Original research

Heritable genetic variants in key cancer genes link cancer risk with anthropometric traits

Matteo Di Giovannantonio,[1,2] Benjamin HL Harris ![ORCID],[1] Ping Zhang,[2] Isaac Kitchen-Smith,[2] Lingyun Xiong,[2] Natasha Sahgal,[2] Giovanni Stracquadanio,[2,3] Marsha Wallace,[2] Sarah Blagden,[4] Simon Lord,[5] David Harris,[6] Adrian H.L. Harris,[7] Francesca M. Buffa ![ORCID],[1] Gareth L. Bond[2,8]

Correspondence to
Prof. Francesca M. Buffa, Computational Biology & Integrative Genomics Lab, Department of Oncology, University of Oxford, Oxford, OX3 7DQ, UK; francesca.buffa@oncology. ox.ac.uk and Dr. Gareth L. Bond, Institute of Cancer & Genomics Sciences, University of Birmingham, Bimingham, B15 2TT, UK; G.Bond.1@bham.ac.uk

MDG and BHH are joint first authors.
FMB and GLB are joint senior authors.

ABSTRACT

Background Height and other anthropometric measures are consistently found to associate with differential cancer risk. However, both genetic and mechanistic insights into these epidemiological associations are notably lacking. Conversely, inherited genetic variants in tumour suppressors and oncogenes increase cancer risk, but little is known about their influence on anthropometric traits.

Methods By integrating inherited and somatic cancer genetic data from the Genome-Wide Association Study Catalog, expression Quantitative Trait Loci databases and the Cancer Gene Census, we identify SNPs that associate with different cancer types and differential gene expression in at least one tissue type, and explore the potential pleiotropic associations of these SNPs with anthropometric traits through SNP-wise association in a cohort of 500,000 individuals.

Results We identify three regulatory SNPs for three important cancer genes, FANCA, MAP3K1 and TP53 that associate with both anthropometric traits and cancer risk. Of particular interest, we identify a previously unrecognised strong association between the rs78378222[C] SNP in the 3' untranslated region (3'-UTR) of TP53 and both increased risk for developing non-melanomatous skin cancer (OR=1.36 (95% 1.31 to 1.41), adjusted p=7.62E^{-63}), brain malignancy (OR=3.12 (2.22 to 4.37), adjusted p=1.43E^{-12}) and increased standing height (adjusted p=2.18E^{-24}, beta=0.073±0.007), lean body mass (adjusted p=8.34E^{-37}, beta=0.073±0.005) and basal metabolic rate (adjusted p=1.13E^{-31}, beta=0.076±0.006), thus offering a novel genetic link between these anthropometric traits and cancer risk.

Conclusion Our results clearly demonstrate that heritable variants in key cancer genes can associate with both differential cancer risk and anthropometric traits in the general population, thereby lending support for a genetic basis for linking these human phenotypes.

INTRODUCTION

Height and other anthropometric measures have been consistently found to associate with differential cancer risk.[1–3] However, mechanistic and genetic insights into these associations are notably lacking.[2 4] Increased height has been associated with a number of malignancies including skin, breast, colon, rectum, endometrium, ovary and kidney cancers, as well as Hodgkin's lymphoma and leukaemia.[2 5] Current theories linking increased height with increased cancer risk centre on the role of growth hormone, insulin-like growth factor and/or insulin as pro-tumourigenic by regulating the Ras-MAPK and PI3K pathways, and hence leading to increased cellular proliferation.[6] It has also been proposed that increased height corresponds to an increased number of cells, resulting in an increased probability of malignancy simply by chance[7]; however, such allometric scaling of body mass explains only part of the observed effect and it is not observed among mammal species of different size (Peto's paradox).[8] Increased body mass index (BMI), a surrogate measure of obesity based on the weight-to-height ratio, has also been associated with an increased risk of various cancers, including postmenopausal breast cancer, colorectal cancer and renal cancer.[9] Increased BMI is partially attributed to increasing accumulation of fat mass. It is hypothesised that adiposity causes a state of systemic inflammation, a shift of metabolite and adipokine release and an increase of circulating insulin, secondary to insulin resistance.[10] This combination may increase cancer risk. However, the mechanism(s) behind this association are not well understood. Thus, links between anthropometric measures and cancer remain an area of intense research.

Anthropometric traits are largely determined by genes that control cellular proliferation, metabolism and apoptosis, processes that are also required for immortalisation of cancer cells and development of tumours.[11] It is well established that germline mutations within tumour suppressor genes and oncogenes affect cancer risk.[12] Studies on mouse models and rare diseases indicate that mutations in some of these genes may influence body mass composition.[13–15] For example, the tumour suppressor gene ARH1 inhibits cell growth. A deletion in ARH1 is commonly associated with breast and ovarian carcinoma, while mice that are engineered to overexpress ARH1 are significantly smaller than wild-type counterparts.[16] Mice possessing knockouts of genes within the tumourigenic hypoxia-inducible pathway vary in size and respond to high-fat diets differently from those without the mutations.[17 18] TP53 is the most frequently mutated gene in human cancers and is a key regulator of a number of cellular activities

that prevent tumourigenesis, including maintaining genomic stability, controlling cell growth and metabolism.[13][19] Mouse models of TP53 mutations have demonstrated that reduction of p53 activity can increase cancer risk, and also alter metabolism and occurrence of obesity in a complex and signal dependent manner.[14][15]

Further evidence for a relationship between body size, genetics and cancer comes from human genetic diseases, such as Turner syndrome. Turner syndrome is defined by complete or partial chromosome X monosomy and is associated with a distinct clinical phenotype, including gonadal dysgenesis, cubitus valgus and short stature.[20] Short stature in this condition has been attributed to haplo-insufficiency of the SHOX tumour suppressor gene.[21] Deletions and mutations in SHOX have also been proposed to explain short stature within the general population.[22] Another example comes from patients with Simson-Golabi-Behmel syndrome (SGB), caused by mutations in the GPC3 tumour suppressor gene. Patients with SGB tend to have prenatal and postnatal overgrowth and thus ultimately taller stature together with an increased risk of embryonic tumours. GPC3-null mice also display overgrowth.[23] These effects, in both humans and mice, are thought to be due to unshackling of the hedgehog pathways, and increasing cellular proliferation.[24][25]

Pleiotropy is the phenomenon whereby a single SNP or genetic mutation can influence multiple traits. These rare diseases provide evidence that major differences in the expression of cancer genes can influence both anthropometric traits and cancer risk in specific cohorts of patients. However, it is less clear that such associations occur in the broader population. The availability of large phenotypic and genetically linked datasets has allowed us to deploy genome-wide association studies (GWAS) of common SNPs to understand the link between genetics and anthropometric traits. However, until now, it has been difficult to look for pleiotropy between functional SNPs in genes associated with cancer risk and anthropometric traits, as very few cohorts possess both comprehensive genetic data and detailed anthropometric data in the same population. The UK Biobank provides a unique opportunity to investigate such pleiotropic associations in a large prospective cohort of over 500 000 participants.

