Genes & Health Research Team (2021). MC3R links nutritional state to childhood growth and the timing of puberty. *Nature*, *599*(7885), 436-441. https://doi.org/10.1038/s41586-021-04088-9

MC3R links nutritional state to childhood growth and the timing of puberty

Lam BYH[1*], Williamson A[1,2*], Finer S[3*], Day FR[2], Tadross JA[1,4], Gonçalves Soares A[5], Wade K[5], Sweeney P[6], Bedenbaugh MN[7], Porter DT[6], Melvin A[1], Ellacott KLJ[8], Lippert RN[9], Buller S[1], Rosmaninho-Salgado J[10], Dowsett GKC[1], Ridley KE[11], Xu Z[11], Cimino I[1], Rimmington D[1], Rainbow K[1], Duckett K[1], Holmqvist S[11], Khan A[3], Dai X[12], Bochukova EG[12], Genes & Health Research Team[#], Trembath RC[13], Martin HC[14], Coll AP[1], Rowitch DH[11], Wareham NJ[2], van Heel DA[3,12], Timpson N[5], Simerly RB[7], Ong KK[2,11], Cone RD[6,15], Langenberg C[2,16*], Perry JRB[2*], Yeo GS[1*], O'Rahilly S[1*]

[*]Equal contributions

Correspondence to so104@medschl.cam.ac.uk

1 MRC Metabolic Disease Unit, Wellcome-MRC Institute of Metabolic Science, University of Cambridge, and NIHR Cambridge Biomedical Research Centre, United Kingdom.

2 MRC Epidemiology Unit, Wellcome-MRC Institute of Metabolic Science, University of Cambridge, United Kingdom.

3 Wolfson Institute of Population Health, Barts and the London School of Medicine and Dentistry, Queen Mary University of London, United Kingdom.

4 Department of Pathology, University of Cambridge, United Kingdom.

5 MRC Integrative Epidemiology Unit and Population Health Sciences, Bristol Medical School, University of Bristol, United Kingdom.

6 Life Sciences Institute, University of Michigan, Ann Arbor, MI USA.

7 Department of Molecular Physiology and Biophysics, School of Medicine, Vanderbilt University, Nashville, TN USA.

8 Institute of Biomedical and Clinical Sciences, University of Exeter Medical School, Exeter, United Kingdom.

9 Department of Neurocircuit Development and Function, German Institute of Human Nutrition, Potsdam Rehbrücke, Potsdam Germany.

10 Medical Genetics Unit, Hospital Pediátrico, Centro Hospitalar e Universitário de Coimbra, Portugal.

11 Department of Paediatrics, University of Cambridge, United Kingdom.

12 Blizard Institute, Barts and The London School of Medicine and Dentistry, Queen Mary University London, London, United Kingdom.

13 School of Basic and Medical Biosciences, Faculty of Life Sciences and Medicine, King's College London, London, United Kingdom.

14 Wellcome Sanger Institute, Hinxton, Cambridge, United Kingdom.

15 Department of Molecular and Integrative Physiology, School of Medicine, University of Michigan, Ann Arbor, MI USA.

16 Computational Medicine, Berlin Institute of Health at Charité – Universitätsmedizin Berlin, Germany.

[#] A full list of members and their affiliations appears in the Supplementary Information.

Summary

The state of somatic energy stores in metazoans is communicated to the brain, which regulates key aspects of behaviour, growth, nutrient partitioning and development[1]. The central melanocortin system acts through Melanocortin-4 Receptor (MC4R) to control appetite, food intake and energy expenditure[2]. We now present evidence that the Melanocortin-3 Receptor (MC3R) regulates the timing of sexual maturation, the rate of linear growth and the accrual of lean mass, all energy-sensitive processes. We found that humans who carry loss-of-function mutations in *MC3R*, including a rare homozygote, have a later onset of puberty. Consistent with previous findings in mice, they also had reduced linear growth, lean mass and IGF-1 levels. Mice lacking *Mc3r* had delayed sexual maturation and an insensitivity of reproductive cycle length to nutritional perturbation. The expression of *Mc3r* is enriched in hypothalamic neurons controlling reproduction and growth and increases during post-natal development in a manner consistent with a role in regulation of sexual maturation. These findings suggest a bifurcating model of nutrient sensing by the central melanocortin pathway with signalling through MC4R controlling the acquisition and retention of calories, while MC3R primarily regulates their disposition into growth, lean mass and the timing of sexual maturation.

Introduction

Pro-opiomelanocortin (POMC), encoding several melanocortin peptides, is expressed in neurons of the hypothalamic arcuate nucleus[2] which are activated by key hormonal signals of caloric balance, leptin[3] and insulin[4]. These hormones also negatively regulate the activity of neurons releasing the melanocortin receptor antagonist, agouti-related peptide (AgRP)[5]. The actions of alpha- and beta- melanocyte-stimulating hormone (MSH) on the Melanocortin-4 Receptor (MC4R) are necessary for the normal control of food intake and energy expenditure[6,7]. Humans (and mice) lacking *MC4R* are obese and hyperphagic and have reduced basal energy expenditure[6,8-12]. However they have normal or even accelerated early linear growth and no retardation of pubertal development[11], both of which are impaired by caloric deprivation[13] or leptin deficiency[14,15]. This suggests either that POMC-derived peptides are not responsible for transmitting nutritional signals to those particular downstream processes or that a different melanocortin receptor is involved. The Melanocortin-3 Receptor (MC3R) is the only other melanocortin receptor that is predominantly expressed in the brain[16,17]. Mice lacking *Mc3r* have been reported to have a normal reproductive development, fertility and no change in food intake, but develop an altered body composition with a high ratio of fat-to-lean mass and impaired linear growth[18-20]. Human genome-wide association studies (GWAS) have identified common variants in the vicinity of *MC3R* which are associated with both adult height[21] and with age at menarche[22]. While rare functionally compromised heterozygous variants in *MC3R* have been reported in humans, no consistent phenotype has been reported[23], though associations with height[24] and obesity[25-27] have been suggested. We set out to establish the role of MC3R in human physiology by seeking naturally occurring mutations which resulted in functional impairment of the receptor, and studying the relationship with relevant human phenotypes. We identify a strong and previously unreported impact of *MC3R* loss-of-function (LoF) mutations on pubertal timing in humans, and provide evidence for the conservation of this pathway in mice. Consistent with phenotypes previously described in *Mc3r* deficient mice, we report that human *MC3R* deficiency is also associated with reduced childhood growth, adult height and lean mass.

Heterozygous *MC3R* phenotypes

Using whole exome sequence (WES) data from ~200,000 UK Biobank (UKBB) participants, we found that 0.82% of individuals carried at least one rare (minor allele frequency, MAF <0.2%), predicted deleterious variant in *MC3R* (**Table S1**). We undertook aggregated gene burden tests focused on traits relevant to growth, body composition and pubertal timing. The 812 female *MC3R* rare mutation carriers had a 4.7-month delay in age at menarche compared to non-carriers (beta=0.39 years, $P=6.4x10^{-12}$), an effect size ~3 times larger than the most significantly associated common variant in the genome (*LIN28B* locus)[22]. The *MC3R* gene burden score was also associated with delayed voice breaking in males, shorter adult and childhood stature, lower sitting height, lower circulating IGF1 levels, lower total body lean mass, and lower appendicular lean mass (ALM)/BMI ratio, an established measure of sarcopenia[28,29] (**Table S2**).

In order to determine whether there was a quantitative relationship between the degree of functional impact of individual non-synonymous mutations and phenotypic outcomes, we selected three missense mutations which were sufficiently common in the full ~500K UKBB sample to allow robust testing of association with phenotypes (MAF ≥ 0.05% using array genotypes). We identified p.F45S and p.R220S with MAFs of 0.06%, and 0.19% respectively (**Extended Data** (**ED**) **Fig 1**). The third variant, p.V44I (MAF = 10.09%; **ED Fig 1**) is in strong linkage disequilibrium (LD, $r^2 = 0.97$) with a previously identified GWAS signal for age at menarche (rs3746619 located in the 5' untranslated region (UTR))[22]. We measured the ability of these mutants to generate cAMP in Human Embryonic Kidney (HEK293) cells *in vitro*, upon stimulation by [Nle_4, D-Phe_7]-α-melanocyte-stimulating hormone (NDP-MSH). p.F45S exhibited severely impaired signalling compared to MC3R wild-type (WT), p.R220S showed partial LoF and p.V44I was indistinguishable from WT (**Fig 1a-d, Table S3**).

While all three variants were individually and jointly associated with delay of pubertal onset in both females and males (**Fig 1e, Table S2&S4**), individuals heterozygous for the rarer p.F45S and p.R220S variants, which result in a more substantial disruption of cAMP signalling, had a greater delay in pubertal onset, with female carriers of p.F45S mutation having a 5.16-month delay (**Fig 1f**). These variants were also associated with reduced growth, as indicated by shorter total and sitting height in adults (**Fig 1e&g, Table S2**), and shorter relative childhood height at age 10 years (**Fig 1e, Table S2**). The much more common p.V44I variant was also significantly associated with age at puberty and height, albeit with a substantially smaller effect size (**Fig 1e-g**). Although this variant exhibits no significant difference from WT in the cAMP assay, we hypothesise that the large numbers of carriers (~50,000) allowed us to discern a phenotypic impact of a reduction in signalling resulting from this mutation not discernible in a heterologous over-expression system (~94% of WT, **Fig 1b&d**). Alternatively, the effect may be explained by its LD with the 5'UTR variant (or other non-coding variants) that could affect the expression of *MC3R.* Carriers of these three variants also had lower total body lean mass and a reduced ALM/BMI ratio (**Fig 1e&h, Table S2**). There was some heterogeneity between individual variant associations – notably, associations with childhood height were more consistent than with adult height (**Fig 1e, Table S2**) and p.R220S was also associated with lower circulating IGF1 levels (**Fig 1e&i, Table S2**). Of note, no variant showed any association with BMI, waist-to-hip ratio, fat mass, type 2 diabetes, HbA1c or random glucose (**Fig 1e, Table S2**). Phenome-wide association analyses across publicly available GWAS summary statistics in UKBB and additional cohorts demonstrated that pubertal onset and height had the strongest associations (**ED Fig 2**), with no other traits reaching significance after multiple test correction.

In order to study the impact of MC3R LoF throughout development, we studied 5,993 unrelated participants from the Avon Longitudinal Study of Parents and Children (ALSPAC)[30]. Using a pooled amplicon next-generation sequencing approach[31], we identified seven rare, non-synonymous variants in *MC3R* that were predicted deleterious *in silico* by SIFT and Polyphen2 (**ED Fig 1, Table S3**) and found three variants: p.F45S, p.L53R and p.A214P, which all exhibited complete LoF in generating cAMP (**Fig 1a-d, Table S3).** We then used Sanger

sequencing to identify a total of six heterozygous carriers of any one of the three LoF mutations and performed an aggregated burden test on anthropometric trajectories and pubertal timing. We found that despite the small sample size (n=6), *MC3R* LoF mutations were associated with lower height throughout childhood, adolescence and early adulthood, with a trend towards lower lean mass and lower weight (**Table S5, ED Fig 3,** also see **Supplementary Information, Tables S14-S15)**. No effect on pubertal onset was discernible in this small group (**Table S6**).

To explore the effect of *MC3R* variants on the plasma proteome and metabolome, we used data from the Fenland Study[32] and EPIC Norfolk[33,34], respectively. We identified IGFBP1, a liver-derived protein which is known to be suppressed by growth hormone (GH)[35], as the most strongly associated target (**Table S7**). The two most strongly associated metabolites with *MC3R* p.F45S, pipecolate (beta = 1.1, SE=0.33, P=9.6x10^{-4}) and 4-hydroxyphenylpyruvate (beta=0.96, SE=0.31, P=0.0025) are metabolites of lysine and tyrosine respectively and likely reflect increased proteolysis (**Table S8**). These associations, while potentially illuminating, did not reach stringent, multiple test corrected thresholds (see methods).

