacologically delay the onset of puberty. No participants with Dup7 received hormonal treatment to alter the timing of puberty, and therefore none were excluded for this reason. A one-way ANOVA was used to calculate group differences in AAM in SPSS across WS, TD, and Dup7 participants. Post-hoc Tukey's honest significant difference tests were performed between each group to assess whether AAM significantly differed between WS and TD groups and between Dup7 and TD groups. Additionally, a one-way ANOVA was also used to examine other common genetic and environment factors that may influence the timing of puberty across WS siblings, TD, and Dup7 siblings.

Pituitary Volume Size Differences with 7q11.23 CNVs

For age-matched groups of 30 individuals with WS, 50 TD children, and 16 individuals with Dup7, investigators blinded to diagnostic group determined the volume of the pituitary by manual segmentation on T1-weighted ME-MPRAGE structural scans (TR/TE: 10.5/1.8 ms, Flip Angle: 7°, voxel size: 1 × 1 × 1 mm, 176 slices) collected on a GE 3T MRI scanner. Group differences in pituitary volumes across 7q11.23 copy number groups were assessed in SPSS using a general linear model, controlling for age at scan and sex. Post-hoc pairwise comparisons of estimated marginal means were also reported from the general linear model in SPSS. To further examine the developmental patterns of the pituitary volume across 7q11.23 CNV, 226 longitudinal visits (1–5 visits per participant, age range: 5–24 years old) of these same participants (all children returned approximately every two years for the study; WS: N = 30, 87 visits; TD: N = 50, 102 visits; Dup7: N = 16, 37 visits) were used to evaluate the development of pituitary size. R's *gamm4* statistical package was used to carry out a mixed-effects penalized-spline analysis to model age-dependent developmental trajectories within these three diagnoses groups.

Differential expression analysis of STAG3L2 in 7q11.23 CNVs

RNA was extracted from lymphocyte cell lines for 23 children with WS, 40 TD children, and 13 children with Dup7, and RNA sequencing was performed at the NIH Intramural Sequencing Center. Stranded Poly-A selected mRNA libraries were constructed from 1 µg total RNA for each sample using the TruSeq Stranded mRNA Kit (Illumina) according to the manufacturer's instructions. Amplification was performed using 10 cycles to minimize over-amplified product. Unique dual-indexed barcode adapters were applied to each library. Libraries were pooled in an equimolar ratio for sequencing. The pooled libraries were sequenced on an S4 flow cell on a NovaSeq 6000 using version 1.0 chemistry to achieve a minimum of 49 million 150 base read pairs. The data were processed using RTA version 3.4.4. Alignment of resulting RNASeq fastq files was performed using STAR version 2.6.1 to the GRCh37 genome build. Aligned reads were further processed using QoRTs version 1.3.6 and expression values of *STAG3L2* were analyzed using DESeq2 to test for stepwise group differences. Post-hoc pairwise t-tests of extracted normalized count values were conducted in SPSS to test for differences in *STAG3L2* expression between groups.

QUANTIFICATION AND STATISTICAL ANALYSIS

For the analysis of AAM in TD children, the relation between PGS-AAM scores and clinician-ascertained AAM was determined using Pearson correlation in SPSS. For the analyses of AAM in 7q11.23 CNVs, one-way ANOVAs were used to calculate group differences in AAM in SPSS, and post-hoc Tukey's honest significant difference tests were performed. For the cross-sectional analyses of pituitary volume, a general linear model was performed in SPSS, controlling for age at scan and sex. For the longitudinal pituitary volume data, mixed-effects penalized-spline modeling was analyzed using R's *gamm4* package. For the differential expression analysis, expression values were analyzed using DESeq2 to test for stepwise group differences and post-hoc pairwise t-tests of extracted normalized count values were conducted in SPSS.

ADDITIONAL RESOURCES

This work was part of two clinical trials, NCT01434368 and NCT01132885.