METHODS
The UK Biobank
The UK Biobank is a cohort of ~500 000 UK residents who volunteered to have their clinical, lifestyle, anthropometric and genetic data collected for research. Data were collected in a number of ways, including by face-to-face interview, touchscreen assessment and from centralised clinical registers (eg, the Cancer register and Death register). Participants were aged between 40 and 69 years at recruitment. All baseline data used in this study were collected in 22 UK centres, between 2006 and 2010.

Anthropometric data
During the baseline assessment, participants had various anthropometric traits measured directly or by bioelectrical impedance. Bioimpedance data were measured using a Tanita BC418MA body composition analyser. Participants stood barefoot on the analyser and held the metal handles. This device produced measurements of fat mass, fat-free mass and basal metabolic rate. Further information on anthropometric data collected is found here.[26]

Cancer data
Cancer occurrences were defined by presence of a cancer International Classification of Diseases (ICD) code in the UK Cancer register or the UK Death register. To maximise the number of individual cancers cases, we combined ICD9 and ICD10 codes of identical cancers. This was carried out by a clinician to ensure the matching was accurate: for example, malignant neoplasm of brain was defined as C71 (ICD10) and 191 (ICD9). To increase statistical power, when appropriate we merged ICD codes into clinically relevant groups. We ran both individual cancers (combined ICD9 and ICD10 codes) and clinically relevant groups in our analysis.

Genetic data
Blood samples were collected when participants were recruited, and DNA extracted.[27] DNA was then genotyped on either the Affymetrix UK BiLEVE Axiom array or the Affymetrix UK Biobank Axiom array (Santa Clara, California, USA). Imputation was based on a merged reference panel of ~90 million biallelic variants, from the 1000 Genomes Phase III[28] and the UK10K[29] haplotype panels. Imputation was performed using IMPUTE2 as described,[30] producing 488 295 genotyped participants.

Sample quality control
In addition to the standard quality control, we used further quality control steps to ensure robustness of our analyses. We excluded individuals based on (i) a mismatched value between self-reported and genetic sex (data-field: 22 001 and 31); (ii) level of genotype missingness of >0.05 (data-field: 22005); (iii) genetic relatedness factor with kinship coefficient of >0.0442; (iv) sex chromosome aneuploidy (data-field: 22019) and (v) outliers for heterozygosity or missing rate (data-field: 22027). We selected the European population based on self-reported ethnicity (data-field: 21000) by excluding individuals with non-white ethnic backgrounds. This left a study population of 379 358 suitable genotyped individuals.

Identification of functional cancer gene SNPs
We identified genes confirmed to be involved in carcinogenesis from the COSMIC Cancer Gene Census (release V.88, 19 March 2019), the current reference record of genes containing cancer driver mutations. We selected all SNPs annotated to these genes that have been significantly associated with differential cancer risk in the GWAS Catalog.[31] Significant associations with cancer risk were defined as p values of <5E[-08] cut-off. Finally, we specified that these SNPs were also associated with differential expression of the cancer gene in at least one expression quantitative trait locus (eQTL) database. cis-eQTL databases used were GTEX, NESDA/NTR and PancanQTL.[32–34]

SNP quality control
SNPs were excluded if Hardy-Weinberg equilibrium had p values $<1E^{-10}$, there was a minor allele frequency <0.0001, level of missingness >0.05 or an imputation score <0.8 (as per http://www.nealelab.is/uk-biobank/).

Analysis
To carry out the SNP-wise analysis on all functional-CG SNPs, we used SNPTEST (V.2.5.4),[35] and employed the frequentist approach under dominant, additive and recessive inheritance models, using sex, age and genetic principal components (1–20 PCs) as covariates. Genetic PCs were included as covariates to control for hidden population structure. We controlled for

genotype uncertainty by implementing missing data likelihood score test. Within SNPTEST, we used the frequentist approach under an additive inheritance model. P values of SNP-wise association were adjusted by the stringent Bonferroni correction based on number of tested traits (28 anthropometric traits, 50 cancer types) multiplied by the number of eSNPs (100). Significant associations were defined as Bonferroni correction p values below the threshold of $1E^{-5}$.

Linkage and lead SNPs

R^2 and D' coefficients were calculated to evaluate the degree of linkage disequilibrium (LD) between different loci using LDlink V.3.7 (D'>0.8, R^2 >0.4 https://ldlink.nci.nih.gov/).[36] Lead SNPs were defined as the SNPs that were most strongly associated with the traits in question for each locus.[37] Due to the low value of R^2 in some of the potential LD eSNPs, we carried out a leave-one-out analysis, as described here.

► For each lead SNP:
  – All participants from the cohort carrying the lead SNP were removed.
  – The association for each of the potential LD eSNPs was performed (eg, lead SNP rs78378222 carriers were

removed while rs35850753 and rs8753 were tested for their associations with cancer and anthropometric traits).
  – Adjusted p values from the whole cohort association analysis with those obtained in this analysis were compared.
  – If the association of the potential LD eSNPs was not significant after the removal of lead SNP carriers (p>0.05), we considered the eSNP in LD.