MC3R LoF homozygous phenotype

In the exome data from participants in the Genes & Health (G&H) study, in whom 18.8% report parental relatedness[36], we found two rare, homozygous non-synonymous mutations p.M97I and p.G240W (**ED Fig 1**), each in one participant. While p.M97I signalled normally, the p.G240W mutant receptor was completely unresponsive (**Fig 1a-d, Table S3**).

The participant carrying p.G240W was invited for phenotypic assessment under ethically approved recall protocols, and gave informed consent for publication of results. He is a male of Bangladeshi origin, in his early 40s whose parents are second cousins. The mutation is in an 8.3Mb genomic region of homozygosity, consistent with consanguineous inheritance.

He reported a history of significantly delayed puberty, starting in his early 20s after which he subsequently fathered children. He was of markedly short stature, -2.95 SDs of the mean by WHO reference[37]. His sitting height ratio was below the normal range for South Asians and had reduced circulating levels of IGF1 (**Table S9**). In contrast to the finding in heterozygotes, he has been overweight/obese since early childhood and currently has a BMI of 40.4 kg/m^2 (**Table S9**), accompanied by type 2 diabetes and hypertension, both well-controlled. Inspection of his exome sequence for all known monogenic obesity genes did not reveal any pathogenic mutations.

Whole body dual-energy X-ray absorptiometry (DEXA) scanning (**Fig 2a**) revealed a high percentage of body fat at 48.5% (**Fig 2a-c**), but a low total lean mass for his level of BMI (**Fig 2d**). His ALM/BMI ratio, an index of sarcopenia, was below normal (**Fig 2e, Table S9**).

Conserved role of MC3R in mice

Male mice lacking *Mc3r* had a 2-day delay in the onset of sexual maturation compared to WT littermates (**Fig 3a**), with females showing a similar trend (**Fig 3b**). In mature female *Mc3r* null mice, the oestrus cycle length was significantly prolonged (**Fig 3c&d**). To establish

whether the known impact of acute caloric deficiency on cycle length required MC3R, WT and *Mc3r* deficient mice were subjected to an overnight fast. In WT mice, this resulted in a > 2-fold prolongation of oestrous cycle length. Strikingly, in the absence of *Mc3r*, the effect of fasting on cycle length was abolished (**Fig 3c&d, ED Fig 4a&b**).

Mc3r expression in the hypothalamus

Using a single-cell RNA sequencing dataset of the arcuate nucleus (ARC)[38] Sweeney *et al* [39] recently reported that *Mc3r* expression was significantly enriched in neurons expressing Kisspeptin, Neurokinin B and Dynorphin (so called KND*y* neurons) and in Growth Hormone-releasing Hormone (GHRH) neurons. We undertook an expanded analysis including three additional studies[40-42] (**Table S10, ED Fig 5a, gene markers in Table S11**) increasing the number of neurons interrogated to 18,427; 1,166 of which expressed *Mc3r* (**Fig 4a**, **ED Fig 5b, gene markers in Table S12**). This analysis strengthened evidence for co-expression of *Mc3r* in KND*y* neurons (controlling reproduction) and GHRH neurons (controlling growth) (**Fig 4b**). Using single-molecule *in situ* hybridisation (smFISH), we validated the co-expression of *Mc3r+Tac2* (**Fig 4c&d**), *Mc3r+Kiss1* (**ED Fig 5c-e**) and *Mc3r+Ghrh* (**Fig 4c&d**) in the ARC. Leptin regulates the activity of both KND*y*[43] and GHRH neurons[44,45], we therefore assessed the expression of the leptin receptor gene (*Lepr*), *Mc3r* and *Mc4r* in the KND*y* and GHRH neurons from the full dataset of 18.4K cells. Both clusters expressed more *Mc3r* compared to *Lepr* and *Mc4r* (**ED Fig 6a-c**). We also established that the expression of *MC3R* in KND*y* and GHRH neurons is conserved in humans by smFISH (**ED Fig 7a&b**). Finally, we studied female mice at P16 (infantile), P28 (juvenile) and P48 (sexual maturation), and found that *Mc3r* mRNA was detected in 40-60% of *Kiss1*-expressing KND*y* neurons in the ARC with no change in proportion with age (**Fig 4e**). In contrast, in the *Kiss1* neurons of the anteroventral periventricular nucleus (AVPV), necessary for the pre-ovulatory GnRH surge[46], there was a significant increase in the number of *Kiss1* and *Mc3r* co-expressing cells from P28 to P48 (**Fig 4f, ED Fig 8a-c**).

Summary and Conclusions

Caloric deprivation is associated with reduced linear growth and delay in the onset of puberty[13], whereas over-nourished children tend to grow more rapidly and enter puberty earlier[47]. Increased macronutrient availability is thought to underpin the progressive increase in height and decrease at age of onset of puberty that has occurred globally over the past century or more[48]. Leptin and insulin provide signals of nutritional sufficiency to hypothalamic neurons expressing melanocortin agonists and antagonists. While these act on MC4R to control food intake and energy expenditure, no such clarity has existed regarding the link between nutritional status and the control of linear growth or the onset of puberty.

The robust association between *MC3R* LoF mutations and pubertal delay found in our study indicates a role for MC3R in the control of the human reproductive axis. The striking insensitivity of *Mc3r* deficient mice to the reproductive impact of a period of fasting and the evidence that these mice have delayed sexual maturation indicates conservation of this biology

across species. *Mc3r* deficient mice have been previously reported to be reproductively unimpaired, but those studies did not subject the mice to fasting and may have failed to detect a subtle delay in the timing of sexual maturation[18,19]. Obese human females with homozygous mutations disrupting POMC do not initiate pubertal development[49]. When treated with setmelanotide, an agonist with 10X selectivity for MC4R over MC3R, they lose weight but remain hypogonadal[49].

MC3R's effects on the reproductive axis may involve direct action on GnRH neurons[50]. We provide evidence that *Mc3r* expression is enriched in KND*y* neurons in the ARC, a site where inhibition of kisspeptin neurons has been shown to impair gonadotropic responses to melanocortins[51]. *Mc3r* expression was also high in kisspeptin neurons in the AVPV, known to be important for the pre-ovulatory surge of gonadotropins[46]. In the latter population, *Mc3r* expression increased with post-natal development in a manner consistent with a role in the timing of sexual maturation.

Consistent with reports of reduced femoral length in mice lacking *Mc3r*[19], we found that humans defective in MC3R signalling have reduced linear growth, correlating with the severity of receptor dysfunction. MC3R status also appears to influence the accrual of lean mass in humans, mirroring previous reports in mice of a low ratio of lean-to-fat tissue[18,19]. The involvement of the GH-IGF1 axis in this phenotype seems likely as, consistent with previous findings in *Mc3r* null mice[19], IGF1 levels were reduced in human mutation carriers. In mice and humans, subpopulations of GHRH neurons express *MC3R*.

The impact of *MC3R* deficiency on height is disproportionate, with greater impact on trunk than leg length. We hypothesise that this occurs because a state of relative GH deficiency throughout childhood and adolescence is partially offset by a longer period of limb growth due to the later onset of puberty, which delays epiphyseal fusion, permitting an extended period of long bone growth.

Consistent with what has been described in *Mc3r* deficient mice[18,19], humans with impaired MC3R signalling have evidence for reduced lean mass. GH is known to influence body composition[52] and is a candidate for this effect, but we cannot exclude additional MC3R-dependent pathways. In that regard the association of MC3R dysfunction with raised circulating levels of breakdown products of amino acid metabolism is notable.

Whether mutations in *MC3R* predispose to human obesity has been unclear[23]. While *Mc3r* null mice have a high ratio of fat-to-lean mass, they are not markedly obese, and heterozygous mice have no alterations in their weight or body composition[18,19]. Consistent with this, heterozygous human carriers of LoF mutations do not have elevated fat mass. In contrast, our homozygous null proband has been obese since early childhood, with no evidence for mutations in known obesity genes. *MC3R* is expressed on both POMC and AgRP neurons and could influence their function in controlling energy balance[39]. Resolution of this question will require the identification of additional humans homozygous for LoF *MC3R* mutations.

We have described a new clinical syndrome of *MC3R* deficiency. Analysis of the *MC3R* gene should become part of the routine genetic analysis of patients delayed puberty, short stature and low IGF1. Our data suggest the potential utility of MC3R agonists in some patients with delayed puberty and/or short stature and also potentially in sarcopenia, a condition where low lean mass, including muscle, contributes to disability in various chronic disorders[53].

In summary, across the animal kingdom, nutritional status is a critical determinant of linear growth and the timing of reproductive maturity[54]. MC3R appears to play an important role in linking signals of caloric sufficiency that act through POMC expressing neurons to the control of growth and reproduction. This provides a plausible mechanistic basis for the global secular trends towards taller human height and earlier onset of puberty that have accompanied higher levels of caloric availability[48].

References

1 Friedman, J. M. The function of leptin in nutrition, weight, and physiology. *Nutr Rev* **60**, S1-14; discussion S68-84, 85-17, doi:10.1301/002966402320634878 (2002).

2 Cone, R. D. Anatomy and regulation of the central melanocortin system. *Nat Neurosci* **8**, 571-578, doi:10.1038/nn1455 (2005).

3 Cowley, M. A. *et al.* Leptin activates anorexigenic POMC neurons through a neural network in the arcuate nucleus. *Nature* **411**, 480-484, doi:10.1038/35078085 (2001).

4 Hill, J. W. *et al.* Direct insulin and leptin action on pro-opiomelanocortin neurons is required for normal glucose homeostasis and fertility. *Cell Metab* **11**, 286-297, doi:10.1016/j.cmet.2010.03.002 (2010).

5 Varela, L. & Horvath, T. L. Leptin and insulin pathways in POMC and AgRP neurons that modulate energy balance and glucose homeostasis. *EMBO Rep* **13**, 1079-1086, doi:10.1038/embor.2012.174 (2012).

6 Chen, A. S. *et al.* Role of the melanocortin-4 receptor in metabolic rate and food intake in mice. *Transgenic Res* **9**, 145-154, doi:10.1023/a:1008983615045 (2000).

7 Fan, W., Boston, B. A., Kesterson, R. A., Hruby, V. J. & Cone, R. D. Role of melanocortinergic neurons in feeding and the agouti obesity syndrome. *Nature* **385**, 165-168, doi:10.1038/385165a0 (1997).

8 Vaisse, C., Clement, K., Guy-Grand, B. & Froguel, P. A frameshift mutation in human MC4R is associated with a dominant form of obesity. *Nat Genet* **20**, 113-114, doi:10.1038/2407 (1998).

9 Yeo, G. S. *et al.* A frameshift mutation in MC4R associated with dominantly inherited human obesity. *Nat Genet* **20**, 111-112, doi:10.1038/2404 (1998).

10 Huszar, D. *et al.* Targeted disruption of the melanocortin-4 receptor results in obesity in mice. *Cell* **88**, 131-141, doi:10.1016/s0092-8674(00)81865-6 (1997).

11 Farooqi, I. S. *et al.* Clinical spectrum of obesity and mutations in the melanocortin 4 receptor gene. *N Engl J Med* **348**, 1085-1095, doi:10.1056/NEJMoa022050 (2003).

12 Krakoff, J. *et al.* Lower metabolic rate in individuals heterozygous for either a frameshift or a functional missense MC4R variant. *Diabetes* **57**, 3267-3272, doi:10.2337/db08-0577 (2008).

13 Brown, P. I., Brasel, J. in *The Malnourished Child Nestlé Nutrition Workshop Series* (ed Lewinter-Suskind Leslie Suskind Robert M) 213-228 (Nestlé Nutrition Institute and Vevey/Raven Press, 1990).

14 Clement, K. *et al.* A mutation in the human leptin receptor gene causes obesity and pituitary dysfunction. *Nature* **392**, 398-401, doi:10.1038/32911 (1998).

