Prevalence of deleterious variants in *MC3R* in patients with constitutional delay of growth and puberty

Katie Duckett[1,2], Alice Williamson[1,2], John W.R. Kincaid[1], Kara Rainbow[1], Laura J. Corbin[3,4], Hilary C. Martin[5], Ruth Y. Eberhardt[5], Qin Qin Huang[5], Matthew E. Hurles[5], Wen He[6], Raja Brauner[7], Angela Delaney[8], Leo Dunkel[9], Romina P. Grinspon[10], Janet E. Hall[11], Joel N. Hirschhorn[6], Sasha R. Howard[9], Ana C. Latronico[12], Alexander A.L. Jorge[12], Ken McElreavey[13], Verónica Mericq[14], Paulina M. Merino[14], Mark R. Palmert[15], Lacey Plummer[16], Rodolfo A. Rey[10], Raíssa C. Rezende[12], Stephanie B. Seminara[16], Kathryn Salnikov[16], Indraneel Banerjee[17], Brian Y.H. Lam[1], John R.B. Perry[1,2], Nicholas J. Timpson[3,4], Peter Clayton[18], Yee-Ming Chan[6], Ken K. Ong[2,19*] & Stephen O'Rahilly[1*]

[1]Wellcome-MRC Institute of Metabolic Science–Metabolic Research Laboratories, University of Cambridge, Cambridge, UK

[2]MRC Epidemiology Unit, Wellcome–MRC Institute of Metabolic Science, University of Cambridge, Cambridge, UK

[3]MRC Integrative Epidemiology Unit at University of Bristol, Bristol BS8 2BN, UK

[4]Population Health Sciences, Bristol Medical School, University of Bristol, Bristol BS8 2BN, UK

[5]Human Genetics Programme, Wellcome Sanger Institute, Wellcome Genome Campus, Hinxton, UK

[6]Division of Endocrinology, Department of Pediatrics, Boston Children's Hospital, Boston, USA; Department of Pediatrics, Harvard Medical School, Boston, USA; Broad Institute of Harvard and MIT, Cambridge, USA

[7]Hôpital Fondation Adolphe de Rothschild and Université Paris Cité, Paris, France

[8]Division of Endocrinology, St. Jude Children's Research Hospital, Memphis, TN, USA

[9]Centre for Endocrinology, William Harvey Research Institute Barts and the London School of Medicine and Dentistry, Queen Mary University of London, London UK

[10]Centro de Investigaciones Endocrinolègicas "Dr. César Bergadá" (CEDIE), CONICET - FEI - Divisièn de Endocrinología, Hospital de Niños Ricardo Gutiérrez, Buenos Aires, Argentina

[11]Division of Intramural Research, National Institute of Environmental Health Sciences, NIH, Research Triangle Park, NC, USA

[12]Unidade de Endocrinologia do Desenvolvimento, Laboratório de Hormônios e Genética Molecular LIM/42, Hospital das Clínicas da Faculdade de Medicina da Universidade de São Paulo, São Paulo, Brasil

[13]Human Developmental Genetics, CNRS UMR3738, Institut Pasteur, Paris, France

[14]Institute of Maternal and Child Research (IDIMI), Faculty of Medicine, University of Chile, Santiago, Chile

[15]Division of Endocrinology, Hospital for Sick Children; Departments of Pediatrics and Physiology, University of Toronto, Toronto, ON, Canada

[16]MGH-Harvard Center for Reproductive Medicine, Massachusetts General Hospital, Boston, MA, USA

[17]Department of Paediatric Endocrinology, Royal Manchester Children's Hospital, Manchester UK

[18] Developmental Biology & Medicine, Faculty of Biology, Medicine & Health, University of Manchester, Manchester, UK

[19]Department of Paediatrics, University of Cambridge, Cambridge, UK

* Equal contributions

Correspondence to so104@medschl.cam.ac.uk (https://orcid.org/0000-0003-2199-4449) or ken.ong@mrc-epid.cam.ac.uk (https://orcid.org/0000-0003-4689-7530).

Disclosure Summary

S.O. has undertaken remunerated consultancy work for Pfizer, AstraZeneca, GSK, ERX Pharmaceuticals and Third Rock Ventures. M.E.H. is a co-founder, consultant and non-executive director of Congenica Ltd. J.R.B.P is an employee and shareholder of Adrestia Therapeutics. A.A.L.J. has received an independent research grant from BioMarin and consulting fees from Novo Nordisk. V.M. is an associate editor of JCEM. All other authors have nothing to disclose.

Abstract

Context: The melanocortin 3 receptor (MC3R) has recently emerged as a critical regulator of pubertal timing, linear growth and the acquisition of lean mass in humans and mice. In population-based studies, heterozygous carriers of deleterious variants in *MC3R* report a later onset of puberty than non-carriers. However, the frequency of such variants in patients who present with clinical disorders of pubertal development is currently unknown.

Objective: To determine whether deleterious *MC3R* variants are more frequently found in patients clinically presenting with constitutional delay of growth and puberty (CDGP) or normosmic idiopathic hypogonadotropic hypogonadism (nIHH).

Design, Setting and Participants: We examined the sequence of *MC3R* in 362 adolescents with a clinical diagnosis of CDGP and 657 patients with nIHH, experimentally characterised the signalling properties of all non-synonymous variants found and compared their frequency to that in 5774 controls from a population-based cohort. Additionally, we established the relative frequency of predicted deleterious variants in individuals with self-reported delayed vs normally timed menarche/voice breaking in the UK Biobank cohort.