RESULTS

Through integration and curation of the GWAS Catalog, eQTL databases and the Cancer Gene Census, we identified 100 SNPs which have been shown to associate with a differential risk of developing a total of 21 different cancer types, and show differential gene expression in at least one tissue type of (i) proto-oncogenes (8 genes), (ii) oncogenic fusion proteins (8 genes), (iii) tumour suppressors (16 genes) and (iv) 15 genes that span two or more of these groups (online supplementary tables 1 and 2). We define these SNPs as *cancer eSNPs*. In our analyses of the UK Biobank cohort, we found 13 cancer eSNPs associated with differences in anthropometric traits and 31 with differential cancer risk. Interestingly, seven of these SNPs overlapped, thus displaying some level of pleiotropy between cancer risk and
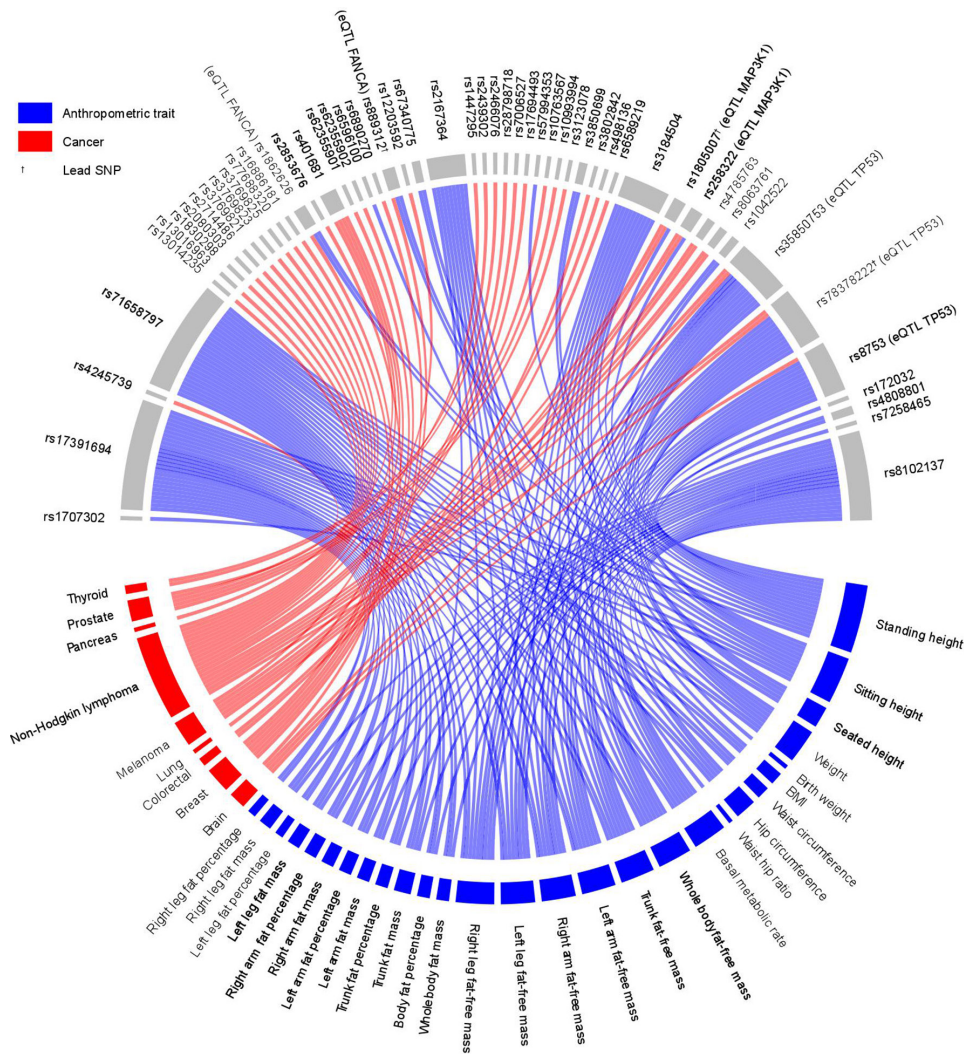
Figure 1 A circos plot indicating the functional SNPs that significantly associate with anthropometric traits and cancer risk. SNPs are located on the top half of the plot while cancer types and anthropometric traits are on the bottom half. Significant associations (Bonferroni corrected p values <1E⁻⁵) are shown as solid lines. Blue lines indicate significant associations with anthropometric traits. Red lines depict significant associations with differential cancer risk.
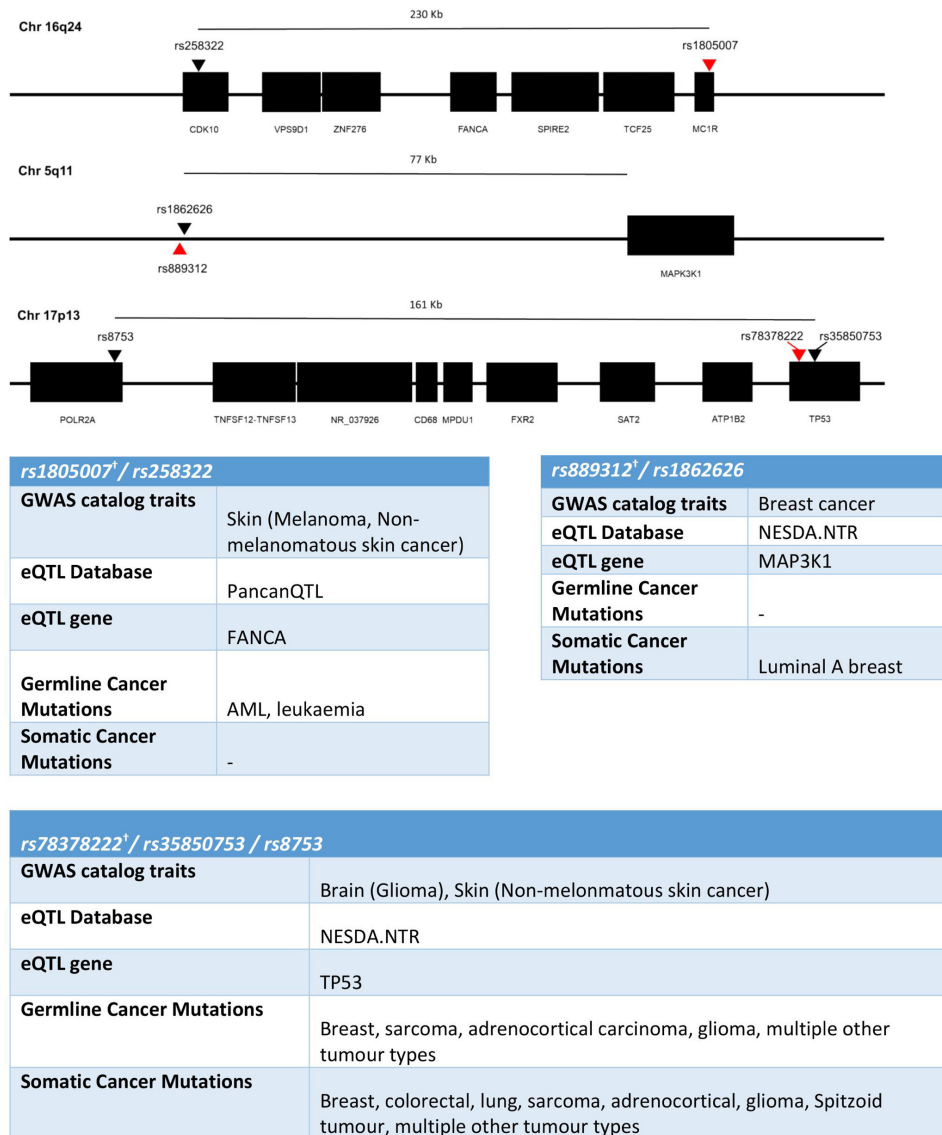
Figure 2 A gene map of the chromosomal regions containing the pleiotropic functional SNPs. Lead SNPs are highlighted with red arrowhead. Tables summarise salient genome-wide association study (GWAS) Catalog, expression quantitative trait loci (eQTL) and Caner Gene Census information for each variant.

