THE GENETICS OF PUBERTAL GROWTH AND TIMING

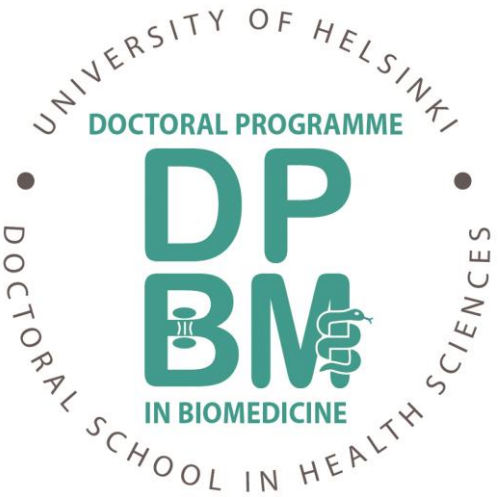
Diana L. Cousminer

ACADEMIC DISSERTATION

To be presented with the permission of the Faculty of Medicine of the University of Helsinki for public examination in Lecture Hall 3, Biomedicum Helsinki 1, on 17 April 2015 at 1 PM.

Institute for Molecular Medicine Finland (FIMM)
Faculty of Medicine
and
Doctoral Programme in Biomedicine (DPBM)
University of Helsinki
Helsinki, Finland

Helsinki 2015



Dissertationes Scholae Doctoralis Ad Sanitatem Investigandam Universitatis Helsinkiensis No. 21/2015

Cover art: Munch, Edvard. *Pubertet*. 1894. National Gallery, Oslo.

ISBN 978-951-51-0974-3 (paperback)
ISBN 978-951-51-0975-0 (PDF)

Hansaprint Oy
Helsinki 2015

Supervised by

Docent Elisabeth Widén, MD, PhD
Academy Research Fellow
Institute for Molecular Medicine, Finland (FIMM)
Helsinki, Finland

Reviewed by

Professor Hannes Lohi, PhD
Research Program Unit, Molecular Neurology, Faculty of Medicine
Department of Veterinary Biosciences, Faculty of Veterinary Medicine
University of Helsinki
Folkhälsan Research Center
Helsinki, Finland

Professor Taneli Raivio, PhD
Department of Physiology, Faculty of Medicine
University of Helsinki
Children's Hospital, Helsinki University Central Hospital (HUCH)
Helsinki, Finland

Opponent

Professor Gregory Gibson, PhD
School of Biology
Georgia Institute of Technology
Atlanta, GA, USA

“I may not have gone where I intended to go, but I think I have ended up where I intended to be.”

-Douglas Adams

For Dad

Abstract

Puberty is a highly variable developmental stage marked by the development of secondary sex characteristics and the attainment of reproductive maturity. Variation during childhood developmental phases correlates with altered disease risk in adulthood; variation in pubertal growth and timing, in particular, correlates with adult risk for type 2 diabetes, obesity, adverse cardiovascular heath, and hormone-dependent cancers. While normal variation in age at menarche (AAM) has recently been investigated in large-scale genome-wide association (GWA) studies, the genetic regulation of male puberty remains less understood. Moreover, extreme variation in pubertal timing is a common cause for referral to pediatric specialists, while the underlying genetic factors are largely unknown.

This work aimed to identify both common and rare genetic variants influencing pubertal growth and timing in both sexes. Utilizing Finnish population-based samples with frequent height measurements across puberty, we ran GWA of growth during late puberty and uncovered an association for variants near *LIN28B*, the most robust menarche-associated locus. Investigation of the longitudinal effects of two partly-correlated markers, one tagging a pubertal timing effect and one tagging an effect on adult stature, revealed distinct sex-specific association patterns with height growth from birth until adulthood. Thus, the *LIN28B* locus tags an important developmental regulator of both growth and maturational development.

We then expanded to include European-wide samples within the Early Growth Genetics (EGG) Consortium. Genetic mapping of three pubertal growth traits revealed 9 novel pubertal growth variants in addition to *LIN28B*, 5 of which also associated with pubertal timing, and one which associated with childhood and adult body mass index (BMI). Longitudinal investigation of these variants showed diverse patterns of association with height growth, some of which contradicted epidemiological correlations between rapid prepubertal growth, advanced puberty, and reduced final adult stature. Given the complex relationships between these traits, tracking individual unique effects across multiple periods of growth may help uncover the pathways linking childhood development with adult health outcomes.

Also within the EGG Consortium, GWA meta-analysis of Tanner genital and breast staging data uncovered the first robust male puberty locus on chromosome 16 near *MKL2*, a locus which also associates with advanced menarche, decreased pubertal growth, and shorter adult stature. Furthermore, part of the genetic architecture underlying the onset of puberty is shared between males and females, evidenced by the high correlation between menarche-advancing alleles and earlier male genital development. However, while BMI-increasing alleles strongly correlate with advanced breast development in girls, our data shows that these variants play a more complex role in male puberty.

Finally, we performed targeted sequencing of the pericentromeric region of chromosome 2 previously robustly linked with constitutional delay of growth and puberty (CDGP), an extreme delay in normal pubertal timing, in multiply affected Finnish families. Analysis of shared low-frequency variation in genes and regulatory regions of the best functional candidate genes revealed 6 protein-altering variants in a single gene, *DNAH6*, in 10 of the families. However, follow-up sequencing in an additional 135 Finnish CDGP cases failed to provide evidence for enrichment of *DNAH6* mutations compared to a large, unique set of SISu Finnish population controls. *DNAH6* is potentially an appropriate candidate gene that may be involved in the regulation of steroid hormone biosynthesis by the cytoskeleton. This study highlights the difficulties of detecting susceptibility variants under a linkage signal for complex traits.

Taken together, these results advance our understanding of the genetics of pubertal timing and development in both sexes. However, more work is needed to understand how each genetic locus functions in the biology of puberty and childhood growth, and further study of the genetic loci highlighted in this work may help pinpoint the mechanisms that link the timing of this important developmental stage with adult health and risk for common diseases.

Keywords: puberty, development, growth, genome-wide association studies (GWAS), targeted sequencing, constitutional delay of growth and puberty (CDGP)

Contents

List of original publications

This thesis is based on the following original publications that are referred to by the Roman numerals I-IV in the text:

I. Widén E, Ripatti S, **Cousminer DL**, Surakka I, Lappalainen T, Järvelin MR, Eriksson JG, Raitakari O, Salomaa V, Sovio U, Hartikainen AL, Pouta A, McCarthy MI, Osmond C, Kajantie E, Lehtimäki T, Viikari J, Kähönen M, Tyler-Smith C, Freimer N, Hirschhorn JN, Peltonen L, Palotie A. Distinct variants at LIN28B influence growth in height from birth to adulthood. *Am J Hum Genet*. 2010 May 14;86(5):773-82.

II. **Cousminer DL**, Berry DJ, Timpson NJ, Ang W, Thiering E, Byrne EM, Taal HR, Huikari V, Bradfield JP, Kerkhof M, Groen-Blokhuis MM, Kreiner-Møller E, Marinelli M, Holst C, Leinonen JT, Perry JR, Surakka I, Pietiläinen O, Kettunen J, Anttila V, Kaakinen M, Sovio U, Pouta A, Das S, Lagou V, Power C, Prokopenko I, Evans DM, Kemp JP, St Pourcain B, Ring S, Palotie A, Kajantie E, Osmond C, Lehtimäki T, Viikari JS, Kähönen M, Warrington NM, Lye SJ, Palmer LJ, Tiesler CM, Flexeder C, Montgomery GW, Medland SE, Hofman A, Hakonarson H, Guxens M, Bartels M, Salomaa V; ReproGen Consortium, Murabito JM, Kaprio J, Sørensen TI, Ballester F, Bisgaard H, Boomsma DI, Koppelman GH, Grant SF, Jaddoe VW, Martin NG, Heinrich J, Pennell CE, Raitakari OT, Eriksson JG, Smith GD, Hyppönen E, Järvelin MR, McCarthy MI, Ripatti S, Widén E; Early Growth Genetics (EGG) Consortium. Genome-wide association and longitudinal analyses reveal genetic loci linking pubertal height growth, pubertal timing and childhood adiposity. *Hum Mol Genet*. 2013 Jul 1;22(13):2735-47.

III. **Cousminer DL**, Stergiakouli E, Berry DJ, Ang W, Groen-Blokhuis MM, Körner A, Siitonen N, Ntalla I, Marinelli M, Perry JR, Kettunen J, Jansen R, Surakka I, Timpson NJ, Ring S, Mcmahon G, Power C, Wang C, Kähönen M, Viikari J, Lehtimäki T, Middeldorp CM, Hulshoff Pol HE, Neef M, Weise S, Pahkala K, Niinikoski H, Zeggini E, Panoutsopoulou K, Bustamante M, Penninx BW; ReproGen Consortium, Murabito J, Torrent M, Dedoussis GV, Kiess W, Boomsma DI, Pennell CE, Raitakari OT, Hyppönen E, Davey Smith G, Ripatti S, McCarthy MI, Widén E; Early Growth Genetics (EGG) Consortium. Genome-wide association study of sexual maturation in males and females highlights a role for body mass and menarche loci in male puberty. *Hum Mol Genet*. 2014 Aug 15;23(16):4452-64.

IV. **Cousminer DL**, Leinonen JT, Sarin AP, Chheda H, Surakka I, Wehkalampi K, Ellonen P, Ripatti S, Dunkel L, Palotie A, and Widén E. Targeted sequencing of the pericentromere of chromosome 2 linked to constitutional delay of growth and puberty. *Submitted to PLoS ONE.*

Abbreviations

1000G	1000 Genomes
A	adenine
AAM	age at menarche
ADAMTSL3	ADAMTS-like 3
ADCY3	adenylyl cyclase 3
ALSPAC	Avon Longitudinal Study of Parents and Children
BC58-T1DGC	1958 British Birth Cohort- Type 1 Diabetes Genetics cohort
BC58-WTCCC	1958 British Birth Cohort- Wellcome Trust Case Control cohort
BMI	body mass index
C	cytosine
CABLES1	Cdk5 And Abl Enzyme Substrate 1
CAMTA1	Calmodulin Binding Transcription Activator 1
CDGP	constitutional delay of growth and puberty
CHH	congenital hypogonadotropic hypogonadism
CHOP	Children's Hospital of Philadelphia
CI	confidence interval
CNV	copy number variant
COPSAC	Copenhagen Studies on Asthma in Childhood
DAX1	dosage-sensitive sex reversal, adrenal hypoplasia critical region, on chromosome X, gene 1
DHEA	dehydroepiandrosterone
DHEA-S	dehydroepiandrosterone sulfate
DILGOM	Dietary, life style, and genetic determinants of obesity and metabolic syndrome
DNAH6	dynein, axonemal, heavy chain 6
DNA	deoxyribonucleic acid
ECM	extra-cellular matrix
EFEMP1	EGF-containing fibulin-like extracellular matrix protein 1
EGG	Early Growth Genetics
ENCODE	Encyclopedia of DNA Elements
eQTL	expression quantitative trait locus
ER	endoplasmic reticulum
ERK1	Extracellular signal-regulated kinase 1
ExAC	Exome Aggregation Consortium
FGF	fibroblast growth factor
FGFR	fibroblast growth factor receptor
FIMM SeqLab	Institute for Molecular Medicine Finland Sequencing Laboratory
FIN-1000G	Finnish 1000 genomes subset
FSH	follicle-stimulating hormone
G	guanine

GABA	gamma-aminobutyric acid
GC	genomic control
GENEVAR	GENe Expression VARiation
g:GOSt	Gene Ontology Statistics
GH	growth hormone
GHSR	growth hormone secretagogue receptor
GIANT	Genetic Investigation of ANthropometric Traits
GINIplus	German Infant Study on the influence of Nutrition Intervention PLUS environmental and genetic influences on allergy development
GNA12	Guanine nucleotide-binding protein subunit alpha-12
GnRH	gonadotropin-releasing hormone
GNRHR	Gonadotropin-releasing hormone receptor
GOYA	Genome-Wide Population-Based Association Study of Extremely Overweight Young Adults
GPR54	G protein-coupled receptor 54
GSE	gene set enrichment
GWA	genome-wide association
FANCL	Fanconi anemia complementation group L
FDR	false discovery rate
FTCS	Finnish Twin Cohort Study
H2000	Health 2000
HBCS	Helsinki Birth Cohort Study
HDL	high-density lipoprotein
HESX-1	Homeobox, ES Cell Expressed 1
HLA	human leukocyte antigen
HPG	hypothalamus-pituitary-gonad
HR	hazard ratio
HSL	hormone-sensitive lipase
IA	intracranial aneurysm
IBD	identity-by-descent
IBS	identity-by-state
IGF-1	insulin-like growth factor 1
indel	insertion-deletion
IPP	idiopathic precocious puberty
KAL1	X-linked Kallmann syndrome
KISS1R	kisspeptin receptor
KS	Kallmann syndrome
LD	linkage disequilibrium
LEIPZIG	Leipzig childhood cohort
LEP	leptin
LEPR	leptin receptor
LH	luteinizing hormone
LHRH	luteinizing hormone-releasing hormone

LHX3	LIM homeobox 3
LIN28A, LIN28B	Lin-28 homolog A/B
LISAplus	Influence of life-style factors on the development of the immune system and allergies in East and West Germany Plus the influence of traffic emissions and genetics
LOD	logarithm-of-odds
MAF	minor allele frequency
MAPK3	mitogen-activated protein kinase 3
MAZR	MAZ-related factor
MC4R	melanocortin 4 receptor
MDS	multidimensional scaling
miRNA	microRNA
MKL2	myocardin-like 2
MKRN3	makorin ring finger protein 3
MTCH2	mitochondrial carrier 2
NFBC1966	Northern Finland Birth Cohort 1966
NFBC8586	Northern Finland Birth Cohort 1985-86
NHS	Nurses' Health Study
nIHH	normosmic idiopathic hypogonadotropic hypogonadism
NTR	Netherlands Twin Registry
PATZ1	POZ (BTB) and AT hook containing zinc finger 1
PAX-3	paired box 3
PCs	principal components
PCR	polymerase chain reaction
PDS	pubertal development scale
PHV	peak height velocity
PIAMA	Prevention and Incidence of Asthma and Mite Allergy birth cohort study
POMC	proopiomelanocortin
PROK2	prokineticin 2
PROKR2	prokineticin 2 receptor
PROP1	PROP paired-like homeobox 1
PWM	position weight matrix
PYY	peptide YY
PXMP3	Peroxisome assembly factor 1
QIMR	Queensland Institute of Medical Research
RAINE	Western Australian Pregnancy Cohort Study
RORA	RAR-Related Orphan Receptor A
RR	relative risk
SD, SDS	standard deviation, standard deviation score
SE	standard error
SF-1	splicing factor 1
SISU	Sequencing Initiative Suomi
SNP	single nucleotide polymorphism

SOCS2	suppressor of cytokine signaling 2
SPOCK	sparc/osteonectin, cwcv and kazal-like domains proteoglycan
STRIP	Special Turku Coronary Risk Factor Intervention Project
T	thymine
TAC3	Tachykinin-3
TAC3R	Tachykinin-3 receptor
TEENAGE	TEENS of Attica: Genes and Environment Study
TGF	Transforming growth factor
TF	transcription factor
TMEM38B	transmembrane protein 38B
TOB1	transducer of ERBB2, 1
TRH	thyrotropin-releasing hormone
VCP	variant calling pipeline
VEP	Variant Effect Predictor
VGLL3	vestigial-like family member 3
WT1	Wilms tumor 1
YFS	Cardiovascular Risk in Young Finns Study
ZBTB38	zinc finger and BTB domain containing 38

1 Introduction

In the news, stories report that puberty is advancing for both girls and boys. These reports often link the advancement of puberty to rising levels of obesity, and sometimes to endocrine-disrupting chemicals in the environment. While the timing of puberty has captured the public's attention because of the social implications of earlier-maturing teens, earlier puberty also has important implications for health later in life. Yet when it comes down to it, the basic biology of what triggers pubertal onset, and how the timing of this developmental phase relates to adult health, is quite poorly understood. While it is clear that changing environmental factors are important for driving secular advances in puberty seen over the last century, it is shared genetic mechanisms that likely play a large role in the links between early development, pubertal timing, and health in adulthood.

Puberty is a developmental stage that all adolescents go through. However, the most commonly used marker of puberty is age at menarche, when girls experience their first period. Since there is no non-invasive, easy-to-assess marker similar to menarche in boys, male puberty is understudied. One trait that occurs in both sexes is the pubertal growth spurt. The timing and magnitude of the pubertal growth spurt correlate closely with the central onset of puberty. Additionally, the pubertal growth spurt contributes up to 15-20% of final adult stature (1), so genetic variants associating with the growth spurt may also represent genetic height growth potential (2, 3). Finally, the pubertal growth spurt is associated with adult health outcomes, making it a promising avenue for the exploration of the genetic mechanisms underlying puberty, growth, and their associated health risks later in life.

Another measure of pubertal development that is collected in some childhood cohort studies is the Tanner pubertal development scale, a 5-stage scale of male genital development and female breast development ranging from childlike to adult morphology. Both genital development and breast development begin early in puberty, closely following central activation in the hypothalamus. These markers are likely to tag the timing of pubertal onset more closely than age at menarche, since the tempo of pubertal development varies and because menarche occurs on average 2 years after puberty begins. Thus, studying these traits may reveal genetic pathways that are more closely related to the mechanisms which trigger central pubertal onset.

Finally, the most common cause of referral to pediatric specialists is extremely early or delayed puberty. For pubertal delay in particular, very little is known about the molecular mechanisms involved. At the extreme late end of the normal population distribution, constitutional delay of growth and puberty (CDGP) represents a little over 2% of the population, and carries risks for short adult stature, decreased bone mineral density, and

psychosocial problems. Despite the prevalence of late growth and puberty, only a small handful of genetic variants are known which predispose this delay in pubertal onset.

In this thesis, our primary aims were to map common variants associating with pubertal growth and timing in both males and females, and to localize less common variation that might predispose late puberty in CDGP families. Since the timing of this important developmental phase may have psychosocial implications and correlates with risks for adult obesity, type 2 diabetes, hormone-dependent cancer, and cardiovascular health, determining which genetic variants impact developmental trajectories between puberty and adverse health outcomes is an important public health question.

2 Literature review

2.1 Human genetics

Genetic variation contributes in some way to nearly every human ailment. Inherited factors, interacting with diet, environment, and lifestyle, contribute substantially to our risk for developing common diseases such as cancer, diabetes, and cardiovascular disease, which together constitute the majority of the adverse health burden in the United States (www.genome.gov). As genomic research advances, scientists are gaining powerful tools for identifying and characterizing these genetic factors, which ultimately enable medical researchers to improve health care at a personal level. While translating genetic findings from the laboratory to the clinic takes time and effort, the pace of discovery over the past few decades has been rapid, and genomic findings are beginning to find their way into medical practice. For example, genetically inherited variation can affect how a person responds to drug treatments, and knowledge gained from personal genetic profiles of variants in genes encoding drug-metabolizing enzymes and drug transporter proteins will allow better prediction and management of adverse drug reactions (4). As another example, accurate prediction of tumor phenotypes in cancer using genomic data will have a substantial impact on the diagnosis, personalised treatment, and prevention of cancer (5). These are only a couple of examples of the promise that genetic discovery holds for the future of medical science.

2.1.1 A brief history

Although the word *genetics* wasn't coined until 1905, the history of genetics is often traced back to the work of the Augustinian friar Gregor Mendel, who worked out the basic laws of genetic inheritance in his studies of garden peas. Presented to the Natural History Society of Brünn in 1865, Mendel's *Experiments on Plant Hybridization* showed that blocks of inheritance come in discrete units, some of which are dominantly expressed over other variations of the trait (6). His important discoveries remained largely underappreciated for nearly 40 years, but once they were rediscovered, renewed interest in the study of the inheritance of traits began.

In the 19th century, there were other important advances. In 1871, Friedrich Miescher discovered "nuclein," now known as nucleic acids (7), which we now know are the building blocks of DNA. Chromosomes and mitosis were discovered in 1882 by Walther Flemming (8), providing the first glimpse at how cells replicated and divided material into daughter cells. Furthermore, Francis Galton developed a basic statistical model for inheritance, proposing that offspring receive half their inherited characteristics from each parent and a quarter from each grandparent (9).

The first half of the 20th century witnessed countless more seminal discoveries. Working with *Drosophila* as a model organism, Thomas Hunt Morgan developed gene theory– that traits are passed from parent to offspring via discrete genes – and the principal of linkage – that certain alleles are somehow coupled and inherited together (10); Barbara McClintock's work in maize demonstrated chromosomal crossing-over, which is the exchange of genetic material between homologous chromosomes (11); Avery, MacLeod, and McCarty showed that the agent of inheritance was deoxyribonucleic acid (DNA) and not protein (12); and Watson and Crick famously explained both the shape and function of DNA simultaneously in their double helix model, in which DNA consists of two backbones joined by the complimentary nucleic acid base pairs adenosine (A) – tyrosine (T) and cytosine (C) – guanine (G) (13). In 1961 scientists cracked the genetic code, showing that each codon (a triplet of bases) translates into a different amino acid (14).

The pace of advancement quickened further with the refinement of the polymerase chain reaction (PCR) in 1985 (15), a method which uses successive cycles of heating and cooling to amplify a tiny DNA sample exponentially in a short timeframe. Using this technology, the first positionally cloned gene, *CFTR*, was published for cystic fibrosis in 1989 (16). This essential tool enabled DNA technology to speed forward, spurring the launch of the Human Genome Project in 1990, a planned 15-year project aiming to fully sequence the human genome and several model organism genomes (http://web.ornl.gov/sci/techresources/Human_Genome/index.shtml). With competition heating up between the National Institutes of Health and the privately funded Celera Genomics in the late '90s, the human genome first draft was published ahead of schedule in 2000, and the final draft in 2003 (17, 18). In the meantime, the International HapMap project (http://hapmap.ncbi.nlm.nih.gov/) was launched in 2002, aiming to more fully understand genetic variation among individuals and populations, and was the first step towards identifying genetic variation relevant for common complex diseases. Also following the publication of the human reference genome, the Encyclopedia Of DNA Elements (ENCODE) and Functional Annotation of the Mammalian Genome (FANTOM) projects were created to provide high-quality, comprehensively annotated candidate functional and regulatory elements in the human and mammalian genomes for hundreds of different cell types, eventually providing scientists with a detailed map of how the genome is built and functions (19, 20).

While genetics in the latter part of the 20th century saw success in localizing and identifying single genetic mutations leading directly to disease phenotypes ("Mendelian" diseases), the cataloging of single nucleotide polymorphisms (SNPs) by the HapMap project allowed dense genotyping chips to be developed, cost-effectively resolving the variants present at thousands of positions along the genome in large-scale population or birth cohorts. The combined availability of genotype data and densely phenotyped

populations resulted in an explosion of phenotype-genotype associations for complex traits and diseases. Additionally, advancements in sequencing technology brought down both the cost and time needed to perform high-throughput sequencing, making whole-exome and whole-genome sequencing common tools for detecting variants associated with traits and diseases (21). Further, sequencing technology advancements have spurred large-scale projects like the 1000 Genomes Project (22), aiming to discover lower-frequency and structural variants in diverse global populations, and the International Cancer Genome Consortium (23), coordinating cancer genome projects worldwide. These large, globally collaborative projects promise to usher in a new era of scientific cooperation and personalized genomic medicine (21).

2.1.2 A tour of the human genome

The human genome consists of 46 chromosomes, of which 23 are inherited from each parent. These consist of 22 autosomes and 2 sex chromosomes (X and Y). The nucleus of almost every cell in the human body contains the complete complement of chromosomes, which are structures made up of DNA and tightly organized by proteins. Each of our cells also contains a small amount of extranuclear DNA, contained in an organelle called the mitochondria.

DNA is a chemical compound made up of a double helix consisting of two linear chains of nucleotide bases, adenine (A), thymine (T), guanine (G), and cytosine (C). Each base makes a pair with its complement in a predictable manner – A always pairs with T and G always pairs with C. The human genome is made up of over 3 billion base pairs, and the order of these base pairs determines the meaning of our genetic sequence. The minority of this sequence is coding, containing the genes which are instructions for making proteins, and other parts of the genome are non-coding, but still contain meaningful instructions, directing when and how genes are expressed.

The publication of the complete human reference genome revealed some surprises about the structure of the genome. Current estimates place the number of genes in the genome at around 23,000, many fewer than previously thought, because the same gene can be transcribed into multiple different proteins and adapted for use in different biological contexts. The vast majority of these genes, while predicted computationally, correspond to orthologous genes in the mouse genome, suggesting that they are truly biologically relevant sequences. Many more ribonucleic acid (RNA) genes (around 6,500) and pseudogenes (over 12,000) also populate the human genome (21).

Repetitive sequences account for at least half of the genome, a category made up of both simple sequence repeats and segmental duplications (SDs), or copy number variants

(CNVs). The breadth and distribution of SDs came as a surprise, representing as much as 5% of the genome in extensive inter- and intra-chromosomal duplications (21). Specific SDs and CNVs are relevant for disease; for example, these have been associated with autism, schizophrenia, and epilepsy (24).

Finally, the ENCODE project is beginning to reveal how the genome functions. A functional element is defined as a region of the genome encoding a defined product, like a protein, or containing a reproducible biochemical signature (such as transcription, or a specific chromatin structure). These signatures are thought to mark genomic sequences that have important functions, like exons (the coding part of genes), sites of RNA processing, or regulatory elements like promoters, enhancers, silencers, or insulators. While it is possible to associate certain signatures with specific functions, the ultimate biological function or role of a given genomic element will require functional validation (19) that could take decades to complete. Furthermore, biochemically active regions cover a larger proportion of the genome than evolutionarily conserved regions, so a combination of complementary approaches will be needed to sort out which elements really do function in human biology and disease (25).

2.1.3 Gene mapping: methods for gene localization

Gene mapping is the process of localizing and identifying genes that underlie inherited traits or diseases. Methods used for gene mapping are largely based on correlation between measurable features of a trait or disease (known as phenotypes) and genetic variants (genotypes) across the genome. Alternatively, a case/control approach may be used which compares the frequency of genetic variants between two groups of study subjects, those who have a disease (or surpass an arbitrary cut-off threshold of a quantitative trait) and those who do not.

Since the 1980s, naturally occurring sites of genetic variation have been used as marker tags for gene mapping. Types of genetic variation that can contribute to variation in the development of a trait or disease include, but are not limited to, SNPs, or single base pairs that are variable in the population, CNVs, short tandem repeats, and insertions-deletions (indels). Early studies mapped many genes harbouring variants that caused rare Mendelian (single-gene) disorders (26), and newer methods allow the discovery of genes containing variants that associate with susceptibility to common diseases and variation in complex traits, or quantitative trait loci (QTL). The type and number of markers used to tag the genome has evolved from several hundred microsatellite tags to current “SNP chips,” which genotype hundreds of thousands or even millions of SNPs per sample quickly and cost-effectively in a single experiment. Along with imputation to fully sequenced reference genomes, these advances bring the number of polymorphisms per individual to more than ten million, densely tagging the entire genome. Additionally, the methods used

have developed, allowing mapping not only of rare variants with strong phenotypic effects, but also more common variants in genes contributing to complex, multigenic traits, including common diseases such as diabetes and heart disease.

Approaches for mapping genes can be hypothesis-free or candidate gene-based. Two examples of hypothesis-free methods are linkage and association. Linkage analysis was first developed in the early 20th century for mapping genes in fruit flies. With the advent of microsatellite markers and genome maps, this method was eventually successful for mapping genes in humans with the discovery of the gene for cystic fibrosis in 1989 (16) and for Huntington's disease in 1993 (27). Linkage analysis follows the inheritance patterns of genotyped markers in families, and localizes genomic regions harbouring disease genes by comparing the segregation of these markers with the transmission of the phenotype of interest. This method unearthed genetic loci underlying many Mendelian diseases, as well as familial versions of common diseases. However, linkage has met with more limited success as a method for detecting genes involved in polygenic or complex traits, except in the identification of rare, highly penetrant risk alleles in families with strong aggregation of specific complex diseases or alleles with large effects on quantitative traits (28).

More recently, genome-wide association (GWA) has become widely used for QTL mapping. In this method, association between common genetic variants and phenotypic variation in a trait of interest are detected by linear regression analysis of hundreds of thousands of genotyped, or millions of imputed, variants across the genome against phenotypic measurements. For example, in 2010, the Genetic Investigation of ANthropometric Traits (GIANT) Consortium published a GWA study of adult stature in over 180,000 individuals in which 180 common variants were associated with normal variation in final height (29). Other studies have used a case/control design, for example a recent study which reported 10 new common susceptibility loci for type 2 diabetes by analyzing approximately 35,000 cases and 115,000 controls of European descent (4).

Hypothesis-free methods have the advantage that no previous biological knowledge is necessary for detecting genetic associations, and thus also provide an avenue for discovering unknown biological pathways and mechanisms. Indeed, candidate-gene based approaches have often been severely limited by the dearth of current relevant knowledge. Still, both approaches have vastly increased our knowledge of the genetics of both simple and complex traits and diseases over the past few decades. As of November 2014, the Catalog of Published Genome-Wide Association Studies contained 2056 publications and 14,854 SNPs with association data (www.genome.gov/gwastudies), and the number of genes with a phenotype-causing mutation was 3,299 (http://www.omim.org/statistics/geneMap).

2.2 Pubertal development

2.2.1 The physiology of puberty

Puberty is the developmental process by which children become reproductively competent, involving both physiological and behavioral changes. This period of maturation generally spans 3-4 years, normally initiating between the ages of 9 and 14, with girls entering puberty an average of two years earlier than boys. At the population level, the timing of pubertal onset falls under a continuous distribution. Although genetic variation plays a major role in the timing of pubertal onset, ethnicity, health status,

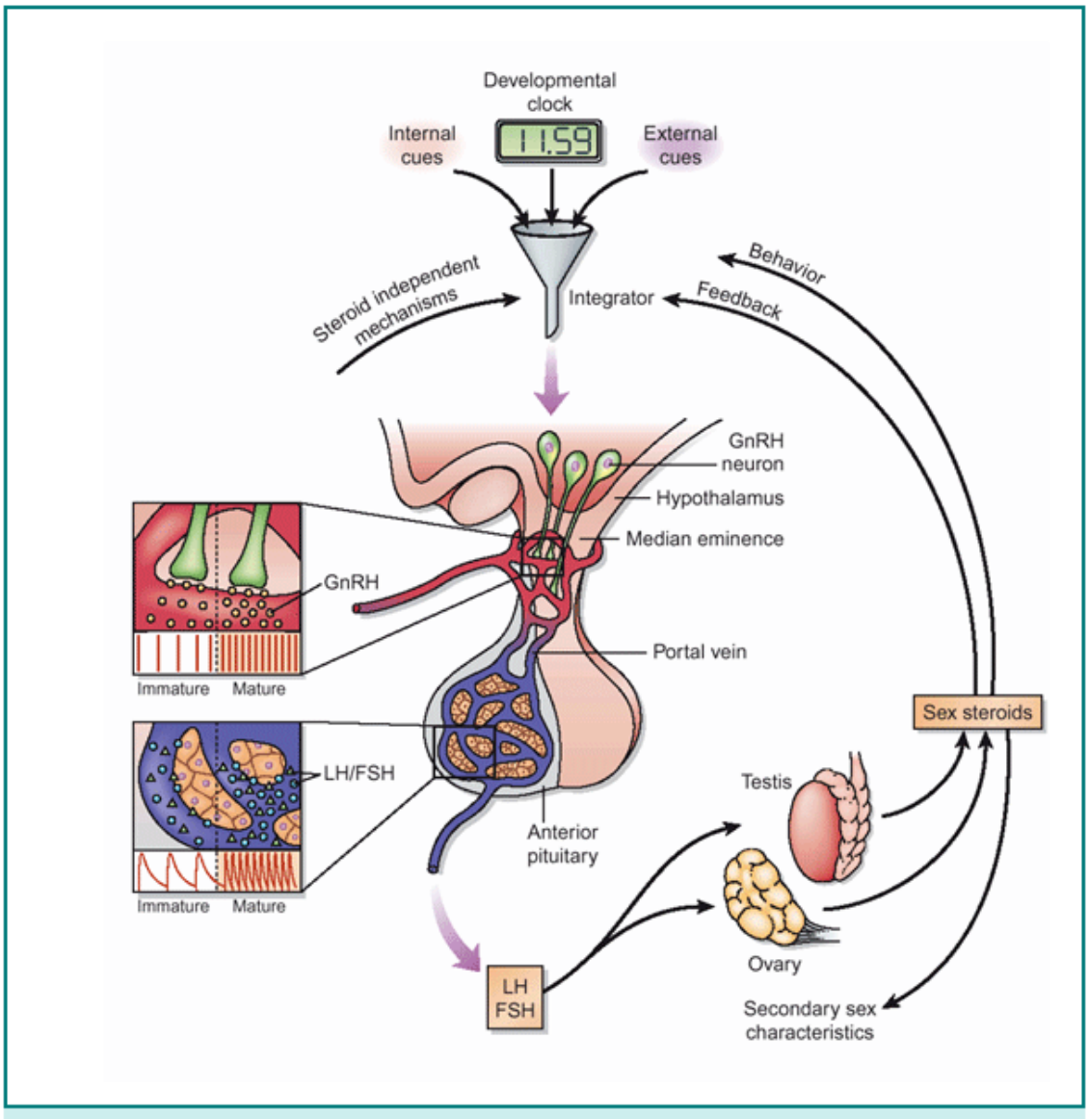


Figure 1. *Signal integration and the hypothalamus-pituitary-gonadal axis. Reprinted from Sisk and Foster (2004) with permission from Nature Publishing Group.*

nutrition, activity level, and other environmental agents can all modulate the timing of puberty (30).