15 Strobel, A., Issad, T., Camoin, L., Ozata, M. & Strosberg, A. D. A leptin missense mutation associated with hypogonadism and morbid obesity. *Nat Genet* **18**, 213-215, doi:10.1038/ng0398-213 (1998).
16 Roselli-Rehfuss, L. *et al.* Identification of a receptor for gamma melanotropin and other proopiomelanocortin peptides in the hypothalamus and limbic system. *Proc Natl Acad Sci U S A* **90**, 8856-8860, doi:10.1073/pnas.90.19.8856 (1993).
17 Gantz, I. *et al.* Molecular cloning of a novel melanocortin receptor. *J Biol Chem* **268**, 8246-8250 (1993).
18 Butler, A. A. *et al.* A unique metabolic syndrome causes obesity in the melanocortin-3 receptor-deficient mouse. *Endocrinology* **141**, 3518-3521, doi:10.1210/endo.141.9.7791 (2000).
19 Chen, A. S. *et al.* Inactivation of the mouse melanocortin-3 receptor results in increased fat mass and reduced lean body mass. *Nat Genet* **26**, 97-102, doi:10.1038/79254 (2000).
20 Renquist, B. J. *et al.* Melanocortin-3 receptor regulates the normal fasting response. *Proc Natl Acad Sci U S A* **109**, E1489-1498, doi:10.1073/pnas.1201994109 (2012).
21 Wood, A. R. *et al.* Defining the role of common variation in the genomic and biological architecture of adult human height. *Nat Genet* **46**, 1173-1186, doi:10.1038/ng.3097 (2014).
22 Day, F. R. *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* **49**, 834-841, doi:10.1038/ng.3841 (2017).
23 Demidowich, A. P., Jun, J. Y. & Yanovski, J. A. Polymorphisms and mutations in the melanocortin-3 receptor and their relation to human obesity. *Biochim Biophys Acta Mol Basis Dis* **1863**, 2468-2476, doi:10.1016/j.bbadis.2017.03.018 (2017).
24 Marouli, E. *et al.* Rare and low-frequency coding variants alter human adult height. *Nature* **542**, 186-190, doi:10.1038/nature21039 (2017).
25 Mencarelli, M. *et al.* Rare melanocortin-3 receptor mutations with in vitro functional consequences are associated with human obesity. *Hum Mol Genet* **20**, 392-399, doi:10.1093/hmg/ddq472 (2011).
26 Zegers, D. *et al.* Identification of three novel genetic variants in the melanocortin-3 receptor of obese children. *Obesity (Silver Spring)* **19**, 152-159, doi:10.1038/oby.2010.127 (2011).
27 Lee, Y. S., Poh, L. K. & Loke, K. Y. A novel melanocortin 3 receptor gene (MC3R) mutation associated with severe obesity. *J Clin Endocrinol Metab* **87**, 1423-1426, doi:10.1210/jcem.87.3.8461 (2002).
28 Studenski, S. A. *et al.* The FNIH Sarcopenia Project: Rationale, Study Description, Conference Recommendations, and Final Estimates. *The Journals of Gerontology: Series A* **69**, 547-558, doi:10.1093/gerona/glu010 (2014).
29 Kim, T. N. *et al.* Comparisons of three different methods for defining sarcopenia: An aspect of cardiometabolic risk. *Sci Rep* **7**, 6491, doi:10.1038/s41598-017-06831-7 (2017).
30 Boyd, A. *et al.* Cohort Profile: the 'children of the 90s'--the index offspring of the Avon Longitudinal Study of Parents and Children. *Int J Epidemiol* **42**, 111-127, doi:10.1093/ije/dys064 (2013).
31 Wade, K. H. *et al.* Loss-of-function mutations in the melanocortin 4 receptor in a UK birth cohort. *Nat Med* **27**, 1088-1096, doi:10.1038/s41591-021-01349-y (2021).
32 Lotta, L. A. *et al.* A cross-platform approach identifies genetic regulators of human metabolism and health. *Nat Genet* **53**, 54-64, doi:10.1038/s41588-020-00751-5 (2021).
33 Khaw, K. T. *et al.* Combined impact of health behaviours and mortality in men and women: the EPIC-Norfolk prospective population study. *PLoS Med* **5**, e12, doi:10.1371/journal.pmed.0050012 (2008).

34 Pietzner, M. *et al.* Plasma metabolites to profile pathways in noncommunicable disease multimorbidity. *Nat Med* **27**, 471-479, doi:10.1038/s41591-021-01266-0 (2021).
35 Tapanainen, J. *et al.* Short and long term effects of growth hormone on circulating levels of insulin-like growth factor-I (IGF-I), IGF-binding protein-1, and insulin: a placebo-controlled study. *J Clin Endocrinol Metab* **73**, 71-74, doi:10.1210/jcem-73-1-71 (1991).
36 Finer, S. *et al.* Cohort Profile: East London Genes & Health (ELGH), a community-based population genomics and health study in British Bangladeshi and British Pakistani people. *Int J Epidemiol* **49**, 20-21i, doi:10.1093/ije/dyz174 (2020).
37 de Onis, M. *et al.* Development of a WHO growth reference for school-aged children and adolescents. *Bull World Health Organ* **85**, 660-667, doi:10.2471/blt.07.043497 (2007).
38 Campbell, J. N. *et al.* A molecular census of arcuate hypothalamus and median eminence cell types. *Nat Neurosci* **20**, 484-496, doi:10.1038/nn.4495 (2017).
39 Sweeney, P. *et al.* The melanocortin-3 receptor is a pharmacological target for the regulation of anorexia. *Sci Transl Med* **13**, doi:10.1126/scitranslmed.abd6434 (2021).
40 Lam, B. Y. H. *et al.* Heterogeneity of hypothalamic pro-opiomelanocortin-expressing neurons revealed by single-cell RNA sequencing. *Mol Metab* **6**, 383-392, doi:10.1016/j.molmet.2017.02.007 (2017).
41 Romanov, R. A. *et al.* Molecular interrogation of hypothalamic organization reveals distinct dopamine neuronal subtypes. *Nat Neurosci* **20**, 176-188, doi:10.1038/nn.4462 (2017).
42 Chen, R., Wu, X., Jiang, L. & Zhang, Y. Single-Cell RNA-Seq Reveals Hypothalamic Cell Diversity. *Cell Rep* **18**, 3227-3241, doi:10.1016/j.celrep.2017.03.004 (2017).
43 Backholer, K. *et al.* Kisspeptin cells in the ewe brain respond to leptin and communicate with neuropeptide Y and proopiomelanocortin cells. *Endocrinology* **151**, 2233-2243, doi:10.1210/en.2009-1190 (2010).
44 Cocchi, D., De Gennaro Colonna, V., Bagnasco, M., Bonacci, D. & Muller, E. E. Leptin regulates GH secretion in the rat by acting on GHRH and somatostatinergic functions. *J Endocrinol* **162**, 95-99, doi:10.1677/joe.0.1620095 (1999).
45 Tannenbaum, G. S., Gurd, W. & Lapointe, M. Leptin is a potent stimulator of spontaneous pulsatile growth hormone (GH) secretion and the GH response to GH-releasing hormone. *Endocrinology* **139**, 3871-3875, doi:10.1210/endo.139.9.6206 (1998).
46 Wang, L. & Moenter, S. M. Differential Roles of Hypothalamic AVPV and Arcuate Kisspeptin Neurons in Estradiol Feedback Regulation of Female Reproduction. *Neuroendocrinology* **110**, 172-184, doi:10.1159/000503006 (2020).
47 Dunger, D. B., Ahmed, M. L. & Ong, K. K. Effects of obesity on growth and puberty. *Best Pract Res Clin Endocrinol Metab* **19**, 375-390, doi:10.1016/j.beem.2005.04.005 (2005).
48 Hauspie, R. C., Vercauteren, M. & Susanne, C. Secular changes in growth and maturation: an update. *Acta Paediatr Suppl* **423**, 20-27, doi:10.1111/j.1651-2227.1997.tb18364.x (1997).
49 Kuhnen, P. *et al.* Proopiomelanocortin Deficiency Treated with a Melanocortin-4 Receptor Agonist. *N Engl J Med* **375**, 240-246, doi:10.1056/NEJMoa1512693 (2016).
50 Roa, J. & Herbison, A. E. Direct regulation of GnRH neuron excitability by arcuate nucleus POMC and NPY neuron neuropeptides in female mice. *Endocrinology* **153**, 5587-5599, doi:10.1210/en.2012-1470 (2012).
51 Manfredi-Lozano, M. *et al.* Defining a novel leptin-melanocortin-kisspeptin pathway involved in the metabolic control of puberty. *Mol Metab* **5**, 844-857, doi:10.1016/j.molmet.2016.08.003 (2016).
52 Salomon, F., Cuneo, R. C., Hesp, R. & Sonksen, P. H. The effects of treatment with recombinant human growth hormone on body composition and metabolism in adults with

growth hormone deficiency. *N Engl J Med* **321**, 1797-1803, doi:10.1056/NEJM198912283212605 (1989).
53 Doherty, T. J. Invited review: Aging and sarcopenia. *J Appl Physiol (1985)* **95**, 1717-1727, doi:10.1152/japplphysiol.00347.2003 (2003).
54 McCance, R. A. & Widdowson, E. M. The determinants of growth and form. *Proc R Soc Lond B Biol Sci* **185**, 1-17, doi:10.1098/rspb.1974.0001 (1974).

Figure Legends

Figure 1 Non-synonymous variants of *MC3R* and association with phenotypes

(a-b) Dose-dependent cAMP accumulation activity of MC3R mutants stimulated by NDP-MSH. MC3R mutants are grouped by their functional classification: Loss-of-function (LoF) **(a)**; and wild-type like (WT-like) **(b)**. Data is normalised to % WT (black). Mean±SEM shown, N and p-values are listed in Table S3.

(c-d) Log_{10} half maximal effective concentration ($LogEC_{50}$) **(c)** and maximal relative efficacy (E_{max}) values **(d)** of MC3R mutants. Mean ± SEM shown, * indicates Bonferroni $p<0.05$ using one-way ANOVA, N and p-values are listed in Table S3.

(e) Heatmap showing the phenotypic association of *MC3R* variants in UKBB. Bonferroni threshold=0.0025.

(f-i) The effect of UKBB *MC3R* variants on the age at menarche (years) **(f)**, adult height (cm) **(g)**, ALM/BMI Ratio (m^2x1000) **(h)**, and plasma IGF1 level (nmol/L) **(i)**. Beta ± 95% CI shown, MAF and p-values are listed in Table S2.

Figure 2 Characteristics of Human homozygous for MC3R p.G240W mutation

(a) Whole body DEXA image of the MC3R p.G240W proband.

(b-c) Boxplots of % fat and lean mass in proband (orange) compared to South Asians males (n=36) **(b)** and European males aged 44-54 (n=417) **(c)** in UKBB. Centre=median; box=interquartile range (IQR); whiskers=1.5*IQR.

(d-e) Proband's total lean mass (kg, black circle) **(d)** and ALM/BMI (m^2) **(e)** compared to UKBB males with BMI from 18-52 (mean=blue, grey shade=95% CI; N=2356).

Figure 3 The role of MC3R in sexual maturation and oestrous cycle regulation

(a-b) Day of pubertal maturation as measured by the preputial separation in males **(a)** (N:WT=4; $Mc3r^{+/-}$=14; $Mc3r^{-/-}$=9, Kruskal-Wallis test, P=0.015); and first oestrous in females **(b)** (N:WT=4; $Mc3r^{+/-}$ =15; $Mc3r^{-/-}$ =7, Kruskal-Wallis test, P=0.280). Mean ± SEM shown.

(c) Quantification of the length of the oestrous cycle in WT (N:Fast=19, Fed=38) & $Mc3r^{-/-}$ mice (N:Fast=14,Fed=19) in either *ad libitum* fed and fasted conditions. Mean ± SEM shown. (2-way ANOVA with Bonferroni post-hoc, ns - not significant; * - $p\leq0.05$; **** - $p\leq0.0001$).