anthropometric traits (figure 1, online supplementary tables 3 and 4). These seven cancer eSNPs are found on three different chromosomes and in LD (see 'Methods' section) with at least one other identified cancer eSNP (figure 2). They have been found to be associated with differential expression levels of (i) Fanconi anaemia, complementation group A (FANCA) (rs1805007, rs258322), (ii) mitogen-activated protein kinase kinase kinase 1 (MAP3K1: rs889312, rs1862626) and (iii) tumour suppressor p53, TP53 (rs78378222, rs35850753, rs8753). The two cancer eSNPs associated with FANCA expression reside in neighbouring genes; rs258322 is a non-coding variant within CDK10 and rs1805007 is a missense variant within MC1R. eSNPs associated with MAP3K1 expression are found in an intergenic region close to ZNF296 and MAP3K1 (5q11.2). In contrast, the two cancer eSNPs associated with differential TP53 expression are found in untranslated regions (UTRs) of the TP53 gene itself. rs78378222 is found in the 3'-UTR of TP53, and rs35850753 is found in the 5'-UTR of the d133 isoform of TP53. The third TP53 eSNP lies in the 3'-UTR of the neighbouring POLR2A gene. The cancer eSNPs showing the strongest association between

anthropometric measures and cancer risk within their loci (lead cancer eSNP) were rs1805007 (FANCA, C>T, minor allele frequency in UK Biobank[40]=0.102), rs78378222 (TP53, A>C, MAF=0.012) and rs889312 (MAP3K1, C>A, MAF=0.284).