Puberty begins with a precisely orchestrated sequence of events triggered by the integration of multiple genetic and environmental signals by the body's "developmental clock" **(Figure 1)** (31). While the mechanisms that integrate these various internal and external signals remain largely unclear, the result is activation of the hypothalamic-pituitary-gonadal (HPG) axis, beginning with the pulsatile release of gonadotropin-releasing hormone (GnRH) at a characteristic frequency from the hypothalamus to the anterior pituitary gland (32) after a period of childhood quiescence. In the pituitary, pulsatile GnRH stimulates the release of the gonadotropins luteinizing hormone (LH) and follicle stimulating hormone (FSH), which travel through the blood stream to the gonads (ovaries and testis), where they stimulate gonadarche, the production of mature gametes and sex steroids (30). Steroid hormones released from the gonads return through a positive feedback loop to the brain where they influence GnRH secretion and sexual behavior (33). This complex system of communication between the hypothalamus, pituitary gland, and gonads perpetuates hormonal signals, triggering the physiological changes occurring during pubertal development, such as the pubertal growth spurt, breast development (thelarche) and menarche in girls, and voice break and testicular enlargement in boys.

GnRH neurons develop in the olfactory placode (nasal compartment) from progenitor cells and migrate into the basal hypothalamus during 6-16 weeks of gestation in humans (34). During this migration, many factors are likely to be important for successful development, including molecules related to cell-surface interactions, for example membrane receptors and adhesion molecules, and those related to distance communication, like neurotransmitters (34). The GnRH neural network and the HPG axis are partly functional at birth (the "minipuberty" of infancy), become restrained by unknown mechanisms, and are reactivated at pubertal initiation (33, 35, 36). Although secretion of GnRH is restrained, release of the hormone from the hypothalamus in a pulsatile fashion is continuous throughout childhood. During pubertal initiation, the frequency of GnRH release speeds up and approaches a characteristic frequency that triggers the pubertal development program to begin. Both transsynaptic and glial-neuronal communication are crucial to this change; neuronal and glial input to the GnRH neuronal network increase, including excitatory signals like glutamate and kisspeptin, while inhibitory input, most importantly from γ-aminobutyric acid (GABA) and opiatergic neurons, decreases (37).

Kisspeptins are, in fact, critical for the activation of GnRH neurons during pubertal onset. Kisspeptins are a family of neuropeptides encoded by the *KISS1* gene, and act by binding to the G-protein coupled receptor 54 (GPR54, otherwise known as KISS1R) (38). Kisspeptin is produced by neurons in the hypothalamus that have direct connections with GnRH neurons (39). Kisspeptin activity not only activates GnRH production, but also upregulates KISS1R in GnRH neurons.

Some physical features associated with puberty, such as pubic hair, develop independently of the HPG axis. Adrenarche, which is defined as the onset of dehydroepiandrosterone (DHEA) and DHEA-sulfate (DHEA-S) production from the adrenal gland beginning around age 6, results in the growth of androgen-dependent hair (pubarche) in both sexes prior the activation of the HPG axis. Pubarche occurs around age 8 in both sexes, although patterns of hair growth are sexually dimorphic (40).

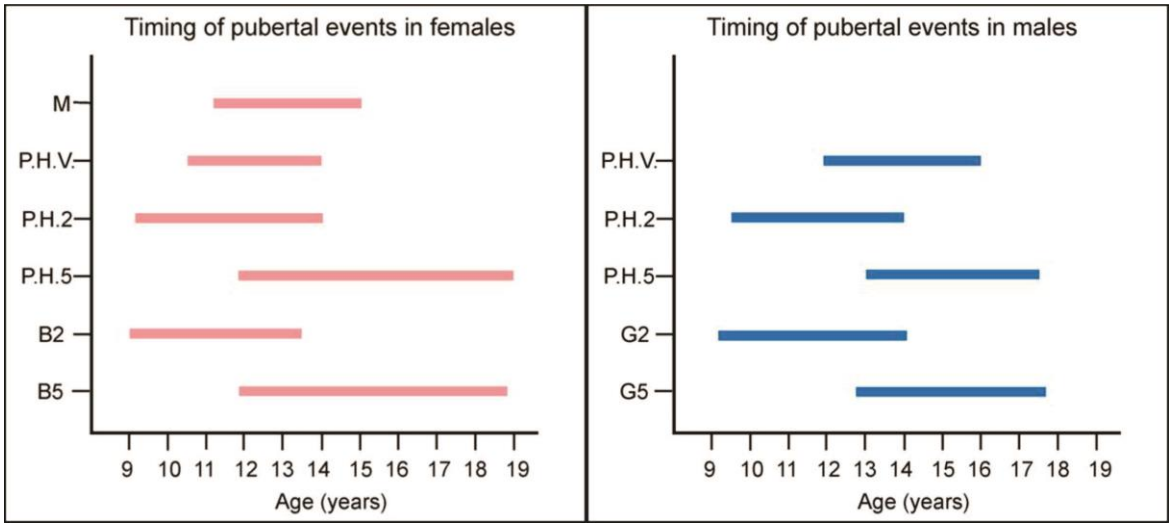


Figure 2. *A schematic representing the timing that the majority of adolecents experience pubertal events. M, menarche; P.H.V., peak height velocity; P.H.2, Tanner pubic hair stage 2; P.H.5, Tanner pubic hair stage 5; B2, Tanner breast stage 2; B5, Tanner breast stage 5; G2, Tanner genital stage 2; G5, Tanner genital stage 5 Adapted from* Brook's Clinical Pediatric Endocrinology, *6th edition (Ed. Brook 2009) and* The Cambridge Encyclopedia of Human Growth and Development, *(Ed. Ulijaszek 1998) with permission.*

2.2.2 Clinical assessment of normal pubertal development

The onset of puberty can be measured indirectly by monitoring the development of secondary sex characteristics, height growth velocities, and onset of menstruation (age at menarche, AAM). Development of breast tissue is the first clinical sign of central pubertal onset in girls (41), while in boys, the first physical indication is enlargement of the testis and reddening of the scrotal skin (42) **(Figure 2)**. While the first menstruation is the most commonly assessed pubertal marker in large-scale epidemiological studies, it is a relatively late event in puberty, occurring on average a little over 2 years after the onset of breast development, and its timing varies between populations and ethnic groups. The onset of puberty can also be estimated by charting the pubertal growth spurt, which almost doubles the velocity of growth after a childhood phase of slower, more gradual growth.

The take-off phase of this growth spurt, as well as the period of peak height growth velocity (PHV) and age at which final adult height (FH) is attained, are dependent on the central (hypothalamic) activation of puberty, so they occur earlier in individuals maturing early, and are delayed in those maturing late (43). Thus, timing of the growth spurt can give an estimation of the timing of puberty and may prove useful as a simple phenotypic measure in studies aimed at detecting genes relating to pubertal timing in both sexes (44).

The measurement of pubertal status can be imprecise, as puberty is a complex process that proceeds at different ages and at a unique tempo in each adolescent. Problems in ascertainment of precise pubertal phenotyping can be due to these individual differences, a lack of precise measurement techniques, or difficulty in obtaining permission to use the best measurement techniques available, which are sometimes thought to be invasive to one's privacy, particularly the measurement of testicular volume (45).

Several non-invasive scales have been developed to assess the stage of maturation, either by clinician assessment or self-assessment. The most widely used scale, the Tanner scale, was developed by Marshall and Tanner in the late 1960's based on British girls and boys participating in the Harpenden Growth Study who were followed every 3 months during adolescence. The researchers took photographs of secondary sex characteristics (pubic hair in both sexes, breasts in girls, and genitals in boys) and classified each trait into five developmental stages **(Figure 3, Box 1)**, ranging from pre-adolescent childlike morphology to fully mature (41, 42). Alternative assessment systems have also been developed, for example the Pubertal Development Scale (PDS) (45), as a way for adolescents to self-assess their pubertal status.

The most accurate assessment of secondary sexual characteristics is likely to be performed by trained health professionals (46). However, clinical assessments are not always practically feasible, so self- or parental assessment is sometimes necessary. The validity of these alternative assessments remain debated, with some studies concluding that self-assessments are accurate and agree highly with physician rating (47, 48). Other authors present concerns that subjects might overestimate their development at the early maturational stages and underestimate their development at later stages (49, 50), but suggest that self-assessment can be used in research where approximations of the timing of sexual maturation are sufficient (50). Furthermore, pubertal self-assessment is not reliable in obese children (51).

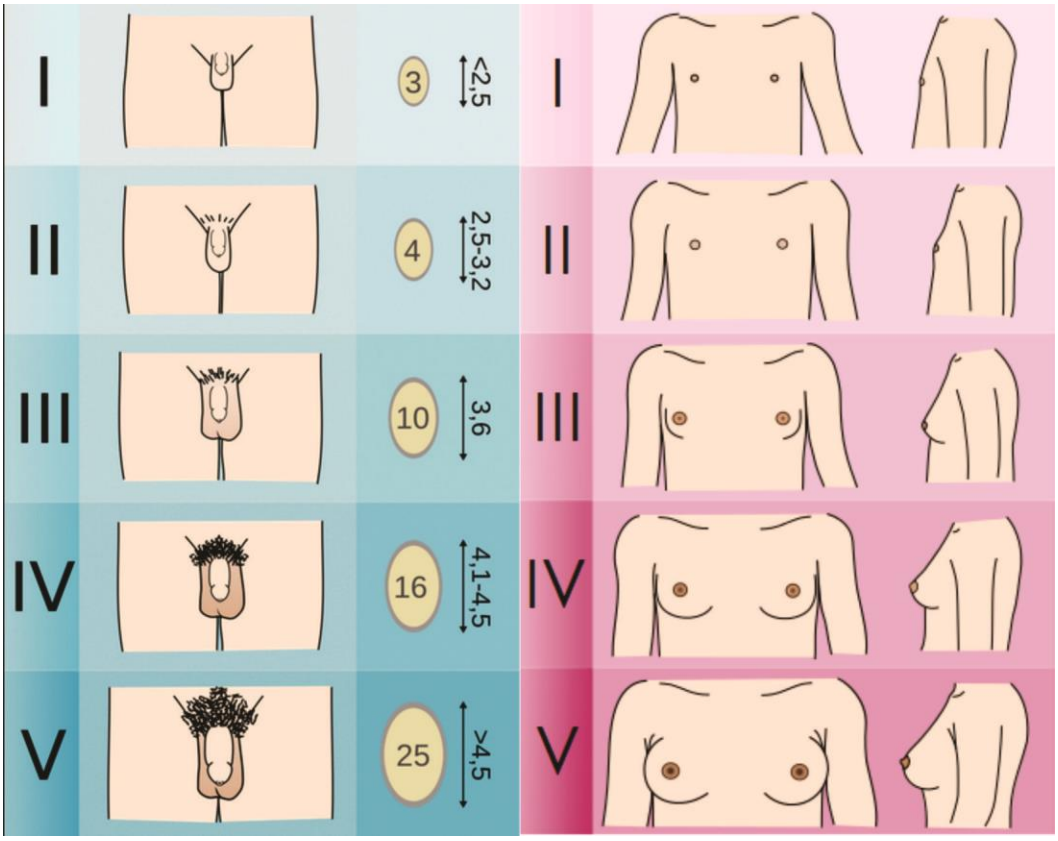


Figure 3. *Tanner stages for male genital development (left) and female breast development (right). For boys, the numbers inside the ovals represent average testicular volume in ml, and numbers to the right represent testicular size. Illustrations by Michal Komorniczak. Shared under the Creative Commons license.*

2.2.3 Heritability of pubertal timing

The high correlation within racial and ethnic populations, within families, and between monozygotic twin pairs (52, 53) suggests that inherited factors play a major role in the variability of pubertal timing. Heritability estimates place the genetic contribution to pubertal variation at 50 - 80% (54–57). A recent study which examined the familial resemblance of age at menarche in 25,970 females estimated heritability at 57% (95% confidence interval 53%, 61%) (55). Heritability estimates of other markers of central pubertal onset were much higher; for example, a study of the pubertal height growth spurt in the Swedish Young Male Twins Study estimated the heritability of the age of onset of the growth spurt to be 91% and the age at peak height velocity to be 93% (58). These studies all show that genetic variation plays a substantial role in determining the timing of pubertal initiation.

Box 1. Tanner puberty scale

Boys- genital stages

I. Preadolescent; testes, scrotum, and penis are of approximately the same size and proportion as in early childhood.
II. The scrotum and testes have enlarged, and there is a change in the texture of the scrotal skin. There is also some reddening of the scrotal skin.
III. Growth of the penis has occurred, at first in length and then in breadth. There has been further growth of the scrotum and testes.
IV. Penis further enlarged in length and breadth with development of glans. Testes and scrotum are further enlarged, with further darkening of the scrotal skin.
V. Adult in size and shape.

Girls- breast stages

I. Preadolescent; elevation of papilla only.
II. Breast bud stage; elevation of breast and papilla as a small mound, enlargement of areola diameter.
III. Further enlargement of breast and areola, with no separation of their contours.
IV. Projection of areola and papilla to form a secondary mound above the level of the breast.
V. Mature stage; projection of papilla only, owing to recession of the areola to the general contour of the breast.

Adapted with permission from Ahmed, et al (2009), Cell Press.

2.2.4 Early life and environmental influences on the timing of puberty

Environmental influences during childhood and as early as during the prenatal period also influence the timing of puberty. Two recent large-scale studies, the Sister Study (n = 33,501 women (59)) and the Breakthrough Generations Study (n = 81,606 (60)), both found strong associations between earlier menarche and low birth weight and maternal smoking, a positive association with maternal age at birth, and an inverse association with birth order. Additionally, prenatal exposure to maternal pre-eclampsia (60), pregnancy-related hypertensive disorder, and maternal pre-pregnancy diabetes (59) were correlated with advanced menarcheal timing in daughters. During infancy or childhood, not being breastfed and being heavy or tall at age 7 were correlated with advanced menarche (60), and soy formula was associated with both early and late menarche (59). Another study noted similar correlations between intrauterine and early postnatal growth and age at take-off of the pubertal growth spurt and age at peak height velocity in both boys and girls (61).

While overweight and obesity have received much attention in recent years, nutrition more specifically may also contribute to altered pubertal timing. Observational studies show that children with the highest intake of vegetable-based protein experienced puberty up to 7

months later, while children who ate more animal-based protein began puberty up to 7 months earlier (62). High flavone or isoflavone (phytoestrogens which may perform hormone-like activities) intake also correlated with later breast development (63) and peak height velocity in girls (64). Thus, the type and quality of diet during childhood may influence pubertal timing and related later-life health outcomes (65).

Finally, energy metabolism and physical activity in childhood also play a role in pubertal timing. Gymnasts are a well-known example; young female gymnasts who train extensively often experience delayed menarche (66). Conversely, children who exercised little during childhood more often had earlier menarche (60). These effects probably act through nutritional cues and metabolic control via peripheral hormones such as leptin and ghrelin that sense the body's metabolic state and communicate this information back to the reproductive HPG axis (67, 68).

2.2.5 Secular trends toward earlier maturity

The majority of studies looking at pubertal timing trends over the last century or so have used menarcheal age as the standard unit of measurement. In the first half of the 20th century, AAM declined until reaching a plateau, but recent evidence suggests that the trend toward earlier puberty has resumed (69). In the Breakthrough Generations Study (n = 94,170 women), the average age at menarche declined from women born between 1908-19 (mean AAM 13.5 yrs) and 1945-49 (mean AAM 12.6 yrs), stabilized, and then resumed the downward trend for girls born in 1990-93 (average AAM 12.3 yrs) (69). Studies of other populations show a similar trend towards earlier puberty for current datasets (70).

Observations of earlier onset of pubertal markers other than age at menarche also show that puberty in general is beginning earlier, possibly as a consequence of increased adiposity. A study which looked at the onset of breast development in a longitudinal cohort noted a trend toward earlier thelarche consistent with temporal changes in body mass index (BMI) (71).

Recent studies have also noted a possible trend in the advancement of puberty in boys. One study found a long-term steady decline in the age of male sexual maturity since the mid-eighteenth century, a trend which paralleled that seen in girls and suggests the presence of common environmental cues that have influenced both male and female pubertal timing (72). Other small studies either support (73–76) or refute (77) these observations in various populations, though what remains clear is that more studies are needed to determine whether there is a contemporary secular trend towards earlier puberty in boys (78).

2.3 Disorders of pubertal timing

2.3.1 Early puberty

Normal puberty can occur as early as 9 years in boys and 8 years in girls, or as early as 7 years in non-Hispanic African American or Mexican girls (79). When the development of secondary sexual characteristics happens before these ages, it may be due to precocious puberty **(Figure 4)**. In girls, central onset of puberty at a young age is most often idiopathic and accounts for 90% of precocious puberty in girls (79), while in boys there is more likely to be an underlying intracranial cause (80) as precocious puberty is idiopathic in only 50% of cases in boys (79). In idiopathic precocious puberty (IPP), the HPG axis is activated and puberty progresses according to a normal pattern. In contrast, in peripheral precocious puberty, secondary sex characteristics can develop without central gonadotropin release, most often caused by a gonadal or adrenal gland disorder (80).

2.3.2 Delayed puberty

Delayed puberty is defined as the absence of pubertal signs (breast or testicular development) after the age of 13 years in girls and 14 years in boys. Pubertal delay is thought to be more common in boys (though this may be due to referral bias), with constitutional delay of growth and puberty (CDGP) accounting for most cases of male pubertal delay. CDGP is characterized by a delay in height growth, skeletal maturation, and pubertal development, and the lack of normal adolescent growth is often the reason for specialist referral. Levels of the sex hormones and gonadotropins correlate with bone age rather than chronological age, and spontaneous secretion of growth hormone (GH) is reduced, while response to GH is normal (81).

CDGP represents the extreme late end of the normal pubertal spectrum, and puberty eventually begins spontaneously. However, delayed puberty can be difficult to distinguish from absent puberty due to a congenital lack of gonadotropin (hypogonadotropic hypogonadism, HH). Delayed puberty may also be due to chronic endocrine, metabolic, or systemic disorders, or simply to undernutrition (79).

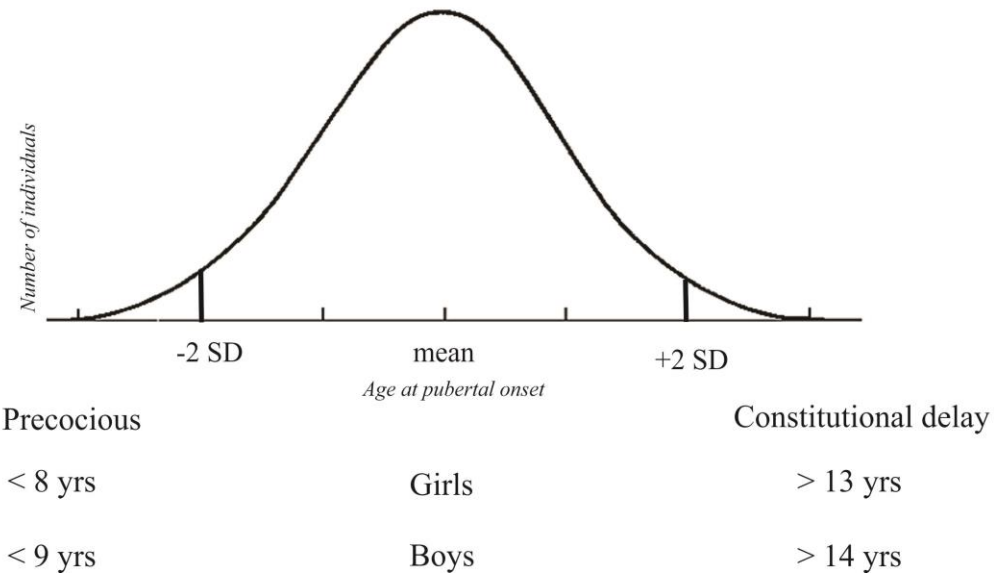


Figure 4. *Pubertal onset is a complex trait that is normally distributed in the population. Ages for precocious and delayed puberty are sourced from Prentice and Williams (2013).*

2.3.3 Absent puberty

If there is a lack of pubertal initiation by age 18, it is considered pathologic (79). Congenital hypogonadotropic hypogonadism (CHH) is difficult to distinguish from CDGP, but it is characterized by the inability to produce the gonadotropic hormones LH and FSH in the pituitary gland (82). In this case, the gonads are normal. As there is no definitive test to distinguish CHH from CDGP, patients require close monitoring. In around 60% of CHH patients, the disorder is associated with anosmia or hyposmia (lack or limited sense of smell), when it is known as Kallmann syndrome (KS). In KS, the normal migration of GnRH neurons from the olfactory placode to the ventral hypothalamus during embryonic development fails. Although the precise prevalence is unknown, there is a clear difference in estimates between men and women. In Finland, for example, the incidence rate was found to be 1:30,000 men and 1:125,000 women (83). This may be partly because one gene associated with KS, *KAL1*, is located on the X chromosome.

Alternatively, hypogonadism can be hypergonadotropic, with elevated levels of LH and FSH, indicating a lack of negative feedback to the hypothalamus (82). In this case, there is a failure of the gonads to produce estrogen or testosterone, and follicular maturation or sperm production fail to occur.

In some cases, the reproductive phenotype may be quite severe, while in others, HH may reverse later in life (84). CHH is generally treated with lifelong hormonal therapy, but research indicates that for some individuals, this may not be necessary.

Finally, HH may be acquired rather than congenital. The cause may be tumors in the hypothalamus or pituitary, autoimmune diseases, or to nutritional deficit (as in anorexia nervosa) or intense physical activity. Additionally, radiotherapy for the treatment of leukemia or brain tumors may cause hypogonadism (85).

2.4 Links between puberty, childhood growth, and adult disease

Puberty is one developmental stage along a continuum of growth events from conception until adulthood. There are substantial correlations between childhood body size (both in height and weight) and pubertal timing, as well as between pubertal timing and growth during the pubertal growth spurt and final adult stature. Furthermore, growth patterns and pubertal timing have been linked to various adult health outcomes and risks, suggesting that there may be shared mechanisms underlying the etiology of developmental events and risk of developing certain diseases.

2.4.1 Height growth

Postnatal height growth is a complex process, consisting of the partly overlapping phases of rapid infant growth, slowly diminishing childhood growth and a growth spurt during puberty **(Figure 5)**. In particular, the pubertal growth spurt accounts for 15-20% of final adult stature (1), and represents both the activation of central puberty and local growth factors (2, 3). Final stature represents both genetic potential and the timing of pubertal growth spurt, which is dependent on the timing of the central activation of puberty.

The majority of children remain at approximately the same centile of height (which is largely genetically determined) compared to their peers during most of their growth (43), except during puberty. During this time, early maturing individuals move to a higher centile in comparison to their peers, since they have entered the pubertal growth spurt earlier than others, and when they mature they tend to drop back to their pre-puberty centile. In contrast, a late developer drops below their previous growth centile and catches up once the pubertal growth spurt occurs.

The timing of puberty can also impact the total amount of pubertal growth, and thus final adult stature. In girls, earlier menarche, breast development, and pubertal growth spurt have been associated with an earlier cessation of growth and shorter adult stature (86–88). In boys, however, the opposite correlation has been seen, with an earlier growth spurt relating to slower progression through puberty and taller stature at age 18 or adulthood (58, 86).

As for adolescents with constitutional delay of growth and puberty, final height is sometimes significantly shorter than the mean of healthy adults (81). This is true for both boys and girls (81, 89–91). This reduction in final height for some CDGP patients may be due to a reduction in height SDS already prior to puberty (between ages 3 – 9 years), which does not seem to occur in the subset of patients who do reach their genetic height potential (92).

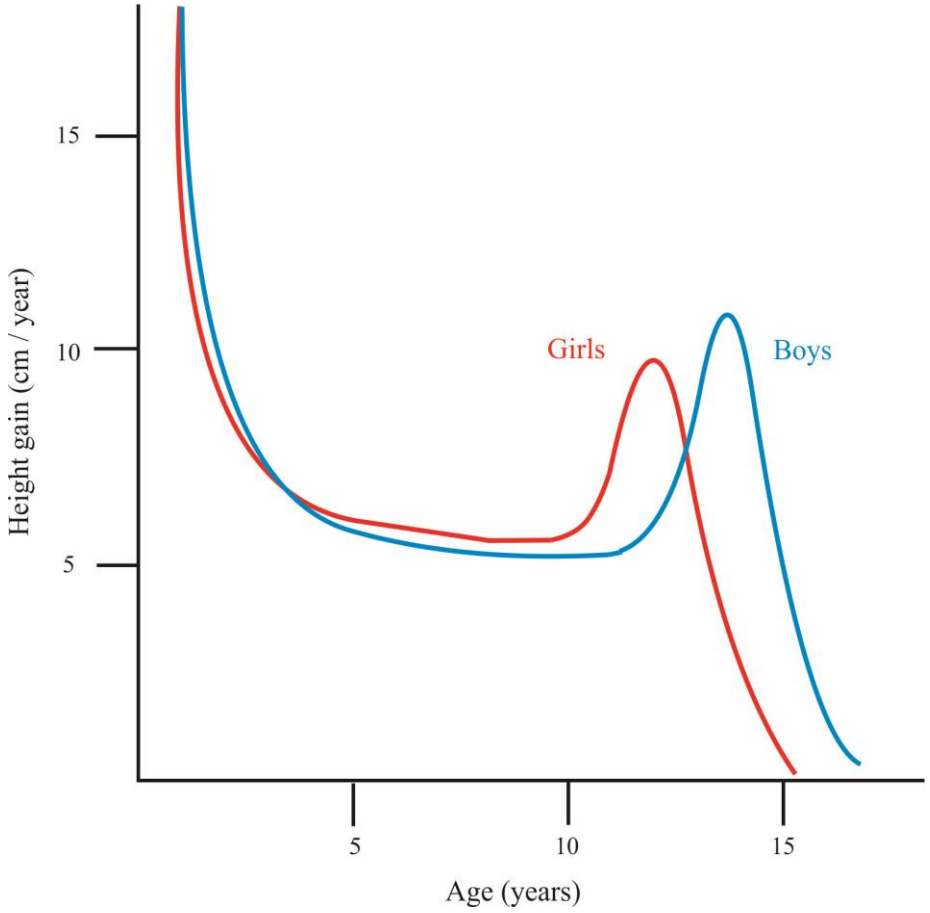


Figure 5. *Height growth velocity from birth until adolescence showing rapid growth during infancy, a decline in height velocity over childhood, and the pubertal height growth spurt in girls (red) and boys (blue).*

2.4.2 Overweight and obesity

There are well-documented links between obesity in childhood, pubertal timing, and height growth, indicating a complex relationship between body size and maturational development. Overweight or obese children are likely to have a higher growth velocity and be taller in childhood, but subsequently grow less during puberty while achieving normal adult stature (93–96). This increase in growth appears to be GH-independent,

suggesting that other explanations, for example increased leptin and insulin levels, adrenal androgens, insulin-like growth factor (IGF)-1, IGF-binding protein-1 and GH-binding proteins, may explain why obese children have accelerated growth (97). In fact, GH levels are reduced in obese children, which is suggested to cause the growth reduction seen during puberty (98). Further, while there are clear correlations between increased adiposity and early puberty in girls, some studies in boys show that overweight/obesity may link with early puberty, while others link increased adiposity to pubertal delay (95, 96, 99, 100). Some studies show that these effects are mediated via levels of pubertal hormones (97), but it is clear that the underlying mechanisms are complex and partly sex- specific.

2.4.3 Altered pubertal timing and adult health risks

Alterations in normal pubertal timing have disease risk implications, both during adolescence as well as for the risk of disease in adulthood. A range of effects may be related to both early and late development (101). Studies of precocious puberty have noted the most adverse effects, including accelerated skeletal maturation and short adult height, early sexual debut, potential sexual abuse and difficulties such as depression and eating disorders (102), as well as higher risk of reproductive tract cancers (103). Premature adrenarche as well as early menarche are associated with higher risk of developing metabolic syndrome-related problems, such as obesity, type 2 diabetes, and cardiovascular disease (103), in addition to increased overall mortality in women (104). Additionally, delayed puberty can lead to stressful psychosocial effects (101), and delayed menarche is associated with an increased risk for bone fracture and low mineral density in the postmenopausal period (105).

2.4.4 Cancer

The timing of puberty is a risk factor for the development of hormone-dependent cancer later in life. For example, much attention has been focused on the potential links between the timing of puberty and risk for breast cancer. A recent study investigated the timing of pubertal stages and later life breast cancer in over 100,000 women and found that breast duct development may be a period of heightened susceptibility for cancer risk; specifically, increased breast cancer risk was associated with earlier thelarche (hazard ratio (HR) [95% CI] = 1.23 [1.02, 1.48], 1 [referent] and 0.80 [0.69, 0.93] for ≤10, 11-12 and ≥13 years respectively) and earlier menarche (1.06 [0.93, 1.21], 1 [referent] and 0.78 [0.62, 0.99] for ≤12, 13-14 and ≥15 years) (106). A correlation between menarcheal timing and breast cancer risk was also found in a huge study of almost 400,000 women, where risk of breast cancer increased by a factor of 1.05 (95% CI 1.04-1.06; $P < 0.0001$) for every year younger menarche (107). More specifically, the risk for both estrogen receptor-

negative and positive (ER-PR- and ER+PR+) malignancies were inversely associated with increasing age at menarche (108, 109).

Other potentially hormone-dependent cancers have been linked to pubertal timing in both sexes; for example, the timing of menarche is inversely associated with ovarian cancer risk (oldest compared to youngest AAM, relative risk (RR) [95% CI] = 0.85 [0.75, 0.97]) (110). Also, lifetime growth, specifically during puberty, is associated with testicular cancer risk, possibly mediated through environmental factors that affect growth over the lifecourse (111).

The timing of menarche has further been investigated as a risk factor for other types of cancer that are not likely to be hormone-dependent, ie. kidney cancer (112), but such links are more tentative and require further study.

2.4.5 Metabolic disorder and related disease

Metabolic syndrome is a disorder of energy use and storage characterized by central obesity, high blood pressure, elevated fasting glucose levels, high serum triglycerides, and low high-density cholesterol (HDL) levels. Metabolic syndrome increases risk for developing cardiovascular disease and diabetes, and represents a major health care burden in the US and developed countries.

Altered pubertal timing is one developmental event that has been associated with risk for metabolic-syndrome related disease later in life. Advanced puberty (early AAM) is correlated with increased risk for cardiovascular disease events, cardiovascular disease mortality, and mortality overall in women, and these findings were only partly mediated by increased adiposity. Women with AAM earlier than 12 years had a higher RR compared to other women of hypertension [1.13 (95% CI: 1.02-1.24)], incident cardiovascular disease [1.17 (1.07-1.27)], incident coronary heart disease [1.23 (1.06-1.43)], all-cause mortality [1.22 (1.07-1.39)], cardiovascular disease mortality [1.28 (1.02-1.62)] and cancer mortality [1.25 (1.03-1.51)], adjusted for many potential confounding factors such as age, smoking, physical activity, hormone use, and measures of body size like BMI and waist circumference (104). Early AAM has also been associated with increased risk for type 2 diabetes later in life in a large study of over 100,000 women, also partly mediated through excessive adult adiposity and increased exposure to sex hormones. RR of type 2 diabetes mellitus across AAM categories ($\leq$ 11, 12, 13, 14, and $\leq$ 15 years) was 1.18 (95% CI: 1.10, 1.27), 1.09 (1.02, 1.17), 1.00 (referent), 0.92 (0.83, 1.01), and 0.95 (0.84, 1.06), respectively, in the Nurses' Health Study (NHS) (*P* for trend < 0.0001) and 1.40 (1.24, 1.57), 1.13 (1.00, 1.27), 1.00 (referent), 0.98 (0.82, 1.18), and 0.96 (0.78, 1.19), respectively, in NHS II (*P* for trend < 0.0001) (113).