Results: *MC3R* loss-of-function variants were infrequent but overrepresented in patients with CDGP (8/362 (2.2%), OR=4.17, p=0.001). There was no strong evidence of overrepresentation in patients with nIHH (4/657 (0.6%), OR=1.15, p=0.779). In 246,328 women from UK Biobank, predicted deleterious variants were more frequently found in those self-reporting delayed (≥16 years) vs normal age at menarche (OR=1.66, p=3.90E-07).

Conclusions: We have found evidence that functionally damaging variants in *MC3R* are overrepresented in individuals with CDGP but are not a common cause of this phenotype.

Keywords: delayed puberty, constitutional delay, ALSPAC, UK Biobank, MC3R

Introduction

Delayed puberty is a common clinical presentation which often causes psychological distress to affected individuals and is associated with adverse health outcomes later in life (1,2). Its most common form is self-limited delayed puberty, in which pubertal onset occurs at the extreme end of normal and before age 18 years and is followed by normal reproductive hormone activity (3). This pattern may be accompanied by reduced linear growth that is already apparent from early childhood ('constitutional delay of growth and puberty', CDGP). In contrast, individuals with hypogonadotropic hypogonadism (HH) are unable to progress through puberty without hormone replacement therapy. Self-limited delayed puberty is often difficult to distinguish from HH during adolescence. Therefore, understanding the aetiology may allow the earlier diagnosis of delayed puberty and HH, as well as inform the development of new treatments (4).

Delayed puberty is highly heritable, based on analysis of affected pedigrees. The pattern of inheritance is often autosomal dominant with incomplete penetrance (5–9). Rare variants in genes that regulate hypothalamic-pituitary function and embryonic migration of gonadotrophin-releasing hormone neurons have been identified in patients with HH (8–12). Furthermore, genome-wide association studies in large population-based cohorts have identified hundreds of common variants associated with normal variation in puberty timing in both males and females (13,14), and these also implicate many genes involved in the development and functioning of the hypothalamic-pituitary-gonadal (HPG) axis.

Protein-altering variants in the melanocortin 3 receptor gene (*MC3R*) were recently shown to confer later puberty timing in males and females from UK population-based cohorts, and a single male who was homozygous for a severely disruptive variant entered puberty after 20 years of age (15). *MC3R* is highly

expressed in the KNDy neurons of the hypothalamic arcuate nucleus, which produce kisspeptin, the key central activator of the sex hormone axis (15–18). We proposed that MC3R signalling via the leptin-POMC-melanocortin pathway is a permissive signal that activates pubertal hormone activity in response to adequate nutritional status (15).

Here, we hypothesized that deleterious *MC3R* variants might also be implicated in the aetiology of CDGP. To test this hypothesis, we first experimentally characterised all *MC3R* protein-coding variants which have not been previously reported in a population-based cohort, the Avon Longitudinal Study of Parents and Children (ALSPAC). We then further experimentally characterised all *MC3R* protein-coding variants in 2 cohorts of patients with CDGP and a cohort of patients with normosmic idiopathic HH (nIHH) and used these results to compare the frequency of deleterious variants to the background population. Finally, we set out to determine whether individuals at the extreme ends of self-reported pubertal timing within the general population are more likely to carry a deleterious *MC3R* variant, using the UK Biobank population-based cohort.

Methods

Patient Cohorts and Genetic Sequencing

Three cohorts of patients were examined:

- 278 individuals (169 males) of diverse ancestries clinically diagnosed with CDGP from the Delayed Puberty Genetic Consortium (DPGen) for whom whole-exome sequence (WES) data were available (with sequencing performed as described in Zhu et al (8)). Inclusion criteria for a CDGP diagnosis are as described in Jonsdottir-Lewis et al (19), with the additional criterion that patients did not have a relative diagnosed with IHH.

- 84 individuals (80 males) of majority European ancestry clinically diagnosed with CDGP in Manchester, UK (20) in whom targeted sequencing of *MC3R* was performed. Full inclusion criteria are described in Banerjee et al (20).

- 657 individuals clinically diagnosed with nIHH (18 years or older, sex steroids below the adult reference range, gonadotropins not elevated above the reference range). Patients with nIHH were enrolled in a genetic study at Massachusetts General Hospital as previously described (21). WES was performed using the Broad Institute Genomics platform in the 657 patients, and variant calling was performed as previously described (22).

All human subjects research was approved by the ethics boards of each institution, and written informed consent was obtained for all study participants.

Population Cohorts and Genetic Sequencing

ALSPAC cohort

Individuals from Generation 1 (G1) of the UK Avon Longitudinal Study of Parents and Children (ALSPAC) birth cohort study were included as an unselected population cohort of majority European descent (23–25). Ethical approval for the study was obtained from the ALSPAC Ethics and Law Committee and the Local Research Ethics Committees. Consent for biological samples has been collected in accordance with the Human Tissue Act (2004).

Two sources of genetic data were used in the estimation of *MC3R* deleterious variant frequencies. Firstly, a pooled high-throughput sequencing approach was used to identify all non-synonymous coding variants in *MC3R* in 5774 individuals for which DNA was available (as described in Lam et al (15)). Briefly, 5992 individual DNA samples (representing 5774 unique individuals) were combined into pools of 50 and sequenced at the *MC3R* locus. Individual DNA present in pools containing three identified complete

loss-of-function (CLoF) variants were Sanger sequenced to determine the number of carriers of CloF variants. Secondly, to determine the frequency of a partial loss-of-function (PloF) missense mutation in *MC3R*, R220S (rs61735259), this variant was assessed in the wider ALSPAC cohort using a combination of WES data (available for 8301 participants) and Haplotype Reference Consortium (HRC) -imputed data (available for an additional 1665 participants). A description of the derivation and quality control of WES data can be found in the Supplementary Materials (26).