(d) Number of oestrous cycles per 15 days in WT and $Mc3r^{-/-}$ mice in *ad libitum* fed and fasted conditions. Mean ± SEM shown. (2-way ANOVA with Bonferroni post-hoc, ns - not significant; * - $p\leq0.05$; **** - $p\leq0.0001$).

Figure 4 *Mc3r* expression in murine hypothalamus

(a) Single-cell RNA sequencing of 1,166 *Mc3r*-expressing neurons reveal 11 distinct clusters, tSNE plot shown.

(b) tSNEs showing normalised expression of *Mc3r* (dark red), *Ghrh* (dark green), *Tac2* & *Kiss1* (dark blue),.

(c) smFISH showing co-localisation of *Mc3r* (white) with *Tac2* (magenta) and *Ghrh* (yellow) in a representative mediobasal hypothalamus (n=6 mice). Selected *Mc3r+Tac2* +ve and *Mc3r+Ghrh* double-positive neurons are indicated by red and green arrows, respectively. (3V - 3rd ventricle, Arc - arcuate nucleus, scale bars: left=500μm, right=50μm).

(d) The quantitation of *Mc3r* mRNA expression in *Tac2* and *Ghrh* neurons. Co-expression percentage±SEM is shown at the top (N=6 mice).

(e) The number of arcuate *Mc3r+Kiss1* co-expressing neurons at post-natal day (P)16, P28 and P48. (One-way ANOVA with Tukey's post-hoc, ns=$p>0.05$, N=3 mice for all age groups).

(f) There was an increase in the number of *Mc3r+Kiss1* co-expressing neurons with age in the AVPV. (2-tailed Student's unpaired t-test, ***$p<0.001$, N=4 mice for both P28 and P48).

Methods

In-vitro cyclic-AMP (cAMP) accumulation assay

Human Embryonic Kidney (HEK293) cells were obtained from lab stock and maintained with Dulbecco's Modified Eagle Medium High Glucose (Invitrogen, Carlsbad, CA, USA), supplemented with 10% fetal bovine serum (Invitrogen), 1% Glutamax (Invitrogen), 100 U/ml penicillin and 100 mg/ml streptomycin (Sigma-Aldrich, IL, USA). HEK293 cells were kept in 37˚C humidified air with 5% CO_2. The cell line was tested negative for mycoplasma contamination, it was not a commonly misidentified cell line and not authenticated.

Site-directed mutagenesis on wild-type Human N-FLAG-MC3R pcDNA3.1(+) was performed using Agilent QuikChange Lightning kit (Santa Clara, CA, USA) to generate all *MC3R* variants for cAMP activity measurement.

10ng of plasmid carrying *MC3R* wild-type (WT) and variants were transfected into HEK293 cells using Lipofectamine 3000 (Invitrogen) 48 hours prior to starting the cAMP assay. An increasing dose of [Nle^4, $D\text{-}Phe^7$]-α-melanocyte-stimulating hormone (NDP-αMSH, Bachem, Bubendorf, Switzerland) from 10^{-13} to 10^{-5}M was administered the following day for 2 hours in phosphate buffered saline (PBS, Sigma-Aldrich) before intracellular cAMP concentration measurement using a luminescence based HitHunter cAMP Assay for Small Molecules (Cat# DiscoverX 90- 0075SM25, Eurofins DiscoverX, Fremont, CA, USA) and Tecan Infinite M1000 Pro microplate reader. cAMP standard curve was measured for each experiment following the standard manufacturer protocol and used to transform luminescence values to cAMP concentrations for downstream analyses.

The baseline and maximal cAMP concentrations were normalised to MC3R WT from the same experiment and a 3-point sigmoidal dose-response curve was fitted to each individual replicate to determine the E_{max} and the $logEC_{50}$. The average relative maximal efficacy (E_{max}) and log half maximal effective concentration ($logEC_{50}$) values were used for LoF determination. The $logEC_{50}$ was not used for mutants which exhibited no response. One-way ANOVA was used to compare the E_{max} and $logEC_{50}$ of each MC3R mutation to the MC3R WT response. All calculations were performed with GraphPad Prism 7.

The LoF classifications were defined as follows:

complete LoF (cLoF): $E_{max} \leq 25\%$ WT or $EC_{50} \geq 50x$ WT
partial LoF (pLoF): $25\% WT < E_{max} \leq 75\%$ WT or $5x WT \leq EC_{50} < 50x$ WT
WT-like: $125\% WT > E_{max} > 75\%$ WT or $0.2x WT < EC_{50} < 5x$ WT

UK Biobank Phenotype association

Cohort Information

UK Biobank is a large and prospective study of ~500,000 participants aged 40-69 years, recruited between 2006 and 2010[55]. All analyses conducted using the UK Biobank Resource were done under application numbers 32974 and 44448.

Phenotype Measurements

We considered candidate anthropometric, puberty timing, and metabolic traits. The following specific filters were used: age at menarche was filtered to correspond to the ReproGen consortium definition[56] for analyses conducted with genotyping array and imputed data and the full cohort (~100,000 female participants) was used for the whole-exome sequencing (WES) data. Type 2 diabetes was identified on the basis of probable diabetes[57] plus any mention of E11 in HES (main or secondary) or death (underlying or contributory cause); body composition variables (total lean mass and appendicular lean mass) were derived from prediction equations based on demographic, anthropometric and bioelectrical impedance values[58]. Waist-to-hip ratio was adjusted for BMI and the residuals from this were rank-based inverse normally transformed.

UK Biobank WES data processing and QC

The VCF and PLINK files for whole exome sequencing (WES) data of 200,643 UK Biobank participants, made available in October 2020, were downloaded and used for the analysis. The data processing and QC were performed as previously described[59]. The QC filters used were: QUAL (variant site-level quality score); and AQ (variant site-level allele quality score) between 20-99. We also defined a heterozygous genotype call as imbalanced if allelic balance ≤ 0.25 or ≥ 0.8 and excluded it from the analysis.

UKBB WES Variant annotation

We annotated the *MC3R* variants using Ensembl Variant Effect Predictor (VEP) tool release 99 based on Human genome build GRCh38. CADD v1.6 VEP plugin was used to provide prediction scores for deleteriousness.

WES Gene Burden Tests

We selected all rare alleles (MAF<0.2%) in *MC3R* which were annotated as "HIGH" or "MODERATE" impact by VEP, excluding those that were annotated as benign by PolyPhen2. Gene burden scores were created by collapsing variants above to define a binary call denoting whether an individual carries none versus one or more rare, predicted damaging alleles in *MC3R*. The reported effect estimates represent the trait difference between *MC3R* mutation carriers and non-carriers. These dummy variables were then transformed into BGEN genotype call format for association testing using BOLT-LMM[60]. Only common, autosomal variants that passed the QC and were present on both genotyping array types in UKBB were included in the genetic relationship matrix (GRM). Genotyping array type, age at baseline and first ten genetically derived principal components were included as covariates. Samples were excluded from analysis if they failed UK Biobank QC, were of non-European ancestry or if the participant withdrew consent from the study.

Selection of variants from UKBB

In order to identify directly genotyped variants covered on the UK Biobank Axiom array (Affymetrix) we extracted genotype counts in the coding region of *MC3R* available in UK Biobank using plink v1.9[61]. Genotyping quality was assessed using plink v1.9 and cluster plots of raw genotype intensity data. Variants which have a MAF > 0.05% were taken forward in

analysis **(Table S3).** Variant Effect Predictor (VEP v99)[62] and CADD (v1.6)[63] were used to annotate the extracted variants and assess their predicted deleteriousness.

Genotype Measurements

Genotypes: Imputed genotype data were used for 2 variants – rs3827103 and rs61735259, to maximise sample size (Info score >0.96). Directly genotyped data was used for rs143321797 due to its low MAF (0.06%), and genotype cluster plots were manually inspected to ensure genotype reliability[64]. Furthermore, genotype concordance for non-reference carriers was examined across WES, genotyping array and imputation for rs143321797 and rs61735259 **(Table S13).**

Statistical analyses

Individual variant associations with outcomes were assessed under additive genetic models. For the individual variants, associations were tested using mixed linear models implemented in BOLT-LMM[60], which allow the inclusion of related individuals. Phenome-wide analyses were performed in up to 451,301 individuals. The variant- based models were performed adjusted for age, sex (where appropriate) and the first 10 genetic principal components as provided by UKBB[65], with two outcomes additionally controlling for height where this is stated.

Phenome-wide association study

A phenome-wide association study (pheWAS) was conducted using publicly available genome-wide analysis (GWAS) summary statistics from five different repositories: GWAS of 633 ICD10-coded disease phenotypes from the UKBB provided by the Neale lab (http://www.nealelab.is/uk-biobank) where data was systematically coded using an algorithm based approach to determine the most appropriate analysis[66], Open Targets Genetics[67], Open GWAS IEU[68], Global Biobank Engine[69], and Phenoscanner[70]. Summary statistics were extracted for the three coding *MC3R* variants rs3827103, rs143321797 and rs61735259.

We considered studies with > 5,000 individuals and excluded binary traits where there were less than 0.1% of cases in the cohort. We manually pruned the list of phenotypes to retain only non-redundant traits by choosing the largest available study covering all variants in cases where a phenotype available included in multiple data sets.

We used *grs.summary()* function from R package gtx (v0.0.8; http://cran.r-project.org/web/packages/gtx) , which enables multi-SNP genetic risk score analysis using single SNP summary statistics, across 478 traits for which summary statistics for all three variants were available. Weights for each variant's CADD Phred-score (v1.6) were used in the analysis: rs143321797-C=26.2; rs6173525-A=23; rs3827103-A=19. We used a Bonferroni significance threshold to control for multiple testing ($P< 1.046x10^{-4}$).

The Avon Longitudinal Study of Parents and Children (ALSPAC)

Cohort information

The Avon Longitudinal Study of Parents and Children (ALSPAC) is a prospective birth cohort from the southwest of England established to study environmental and genetic characteristics

that influence health, development and growth of children and their parents[71,72]. Full details of the cohort and study design have been described previously and are available at http://www.alspac.bris.ac.uk.

Ethical approval for the study was obtained from the ALSPAC Ethics and Law Committee and the Local Research Ethics Committees. Written informed consent was obtained from mothers at recruitment, from the main carers (usually the mothers) for assessments on the children from ages 7 to 16 years and, from age 16 years onwards, the children gave written informed consent at all assessments. Consent for biological samples has been collected in accordance with the Human Tissue Act (2004) and informed consent for the use of data collected via questionnaires and clinics was obtained from participants following the recommendations of the ALSPAC Ethics and Law Committee at the time.

Measurements

Weight and length of each participant were measured at birth and at 4, 8, 12 and 18 months. Weight (to the nearest 50g) and height (to the nearest millimetre) were measured from 25 months to 24 years. For weight, the participant was encouraged to pass urine and undress to their underclothes. For height, children were positioned with their feet flat and heels together, standing straight so that their heels and shoulders came into contact with the vertical backboard. Equipment used for each measurement were comparable (e.g. Fereday 100kg combined scale, Soenhle scale, Seca scale and Tanita Body Fat Analyser for weight, and Harpenden Neonatometer or Stadiometer, Kiddimetre and Leicester measure for height). Growth trajectories were carried out using linear spline multilevel modelling of weight and height from birth to when participants were 10 years. Any missing clinic values were replaced with age-specific predicted values from growth trajectories[73].

Fat and lean masses (kg) were measured when participants were a mean age of 10, 12, 14, 15, 18 and 24 years using the Lunar prodigy narrow fan beam densitometer dual energy x-ray absorptiometry (DEXA) scanner.

Puberty onset was defined by age at menarche in females and age at peak height velocity (APV) in males. Age at menarche was assessed from up to nine annual postal questionnaires relating to pubertal development completed by the participants from the age of 8 to 17 years. Each questionnaire asked whether menstrual periods had started and, if so, at what age. Earlier questionnaires were completed by the study mothers on their daughter's behalf and, from about age 15 years, the questionnaires were completed by the study child. The first-reported age at menarche was used. APV was estimated using Superimposition by Translation and Rotation (SITAR) growth curve analysis, using height measurements taken between ages 5 and 20 years[74].