Until recently, studies relied on AAM to investigate the correlation between the timing of puberty and later development of metabolic-related disorder. Using the timing of pubertal growth as a proxy for pubertal timing, it has also been possible to look at the link between puberty and cardiometabolic health in both sexes. Specifically, earlier pubertal timing was associated with higher adult BMI, fasting insulin, diastolic blood pressure, and decreased HDL cholesterol, all of which are metabolic syndrome-related derangements, in both men and women (114).

Common to all of these observations is that puberty is a key developmental period during the lifecourse which impacts health in later life. Therefore, it is critical that we gain an understanding of the genetic and biological mechanisms that link puberty with adult health, and investigate how health during adolescence might be modified to lower the risks of developing disease in adulthood.

2.5 The genetics of pubertal timing

2.5.1 Single-gene (Mendelian) mutations

Single mutations in more than 21 genes have been associated with delayed or absent puberty (115). Some of these genes are involved in the development of the olfactory nervous system, with mutations typically resulting in anosmia or hyposmia (reduction in the ability to smell) and HH. Another set of mutations affect the kisspeptin-KISS1R signaling pathway, which mediates the release of GnRH by modulating the function of GnRH neurons and pituitary gonadotropins. These mutations together explain the cause of abnormal puberty in approximately 30% of cases (36). Finally, there are genetic variants involving HH and obesity as well as the development of the HPG axis (115).

Based on the mode of inheritance, there are three types of Kallmann syndrome (KS): X-linked (KAL1), autosomal dominant (usually with incomplete penetrance), and autosomal recessive (116), although occasionally digenic inheritance has also been noted (117).

X-chromosomal KS is caused by mutations in the gene *KAL1*, encoding an extracellular matrix (ECM) glycoprotein called anosmin-1 that plays a key role in the migration of GnRH neurons and olfactory nerves to the hypothalamus. Mutations in *KAL1* prevent GnRH neurons from entering the brain, which results in a lack of pubertal development and infertility (118). Several different genes have been implicated in the autosomal dominant form of KS. These include mutations in various fibroblast growth factor 8 (*FGF8*) (119) and the fibroblast growth factor receptor 1 (*FGFR1*) (120). Mutations in the prokineticin 2 (*PROK2*) signaling pathway underlie recessive and compound heterozygous cases (121). The tyrosine kinase fibroblast growth factor receptor 1 (FGFR1) is required

for the development of the olfactory system and interacts with anosmin-1 (116). Although patients with the KAL1 form of KS have a severe phenotype with absence or extreme delays in pubertal onset, patients with *FGFR1* mutations have a wider spectrum of phenotypes, some of which may also cause reversible hypogonadotropic hypogonadism. This suggests that modifier genes exist which have not yet been identified (84, 122). Mutations in the prokineticin-2 gene (*PROK2*) and its receptor (*PROKR2*) were uncovered in a study of KS patients, which implicates an important role for prokineticin signaling in olfactory and HPG axis development (115). In total, the genetic origin only around 30% of KS cases has been explained (117).

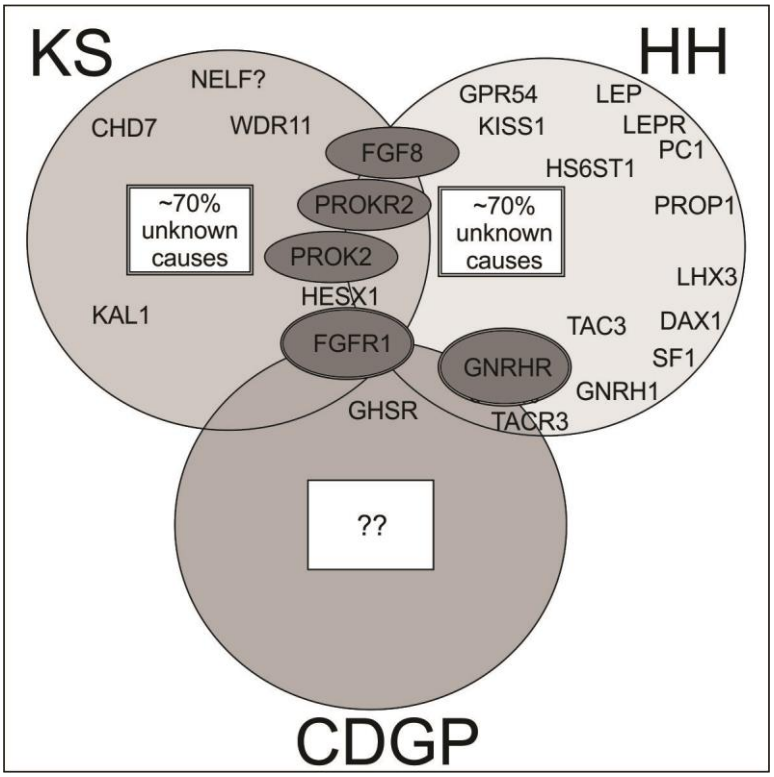


Figure 6. *The genetic basis underlying Kallmann syndrome (KS), hypogonadotropic hypogonadism (HH), and constitutional delay of growth and puberty (CDGP). Adapted from Gajdos, et al (2010) with permission from the authors.*

After GnRH neurons have completed migration, numerous genes are involved in their regulation. Mutations in these genes and those which regulate pituitary development and function can result in CHH without anosmia, for example in the gonadotropin-releasing hormone receptor (*GNRHR*) gene (123), which is estimated to account for up to 40% of all autosomal recessive cases of IHH and should be the first candidate gene screened in the clinic (124).

A delay in puberty diagnosed as CHH may also result from mutations in genes that affect gonadotropins, such as *GnRHR, LH,* and *FSH*, or in genes involved in HPG development, like *LHX3, PROP1, HESX-1, DAX1,* and *SF-1*. Gonadotropins are heterodimers made up of an α-subunit shared by all glycoprotein hormones and a unique β-subunit. Delayed puberty has also been reported in patients with mutations in the *FSH* and *LH* β-subunit genes (116, 123).

Additionally, mutations in *KISS1R* have been detected in HH patients (125). KISS1R is the G-protein coupled receptor for a family of structurally related peptides called kisspeptins, which are encoded by the *KISS1* gene (126). The KISS1-KISS1R system is a critical part of the HPG axis and plays a key role in pubertal onset (125). Kisspeptin-GPR54 signaling plays an ongoing role in the HPG axis as a regulator of LH and GnRH secretion as well in sex steroid feedback in adults (52).

Other genes are responsible for a link between obesity and HH, including mutations in prohormone convertase 1 (*PC1*), an endopeptase required for post-translational processing of some prohormones and neuropeptides, which have been found in patients with obesity and HH (116).

Finally, mutations in leptin (*LEP*) or the leptin receptor (*LEPR*) have also been seen in HH patients (127). In mice, leptin deficincy leads to a complex phenotype with endocrine and metabolic abnormalities. In one study, 3 homozygotes for a missense mutation were infertile due to hypothalamic-pituitary hormone deficiency, 2 of whom were adults at the time the paper was published. The female had primary amenhorrea, and the male had hypogonadism; both were morbidly obese (127).

Leptin is secreted by white adipose tissue in proportion to the amount of fat stores, and normally acts as a satiety indicator and regulates body weight (128). Leptin also seems to have a permissive role in pubertal onset. For example, in women with hypothalamic amenorrhea due to excess exercise or nutritional deficit, leptin treatment has restored reproductive function (129). AAM is inversely related to the serum concentration of leptin (130), and, in both sexes, leptin concentration rises before the pubertal transision, followed by an increase in FSH, LH, and sex steroids (131). Leptin functions as a pleiotropic mediator in a range of neuroendocrine systems, regulating neurotransmitters such as polypeptide YY (PYY), which are believed to regulate GnRH secretion (116, 128). However, GnRH neurons do not express leptin receptors themselves (132). Recently, it has been shown that leptin may stimulate the secretion of kisspeptins, which then act on GnRH neurons (67). Additionally, a common variant near *LEPR* was associated with the timing of menarche in the general population ($P = 3.5 \times 10^{-8}$) (133), lending support to the notion that adequate energy storage is an important positive signal which allows puberty to commence.

In the past year, focus has shifted to the role of epigenetic mechanisms in puberty. Four heterozygous mutations in makorin RING-finger protein 3 (*MKRN3*) were found by whole-exome sequencing in families with central precocious puberty. Interestingly, the gene lies in an imprinted region of chromosome 15 critical for Prader-Willi syndrome, and all of the mutations were transmitted from the affected patients' fathers. Studies in mice hypothalami confirmed that the expression of the gene decreased immediately prior to puberty, suggesting that *MKRN3* normally plays a repressive role (134).

Recently, a study investigated common variants in *KAL1, FGFR1, PROK2* and *PROKR2* for potential links to early or late menarche and found no association, suggesting that although these genes are important in puberty, they don't play a major role in the normal variation of pubertal timing in the general population (135). Another study did not find that genetic variation in *GNRH* or *GNRHR* is a common cause of late puberty in the general population (115). Conversely, the most recent GWA study of AAM did find that common variants associated with menarche were enriched in or near genes for Mendelian disorders of puberty (enrichment $P = 0.05$) (133). These genes included *MKRN3, LEPR, TACR3*, and *GNRH1*. The recent GWA study needed to include over 180,000 women to detect these effects, so one reason why these studies came to different conclusions is likely that candidate studies were simply too small to detect subtle effects in the general population.

2.5.2 Constitutional delay of growth and puberty: a grey area

The genetics of constitutional delay of growth and puberty likely represents a grey area between rare Mendelian alterations that underlie severe pubertal delay or absence and more subtle variation underlying normal population variability in pubertal timing. One or a few genetic variants with large effect size likely drive the tail ends of normally distributed population traits (136). Thus far, however, very little is known about the actual genes which cause pubertal delay. In fact, only a handful of variants have been found in CDGP patients, including a homozygous mutation in *GNRHR* (137), several heterozygous point mutations in the GH secretagogue receptor gene (*GHSR*) (138), and a nonsynonymous variant in tachykinin receptor 3 (*TAC3R*) (139). Other genes known to underlie disorders of puberty (*FGFR1, GNRHR*, and *TAC3*) (140), and variation in a gene strongly associated with pubertal timing in the general population (*LIN28B*) (141), were not found to harbor mutations in CDGP patients.

Two studies investigating the inheritance patterns of CDGP showed that many families displayed an apparent autosomal dominant transmission pattern for a susceptibility variant (142, 143). Other inheritance patters were also seen, which suggested that there are likely

single genes with major effects whose penetrance is affected by genetic or environmental modifiers (142). Subsequently, a linkage analysis located a region of linkage on the pericentromere of chromosome 2 (144), strongly suggesting that this region of the genome harbors a gene predisposing pubertal delay in multiply affected families.

2.5.3 Common variants influencing pubertal timing in the population

Prior to the era of GWA studies, little was known about common variants which influence the timing of puberty in the general population, despite the wide variability in timing and tempo between individuals. In 2008, a genome-wide linkage study found no evidence for highly penetrant common variants influencing AAM in women from Australia, The Netherlands, and the United Kingdom (102). In 2009, a small GWA study identified associated variants in the gene *SPOCK*, but subsequent studies failed to replicate this locus (145).

This was the current state of knowledge about the genetics of puberty in the general population in 2009, when we began work on the first study included in this thesis, a gene mapping study of the pubertal growth spurt in several large Finnish population birth cohort studies. Later that same year, four studies, including between 5,000 and 25,000 study subjects each, were published **(Table 1)**. These studies identified a robust association between SNPs near *LIN28B*, a heterochronic regulator of developmental timing, and AAM (146–149), the same locus which robustly associated with pubertal growth in our study. Variation near the gene also associated with earlier breast development in girls, earlier pubic hair development and voice break in boys, a faster tempo of height growth and shorter adult stature in both sexes (146). A secondary locus on chromosome 9 was also highlighted by these studies.

Realizing that increasing the sample size would also increase power to detect common variants with smaller effect sizes, we joined forces with other teams interested in the genetics of reproductive traits within the ReproGen Consortium, which came to include 175 researchers from 104 international research institutions. In 2010, we contributed data from several Finnish cohorts to a study of AAM including 87,802 women of European descent, which identified 30 new loci at the genome-wide significant level, including 4 that had previously been associated with body mass index, 3 loci in or near genes for energy homeostasis, and 3 loci implicated in hormonal regulation (150). A further 10 loci were associated at the suggestive level.

The most recent large-scale GWA study of AAM published by the ReproGen consortium included over 180,000 women of European ancestry, from 57 contributing cohort studies including our contribution from the Finnish population cohorts, and mapped 123 signals at 106 genomic loci as associated with the timing of menarche (133). All but one of the

previously published signals showed association with menarche in this study, and 2.71% of the variance in AAM was explained by these 123 SNPs in an independent sample of 8,689 women. Interestingly, six independent signals were located in imprinted regions of the genome, and parent-of-origin effects were found at 14q32 which harbors the reciprocally imprinted genes *DLK1* and *MEG3*, 15q11-13 in the imprinted region critical for Prader-Willi syndrome, and at *KCNK9*, representing an enrichment when compared to all published genome-wide significant alleles for any trait (6/123, 4.8% vs. 75/4332, 1.7%, $P = 0.017$).

Studies investigating the association between the timing of menarche in non-European populations have revealed a large degree of overlap among distant populations. The majority of these studies aimed to replicate European menarche loci in different populations. 8 loci were replicated in 3468 Hispanic women from the United States as part of the Women's Health Initiative SNP Health Association Resource (151), and 25 (60%) of 42 European loci also contained significant SNPs associated with the timing of menarche in African American women (n = 18,089) (152). One of these was the locus at *RORA*, which was suggestively associated in European women, and also associated with Tanner brest development stage in our third study included in this thesis. In 6,929 Chinese women from Shanghai, China, 32 variants had the same direction of effect as in European women, and 9 of them were statistically significant (153). Finally, a study of over 15,000 Japanese women showed that SNPs at two European loci also associated at a statistically significant level with menarche in this population, and that six more loci were suggestively associated (154).

Other markers of puberty, including measurements of the pubertal height growth spurt and secondary sex characteristics, have not been studied comprehensively in GWA studies. Candidate studies of the association between adult height-associated SNPs and the pubertal growth spurt have only revealed association for SNPs at *LIN28B* (155) and *SOCS2* (156). Consequently, very little is known about the genetic underpinnings of male pubertal maturation and how menarche loci relate to pubertal initiation, tempo, or other markers of puberty.

In conclusion, it is clear that advances are being made in our understanding of the genetic architecture of pubertal timing in the general population, both in Europeans as well as other ethnicities. However, one limitation of these studies is that they are all primarily investigating the timing of menarche, which is a female-specific trait which occurs an average of two years after the onset of puberty. In the studies included in this thesis, we have looked at phenotypes that are present in both sexes, such as the pubertal growth spurt, and sexually dimorphic traits which closely follow the onset of puberty, female breast and male genital development. Therefore, our work has enabled the investigation of common variants influencing puberty in males, as well as allowing us to study the genetic

overlap of different phases of pubertal development, both important contributions to the study of pubertal genetics.

2.5.4 Genes with common variants: their role in rare extremes of puberty

Few studies have examined whether genes located in GWA studies contain rare variants which underlie extreme pubertal delay or precocity, but those which have did not find evidence for association. *LIN28B*, and its homolog *LIN28A*, were investigated in precocious (157, 158) and delayed (141) puberty, but no mutations were found. Clearly, more extensive studies are needed to determine whether genes with subtle effects on pubertal timing in the population might harbor rare mutations that would predispose a more extreme phenotype.

Table 1. *Published GWA studies to date that have investigated age at menarche*

Year	First author	Population	Sample size	Replication sample size	Number of significant loci	Top locus	Top SNP	*P*	Effect size
2014	Perry	European*	182,416		123	*LIN28B*	rs7759938	7.8×10^{-110}	1.4 months
2013	Pyun	Korean	3,437		0				
2013	Tanikawa	Japanese	15,495		0	*LIN28B*	rs364663	5×10^{-7}	1.1 months
2013	Demerath	African American	18,089	2,850 African Americans, 87,802 Europeans	1	*ZNF483*	rs10441737	4×10^{-15}	
2010	Elks	European*	86,142	12,813	30	*LIN28B*	rs7759938	5×10^{-60}	1.5 months
2009	Ong	European	4,714	16,373	1	*LIN28B*	rs314276	4×10^{-16}	1.7 months
2009	Perry	European	17,510		2	*LIN28B*	rs7759938	7×10^{-9}	1.1 months
2009	Sulem	European	15,297	10,040	1	*LIN28B*	rs314280	2×10^{-14}	1.2 months
2009	He	European	17,438		2	*LIN28B*	rs314277	2.7×10^{-13}	1.9 months
2009	Liu	European	477	854 European siblings, 762 Europeans, 1,387 Chinese	0				

**We have contributed data from several Finnish cohorts towards these large meta-analyses.*

3 Study aims

While puberty is a key developmental window associated with both early life events and later life health risks, the genetic mechanisms regulating pubertal timing and growth are understudied, especially in boys. Additionally, little is known about the genetic architecture underlying extreme pubertal delay, which represents the tail end of the population distribution and is the most common cause for referral to pediatric specialists. We hypothesized that many genetic variants, both common and more rare, influence pubertal timing. Thus, the general aim of this thesis is to address these knowledge gaps by identifying common genetic variants influencing pubertal growth and timing in the general population and less common variation influencing pubertal delay, in both sexes.

Specifically, this work aimed to:

1. Identify common genetic variants influencing the timing of the pubertal growth spurt and sexual maturation in Finnish and European-descent populations,

2. Characterize the genetic relationship between the discovered variants and related puberty and anthropometric traits to place them into a larger developmental context, and

3. Identify a susceptibility locus predisposing to pubertal delay in families segregating constitutional delay of growth and puberty.

4 Study design, subjects, and methods

4.1 Study design

To address Aim 1, we used a genome-wide association approach for locus discovery. We investigated several different models of the pubertal growth spurt in Finns (Study I) and Europeans (Study II), as well as the timing of sexual maturation using Tanner genital and breast development scale data in subjects of European descent (Study III). First, we looked at late pubertal growth in a GWA study including more than 4,300 Finnish population samples (Study I) for locus discovery and a further three Finnish population cohorts for replication and follow-up. In Study II, we expanded in scope to include three models of the pubertal growth spurt, including the take-off phase, late pubertal growth, and the total magnitude of the growth spurt in up to 18,700 samples of European descent. In Study III, we investigated Tanner breast and genital stage data to detect variants influencing the timing of pubertal onset in more than 11,200 European samples.

For Aim 2, we characterized the effects of the discovered variants on related traits using regression models to study their influence on the timing of menarche (Studies I-III), postnatal growth (Studies I and II), pubertal growth (Study III), adult stature (Studies I-III), and body mass index (Studies II and III).

For Aim 3, we sequenced the pericentromeric region of chromosome 2 in 13 families segregating constitutional delay of growth and puberty contributing the most to a previously published, robust linkage signal (Study IV). In this study, we looked for shared low-frequency variants in genes and regulatory regions surrounding the best candidate genes from 79-124 Mb along chromosome 2. Segments of interest that passed filtering criteria were sequenced in an additional 135 unrelated affected probands.

Table 2. *Summary of the four studies included in this thesis.*

Study	Primary phenotype	Description	Sex	Population	Sample size
I	Late pubertal growth	Difference in height SDS 14 yrs - adult	males & females	Finnish	Up to 11,573
II	Pubertal growth	Height SDS at age 10	Females	European	6,879
		Height SDS at age 12	males	European	7,161
		Difference in height SDS 8 yrs - adult	males & females	European	10,799
		Difference in height SDS 14 yrs - adult	males & females	European	9,228
III	Tanner puberty stage	Genital stage	males	European	3,769
		Breast stage	females	European	6,147
IV	Constutitional delay of growth and puberty	Normal puberty ≥2 SD from the mean	males & females	Finnish	13 families

4.2 Phenotypes and study subjects

Studies I and II: Pubertal growth

In Study I, the main phenotype investigated was the difference in height standard deviation (SD) score between age 14 and adult, approximating the timing of the pubertal growth spurt **(Table 2)**. This study included subjects from birth cohorts collected from the Finnish population. The primary discovery set consisted of samples from the Northern Finland Birth Cohort 1966 (NFBC1966; maximum n = 5,087). Replication and follow-up utilized samples from the Helsinki Birth Cohort Study (HBCS; n = 1,718), the Cardiovascular Risk in Young Finns study (YFS; n = 2,639) and Health 2000 (H2000; n = 2,129) **(Table 3)**.

Study II expanded upon the phenotype utilized in Study I. Because there is a large amount of variation in the timing and tempo of the pubertal growth spurt, an accurate model requires very frequent height measurements spanning a large age range. However, these are quite difficult to obtain in cohort studies. Also, girls enter puberty an average of 2 years prior to boys, and thus begin their growth spurt earlier than males. In this study, we took these challenges into consideration and aimed to characterize pubertal growth loci by leveraging heterogenous height data taken at various ages during adolesence in the participating cohorts to maximize power. Thus, we modelled the growth spurt using three partly correlated measures of pubertal growth **(Figure 7)** that partly reflect pubertal timing (44). These were:

1. A single height measurement ("Analysis I") which approximated the timing of the take-off phase of the growth spurt: girls with height measurements available at age 10 (±1 year) and boys with height measured at age 12 (±1 year). Sex-specific SD scores were calculated within each study by dividing the within-population height mean by the SD. 14,040 samples (7,161 males and 6,879 females) from 9 European-descent cohorts were included.

2. The total amount of pubertal growth ("Analysis II"): we included individuals with a childhood height measurement at age 8 (±1 year) and at adulthood (≥18 years old). The height difference was calculated between the two measurements and sex-specific SD scores of this difference were calculated within each study by dividing the within-population mean difference in height by the SD. 6 cohorts including as many as 10,799 samples (5,043 males and 5,756 females) contributed to this analysis.

3. The amount of late pubertal growth ("Analysis III"), approximating the timing of the pubertal growth spurt: similar as in Study I, the height difference was calculated between height at age 14 (±1 year) and at adulthood (≥18 years old) and logarithm-

transformed sex-specific Z-scores were calculated within each study. Log transformation was performed prior to SD score calculation, again by dividing the population-specific mean log-transformed height difference by the SD. 5 cohorts with up to 9,228 subjects (4,282 males and 4,946 females) were included.

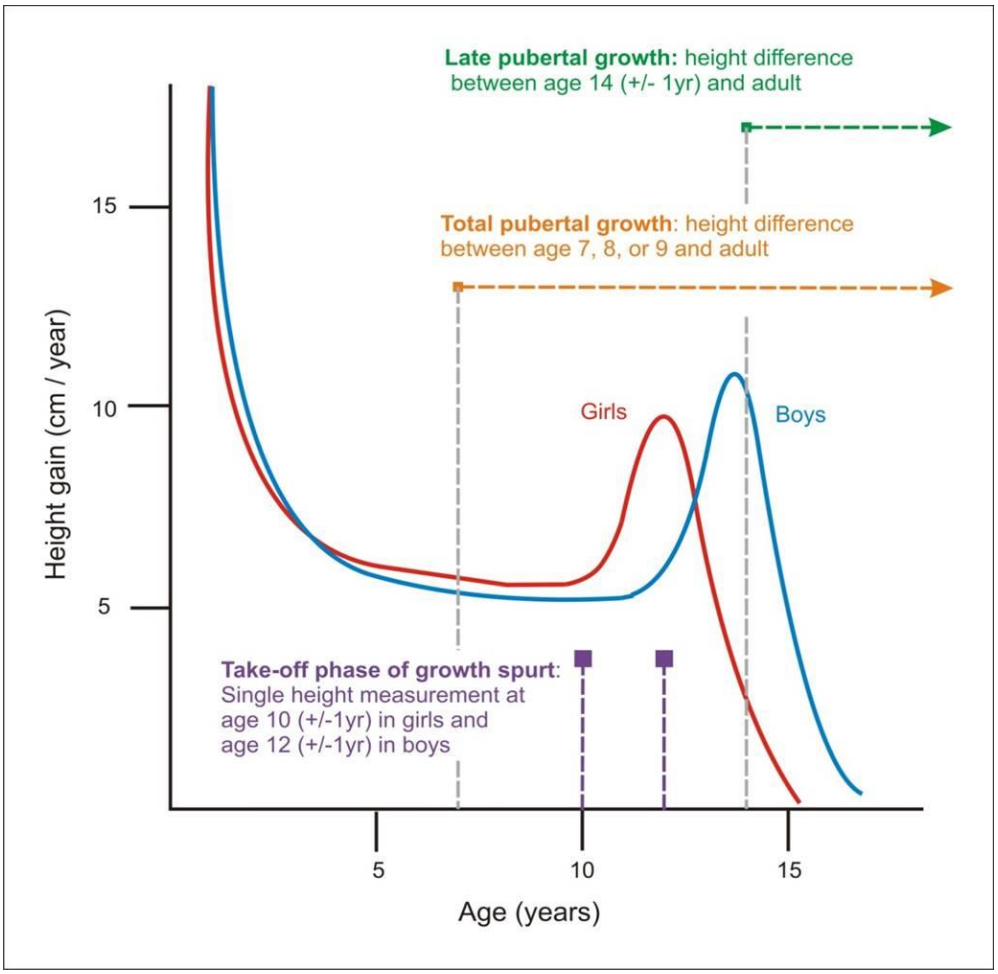


Figure 7. *An overview of the three measures of the pubertal growth spurt utilized in Study II.*

Cohorts contributing to Study II were the Avon Longitudinal Study of Parents and Children (ALSPAC), 1958 British Birth Cohort (BC58-T1DGC and BC58-WTCCC), YFS, HBCS, Lifestyle- Immune System- Allergy Plus Environment and Genetics Study (LISAplus), NFBC1966, Queensland Institute of Medical Research (QIMR), and Western Australia Pregnancy Study (RAINE) **(Table 3)**. Follow-up replication or genotyping for Study II was performed in the ALSPAC follow-up sample, Children's Hospital of Philadelphia (CHOP), Finnish Twin Cohort Study (FTCS), Genome-Wide Population-Based Association Study of Extremely Overweight Young Adults (GOYA), and Lifestyle-Immune System- Allergy study & German Infant Study on the influence of Nutrition Intervention plus environment and genetics (LISAplus & GINIplus; follow-up sample).

Analysis of early height was performed for select markers in CHOP, Copenhagen Study on Asthma in Childhood (COPSAC), Generation R Study (Generation R), HBCS, INMA, LISAplus & GINIplus, NTR, NFBC1966, Prevention and Incidence of Asthma and Mite Allergy birth cohort study (PIAMA), and RAINE.

Study III: Sexual development

In Study III, we focused on Tanner pubertal stage data (male genital stage and female breast stage) as our primary phenotypes. Tanner breast stage or genital stage was treated as a quantitative trait on a scale of 1-5. We chose separate age ranges for boys and girls at which the Tanner scale data were approximately normally distributed: females with a Tanner breast stage assessment within 10.5 - 12.5 years of age were included in the female discovery set (n = 6,147), and males with a Tanner genital stage assessment between 12.6 - 15 years of age were included in the male discovery analysis (n = 3,769).

In Study III, ALSPAC, BC58-T1DGC and BC58-WTCCC, YFS, Netherlands Twin Registry (NTR), TEENs of Attica: Genes and Environment (TEENAGE) study, and RAINE were included as discovery cohorts **(Table 3)**. Follow-up was performed in Infancia y Medio Ambiente (INMA), Special Turku Coronary Risk Factor Intervention Project (STRIP), and the Leipzig Childhood Cohort (LEIPZIG).

In some cohorts, Tanner stage was assessed by a clinician or trained researcher, and in others, Tanner stage was self-assessed using schematic drawings. Tanner stage was measured by self-assessment in ALSPAC, RAINE, and TEENAGE. Self-assessment or maternal assessment was performed in the follow-up cohort INMA. Studies in which Tanner stage was assessed by a medical professional were B58C-WTCCC, B58C-T1DGC, YFS, LEIPZIG and STRIP. In NTR, data collection was performed in several subprojects.

Study IV: Constitutional delay of growth and puberty

In Study IV, we investigated the region of chromosome 2 previously robustly linked with constitutional delay of growth and puberty in a set of Finnish families. Probands were drawn from patients referred to pediatric specialists due to delayed puberty who were identified from hospital discharge records from Helsinki, Kuopio, Tampere, and Turku University Hospitals as well as two municipal hospitals in southern Finland. Patients were included if they fulfilled diagnostic criteria for CDGP, which was defined as pubertal development 2 SD later than the population mean (specifically, older than 13.5 years in boys for Tanner genital stage II and later than 13.0 years in girls for Tanner breast stage II), and if other causes of pubertal delay (for example chronic illness or HH) could be excluded. Growth charts were also examined, and patients were included if their pubertal growth spurt was 2 SD later than the population average (age at takeoff of pubertal growth

Table 3. *Summary of the cohorts that contributed data to Studies I- III.*

Cohort abbreviation	Cohort Name	Ethnicity	Maximum N (rounded)	Study
NFBC1966	Northern Finland Birth Cohort 1966	Finnish	4300	I, II
HBCS	Helsinki Birth Cohort Study	Finnish	1700	I, II
YFS	Cardiovascular Risk in Young Finns Study	Finnish	2600	I, II, III
H2000	Health2000	Finnish	2100	I
ALSPAC	Avon Longitudinal Study of Parents and Children	British	6000	II, III
BC58-T1DGC	1958 British Birth Cohort Study	British	2000	II, III
BC58-WTCCC	1958 British Birth Cohort Study	British	2000	II, III
LISAplus	Lifestyle, Immune System, Allery Plus Environment and Genetics Study	German	440	II
QIMR	Queensland Institute of Medical Research	European	1100	II
RAINE	Western Australia Pregnancy Study	European	760	II, III
CHOP	Children's Hospital of Philadelphia	European	1000	II
FTCS	Finnish Twin Cohort Study	Finnish	260	II
GOYA	Genome-wide Population-Based Association Study of Extremely Overweight Young Adults	Danish	180	II
GINIplus	German Infant Study on the influence of Nutrition Intervention plus Environment and Genetics	German		II
NTR	Netherlands Twin Registry	Dutch	600	III
TEENAGE	TEENs of Attica: Genes and Environment	Greek	300	III
INMA	Infancia y Medio Ambiente	Spanish	850	III
STRIP	Special Turku Intervention Project	Finnish	400	III
LEIPZIG	Leipzig Childhood Cohort	German	2200	III
NFBC8586	Northern Finland Birth Cohort 1985-86	Finnish	4300	II, III
DILGOM	Dietary, Lifestyle, and Genetic determinants of Obesity and Metabolic syndrome	Finnish	513	II, III
COPSAC	Copenhagen Study on Asthma in Childhood	Danish	360	II
Generation R	Generation R Study	Dutch, European	2000	II
PIAMA	Prevention and Incidence of Asthma and Mite Allergy birth cohort study	Dutch	350	II

later than 13.8 years in boys and 12.2 years in girls, and age at PHV later than 15.6 years in boys and 13.7 years in girls) (2).

Invitation letters were sent to the family members of probands, and respondants were given structured interviews to determine their eligibility. Archived height measurements were used to draw growth charts, where available, and the timing of the pubertal growth spurt was used to determine the timing of puberty. Family members were considered to have CDGP if the age at take-off or PHV of the pubertal growth spurt occurred 1.5 SD or later than the population mean, or if they attained adult height older than 18 years for males or 16 years for females (144).

We included the probands and parents from the 13 families that contributed most to the previously published linkage signal based on family-specific LOD scores in the linkage region on chromosome 2p13-2q13 (144) and investigated leading candidate regions in an additional set of 135 unrelated probands **(Table 4)**. All participants and/or their guardian gave written informed consent. The study design and sample collection were approved by the Ethics Committee for Pediatrics, Adolescent Medicine, and Psychiatry of the Hospital District of Helsinki and Uusimaa.

We also utilized several sets of control samples in Study IV. The first was a set of whole-genome genotyped samples from the neurological services at the Kuopio University Hospital and the Helsinki University Hospital that were originally collected as controls for a study on intracranial aneurysm (IA controls; n = 730). These controls were collected from anonymous Finnish patients who gave blood for causes unrelated to IA. Genomic DNA was extracted from these blood samples, and genotypes were ascertained by the Institute for Molecular Medicine Technology Centre (Helsinki, Finland) using Illumina HumanCNV370-duo chips (141).