High concordance was observed between carriers identified in the ALSPAC HRC-imputed dataset and carriers identified in the WES data (Supplementary Table 1 (26)). All individuals who were labelled as carriers in the HRC dataset had this status confirmed in the WES data (where they existed). Given high concordance, genotype data were combined to give a maximised final genotype designation.

UK Biobank cohort

436,865 unrelated individuals (246,328 women and 190,547 men) of majority European descent for whom age at menarche or voice breaking data was available were included in this analysis (27). WES data for these individuals underwent variant-level quality control as described in Gardner et al (28). No other exclusion criteria were applied.

Sequencing the *MC3R* locus

For 90 CDGP patients without WES data, DNA was amplified in the *MC3R* coding region and Sanger sequenced as detailed in Lam et al, 2021. For full coverage of the *MC3R* coding sequence, the following primers were used:

MC3R_F1: 5'-TGGAACAGCAAAGTTCTCCCT-3'

MC3R_R3: 5'-CCTCACGTGGATGGAAAGTC-3'

MC3R_2F: 5'-CAGCATCATGACCGTGAGGAA-3'

MC3R_1R: 5'-CGAAGGTCAGGTAGTCGCTG-3'

MC3R_2R: 5'-TGCATGAGTGTTGCTGTGGG-3'

Functional assays for cyclic AMP activity

MC3R coding variants were cloned into a wild-type (WT) human N-FLAG-MC3R pcDNA3.1(+) vector using site-directed mutagenesis (Agilent Quikchange Lightning kit). HEK293 cells (ATCC) were plated into a 96-well plate (Greiner) at a density of 20,000 cells per well and transfected with 10ng WT or variant construct in duplicate using Lipofectamine 3000 (Invitrogen) as per manufacturer's protocol. After 48h, cells were treated with a final concentration range of 10^{-14} to 10^{-5}M NDP-α-MSH (Bachem) in phosphate buffered saline for 2h. cAMP levels were measured using Eurofins DiscoverX Hithunter cAMP assay for Small Molecules as per manufacturer's protocol. Luminescence readings were taken on a Tecan M1000 Pro plate reader.

For all constructs, raw luminescence values were converted to cAMP concentration (µM) using a cAMP standard curve (approx. 0-10µM) and three-parameter dose-response curves were fitted using Graphpad Prism v.6.0 to calculate V_{max} and $logEC_{50}$ values for WT/variants. All values were normalised to WT V_{max} (100%) and replotted for LoF determination. For both V_{max} and $logEC_{50}$ non-normalised values, statistical difference from WT was determined using two-way ANOVA for all variants except I50T, with variant, plate number and day of assay as covariates in the model, followed by post-hoc Dunnet's test to determine significance for each variant (calculated using R v4.1.3). I50T was analysed separately using the same workflow as it was assayed at a different time separately to all other variants.

Variants were classified as WT-like (75% WT<V_{max}≤120% WT or 0.2xWT≤EC_{50}<5xWT), PloF (25% WT<V_{max}≤75% WT or 5xWT≤EC_{50}<50xWT), or CloF (V_{max}≤25% WT or EC_{50}≥50xWT).

Odds ratios calculations in patient cohorts

The number of carriers of functionally characterised *MC3R* LoF variants was counted for each of the patient cohorts and the unselected ALSPAC cohort. Exact odds ratios for carrying PloF, CloF or combined PloF/CloF were calculated in the patient cohorts compared to ALSPAC and p-values calculated using Fisher's Exact test (*fisher.test()* function in base R *stats* package v4.1.0) .

Odds ratios calculations in UK Biobank

MC3R protein-altering variants present in UK Biobank (n=171) were grouped into variant classes according to their bioinformatically predicted functional impact:

- HIGH = predicted high impact by Variant Effect Predictor (VEP; v104) (premature stop variants, frameshift variants, start-lost variants, stop-lost variants) (29).
- MODERATE excl. benign = variants predicted as moderate impact by VEP (missense/inframe insertions/deletions) excluding those predicted 'benign' by SIFT or 'non-damaging' by PolyPhen-2 (30,31).
- MODERATE/HIGH excl. benign = variants fitting into the above two classes combined.
- CADD ≥25 = variants with a predicted damaging score over 25 by Combined Annotation-Dependent Depletion (CADD) v1.6 (32).

Women in UK Biobank self-reported their age at menarche to the nearest whole year (UKB field ID =2714). We categorised these responses as: delayed (≥16yrs), early (<10yrs), or within the typical range – defined as 'normal' (10 to 15 inclusive). Men in UK Biobank self-reported their age at voice breaking (UKB field ID =2385) as 'later than average', 'earlier than average' or 'about average' compared to peers. Exact odds ratios of carriers vs non-carriers were calculated and p-values were calculated using a Fisher's Exact test through use of the *oddsratio()* function in R package *epitools* (v0.5-10.1).

Results

Frequency of deleterious *MC3R* variants in ALSPAC

To assess potential overrepresentation of deleterious *MC3R* variants in delayed puberty cohorts, we first established the number of carriers of these variants in the population based ALSPAC birth cohort to indicate their background frequency.