In the UK Biobank cohort, the minor allele of rs1805007 (FANCA) was associated with an increased risk of melanoma (OR=1.63 (95% CI 1.52 to 1.75), adjusted $p=2.57E^{-41}$) and non-melanomatous skin cancer (OR=1.36 (1.31 to 1.41), adjusted $p=7.62E^{-63}$). The minor allele of rs78378222 (TP53) was associated with an increased risk of brain malignancy (OR=3.12 (2.22 to 4.37), adjusted $p=1.43E^{-12}$) and non-melanomatous skin cancer (OR=1.46 (1.34 to 1.60), adjusted $p=5.20E^{-18}$). Post hoc analysis of the histology in the carriers of rs78378222 revealed brain cancers tended to be gliomas (81%) and notably, 16% had no histology recorded. The minor allele of rs889312 (MAP3K1) was associated with an increased risk of breast cancer (OR=1.1 (1.07 to 1.13), adjusted $p=2.82E^{-11}$). Of note, as we selected these SNPs due to their noted association with differential cancer susceptibility in GWAS studies, our results provide an independent
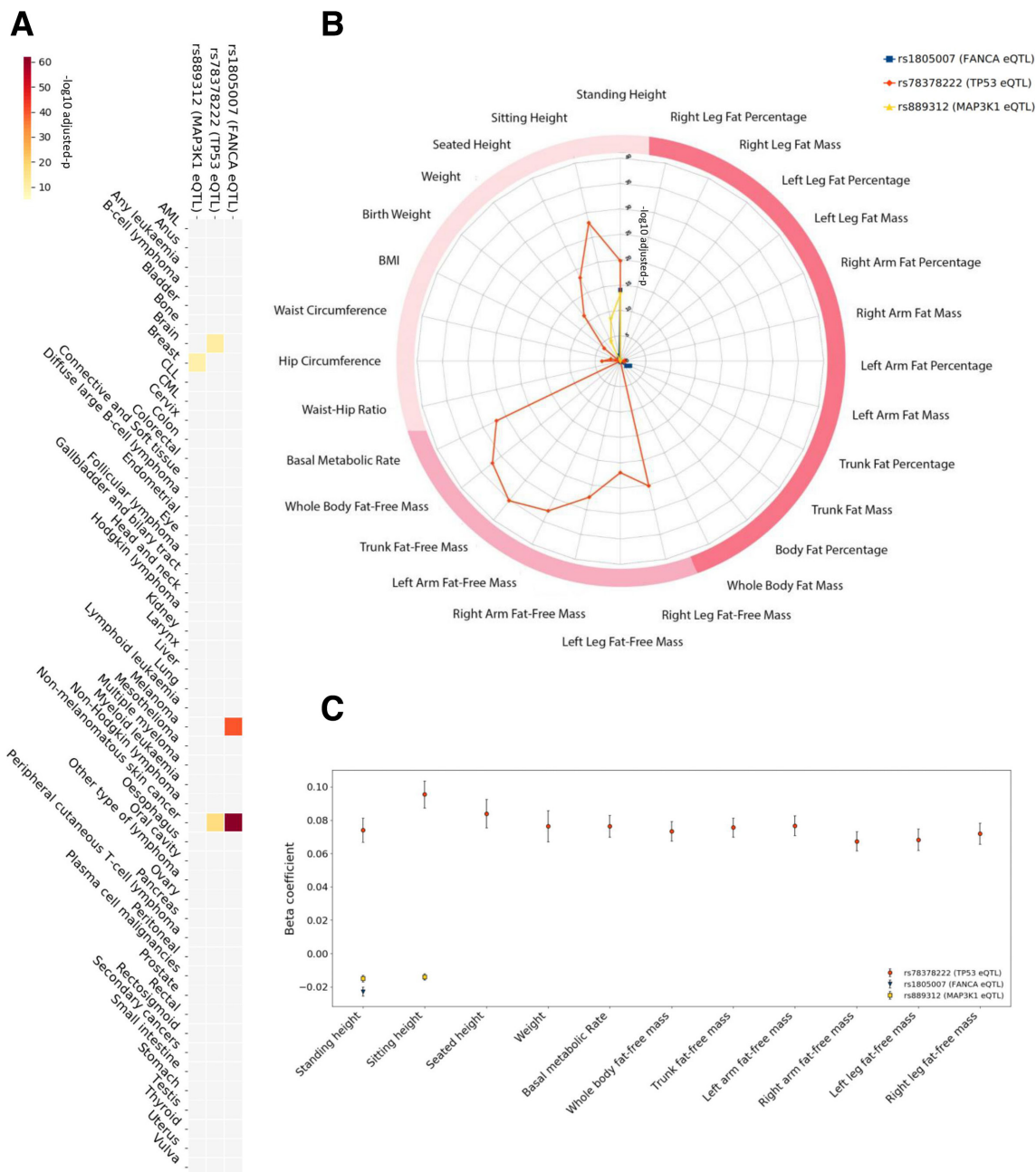
Figure 3 Detailed association results for cancer eSNPs with anthropometric traits and cancer risk in the UK Biobank cohort. A heatmap (A) depicting the significance of the associations of the three lead pleiotropic SNPs with differential cancer risk. The colour scale represents the −log10 adjusted p: with the darkest red on the scale being $7.62E^{-63}$ and the lightest yellow being $1E^{-5}$. A radar plot (B) illustrating the level of significance of the associations of the three lead pleiotropic SNPs with anthropometric traits (radial axis: −log10 adjusted p). The darkest two pink categories are bioelectrical impedance measures (darkest being measures of fat). Traditional bedside anthropometric measures are in blush. SNPs related to FANCA (rs1805007†) and MAP3K1 (rs889312†) associate with standing height. The lead functional SNP related to TP53, rs78378222†, markedly associates with multiple measures of height and lean body mass, as well as basal metabolic rate. An error bar plot (C) of the beta coefficients (Y-axis) of the significant associations with anthropometric traits (X-axis). Error bars denote the 95% CI of the beta value. †Lead cancer eSNP.

validation of these associations (figure 3A, online supplementary table 5).

Considering the anthropometric measures, we identified an unexpectedly high level of pleiotropy, and very strong association, between the two TP53 cancer eSNPs and three measures of height (standing height, sitting height and seated height), overall body weight, six of lean body mass (right arm fat-free mass, left arm fat-free mass, right leg fat-free mass, left leg fat-free mass, trunk fat-free mass and whole body fat-free mass), as well as basal metabolic rate, with p values as low as $7.54E^{-36}$ (figure 3B).

In support of our findings, significant associations of two TP53 cancer eSNPs (rs78378222, rs35850753) with the lean body mass measures (whole body fat-free mass and appendicular fat-free mass) using similar methodologies, yet with different covariates and different subpopulations of the UK Biobank cohort, have been presented in studies of different diseases, for instance, anorexia nervosa, providing further evidence for our results.[38 39] After quality control (QC) procedures, we identified 59 participants homozygous for the minor allele (increased cancer risk) of the lead TP53 eSNP (rs78378222), 9253 heterozygous for the

minor allele and 370 046 homozygous for the major allele. The minor allele carriers were on average taller and leaner, and had higher basal metabolic rate. The associations of these anthropometric traits with the minor allele of the rs78378222 SNP were markedly more significant (almost threefold greater) and with greater effect sizes (up to almost fivefold greater) than for the other pleiotropic eSNPs (figure 3C, online supplementary table 3).

The cancer eSNPs for TP53 showed strikingly strong associations with 10 different anthropometric traits in the UK Biobank cohort. In contrast, the cancer eSNPs related to FANCA (rs1805007) and MAP3K1 (rs889312) were associated with standing height but not lean body mass measures. After QC, we identified 4271 participants homozygous for the minor allele of the lead FANCA eSNP, 68 956 heterozygous for the minor allele and 306 131 homozygous for the major allele. The minor allele for the lead MAP3K1 eSNP was found in 185 184 participants (30 394 homozygous, 154 790 heterozygous). The FANCA eSNP (rs1805007) associated with standing height only (beta=−0.02±0.002, adjusted p=9.20E^{-15}), while the MAP3K1 eSNP (rs889312) associated with broader range of height measures (standing height beta=−0.02±0.004, adjusted p=6.11E^{-14} and sitting height beta=−0.01±0.001, adjusted p=3.34E^{-9}). In order to further validate the significance and magnitude of these associations, we applied a different algorithm, PLINK,[40] to the lead cancer eSNPs. As in the initial analysis, we included the covariates sex, age and genetic principal components. Importantly, the PLINK analysis confirmed the observed associations and yielded similar adjusted p and beta values for the lead cancer eSNPs (online supplementary tables 6–8). In addition, we sought validation for these associations with standing height in the independent GIANT consortium dataset. Here, we examined all lead and linked SNPs for each pleiotropic eSNP in the results of the meta-analysis conducted by Wood et al.[41] We were able to identify eSNPs for MAP3K1 (rs889312, rs1862626) and FANCA (rs1805007, rs258322), but none for TP53, presumably due to the low minor allele frequency. Three of these eSNPs significantly associated with height in the meta-analysis by Wood et al: (i) rs258322 FANCA eQTL (p=1.5E^{-09}, beta=−0.029), (ii) rs889312 MAP3K1 eQTL (p=2.3E^{-08}, beta=−0.018) and (iii) rs1862626 MAP3K1 eQTL (p=3.4E^{-08}, beta=−0.018). Interestingly, both linked MAP3K1 eSNPs and a FANCA eSNP passed the significance threshold and, reassuringly, the directions of the allelic associations are consistent with our findings. The other FANCA eSNP fell just short of the GWAS significance threshold (rs1805007, 4.3E^{-04}, beta=−0.024). Together, all these data clearly link eSNPs for MAP3K1 and FANCA with height.