Additionally, we ascertained the population frequency of sequenced variants using the complete 1000 Genomes (1000G) reference set, the Finnish population-specific data from the 1000G reference (FIN-1000G) (www.1000genomes.org/), and a unique resource of Finnish exome sequences (n = 3,325) from the SISu project (http://www.sisuproject.fi/), which are both publicly available. In the SISu project, whole-exome sequencing was done at the Wellcome Trust Sanger Institute using the Illumina platform. Variant recalling was performed jointly with ~23,000 exomes from other studies at The Broad Institute. Variants failing the genotype calling pipeline quality control and variants that were not within the bait regions of the exome capture kit were removed. Sample-wise quality control was performed after the removal of poor quality SNPs and individual genotypes, and variants with a call rate < 90 % over all samples were removed.

Table 4. *Summary of samples included in Study IV.*

Sample set	**Abbreviation**	**Ethnicity**	**N**
Constitutional delay of growth and puberty	CDGP	Finnish	13 probands; 12 affected parents; 135 unrelated affected patients
Intracranial aneurysm controls	IA controls	Finnish	730
1000 Genomes controls	1000G	European, African, Asian	~1000
1000 Genomes Finnish controls	FIN-1000G	Finnish	~100
Sequencing Initiative Suomi controls	SISu	Finnish	2028 - 3325

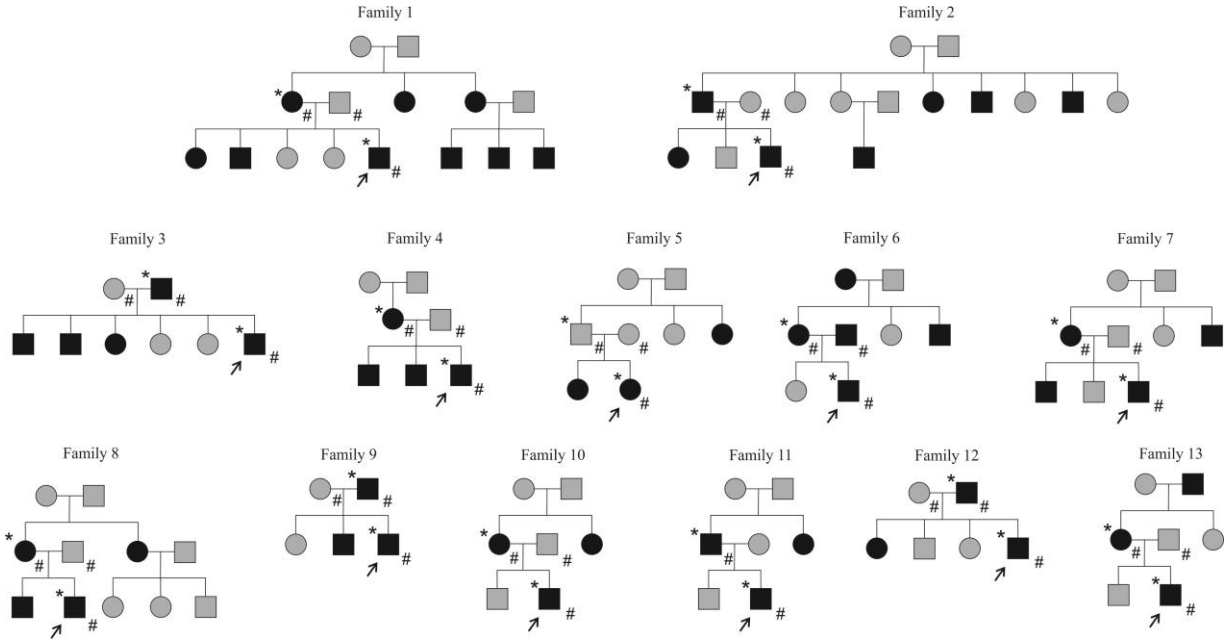


Figure 8. *Pedigrees of the top 13 CDGP families contributing most to the previously published linkage signal. The arrow refers to the proband in each family. Filled circles and squares are classified as affected, while the pubertal status of grey individuals is unknown. #, genotyped; *, sequenced in the present study.*

4.3 Genotyping, imputation and quality control

The same general genotyping and quality control protocol was utilized for Studies I-III according to standard methods. Genome-wide genotypes were obtained by individual cohorts using high-density SNP arrays on Illumina and Affymetrix platforms. Imputation was not performed for genotypes included in Study I, but was done for samples included in Studies II and III. Prior to imputation, SNPs with a minor allele frequency (MAF) < 1%, call rate < 95%, and Hardy-Weinberg frequency of $P < 1 \times 10^{-6}$ were excluded. Samples were also excluded if they contained duplicates, excess heterozygosity, non-European ancestry, or ambiguous sex. The MACH (159), IMPUTE (160), or BEAGLE (161) programs were used to impute approximately 2.5 million SNPs against HapMap Phase II (release 21/22). Imputed SNPs were filtered before meta-analysis to exclude poorly imputed SNPs (IMPUTE filter PROPER INFO < 0.4, MACH filter $r^2 < 0.3$, or Plink INFO < 0.3).

In Study IV, genotyping of the proband and both parents (trios) of the top 13 families that contributed the most to the previously published linkage peak was performed on the Illumina Infinium platform at the FIMM Technology Centre. Eight samples (the two top trios plus two probands from additional families) were genotyped using the HD

HumanCNV370-Quad Beadchip, and thirty samples (eight trios, the parents of the probands that were genotyped on the HumanCNV370 chip, plus one additional proband and their affected parent) were genotyped using the Human610-Quad BeadChip. The BeadStudio Genotyping Module v3.3.7 was used to call genotypes, which were then reviewed manually. SNPs with a success rate of < 95% were excluded. Genotypes were retrieved for markers spanning all autosomes and the X-chromosome. The three batches of genotyping had success rates of 0.96, 0.93, and 0.97, respectively.

4.4 Genome-wide analyses

Studies I-III: Genome-wide association

In Study I, we performed GWA analysis of growth in height during late adolescence (the difference in height SDS between age 14 and adult) in NFBC1966 (n = 2,073 males and 2,248 females). The phenotype was log-transformed prior to the calculation of sex-specific Z-scores. Linear regression was used in PLINK for all autosomes, assuming additive inheritance and including the first two principal components (PCs) as covariates, to account for population substructure. Males and females were meta-analyzed using a fixed-effect model using the R package (www.r-project.org/) MetABEL.

In Studies II and III, we performed genome-wide association analysis and meta-analysed the results across participating cohort studies. Each cohort performed genome-wide association analysis individually by linear regression using an additive model across genotyped and imputed SNPs, separately for males and females. In Study II, age at the adolescent measurement was included as a covariate and the first two PCs were adjusted within each study sample if necessary. In Study III, covariates included the first PCs if necessary and age at measurement (in years, to the nearest month, if available). Programs used to perform the GWA analyses in each cohort were MACH2QTL (159), QUICKTEST (162), SnpTest (163), Plink (164), or ProbABEL (160).

For Study II, a fixed effects inverse-variance meta-analysis model was used to test the effect of each variant on early pubertal height (Analysis I), total growth during puberty (Analysis II), or late pubertal growth (Analysis III) separately for males and females. Sex-specific results from each study were then meta-analyzed for each phenotype in three combined-gender analyses. MetABEL (160) was used for all meta-analyses. MetABEL corrects each individual result for its respective genomic inflation factor (λ) according to the genomic control method for population stratification, and an additional genomic control correction was applied using the overall genomic inflation factor calculated for each of the nine meta-analyzed results. Genome-wide significance was set at a Bonferroni-corrected threshold of $P < 1.67 \times 10^{-8}$, to account for testing three primary phenotypes.

In Study III, prior to meta-analysis, cohort-specific results were filtered to exclude variants with a MAF < 0.03. Meta-analysis of individual cohort results was performed with GWAMA (165) (http://www.well.ox.ac.uk/gwama/index.shtml) for quantitative traits with two genomic control corrections enabled (on individual study results and then on the meta-analysis summary output). Gender-differentiated and gender-heterogeneity analysis options were also enabled.

Study IV: Genome-wide characteristics of CDGP families

Several genome-wide analyses were done in Study IV to explore the general amount of relatedness among the leading 13 CDGP probands. To compare their genetic variance against the IA control set, a population-based sample of random individuals from roughly the same geographical region in Finland, we calculated genome-wide identity-by-descent (IBD) among the 13 probands and the pairwise inbreeding coefficient (F) using PLINK. This analysis was done to confirm that there were no undetected close relationships between the probands' parents.

To visualize the genomic variance between the CDGP probands on a background of IA control subjects, the proportion of alleles shared identity-by-state (IBS) between all pairs of individuals was computed, and the first three PCs were plotted in multidimensional scaling (MDS) plots using R.

4.5 Follow-up analyses of suggestive association signals

Follow-up association analyses

In Study I, we tested the best-associated SNP (rs7759938) and a proxy (rs314268) for association with pubertal growth (height SDS difference between age 15 and adult) in YFS (n = 1,241). We also tested amount of total phenotypic variance explained by rs7759938 by estimation of the regression coefficient (r^2) in NFBC1966 with the normalized phenotype as the outcome in an additive regression model adjusting for the first 2 dimensions obtained from MDS analysis.

In Study II, we followed up 22 signals from Analysis I (the single height measurement at age 10 in girls and 12 in boys) which associated with a *P*-value between 1×10^{-5} and 1.67×10^{-8} and were not previously associated with related traits (adult stature, AAM, or BMI). In-silico analysis was done by ALSPAC (follow-up sample), CHOP, FTCS, and GINIplus (follow-up sample). Association results for a marker showing borderline significance, rs281379, were also provided by NTR. De novo genotyping was done for selected markers (success rate > 98%) from Northern Finland Birth Cohort 85-86 (NFBC8586) with TaqMan Pre-Designed SNP Genotyping Assays on the LightCycler 480 Real-Time PCR

System (Roche) according to the manufacturer's instructions at the FIMM Technology Centre. Statistical analysis in the replication samples was performed similarly as in the discovery analyses, using PLINK (164), ProbABEL (160), or SNPtest (163), by linear regression for each of the 22 markers under an additive model, with age at adolescent measurement and correction for population substructure as optional covariates (first two PCs). Genomic control-corrected discovery results were meta-analyzed together with the individual linear regression results from contributing cohorts for each SNP, using MetABEL.

For Study III, markers with $P < 5x10^{-7}$ in the discovery analyses were followed up: rs246185 (*MKL2*) suggestively associated with male genital stage, and rs1149336 and rs1129332 (*CAMTA1*) suggestively associated in the male genital and female breast combined analysis. Genotyping of STRIP samples was performed with TaqMan Pre-Designed SNP Genotyping Assays on the LightCycler 480 Real-Time PCR System (Roche) at the FIMM Technology Centre with a success rate of 0.94, 0.87, and 0.84, respectively. LEIPZIG sample genotyping was performed using the TaqMan Pre-Designed SNP Genotyping Assay run on the Applied Biosystem 7500 Real Time PCR System, applying the autocaller algorithm of 7500 Software V.2.0.3 with calling rates of 0.98, 0.98, and 0.93, respectively. Association analyses of genotype with Tanner stage were performed by each contributing cohort, and meta-analysed together with the primary discovery summary statistics using linear regression in PLINK (164).

In Study III, we also repeated the female Tanner breast stage analysis with a BMI cutoff to exclude the possibility of obesity interfering with breast stage measurements. Females in the lower 80^{th} percentile of BMI were included from the contemporary cohorts ALSPAC, YFS, and NTR (n = 2,637). The top SNPs from the *LIN28B* locus were then extracted and compared to the original female meta-analysis results.

Conditional analysis

In Study II, several of our leading signals overlapped previously published adult height, AAM, or BMI loci. To determine whether our leading signals represented independent effects on growth during puberty, we performed conditional analysis by linear regression using an additive model on the primary pubertal growth phenotypes, adjusting each marker (imputed genotype dose) for the previously reported marker. Age at adolescent measurement (where available) and optional adjustment for population substructure were included as covariates.

CNV analysis of 16p11.2

The leading novel signal (rs4788196 at *MAPK3*) in Study II fell in a region of the genome where rare recurrent CNVs have been associated with early onset obesity (166, 167).

Thus, we explored whether CNVs could explain the association signal between growth during puberty and the leading variant, since advanced pubertal timing has also been correlated with increased pre-pubertal body mass. CNVs were genotyped using signal intensity distributions and B-allele frequencies of the genotyping probes with PennCNV software (168) and adjusted for genomic waves according to genomic GC content (169). The CNV scan was performed for 2,310 individuals in YFS and 4,931 subjects in NFBC1966 (170).

1000 Genomes imputation

In Study III, the leading signal associated with male genital stage (rs246185 at *MKL2*) also fell on chromosome 16. To fine map this locus in an attempt to further refine the association signal, we imputed against 1000 Genomes data in ALSPAC, YFS, RAINE, and TEENAGE. Imputation was performed with IMPUTE (171), MaCH (159) or Minimac (172). Association analysis was then performed in each cohort using SNPtest V2 (171), MACH2QTL (159), or QUICKTEST (162). After cohort-specific association, the results were filtered based on imputation quality (PROPER INFO <0.4, r2<0.3, or INFO < 0.3), and meta-analysis was performed using GWAMA (165). The 1 Mb region surrounding rs246185 was extracted from the complete chromosome 16 results.

Functional investigation of variants in LD

Also in Study III, we investigated variants that were in high linkage disequilibrium (LD) with rs246185 for evidence of involvement in transcription factor (TF) binding sites. SNPs nearby rs246185 were extracted from the 1000 Genomes Pilot I dataset in the online SNAP proxy search tool (http://www.broadinstitute.org/mpg/snap/ldsearch.php) with an r^2 threshold of 0.6. The resulting 12 SNPs (plus rs246185) were then entered into RegulomeDB (173) (http://www.regulome.stanford.edu/index), which annotates variants with known and predicted regulatory elements, including regions of DNAase hypersensitivity, transcription factor binding sites, and biochemically characterized promoter regions that affect transcription, in intergenic genomic regions. These data are sourced from public datasets.

Expression quantitative trait loci (eQTLs)

Expression quantitative trait loci (eQTLs) were investigated in Studies II-III to attempt to link the leading GWA signals to nearby gene expression levels. For Study II, we queried significant SNPs from the Dietary, Lifestyle, and Genetic determinants of Obesity and Metabolic syndrome (DILGOM) study (174). In this study, whole blood was extracted from 518 unrelated individuals and genotyped on the Ilumina 610-Quad SNP array. In parallel, mRNA expression was quantified with Illumina HumanHT-12 Expression

BeadChips, and the association between transcript expression levels and each SNP was tested by linear regression.

In Study III, whole blood eQTLs were also queried from the DILGOM study (174), NTR, and the Netherlands Study of Depression and Anxiety (175, 176). GENEVAR (177) (GENe Expression VARiation; http://www.sanger.ac.uk/resources/software/genevar/) was used to look up data on rs246185 from lymphoblastoid cell lines (178, 179), skin, and adipose tissue (179).

4.6 Pathway analyses

We performed pathway analyses in Studies II-III to look for biological pathways common to genes nearby the leading associated signals. For Study II, we entered the nearest gene to all signals at $P < 1x10^{-4}$ (one per locus) into the g:Profiler Gene Group Functional Profiler tool (g:GOSt) (180). This online tool queries databases of biological pathways for enrichment of user-entered genes. For the single height analysis (Analysis I), we also entered *MAPK3* because the gene was implicated by eQTL evidence to be functionally relevant, which brought the total of signals queried to 93, corresponding to 0.0003% of all discovery signals for Analysis I. We additionally ran gene set enrichment (GSE) analysis using MAGENTA (181), which calculates a *P*-value for each gene in the genome based on GWA results, and then searches biological databases for pathways showing an enrichment of genes with lower than expected *P*-values. There were no significant findings for Analyses II and III.

In Study III, GSE analysis was performed using MAGENTA (181) on the results for males and females separately and combined. For both Studies II and III, default values were used for the number of gene set permutations, for the upstream and lower boundaries surrounding each gene, and for the gene set size limits. GSE analyses were run for both 75^{th} and 95^{th} percentiles, and genes in the HLA region were removed prior to analysis. Pathways were reported that had a false discovery rate (FDR) of ≤ 0.05 and a nominal GSE $P < 1x10^{-3}$.

4.7 Association analyses of AAM, height, and BMI

Pubertal timing (timing of the growth spurt and menarche)

In Studies I-II, our primary phenotypes were not directly pubertal traits, but of measures of the growth spurt that occurs during puberty, which reflects both pubertal timing and genetic height growth potential. Thus, it was important to assess the relevance of pubertal growth-associated variants for their effect on more direct assessments of pubertal timing.

In Study I, we first investigated whether markers at the *LIN28B* locus associated with late pubertal growth reflected an effect on the timing of the pubertal growth spurt. To do this, we looked at the association between the leading markers and growth during early and mid-puberty (9-12 years and 12-14/15 years in YFS and NFBC1966, and 9-12 years in HBCS). Additionally, we investigated the association between rs7759968 and AAM in the Finnish datasets (n = 4,379 Finns). Finally, we performed multiple regression analysis including rs7759938 and height growth during early puberty (9-12 years old) to determine whether the pubertal height growth and menarche effects were independent.

In Study II, the leading signals not previously implicated in the timing of menarche were queried from in silico meta-analysis data of 87,802 women published by the ReproGen Consortium (150).

Additionally, in Study III, the effects of the male genital stage-associated marker rs246185 on female puberty and pubertal growth were estimated by extracting this variant from the female Tanner GWA analysis results as well as from previously published GWA studies on age at menarche by the ReproGen Consortium (150) and three pubertal height growth phenotypes (182).

Height and BMI

In Study I, an association to final height was previously published at two *LIN28B* markers only partly correlated with each other (rs314277 and rs314268, r^2 = 0.26). rs314268 overlaps with a pubertal timing effect (r^2 = 0.94 with rs7759938), thus potentially mediating an adult height effect through the timing of the pubertal growth spurt. In the Finnish cohorts, we tested for two separate height effects by considering the effect of rs7759938 and rs314277 separately and then by including both markers into a regression model against final adult stature. Finally, we performed multiple regression analysis including both SNPs, sex, and all possible interactions into the model. We then plotted the marker-specific beta and SE from the multiple regression analyses of height across the complete growth trajectory from infancy to adulthood, including both markers in the models (standardized height at 1-year intervals between ages 1-12, standardized height at age 14, and adult stature in NFBC1966 and HBCS) to visualize the effects of these markers across postnatal growth.

Similarly, in Study II, eight of the leading loci were previously identified as associated with adult stature (29) and one was previously associated with BMI (183, 184). To characterize their effects longitudinally across postnatal life, we divided height or BMI measurements from childhood to adulthood into 6 age bins: 1) Pre-puberty (6.5 - 8.5 years old), 2) early puberty (8.6 - 10.5 yrs old), 3) mid-puberty for females (10.6 - 12.5 yrs old), 4) mid-puberty for males (12.6 - 14.5 yrs old), 5) late puberty (14.6 - 17.5 yrs old) and finally 6) adult (> 17.6 yrs old). In each contributing cohort, each marker of interest

(imputed genotype dosage) was tested for association with height or BMI SDS for males and females separately for all age bins available, using linear regression with an additive model and adjustment for exact age at measurement (to the nearest month), along with correction for population stratification if needed. One measurement was included per subject per bin, with the age closest to the mean used when more than one measurement was available. Summary statistics were meta-analyzed like the primary analyses in each age bin, separately for males and females, for both height and BMI distinctly. Effect sizes were plotted against age to visualize their effects longitudinally across postnatal growth for the 10 significantly associated pubertal growth SNPs.

Additionally, in Study II we investigated the association with early growth for five menarche-associated pubertal growth variants. Cohorts that contributed had height measurements available at 1, 2, 3, or 4 years old. Length was measured at 12 months (range 6-18 months) and height was measured at 24 (range 18-30), 36 (range 30-42) and 48 (range 42-54) months. When multiple measurements per individual were available, those closest to 12, 24, 36 or 48 months were used. Sex- and age-adjusted SD scores were calculated in each study using Growth Analyser 3.0 (Dutch Growth Research Foundation, Rotterdam, the Netherlands) (185). The association between each marker genotype and length or height SDS was assessed using linear regression for males and females separately, assuming an additive model. Imputed genotypes were used where directly-assayed genotypes were unavailable, and meta-analysis of the within-cohort results was done using the inverse-variance method. A fixed-effects model was assumed, using RMeta in R.

For Study III, the leading male genital signal rs246185 was queried from the previously published GIANT-consortium study on adult stature (29).

Overlap of published menarche and BMI SNPs with Tanner staging

In Study III, we looked at the association of published menarche and BMI loci in our Tanner pubertal staging data. 32 previously published menarche loci, 10 possible menarche loci (150), and 2 pubertal growth loci with evidence for menarche-association from Study II were extracted from the Tanner GWA discovery analysis results sets, for both males and females separately and combined. Also, we extracted 31 SNPs associated with adult (183) and/or childhood (184) BMI from the three Tanner analysis results. We then plotted the menarche effect size against the Tanner effect size for the puberty-advancing allele for males and females both separately and combined. Similarly, we plotted the BMI effect size against the Tanner effect size for the BMI-increasing allele for males and females both separately and combined. Finally, we calculated the correlation between the menarche and Tanner effect sizes, and between the BMI and Tanner effect sizes for the loci in each plot, for both males and females separately and combined.

4.8 Sequencing (Study IV)

In Study IV, we performed targeted sequencing of the pericentromeric region (79,171,971 – 124,250,162 Mb; genome build 37) of chromosome 2 in 13 CDGP probands and their suspected affected parent (n = 26). Sequencing was done on the Illumina HiSeq2000 platform with 100bp paired-end reads. Variant calling, sequencing alignment, SNP calling, and indel calling were done using a variant calling pipeline (VCP) that was developed at the FIMM Technology Centre for quality control, short read alignment, and variant calling and annotation (186). The sequencing and bioinformatics were performed at the FIMM Technology Centre.

Follow-up sequencing in additional CDGP cases

We assessed the presence of low-frequency variants in two sets of three consecutive *DNAH6* exons (23, 24, 25, 46, 47 and 48) in probands with CDGP (n = 135) by Sanger sequencing. The samples were amplified using Thermo Scientific DreamTaq Green PCR Master Mix (2X) (Thermo Fisher Scientific Inc, Waltham, MA) according to the manufacturer's instructions. Primers were designed with OligoArchitect™ Online v4.0 (Sigma Aldrich, St Louis, MO), and the corresponding oligonucleotides were ordered from Sigma Aldrich. The primer sequences were as follows (5' to 3'): 23F: CCATGACCAGTATAATTG, 23R: TATGCTTAGAGTGAGAAT, 24F: ATAGTGGAATGTCAATAG, 24R: ATGTTTCTTAAATATGTGAT, 25F: GTAACTCACACTCACATT, 25R: TGTCAGAGCATTAGAATT, 46F: TTGCTATGTTAGAACTTC, 46R: AATACAAAGGAAACCAAT, 47F: TATCTACTATGCTGACAT, 47R: TCTCTATATGAATAAATTCCT, 48F: TTATTGAAATGACACAAC, 48R: GAGAATGGACTAATACAG. We used the following cycling conditions: 95°C for 1 min, 35x (95°C for 30s, 52°C for 25s, 72°C for 30s), and 72°C for 10min. The samples were purified with ExoSAP-IT® (Affymetrix, Santa Clara, CA) according to the manufacturer's instructions, and were then capillary-sequenced by the Genome Analyzer II (Illumina, San Diego, CA) platform. Sequence analysis was performed with novoSNP3.0.1 (187). We used the UCSC genome browser (hg19) (188) as a reference sequence.

4.9 Variant filtering and analysis (Study IV)

We only considered variants falling within the bait region of 79-124 Mb. We prioritized variants with predicted consequences that affect protein structure (nonsynonymous, splice site, frameshift or premature stop codons) or fell in the untranslated regions upstream or downstream of the gene (5' or 3' UTRs) or were intronic. Variants in family 1 were often prioritized since this is a large pedigree consisting of 15 individuals, 9 of whom were affected, that contributed the most to the linkage finding (144). By definition, CDGP is

present in around 2% of the general population in whom puberty begins 2 SD later than the population mean, so the predisposing variant or variants are expected to be present in the Finnish population at a low frequency. Additionally, since the penetrance of the causative allele(s) is unknown, we chose to include variants with a lenient allele frequency threshold of 10% or less.

First, we examined whether family 1 shared identical variants or different variants in the same gene with known 1000G frequencies of <10% with any of the other families. Only genes with variants in 5 or more families were considered. Second, we repeated this analysis for variants of unknown frequency. Without frequency information to use as a filter, we prioritized variants found in family 1. Last, we considered genic variants transmitted from either parent because bilineal inheritance could not be ruled out completely. To do this, we extracted variants transmitted from either parent at < 10% in 1000G and looked for identical variants or different variants in the same gene that were shared by family 1 and at least 4 others.

Variants of interest were followed up in several populations to confirm that their frequencies are less than 10%. We looked at American of European descent (Exome Variant Server; http://evs.gs.washington.edu/EVS/) and the Finnish (FIN-1000G and SISU) population to confirm their rarity. To investigate the potential deleteriousness of these variants, we examined SIFT (189) and PolyPhen (Polymorphism Phenotyping; http://genetics.bwh.harvard.edu/pph2/) (190) scores for each variant using the Variant Effect Predictor (191) (http://useast.ensembl.org/info/docs/tools/vep/index.html). SIFT and PolyPhen predict whether a codon which alters the amino acid affects protein structure and function. We also examined each variant's score in RegulomeDB (http://regulome.stanford.edu/index), which identifies DNA features and regulatory elements in intergenic regions and assigns each variant a score which reflects its predicted ability to disrupt transcription factor binding.

Candidate gene analysis

A change in important regulatory element sequences affecting gene expression might be another way that a variant could impart a delay on pubertal onset. Thus, we created a list of plausible candidate genes within the linked region of 79-124 Mb on chromosome 2 using Ensemble Biomart (http://www.ensembl.org/biomart/martview/). We focused on genes annotated with appropriate gene ontology (GO) terms relating to puberty, sexual maturation, and growth **(Table 5)**. We also manually searched each gene on the list from AceView (http://www.ncbi.nlm.nih.gov/IEB/Research/Acembly/) and PubMed (http://www.ncbi.nlm.nih.gov/pubmed) for additional evidence of involvement in puberty. This resulted in a short list of the best functional candidates that had evidence of involvement in both reproduction and growth or are directly involved in the HPG axis. The genes on this short list were *GLI2, PAX8, INHBB, KDM3A, NPHP1, BCL2L11, MERTK,* and *IL1B*.

Next, we looked at the possible regulatory regions surrounding these 8 best functional candidate genes. To do this, we looked at transmitted variants within 400kb around these genes (200kb from each gene's start and terminus). We used RegulomeDB to see if shared variants in these regions potentially affect protein regulation by altering transcription factor binding sites.

AAM-associated region near rs6758290

A recent genome-wide association study pinpointed a variant associated with normal variation in age at menarche (rs6758290 at 105.8 Mb) (148), which falls in the region of linkage on chromosome 2. We investigated whether low-frequency variants in LD with rs6758290 show evidence for involvement in pubertal delay in our CDGP families. The LD block ($r2 \geq 0.1$) surrounding this marker extends approximately 500 kb upstream and 300 kb downstream from rs6758290 in European populations, so we extracted all variants from 105,364,800 – 106,164,800 (GRCh37) along chromosome 2 and looked for transmitted shared variants in our families. We also looked at the RegulomeDB score for all shared non-exonic variants.

Simulation analysis

Six variants in *DNAH6* were found in 10 out of 13 CDGP probands. These six low-frequency variants were predicted to affect the protein's structure (five nonsynonymous amino acid changes and one stop codon). Using the SISu exome data, we investigated the probability of finding protein-altering variants in 10 out of 13 random sets of Finnish individuals. We first detected the number of nonsynonysmous or stop codons in *DNAH6* in Finnish SISu exomes as part of the Exome Aggregation Consortium (ExAC) browser (n ~5,000). There were a total of 82 variants, including 79 SNPs and 3 indels, of which 29 were nonsynonymous variants and one was a stop-gain variant in Variant Effect Predictor (VEP; http://grch37.ensembl.org/Homo_sapiens/Tools/VEP). Many of the Finnish SISu individuals had missing genotypes for one or more of these variants, so they were removed (final n = 2,028). One variant had a high alternative allele frequency (0.96), so it was flipped. We then ran a simulation by randomly selecting 13 individuals and counting how frequently 10 of those individuals contained at least one of the 30 protein-altering variants. We ran the simulation with 10,000 and 100,000 iterations.

Table 5. *Candidate genes on chromosome 2 between 79-124 Mb annotated with GO terms involving growth and reproductive function (Study IV).*

Gene	Position (Mb) on chr 2	General Function (Aceview)	GO term(s)
BCL2L11	111.8	Apoptotic activator	Development of primary sexual characteristics, developmental process involved in reproduction, gonad development, multicellular organism reproduction, reproduction, sexual reproduction, regulation of growth, response to growth factor stimulus
GLI2	121.5	Mediator of Sonic hedgehog (Shh) signaling, embryogenesis including pituitary development	Developmental process involved in reproduction, reproduction, growth, developmental growth
IL1B	113.5	Cell proliferation, differentiation, and apoptosis	Reproduction, growth factor activity, growth factor receptor binding
INHBB	121.1	Inhibin beta B subunit joins the alpha subunit to form a pituitary FSH secretion inhibitor	Cellular process involved in reproduction in multicellular organism, development of primary sexual characteristics, developmental process involved in reproduction, gonad development, multicellular organism reproduction, positive regulation of gonadotropin secretion, regulation of gonadotropin secretion, reproduction, sexual reproduction, negative regulation of gonadotropin secretion, regulation of follicle-stimulating hormone secretion, growth, growth factor activity
KDM3A	86.6	Hormone-dependent transcriptional activation	Cellular process involved in reproduction in multicellular organism, developmental process involved in reproduction, multicellular organism reproduction, reproduction, sexual reproduction
MERTK	112.6	Receptor tyrosine kinase that transduces signals from the extracellular matrix into the cytoplasm by binding to several ligands. Induces phosphorylation of MAPK1. Autosomal recessive retinitis pigmentosa	Development of primary sexual characteristics, developmental process involved in reproduction, multicellular organism reproduction, reproduction, sexual reproduction
NPHP1	110.9	Control of cell division, as well as in cell-cell and cell-matrix adhesion signaling, likely as part of a multifunctional complex localized in actin- and microtubule-based structures	Cellular process involved in reproduction in multicellular organism, developmental process involved in reproduction, multicellular organism reproduction, reproduction, sexual reproduction
PAX8	113.9	Thyroid follicular cell development and expression of thyroid-specific genes	Response to gonadotropin stimulus, cellular response to gonadotropin stimulus

5 Results and Discussion

5.1 rs7759968 and rs314277 at *LIN28B* (Studies I-III)

All three of our GWA approaches were successful for detecting common variants in the Finnish and European populations. In particular, each study picked up variants near the gene *LIN28B.* In the first study, investigating growth during late adolescence in the Finnish population, we showed that the G allele at rs7759938 associated not only with growth from age 14 to adult, but also tagged an effect delaying the pubertal growth spurt. We saw the same association when we expanded the study population to Europeans within the EGG Consortium in Study II. In Study I, NFBC1966 was used as the discovery cohort, with follow-up association tests in YFS. Both of these cohorts also contributed to the discovery GWA in Study II.

Table 6. *Association results of GWA of pubertal growth from age 14 to adult for rs7759938 and rs314268 at the* LIN28B *locus for NFBC1966 and YFS in females and males (Study I).*

NFBC1966 Females					**YFS Females**		
SNP	Allele	N	Beta (SE)	P	N	Beta (SE)	P
rs7759938	G	2242	0.16 (0.032)	7.2×10^{-7}	664	0.17 (0.058)	0.004
rs314268	G	2240	0.14 (0.032)	1.4×10^{-5}	666	0.14 (0.058)	0.02
NFBC1966 Males					**NFBC1966 Males**		
rs7759938	G	2068	0.093 (0.033)	0.005	569	0.14 (0.066)	0.03
rs314268	G	2065	0.095 (0.033)	0.004	570	0.12 (0.066)	0.07

Specifically, in Study I, 5 markers in a single LD block spanning 111 kb on chromosome 6 were significantly associated with late pubertal growth (the difference in height SDS between age 14 and adulthood; $P < 5 \times 10^{-7}$, n = 4,321 Finns). These markers covered the 5' end and upstream region of the gene *LIN28B*. Follow-up analysis of the leading SNP, rs7759938, and a proxy, rs315268, in YFS yielded further support for the locus (combined *P* for rs7759968 = 5.0×10^{-11}, n = 5,562 Finns; **Table 6**), and the strongest associated marker explained ~1% of the total phenotypic variation in females and 0.5% in males. In Study II, rs7759938 also associated with late pubertal growth ($P = 3.9 \times 10^{-9}$, n = 8,863 Europeans), with a stronger signal seen in females than in males.