We previously reported 20 non-synonymous coding *MC3R* variants found within 5774 individuals in ALSPAC; 3 variants present in 6 individuals were functionally characterised as CloF, 1 variant present in an unknown number of individuals was characterised as PloF, and 7 were WT-like (15). In this work we experimentally characterised the 9 remaining variants found in these individuals to determine if any further LoF variants were present. All 9 variants were categorised as WT-like based on their EC_{50} and V_{max} values (Figure 1, Supplementary Figure 1 and Supplementary Table 2 (26)). Overall, we found the frequency of *MC3R* CloF variant carriers in ALSPAC to be 1.0/1000 participants (Table 1). To determine the number of carriers of the single PloF variant found in ALSPAC, R220S, we combined WES data with HRC-imputed derived R220S genotypes available for 9966 participants. In total, 42 were designated as PloF carriers (36 WES and 6 HRC-imputed), giving an overall PloF carrier frequency of 4.2/1000 participants (Table 1).

Frequency of deleterious *MC3R* variants in CDGP

Of the 362 patients with CDGP (84 from Manchester UK CDGP and 278 from DPGen), 8 patients carried a heterozygous non-synonymous *MC3R* variant (F45S, D121Y, L212F, R220S, C254S and a stop-lost variant – stop324Q). C254S and D121Y were functionally characterised as CloF, and L212F as PloF (Figure 1). F45S has been previously reported as CloF and R220S as PloF (15). The patient with the F45S variant had inherited this from his father who also had a history of delayed puberty.

The stop-lost variant (stop324Q) was predicted to lead to an additional 7 amino acids at the C-terminus followed by an alternative stop codon, based on the transcribed 3' UTR sequence (33,34). This C-terminally extended version of the receptor was cloned and functionally characterised and classified as PloF (Figure 1). However, we cannot be sure that this variant is successfully transcribed and translated in the *in vivo* context as it may impact mRNA and/or protein stability leading to further loss of function.

Overall, the frequency of *MC3R* LoF variant carriers was 22.1/1000 patients with CDGP, which is 4.17-fold higher (p=0.001) than was found in the population-based ALSPAC study (Table 1).

As 162/278 participants within DPGen were of non-European ancestry, we performed a sensitivity analysis limiting the DPGen dataset to the 116 individuals of non-Finnish European (NFE) descent, reflective of the ALSPAC control cohort, and very similar LoF carrier frequency and odds ratio were observed (26.1/1000 patients, OR=4.66, p=0.031) (Supplementary Table 3 (26)).

In contrast to patients with CDGP, the frequency of *MC3R* LoF variant carriers among patients with nIHH (6.1/1000) was not largely different to that in controls (Table 1). 2/657 nIHH patients carried the R220S variant, 2/657 patients carried variants functionally characterised as CLoF (D121Y and E305K), and 2/657 patients carried variants characterised as WT-like (V158L and V166L) (Fig 1 and Supplementary Table 2 (26)).

MC3R associations with delayed puberty in UK Biobank

Among 246,328 women from the UK Biobank, *MC3R* variant carrier status was associated with higher odds of self-reported delayed menarche (age at menarche ≥16 yrs), with stronger estimated effects seen for more deleterious variants (Table 2). High-impact variants were 2.3-fold more frequent (p=0.0098) in individuals with delayed menarche compared to those with a normal age at menarche and predicted deleterious (CADD≥25) variants were 1.7-fold more frequent (p=3.90E-07). In contrast, individuals with early menarche (<10 yrs) were less likely to carry a moderate-impact variant (OR=0.67, p=0.043).

While providing a precise estimate for pubertal onset is more challenging in males, the 190,547 men in UK Biobank who reported an "older than average" age at voice breaking were 1.6-fold more likely to carry a moderate-impact *MC3R* variant compared to men with an average age at voice breaking (p=7.46E-04) (Supplementary Table 4 (26)).

Discussion

These data provide evidence suggesting that deleterious mutations in *MC3R* are more frequently found in patients presenting with CDGP but not in those with nIHH. Although the overall number of *MC3R* variants identified was small, previous studies investigating the presence of LoF variants in HPG genes in patients with self-limited delayed puberty report only small numbers of cases who carry a variant in any individual gene (8–10,12,35).

In the population-based UK Biobank cohort, we observed an increased frequency of predicted damaging *MC3R* variants in women with delayed compared to normal age at menarche. This result is concordant with a previous association of rare, predicted deleterious *MC3R* variants with later age at menarche when assessed as a continuous variable in the smaller 200k UK Biobank exome data release (15). The majority of carriers of High impact *MC3R* variants still report 'normal' timing of menarche, evidence that *MC3R* haploinsufficiency has a low level of penetrance and likely influences pubertal timing within a wider polygenic and environmental context.

We also observed a lower frequency of predicted damaging *MC3R* variants in those with early menarche. These findings suggest that MC3R tone may be relevant to both extremes of the population distribution of pubertal timing. It is still unknown whether MC3R activation alone can promote early pubertal timing. Future discovery and characterisation of the impact of *MC3R* gain-of-function variants on pubertal timing and reproductive function would be informative.

We found that deleterious *MC3R* variants are infrequent and not overrepresented among patients with nIHH. This finding is consistent with the evidence that MC3R is a permissive rather than an essential factor for pubertal timing; *Mc3r* knockout mice still progress through puberty, albeit with significant delay, and the reported male homozygote carrier of *MC3R* LoF variant also eventually progressed through puberty to achieve reproductive capacity (15).

We acknowledge limitations of this study, including the lack of matched control groups to account for the ancestry of patients and sequencing coverage. As most patients are of European ancestry, we use ALSPAC as a control group with an approximately equivalent genetic background. Exclusion of patients with non-European ancestry within the DPGen cohort did not change the overall frequency of LoF *MC3R* carriers, indicating this overrepresentation is not confounded by ancestry differences. The growing body of genetic and phenotypic information on cohorts of non-European ancestry will, in time, allow more robust testing of the role of rare variants in *MC3R* in CDGP across different populations. The 171 *MC3R* variants found in the UK Biobank were classified based on their predicted deleteriousness rather than *in vitro* functional assays due to the large number of variants found; a more accurate estimate of LoF carriage will require a high-throughput approach for functionally characterising all *MC3R* protein-altering variants within this cohort.