DISCUSSION

This is the first comprehensive study providing evidence that functional common genetic variants in oncogenes and tumour suppressor genes can associate with both anthropometric traits and cancer risk in the general European population, and in the same cohort. SNPs exhibiting these pleiotropic associations in our study are found in three different loci: (i) FANCA (two SNPs in LD), (ii) MAP3K1 (two SNPs in LD) and (iii) TP53 (three SNPs in LD). Previous work has noted, rs10061757, a SNP in moderate LD (R^2=0.6, D'=1) with the eSNPs for MAP3K1 (rs889312, rs1862626) has previously been shown to associate with height in the UK Biobank cohort.[42] Furthermore, observations from mouse models designed to alter signalling pathways involving *FANCA* and *TP53*, suggested that such associations

with anthropometric traits in humans might be possible. For example, targeted disruption of exons of the FANCA gene in mice results in altered anthropometric traits, including growth retardation, as well as increased cancer risk.[43 44] Furthermore, mouse models of TP53 mutations have clearly demonstrated that reduction of p53 activity can result in increased cancer risk, altered metabolism and influence obesity in a complex and signal-dependent manner.[14 15]

The strongest associations we observed with both anthropometric traits and cancer risk are loci related to TP53. p53 is a key regulator of a number of cellular activities which prevent tumourigenesis, including maintaining genomic stability, controlling cell growth and metabolism.[19] TP53 is the most frequently mutated gene in human cancers.[13] Moreover, in all families with similar TP53 mutations in their heritable genomes, a dramatic increase in cancer risk is observed (Li-Fraumeni syndrome (LFS)).[12] In recent studies, it has been shown that patients with LFS have an increased risk of developing cancer and an increased capacity for oxidative phosphorylation,[45] providing a potential link with anthropometric traits and basal metabolic rate. Furthermore, the well-tolerated antidiabetic drug metformin, which is thought to inhibit mitochondrial complex 1, increases cancer-free survival in a mouse model of LFS and reduces proliferation in cancer cell lines.[46 47] Metformin is now being trialled in patients with LFS to hopefully provide a preventative option for these high-risk patients (ClinicalTrials.gov number: NCT01981525). Based on the potential link with oxidative phosphorylation, this intervention might also be trialled in those carrying the minor allele of the TP53 cancer eSNP (rs78378222).

The TP53 mutations found in LFS are rare in the general population. However, here we show that relatively frequent SNPs related to p53 also affect cancer risk and anthropometric traits in the same cohort. Notably, the minor allele of an SNP in the polyadenylation signal of p53 (rs78378222[C]) which is found in approximately 1% of populations of European descent, has been shown to impair 3'-end processing of p53 mRNA, resulting in a reduction of p53 protein and an increased risk for glioma and basal cell carcinoma as well as affecting head circumference and intracranial volume.[48 49] Here, we validate these cancer associations in a separate cohort (non-melanomatous skin cancer, OR=1.46 (95% CI 1.34 to 1.60), adjusted p=5.20E^{-18}, brain malignancy, OR=3.12 (2.23 to 4.37), adjusted p=1.43E^{-12}), and also show that carriers of this allele tend to be taller, leaner and have a higher basal metabolic rate (standing height, adjusted p=2.18E^{-24}, beta=0.073±0.007, whole body fat-free mass, adjusted p=8.34E^{-37}, beta=0.073±0.005, basal metabolic rate, adjusted p=1.13E^{-31}, beta=0.076±0.006). Of these, the association between rs78378222 and lean body mass confirms previous results that this SNP has the second-largest positive effect size on whole body fat-free mass and fifth largest positive effect size on appendicular fat-free mass in a population aged 60–74 years.[38 39] Indeed in our analysis of all the cancer eSNPs, the rs78378222 minor allele has the largest effect size on fat-free mass measures and also has the largest effect size on standing, seated and sitting height as well as basal metabolic rate (online supplementary table 3). The next two highest ranking SNPs in fat-free mass, height measures and basal metabolic rate are the other two TP53 eSNPs (rs35850753, rs8753).

Prior to this study, associations between increased height and non-melanomatous skin cancer/brain malignancy had been established, but mechanistic and genetic explanations were lacking.[2 50] Our study proves the concept that functional loci in well-characterised tumour suppressors and oncogenes alter both cancer risk and anthropometric traits. Of particular interest, we

identify a strong, new, association between the rs78378222[C] SNP in the polyadenylation site of p53 with both increased risk for developing non-melanomatous skin cancer/brain malignancy and increased height, body weight, lean body mass and basal metabolic rate. This offers a novel genetic link between anthropometric traits and cancer risk.