The amount of growth during puberty might be related to the timing of puberty, since individuals maturing early tend to complete growth and reach their final height earlier than late-maturing individuals. The effect of rs7759938 on late pubertal growth might thus be due to an effect on the timing of the pubertal growth spurt. In Study I, we addressed this question by looking at growth between ages 9 and 12, a proxy measurement for the timing of the onset of the growth spurt, and found that the G allele of rs7759938 did associate

with later onset of the growth spurt (decreased growth during early puberty in both males and females, $P = 1.7 \times 10^{-4}$). Thus, rs7759938 (G), the leading marker at the *LIN28B* locus, associates with a lag in growth in early adolescence followed by an increase in growth in late adolescence, consistent with the hypothesis that this locus tags a delay in the onset of the pubertal growth spurt.

Additionally, in this study as well as in four other studies which were published as Study I was under preparation (146–149), the allele associated with later onset of the pubertal growth spurt also associated with later age at menarche (in Study I: $\beta = 0.124$, SE = 0.023, $P = 8.3 \times 10^{-8}$, n = 4,379 Finns). Both the growth spurt and menarche are secondary manifestations of puberty, occurring after central activation in the hypothalamus. To clarify whether rs7759938 might have distinct effects on pubertal growth and menarcheal timing or reflect the same underlying mechanism, we performed multiple regression analysis including both the genotype at rs7759938 and height growth from 9-12 years of age (as a proxy for early pubertal growth) into the model in Study I. This resulted in a major reduction of the association signal of rs7759938 with AAM ($\beta = 0.046$, SE = 0.03, $P = 0.13$, n = 1,932 Finns), showing that the effects of this locus on the timing of the pubertal growth spurt and AAM are likely not independent and are probably mediated through a common underlying mechanism.

Further evidence for *LIN28B* as a regulator of central pubertal onset came from Study III, since *LIN28B* was the most robustly associated locus in the combined analysis of Tanner male genital and female breast staging, both early markers of pubertal onset **(Table 7)**. In Study III, only rs314276 ($P = 1.6 \times 10^{-8}$; n = 9,915 Europeans) near *LIN28B* reached genome-wide significance in the combined male and female analysis of Tanner female breast and male genital stage. Again, the association was more significant in females (β (SE) = 0.087 (0.017), $P = 3.7 \times 10^{-7}$) than in males (β (SE) = 0.064 (0.024), $P = 0.0085$). Because variation at the *LIN28B* locus associates with multiple secondary pubertal traits in both boys and girls, it appears that this locus may be an important part of the switch which triggers central activation of pubertal onset in both sexes.

Table 7. *Loci significantly associated with pubertal growth and timing in studies I-III. M, males; F, females; H, height; AAM, age at menarche; BMI, body mass index.*

Nearby gene	SNP	*P*	N	Population	Phenotype	Sex	Previously associated related trait	Study
LIN28B	rs7759938	5.0×10^{-11}	5,038	Finnish	Height change SDS (14-adult)	M + F	H, AAM	I
LIN28B	rs7759938	3.9×10^{-9}	8,863	European	Height change SDS (14-adult)	M + F	H, AAM	II
LIN28B	rs314276	1.6×10^{-8}	9,915	European	Tanner pubertal stage	M + F	H, AAM	III
ZBTB38	rs6764769	4.6×10^{-10}	13,960	European	Height SDS at age 10 and 12	M + F	H	II
PXMP3	rs7846385	5.3×10^{-10}	13,942	European	Height SDS at age 10 and 12	M + F	H, AAM	II
EFEMP1	rs1346789	1.2×10^{-8}	13,960	European	Height SDS at age 10 and 12	M + F	H	II
CABLES1	rs6507528	1.3×10^{-8}	13,160	European	Height SDS at age 10 and 12	M + F	H	II
ADAMTSL3	rs1365198	1.5×10^{-8}	13,946	European	Height SDS at age 10 and 12	M + F	H	II
GNA12	rs960273	5.2×10^{-4}	6,986	European	Height SDS at age 12	M	H	II
ADCY	rs1172294	1.0×10^{-8}	10,799	European	Height change SDS (8-adult)	M + F	H, BMI	II
VGLL3	rs7628864	3.2×10^{-9}	5,756	European	Height change SDS (8-adult)	F	AAM	II
Genome-wide significant after the addition of follow-up samples								
MAPK3	rs4788196	9.5×10^{-11}	18,737	European	Height SDS at age 10 and 12	M + F	-	II
MKL2	rs246185	8.9×10^{-9}	3,977	European	Tanner pubertal stage	M	AAM	III

In each of these three studies, the effect at the *LIN28B* locus was stronger in girls than in boys. At first, we hypothesized that the difference in effect could be due to the fact that the underlying phenotype in Studies I-II (the relative height difference between age 14 and adult) differed between males and females since girls enter puberty an average of two years prior to boys. Girls reach average PHV at around 12 years of age, while this occurs in boys at an average age of 14 years. Thus, it could be that the difference in effect size reflects the fact that measuring growth between 14 and adult targets a slightly later phase of the growth curve in girls as it does in boys. However, since the effect at this locus on breast development also appears to be stronger than the effect on male genital development, it seems likely that these observations reflect a true underlying difference between the sexes.

It is not such a leap to envision that variation at the *LIN28B* locus might be dealt with differently in males and females, via an unknown mechanism. Indeed, we found that two partially-correlated markers, rs7759938 and rs314277 (pairwise r^2 = 0.26), exhibit independent effects on adult stature that differ in males and females. Both of these markers were previously associated with adult stature (29). When analysed individually in our Finnish samples, the markers were both associated with adult stature, but rs314277 contributed more to adult height in females whereas rs7759938 was stronger in males. Subsequent analysis of rs314277 conditional on rs7759938 showed opposite effects in males and females, which suggests that this variant has sex-specific effects. Formal testing indeed revealed evidence for a sex-genotype interaction at rs314277 (P = 0.005).

The effect on final height was not attenuated by controlling for AAM for either marker, showing that prepubertal height growth contributes to the adult stature effects for both markers. We analysed the effects of these variants prior to puberty by plotting the effect size of both markers on postnatal growth at 1-year intervals from birth to adulthood in NFBC1966 and HBCS **(Figure 9)**. This revealed distinct association patterns with height for each marker, and also showed sex-specific association patterns. rs7759938 (G) showed a reduction of the effect on relative height after 9, followed by an increase in the effect after age 12 and 14. This marker influenced the timing of the pubertal growth spurt in both sexes, but whereas the male effect was stable until age 10, there was a female-specific reduction of effect between ages 6-9. The association pattern for rs314277 (A) across postnatal growth was markedly different, with a steady effect on each gender that was distinct, but stable across postnatal growth without any effect on the pubertal growth spurt. Males consistently showed negative effect sizes throughout childhood and into adulthood, while the effect in females was positive from age 2 and reached a *P*-value of 0.01 at ages 9 and 10. Multiple regression analysis also showed suggestive evidence for a genotype-sex interaction for rs314277 on postnatal height growth **(Table 8)**.

Table 8. *Linear regression of final height of two variants at the* LIN28B *locus (Study I) separately (single regression) and including both markers in the analysis (multiple regression)*

		Males			**Females**			**Combined**	
		Beta (SE)	**P**	**N**	**Beta (SE)**	**P**	**N**	**Beta (SE)**	**P**
Single regression	rs7759938	0.078 (0.023)	8×10^{-4}	4211	0.058 (0.023)	0.01	4692	0.074 (0.017)	3×10^{-5}
	rs314277	0.028 (0.029)	0.35	4184	0.091 (0.028)	0.001	4676	0.062 (0.020)	0.003
Multiple regression	rs7759938	0.11 (0.031)	5×10^{-4}	4176	0.021 (0.029)	0.48	4645	0.062 (0.021)	0.003
	rs314277	-0.061 (0.039)	0.12	4176	0.075 (0.037)	0.04	4645	0.011 (0.027)	0.69

Furthermore, while regression analyses of height growth during infancy, childhood, and puberty did not reveal any statistically significant associations, rs314277 showed a positive correlation with growth during all three growth periods in females. Nominal evidence also supported an effect on birth length in both sexes ($\beta = 0.058$, SE = 0.025, $P = 0.02$) which might reflect an effect on fetal growth. Many of the tests performed were not independent, but after applying a conservative Bonferroni correction ($P = 0.005$) to account for 10 independent tests, the main findings remained significant.

Thus, of two partially-correlated adult stature-associated markers at this locus, one tags a pubertal timing effect, while the other marks a more general growth effect. The marker that was not associated with pubertal timing (rs314277) showed a sex-genotype interaction for adult stature, and displayed a sex-specific pattern of association with height growth across the entire postnatal trajectory from age 2.

Although the mechanism through which the *LIN28B* locus affects height growth differently in males and females is as yet unknown, sex bias may occur via sex hormones or sex-specific gene expression that is not mediated hormonally, for example by differential microRNA (miRNA) regulation (192). In fact, *LIN28B*, and its homolog *LIN28A*, inhibit the *let-7* pathway of miRNAs, and LIN28 in turn is downregulated by let-7 in a negative feedback loop (193), showing that it is susceptible to regulation by miRNAs. *LIN28B* and *LIN28A* are homologs of a key developmental regulator initially discovered in *C. elegans*, and mouse studies show that the *Lin28-let-7* pathway is important in postnatal growth and puberty (194), indicating that the associations seen in human GWA studies are very likely mediated through this gene. Futhermore, *let-7* targets

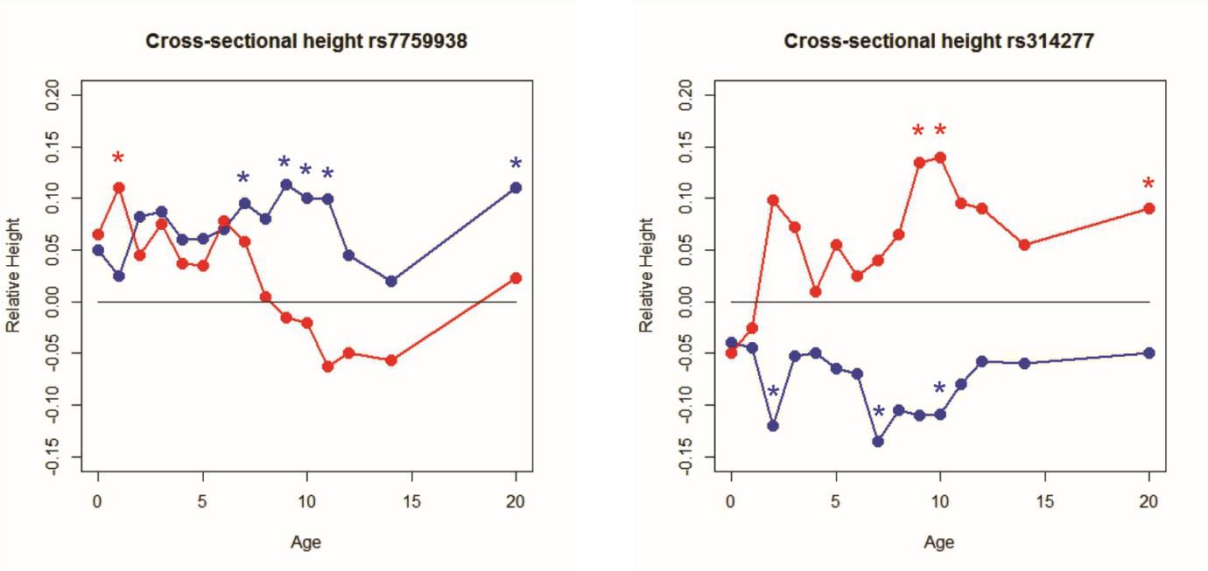


Figure 9. *Linear regression evaluating the independent effects of two markers at the* LIN28B *locus on postnatal height growth. Beta estimates for the association of length/height is plotted at each age. Regression analyses were done in NFBC1966 and HBCS. Birth length was adjusted for gestational age. Males are presented in blue, while females are in red. * indicates a* P*-value of <0.05. Adapted from Widén, et al (2010).*

are enriched in genes nearby height-associated markers (195), suggesting that the height effect of *LIN28B* may act through down-regulation of *let-7* binding targets.

5.2 rs4788196 near MAPK3 (Study II)

In Study II, we again used a simple approach to model three aspects of the pubertal growth spurt: relative height at the take-off phase (Analysis I: height SDS at age 10 yrs in girls and 12 yrs in boys), the total amount of growth during puberty (Analysis II: difference in height SDS between age 8 and adult), and growth during late puberty as in Study I (Analysis III: difference in height SDS between age 14 and adult). This strategy was successful for detecting common genetic variants in more than 18,000 study subjects of European descent. In addition to *LIN28B*, there were 8 loci significantly associated with pubertal growth that were previously associated with final adult stature, AAM, or BMI (addressed below). While there were also many signals that fell beneath the genome-wide significance threshold, additional samples were only available for the follow-up of variants suggestively associated with Analysis I, the single height measurement. Joint analysis of discovery and follow-up stages robustly confirmed a single novel variant not previously associated with related traits (AAM, height, or BMI), rs4788196 (P = 9.49 x 10^{-11}, n = 18,737 Europeans), on chromosome 16. Subsequently, eQTL analysis in whole blood (174) showed that the height-increasing allele (G) at rs4788196 correlated with

decreased expression of *MAPK3* (extracellular signal-related kinase 1, also known as *ERK1*; β =-0.45, $P = 3.6 \times 10^{-5}$), consistent with a mouse study linking deactivation of the gene with increased bone growth (196).

Rare recurrent CNVs near *MAPK3* on chromosome 16p11.2 have been shown to associate with early onset obesity (166, 167), bringing up the possibility that a CNV might mediate the adolescent height effect. This did not appear to be the case. The association results from two analyses, including and excluding CNV carriers, were nearly identical, suggesting that CNVs in this region do not mediate the effect on relative height. The total number of individuals harboring CNVs in this region from NFBC1966 and YFS was 15, including 3 subjects with a previously published deletion of roughly 500kb associated with extreme obesity (166).

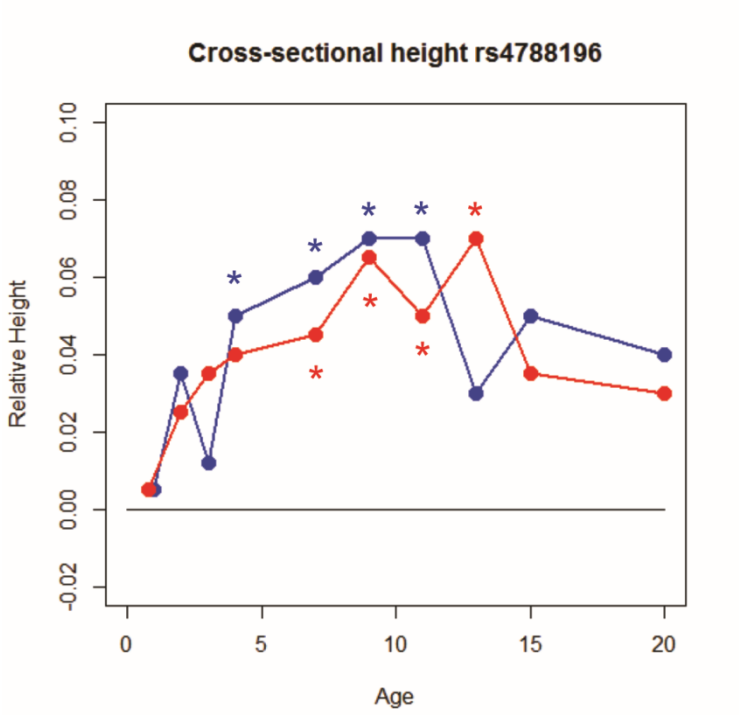


Figure 10. *Linear regression of postnatal height for rs2788196 at the* MAPK3 *locus. * indicates a* P*-value of <0.05. Adapted from Cousminer, et al (2013).*

To characterize the effect of the *MAPK3* variant on height growth in more depth, we plotted the effect size from 8 years to adult for both height and BMI, and on early childhood length yearly from ages 1 – 4 **(Figure 10)**. These analyses revealed a transient effect on height growth for the G allele from age 4 in both sexes that was diluted by adulthood, with no apparent effect on BMI. Rapid growth during childhood and adolescence correlates with early timing of sexual maturation (2), so we also tested rs4788196 for association with AAM (150), and found that the height-increasing allele

associated with earlier menarche (P = 1.42 x 10^{-4}, surpassing the Bonferroni-corrected significance threshold of 0.007). Thus, this locus associates transiently with increased childhood growth in both sexes and earlier menarche in girls, effects that are independent of body mass. GnRH activates the gene, providing a putative biological link between rs4788196 and pubertal timing (197, 198).

5.3 rs246185 near MKL2 (Study III)

In Study III, we were able to assess a direct measurement of male puberty for the first time in a genome-wide association study. Male puberty is challenging to address due to the lack of an easy-to-assess non-invasive marker. However, in the EGG Consortium, nearly 4,000 boys had male genital data assessed by the Tanner pubertal development scale. While there were no genome-wide significant signals after the primary meta-analysis of Tanner genital stage data in boys, one marker at chromosome 16p13.12 (rs246185) suggestively associated at P = 5.0 x 10^{-7}, and surpassed genome-wide significance with the addition of a further 208 samples from the STRIP cohort (P = 8.9 x 10^{-9}). Although our follow-up sample was limited, we found no evidence for heterogeneity of effects between the discovery and follow-up studies (Cochran's statistic Q = 0.23, P = 0.99) and sample characteristics (i.e., ages at phenotype measurement and Tanner stage means) were similar between the discovery and follow-up study subjects.

While multiple potential candidate genes lie in the region, the nearest gene to rs246185, *myocardin-like 2* (*MKL2*), falls 38 kb upstream from the SNP and is a transcriptional activator involved in cell differentiation and development. We were unable to implicate *MKL2*, though, since eQTL analyses from whole blood (175, 176), lymphoblastoid cell lines (178, 179), and skin and adipose tissue (179) did not reveal any links between genotype at the leading SNPs and nearby gene expression levels. This may be partly due to the lack of eQTL-gene expression studies in tissues relevant for sexual maturation, such as those in the HPG-axis, nor during temporally relevant periods.

On the other hand, this locus falls in a predicted promoter region containing a cluster of TF binding sites, so a variant in the region could potentially disrupt a binding site with implications on the expression of a nearby gene important in the timing of pubertal development. Using a denser map to further investigate the region surrounding rs246185, we fine-mapped the region in the four cohorts which contributed to the primary male analysis (ALSPAC, RAINE, YFS, and TEENAGE) by imputing against the 1000 Genomes reference set (22). Only a handful of newly imputed markers showed evidence for association of similar strength as the leading GWA SNP, and none were predicted to affect nearby gene structure. However, rs193536, (r^2 = 0.85 with rs246185), had a slightly stronger association with male genital stage, and is predicted to be likely to affect TF

binding by the RegulomeDB web tool (173) based on multiple lines of evidence, including TF binding motifs, DNase footprints, and DNase peaks.

We also queried additional SNPs in high LD (here defined as $r^2 > 0.6$) with rs246185 but not present in the 1000 Genomes imputed data and detected evidence for another nearby marker, rs74755650 ($r^2 = 0.75$ with rs246185), as potentially impacting TF binding. For each of these two polymorphisms (rs193536 and rs74755650), two TF binding motifs were predicted to be affected by variation in their target sequences based on position weight matrices (PWM) and footprinting, which combines genome sequence information with experimental data to map bound TF binding sites (199, 200). Variation at rs193536 ($r^2 = 0.85$) is predicted to affect binding of PAX-3 and ER, and variation at rs74755650 is predicted to affect binding of PATZ1 (also known as MAZR) and WT1, each supported by evidence from multiple cell types (199, 200). The consensus binding motifs at these sites show that the polymorphisms at the PATZ1 and PAX-3 target sequences may have greater impact on the ability of these factors to recognize their target sites and bind to DNA than the polymorphisms at the WT1 and ER binding sites. For the PATZ1 binding motif, the G allele is strongly preferred to a T or A variant at rs74755650, while the PAX-3 transcription factor strongly prefers a C to an A allele at rs193536. It is therefore possible that variation at these TF binding sites may influence variability in nearby gene expression.

Both PATZ1 and PAX-3 are compelling candidates as mediators of variation in the timing of sexual development since they are both important for morphological development during embryogenesis (201, 202). However, PATZ1 is both a transcriptional repressor (203) and activator (204) that has a critical role in spermatogenesis (205), embryonic and postnatal growth (201), and is a corepressor of androgen receptor-dependent transcription (206), which is important for normal puberty in both males and females (207). Also, both male and female knock-out mice for the *Patz1* gene are infertile (205), indicating a potentially broad role for this factor in the sexual development of both sexes. Still, further study is needed to investigate whether rs74755650, the polymorphism predicted to affect PATZ1 binding, is the underlying functional variant at this locus, and which nearby gene it may affect.

Having established a novel locus for male genital development, we wanted to assess whether this locus might play a role in central pubertal activation in both sexes. Thus, we queried our GWA data on breast development as well as previously published data on menarcheal timing, the pubertal growth spurt, and adult stature for association with rs246185. The effect on breast development was not significant, despite a consistent direction of effect. Still, rs1659127 ($r^2 = 0.84$ with rs246185) is a previously published AAM SNP, so we looked up rs246185 in the complete ReproGen dataset (150) and saw that the T allele associated with 2.1 weeks earlier menarche ($P = 1.0 \times 10^{-8}$), consistent

with our data showing association between the same allele and advanced genital development (higher Tanner score) in boys.

We also found that the same allele also associated with decreased total pubertal growth (Analysis II in Study II), and decreased late pubertal growth (Analysis III in Study II) (182). The early cessation of growth during puberty results in shorter adult stature, evidenced by the significant association between rs1659127 and reduced final adult height (P = 1.1 x 10^{-11}) (29). rs246185 also associated with adult stature, although to a lesser degree (P = 7.5 x 10^{-6}). While adult height SNPs may affect growth during multiple growth periods (156, 182, 208, 209), this locus primarily affects early puberty, diminished growth during puberty, and shorter adult height, consistent with epidemiological observations (210, 211).

5.4 Additional pubertal growth loci (Study II)

In Study II, there was significant deviation from the expected *P*-value distributions for all three combined-gender analyses (I, II, and III), males and females separately for Analysis I (height SDS at age 10 yrs), and females only for Analysis II (difference in height SDS between age 8 and adult). All three models resulted in genome-wide significant loci, and 8 loci (in addition to rs7759938 near *LIN28B* and rs4788196 near *MAPK3*, mentioned above) contained markers that reached *P*-values below the genome-wide significance threshold corrected for testing three primary phenotypes ($P < 1.67 \times 10^{-8}$, after genomic control).

As previously discussed, epidemiological studies support phenotypic correlations between earlier pubertal timing, increased obesity and decreased final height. Indeed, there was substantial overlap between pubertal height growth-associated loci and these related traits **(Figure 11)**. All three pubertal growth approaches detected loci which also associated with pubertal timing (near *MAPK3, PXMP3, VGLL3, ADCY3-POMC,* and *LIN28B*) and 8 signals overlapped with adult stature loci (29). Also, the adult height and menarche-associated locus at *ADCY3-POMC* has further been implicated in childhood (184) and adult obesity (183). One of our signals may be independent from the previously reported signal, since conditioning for the previously reported marker in partial LD with our signal did not attenuate the association nearby *CABLES1* (rs6507528; P = 0.00011).

Comparing the height effects of the leading loci in Study II longitudinally across puberty, we noticed that they could be divided into two groups based on their association with pubertal timing. One group of loci (near *ZBTB38, EFEMP1*, *CABLES1*, *ADAMTSL3*, and *GNA12*) did not associate with AAM, but all impacted height SDS across multiple growth phases, strongly and steadily from prepuberty to adulthood. These loci seem to affect overall growth potential, rather than specific effects during the pubertal growth phase. In contrast, the five variants that did associate with menarche (*LIN28B, MAPK3, ADCY3,*

VGLL3 and *PXMP3*) had diverse effects on the timing and tempo of growth, both before and during puberty.

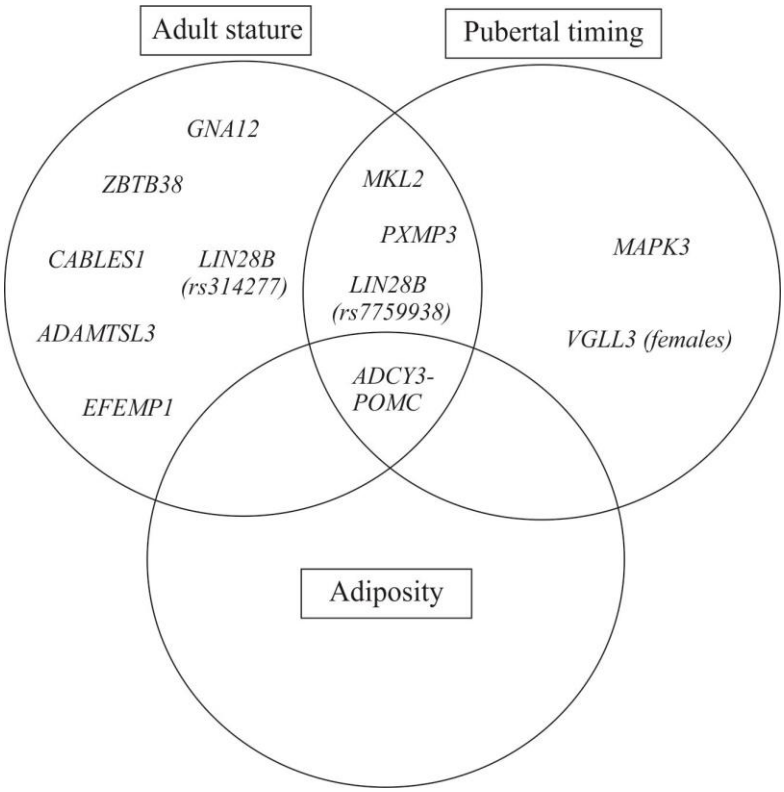


Figure 11. *Variants associated with growth during puberty have distinct association patterns with related traits. Some associate only with pubertal timing (*MAPK3 *and* VGLL3- *Study II), and some have postnatal growth effects across puberty resulting in an influence on final adult height (the effect tagged by rs314277 at* LIN28B *(Study I),* GNA12, ZBTB38, CABLES1, ADAMTSL3, *and* EFEMP1- *Study II). Some pubertal timing effects also result in adult height effects, either due to accelerated growth before puberty (the effect tagged by rs7759938 at* LIN28B *(Studies I and II) and* PXMP3 *(Study II)), or due to a reduction of height growth during the pubertal growth spurt (*MKL2 *(Study III)). Only* ADCY3-POMC *(Study II) associated with pubertal timing, adult height, and an increase in adiposity. Adapted from Cousminer et al (2013).*

Epidemiological studies have demonstrated a developmental pattern linking taller prepubertal stature with earlier puberty, accelerated skeletal maturation, and shorter adult stature due to earlier cessation of growth. While the majority of loci did show the expected association between early menarche and decreased overall pubertal height growth **(Table 9)**, the prepubertal height effects varied for each locus. Three variants (near the genes *MAPK3*, *PXMP3,* and *VGLL3*) linked tall prepubertal stature with earlier menarche, while

the early puberty allele (T) at rs7759938 (*LIN28B*) correlated with shorter prepubertal height, as seen in Study I.

Table 9. *Epidemiologically observed links between BMI, height growth, and puberty, and associations seen at individual loci (Studies I-III).*

	Pre-puberty	**Puberty**	**Adult**
Epidemiological observations			
	Tall or average stature Higher BMI	Earlier menarche Less pubertal growth	Shorter final stature Higher BMI
Individual genetic loci			
LIN28B (rs7759938 G)	Average stature (females) Taller stature (males)	Later menarche Later pubertal growth spurt Later breast development Later male genital development	Taller adult stature (males)
LIN28B (rs314277 A)	Taller stature (females)	No effect	Taller stature (females)
MAPK3 (rs4788196 G)	Taller stature	Earlier menarche	No effect
PXMP3 (rs7846385 C)	Taller stature	Earlier menarche Diminished pubertal growth	No effect
VGLL3 (rs7628854 G)	Taller stature (females)	Earlier menarche Diminished pubertal growth (females)	No effect
ADCY3 (rs1172294 G)	Average stature Higher BMI	Earlier menarche Diminished pubertal growth	Shorter adult stature Higher BMI
MKL2 (rs246185 T)	Average stature	Earlier menarche Earlier male genital development Diminished pubertal growth	Shorter adult stature

Epidemiological studies also correlate early puberty with shorter adult height, presumably due to a shorter overall period of height growth. In contrast, two genetic studies found that alleles associated with early puberty may correlate with either increased or decreased stature (133, 150). In Study II, contradictory to the predicted pattern, at rs7846385 (*PXMP3*) the early menarche allele associates with increased adult stature. A plot of the height growth effects longitudinally across childhood show that this is because the early-menarche allele also associates with tall prepubertal height and a limited reduction in total pubertal growth. This observation agrees with a recent study suggesting that adult height loci mediate their effect by influencing height growth during the prepubertal period (155). However, we find that pubertal growth variants are associated with diverse and unique postnatal growth patterns, and that not all loci which influence growth during puberty also impact adult stature.

We also plotted the effect size of BMI SDS against six age bins across puberty in Study II, but only the previously published menarche and BMI signal at *ADCY3-POMC* was significantly associated with BMI. Epidemiological studies predict that this marker, rs1172294, would associate with increased prepubertal stature, but we did not find that to be the case. Still, the BMI-increasing allele (G) was associated with earlier AAM (P = 8.64 x 10^{-8}) and a decline in pubertal growth in both males and females. Thus, the links between prepubertal height, BMI, and pubertal timing may also vary at each locus from epidemiologically accepted patterns. Consistently, other variants previously associated with childhood obesity (184) showed a similar parallel between elevated BMI and diminished pubertal growth. rs3817334 (*MTCH2*), previously associated with adult (183) but not childhood BMI, also associated with the same decrease in overall pubertal growth despite the lack of association with body mass in childhood, suggesting the presence of some other shared mechanism.

5.5 BMI and Tanner breast stage in girls (Study III)

The GWA meta-analysis of Tanner female breast stage and the female and male combined analysis did not reveal any novel associated loci. Obesity (51) and self-assessment rather than clinician-assessment (212, 213) may both interfere with an accurate evaluation of breast development, so we excluded girls in the top 20^{th} percentile of BMI in contemporary cohorts and re-ran the Tanner breast stage association analysis. However, we observed only a slight attenuation of the effect at the *LIN28B* locus (full dataset, rs314267 β (SE) = 0.09 (0.02), 20^{th} percentile of BMI cutoff β (SE) = 0.06 (0.03)), showing that obesity only slightly affected our results. We also observed a high degree of consistency between menarche-advancing alleles and association to advanced breast development, addressed below. Thus, the lack of genome-wide significant associations is likely due to low power rather than an incorrectly assessed phenotype.

5.6 Genetic overlap of menarche loci with Tanner stage (Study III)

Female breast and male genital development are both early markers of pubertal development occurring soon after central activation. Data on these phenotypes therefore allowed us to assess the overlap between early and late measures of pubertal development by evaluating the proportion of menarche signals that also associate with measures more closely tagging the onset of puberty. Additionally, with data on male genital development, we were able to assess the overlap between regulators of male and female puberty.

While our data provided limited power to detect significant associations, we did see a deviation from the expected *P*-value distribution for menarche variants in both the male and female Tanner stage GWA meta-analysis results. When we performed regression

analyses comparing the association effect on pubertal timing between the AAM and Tanner associations, we found them to be highly consistent, with the menarche-advancing allele also associating with higher Tanner score and thus earlier development (females P = 0.02, males P = 0.006, combined P = 0.0009) for most loci. In addition to *LIN28B* and *MKL2*, the AAM locus at *TMEM38B* was significantly associated with Tanner pubertal stage for both sexes combined ($P = 1.8 \times 10^{-5}$; significance threshold accounting for querying 44 SNPs = 0.001). Additionally, rs3743266 at *RORA*, previously a possible menarche locus that did not surpass genome-wide significance (150) but was associated with menarche in African American women (152), was associated with Tanner stage in our data (P = 0.0008).

Thus, like the loci at *LIN28B* and *MKL2*, we find that other menarche loci are important for pubertal initiation in both sexes and in general tag biological effects upstream of sex-specific effects, as evidenced by the high correlation between menarche-advancing effects and earlier sexual maturation in both boys and girls. While some specific alleles may have effects that vary between boys and girls, the majority of the overall genetic architecture regulating pubertal initiation is similar between them, consistent with a recent study which found a high degree of genetic overlap between genes involved in the first pubertal processes, such as increased hormone secretion, and later development of secondary sex characteristics (214). After our study was published, the 106 loci recently implicated as AAM loci (133) were queried from our Tanner GWA data, and similar results were found: 90 of 106 menarche loci showed consistent direction of effect with the male and female combined Tanner data.