The reproductive role of MC3R is tightly linked to whole body nutritional status as MC3R activation is regulated by the leptin-proopiomelanocortin pathway, and *MC3R* variants are also associated with shorter stature, lower circulating IGF-1 and reduced indices of lean mass (15,36). Whilst the focus of this study was delayed puberty diagnoses, future prospective studies which include a systematic measure of these phenotypes in patients with delayed puberty would be valuable in ascertaining if patients carrying *MC3R* variants also exhibit nutritionally related phenotypes.

In conclusion, *MC3R* loss-of-function variants are more frequently found in patients presenting with CDGP, and within the general population predicted damaging variants are more common in those with self-reported delayed puberty compared to normally timed puberty.

Acknowledgements and Funding

This research was funded in part, by Wellcome [214274/Z/118/Z and 20863/Z/17/Z]. For Open Access, the author has applied a CC BY public copyright licence to any Author Accepted Manuscript version arising from this submission.

MC3R genetic analysis, next-generation sequencing and functional analysis were supported by the UK Medical Research Council (MRC) Metabolic Diseases Unit (MC_UU_00014/1), Wellcome (WT 095515/Z/11/Z) and the National Institute for Health Research (NIHR) Cambridge Biomedical Research Centre. K.D., K.R, J.W.R.K. and B.Y.H.L are supported by the MRC Metabolic Disease Unit (MC_UU_00014/1) and Wellcome (WT 095515/Z/11/Z). S.O. is supported by a Wellcome Investigator award (WT 095515/Z/11/Z) and the NIHR Cambridge Biomedical Research Centre. J.R.B.P. and K.K.O. are supported by the MRC (Unit programme MC_UU_00006/2). A.W. holds a PhD studentship supported by Wellcome. S.R.H. is supported by a Wellcome fellowship (WT 222049/Z/20/Z). Next-generation sequencing was performed via Wellcome–MRC IMS Genomics and transcriptomics core facility supported by the MRC (MC_UU_00014/5) and Wellcome (208363/Z/17/Z) and the Cancer Research UK Cambridge Institute Genomics Core.

Activities of the Delayed Puberty Genetics Consortium were supported by the National Institutes of Health (R01 HD090071, R01 HD048960, P50 HD104224, M01RR002172, K23 RR15544 and M01 RR000080) and the Doris Duke Charitable Institute (2013110).

The MRC, Wellcome (217065/Z/19/Z) and the University of Bristol provide core support for ALSPAC. GWAS data was generated by sample logistics and genotyping facilities at Wellcome Sanger Institute and

LabCorp (Laboratory Corporation of America) using support from 23andMe. N.J.T. is a Wellcome Investigator (202802/Z/16/Z), is the principal investigator of the ALSPAC (MRC & WT 217065/Z/19/Z), is supported by the University of Bristol NIHR Biomedical Research Centre (BRC-1215-2001), the MRC Integrative Epidemiology Unit (MC_UU_00011) and works within the Cancer Research UK Integrative Cancer Epidemiology Programme (C18281/A19169). L.J.C. is supported by a Wellcome Investigator grant (202802/Z/16/Z). We are extremely grateful to all of the families who took part in the ALSPAC study, the midwives for their help in recruiting them, and the whole ALSPAC team, which includes interviewers, computer and laboratory technicians, clerical workers, research scientists, volunteers, managers, receptionists and nurses. We would also like to thank Allan Daly and Vivek Iyer from the Sanger Human Genetics Informatics team for their contribution to the quality control of the ALSPAC WES dataset.

Data availability

All data used in the UKB association analysis are available from the UKB upon application (https://www.ukbiobank.ac.uk/). This study has been conducted under UK Biobank application number 9905.

The ALSPAC study website contains details of all the data that is available through a fully searchable data dictionary and variable search tool (https://www.bristol.ac.uk/alspac/researchers/our-data/). This study utilised data under project numbers B2891 and B4033.

Some datasets analysed during the current study are not publicly available but are available from the corresponding author on reasonable request.

References

1. Palmert MR, Dunkel L. Delayed Puberty. *N Engl J Med*. 2012;366(5):443-453. doi:10.1056/NEJMcp1109290

2. Day FR, Elks CE, Murray A, Ong KK, Perry JRB. Puberty timing associated with diabetes, cardiovascular disease and also diverse health outcomes in men and women: the UK Biobank study. *Sci Rep*. 2015;5(1):11208. doi:10.1038/srep11208

3. Howard SR, Dunkel L. Delayed Puberty—Phenotypic Diversity, Molecular Genetic Mechanisms, and Recent Discoveries. *Endocr Rev*. 2019;40(5):1285-1317. doi:10.1210/er.2018-00248

4. Harrington J, Palmert MR. Distinguishing Self-limited Delayed Puberty From Permanent Hypogonadotropic Hypogonadism: How and Why? *J Clin Endocrinol Metab*. 2021;106(12):e5264-e5266. doi:10.1210/clinem/dgab498

5. Sedlmeyer IL, Palmert MR. Delayed Puberty: Analysis of a Large Case Series from an Academic Center. *J Clin Endocrinol Metab*. 2002;87(4):1613-1620. doi:10.1210/jcem.87.4.8395