Full details of all measures used in this study are available on the online dictionary: http://www.bristol.ac.uk/alspac/researchers/our-data/.

Pooled high-throughput sequencing (HTS) of MC3R in ALSPAC

The pooled sequencing workflow of *MC3R* was conducted as previously described[75]. Briefly 20ng DNA samples representing 5,993 unrelated individuals used in analyses were randomly combined into pools of 50 at the Medical Research Council Biorepository Unit. 10ng of pooled DNA was used for *MC3R* single-exon PCR with Q5 Hot Start High-Fidelity DNA Polymerase (NEB, Ipswich MA, USA) and *MC3R* exon primers -331 bp upstream (5'-TGGAACAGCAAAGTTCTCCCT-3') and +61 bp downstream (5'-CCTCACGTGGATGGAAAGTC-3') of the protein coding region yielding a PCR product of 1375bp. The PCR product was purified using Agencourt Ampure XP beads (Beckman Coulter, Brea CA, USA), and quantified with QuantiFluor dsDNA system (Promega, Wadison, WI, USA) using Tecan Infinite M1000 Pro plate reader (Männedorf, Switzerland). 1ng of purified PCR product was used to construct the sequencing libraries with Nextera XT Library Preparation Kit and Nextera XT Index V2 barcodes (Illumina, San Diego, CA, USA) according to manufacturer's instruction. Ampure Xp beads were used to purified the libraries, which were then quantified using Kapa Library quantification kit (Roche, Basel, Switzerland) and Quantstudio 7 Flex Real Time PCR instrument (ThermoFisher Scientific, Waltham, MA, USA). All libraries were combined at 10nM for paired-end sequencing at 150bp (PE150) on the Illumina HiSeq4000 instrument at the CRUK Cambridge Institute Genomics Core. An even coverage was achieved, with a mean per-pool per-base sequencing depth at 45,490 ± 436-fold (SEM, data not shown), throughout the protein coding region of *MC3R*.

HTS Sequencing Bioinformatics

BWA MEM (0.7.12)[76] was used to align the sequence reads to Human GRCh38 (hg38) genome. PCR de-duplication was performed using Picard 1.127 (https://broadinstitute.github.io/picard/). GATK 3.8 (https://gatk.broadinstitute.org/) was used to perform indel realignment and base quality score recalibration according to GATK Best Practices. The variants were called by mpileup2snp and mpileup2indel function from Varscan 2.4.2[77] with the following parameters: variant coverage ≥100X, 'Strand Filter' = 'ON', allele frequency (VAF) ≥0.6% and p-value <0.05.

ALSPAC MC3R LoF variant selection

Using the HTS we initially identified 20 non-synonymous variants in *MC3R* (data not shown). Seven variants were predicted deleterious by SIFT, Polyphen-2 and CADD v1.6 (**Table S3**) and were taken forward for functional characterisation for their cAMP activity. We identified 3 complete loss-of-function (cLoF) variants p.F45S, p.L53R, p.A214P. Subsequently we went back to the original ALSPAC DNA samples and validated 6 heterozygous carriages: 4X p.F45S, 1X p.L53R and 1X p.A214P via traditional Sanger sequencing described below.

Sanger Sequencing for variant validation and carrier identification

Original DNA samples from participants were validated by Sanger sequencing. The *MC3R* coding region was first amplified, using GoTaq Green (Promega) Master Mix with 10ng DNA per 10µl PCR reaction and *MC3R* exon primers (as above). *MC3R* PCR cycling conditions were as follows: one cycle of 95˚C for five minutes; 35 cycles of 95˚C for 30 seconds, 60˚C for 30 seconds and 72˚C for 2 minutes and one cycle of 72˚C for five minutes.

Unincorporated primers and dNTPs were removed using 20 units Exonuclease I (Exo) (NEB) and 1 unit of Shrimp Alkaline Phosphatase (SAP) (NEB) at 37˚C for 20 minutes and then 80˚C for 15 minutes. 1µl of the Exo/SAP product was used in the Sanger sequencing reactions with 0.5µl of BigDye Terminator v3.1, 2µl 5x Sequencing buffer, 0.5µM sequencing primer and up to 10µl using Nuclease free water. The Sanger sequencing cycling conditions were 24 cycles of 95˚C for 10 seconds, 50˚C for five seconds, 60˚C for four minutes.

Sanger sequencing reactions was purified using AxyPrep MAG PCR Clean-Up Kit (Axygen, Corning Inc., Somerville, MA, USA) according to manufacturer's instructions. Purified sequencing products were resuspended in 30µl nuclease free water and analysed on a 3730 DNA Analyzer (Thermofisher). The data was analysed on Sequencer 4.8 Build 3767 (Gene Codes Corporation, Ann Arbor, MI, USA).

Associations between MC3R and anthropometric traits and puberty onset

Of the 5,993 individuals sequenced, five individuals had missing identifier information for linkage with the wider ALSPAC data set and 214 individuals were duplicated; therefore, these exclusions left 5,774 participants in the sequenced set. After merging in all required clinic and questionnaire data from the ALSPAC cohort and excluding related individuals (details on these exclusions are available[78]), 5,724 remained in the sequenced set for all analyses, 5,717 of which had complete information on sex, comprising the final sample for analyses. We grouped the six *MC3R* mutation carriers of three identified *MC3R* cLoF mutations into the 'MC3R mutations' group.

The associations of the *MC3R* LoF with body mass index (BMI), height, weight, lean mass, and fat mass at different ages, and age at puberty onset were assessed using linear regression. All analyses and estimates, except for age at puberty onset, were adjusted for age and sex.

The Fenland Study

Cohort Information

The Fenland study is a population-based cohort of 12,435 participants born between 1950 and 1975 who underwent detailed phenotyping at the baseline visit between 2005-2015, which has previously described in detail[79]. The study was approved by the Cambridge Local Research Ethics Committee (ref. 04/Q0108/19) and all participants provided written informed consent. Briefly, the participants were recruited from general practice surgeries in the Cambridgeshire region in the UK. Individuals were not enrolled in the cohort if they were clinically diagnosed diabetes mellitus, a terminal illness or psychotic disorder, unable to walk unaided, or were pregnant or lactating.

Measurements

Proteomic profiling has previously been described[80,81]. Proteomics profiling was performed on fasted EDTA samples collected at baseline by SomaLogic Inc. (Boulder, CO, USA) using DNA aptamer-based technology. Relative protein abundances of 4,775 human protein targets were evaluated by 4,979 aptamers (SomaLogic V4).

Statistical analyses

10,708 Fenland participants had both phenotypes and genetic data after excluding ancestry outliers and related individuals. Association analyses for variants of interest was performed as described previously[81]. Briefly, within the 3 genotyping subsets, aptamer abundances were transformed to follow a normal distribution using rank-based inverse normal transformation, and were then adjusted for age, sex, sample collection site and first 10 genetic principal components. The residuals were then used as input for the genetic association analyses using an additive model with BGENIE (v1.3)[65]. The results for the three genotyping arrays were combined in a fixed-effects meta-analysis in METAL (v 2011-03-25)[82].

We first prioritised a total of 14 proteins from the insulin-like growth factor family of proteins targeted by 15 aptamers at a rigorous Bonferroni significance threshold ($p< 0.0033$). We further considered all proteins targeted by the platform at a lenient multiple testing threshold of $P < 1x10^{-4}$.

The European Prospective Investigation of Cancer (EPIC)-Norfolk study

Cohort Information

The European Prospective Investigation of Cancer (EPIC)-Norfolk study is a prospective cohort of 25,639 individuals aged between 40 and 79 and living in the county of Norfolk in the United Kingdom at recruitment. The study was approved by the Norfolk Research Ethics Committee (REC 500 ref. 98CN01) and all participants gave their written consent before entering the study[83].

Measurements

Genotyping, imputation and untargeted metabolite profiling of baseline non-fasted serum samples from 9,712 unrelated European individuals in the EPIC-Norfolk cohort was performed using the Discovery HD4 platform (Metabolon, Inc., Durham, USA), as previously described[84,85].

Statistical Analysis

Linear regression models adjusted for age, sex, time of blood sample, time fasting and the first 10 genetic PCs were run for each *MC3R* variant and metabolite pair in R (v3.6.0). A total of 656 metabolites with a known chemical identity were included in the analysis. Statistical significance was considered at a Bonferroni significance threshold of $P < 7.6x10^{-5}$.

Genes & Health

Cohort Information

Genes & Health is an ongoing community-based population study comprising (at 31st August 2021) 48,960 British Bangladeshis and British Pakistanis[86]. Genes & Health operates under approval from the National Research Ethics Committee (London and South-East), and Health Research Authority (reference 14/LO/124), and Queen Mary University of London is the Sponsor. Genes & Health incorporates Stage 1 (health record access, saliva DNA collection) on all volunteers and Stage 2 (focused recall studies) procedures on selected volunteers, including recall-by-genotype. Exome sequencing has been performed on all volunteers

reporting parental relatedness (n=5,236) and genotyping (Illumina GSAv3EA+MD chip) on all. Informed consent is taken at both Stage 1 and Stage 2, and allows analysis of health and genetic data and publication of results.

Identification of MC3R variants in Genes & Health

Non-synonymous variants for *MC3R* were identified from public exome data available on the Genes & Health (G&H) website (https://genesandhealth.org), summary data downloaded in September 2019. The exome sequencing of G&H is described in[87].

Genes & Health clinical recall and measurements

The Genes & Health proband was recruited and recalled to the study under Stage 2 procedures described above. Clinical assessment was performed using standard operating protocols designed for metabolic phenotyping in the Genes & Health study, and were performed by qualified medical staff and a bilingual research assistant. All measurements were taken wearing light clothing and with footwear removed, and after voiding urine and after a 10-hour fast. Height was measured in cm (to the nearest 0.5cm) using a stadiometer, with feet spaced slightly apart with the back of heels and buttocks touching the stadiometer and facing straight ahead. Weight was measured (to the nearest 0.1kg) using a Tanita TBF-300 scales and body composition analyser. Blood pressure was measured (to the nearest 1mm/Hg) using a GE Carescape V100 automated blood pressure monitor.

Whole body DEXA scanning (Hologic, Horizon W, S/N 100091, Auto Whole Body protocol), was performed as part of routine clinical care within the National Health Service, one month after the research clinical assessment. Height (155.0cm) and weight (96.96kg) were re-measured at the time of scanning were consistent with the research assessment (height 155.0cm and 97.8kg). The DEXA-derived values have been used to compute all DEXA-based measurements, including lean and fat mass. We calculated sitting height and sitting height ratio the skeletal views from the DEXA scan. Anatomical landmarks were used to calculate the sitting height (upper border of the skull to the superior border of the greater trochanter), and the standing height (upper border of the head to the base of the calcaneum, proportioned to clinical height measurement).

Venepuncture was undertaken after a 10 hour overnight fast, using a Vacutainer system. Blood plasma was separated from lithium heparin tubes, collected and stored on ice, for insulin, c-peptide, leptin and adiponectin assays. Blood serum was obtained using serum separator tubes for lipid and bone profile, liver and renal function, follicle-stimulating and luteinising hormone, testosterone, thyroid function tests, sex-hormone binding globulin, cortisol (collected at room temperature), and insulin-like growth factor-1 (collected on ice). Adrenocorticotropic hormone was assayed from plasma collected using an EDTA tube on ice. Full blood count and haemoglobin A1c were assayed from EDTA whole blood, and plasma glucose from a fluoride oxalate tube. All samples were assayed at the University of Cambridge Core Biochemistry Assay Laboratory.