5.7 Genetic overlap of BMI loci and Tanner stage (Study III)

While there is a clear epidemiological correlation between higher BMI and advanced pubertal timing in girls, the relationship between body mass and pubertal timing remains controversial in boys. Most studies show obesity to be associated with earlier puberty, but a subset of overweight boys have pubertal delay (215). In Study III, we were able to address the association between BMI variants and puberty in both sexes by extracting all known BMI-increasing variants (n = 31 (183, 184)) from the Tanner GWA meta-analysis results. Regression analyses of BMI loci also associated with AAM showed that the majority of BMI-increasing alleles associated with earlier pubertal development, especially in girls (female P = 0.019). These findings are consistent with studies of BMI loci and menarcheal timing (150, 216). rs3817334 (T) at *MTCH2*, a locus suggestively associated with decreased growth across puberty (182) but not AAM, associated with advanced breast development (P = 0.0008; threshold accounting for testing 31 loci = 0.0016; sex difference P = 0.003).

Although the QQ-plot for BMI loci in the male genital analysis also deviated from the line of null association similarly as in females, the correlation between the previously reported

BMI-increasing effects and pubertal timing effects was different between boys and girls. While some BMI-increasing loci showed the same trend towards association with earlier puberty, several of these alleles robustly displayed the opposite association. For example, the BMI-increasing allele (A) at rs571312 (*MC4R*), a locus not previously associated with puberty, associated with delayed genital development (P = 0.0009; sex difference P = 0.003). Of the menarche-associated BMI alleles, the T allele at rs887912 (*FANCL*) also associated with delayed sexual development in boys, but the association was not significant after correction for multiple testing (P = 0.005; sex difference P = 0.004). These results provide the first suggestion of the genetic architecture underlying the epidemiological observation that a subset of overweight boys experience pubertal delay.

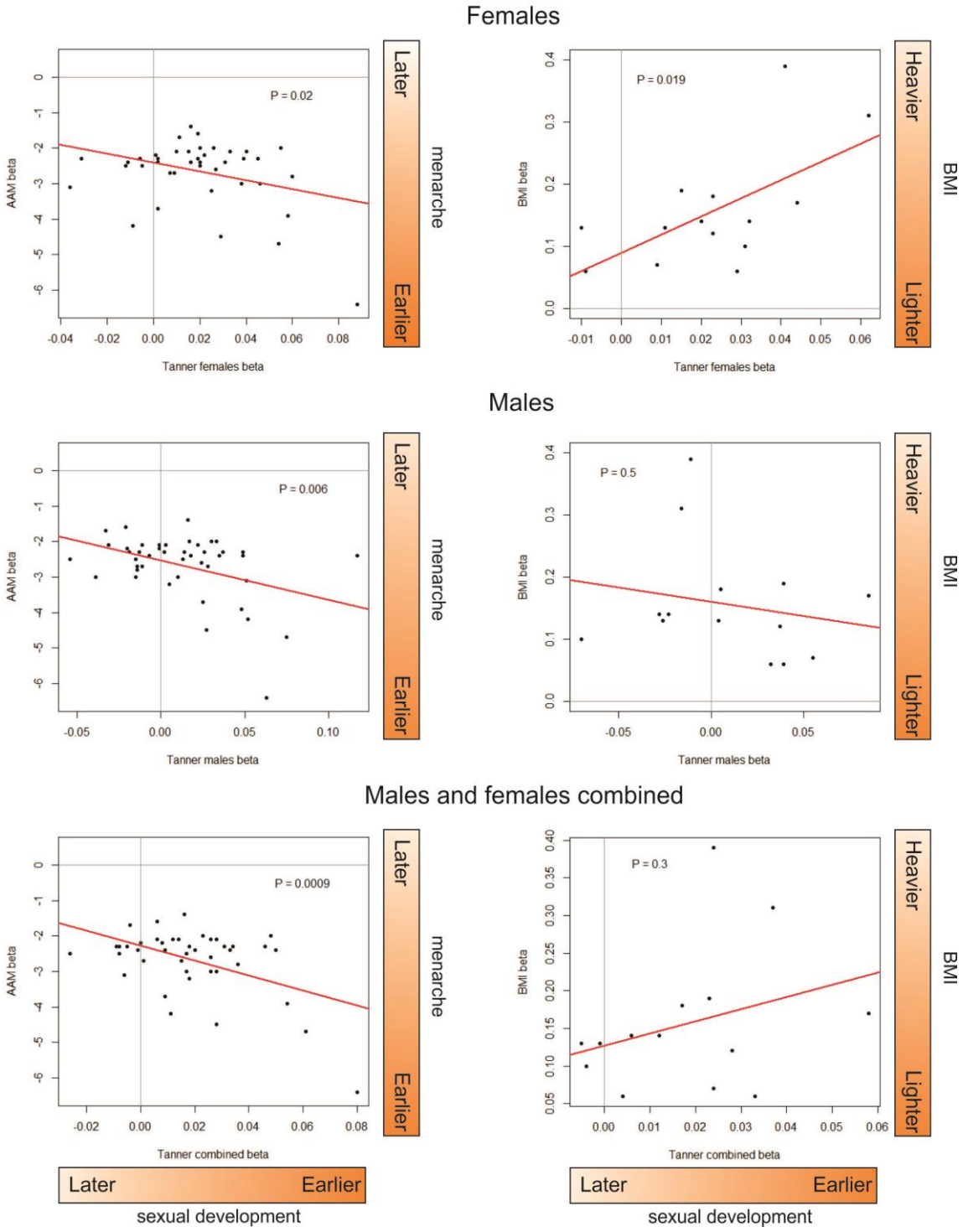


Figure 12. *Association between Tanner stage and menarche-advancing variants (left) and between Tanner stage and BMI-increasing variants (right) for females, males, and both sexes combined. Adapted from Cousminer et al. (2014).*

5.8 Pathway analyses (Studies II and III)

In Studies II and III, we performed several pathway analyses to get a glimpse of the biological pathways that are enriched for genes with markers associated with pubertal growth and Tanner puberty stage.

In Study II, we performed these analyses using the g:Profiler Gene Group Functional Profiler tool (g:GOSt (180)) and MAGENTA GSE analyses (181). These approaches commonly highlighted the TGF-beta signalling pathway and pathways in cancer for loci identified in Analysis I (height SDS at age 10 yrs in girls and 12 yrs in boys). Whereas g:Profiler identified the MAPK-pathway, the GSE analysis showed enrichment of lower than expected *P*-values for genes belonging to the *TOB1* pathway, but the individual implicated gene regions were only suggestively associated and none surpassed the genome-wide significance threshold.

For Study III, we used MAGENTA to perform GSE analysis (181). In the combined-sex GWA results, pathways that were enriched for genes near or containing associated SNPs included aminopeptidase activity, hormone sensitive lipase (HSL) mediated triacylglycerol hydrolysis, and apoptosis. For the female analysis, enriched pathways included thyrotropin-releasing hormone receptor signalling, cardiomyopathy, cancer (acute myeloid leukemia, endometrial cancer, prostate cancer, and the PI3 pathway), as well as apoptosis and the regulation of the actin cytoskeleton. For the male-specific analysis, only the protease inhibitor pathway surpassed an FDR of < 0.05.

These analyses picked up both expected and novel pathways. Some of the expected pathways involved in female breast and male genital development included apoptosis, which is a hallmark of tissue remodeling; hypothalamic-pituitary pathways, such as the TRH (thyrotropin-releasing hormone) receptor signalling pathway (thyroid hormones are essential for normal sexual development); and steroid hormone biosynthesis and hormone biosynthetic process in the secondary tier of results, which show that multiple real signals can be expected below the significance threshold. Also, energy metabolism was picked up in the analysis of both sexes, as seen by the enriched pathway HSL-mediated triacylglycerol hydrolysis. Triacylglycerol provides cholesterol for steroid biosynthesis and is an important energy store (217). Also, in mice, targeted disruption of HSL resulted in male sterility (218), showing the importance of this pathway in sexual development.

For the early pubertal growth analysis, we detected the TGF-beta signalling pathway, which is involved in many important cellular processes from cell growth and differentiation to homestasis. Also, we picked up the MAPK pathway, suggesting that other members of this pathway may be important for growth during puberty in addition to our novel signal near *MAPK3*. Finally, the TOB1 pathway was enriched for genes containing markers with low *P* values. Members of the TOB1 family regulate cell growth,

making it an intuitive pathway for developmental growth. Additionally, it appears to interact with the MAPK pathway (219, 220) and functions as a tumor suppressor (221).

Notably, pathways involved in cancer were picked up for both the early pubertal height analysis and the sexual development analysis, providing clues towards the individual genes which mediate a shared involvement in pubertal timing and cancer risk. Some of the genes nearby the leading associated signals are known to be involved in cancer, such as *LIN28B* and *MAPK3*. For example, *LIN28* overexpression appears to promote malignancy in many cancer types via repression of *let-7* (222). In fact, the same genetic variant associated with adolescent phenotypes has also been linked with cancer severity (223). Also, studies link MAP-kinase expression with cancer (224), and, *MAPK3* expression predicts outcome in hormone-treated breast cancer (225).

Other large GWA studies of AAM have also performed pathway analyses. In the study which included approximately 80,000 women, Ingenuity pathway analysis revealed two functional networks consisting of 16 and 11 genes, respectively (150). The first network was related to gene expression, cellular growth, and cellular function and maintenance, while the second was involved in lipid metabolism, small molecule biochemistry, and molecular transport, showing that genes involved in developmental timing are fundamental to basic biological processes. The subsequent AAM GWA study including more than 180,000 women found that menarche signals were enriched in or near genes underlying rare Mendelian pubertal disorders (133), despite previous failed attempts to find common variation in genes related to HH (135). This is consistent with another candidate gene study which found that genes involved in steroid hormone metabolism and biosynthesis (278 genes in 24,341 women) were also significantly associated with AAM (226). In the large AAM GWA study, genes were also enriched for pituitary development or function, hormonal functions, and nuclear hormone receptors, co-activators, or co-repressors (133), similar to the results of our pathway analyses. Additionally, each of the 9 genes in the GABA-receptor II signalling pathway contained SNPs suggestively associated with AAM, but below the genome-wide significance threshold. The GABA pathway is not unexpected; in animal models, activation of the $GABA_B$ receptor inhibits the secretion of GnRH from the hypothalamus (227).

5.9 Genotyping of CDGP trios (Study IV)

We performed genome-wide analyses to identify genetic outliers as well as detect previously unidentified close relationships among the families contributing to the previously published chromosome 2 linkage peak (144). First, analysis of identity-by-descent (IBD) aimed to detect individuals who appeared more similar to each other than expected in a random population. The IBD estimate (the proportion of alleles predicted to be shared by descent from a common ancestor) for all pairwise combinations of affected

probands was between 0.05-0.08 (0-0.08 in the unselected IA controls; 8 control pairs had values from 0.09-0.5 and thus one member of each pair was excluded from further analyses). Second, the F inbreeding coefficient was -0.03 to 0.02 (-0.06 to 0.1 in controls). These values indicate that the affected probands are not more closely related to each other than expected, nor is there a high probability of consanguinity among their recent ancestors.

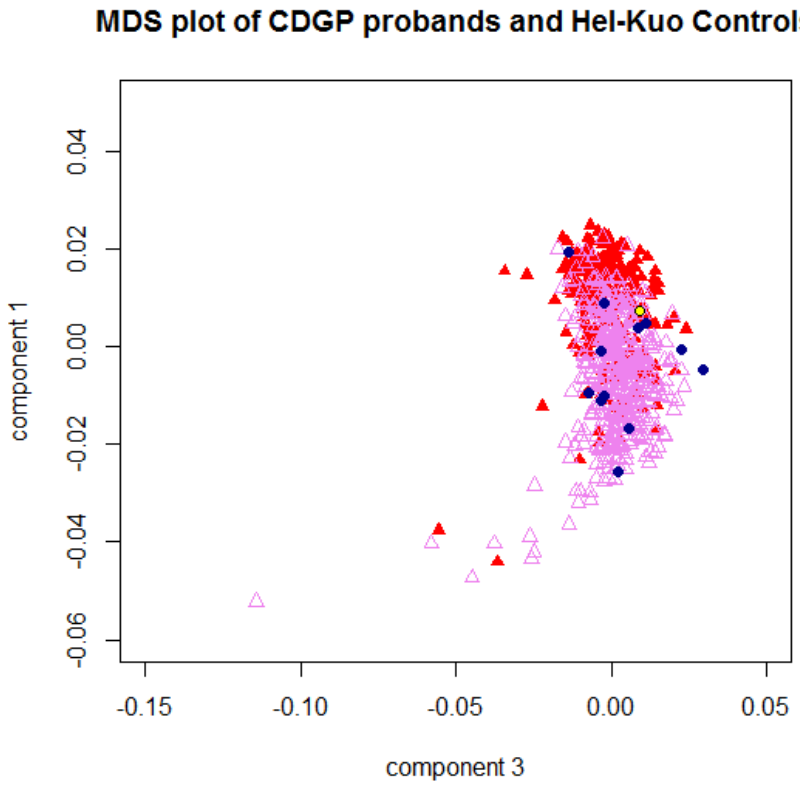


Figure 13. *Multidimensional scaling plot showing the 13 CDGP probands (family 1 in yellow, families 2-13 in blue) against a background of Finnish controls from Helsinki (pink) and Kuopio (red).*

The first three PCs were calculated and used to create a matrix for MDS against the IA control samples. We immediately noticed that the first versus third PC mimics a geographical northeast to southwest axis **(Figure 13)**, with individuals from the Kuopio region at the upper part of the graph, and individuals from Helsinki spread towards the lower part. The Finnish controls from Kuopio and Helsinki heavily overlap, in concordance with the post-World War II migration of Finns into the capital region (228). Although a couple of the probands lie close to one another, in general they all fit homogenously within the background of these controls, so we can conclude that none of the probands were either outliers or more closely related than previously expected.

5.10 Sequencing of CDGP probands and affected parents (Study IV)

Targeted sequencing of chr 2: 79-124 Mb was performed in 26 samples (the affected proband and their suspected affected parent). The parent classified as affected was equally the proband's mother (n=7) or father (n=6). The mean coverage depth in the target region across all samples was 60.6X and in the protein-coding regions was 41X, with an average of 44,826,375 reads per sample. On average, 98.2% of reads aligned to the reference genome, and in the targeted region, 82.2% of annotated exons (Ensembl protein-coding genes, n = 229) were covered at least 10X.

5.11 DNAH6 (Study IV)

Of the gene-centered analyses, low-frequency variants following the expected transmission pattern were found in three genes (*DNAH6, KDM3A,* and *AC013402.2/LOC150568*). Low-frequency variants were inherited from either parent within the gene *DDX18,* but some of these turned out to be fairly common in the Finnish population (MAF ~0.18). However, only *DNAH6* (Dynein, Axonemal, Heavy Chain 6), contained variants that met frequency and sharing criteria and had predicted protein-altering consequences, although these variants wee transmitted from both the affected and unaffected parents. *DNAH6* harbored 6 variants in the top 13 families at low frequency with predicted stop or missense codons **(Table 10)**. All 6 variants were present at less than ~6% in the Finnish population. In family 1, rs184604697 was transmitted from the affected parent to the proband and resulted in a premature stop codon (c.7689C>A, p.Tyr2563Ter). Two of the other 5 variants shared by other families were predicted to result in changes that would be damaging or deleterious by SIFT and/or PolyPhen-2, although these variants were equally likely to be transmitted from the affected parent or the alternative parent. The six variants were located in two sets of three consecutive exons (exons 23-25 and 46-48).

These 6 exons were chosen for follow-up sequencing in an additional 135 Finnish CDGP patients **(Table 11)**. We detected 6 variants in these samples, 5 of which were seen in the top 13 families. The stop variant seen in family 1 was not detected in other CDGP probands. One novel mutation was observed that was not seen in the Finnish population controls. However, none of the variants were enriched in the CDGP probands compared to Finnish controls. Unfortunately, there was no data available on pubertal timing for the Finnish population sample to check for delayed puberty in carriers of these variants.

Table 10. *Variants in* DNAH6 *found in the CDGP families (Study IV).*

Variant	Position on chr 2(GRCh37)	Consequence	Predicted Effect (SIFT/PolyPhen)	Family	Transmission
rs184604697	84926729	Stop gained c.7689C>A, p.Tyr2563Ter	NA	Family 1	Affected parent
rs61743118	84846930	Nonsynonymous c. 3694A>G, p.Met1232Val	Tolerated (1) / benign (0.002)	Family 2 Family 5 Family 6 Family 11	Affected parent (Families 6, 11), other parent or de novo (Families 2, 5)
rs146306207	84926746	Nonsynonymous c.7706G>A, p.Arg2569His	Deleterious (0)/ probably damaging (0.958)	Family 4	Other parent or de novo
rs61733547	84848596	Nonsynonymous c.3992G>A, p.Arg1331His	Tolerated (0.1)/ benign (0.023)	Family 8 Family 13	Affected parent (Family 8), other parent or de novo (Family 13)
rs114514726	84924894	Nonsynonymous c.1636G>A, p.Val546Ile	Tolerated (0.89)/ unknown	Family 8	Other parent or de novo
rs200844717	84928399	Nonsynonymous c.7997C>A, p.Ser2666Tyr	Deleterious (0)/ probably damaging (0.998)	Family 9 Family 13	Affected parent (both)

DNAH6 encodes a dynein which functions as a microtubule-associated motor protein. Dyneins can be axonemal, providing the motive force for cilia and flagella, or cytoplasmic, where they assist intercellular movement and cytoskeletal remodeling (229). Although this particular gene is predicted by similarity-based approaches to localize in the cilium axoneme (http://www.uniprot.org/), there is a potential role for dyneins in steroid hormone metabolism via cytoskeletal trafficking of substrates (230), and heavy chain dyneins are expressed in both the brain and testis (231). *DNAH6* is a large gene spanning 300kb, and the largest of four isoforms is made up of 4,158 amino acid residues. The stop codon transmitted in family 1 is predicted to truncate a large portion of the protein, including several ATP-binding motor subunits and the microtubule-binding stalk, but the functional implication of this truncation is not known. Not surprising for a gene of its size, a substantial amount of variation has been documented within the gene, including 58 stop codons. Furthermore, when we ran simulation analyses, we found that approximately 13% of the time, 10 or more nonsynonymous or stop codons could be found in random sets of 13 Finnish individuals from the SISu exome dataset, showing that the high frequency of protein-altering variants in *DNAH6* in our CDGP probands is likely due to chance.

Table 11. *Frequencies of* DNAH6 *variants in the 1000G population, in the SISu Finnish population, and in a set of 135 unrelated Finnish CDGP patients, showing that these variants are not enriched in Finns or in subjects with pubertal delay.*

Variant	Global 1000G frq	SISu exome frq	CDGP frq (n variant alleles/ total alleles)	*P*
rs184604697	0.001	0.004	0	NA
rs61743118	0.011	0.052	0.049 (12/246)	0.87
rs146306207	0.001	0.009	0.015 (4/270)	0.53
rs61733547	0.026	0.047	0.049 (10/204)	1.0
rs114514726	0.017	0.029	0.02 (5/246)	0.68
rs200844717	NA	0.01	0.008 (2/252)	1.0

5.12 Menarche-associated region surrounding rs6758290 (Study IV)

We also investigated a new locus associated with AAM that falls in the region under the chromosome 2 linkage signal, tagged by rs6758290 at 105.8 Mb (133). Although we examined our families for shared variants in this region, our results did not detect low-frequency mutation that would explain the tendency toward pubertal delay, in part due to the lack of publically available frequency data for part of this region. With access to raw sequencing data from the SISu exome project, we were able to determine that this region is likely difficult to sequence. In the SISu data, only one variant in LD with rs6758290 could be detected, and it had a low call rate of 19% and an average read depth of 2. In our CDGP subjects, however, SNPs in this region appear to have been sequenced successfully, with an average call depth across all families of 13 – 54 reads per marker. While several of these variants were shared among the CDGP probands, none were predicted to be protein-damaging or to affect nearby gene regulation. Of these variants, 3 did have population frequency data, and all were common in the Finnish population, with a MAF of ~0.2.

While we could not detect compelling evidence for variation predisposing pubertal delay in our study, this region remains an interesting candidate for future study. The nearest gene to the AAM signal is *GPR45*, which is a G protein-coupled receptor. Little is known about the gene, but a BlastP search (in AceView; http://www.ncbi.nlm.nih.gov/IEB/Research/Acembly/) showed that the most closely related gene in *C. elegans* is *npr-1*, a neuropeptide receptor gene that is homologous to the mammalian neuropeptide Y (NPY) receptor. NPY appears to have an inhibitory function in the HPG axis, restraining the onset of puberty (232). Also, NPY levels were higher in girls with CDGP than girls with normal pubertal timing in one study (233), suggesting that it may play an important role in pubertal delay if deregulated.

5.13 Other variant detection approaches (Study IV)

The other approaches taken to detect variants in this study that might predispose pubertal delay were not successful. While many genes contained shared transmitted variants with no frequency annotations in the top 5 families, little data could be gathered to prioritize any of these as plausible targets for follow-up studies. Most nonsynonymous variants were not predicted to be protein-damaging, and many apparently shared variants were flagged as suspect or found in multiple genomic locations, suggesting that these regions are difficult to sequence correctly, and that future improvements in sequence annotation is necessary before further assessment can be made.

As for regulatory variants surrounding the best functional candidate genes, nearly all shared variants lacked publically available frequency information and could not be assessed further. For many variants, information on their potential to fall in regulatory regions was also limited. In any case, verifying a regulatory variant's interaction with pubertal timing is not straightforward.

Many challenges exist that might affect our ability to detect susceptibility variants associated with CDGP. First, if the variant is regulatory, it may be quite subtle and very difficult to identify. Although our overall sequencing coverage was good, the causative mutation could lie in a gene that was not sequenced well, as difficult-to-sequence regions clearly exist in this part of the genome. Also, variant annotation depends on the choice of transcript set as well as the software used for annotation (234), so if variants were not annotated appropriately, they may have been excluded from further analysis. Our knowledge of regulatory regions of the genome is patchy at best, so the causative variant might lie in a regulatory region that is of unknown function or regulating a gene that is not an obvious functional candidate based on current knowledge. Finally, other studies have encountered difficulties in detecting variants under linkage signals, which suggests that many linkage findings may be false positive results. Thus, variation predisposing delayed growth and puberty remains intractable to current methods, and future advances may help to detect the genetic mechanisms underlying the tail late end of normal pubertal development.

6 Conclusions and future directions

When the work in this thesis began, little was known about genetic variants which contribute to pubertal development in the normal population, and most genetic studies focused primarily on age at menarche. By taking a broad view of the pubertal phenotypes available, we expanded the phenotypes investigated to include measures of the pubertal growth spurt as well as female breast and male genital development. Utilizing growth as a mapping target, we showed that the menarche-associated genomic region at *LIN28B* regulates postnatal growth, with a strong influence on growth during puberty. Using a longitudinal approach made possible by frequent height assessments in Finnish population-based cohorts across adolescence, we described independent and sex-specific effects for several variants at this locus. These findings provided an early suggestion that *LIN28B* is an important switch regulating multiple aspects of growth and development, a hypothesis which was subsequently confirmed in mouse studies. *LIN28B* is a highly conserved regulator of developmental timing first seen in *c. elegans*, and is reciprocally controlled in a double-negative feedback loop by the *let-7* family of miRNAs. Due to its complex involvement in normal cellular growth and division as well as in cellular transformation when overexpressed, functional work remains to be done to fully understand how the *LIN28-let-7* pathway links developmental pathways with adult health outcomes in humans.

By tracking genetic effects across multiple growth periods, we uncovered 10 additional variants significantly associated with pubertal growth, some of which also influence pubertal timing, adiposity, or final height. While genetically defined association patterns do not always follow epidemiologically predicted correlations between these traits, they may pinpoint the molecular mechanisms and pathways which underlie connections between related traits. In turn, the characterization of these pathways may provide new insights into the links between developmental patterns and adult health outcomes. These links are likely to be quite complex, and detailed functional studies are required to investigate each locus in depth. These studies will need to determine which nearby genes convey alterations in pubertal growth and timing, and how these same genes might influence cardiovascular health, energy metabolism, or cancer risk later in life.

This work also expanded the study of the genetic architecture of puberty to include males, previously overlooked in genetic association studies, and in doing so shed light on the genetic overlap between male and female pubertal maturation. We found that while the leading association signals vary between the sexes, there is likely a substantial overlap between molecular mechanisms that regulate pubertal initiation in males and females. Conversely, the genetic relationship between body fatness and puberty appears to be partially distinct in males and females. Further studies are required to understand the complex sex-specific correlations between energy metabolism and pubertal timing, and how they differ between boys and girls.

Meanwhile, genetic variation which leads to pubertal delay continues to be difficult to detect. In our work, we investigated a previously published locus on chromosome 2 with compelling evidence for a susceptibility locus linked in Finnish constitutional delay of growth and puberty families. Despite the advantages of studying CDGP in Finnish families – a relatively homogenous genetic structure, detailed phenotype data with strict criteria for the inclusion of family members, and a unique resource of population-based sequences in the SISu set – we were unable to pinpoint genomic variation that would predispose to pubertal delay in these families. It is possible that the causative locus was missed due to gaps in sequencing coverage, but it is also possible that the variant may be subtle, or regulatory, in nature. Thus, genetic studies attempting to localize causative variants under a linkage peak remain challenging, partly due to gaps in the functional annotation of the genome as well as limitations of current sequencing technology to deal with difficult-to-sequence regions. Advancements in annotation of gene functions and regulatory regions of the genome will be necessary for facilitating studies of complex phenotypes like CDGP.

7 Acknowldegements

This work was carried out at the Institute for Molecular Medicine Finland (FIMM). Director Olli Kallioniemi is heartily thanked for his leadership and for providing an excellent atmosphere for doing science. For financial support, I gratefully acknowledge the Academy of Finland and the Helsinki Biomedical Graduate Program, now the Doctoral Programme in Biomedicine (DPBM).

I would like to thank Prof. Greg Gibson for acting as my opponent and rearranging a busy schedule to include my defense. Thank you for helping to create an atmosphere of anticipation, rather than fear, for this great adventure! I also wish to thank my pre-examiners Prof. Hannes Lohi and Prof. Taneli Raivio for taking the time to give a thorough review of my thesis. You both provided valuable expertise and critique which served to improve this thesis.

I have been blessed to know several strong female role models in my life. Although I didn't know late Academician Leena Palotie well, she was certainly one of them, and my thanks go to her for being a trailblazer. The most important message I took away from our brief encounters was that you can have both a career and a family, and you can do it with style.

I owe a huge debt of gratitude to my supervisor these past years, Docent Elisabeth Widén, for guiding me scientifically, professionally, and personally. Thank you for believing in me, putting me on the path of so many opportunities, giving me the time to work things out in my own way, and for the flexibility to continue my work from the other side of the world. I would not be who I am and where I am today without your dedicated and patient support and guidance.

My collaborators, both in Finland and abroad, have made this work possible. I would especially like to thank Aarno Palotie for guidance and for being so approachable, Mark McCarthy for skillfully running the EGG Consortium, and everyone involved in the many cohort studies on which this work is based. In particular, all of the people who made the Finnish cohort studies possible are gratefully acknowledged. Also, many thanks go to everyone at the FIMM Technology Center, particularly in the Sequencing, Genomics, and IT units.

A big thank you for the invaluable consulting services from the statisticians next door, especially Samuli Ripatti and Ida Surakka. Not having a statistical mind myself, you guys are like magicians to me. Ida, thank you a million times for all the R scripts. Maybe one day, when I grow up, I'll learn how to write one all by myself. Thanks to Antti, Emmi, Himanshu, and Johannes, who were always there to lend a hand, answer a question, or collaborate. To fellow lab member Jaakko, I want to say thank you for the opportunity to both teach and learn together with you. I felt like we were "partners in crime" as the only

students in our small group. Thanks for bringing energy and dedication to work every day, with a wonderful and motivating positive attitude.

To my "adopted home" at SickKids in Toronto – huge thanks to Mark Palmert for taking me into the group and allowing me to hang around. I learned so much from joining your group, experiencing another style of supervision, and gaining a different perspective on how to study genetics and puberty. I especially want to thank you for taking the time to meet with me regularly, and for giving me career and personal guidance. A very special thanks to Christina Corre for teaching me all about mice and for the lunchtime therapy sessions. Sharing my days with a Swede made the transition from Finland to Canada so much easier in so many ways.

All the wonderful friends and colleagues at FIMM and around Meilahti, you made it such a great place to be and have already been sorely missed. I hope our paths continue to cross in the future. There are too many to name you all, but I especially want to mention Ansku, Annu, Jenni, Maral, Mari, Marjis, Olli, Perttu, Taru, Tiia, Tuuli, Virpi, Verneri, Will, and of course so many others. I would also like to thank the ladies who make the wheels turn and have always had patient answers to my questions and solutions to any problem: Susanna Rosas, Reetta Niemelä, and Sari Kivikko, in particular.

To all my friends in Finland, most of you from Thespians Anonymous, thanks for all the great times over the years and for keeping me sane. A very special thank you to Rachel: I could not have completed this work without your companionship these last couple of years. You and I are two of a kind. Miro, old friend, who'd have guessed we'd both make it this far. I ultimately have you to thank for setting me on this journey!

To Paula and Veltsa, your courage and strength, kindness and love, support and caring have continued to amaze me all these years. You have always been there for me and I am grateful for every experience we have shared together. Oma Elke, thank you for all you do to take care of us, and for your sweet disposition that always brings a smile to my face.

The greatest role model in my life is, and always has been, Mom. You taught me what it means to be *a force to be reckoned with*. You are the smart, beautiful, extremely talented, and accomplished woman I have to thank for showing me what kind of person I should become. I am so proud to be your daughter.

To Dad, I hope you got to have those conversations with Darwin and tell him all about modern science. Someday I'll fill you in on what you've missed, too.

Finally, to my dear Ami, Elliot, and Misha: you make my world go around. And from now on, I will only answer to "*Doctor* Mom." Better get used to it.

References

1. Luo,Z.C. and Karlberg,J. (2000) Critical growth phases for adult shortness. *Am. J. Epidemiol.*, **152**, 125–131.
2. Tanner,J.M., Whitehouse,R.H., Marubini,E. and Resele,L.F. (1976) The adolescent growth spurt of boys and girls of the Harpenden growth study. *Ann. Hum. Biol.*, **3**, 109–126.
3. Perry,R.J., Farquharson,C. and Ahmed,S.F. (2008) The role of sex steroids in controlling pubertal growth. *Clin. Endocrinol. (Oxf.)*, **68**, 4–15.
4. Su,S.-C., Chung,W.-H. and Hung,S.-I. (2014) Digging Up the Human Genome: Current Progress in Deciphering Adverse Drug Reactions. *BioMed Res. Int.*, **2014**, 824343.
5. Wang,E., Zaman,N., McGee,S., Milanese,J.-S., Masoudi-Nejad,A. and O'Connor,M. (2014) Predictive genomics: A cancer hallmark network framework for predicting tumor clinical phenotypes using genome sequencing data. *Semin. Cancer Biol.*, 10.1016/j.semcancer.2014.04.002.
6. Weiling,F. (1967) Mendel, G.: Experiments in plant hybridisation. Mendel's original paper in English translation with Commentary and Asessment by the late Sir Ronald A. Fisher, edited by J. H. Bennett. Oliver & Boyd, Edinburgh and London 1965; IX, 95 S., Preis 21 s. *Biom. Z.*, **9**, 208–208.
7. Dahm,R. (2005) Friedrich Miescher and the discovery of DNA. *Dev. Biol.*, **278**, 274–288.
8. Paweletz,N. (2001) Walther Flemming: pioneer of mitosis research. *Nat. Rev. Mol. Cell Biol.*, **2**, 72–75.
9. Galton,S.F. (1889) Natural Inheritance MacMillan.
10. Benson,K.R. (2001) TIMELINE: T. H. Morgan's resistance to the chromosome theory. *Nat. Rev. Genet.*, **2**, 469–474.
11. Creighton,H.B. and McClintock,B. (1935) The Correlation of Cytological and Genetical Crossing-Over in Zea Mays. A Corroboration. *Proc. Natl. Acad. Sci. U. S. A.*, **21**, 148–150.
12. Avery,O.T., Macleod,C.M. and McCarty,M. (1944) STUDIES ON THE CHEMICAL NATURE OF THE SUBSTANCE INDUCING TRANSFORMATION OF PNEUMOCOCCAL TYPES : INDUCTION OF TRANSFORMATION BY A DESOXYRIBONUCLEIC ACID FRACTION ISOLATED FROM PNEUMOCOCCUS TYPE III. *J. Exp. Med.*, **79**, 137–158.
13. WATSON,J.D. and CRICK,F.H. (1953) Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. *Nature*, **171**, 737–738.
14. CRICK,F.H., BARNETT,L., BRENNER,S. and WATTS-TOBIN,R.J. (1961) General nature of the genetic code for proteins. *Nature*, **192**, 1227–1232.
15. Saiki,R.K., Bugawan,T.L., Horn,G.T., Mullis,K.B. and Erlich,H.A. (1986) Analysis of enzymatically amplified beta-globin and HLA-DQ alpha DNA with allele-specific oligonucleotide probes. *Nature*, **324**, 163–166.
16. Riordan,J.R., Rommens,J.M., Kerem,B., Alon,N., Rozmahel,R., Grzelczak,Z., Zielenski,J., Lok,S., Plavsic,N. and Chou,J.L. (1989) Identification of the cystic fibrosis gene: cloning and characterization of complementary DNA. *Science*, **245**, 1066–1073.
17. Lander,E.S., Linton,L.M., Birren,B., Nusbaum,C., Zody,M.C., Baldwin,J., Devon,K., Dewar,K., Doyle,M., FitzHugh,W., *et al.* (2001) Initial sequencing and analysis of the human genome. *Nature*, **409**, 860–921.
18. Venter,J.C., Adams,M.D., Myers,E.W., Li,P.W., Mural,R.J., Sutton,G.G., Smith,H.O., Yandell,M., Evans,C.A., Holt,R.A., *et al.* (2001) The sequence of the human genome. *Science*, **291**, 1304–1351.
19. ENCODE Project Consortium (2011) A user's guide to the encyclopedia of DNA elements (ENCODE). *PLoS Biol.*, **9**, e1001046.