6. Sedlmeyer IL, Hirschhorn JN, Palmert MR. Pedigree Analysis of Constitutional Delay of Growth and Maturation: Determination of Familial Aggregation and Inheritance Patterns. *J Clin Endocrinol Metab*. 2002;87(12):5581-5586. doi:10.1210/jc.2002-020862

7. Wehkalampi K, Widén E, Laine T, Palotie A, Dunkel L. 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*. 2008;93(3):723-728. doi:10.1210/jc.2007-1786

8. Zhu J, Choa REY, Guo MH, et al. A shared genetic basis for self-limited delayed puberty and idiopathic hypogonadotropic hypogonadism. *J Clin Endocrinol Metab*. 2015;100(4):E646-654. doi:10.1210/jc.2015-1080

9. Howard SR, Guasti L, Ruiz-Babot G, et al. IGSF10 mutations dysregulate gonadotropin-releasing hormone neuronal migration resulting in delayed puberty. *EMBO Mol Med*. 2016;8(6):626-642. doi:10.15252/emmm.201606250

10. Tusset C, Noel SD, Trarbach EB, et al. Mutational Analysis of TAC3 and TACR3 Genes in Patients with Idiopathic Central Pubertal Disorders. *Arq Bras Endocrinol Metabol*. 2012;56(9):646-652.

11. Miraoui H, Dwyer AA, Sykiotis GP, et al. Mutations in FGF17, IL17RD, DUSP6, SPRY4, and FLRT3 Are Identified in Individuals with Congenital Hypogonadotropic Hypogonadism. *Am J Hum Genet*. 2013;92(5):725-743. doi:10.1016/j.ajhg.2013.04.008

12. Cassatella D, Howard SR, Acierno JS, et al. Congenital hypogonadotropic hypogonadism and constitutional delay of growth and puberty have distinct genetic architectures. *Eur J Endocrinol*. 2018;178(4):377-388. doi:10.1530/EJE-17-0568

13. Hollis B, Day FR, Busch AS, et al. Genomic analysis of male puberty timing highlights shared genetic basis with hair colour and lifespan. *Nat Commun*. 2020;11(1):1536. doi:10.1038/s41467-020-14451-5

14. Day FR, Thompson DJ, Helgason H, et al. Genomic analyses identify hundreds of variants associated with age at menarche and support a role for puberty timing in cancer risk. *Nat Genet*. 2017;49(6):834-841. doi:10.1038/ng.3841

15. Lam BYH, Williamson A, Finer S, et al. MC3R links nutritional state to childhood growth and the timing of puberty. *Nature*. 2021;599(7885):436-441. doi:10.1038/s41586-021-04088-9

16. Campbell JN, Macosko EZ, Fenselau H, et al. A molecular census of arcuate hypothalamus and median eminence cell types. *Nat Neurosci*. 2017;20(3):484-496. doi:10.1038/nn.4495

17. Sweeney P, Bedenbaugh MN, Maldonado J, et al. The melanocortin-3 receptor is a pharmacological target for the regulation of anorexia. *Sci Transl Med*. 2021;13(590):eabd6434. doi:10.1126/scitranslmed.abd6434

18. Lam BYH, Cimino I, Polex-Wolf J, et al. Heterogeneity of hypothalamic pro-opiomelanocortin-expressing neurons revealed by single-cell RNA sequencing. *Mol Metab*. 2017;6(5):383-392. doi:10.1016/j.molmet.2017.02.007

19. Jonsdottir-Lewis E, Feld A, Ciarlo R, Denhoff E, Feldman HA, Chan YM. Timing of Pubertal Onset in Girls and Boys With Constitutional Delay. *J Clin Endocrinol Metab*. 2021;106(9):e3693-e3703. doi:10.1210/clinem/dgab270

20. Banerjee I, Hanson D, Perveen R, Whatmore A, Black GC, Clayton PE. Constitutional delay of growth and puberty is not commonly associated with mutations in the acid labile subunit gene. *Eur J Endocrinol*. 2008;158(4):473-477. doi:10.1530/EJE-07-0769

21. Guo MH, Plummer L, Chan YM, Hirschhorn JN, Lippincott MF. Burden Testing of Rare Variants Identified through Exome Sequencing via Publicly Available Control Data. *Am J Hum Genet*. 2018;103(4):522-534. doi:10.1016/j.ajhg.2018.08.016

22. Davis EE, Balasubramanian R, Kupchinsky ZA, et al. TCF12 haploinsufficiency causes autosomal dominant Kallmann syndrome and reveals network-level interactions between causal loci. *Hum Mol Genet*. 2020;29(14):2435-2450. doi:10.1093/hmg/ddaa120

23. Boyd A, Golding J, Macleod J, et al. Cohort Profile: The 'Children of the 90s' —the index offspring of the Avon Longitudinal Study of Parents and Children. *Int J Epidemiol*. 2013;42(1):111-127. doi:10.1093/ije/dys064

24. Fraser A, Macdonald-Wallis C, Tilling K, et al. Cohort Profile: The Avon Longitudinal Study of Parents and Children: ALSPAC mothers cohort. *Int J Epidemiol*. 2013;42(1):97-110. doi:10.1093/ije/dys066

25. Northstone K, Lewcock M, Groom A, et al. The Avon Longitudinal Study of Parents and Children (ALSPAC): an update on the enrolled sample of index children in 2019. *Wellcome Open Res*. 2019;4:51. doi:10.12688/wellcomeopenres.15132.1

26. Supplementary Material to "Prevalence of deleterious variants in MC3R in patients with constitutional delay of growth and puberty." https://doi.org/10.17863/CAM.93605

27. Backman JD, Li AH, Marcketta A, et al. Exome sequencing and analysis of 454,787 UK Biobank participants. *Nature*. 2021;599(7886):628-634. doi:10.1038/s41586-021-04103-z

28. Gardner EJ, Kentistou KA, Stankovic S, et al. Damaging missense variants in IGF1R implicate a role for IGF-1 resistance in the etiology of type 2 diabetes. *Cell Genomics*. 2022;2(12):100208. doi:10.1016/j.xgen.2022.100208

29. McLaren W, Gil L, Hunt SE, et al. The Ensembl Variant Effect Predictor. *Genome Biol*. 2016;17(1). doi:10.1186/s13059-016-0974-4

30. Ng PC, Henikoff S. SIFT: predicting amino acid changes that affect protein function. *Nucleic Acids Res*. 2003;31(13):3812-3814.