G&H Proband comparison to UKBB

The G&H proband was compared to males who have DEXA imaging data available in UKBB. This cohort was further stratified by self-reported ethnicity (field 21000) into European males, the majority of the cohort, (N=2,367; 2,356 with both BMI and DEXA measures) and South Asian males (N=36) to allow matched assessment of the proband with individuals of the same ethnic background. South Asian ethnicity was defined as individuals who reported to be of Indian, Bangladeshi and Pakistani ethnicity. Total lean and total fat percentage were compared with age matched males of European ethnicity to account for age effect. These included males within a 10-year span closest to that of the proband - aged 44-54 at the second study visit when DEXA images were obtained (N=417).

Individuals with missing data were removed from the comparison. Percentage of lean and fat mass were calculated using DEXA total lean and fat mass variables and total mass as defined by the DEXA measurements. BMI at the second health check was used to allow comparison across different BMI ranges, to match the study visit when DEXA images were obtained. Appendicular lean mass was calculated using the sum of lean mass from legs and arms in kilograms, divided by BMI. Z-scores of these measures were calculated to aid cross-trait comparison within these subgroups of interest.

Laboratory animals

Mouse strains used in the reproductive function of MC3R included C57BL/6J (the Jackson laboratory) and *Mc3r*-knockout (bred in-house at the University of Michigan). Male and female mice were group-housed at 20-24°C with a 12-hour light/12-hour dark cycle and provided *ad libitum* access to food. The experiments were previously approved by the University of Michigan and Vanderbilt University Institutional animal care and use offices (Institutional Animal Care and Use Committee).

Mouse studies performed in Cambridge was in accordance with UK Home Office Legislation regulated under the Animals (Scientific Procedures) Act 1986 Amendment, Regulations 2012, following ethical review by the University of Cambridge Animal Welfare and Ethical Review Body (AWERB). For the adult *in situ* hybridisation experiments, 3 adult male + 3 female (*Tac2+Ghrh+Mc3r*) and one female + two males (*Kiss1+Tac2+Mc3r*) C57BL/6J mice at 6-8 weeks were housed in individually ventilated cages in a controlled temperature (20-24°C) facilities with a 12-h light/dark cycle (lights on 06:00–18:00) and *ad libitum* access to food and water in the animal facility at the Anne McLaren Building, University of Cambridge.

Human post-mortem tissue

An anonymised human hypothalamic tissue sample was provided by the Cambridge Brain Bank from a female donor aged 95 at the time of death. The donor gave informed written consent for the use of tissue for research, and samples obtained were used in accordance with the Research Ethics Committee Approval number 10/H0308/56.

Assessment of puberty onset and fertility

Puberty onset in wild type, *Mc3r*$^{+/-}$, and *Mc3r*$^{-/-}$ was determined by daily examination for preputial separation in males. First oestrous in females was identified by daily vaginal smears. To visualize first oestrous, the vaginal cells were flushed by introducing 100µl of sterile saline using a sterile transfer pipette. The saline was slowly released into the vagina and drawn back into the tip; this was repeated 4 to 5 times in the same sterile pipette and the cell suspension was then transferred into a 24 well plate. The fluid was then mounted onto a glass slide and the smear was viewed on an inverted compound light microscope.

For the fasting study, animals were randomised and were either fasted or left *ad libitum* fed for one overnight before the assessment of their oestrous cycle progression.

The researchers were blinded to the genotype/treatment for the experiments. Power calculation was performed, N is shown in the corresponding figure legends.

Single-molecule fluorecent in-situ hybridisation (smFISH)

For the *Mc3r* expression in adult mice, animals were euthanised with a lethal administration of sodium pentobarbital of (50mg/kg) intraperitoneally and were perfused with 10% formalin in PBS. The brains were excised after the perfusion and further fixed in 10% formalin in for 24hrs at 4°C. The following day the brains were immersed in 25% sucrose and ProClin 300 (1:2000, Sigma) in PBS solution and kept at 4°C. After 24hrs the brains were embedded in optimal cutting temperature (OCT) compound and frozen in Novec 7000 (Sigma) and dry ice, followed by -80°C storage until use.

16µm cryosections containing the hypothalamus were prepared on a Leica CM1950 cryostat (Wetzlar, Germany) at -12°C. For smFISH, sections were baked at 65°c for 1 hour and fixed in 4% PFA solution at 4°C for 15mins. Slides were then washed and dehydrated in PBS and ethanol gradients from 50% to 100% for a total of 30 mins. Slides were air dried.

For the human smFISH, a fresh tissue block of human hypothalamus was fixed in 10% neutral buffered formalin at room temperature for 24h, transferred to 70% ethanol, and processed into paraffin. 6 µm sections were cut and mounted onto Superfrost Plus slides (ThermoFisher) in an RNase free environment, and then dried overnight at 37°C. Sections containing the mediobasal hypothalamus were deparaffinised, rehydrated.

Multiplex smFISH was performed as previously described[88] on a Leica Bond RX automated stainer, using RNAScope Multiplex Fluorescent V2 reagents (Advanced Cell Diagnostics (ACD), Newark, CA, USA). Slides underwent heat-induced epitope retrieval with Epitope Retrieval Solution 2 (Leica) at 95°C for 5 mins. Slides were then incubated in RNAScope Protease III reagent at 42°C for 15 mins, before being treated with RNAScope Hydrogen Peroxide for 10 mins at RT to inactivate endogenous peroxidases. Double-Z mRNA probes for mouse *Ghrh* (Mm-Ghrh-C2), *Tac2* (Mm-Tac2-C3), *Kiss1*(Mm-Kiss1-C4), *Mc3r* (Mm-Mc3r), and human *MC3R* (Hs-MC3R) *GHRH* (Hs-GHRH-C2), and *KISS1* (Hs-KISS1-C3) were designed by ACD for RNAScope on Leica Automated Systems. Slides were incubated in

RNAScope 2.5 LS probes for 2 hours at RT. DNA amplification trees were built through consecutive incubations in AMP1 (preamplifier), AMP2 (background reduction) and AMP3 (amplifier) reagents for 15 to 30 mins each at 42°C. Slides were washed in LS Rinse buffer between incubations. After amplification, probe channels were detected sequentially via HRP–TSA labelling. To develop the C1–C3 probe signals, samples were incubated in channel-specific horseradish peroxidase reagents for 30 mins, TSA fluorophores for 30 min and HRP-blocking reagent for 15 min at 42 °C. The probes in C1, C2 and C3 channels were labelled using Opal 520 (Akoya Biosciences, Marlborough, MA, USA), Opal 570 (Akoya), and Opal 650 (Akoya) fluorophores (diluted 1:500) respectively. Samples were then incubated in DAPI (Sigma-Aldrich, 0.25µg /ml) for 20 mins at room temperature to mark cell nuclei. Slides were mounted using ~90 µl of Prolong Diamond Antifade (ThermoFisher) and standard coverslips (24×50 mm^2; ThermoFisher). Slides were dried at RT for 24 hrs before storage at 4°C. Image were acquired using a Perkin Elmer (Waltham, MA, USA) CLS Operetta high-content screening confocal microscope using 5X and 40X objectives with Harmony software version 4.9. Randomisation and blinding were not relevant as these were observational for *in situ* studies with no sample groups. No prior power calculation was performed.

For the study of *Mc3r* expression in the hypothalamic arcuate nucleus (ARC) and anteroventral periventricular nucleus (AVPV) in female mice from a prepubertal to a postpubertal state, the animals were randomised and brains harvested at age post-natal day (P)16, P28 and P48. No prior power calculation was performed, N is shown in the corresponding figure legends. The animals were anesthetized with tribromoethanol and perfused transcardially with saline followed by fixative (4% paraformaldehyde in borate buffer, pH 9.5). Brains were post-fixed in a solution of 20% sucrose in fixative and cryoprotected in 20% sucrose in 0.2M potassium phosphate buffered saline (KPBS). Four series of 20 µm-thick frozen sections were collected using a sliding microtome. Sections containing the ARC or AVPV were mounted onto SuperFrost Plus slides (ThermoFisher), and *in situ* hybridization was performed according to the RNAscope fluorescence multiplex kit user manual for fixed frozen tissue (ACD) using RNAscope Probe (Mm-Mc3r-C) and (Mm-Kiss1-O1-C3). Images of the ARC and AVPV of each animal were obtained using a laser scanning confocal microscope (Zeiss LSM 800). Confocal image stacks were collected through the z-axis at a frequency of 0.8 µm using a 20x objective (NA 0.8). The researcher was blinded to the age of the animals for this experiment.

Imaging analysis

For the adult mouse and human study, data from Harmony (v4.9) was converted into OME TIFF pyramidal format. Individual imaging fields were collapsed along z-axis into max projections and subsequently flatfield corrected. Microscope registered coordinates were then used to tile mosaics of all imaging fields in the dataset. OME TIFF files were then read into QuPath v0.2.3[89] for analysis. Hypothalamic regions were annotated in QuPath, within which StarDist[90] was used for nuclear segmentation using the pre-trained 'dsb2018_heavy_augment' machine learning model with the default settings. Segmented nuclei were expanded by 2.5 µm to estimate the cell boundary. Cells were classified as *Ghrh* or *Tac2* positive by based on

median channel intensity within the nuclear region, and the subcellular detection algorithm was use to count the number of *Mc3r* spots within each cell. The data were exported into .csv format for downstream analysis.

For the developmental study in mice, three-dimensional representations of labeled cells were digitally rendered using Imaris software (version 9.2.0, Bitplane). To determine overall *Mc3r* mRNA abundance, a region of interest (ROI) was placed around either the ARH or AVPV and the total density of *Mc3r* labeling was quantified using the spots function. Total numbers of labeled *Kiss1* neurons in the ARC, *Kiss1* in the AVPV, as well as numbers of these neuronal populations that co-express *Mc3r*, were counted manually in each image stack, aided by Imaris software (Bitplane, v9.3). Only neurons with labeling that was 3 times that of background were considered positively labeled for *Mc3r* mRNA. Background for each section was determined by placing ten cell-sized ROIs in user-defined areas, where *Mc3r* labeling appeared to be lacking, and averaging the number of spots counted in each background ROI.

Single-cell RNA Sequencing data analysis

Raw sequence reads from published murine hypothalamic single-cell studies were obtained from Gene Expression Omnibus (GEO accessions GSE93374, GSE87544, GSE92707 and GSE74672, https://www.ncbi.nlm.nih.gov/geo/). Experimental details for the datasets are listed in **Table S10**.

For dropseq experiments GSE93374 and GSE87544, the 3' adaptor of the biological read was first trimmed with Cutadapt 1.16 using 'AAAAAA', the trimmed read was subsequently mapped with RNA STAR 2.7.5b[91] to the mouse GRCm38 genome. Read 1, which contained the cell barcode (12nt) and the UMI (8nt) was first split using fastxtrimmer (http://hannonlab.cshl.edu/fastx_toolkit/) and then Fgbio 1.1.0 (http://fulcrumgenomics.github.io/fgbio/) was used to attach information back onto the mapped data generated from Read 2. Gene-level unique molecular identifer (UMI) count was performed using Dropseq tools 2.3.0 (https://github.com/broadinstitute/Drop-seq/) with a modified gene model from Ensembl V100, where the predicted gene *Gm28040* was removed to recover reads for *Kiss1*. For smart-seq2 experiments GSE92707 and GSE74672, reads were mapped to the mouse GRCm38 genome and gene-level expression was counted using STAR 2.5.0a with Ensembl V100 gene model.

Gene-level counts from all 4 datasets were processed separatedly using Seurat v3.2[92]: Count data was normalised and scaled using the default options. Variable gene expression was determined using 'VST' selection method and cell clustering was performed using shared nearest neighbours (SNN) algorithm using the defaults. Clusters with high *Snap25* and *Syt1* expression were considered neuronal and were extracted for subsequent integration analysis. Pre-integration, cells with detectable *Olig1* in each of the datasets were removed. For GSE93374, we detected contaminating red blood cells and they were removed using the expression of *Hba-a1*, *Hba-a2*, *Hbb-bs* and *Hbb-bt*). For GSE74672, cells from animals treated with PFA were also removed from the downstream analysis.