20. Arner,E., Daub,C.O., Vitting-Seerup,K., Andersson,R., Lilje,B., Drabløs,F., Lennartsson,A., Rönnerblad,M., Hrydziuszko,O., Vitezic,M., *et al.* (2015) Transcribed enhancers lead waves of coordinated transcription in transitioning mammalian cells. *Science*, **347**, 1010–1014.
21. Naidoo,N., Pawitan,Y., Soong,R., Cooper,D.N. and Ku,C.-S. (2011) Human genetics and genomics a decade after the release of the draft sequence of the human genome. *Hum. Genomics*, **5**, 577–622.
22. Consortium,T. 1000 G.P. (2012) An integrated map of genetic variation from 1,092 human genomes. *Nature*, **491**, 56–65.
23. International Cancer Genome Consortium, Hudson,T.J., Anderson,W., Artez,A., Barker,A.D., Bell,C., Bernabé,R.R., Bhan,M.K., Calvo,F., Eerola,I., *et al.* (2010) International network of cancer genome projects. *Nature*, **464**, 993–998.
24. Mefford,H.C. and Eichler,E.E. (2009) Duplication Hotspots, Rare Genomic Disorders and Common Disease. *Curr. Opin. Genet. Dev.*, **19**, 196–204.
25. Kellis,M., Wold,B., Snyder,M.P., Bernstein,B.E., Kundaje,A., Marinov,G.K., Ward,L.D., Birney,E., Crawford,G.E., Dekker,J., *et al.* (2014) Defining functional DNA elements in the human genome. *Proc. Natl. Acad. Sci.*, 10.1073/pnas.1318948111.
26. Altshuler,D., Daly,M.J. and Lander,E.S. (2008) Genetic Mapping in Human Disease. *Science*, **322**, 881–888.
27. MacDonald,M.E., Ambrose,C.M., Duyao,M.P., Myers,R.H., Lin,C., Srinidhi,L., Barnes,G., Taylor,S.A., James,M., Groot,N., *et al.* (1993) A novel gene containing a trinucleotide repeat that is expanded and unstable on Huntington's disease chromosomes. *Cell*, **72**, 971–983.
28. Bailey-Wilson,J.E. and Wilson,A.F. (2011) Linkage analysis in the next-generation sequencing era. *Hum. Hered.*, **72**, 228–236.
29. Lango Allen,H., Estrada,K., Lettre,G., Berndt,S.I., Weedon,M.N., Rivadeneira,F., Willer,C.J., Jackson,A.U., Vedantam,S., Raychaudhuri,S., *et al.* (2010) Hundreds of variants clustered in genomic loci and biological pathways affect human height. *Nature*, **467**, 832–838.
30. Pinyerd,B. and Zipf,W.B. (2005) Puberty-timing is everything! *J. Pediatr. Nurs.*, **20**, 75–82.
31. Sisk,C.L. and Foster,D.L. (2004) The neural basis of puberty and adolescence. *Nat. Neurosci.*, **7**, 1040–1047.
32. Ojeda,S.R., Lomniczi,A., Sandau,U. and Matagne,V. (2010) New concepts on the control of the onset of puberty. *Endocr. Dev.*, **17**, 44–51.
33. Ebling,F.J.P. (2005) The neuroendocrine timing of puberty. *Reprod. Camb. Engl.*, **129**, 675–683.
34. Tobet,S.A., Bless,E.P. and Schwarting,G.A. (2001) Developmental aspect of the gonadotropin-releasing hormone system. *Mol. Cell. Endocrinol.*, **185**, 173–184.
35. Terasawa,E. and Fernandez,D.L. (2001) Neurobiological mechanisms of the onset of puberty in primates. *Endocr. Rev.*, **22**, 111–151.
36. Herbison,A.E. (2007) Genetics of puberty. *Horm. Res.*, **68 Suppl 5**, 75–79.
37. Ojeda,S.R., Lomniczi,A., Mastronardi,C., Heger,S., Roth,C., Parent,A.-S., Matagne,V. and Mungenast,A.E. (2006) Minireview: the neuroendocrine regulation of puberty: is the time ripe for a systems biology approach? *Endocrinology*, **147**, 1166–1174.
38. Kotani,M., Detheux,M., Vandenbogaerde,A., Communi,D., Vanderwinden,J.-M., Poul,E.L., Brézillon,S., Tyldesley,R., Suarez-Huerta,N., Vandeput,F., *et al.* (2001) The Metastasis Suppressor Gene KiSS-1 Encodes Kisspeptins, the Natural Ligands of the Orphan G Protein-coupled Receptor GPR54. *J. Biol. Chem.*, **276**, 34631–34636.
39. Calley,J.L. and Dhillo,W.S. (2014) Effects of the Hormone Kisspeptin on Reproductive Hormone Release in Humans. *Adv. Biol.*, **2014**, e512650.
40. Auchus,R.J. and Rainey,W.E. (2004) Adrenarche - physiology, biochemistry and human disease. *Clin. Endocrinol. (Oxf.)*, **60**, 288–296.

41. Marshall,W.A. and Tanner,J.M. (1969) Variations in pattern of pubertal changes in girls. *Arch. Dis. Child.*, **44**, 291–303.
42. Marshall,W.A. and Tanner,J.M. (1970) Variations in the pattern of pubertal changes in boys. *Arch. Dis. Child.*, **45**, 13–23.
43. Tanner,J.M., Whitehouse,R.H. and Takaishi,M. (1966) Standards from birth to maturity for height, weight, height velocity, and weight velocity: British children, 1965. I. *Arch. Dis. Child.*, **41**, 454–471.
44. Wehkalampi,K., Silventoinen,K., Kaprio,J., Dick,D.M., Rose,R.J., Pulkkinen,L. and Dunkel,L. (2008) Genetic and environmental influences on pubertal timing assessed by height growth. *Am. J. Hum. Biol. Off. J. Hum. Biol. Counc.*, **20**, 417–423.
45. Petersen,A.C., Crockett,L., Richards,M. and Boxer,A. (1988) A self-report measure of pubertal status: Reliability, validity, and initial norms. *J. Youth Adolesc.*, **17**, 117–133.
46. COLEMAN,L. and COLEMAN,J. (2002) The measurement of puberty: a review. *J. Adolesc.*, **25**, 535–550.
47. Duke,P.M., Litt,I.F. and Gross,R.T. (1980) Adolescents' self-assessment of sexual maturation. *Pediatrics*, **66**, 918–920.
48. Morris,N.M. and Udry,J.R. (1980) Validation of a self-administered instrument to assess stage of adolescent development. *J. Youth Adolesc.*, **9**, 271–280.
49. Taylor,S.J., Whincup,P.H., Hindmarsh,P.C., Lampe,F., Odoki,K. and Cook,D.G. (2001) Performance of a new pubertal self-assessment questionnaire: a preliminary study. *Paediatr. Perinat. Epidemiol.*, **15**, 88–94.
50. Schlossberger,N.M., Turner,R.A. and Irwin,C.E.,Jr (1992) Validity of self-report of pubertal maturation in early adolescents. *J. Adolesc. Health Off. Publ. Soc. Adolesc. Med.*, **13**, 109–113.
51. Bonat,S., Pathomvanich,A., Keil,M.F., Field,A.E. and Yanovski,J.A. (2002) Self-assessment of pubertal stage in overweight children. *Pediatrics*, **110**, 743–747.
52. Kaminski,B.A. and Palmert,M.R. (2008) Genetic control of pubertal timing. *Curr. Opin. Pediatr.*, **20**, 458–464.
53. Sharma,J.C. (1983) The genetic contribution to pubertal growth and development studied by longitudinal growth data on twins. *Ann. Hum. Biol.*, **10**, 163–171.
54. Towne,B., Czerwinski,S.A., Demerath,E.W., Blangero,J., Roche,A.F. and Siervogel,R.M. (2005) Heritability of age at menarche in girls from the Fels Longitudinal Study. *Am. J. Phys. Anthropol.*, **128**, 210–219.
55. 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.
56. Van den Berg,S.M. and Boomsma,D.I. (2007) The familial clustering of age at menarche in extended twin families. *Behav. Genet.*, **37**, 661–667.
57. 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.
58. Silventoinen,K., Haukka,J., Dunkel,L., Tynelius,P. and Rasmussen,F. (2008) Genetics of pubertal timing and its associations with relative weight in childhood and adult height: the Swedish Young Male Twins Study. *Pediatrics*, **121**, e885–891.
59. D'Aloisio,A.A., DeRoo,L.A., Baird,D.D., Weinberg,C.R. and Sandler,D.P. (2013) Prenatal and infant exposures and age at menarche. *Epidemiol. Camb. Mass*, **24**, 277–284.
60. Morris,D.H., Jones,M.E., Schoemaker,M.J., Ashworth,A. and Swerdlow,A.J. (2010) Determinants of age at menarche in the UK: analyses from the Breakthrough Generations Study. *Br. J. Cancer*, **103**, 1760–1764.
61. Karaolis-Danckert,N., Buyken,A.E., Sonntag,A. and Kroke,A. (2009) Birth and early life influences on the timing of puberty onset: results from the DONALD (DOrtmund Nutritional and Anthropometric Longitudinally Designed) Study. *Am. J. Clin. Nutr.*, **90**, 1559–1565.

62. Günther,A.L.B., Karaolis-Danckert,N., Kroke,A., Remer,T. and Buyken,A.E. (2010) Dietary protein intake throughout childhood is associated with the timing of puberty. *J. Nutr.*, **140**, 565–571.
63. Mervish,N.A., Gardiner,E.W., Galvez,M.P., Kushi,L.H., Windham,G.C., Biro,F.M., Pinney,S.M., Rybak,M.E., Teitelbaum,S.L., Wolff,M.S., *et al.* (2013) Dietary flavonol intake is associated with age of puberty in a longitudinal cohort of girls. *Nutr. Res. N. Y. N*, **33**, 534–542.
64. Cheng,G., Remer,T., Prinz-Langenohl,R., Blaszkewicz,M., Degen,G.H. and Buyken,A.E. (2010) Relation of isoflavones and fiber intake in childhood to the timing of puberty. *Am. J. Clin. Nutr.*, **92**, 556–564.
65. Cheng,G., Buyken,A.E., Shi,L., Karaolis-Danckert,N., Kroke,A., Wudy,S.A., Degen,G.H. and Remer,T. (2012) Beyond overweight: nutrition as an important lifestyle factor influencing timing of puberty. *Nutr. Rev.*, **70**, 133–152.
66. Di Cagno,A., Marchetti,M., Battaglia,C., Giombini,A., Calcagno,G., Fiorilli,G., Piazza,M., Pigozzi,F. and Borrione,P. (2012) Is menstrual delay a serious problem for elite rhythmic gymnasts? *J. Sports Med. Phys. Fitness*, **52**, 647–653.
67. Sanchez-Garrido,M.A. and Tena-Sempere,M. (2013) Metabolic control of puberty: roles of leptin and kisspeptins. *Horm. Behav.*, **64**, 187–194.
68. Tena-Sempere,M. (2013) Interaction between energy homeostasis and reproduction: central effects of leptin and ghrelin on the reproductive axis. *Horm. Metab. Res. Horm. Stoffwechselforschung Horm. Métabolisme*, **45**, 919–927.
69. Morris,D.H., Jones,M.E., Schoemaker,M.J., Ashworth,A. and Swerdlow,A.J. (2011) Secular trends in age at menarche in women in the UK born 1908-93: results from the Breakthrough Generations Study. *Paediatr. Perinat. Epidemiol.*, **25**, 394–400.
70. Talma,H., Schönbeck,Y., van Dommelen,P., Bakker,B., van Buuren,S. and Hirasing,R.A. (2013) Trends in menarcheal age between 1955 and 2009 in the Netherlands. *PloS One*, **8**, e60056.
71. Biro,F.M., Greenspan,L.C., Galvez,M.P., Pinney,S.M., Teitelbaum,S., Windham,G.C., Deardorff,J., Herrick,R.L., Succop,P.A., Hiatt,R.A., *et al.* (2013) Onset of breast development in a longitudinal cohort. *Pediatrics*, **132**, 1019–1027.
72. Goldstein,J.R. (2011) A secular trend toward earlier male sexual maturity: evidence from shifting ages of male young adult mortality. *PloS One*, **6**, e14826.
73. Kryst,Ł., Kowal,M., Woronkowicz,A., Sobiecki,J. and Cichocka,B.A. (2012) Secular changes in height, body weight, body mass index and pubertal development in male children and adolescents in Krakow, Poland. *J. Biosoc. Sci.*, **44**, 495–507.
74. Herman-Giddens,M.E., Steffes,J., Harris,D., Slora,E., Hussey,M., Dowshen,S.A., Wasserman,R., Serwint,J.R., Smitherman,L. and Reiter,E.O. (2012) Secondary sexual characteristics in boys: data from the Pediatric Research in Office Settings Network. *Pediatrics*, **130**, e1058–1068.
75. Ma,H.-M., Chen,S.-K., Chen,R.-M., Zhu,C., Xiong,F., Li,T., Wang,W., Liu,G.-L., Luo,X.-P., Liu,L., *et al.* (2011) Pubertal development timing in urban Chinese boys. *Int. J. Androl.*, **34**, e435–445.
76. Mahachoklertwattana,P., Suthutvoravut,U., Poomthavorn,P., Charoenkiatkul,S., Udomsubpayakul,U. and Rajatanavin,R. (2010) Sexual maturation in Thai boys. *J. Pediatr. Endocrinol. Metab. JPEM*, **23**, 65–71.
77. Papadimitriou,A., Douros,K., Kleanthous,K., Papadimitriou,D.T., Attilakos,A. and Fretzayas,A. (2011) Pubertal maturation of contemporary Greek boys: no evidence of a secular trend. *J. Adolesc. Health Off. Publ. Soc. Adolesc. Med.*, **49**, 434–436.
78. Euling,S.Y., Herman-Giddens,M.E., Lee,P.A., Selevan,S.G., Juul,A., Sørensen,T.I.A., Dunkel,L., Himes,J.H., Teilmann,G. and Swan,S.H. (2008) Examination of US puberty-timing data from 1940 to 1994 for secular trends: panel findings. *Pediatrics*, **121 Suppl 3**, S172–191.

79. Bordini,B. and Rosenfield,R.L. (2011) Normal Pubertal Development: Part II: Clinical Aspects of Puberty. *Pediatr. Rev.*, **32**, 281–292.
80. Prentice,P. and Williams,R. (2013) Disorders of puberty. *Medicine (Baltimore)*, **41**, 604–606.
81. Bierich,J.R. (1992) Constitutional delay of growth and adolescence. *Baillières Clin. Endocrinol. Metab.*, **6**, 573–588.
82. Wolfenden,H. and Ryan,F. (2014) Delayed puberty. *Paediatr. Child Health*, **24**, 124–130.
83. Laitinen,E.-M., Vaaralahti,K., Tommiska,J., Eklund,E., Tervaniemi,M., Valanne,L. and Raivio,T. (2011) Incidence, Phenotypic Features and Molecular Genetics of Kallmann Syndrome in Finland. *Orphanet J. Rare Dis.*, **6**, 41.
84. Raivio,T., Falardeau,J., Dwyer,A., Quinton,R., Hayes,F.J., Hughes,V.A., Cole,L.W., Pearce,S.H., Lee,H., Boepple,P., *et al.* (2007) Reversal of Idiopathic Hypogonadotropic Hypogonadism. *N. Engl. J. Med.*, **357**, 863–873.
85. Bramswig,J. and Dubbers,A. (2009) Disorders of Pubertal Development. *Dtsch. Ärztebl. Int.*, **106**, 295–304.
86. Yousefi,M., Karmaus,W., Zhang,H., Roberts,G., Matthews,S., Clayton,B. and Arshad,S.H. (2013) Relationships between age of puberty onset and height at age 18 years in girls and boys. *World J. Pediatr. WJP*, **9**, 230–238.
87. Georgiadis,E., Mantzoros,C.S., Evagelopoulou,C. and Spentzos,D. (1997) Adult height and menarcheal age of young women in Greece. *Ann. Hum. Biol.*, **24**, 55–59.
88. Helm,P., Münster,K. and Schmidt,L. (1995) Recalled menarche in relation to infertility and adult weight and height. *Acta Obstet. Gynecol. Scand.*, **74**, 718–722.
89. Wehkalampi,K., Päkkilä,K., Laine,T. and Dunkel,L. (2011) Adult height in girls with delayed pubertal growth. *Horm. Res. Pædiatrics*, **76**, 130–135.
90. Crowne,E.C., Shalet,S.M., Wallace,W.H., Eminson,D.M. and Price,D.A. (1991) Final height in girls with untreated constitutional delay in growth and puberty. *Eur. J. Pediatr.*, **150**, 708–712.
91. Crowne,E.C., Shalet,S.M., Wallace,W.H., Eminson,D.M. and Price,D.A. (1990) Final height in boys with untreated constitutional delay in growth and puberty. *Arch. Dis. Child.*, **65**, 1109–1112.
92. Wehkalampi,K., Vangonen,K., Laine,T. and Dunkel,L. (2007) Progressive reduction of relative height in childhood predicts adult stature below target height in boys with constitutional delay of growth and puberty. *Horm. Res.*, **68**, 99–104.
93. Stovitz,S.D., Demerath,E.W., Hannan,P.J., Lytle,L.A. and Himes,J.H. (2011) Growing into obesity: Patterns of height growth in those who become normal weight, overweight, or obese as young adults. *Am. J. Hum. Biol.*, **23**, 635–641.
94. Fennoy,I. (2013) Effect of obesity on linear growth. *Curr. Opin. Endocrinol. Diabetes Obes.*, **20**, 44–49.
95. De Leonibus,C., Marcovecchio,M.L. and Chiarelli,F. (2012) Update on statural growth and pubertal development in obese children. *Pediatr. Rep.*, **4**, e35.
96. De Leonibus,C., Marcovecchio,M.L., Chiavaroli,V., de Giorgis,T., Chiarelli,F. and Mohn,A. (2013) Timing of puberty and physical growth in obese children: a longitudinal study in boys and girls. *Pediatr. Obes.*, 10.1111/j.2047-6310.2013.00176.x.
97. Marcovecchio,M.L. and Chiarelli,F. (2013) Obesity and growth during childhood and puberty. *World Rev. Nutr. Diet.*, **106**, 135–141.
98. Dunger,D.B., Ahmed,M.L. and Ong,K.K. (2006) Early and late weight gain and the timing of puberty. *Mol. Cell. Endocrinol.*, **254-255**, 140–145.
99. Burt Solorzano,C.M. and McCartney,C.R. (2010) Obesity and the pubertal transition in girls and boys. *Reprod. Camb. Engl.*, **140**, 399–410.
100. Ahmed,M.L., Ong,K.K. and Dunger,D.B. (2009) Childhood obesity and the timing of puberty. *Trends Endocrinol. Metab. TEM*, **20**, 237–242.
101. Graber,J.A., Nichols,T.R. and Brooks-Gunn,J. (2010) Putting pubertal timing in developmental context: implications for prevention. *Dev. Psychobiol.*, **52**, 254–262.

102. Anderson,C.A., Zhu,G., Falchi,M., van den Berg,S.M., Treloar,S.A., Spector,T.D., Martin,N.G., Boomsma,D.I., Visscher,P.M. and Montgomery,G.W. (2008) A genome-wide linkage scan for age at menarche in three populations of European descent. *J. Clin. Endocrinol. Metab.*, **93**, 3965–3970.
103. Golub,M.S., Collman,G.W., Foster,P.M.D., Kimmel,C.A., Rajpert-De Meyts,E., Reiter,E.O., Sharpe,R.M., Skakkebaek,N.E. and Toppari,J. (2008) Public Health Implications of Altered Puberty Timing. *PEDIATRICS*, **121**, S218–S230.
104. 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.
105. Eastell,R. (2005) Role of oestrogen in the regulation of bone turnover at the menarche. *J. Endocrinol.*, **185**, 223–234.
106. Bodicoat,D.H., Schoemaker,M.J., Jones,M.E., McFadden,E., Griffin,J., Ashworth,A. and Swerdlow,A.J. (2014) Timing of pubertal stages and breast cancer risk: the Breakthrough Generations Study. *Breast Cancer Res. BCR*, **16**, R18.
107. 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.
108. Ritte,R., Tikk,K., Lukanova,A., Tjønneland,A., Olsen,A., Overvad,K., Dossus,L., Fournier,A., Clavel-Chapelon,F., Grote,V., *et al.* (2013) Reproductive factors and risk of hormone receptor positive and negative breast cancer: a cohort study. *BMC Cancer*, **13**, 584.
109. Ritte,R., Lukanova,A., Tjønneland,A., Olsen,A., Overvad,K., Mesrine,S., Fagherazzi,G., Dossus,L., Teucher,B., Steindorf,K., *et al.* (2013) Height, age at menarche and risk of hormone receptor-positive and -negative breast cancer: a cohort study. *Int. J. Cancer J. Int. Cancer*, **132**, 2619–2629.
110. Gong,T.-T., Wu,Q.-J., Vogtmann,E., Lin,B. and Wang,Y.-L. (2013) Age at menarche and risk of ovarian cancer: a meta-analysis of epidemiological studies. *Int. J. Cancer J. Int. Cancer*, **132**, 2894–2900.
111. Richiardi,L., Vizzini,L., Pastore,G., Segnan,N., Gillio-Tos,A., Fiano,V., Grasso,C., Ciuffreda,L., Lista,P., Pearce,N., *et al.* (2013) Lifetime growth and risk of testicular cancer. *Int. J. Cancer J. Int. Cancer*, 10.1002/ijc.28688.
112. Karami,S., Daugherty,S.E., Schonfeld,S.J., Park,Y., Hollenbeck,A.R., Grubb,R.L.,3rd, Hofmann,J.N., Chow,W.-H. and Purdue,M.P. (2013) Reproductive factors and kidney cancer risk in 2 US cohort studies, 1993-2010. *Am. J. Epidemiol.*, **177**, 1368–1377.
113. He,C., Zhang,C., Hunter,D.J., Hankinson,S.E., Buck Louis,G.M., Hediger,M.L. and Hu,F.B. (2010) Age at menarche and risk of type 2 diabetes: results from 2 large prospective cohort studies. *Am. J. Epidemiol.*, **171**, 334–344.
114. Widén,E., Silventoinen,K., Sovio,U., Ripatti,S., Cousminer,D.L., Hartikainen,A.-L., Laitinen,J., Pouta,A., Kaprio,J., Järvelin,M.-R., *et al.* (2012) Pubertal timing and growth influences cardiometabolic risk factors in adult males and females. *Diabetes Care*, **35**, 850–856.
115. Gajdos,Z.K.Z., Henderson,K.D., Hirschhorn,J.N. and Palmert,M.R. (2010) Genetic determinants of pubertal timing in the general population. *Mol. Cell. Endocrinol.*, **324**, 21–29.
116. Cadman,S.M., Kim,S.-H., Hu,Y., González-Martínez,D. and Bouloux,P.-M. (2007) Molecular pathogenesis of Kallmann's syndrome. *Horm. Res.*, **67**, 231–242.
117. Dodé,C. and Hardelin,J.-P. (2009) Kallmann syndrome. *Eur. J. Hum. Genet. EJHG*, **17**, 139–146.
118. Schwanzel-Fukuda,M., Bick,D. and Pfaff,D.W. (1989) Luteinizing hormone-releasing hormone (LHRH)-expressing cells do not migrate normally in an inherited hypogonadal (Kallmann) syndrome. *Brain Res. Mol. Brain Res.*, **6**, 311–326.

119. Falardeau,J., Chung,W.C.J., Beenken,A., Raivio,T., Plummer,L., Sidis,Y., Jacobson-Dickman,E.E., Eliseenkova,A.V., Ma,J., Dwyer,A., *et al.* (2008) Decreased FGF8 signaling causes deficiency of gonadotropin-releasing hormone in humans and mice. *J. Clin. Invest.*, **118**, 2822–2831.
120. Dodé,C., Levilliers,J., Dupont,J.-M., De Paepe,A., Le Dû,N., Soussi-Yanicostas,N., Coimbra,R.S., Delmaghani,S., Compain-Nouaille,S., Baverel,F., *et al.* (2003) Loss-of-function mutations in FGFR1 cause autosomal dominant Kallmann syndrome. *Nat. Genet.*, **33**, 463–465.
121. Dodé,C., Teixeira,L., Levilliers,J., Fouveaut,C., Bouchard,P., Kottler,M.-L., Lespinasse,J., Lienhardt-Roussie,A., Mathieu,M., Moerman,A., *et al.* (2006) Kallmann syndrome: mutations in the genes encoding prokineticin-2 and prokineticin receptor-2. *PLoS Genet.*, **2**, e175.
122. Banerjee,I. and Clayton,P. (2007) The genetic basis for the timing of human puberty. *J. Neuroendocrinol.*, **19**, 831–838.
123. Layman,L.C., Cohen,D.P., Jin,M., Xie,J., Li,Z., Reindollar,R.H., Bolbolan,S., Bick,D.P., Sherins,R.R., Duck,L.W., *et al.* (1998) Mutations in gonadotropin-releasing hormone receptor gene cause hypogonadotropic hypogonadism. *Nat. Genet.*, **18**, 14–15.
124. Beneduzzi,D., Trarbach,E.B., Min,L., Jorge,A.A.L., Garmes,H.M., Renk,A.C., Fichna,M., Fichna,P., Arantes,K.A., Costa,E.M.F., *et al.* (2014) Role of gonadotropin-releasing hormone receptor mutations in patients with a wide spectrum of pubertal delay. *Fertil. Steril.*, 10.1016/j.fertnstert.2014.05.044.
125. Gajdos,Z.K., Hirschhorn,J.N. and Palmert,M.R. (2009) What controls the timing of puberty? An update on progress from genetic investigation. *Curr. Opin. Endocrinol. Diabetes Obes.*, **16**, 16–24.
126. Tena-Sempere,M. (2010) Kisspeptin signaling in the brain: recent developments and future challenges. *Mol. Cell. Endocrinol.*, **314**, 164–169.
127. Strobel,A., Issad,T., Camoin,L., Ozata,M. and Strosberg,A.D. (1998) A leptin missense mutation associated with hypogonadism and morbid obesity. *Nat. Genet.*, **18**, 213–215.
128. Fernandez-Fernandez,R., Martini,A.C., Navarro,V.M., Castellano,J.M., Dieguez,C., Aguilar,E., Pinilla,L. and Tena-Sempere,M. (2006) Novel signals for the integration of energy balance and reproduction. *Mol. Cell. Endocrinol.*, **254-255**, 127–132.
129. Welt,C.K., Chan,J.L., Bullen,J., Murphy,R., Smith,P., DePaoli,A.M., Karalis,A. and Mantzoros,C.S. (2004) Recombinant Human Leptin in Women with Hypothalamic Amenorrhea. *N. Engl. J. Med.*, **351**, 987–997.
130. Matkovic,V., Ilich,J.Z., Skugor,M., Badenhop,N.E., Goel,P., Clairmont,A., Klisovic,D., Nahhas,R.W. and Landoll,J.D. (1997) Leptin is inversely related to age at menarche in human females. *J. Clin. Endocrinol. Metab.*, **82**, 3239–3245.
131. Garcia-Mayor,R.V., Andrade,M.A., Rios,M., Lage,M., Dieguez,C. and Casanueva,F.F. (1997) Serum leptin levels in normal children: relationship to age, gender, body mass index, pituitary-gonadal hormones, and pubertal stage. *J. Clin. Endocrinol. Metab.*, **82**, 2849–2855.
132. Cunningham,M.J., Clifton,D.K. and Steiner,R.A. (1999) Leptin's actions on the reproductive axis: perspectives and mechanisms. *Biol. Reprod.*, **60**, 216–222.
133. Perry,J.R.B., Felix Day, Cathy E. Elks, Patrick Sulem, Deborah J. Thompson, Teresa Ferreira, Chunyan He, Daniel I. Chasman, Tõnu Esko, Gudmar Thorleifsson, Eva Albrecht, Wei Q. Ang, Tanguy Corre, Diana L. Cousminer, Bjarke Feenstra, Nora Franceschini, Andrea Ganna, Andrew D. Johnson, Sanela Kjellqvist, Kathryn L. Lunetta, George McMahon, Ilja M. Nolte, Lavinia Paternoster, Eleonora Porcu, Albert V. Smith, Lisette Stolk, Alexander Teumer, Natalia Tšernikova, Emmi Tikkanen, Sheila Ulivi, Erin K. Wagner, Najaf Amin, Laura J. Bierut, Enda M. Byrne, Jouke-Jan Hottenga, Daniel L. Koller, Massimo Mangino, Tune H. Pers, Laura M. Yerges-Armstrong, Jing Hua Zhao, Irene L. Andrulis, Hoda Anton-Culver, Femke Atsma, Stefania Bandinelli, Matthias W.

Beckmann, Javier Benitez, Carl Blomqvist, Stig E. Bojesen, Manjeet K. Bolla, Bernardo Bonanni, Hiltrud Brauch, Hermann Brenner, Julie E. Buring, Jenny Chang-Claude, Stephen Chanock, Jinhui Chen, Georgia Chenevix-Trench, J. Margriet Collée, Fergus J. Couch, David Couper, Andrea D. Coviello, Angela Cox, Kamila Czene, Adamo Pio D'adamo, George Davey Smith, Immaculata De Vivo, Ellen W. Demerath, Joe Dennis, Peter Devilee, Aida K. Dieffenbach, Alison M. Dunning, Gudny Eiriksdottir, Johan G. Eriksson, Peter A. Fasching, Luigi Ferrucci, Dieter Flesch-Janys, Henrik Flyger, Tatiana Foroud, Lude Franke, Melissa E. Garcia, Montserrat García-Closas, Frank Geller, Eco E. J. de Geus, Graham G. Giles, Daniel F. Gudbjartsson, Vilmundur Gudnason, Pascal Guénel, Suiqun Guo, Per Hall, Ute Hamann, Robin Haring, Catharina A. Hartman, Andrew C. Heath, Albert Hofman, Maartje J. Hooning, John L. Hopper, Frank B. Hu, David J. Hunter, David Karasik, Douglas P. Kiel, Julia A. Knight, Veli-Matti Kosma, Zoltan Kutalik, Sandra Lai, Diether Lambrechts, Annika Lindblom, Reedik Mägi, Patrik K. Magnusson, Arto Mannermaa, Nicholas G. Martin, Gisli Masson, Patrick F. McArdle, Wendy L. McArdle, Mads Melbye, Kyriaki Michailidou, Evelin Mihailov, Lili Milani, Roger L. Milne, Heli Nevanlinna, Patrick Neven, Ellen A. Nohr, Albertine J. Oldehinkel, Ben A. Oostra, Aarno Palotie, Munro Peacock, Nancy L. Pedersen, Paolo Peterlongo, Julian Peto, Paul D. P. Pharoah, Dirkje S. Postma, Anneli Pouta, Katri Pylkäs, Paolo Radice, Susan Ring, Fernando Rivadeneira, Antonietta Robino, Lynda M. Rose, Anja Rudolph, Veikko Salomaa, Serena Sanna, David Schlessinger, Marjanka K. Schmidt, Mellissa C. Southey, Ulla Sovio, Meir J. Stampfer, Doris Stöckl, Anna M. Storniolo, Nicholas J. Timpson, Jonathan Tyrer, Jenny A. Visser, Peter Vollenweider, Henry Völzke, Gerard Waeber, Melanie Waldenberger, Henri Wallaschofski, Qin Wang, Gonneke Willemsen, Robert Winqvist, Bruce H. R. Wolffenbuttel, Margaret J. Wright, Dorret I. Boomsma, Michael J. Econs, Kay-Tee Khaw, Ruth J. F. Loos, Mark I. McCarthy, Grant W. Montgomery, John P. Rice, Elizabeth A. Streeten, Unnur Thorsteinsdottir, Cornelia M. van Duijn, Behrooz Z. Alizadeh, Sven Bergmann, Eric Boerwinkle, Heather A. Boyd, Laura Crisponi, Paolo Gasparini, Christian Gieger, Tamara B. Harris, Erik Ingelsson, Marjo-Riitta Järvelin, Peter Kraft, Debbie Lawlor, Andres Metspalu, Craig E. Pennell, Paul M. Ridker, Harold Snieder, Thorkild I. A. Sørensen, Tim D. Spector, David P. Strachan, André G. Uitterlinden, Nicholas J. Wareham, Elisabeth Widen, Marek Zygmunt, Anna Murray, Douglas F. Easton, Kari Stefansson, Joanne M. Murabito & Ken K. Ong (2014) Parent-of-origin-specific allelic associations among 106 genomic loci for age at menarche. *Nature*, **advance online publication**.