31. Adzhubei IA, Schmidt S, Peshkin L, et al. A method and server for predicting damaging missense mutations. *Nat Methods*. 2010;7(4):248-249. doi:10.1038/nmeth0410-248

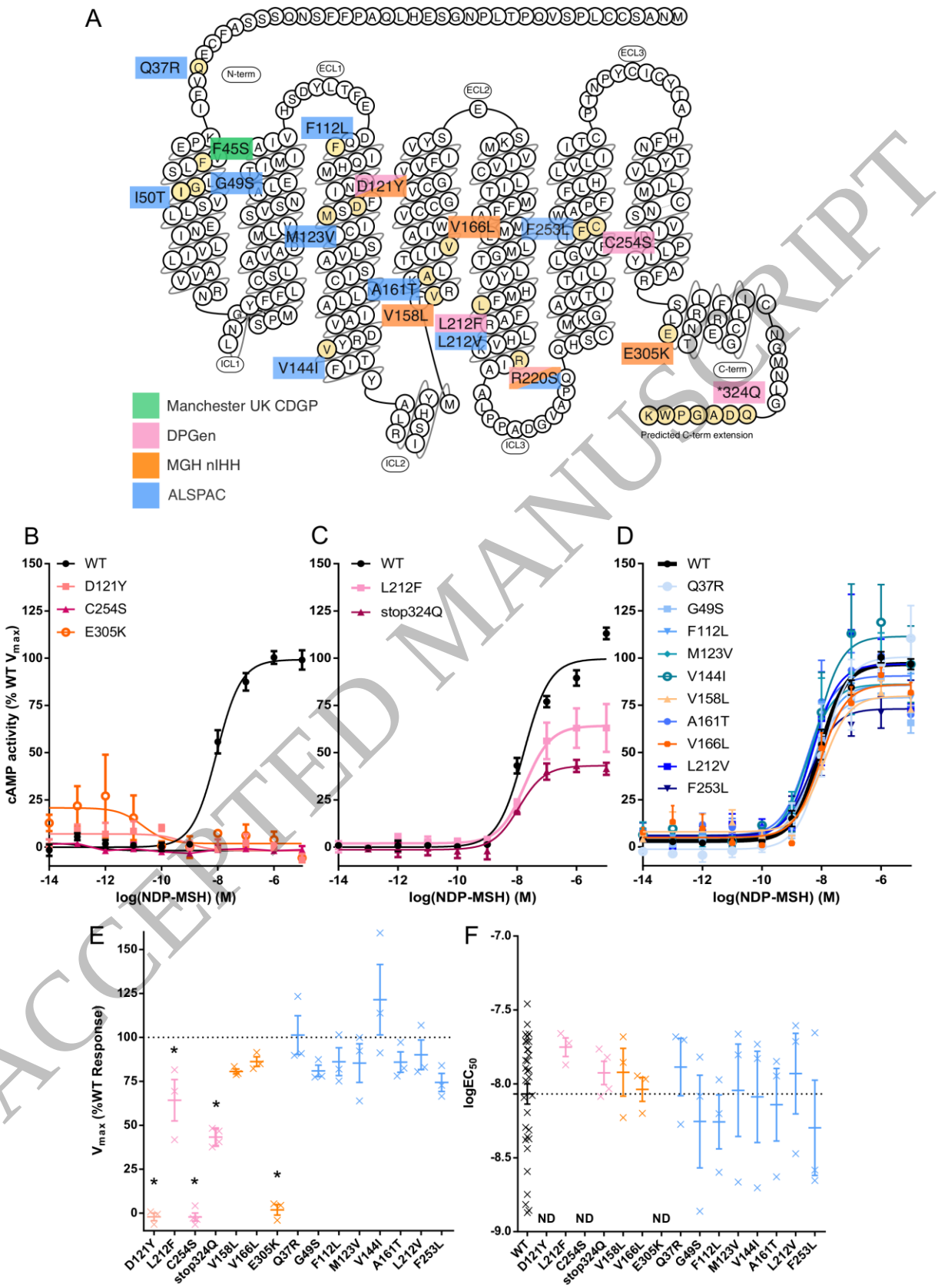
32. Rentzsch P, Witten D, Cooper GM, Shendure J, Kircher M. CADD: predicting the deleteriousness of variants throughout the human genome. *Nucleic Acids Res*. 2019;47(D1):D886-D894. doi:10.1093/nar/gky1016

33. Park J, Sharma N, Cutting GR. Melanocortin 3 Receptor Has a 5′ Exon That Directs Translation of Apically Localized Protein From the Second In-Frame ATG. *Mol Endocrinol*. 2014;28(9):1547-1557. doi:10.1210/me.2014-1105

34. Taylor-Douglas DC, Basu A, Gardner RM, Aspelund S, Wen X, Yanovski JA. Evaluation of Hypothalamic Murineand Human Melanocortin 3 Receptor Transcript Structure. *Biochem Biophys Res Commun*. 2014;454(1):234-238. doi:10.1016/j.bbrc.2014.10.072