The integration of the 4 neuronal datasets was performed using the Seurat v3[92] standard integration workflow: Briefly, the raw counts datasets were renormalised and variable features determined by 'mvp', followed by the use of canonical correlation analysis and mutual nearest neighbours algorithm with 'ndims' = 50 and 'k.filter = 150' to integrate the 4 datasets into a single 18,427-neuron superset. The integrated data was rescaled, 30 PCs were re-calculated via principal component analysis (PCA) and used for t-distributed stochastic neighbor embedding (TSNE) and SNN clustering analysis with 'resolution' = 1 to generate the 28 final clusters. Characteristic gene markers for each cluster were determined using the non-parametric Wilcoxon rank sum test and the marker list is available in **Table S11.**

For *Mc3r* subset, the cells were selected by their expression of *Mc3r* (raw count ≥ 1). Similar to above the subset was re-clustered using 25 PCs and a SNN resolution of 1. Characteristic gene markers for each cluster were determined using the non-parametric Wilcoxon rank sum test and the marker list is available in **Table S12.**

Data Availability

All data used in genetic association analyses are available from the UK Biobank upon application (https://www.ukbiobank.ac.uk).

Data from the Fenland cohort can be requested by bonafide researchers for specified scientific purposes via the study website (https://www.mrc-epid.cam.ac.uk/research/studies/fenland/information-for-researchers/). Data will either be shared through an institutional data sharing agreement or arrangements will be made for analyses to be conducted remotely without the necessity for data transfer.

The EPIC-Norfolk data can be requested by bona fide researchers for specified scientific purposes via the study website (https://www.mrc-epid.cam.ac.uk/research/studies/epic-norfolk/). Data will either be shared through an institutional data sharing agreement or arrangements will be made for analyses to be conducted remotely without the need for data transfer.

ALSPAC data are available through a system of managed open access. Full details of the cohort and study design have been described previously and are available at http://www.alspac.bris.ac.uk. Please note that the study website contains details of all the data that are available through a fully searchable data dictionary and variable search tool (http://www.bristol.ac.uk/alspac/researchers/our-data/). Data for this project were accessed under the project number B2891. The application steps for ALSPAC data access are as follows:(1) Please read the ALSPAC access policy, which describes the process of accessing the data in detail and outlines the costs associated with doing so. (2) You may also find it useful to browse the fully searchable research proposals database, which lists all research projects that have been approved since April 2011. (3) Please submit your research proposal for consideration by the ALSPAC Executive Committee. You will receive a response within 10 working days to advise you whether your proposal has been approved. If you have any questions about accessing data, please email alspac-data@bristol.ac.uk.

Genes & Health: Data is available via http://www.genesandhealth.org/

Publicly available GWAS datasets utilised in pheWAS analyses are available from the Neale Lab: http://www.nealelab.is/uk-biobank, Open Targets Genetics: https://genetics.opentargets.org/, Global Biobank Engine: https://biobankengine.stanford.edu/, Open GWAS IEU: https://gwas.mrcieu.ac.uk/, Phenoscanner: http://www.phenoscanner.medschl.cam.ac.uk/

Mouse single-cell RNA sequencing data is available from GEO accessions GSE93374, GSE87544, GSE92707 and GSE74672 (https://www.ncbi.nlm.nih.gov/geo/)

Code Availability

We wrote programming scripts to assist in the execution of publicly available functions and computer programs in our compute environment. For access to these scripts readers may contact the corresponding author.

Method References

55 Sudlow, C. *et al.* UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS Med* **12**, e1001779, doi:10.1371/journal.pmed.1001779 (2015).

56 Day, F. R. *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* **49**, 834-841, doi:10.1038/ng.3841 (2017).

57 Eastwood, S. V. *et al.* Algorithms for the Capture and Adjudication of Prevalent and Incident Diabetes in UK Biobank. *PLoS One* **11**, e0162388, doi:10.1371/journal.pone.0162388 (2016).

58 Powell, R. M. *et al.* Development and validation of total and regional body composition prediction equations from anthropometry and single frequency segmental bioelectrical impedance with DEXA. *medRxiv*, MEDRXIV/2020/248330 (2020).

59 Zhao, Y. *et al.* GIGYF1 loss of function is associated with clonal mosaicism and adverse metabolic health. *Nat Commun* **12**, 4178, doi:10.1038/s41467-021-24504-y (2021).

60 Loh, P. R. *et al.* Efficient Bayesian mixed-model analysis increases association power in large cohorts. *Nat Genet* **47**, 284-290, doi:10.1038/ng.3190 (2015).

61 Chang, C. C. *et al.* Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience* **4**, 7, doi:10.1186/s13742-015-0047-8 (2015).

62 McLaren, W. *et al.* The Ensembl Variant Effect Predictor. *Genome Biol* **17**, 122, doi:10.1186/s13059-016-0974-4 (2016).

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

64 Van Hout, C. V. *et al.* Exome sequencing and characterization of 49,960 individuals in the UK Biobank. *Nature* **586**, 749-756, doi:10.1038/s41586-020-2853-0 (2020).

65 Bycroft, C. *et al.* The UK Biobank resource with deep phenotyping and genomic data. *Nature* **562**, 203-209, doi:10.1038/s41586-018-0579-z (2018).

66 Millard, L. A. C., Davies, N. M., Gaunt, T. R., Davey Smith, G. & Tilling, K. Software Application Profile: PHESANT: a tool for performing automated phenome scans in UK Biobank. *Int J Epidemiol* **47**, 29-35, doi:10.1093/ije/dyx204 (2018).

67 Ghoussaini, M. *et al.* Open Targets Genetics: systematic identification of trait-associated genes using large-scale genetics and functional genomics. *Nucleic Acids Res*, doi:10.1093/nar/gkaa840 (2020).
68 Elsworth, B. *et al.* The MRC IEU OpenGWAS data infrastructure. *bioRxiv* (2020).
69 McInnes, G. *et al.* Global Biobank Engine: enabling genotype-phenotype browsing for biobank summary statistics. *Bioinformatics* **35**, 2495-2497, doi:10.1093/bioinformatics/bty999 (2019).
70 Kamat, M. A. *et al.* PhenoScanner V2: an expanded tool for searching human genotype-phenotype associations. *Bioinformatics* **35**, 4851-4853, doi:10.1093/bioinformatics/btz469 (2019).
71 Boyd A, G. J., Macleod J, Lawlor DA, Fraser A, Henderson J, Molloy L, Ness A, Ring S, Davey Smith G. Cohort Profile: the 'children of the 90s'--the index offspring of the Avon Longitudinal Study of Parents and Children. *International Journal of Epidemiology* **42**, 111-127 (2013).
72 Fraser A, M.-W. C., Tilling K, Boyd A, Golding J, Davey Smith G, Henderson J, Macleod J, Molloy L, Ness A, Ring S, Nelson SM, Lawlor DA. Cohort Profile: the Avon Longitudinal Study of Parents and Children: ALSPAC mothers cohort. *International Journal of Epidemiology* **42**, 97-110 (2012).
73 Howe, L. D. *et al.* Changes in ponderal index and body mass index across childhood and their associations with fat mass and cardiovascular risk factors at age 15. *PLoS One* **5**, e15186-e15186, doi:10.1371/journal.pone.0015186 (2010).
74 Frysz, M., Howe, L. D., Tobias, J. H. & Paternoster, L. Using SITAR (SuperImposition by Translation and Rotation) to estimate age at peak height velocity in Avon Longitudinal Study of Parents and Children. *Wellcome Open Res* **3**, 90, doi:10.12688/wellcomeopenres.14708.2 (2018).
75 Wade, K. H. *et al.* Loss-of-function mutations in the melanocortin 4 receptor in a UK birth cohort. *Nat Med* **27**, 1088-1096, doi:10.1038/s41591-021-01349-y (2021).
76 Li, H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv:1303.3997 (2013). <https://ui.adsabs.harvard.edu/abs/2013arXiv1303.3997L>.
77 Koboldt, D. C. *et al.* VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. *Genome Res* **22**, 568-576, doi:10.1101/gr.129684.111 (2012).
78 Wade, K. H. *et al.* Loss-of-function mutations in the melanocortin 4 receptor in a UK birth cohort. *Nat Med*, doi:10.1038/s41591-021-01349-y (2021).
79 Lindsay, T. *et al.* Descriptive epidemiology of physical activity energy expenditure in UK adults (The Fenland study). *Int J Behav Nutr Phys Act* **16**, 126, doi:10.1186/s12966-019-0882-6 (2019).
80 Williams, S. A. *et al.* Plasma protein patterns as comprehensive indicators of health. *Nat Med* **25**, 1851-1857, doi:10.1038/s41591-019-0665-2 (2019).
81 Pietzner, M. *et al.* Genetic architecture of host proteins involved in SARS-CoV-2 infection. *Nature Communications* **11**, 6397, doi:10.1038/s41467-020-19996-z (2020).
82 Willer, C. J., Li, Y. & Abecasis, G. R. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* **26**, 2190-2191, doi:10.1093/bioinformatics/btq340 (2010).
83 Day, N. *et al.* EPIC-Norfolk: study design and characteristics of the cohort. European Prospective Investigation of Cancer. *Br J Cancer* **80 Suppl 1**, 95-103 (1999).
84 Pietzner, M. *et al.* Plasma metabolites to profile pathways in noncommunicable disease multimorbidity. *Nat Med* **27**, 471-479, doi:10.1038/s41591-021-01266-0 (2021).
85 Lotta, L. A. *et al.* A cross-platform approach identifies genetic regulators of human metabolism and health. *Nat Genet* **53**, 54-64, doi:10.1038/s41588-020-00751-5 (2021).

86 Finer, S. *et al.* Cohort Profile: East London Genes & Health (ELGH), a community-based population genomics and health study in British Bangladeshi and British Pakistani people. *Int J Epidemiol* **49**, 20-21i, doi:10.1093/ije/dyz174 (2020).
87 Narasimhan, V. M. *et al.* Health and population effects of rare gene knockouts in adult humans with related parents. *Science* **352**, 474-477, doi:10.1126/science.aac8624 (2016).
88 Bayraktar, O. A. *et al.* Astrocyte layers in the mammalian cerebral cortex revealed by a single-cell in situ transcriptomic map. *Nat Neurosci* **23**, 500-509, doi:10.1038/s41593-020-0602-1 (2020).
89 Bankhead, P. *et al.* QuPath: Open source software for digital pathology image analysis. *Sci Rep* **7**, 16878, doi:10.1038/s41598-017-17204-5 (2017).
90 Schmidt, U., Weigert, M., Broaddus, C. & Myers, G. 265-273 (Springer International Publishing).
91 Widmann, J. *et al.* RNASTAR: an RNA STructural Alignment Repository that provides insight into the evolution of natural and artificial RNAs. *RNA* **18**, 1319-1327, doi:10.1261/rna.032052.111 (2012).
92 Stuart, T. *et al.* Comprehensive Integration of Single-Cell Data. *Cell* **177**, 1888-1902 e1821, doi:10.1016/j.cell.2019.05.031 (2019).

Acknowledgements

MC3R genetic analysis, next-generation sequencing and functional analysis were supported by UK Medical Research Council (MRC) Metabolic Diseases Unit funding (MC_UU_00014/1), Wellcome [WT 095515/Z/11/Z] and the NIHR Cambridge Biomedical Research Centre. KR, KD, DR, IC, APC and GSHY are supported by the MRC Metabolic Disease Unit (MC_UU_00014/1) and SOR by a Wellcome Investigator award (WT 095515/Z/11/Z) and NIHR Cambridge Biomedical Research Centre. BYHL is supported by a BBSRC Project Grant (BB/S017593/1). AW and SB hold PhD studentships supported by Wellcome. JAT is supported by an NIHR Clinical Lectureship (CL-2019-14-504). AM holds a PhD studentship supported jointly by the University of Cambridge Experimental Medicine Training Initiative (EMI) programme in partnership with AstraZeneca (EMI-AZ). GKCD is supported by the BBSRC Doctoral Training Programme. Next-generation sequencing was performed via Wellcome-MRC IMS Genomics and transcriptomics core facility supported by the MRC [MC_UU_00014/5] and the Wellcome [208363/Z/17/Z] and the Cancer Research UK Cambridge Institute Genomics Core. The histology core is supported by the MRC [MC_UU_00014/5].