Author affiliations
[1]Computational Biology & Integrative Genomics Lab, Department of Oncology, Medical Science Division, University of Oxford, Oxford, UK
[2]Ludwig Cancer Institute, Medical Sciences Division, University of Oxford, Oxford, UK
[3]Institute of Quantitative Biology, Biochemistry and Biotechnology, University of Edinburgh, Edinburgh, UK
[4]Cancer Therapeutics and mRNA Dysregulation, Department of Oncology, Medical Sciences Division, University of Oxford, Oxford, UK
[5]Early Phase Clinical Trials Unit, Department of Oncology, Medical Siences Division, University of Oxford, Oxford, UK
[6]St Anne's College, University of Oxford, Oxford, UK
[7]Molecular Oncology Laboratories, Department of Oncology, Medical Sciences Division, University of Oxford, Oxford, UK
[8]Institute of Cancer & Genomic Sciences, University of Birmingham, Birmingham, UK

ORCID iDs
Benjamin HL Harris http://orcid.org/0000-0002-8154-8953
Francesca M. Buffa http://orcid.org/0000-0003-0409-406X

REFERENCES
1 Albanes D, Jones DY, Schatzkin A, Micozzi MS, Taylor PR. Adult stature and risk of cancer. *Cancer Res* 1988;48:1658–62.
2 Renehan AG. Height and cancer: consistent links, but mechanisms unclear. *Lancet Oncol* 2011;12:716–7.
3 Bhaskaran K, Douglas I, Forbes H, dos-Santos-Silva I, Leon DA, Smeeth L. Body-Mass index and risk of 22 specific cancers: a population-based cohort study of 5·24 million UK adults. *Lancet* 2014;384:755–65.
4 Park J, Morley TS, Kim M, Clegg DJ, Scherer PE. Obesity and cancer--mechanisms underlying tumour progression and recurrence. *Nat Rev Endocrinol* 2014;10:455–65.
5 Benyi E, Linder M, Adami J, Kieler H, Palme M, Sävendahl L. Adult height is associated with risk of cancer and mortality in 5.5 million Swedish women and men. *J Epidemiol Community Health* 2019;73:730–6.
6 Clayton PE, Banerjee I, Murray PG, Renehan AG. Growth hormone, the insulin-like growth factor axis, insulin and cancer risk. *Nat Rev Endocrinol* 2011;7:11–24.
7 Nunney L. Size matters: height, cell number and a person's risk of cancer. Proceedings of the Royal Society of London. Series B, Biological sciences 285, 2018.
8 Peto R. *Epidemiology, multistage models, and short-term mutagenicity tests*, 1977.
9 Arnold M, Leitzmann M, Freisling H, Bray F, Romieu I, Renehan A, Soerjomataram I. Obesity and cancer: an update of the global impact. *Cancer Epidemiol* 2016;41:8–15.
10 Renehan AG, Zwahlen M, Egger M. Adiposity and cancer risk: new mechanistic insights from epidemiology. *Nat Rev Cancer* 2015;15:484–98.
11 Hanahan D, Weinberg RA. The hallmarks of cancer. *Cell* 2000;100:57–70.
12 Li FP, Fraumeni JF. Soft-Tissue sarcomas, breast cancer, and other neoplasms. A familial syndrome? *Ann Intern Med* 1969;71:747–52.
13 Soussi T, Dehouche K, Béroud C. P53 website and analysis of p53 gene mutations in human cancer: forging a link between epidemiology and carcinogenesis. *Hum Mutat* 2000;15:105–13.
14 Liu S, Kim T-H, Franklin DA, Zhang Y. Protection against High-Fat-Diet-Induced Obesity in MDM2^{C305F} Mice Due to Reduced p53 Activity and Enhanced Energy Expenditure. *Cell Rep* 2017;18:1005–18.
15 Kung C-P, Leu JI-J, Basu S, Khaku S, Anokye-Danso F, Liu Q, George DL, Ahima RS, Murphy ME. The P72R polymorphism of p53 predisposes to obesity and metabolic dysfunction. *Cell Rep* 2016;14:2413–25.
16 Bignon YJ, Chen Y, Chang CY, Riley DJ, Windle JJ, Mellon PL, Lee WH. Expression of a retinoblastoma transgene results in dwarf mice. *Genes Dev* 1993;7:1654–62.
17 Matsuura H, Ichiki T, Inoue E, Nomura M, Miyazaki R, Hashimoto T, Ikeda J, Takayanagi R, Fong G-H, Sunagawa K. Prolyl hydroxylase domain protein 2 plays a critical role in diet-induced obesity and glucose intolerance. *Circulation* 2013;127:2078–87.
18 Zhang X, Lam KSL, Ye H, Chung SK, Zhou M, Wang Y, Xu A. Adipose tissue-specific inhibition of hypoxia-inducible factor 1{alpha} induces obesity and glucose intolerance by impeding energy expenditure in mice. *J Biol Chem* 2010;285:32869–77.
19 Levine AJ. P53, the cellular gatekeeper for growth and division. *Cell* 1997;88:323–31.
20 Al Alwan I, M K, G N, A O, L B, M AD, M B, Khadora M, Nasrat G, Omair A, Brown L, Al Dubayee M, Badri M. Turner syndrome genotype and phenotype and their effect on presenting features and timing of diagnosis. *Int J Health Sci* 2014;8:195–202.
21 Rao E, Weiss B, Fukami M, Rump A, Niesler B, Mertz A, Muroya K, Binder G, Kirsch S, Winkelmann M, Nordsiek G, Heinrich U, Breuning MH, Ranke MB, Rosenthal A, Ogata T, Rappold GA. Pseudoautosomal deletions encompassing a novel homeobox gene cause growth failure in idiopathic short stature and Turner syndrome. *Nat Genet* 1997;16:54–63.
22 Stuppia L, Calabrese G, Gatta V, Pintor S, Morizio E, Fantasia D, Guanciali Franchi P, Rinaldi MM, Scarano G, Concolino D, Giannotti A, Petreschi F, Anzellotti MT, Pomilio M, Chiarelli F, Tumini S, Palka G. Shox mutations detected by FISH and direct sequencing in patients with short stature. *J Med Genet* 2003;40:11e–11.
23 Song HH, Shi W, Xiang Y-Y, Filmus J. The loss of glypican-3 induces alterations in Wnt signaling. *J Biol Chem* 2005;280:2116–25.
24 Filmus J, Capurro M. The role of glypican-3 in the regulation of body size and cancer. *Cell Cycle* 2008;7:2787–90.
25 Neri G, Gurrieri F, Zanni G, Lin A. Clinical and molecular aspects of the Simpson-Golabi-Behmel syndrome. *Am J Med Genet* 1998;79:279–83.
26 UK Biobank. *Anthropometry*, 2014.
27 Biobank UK. *Genotyping of 500,000 UK Biobank participants: description of sample processing workflow and preparation of DNA for genotyping*, 2015.
28 Auton A, Salcedo T. *Assessing rare variation in complex traits*. Springer, 2015.
29 UK10K Consortium, Walter K, Min JL, Huang J, Crooks L, Memari Y, McCarthy S, Perry JRB, Xu C, Futema M, Lawson D, Iotchkova V, Schiffels S, Hendricks AE, Danecek P, Li R, Floyd J, Wain LV, Barroso I, Humphries SE, Hurles ME, Zeggini E, Barrett JC, Plagnol V, Richards JB, Greenwood CMT, Timpson NJ, Durbin R, Soranzo N. The UK10K project identifies rare variants in health and disease. *Nature* 2015;526:82–90.
30 UK Biobank Consortium. *Genotype imputation and genetic association studies of UK Biobank*, 2015.
31 Welter D, MacArthur J, Morales J, Burdett T, Hall P, Junkins H, Klemm A, Flicek P, Manolio T, Hindorff L, Parkinson H. The NHGRI GWAS catalog, a curated resource of SNP-trait associations. *Nucleic Acids Res* 2014;42:D1001–6.
32 GTEx Consortium. The Genotype-Tissue expression (GTEx) project. *Nat Genet* 2013;45:580–5.
33 Gong J, Mei S, Liu C, Xiang Y, Ye Y, Zhang Z, Feng J, Liu R, Diao L, Guo A-Y, Miao X, Han L. PancanQTL: systematic identification of cis-eQTLs and trans-eQTLs in 33 cancer types. *Nucleic Acids Res* 2018;46:D971–6.
34 Jansen R, Hottenga J-J, Nivard MG, Abdellaoui A, Laport B, de Geus EJ, Wright FA, Penninx BWJH, Boomsma DI. Conditional eQTL analysis reveals allelic heterogeneity of gene expression. *Hum Mol Genet* 2017;26:1444–51.
35 Marchini J, Howie B. Genotype imputation for genome-wide association studies. *Nat Rev Genet* 2010;11:499–511.
36 Machiela MJ, Chanock SJ. LDlink: a web-based application for exploring population-specific haplotype structure and linking correlated alleles of possible functional variants. *Bioinformatics* 2015;31:3555–7.
37 Schaid DJ, Chen W, Larson NB. From genome-wide associations to candidate causal variants by statistical fine-mapping. *Nat Rev Genet* 2018;19:491–504.