35. Howard SR, Oleari R, Poliandri A, et al. HS6ST1 Insufficiency Causes Self-Limited Delayed Puberty in Contrast With Other GnRH Deficiency Genes. *J Clin Endocrinol Metab*. 2018;103(9):3420-3429. doi:10.1210/jc.2018-00646

36. Yeo GSH, Chao DHM, Siegert AM, et al. The melanocortin pathway and energy homeostasis: From discovery to obesity therapy. *Mol Metab*. 2021;48:101206. doi:10.1016/j.molmet.2021.101206



1

Figure 1: Functional characterisation of *MC3R* coding mutations in CDGP and population cohorts. A) Schematic showing locations of identified non-synonymous variants in Manchester UK CDGP, nIHH, DPGen and ALSPAC cohorts. F45S has already been characterised as CLoF, and R220S as PLoF. B), C) and D) cAMP dose-response curves upon treatment with NDP-MSH classified CLoF, PLoF and WT-like variants, respectively. Error bars= s.e.m., N reported in Supplementary Materials. E) and F) V_{max} (%WT) and logEC50 values plotted for all variants except I50T (Supplementary Materials) calculated from dose response curves. Dotted line indicates WT response. Individual crosses represent biological replicates. (*) represents Bonferroni $p < 0.05$ via two–way ANOVA compared to wild-type (WT). ND= not determined.

		Partial LoF					Complete LoF					Combined PLoF and CLoF				
Cohort	Total N	N carriers	Frequency (per 1000)	OR vs. ALSPAC	95% CI	p-value	N carriers	Frequency (per 1000)	OR vs. ALSPAC	95% CI	p-value	N carriers	Frequency (per 1000)	OR vs. ALSPAC	95% CI	p-value
Manchester UK CDGP	84	0	0	-	-	-	1	11.9	11.44	0.25-95.9	*0.097*	1	11.9	2.25	0.06-13.4	*0.366*
DPGen	278	5	18.0	4.27	1.31-10.9	*0.009*	2	7.2	6.92	0.68-39.0	*0.050*	7	25.2	4.75	1.81-10.6	*0.001*
Total DP:	362	5	13.8	3.28	1.01-8.34	*0.024*	3	8.3	7.97	1.28-37.5	*0.013*	8	22.1	4.17	1.70-8.91	*0.001*
nIHH	657	2	3.0	0.72	0.08-2.79	*1*	2	3.0	2.93	0.29-16.4	*0.194*	4	6.1	1.15	0.30-3.13	*0.779*
ALSPAC	5774 (9966)	42	4.2	-	-	-	6	1.0	-	-	-	-	5.3	-	0	-

Table 1: Frequencies of validated *MC3R* Loss of Function (LoF) variants in delayed puberty (DP) cohorts. All *MC3R* coding variants were experimentally characterised as Partial LoF (Vmax >25% Vmax(WT) and Vmax ≤75% Vmax(WT)) or Complete LoF (Vmax ≤25% Vmax(WT)). Odds ratios (OR), 95% confidence intervals (CI) and p-values for frequency of LoF mutations compared to ALSPAC controls were calculated using Fisher's Exact test. For CLoF variants, data were available on 5774 ALSPAC individuals, and for pLoF variants 9966 ALSPAC individuals. For combined PLoF and CLoF, a combined frequency in ALSPAC was calculated. CDGP: constitutional delay of growth and puberty; DP: delayed puberty; nIHH = normosmic idiopathic hypogonadotropic hypogonadism

Variant class	Menarche timing	N Noncarriers	N Carriers	Carrier Frequency (%)	OR vs normal timing	95% CI	p-value
HIGH	Delayed (≥16Yrs)	23216	13	0.0560	2.3466	1.17-4.38	*0.0098*
HIGH	Early (< 10Yrs)	9321	2	0.0215	0.8992	0.11-3.42	*1*
HIGH	Normal	213725	51	0.0239	-	-	-
MODERATE excl. benign	Delayed (≥16Yrs)	23088	141	0.6070	1.5155	1.26-1.81	*1.45E-05*
MODERATE excl. benign	Early (< 10Yrs)	9298	25	0.2682	0.6672	0.43-0.99	*0.0430*
MODERATE excl. benign	Normal	212918	858	0.4014	-		-
MODERATE/HIGH excl. benign	Delayed (≥16Yrs)	23076	153	0.6587	1.5527	1.30-1.85	*1.77E-06*
MODERATE/HIGH excl. benign	Early (< 10Yrs)	9296	27	0.2896	0.6802	0.45-1.00	*0.0489*
MODERATE/HIGH excl. benign	Normal	212867	909	0.4252	-	-	-
CADD ≥25	Delayed (≥16Yrs)	23097	132	0.5683	1.6633	1.37-2.01	*3.90E-07*
CADD ≥25	Early (< 10Yrs)	9300	23	0.2467	0.7198	0.45-1.09	*0.1437*
CADD ≥25	Normal	213044	732	0.3424	-	-	-

Table 2: Frequencies of predicted deleterious MC3R variants in women with delayed or early menarche timing. Data are from the UK Biobank 450k exome cohort. Odds ratios were calculated for frequency of carriers for MC3R coding variants in clinically defined categories of self-reported menarche timing. Variant classes were defined by Variant Effect Predictor: HIGH = protein-truncating variants. Moderate = missense variants/in-frame insertion/deletions. Benign variants were excluded as classified by SIFT/Polyphen. Combined Annotation Dependent Depletion (CADD) score cutoff set to ≥ 25.