We thank Peter Barker and Keith Burling of the Cambridge NIHR BRC Clinical Biochemistry Assay Lab for their assistance with biochemical assays.

The EPIC-Norfolk study (doi: 10.22025/2019.10.105.00004) has received funding from the Medical Research Council (MR/N003284/1 and MC-UU_12015/1) and Cancer Research UK (C864/A14136). The genetics work in the EPIC-Norfolk study was funded by the Medical Research Council (MC_PC_13048). Metabolite measurements in the EPIC-Norfolk study were supported by the MRC Cambridge Initiative in Metabolic Science (MR/L00002/1) and the Innovative Medicines Initiative Joint Undertaking under EMIF grant agreement no. 115372. We are grateful to all the participants who have been part of the project and to the many members of the study teams at the University of Cambridge who have enabled this research.

The Fenland Study (doi: 10.22025/2017.10.101.00001) is funded by the Medical Research Council (MC_UU_12015/1). We are grateful to all the volunteers and to the General Practitioners and practice staff for assistance with recruitment. We thank the Fenland Study Investigators, Fenland Study Co-ordination team and the Epidemiology Field, Data and Laboratory teams. We further acknowledge support for genomics and metabolomics from the Medical Research Council (MC_PC_13046). Proteomic measurements were supported and governed by a collaboration agreement between the University of Cambridge and Somalogic.

FRD, NJW, KKO, CL, and JRBP are funded by the Medical Research Council (MC_UU_12015/1, MC_UU_12015/2, MC_UU_00006/1 and MC_UU_00006/2). NJW is an NIHR Senior Investigator.

We are grateful for funding to the BIA prediction equations, supported by the NIHR Biomedical Research Centre Cambridge [IS-BRC-1215-20014]. The NIHR Cambridge Biomedical Research Centre (BRC) is a partnership between Cambridge University Hospitals NHS Foundation Trust and the University of Cambridge, funded by the National Institute for Health Research (NIHR). The views expressed are those of the authors and not necessarily those of the NHS, the NIHR or the Department of Health and Social Care.

We would like to thank A. Mesut Erzurumluoglu, Laura Wittemans, Eleanor Wheeler, Isobel Stewart, Maik Pietzner, Mine Koprulu, Emanuella De Lucia Rolfe, Richard Powell and Nicola Kerrison for providing help with and access to GWAS meta-analysis summary statistics for body composition measures and biomarkers in UKBB, metabolomics measures in the EPIC-Norfolk Study, proteomics measures in the MRC Fenland Study, as well as help with genotype quality control in Fenland and UKBB.

The Medical Research Council UK, Wellcome (217065/Z/19/Z) and the University of Bristol provide core support for ALSPAC. Genome-wide association data was generated by Sample Logistics and Genotyping Facilities at Wellcome Sanger Institute and LabCorp (Laboratory Corporation of America) using support from 23andMe. AGS was supported by the study of Dynamic longitudinal exposome trajectories in cardiovascular and metabolic non-communicable diseases (H2020-SC1-2019-Single-Stage-RTD, project ID 874739). KHW was supported by the Elizabeth Blackwell Institute for Health Research, University of Bristol and the Wellcome Institutional Strategic Support Fund [204813/Z/16/Z]. NJT is a Wellcome Trust Investigator (202802/Z/16/Z), is the PI of the Avon Longitudinal Study of Parents and Children (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 CRUK Integrative Cancer Epidemiology Programme (C18281/A19169).

We are extremely grateful to all 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.

The Rowitch Lab receives funding from Wellcome and the ERC Advanced Grant (REP-789054-1).

Genes & Health is/has recently been core-funded by Wellcome (WT102627, WT210561), the Medical Research Council (UK) (M009017), Higher Education Funding Council for England Catalyst, Barts Charity (845/1796), Health Data Research UK (for London substantive site), and research delivery support from the NHS National Institute for Health Research Clinical Research Network (North Thames). Additional funding for recall was provided by a pump priming award to SF (SCA/PP/12/19) from the Diabetes Research and Wellness Foundation. EGB and XD are supported by the Wellcome (208987/Z/17/Z) and Barts Charity (project grant to EGB). We thank Social Action for Health, Centre of The Cell, members of our Community Advisory Group, and staff who have recruited and collected data from volunteers. We thank the NIHR National Biosample Centre (UK Biocentre), the Social Genetic & Developmental Psychiatry Centre (King's College London), Wellcome Sanger Institute, and Broad Institute for sample processing, genotyping, sequencing and variant annotation. We thank Barts Health NHS Trust, NHS Clinical Commissioning Groups (Hackney, Waltham Forest, Tower Hamlets, Newham), East London NHS Foundation Trust, Bradford Teaching Hospitals NHS Foundation Trust, and Public Health England (especially David Wyllie) for GDPR-compliant data sharing. Most of all we thank all of the volunteers participating in Genes & Health.

RDC receives funding from NIH grants DK070332 and DK126715. PS is funded by F32HD095620 and K99DK127065. RBS receives funding from DK106476 and MNB is funded by F32DK123879.

Authors contributions

BYHL, AW, SF, AM, KW, NJT, KKO, CL, JRBP, GSHY and SOR designed the study. BYHL, AW, KD, AM, SB and JRS planned and performed the in-vitro experiments. BYHL and AW conducted the bioinformatic and genetic analyses in UKBB and G&H. AW, FRD, NJW, KKO, JRBP and CL conducted the genotype-phenotype association in UKBB, Fenland and EPIC. KR and KD conducted the NGS for ALSPAC and Sanger sequencing for ALSPAC and G&H. BYHL and AM conducted genetics and bioinformatic analyses of ALSPAC. AGS and KHW and NJT lead the analysis of phenotypic association in ALSPAC. PS, DTP, KLJE, RNL and RDC performed the study on *Mc3r* null animals. BYHL performed the single-cell data analysis. IC, DR and APC lead the animal studies in Cambridge. JAT, GKCD, KER, SH, ZX, DHR, MNB and RBS conducted the histology, smFISH and imaging analysis. SF, AK, RCT, HCM, DAVH and the G&H team managed the cohort, DAVH, HCM, EGB and XD led the genetic analysis, and SF coordinated and conducted the clinical recall. BYHL, AW, SF, FRD, AGS, KW, NJT, KKO, CL, JRBP, GSHY and SOR wrote the manuscript and it was reviewed by all authors. This publication is the work of the authors and CL, JRBP, GSHY and SOR will serve as guarantors for the contents of this paper.

Competing Interests Statement

SOR has undertaken remunerated consultancy work for Pfizer, AstraZeneca, GSK, and ERX Pharmaceuticals.

DAVH has an unrestricted research grant from Alnylam Pharmaceuticals.

PS and RDC hold equity in Courage Therapeutics Inc. and are inventors of intellectual property optioned to Courage Therapeutics Inc. RDC chairs the Scientific Advisory Board at Courage Therapeutics Inc.

All remaining authors declare no competing interests.

Additional Information

Supplementary Information is available for this paper.

Correspondence and requests for materials should be addressed to Stephen O'Rahilly (so104@medschl.cam.ac.uk). *Nature* expects this identified author to respond to readers' enquiries and requests for materials, and to coordinate the handling of any other matters arising from the published contribution, including corrections complaints. The author named as corresponding author is not necessarily the senior author, and publication of this author's name does not imply seniority. Authors may include more than one e-mail address if essential, in which event *Nature* will communicate with the first-listed address for any post-publication matters arising, and expect that author to coordinate with the other co-authors.

Extended Data Figure Legends

Extended Data Figure 1 Functionally characterised MC3R mutations.

Complete, partial loss-of-function (LoF) and wild-type like mutations are marked in purple, dark yellow and green respectively. Coloured rectangles indicate cohort(s) in which mutations were identified: Red = UK Biobank (UKBB); Blue = Avon Longitudinal Study of Parents & Children (ALSPAC); Light Brown = Genes & Health (G&H).

Extended Data Figure 2 PheWAS analysis of *MC3R* genetic risk score

A CADD-weighted MC3R genetic risk score was constructed (see methods) and used to conduct a phenome-wide analysis (pheWAS) with publicly available summary statistics. Solid black line indicates Bonferroni multiple-testing threshold of $p < 1.046e\text{-}4$, dashed line indicates nominal significance threshold $p < 0.05$.

Extended Data Figure 3 Effect of *MC3R* Loss-of-Function mutations on height (cm) across time

Carriers of MC3R LoF mutations (dark blue) had lower height throughout early life course compared to the reference group (light blue) after adjusting for sex and age. Figures only show results where the mutation group was represented by at least one individual at all time points between birth and 24 years. Mean ± 95% CI shown, N and p-values are listed in Table S4.

Extended Data Figure 4 MC3R is essential for normal cycle length and for fasting-induced suppression of the reproductive axis.

(a & b) Representative traces of progression through the oestrous cycle in WT (a) and *Mc3r*$^{-/-}$ (b) mice following an overnight fast. D=Dioestrous; M=Metoestrous; E=Oestrous.

Extended Figure 5 *Mc3r* is expressed in several cell populations in the mouse hypothalamus

(a) T-SNE plot showing 28 neuronal clusters of the mouse hypothalamus from a combined dataset consisting of 18,427 neurons from 4 published studies.

(b) *Mc3r* is expressed in several neuronal populations (log_2 normalised expression in dark red).

(c) Multiplexed smFISH showing the co-expression of *Mc3r* (white) *Kiss1* (red) and *Tac2* (green) in the arcuate nucleus. (Representative example shown, n=3 mice, scale bar=20μm)

(d) Venn diagram showing the number of cells expressing *Kiss1* (left, red), *Tac2* (right, green), or both (KND*y*, centre).

(e) Violin plots showing the number of Mc3r mRNA punta in *Kiss1* only, KND*y*, and *Tac2* only cells. Mean percentages of cells ± SEM with detected *Mc3r* are shown, data collected from 3 mice.

Extended Data Figure 6 Expression of *Mc3r* and *Lepr* in KNDy and GHRH neurons

(a - b) *Mc3r* expression is more prominent compared to *Mc4r* and *Lepr* in (a) *Tac2* (KNDy) (cluster 7, blue) and (b) GHRH neurons (cluster 15, green).

(c) Violin plots showing expression of *Kiss1, Tac2, Ghrh, Mc3r* and *Lepr* in KNDy and Ghrh neurons in the Campbell[38] and the Chen[42] dataset separately. The Lam[40] and Romanov[41] datasets are not shown due to low cell count (<10).

Extended Data Figure 7 Human smFISH showing the co-expression of *MC3R, KISS1,* and *GHRH* in the human hypothalamic arcuate nucleus

(a) Annotated overview *MC3R and KISS1* co-expression: *MC3R* = grey, *KISS* = magenta and *MC3R+KISS1* = white (scale bar=200μm). High-powered micrograph (squared area) below shows the staining of *MC3R* (white) and *Kiss1* (magenta) mRNA punta in 2 representative cells (teal=DAPI, scale bar=10μm). N=2 slides

(b) Annotated overview of *MC3R and GHRH* co-expression: *MC3R* = grey, *GHRH* = green and *MC3R+KISS1* = white (scale bar=200μm). High-powered micrograph (squared area) below shows the staining of *MC3R* and *GHRH* mRNA punta in a representative cell (teal=DAPI, scale bar=4μm). N=2 slides

Extended Data Figure 8 *Mc3r* expression in kisspeptin neurons in the mouse hypothalamus at P16, P28 and P48

(a - c) Representative smFISH showing the co-expression of *Mc3r and Kiss1* in the anteroventral periventricular nucleus (AVPV) at (a) P16; (b) P28 and (c) P48 (N=3 mice for all age groups): *Mc3r* = green, *Kiss1* = red (scale bar=20μm).