134. 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.
135. Gajdos,Z.K.Z., Butler,J.L., Henderson,K.D., He,C., Supelak,P.J., Egyud,M., Price,A., Reich,D., Clayton,P.E., Le Marchand,L., *et al.* (2008) Association studies of common variants in 10 hypogonadotropic hypogonadism genes with age at menarche. *J. Clin. Endocrinol. Metab.*, **93**, 4290–4298.
136. Manolio,T.A., Collins,F.S., Cox,N.J., Goldstein,D.B., Hindorff,L.A., Hunter,D.J., McCarthy,M.I., Ramos,E.M., Cardon,L.R., Chakravarti,A., *et al.* (2009) Finding the missing heritability of complex diseases. *Nature*, **461**, 747–753.
137. Lin,L., Conway,G.S., Hill,N.R., Dattani,M.T., Hindmarsh,P.C. and Achermann,J.C. (2006) A homozygous R262Q mutation in the gonadotropin-releasing hormone receptor presenting as constitutional delay of growth and puberty with subsequent borderline oligospermia. *J. Clin. Endocrinol. Metab.*, **91**, 5117–5121.
138. Pugliese-Pires,P.N., Fortin,J.-P., Arthur,T., Latronico,A.C., Mendonca,B.B., Villares,S.M.F., Arnhold,I.J.P., Kopin,A.S. and Jorge,A.A.L. (2011) Novel inactivating mutations in the GH secretagogue receptor gene in patients with constitutional delay of growth and puberty. *Eur. J. Endocrinol. Eur. Fed. Endocr. Soc.*, **165**, 233–241.

139. Tusset,C., Noel,S.D., Trarbach,E.B., Silveira,L.F.G., Jorge,A.A.L., Brito,V.N., Cukier,P., Seminara,S.B., Mendonça,B.B. de, Kaiser,U.B., *et al.* (2012) Mutational analysis of TAC3 and TACR3 genes in patients with idiopathic central pubertal disorders. *Arq. Bras. Endocrinol. Metabol.*, **56**, 646–652.
140. Vaaralahti,K., Wehkalampi,K., Tommiska,J., Laitinen,E.-M., Dunkel,L. and Raivio,T. (2011) The role of gene defects underlying isolated hypogonadotropic hypogonadism in patients with constitutional delay of growth and puberty. *Fertil. Steril.*, **95**, 2756–2758.
141. Tommiska,J., Wehkalampi,K., Vaaralahti,K., Laitinen,E.-M., Raivio,T. and Dunkel,L. (2010) LIN28B in constitutional delay of growth and puberty. *J. Clin. Endocrinol. Metab.*, **95**, 3063–3066.
142. Sedlmeyer,I.L., Hirschhorn,J.N. and Palmert,M.R. (2002) Pedigree analysis of constitutional delay of growth and maturation: determination of familial aggregation and inheritance patterns. *J. Clin. Endocrinol. Metab.*, **87**, 5581–5586.
143. Wehkalampi,K., Widén,E., Laine,T., Palotie,A. and Dunkel,L. (2008) Patterns of inheritance of constitutional delay of growth and puberty in families of adolescent girls and boys referred to specialist pediatric care. *J. Clin. Endocrinol. Metab.*, **93**, 723–728.
144. Wehkalampi,K., Widén,E., Laine,T., Palotie,A. and Dunkel,L. (2008) Association of the timing of puberty with a chromosome 2 locus. *J. Clin. Endocrinol. Metab.*, **93**, 4833–4839.
145. Liu,Y.-Z., Guo,Y.-F., Wang,L., Tan,L.-J., Liu,X.-G., Pei,Y.-F., Yan,H., Xiong,D.-H., Deng,F.-Y., Yu,N., *et al.* (2009) Genome-wide association analyses identify SPOCK as a key novel gene underlying age at menarche. *PLoS Genet.*, **5**, e1000420.
146. Ong,K.K., Elks,C.E., Li,S., Zhao,J.H., Luan,J., Andersen,L.B., Bingham,S.A., Brage,S., Smith,G.D., Ekelund,U., *et al.* (2009) Genetic variation in LIN28B is associated with the timing of puberty. *Nat. Genet.*, **41**, 729–733.
147. Sulem,P., Gudbjartsson,D.F., Rafnar,T., Holm,H., Olafsdottir,E.J., Olafsdottir,G.H., Jonsson,T., Alexandersen,P., Feenstra,B., Boyd,H.A., *et al.* (2009) Genome-wide association study identifies sequence variants on 6q21 associated with age at menarche. *Nat. Genet.*, **41**, 734–738.
148. Perry,J.R.B., Stolk,L., Franceschini,N., Lunetta,K.L., Zhai,G., McArdle,P.F., Smith,A.V., Aspelund,T., Bandinelli,S., Boerwinkle,E., *et al.* (2009) Meta-analysis of genome-wide association data identifies two loci influencing age at menarche. *Nat. Genet.*, **41**, 648–650.
149. He,C., Kraft,P., Chen,C., Buring,J.E., Paré,G., Hankinson,S.E., Chanock,S.J., Ridker,P.M., Hunter,D.J. and Chasman,D.I. (2009) Genome-wide association studies identify loci associated with age at menarche and age at natural menopause. *Nat. Genet.*, **41**, 724–728.
150. Elks,C.E., Perry,J.R.B., Sulem,P., Chasman,D.I., Franceschini,N., He,C., Lunetta,K.L., Visser,J.A., Byrne,E.M., Cousminer,D.L., *et al.* (2010) Thirty new loci for age at menarche identified by a meta-analysis of genome-wide association studies. *Nat. Genet.*, **42**, 1077–1085.
151. Chen,C.T.L., Fernández-Rhodes,L., Brzyski,R.G., Carlson,C.S., Chen,Z., Heiss,G., North,K.E., Woods,N.F., Rajkovic,A., Kooperberg,C., *et al.* (2012) Replication of loci influencing ages at menarche and menopause in Hispanic women: the Women's Health Initiative SHARe Study. *Hum. Mol. Genet.*, **21**, 1419–1432.
152. Demerath,E.W., Liu,C.-T., Franceschini,N., Chen,G., Palmer,J.R., Smith,E.N., Chen,C.T.L., Ambrosone,C.B., Arnold,A.M., Bandera,E.V., *et al.* (2013) Genome-wide association study of age at menarche in African-American women. *Hum. Mol. Genet.*, 10.1093/hmg/ddt181.
153. Delahanty,R.J., Beeghly-Fadiel,A., Long,J.R., Gao,Y.T., Lu,W., Xiang,Y.B., Zheng,Y., Ji,B.T., Wen,W.Q., Cai,Q.Y., *et al.* (2013) Evaluation of GWAS-identified genetic variants for age at menarche among Chinese women. *Hum. Reprod. Oxf. Engl.*, **28**, 1135–1143.

154. Tanikawa,C., Okada,Y., Takahashi,A., Oda,K., Kamatani,N., Kubo,M., Nakamura,Y. and Matsuda,K. (2013) Genome wide association study of age at menarche in the Japanese population. *PloS One*, **8**, e63821.
155. Vrieze,S.I., McGue,M., Miller,M.B., Legrand,L.N., Schork,N.J. and Iacono,W.G. (2011) An assessment of the individual and collective effects of variants on height using twins and a developmentally informative study design. *PLoS Genet.*, **7**, e1002413.
156. Sovio,U., Bennett,A.J., Millwood,I.Y., Molitor,J., O'Reilly,P.F., Timpson,N.J., Kaakinen,M., Laitinen,J., Haukka,J., Pillas,D., *et al.* (2009) Genetic determinants of height growth assessed longitudinally from infancy to adulthood in the northern Finland birth cohort 1966. *PLoS Genet.*, **5**, e1000409.
157. Tommiska,J., Sørensen,K., Aksglaede,L., Koivu,R., Puhakka,L., Juul,A. and Raivio,T. (2011) LIN28B, LIN28A, KISS1, and KISS1R in idiopathic central precocious puberty. *BMC Res. Notes*, **4**, 363.
158. Chou,I.-C., Wang,C.-H., Lin,W.-D., Tsai,C.-H. and Tsai,F.-J. (2011) Association study in Taiwanese girls with precocious puberty. *J. Pediatr. Endocrinol. Metab. JPEM*, **24**, 103–104.
159. Li,Y., Willer,C.J., Ding,J., Scheet,P. and Abecasis,G.R. (2010) MaCH: using sequence and genotype data to estimate haplotypes and unobserved genotypes. *Genet. Epidemiol.*, **34**, 816–834.
160. Aulchenko,Y.S., Struchalin,M.V. and van Duijn,C.M. (2010) ProbABEL package for genome-wide association analysis of imputed data. *BMC Bioinformatics*, **11**, 134.
161. Browning,B.L. and Browning,S.R. (2009) A Unified Approach to Genotype Imputation and Haplotype-Phase Inference for Large Data Sets of Trios and Unrelated Individuals. *Am. J. Hum. Genet.*, **84**, 210–223.
162. Kapur,K., Johnson,T., Beckmann,N.D., Sehmi,J., Tanaka,T., Kutalik,Z., Styrkarsdottir,U., Zhang,W., Marek,D., Gudbjartsson,D.F., *et al.* (2010) Genome-Wide Meta-Analysis for Serum Calcium Identifies Significantly Associated SNPs near the Calcium-Sensing Receptor (CASR) Gene. *PLoS Genet.*, **6**.
163. Howie,B.N., Donnelly,P. and Marchini,J. (2009) A Flexible and Accurate Genotype Imputation Method for the Next Generation of Genome-Wide Association Studies. *PLoS Genet*, **5**, e1000529.
164. Purcell,S., Neale,B., Todd-Brown,K., Thomas,L., Ferreira,M.A.R., Bender,D., Maller,J., Sklar,P., de Bakker,P.I.W., Daly,M.J., *et al.* (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.*, **81**, 559–575.
165. Mägi,R. and Morris,A.P. (2010) GWAMA: software for genome-wide association meta-analysis. *BMC Bioinformatics*, **11**, 288.
166. Walters,R.G., Jacquemont,S., Valsesia,A., de Smith,A.J., Martinet,D., Andersson,J., Falchi,M., Chen,F., Andrieux,J., Lobbens,S., *et al.* (2010) A new highly penetrant form of obesity due to deletions on chromosome 16p11.2. *Nature*, **463**, 671–675.
167. Jacquemont,S., Reymond,A., Zufferey,F., Harewood,L., Walters,R.G., Kutalik,Z., Martinet,D., Shen,Y., Valsesia,A., Beckmann,N.D., *et al.* (2011) Mirror extreme BMI phenotypes associated with gene dosage at the chromosome 16p11.2 locus. *Nature*, **478**, 97–102.
168. Wang,K., Li,M., Hadley,D., Liu,R., Glessner,J., Grant,S.F.A., Hakonarson,H. and Bucan,M. (2007) PennCNV: an integrated hidden Markov model designed for high-resolution copy number variation detection in whole-genome SNP genotyping data. *Genome Res.*, **17**, 1665–1674.
169. Diskin,S.J., Li,M., Hou,C., Yang,S., Glessner,J., Hakonarson,H., Bucan,M., Maris,J.M. and Wang,K. (2008) Adjustment of genomic waves in signal intensities from whole-genome SNP genotyping platforms. *Nucleic Acids Res.*, **36**, e126.

170. Pietiläinen,O.P.H., Rehnström,K., Jakkula,E., Service,S.K., Congdon,E., Tilgmann,C., Hartikainen,A.-L., Taanila,A., Heikura,U., Paunio,T., *et al.* (2011) Phenotype mining in CNV carriers from a population cohort. *Hum. Mol. Genet.*, **20**, 2686–2695.
171. Marchini,J., Howie,B., Myers,S., McVean,G. and Donnelly,P. (2007) A new multipoint method for genome-wide association studies by imputation of genotypes. *Nat. Genet.*, **39**, 906–913.
172. Howie,B., Fuchsberger,C., Stephens,M., Marchini,J. and Abecasis,G.R. (2012) Fast and accurate genotype imputation in genome-wide association studies through pre-phasing. *Nat. Genet.*, **44**, 955–959.
173. Boyle,A.P., Hong,E.L., Hariharan,M., Cheng,Y., Schaub,M.A., Kasowski,M., Karczewski,K.J., Park,J., Hitz,B.C., Weng,S., *et al.* (2012) Annotation of functional variation in personal genomes using RegulomeDB. *Genome Res.*, **22**, 1790–1797.
174. Inouye,M., Kettunen,J., Soininen,P., Silander,K., Ripatti,S., Kumpula,L.S., Hämäläinen,E., Jousilahti,P., Kangas,A.J., Männistö,S., *et al.* (2010) Metabonomic, transcriptomic, and genomic variation of a population cohort. *Mol. Syst. Biol.*, **6**, 441.
175. Wright,F.A., Sullivan,P.F., Brooks,A.I., Zou,F., Sun,W., Xia,K., Madar,V., Jansen,R., Chung,W., Zhou,Y.-H., *et al.* (2014) Heritability and genomics of gene expression in peripheral blood. *Nat. Genet.*, **46**, 430–437.
176. Jansen,R., Batista,S., Brooks,A.I., Tischfield,J.A., Willemsen,G., van Grootheest,G., Hottenga,J.-J., Milaneschi,Y., Mbarek,H., Madar,V., *et al.* (2014) Sex differences in the human peripheral blood transcriptome. *BMC Genomics*, **15**, 33.
177. Yang,T.-P., Beazley,C., Montgomery,S.B., Dimas,A.S., Gutierrez-Arcelus,M., Stranger,B.E., Deloukas,P. and Dermitzakis,E.T. (2010) Genevar: a database and Java application for the analysis and visualization of SNP-gene associations in eQTL studies. *Bioinforma. Oxf. Engl.*, **26**, 2474–2476.
178. Stranger,B.E., Montgomery,S.B., Dimas,A.S., Parts,L., Stegle,O., Ingle,C.E., Sekowska,M., Smith,G.D., Evans,D., Gutierrez-Arcelus,M., *et al.* (2012) Patterns of cis regulatory variation in diverse human populations. *PLoS Genet.*, **8**, e1002639.
179. Nica,A.C., Parts,L., Glass,D., Nisbet,J., Barrett,A., Sekowska,M., Travers,M., Potter,S., Grundberg,E., Small,K., *et al.* (2011) The architecture of gene regulatory variation across multiple human tissues: the MuTHER study. *PLoS Genet.*, **7**, e1002003.
180. Reimand,J., Kull,M., Peterson,H., Hansen,J. and Vilo,J. (2007) g:Profiler--a web-based toolset for functional profiling of gene lists from large-scale experiments. *Nucleic Acids Res.*, **35**, W193–200.
181. Segrè,A.V., DIAGRAM Consortium, MAGIC investigators, Groop,L., Mootha,V.K., Daly,M.J. and Altshuler,D. (2010) Common inherited variation in mitochondrial genes is not enriched for associations with type 2 diabetes or related glycemic traits. *PLoS Genet.*, **6**.
182. Cousminer,D.L., Berry,D.J., Timpson,N.J., Ang,W., Thiering,E., Byrne,E.M., Taal,H.R., Huikari,V., Bradfield,J.P., Kerkhof,M., *et al.* (2013) Genome-wide association and longitudinal analyses reveal genetic loci linking pubertal height growth, pubertal timing and childhood adiposity. *Hum. Mol. Genet.*, 10.1093/hmg/ddt104.
183. Speliotes,E.K., Willer,C.J., Berndt,S.I., Monda,K.L., Thorleifsson,G., Jackson,A.U., Lango Allen,H., Lindgren,C.M., Luan,J., Mägi,R., *et al.* (2010) Association analyses of 249,796 individuals reveal 18 new loci associated with body mass index. *Nat. Genet.*, **42**, 937–948.
184. Bradfield,J.P., Taal,H.R., Timpson,N.J., Scherag,A., Lecoeur,C., Warrington,N.M., Hypponen,E., Holst,C., Valcarcel,B., Thiering,E., *et al.* (2012) A genome-wide association meta-analysis identifies new childhood obesity loci. *Nat. Genet.*, **44**, 526–531.
185. Fredriks,A.M., van Buuren,S., Burgmeijer,R.J., Meulmeester,J.F., Beuker,R.J., Brugman,E., Roede,M.J., Verloove-Vanhorick,S.P. and Wit,J.M. (2000) Continuing positive secular growth change in The Netherlands 1955-1997. *Pediatr. Res.*, **47**, 316–323.

186. Sulonen,A.-M., Ellonen,P., Almusa,H., Lepistö,M., Eldfors,S., Hannula,S., Miettinen,T., Tyynismaa,H., Salo,P., Heckman,C., *et al.* (2011) Comparison of solution-based exome capture methods for next generation sequencing. *Genome Biol.*, **12**, R94.
187. Weckx,S., Del-Favero,J., Rademakers,R., Claes,L., Cruts,M., De Jonghe,P., Van Broeckhoven,C. and De Rijk,P. (2005) novoSNP, a novel computational tool for sequence variation discovery. *Genome Res.*, **15**, 436–442.
188. Karolchik,D., Barber,G.P., Casper,J., Clawson,H., Cline,M.S., Diekhans,M., Dreszer,T.R., Fujita,P.A., Guruvadoo,L., Haeussler,M., *et al.* (2014) The UCSC Genome Browser database: 2014 update. *Nucleic Acids Res.*, **42**, D764–770.
189. Kumar,P., Henikoff,S. and Ng,P.C. (2009) Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nat. Protoc.*, **4**, 1073–1081.
190. Adzhubei,I.A., Schmidt,S., Peshkin,L., Ramensky,V.E., Gerasimova,A., Bork,P., Kondrashov,A.S. and Sunyaev,S.R. (2010) A method and server for predicting damaging missense mutations. *Nat. Methods*, **7**, 248–249.
191. McLaren,W., Pritchard,B., Rios,D., Chen,Y., Flicek,P. and Cunningham,F. (2010) Deriving the consequences of genomic variants with the Ensembl API and SNP Effect Predictor. *Bioinforma. Oxf. Engl.*, **26**, 2069–2070.
192. Dai,R. and Ahmed,S.A. (2014) Sexual dimorphism of miRNA expression: a new perspective in understanding the sex bias of autoimmune diseases. *Ther. Clin. Risk Manag.*, **10**, 151–163.
193. Rybak,A., Fuchs,H., Smirnova,L., Brandt,C., Pohl,E.E., Nitsch,R. and Wulczyn,F.G. (2008) A feedback loop comprising lin-28 and let-7 controls pre-let-7 maturation during neural stem-cell commitment. *Nat. Cell Biol.*, **10**, 987–993.
194. Zhu,H., Shah,S., Shyh-Chang,N., Shinoda,G., Einhorn,W.S., Viswanathan,S.R., Takeuchi,A., Grasemann,C., Rinn,J.L., Lopez,M.F., *et al.* (2010) Lin28a transgenic mice manifest size and puberty phenotypes identified in human genetic association studies. *Nat. Genet.*, **42**, 626–630.
195. Lettre,G., Jackson,A.U., Gieger,C., Schumacher,F.R., Berndt,S.I., Sanna,S., Eyheramendy,S., Voight,B.F., Butler,J.L., Guiducci,C., *et al.* (2008) Identification of ten loci associated with height highlights new biological pathways in human growth. *Nat. Genet.*, **40**, 584–591.
196. Sebastian,A., Matsushita,T., Kawanami,A., Mackem,S., Landreth,G.E. and Murakami,S. (2011) Genetic inactivation of ERK1 and ERK2 in chondrocytes promotes bone growth and enlarges the spinal canal. *J. Orthop. Res. Off. Publ. Orthop. Res. Soc.*, **29**, 375–379.
197. Liu,F., Austin,D.A., Mellon,P.L., Olefsky,J.M. and Webster,N.J.G. (2002) GnRH activates ERK1/2 leading to the induction of c-fos and LHbeta protein expression in LbetaT2 cells. *Mol. Endocrinol. Baltim. Md*, **16**, 419–434.
198. Yao,B., Liu,H.-Y., Gu,Y.-C., Shi,S.-S., Tao,X.-Q., Li,X.-J., Ge,Y.-F., Cui,Y.-X. and Yang,G.-B. (2011) Gonadotropin-releasing hormone positively regulates steroidogenesis via extracellular signal-regulated kinase in rat Leydig cells. *Asian J. Androl.*, **13**, 438–445.
199. Pique-Regi,R., Degner,J.F., Pai,A.A., Gaffney,D.J., Gilad,Y. and Pritchard,J.K. (2011) Accurate inference of transcription factor binding from DNA sequence and chromatin accessibility data. *Genome Res.*, **21**, 447–455.
200. Matys,V., Kel-Margoulis,O.V., Fricke,E., Liebich,I., Land,S., Barre-Dirrie,A., Reuter,I., Chekmenev,D., Krull,M., Hornischer,K., *et al.* (2006) TRANSFAC and its module TRANSCompel: transcriptional gene regulation in eukaryotes. *Nucleic Acids Res.*, **34**, D108–110.
201. Valentino,T., Palmieri,D., Vitiello,M., Simeone,A., Palma,G., Arra,C., Chieffi,P., Chiariotti,L., Fusco,A. and Fedele,M. (2013) Embryonic defects and growth alteration in mice with homozygous disruption of the Patz1 gene. *J. Cell. Physiol.*, **228**, 646–653.

202. Buckingham,M. and Relaix,F. (2007) The role of Pax genes in the development of tissues and organs: Pax3 and Pax7 regulate muscle progenitor cell functions. *Annu. Rev. Cell Dev. Biol.*, **23**, 645–673.
203. Fedele,M., Benvenuto,G., Pero,R., Majello,B., Battista,S., Lembo,F., Vollono,E., Day,P.M., Santoro,M., Lania,L., *et al.* (2000) A novel member of the BTB/POZ family, PATZ, associates with the RNF4 RING finger protein and acts as a transcriptional repressor. *J. Biol. Chem.*, **275**, 7894–7901.
204. Kobayashi,A., Yamagiwa,H., Hoshino,H., Muto,A., Sato,K., Morita,M., Hayashi,N., Yamamoto,M. and Igarashi,K. (2000) A combinatorial code for gene expression generated by transcription factor Bach2 and MAZR (MAZ-related factor) through the BTB/POZ domain. *Mol. Cell. Biol.*, **20**, 1733–1746.
205. Fedele,M., Franco,R., Salvatore,G., Paronetto,M.P., Barbagallo,F., Pero,R., Chiariotti,L., Sette,C., Tramontano,D., Chieffi,G., *et al.* (2008) PATZ1 gene has a critical role in the spermatogenesis and testicular tumours. *J. Pathol.*, **215**, 39–47.
206. Pero,R., Lembo,F., Palmieri,E.A., Vitiello,C., Fedele,M., Fusco,A., Bruni,C.B. and Chiariotti,L. (2002) PATZ attenuates the RNF4-mediated enhancement of androgen receptor-dependent transcription. *J. Biol. Chem.*, **277**, 3280–3285.
207. Walters,K.A., Simanainen,U. and Handelsman,D.J. (2010) Molecular insights into androgen actions in male and female reproductive function from androgen receptor knockout models. *Hum. Reprod. Update*, **16**, 543–558.
208. Widén,E., Ripatti,S., Cousminer,D.L., Surakka,I., Lappalainen,T., Järvelin,M.-R., Eriksson,J.G., Raitakari,O., Salomaa,V., Sovio,U., *et al.* (2010) Distinct variants at LIN28B influence growth in height from birth to adulthood. *Am. J. Hum. Genet.*, **86**, 773–782.
209. Paternoster,L., Howe,L.D., Tilling,K., Weedon,M.N., Freathy,R.M., Frayling,T.M., Kemp,J.P., Smith,G.D., Timpson,N.J., Ring,S.M., *et al.* (2011) Adult height variants affect birth length and growth rate in children. *Hum. Mol. Genet.*, **20**, 4069–4075.
210. McIntyre,M.H. (2011) Adult stature, body proportions and age at menarche in the United States National Health and Nutrition Survey (NHANES) III. *Ann. Hum. Biol.*, **38**, 716–720.
211. McIntyre,M.H. and Kacerosky,P.M. (2011) Age and size at maturity in women: a norm of reaction? *Am. J. Hum. Biol. Off. J. Hum. Biol. Counc.*, **23**, 305–312.
212. Rollof,L. and Elfving,M. (2012) Evaluation of self-assessment of pubertal maturation in boys and girls using drawings and orchidometer. *J. Pediatr. Endocrinol. Metab. JPEM*, **25**, 125–129.
213. Desmangles,J.-C., Lappe,J.M., Lipaczewski,G. and Haynatzki,G. (2006) Accuracy of pubertal Tanner staging self-reporting. *J. Pediatr. Endocrinol. Metab. JPEM*, **19**, 213–221.
214. Koenis,M.M.G., Brouwer,R.M., van Baal,G.C.M., van Soelen,I.L.C., Peper,J.S., van Leeuwen,M., Delemarre-van de Waal,H.A., Boomsma,D.I. and Hulshoff Pol,H.E. (2013) Longitudinal study of hormonal and physical development in young twins. *J. Clin. Endocrinol. Metab.*, **98**, E518–527.
215. Wagner,I.V., Sabin,M.A., Pfäffle,R.W., Hiemisch,A., Sergeyev,E., Körner,A. and Kiess,W. (2012) Effects of obesity on human sexual development. *Nat. Rev. Endocrinol.*, **8**, 246–254.
216. Fernández-Rhodes,L., Demerath,E.W., Cousminer,D.L., Tao,R., Dreyfus,J.G., Esko,T., Smith,A.V., Gudnason,V., Harris,T.B., Launer,L., *et al.* (2013) Association of Adiposity Genetic Variants With Menarche Timing in 92,105 Women of European Descent. *Am. J. Epidemiol.*, 10.1093/aje/kws473.
217. Holm,C., Osterlund,T., Laurell,H. and Contreras,J.A. (2000) Molecular mechanisms regulating hormone-sensitive lipase and lipolysis. *Annu. Rev. Nutr.*, **20**, 365–393.

218. Osuga,J., Ishibashi,S., Oka,T., Yagyu,H., Tozawa,R., Fujimoto,A., Shionoiri,F., Yahagi,N., Kraemer,F.B., Tsutsumi,O., *et al.* (2000) Targeted disruption of hormone-sensitive lipase results in male sterility and adipocyte hypertrophy, but not in obesity. *Proc. Natl. Acad. Sci. U. S. A.*, **97**, 787–792.
219. Che,J., Lu,Y.-W., Sun,K.-K., Feng,C., Dong,A.-J. and Jiao,Y. (2013) Overexpression of TOB1 confers radioprotection to bronchial epithelial cells through the MAPK/ERK pathway. *Oncol. Rep.*, **30**, 637–642.
220. Sun,K.-K., Zhong,N., Yang,Y., Zhao,L. and Jiao,Y. (2013) Enhanced radiosensitivity of NSCLC cells by transducer of erbB2.1 (TOB1) through modulation of the MAPK/ERK pathway. *Oncol. Rep.*, **29**, 2385–2391.
221. Lin,S., Zhu,Q., Xu,Y., Liu,H., Zhang,J., Xu,J., Wang,H., Sang,Q., Xing,Q. and Fan,J. (2012) The role of the TOB1 gene in growth suppression of hepatocellular carcinoma. *Oncol. Lett.*, **4**, 981–987.
222. Viswanathan,S.R., Powers,J.T., Einhorn,W., Hoshida,Y., Ng,T.L., Toffanin,S., O'Sullivan,M., Lu,J., Phillips,L.A., Lockhart,V.L., *et al.* (2009) Lin28 promotes transformation and is associated with advanced human malignancies. *Nat. Genet.*, **41**, 843–848.
223. Ye,Y., Madison,B., Wu,X. and Rustgi,A.K. (2012) A LIN28B polymorphism predicts for colon cancer survival. *Cancer Biol. Ther.*, **13**, 1390–1395.
224. Santen,R.J., Song,R.X., McPherson,R., Kumar,R., Adam,L., Jeng,M.-H. and Yue,W. (2002) The role of mitogen-activated protein (MAP) kinase in breast cancer. *J. Steroid Biochem. Mol. Biol.*, **80**, 239–256.
225. Jerjees,D.A., Alabdullah,M., Alkaabi,M., Abduljabbar,R., Muftah,A., Nolan,C., Green,A.R., Ellis,I.O. and Rakha,E.A. (2014) ERK1/2 is related to oestrogen receptor and predicts outcome in hormone-treated breast cancer. *Breast Cancer Res. Treat.*, 10.1007/s10549-014-3066-8.
226. He,C., Kraft,P., Chasman,D.I., Buring,J.E., Chen,C., Hankinson,S.E., Paré,G., Chanock,S., Ridker,P.M. and Hunter,D.J. (2010) A large-scale candidate gene association study of age at menarche and age at natural menopause. *Hum. Genet.*, **128**, 515–527.
227. Constantin,S., Piet,R., Iremonger,K., Hwa Yeo,S., Clarkson,J., Porteous,R. and Herbison,A.E. (2012) GnRH neuron firing and response to GABA in vitro depend on acute brain slice thickness and orientation. *Endocrinology*, **153**, 3758–3769.
228. Jakkula,E., Rehnström,K., Varilo,T., Pietiläinen,O.P.H., Paunio,T., Pedersen,N.L., deFaire,U., Järvelin,M.-R., Saharinen,J., Freimer,N., *et al.* (2008) The genome-wide patterns of variation expose significant substructure in a founder population. *Am. J. Hum. Genet.*, **83**, 787–794.
229. Vaughan,K.T., Mikami,A., Paschal,B.M., Holzbaur,E.L., Hughes,S.M., Echeverri,C.J., Moore,K.J., Gilbert,D.J., Copeland,N.G., Jenkins,N.A., *et al.* (1996) Multiple mouse chromosomal loci for dynein-based motility. *Genomics*, **36**, 29–38.
230. Sewer,M.B. and Li,D. (2008) Regulation of steroid hormone biosynthesis by the cytoskeleton. *Lipids*, **43**, 1109–1115.
231. Maiti,A.K., Mattéi,M.G., Jorissen,M., Volz,A., Zeigler,A. and Bouvagnet,P. (2000) Identification, tissue specific expression, and chromosomal localisation of several human dynein heavy chain genes. *Eur. J. Hum. Genet. EJHG*, **8**, 923–932.
232. Majdoubi,M.E., Sahu,A., Ramaswamy,S. and Plant,T.M. (2000) Neuropeptide Y: A hypothalamic brake restraining the onset of puberty in primates. *Proc. Natl. Acad. Sci. U. S. A.*, **97**, 6179–6184.
233. Błogowska,A., Rzepka-Górska,I. and Krzyżanowska-Świniarska,B. (2004) Is neuropeptide Y responsible for constitutional delay of puberty in girls? A preliminary report. *Gynecol. Endocrinol.*, **19**, 22–25.

234. McCarthy,D.J., Humburg,P., Kanapin,A., Rivas,M.A., Gaulton,K., Consortium,T.W., Cazier,J.-B. and Donnelly,P. (2014) Choice of transcripts and software has a large effect on variant annotation. *Genome Med.*, **6**, 26.