38 Hernandez Cordero AI, Gonzales NM, Parker CC, Sokoloff G, Vandenbergh DJ, Cheng R, Abney M, Skol A, Douglas A, Palmer AA, Gregory JS, Lionikas A. Genome-Wide associations reveal human-mouse genetic convergence and modifiers of myogenesis, CPNE1 and STC2. *Am J Hum Genet* 2020;106.

39 Hübel C, Gaspar HA, Coleman JRI, Finucane H, Purves KL, Hanscombe KB, Prokopenko I, Graff M, Ngwa JS, Workalemahu T, O'Reilly PF, Bulik CM, Breen G, MAGIC investigators, Eating Disorders Working Group of the Psychiatric Genomics Consortium, Major Depressive Disorder Working Group of the Psychiatric Genomics Consortium, Schizophrenia Working Group of the Psychiatric Genomics Consortium, Tourette Syndrome/Obsessive-Compulsive Disorder Working Group of the Psychiatric Genomics Consortium. Genomics of body fat percentage may contribute to sex bias in anorexia nervosa. *Am J Med Genet B Neuropsychiatr Genet* 2019;180:428–38.

40 Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, Maller J, Sklar P, de Bakker PIW, Daly MJ, Sham PC. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* 2007;81:559–75.

41 Wood AR, Esko T, Yang J, Vedantam S, Pers TH, Gustafsson S, Chu AY, Estrada K, Luan Jian'an, Kutalik Z, Amin N, Buchkovich ML, Croteau-Chonka DC, Day FR, Duan Y, Fall T, Fehrmann R, Ferreira T, Jackson AU, Karjalainen J, Lo KS, Locke AE, Mägi R, Mihailov E, Porcu E, Randall JC, Scherag A, Vinkhuyzen AAE, Westra H-J, Winkler TW, Workalemahu T, Zhao JH, Absher D, Albrecht E, Anderson D, Baron J, Beekman M, Demirkan A, Ehret GB, Feenstra B, Feitosa MF, Fischer K, Fraser RM, Goel A, Gong J, Justice AE, Kanoni S, Kleber ME, Kristiansson K, Lim U, Lotay V, Lui JC, Mangino M, Mateo Leach I, Medina-Gomez C, Nalls MA, Nyholt DR, Palmer CD, Pasko D, Pechlivanis S, Prokopenko I, Ried JS, Ripke S, Shungin D, Stancáková A, Strawbridge RJ, Sung YJ, Tanaka T, Teumer A, Trompet S, van der Laan SW, van Setten J, Van Vliet-Ostaptchouk JV, Wang Z, Yengo L, Zhang W, Afzal U, Arnlöv J, Arscott GM, Bandinelli S, Barrett A, Bellis C, Bennett AJ, Berne C, Blüher M, Bolton JL, Böttcher Y, Boyd HA, Bruinenberg M, Buckley BM, Buyske S, Caspersen IH, Chines PS, Clarke R, Claudi-Boehm S, Cooper M, Daw EW, De Jong PA, Deelen J, Delgado G, Denny JC, Dhonukshe-Rutten R, Dimitriou M, Doney ASF, Dörr M, Eklund N, Eury E, Folkersen L, Garcia ME, Geller F, Giedraitis V, Go AS, Grallert H, Grammer TB, Gräßler J, Grönberg H, de Groot LCPGM, Groves CJ, Haessler J, Hall P, Haller T, Hallmans G, Hannemann A, Hartman CA, Hassinen M, Hayward C, Heard-Costa NL, Helmer Q, Hemani G, Henders AK, Hillege HL, Hlatky MA, Hoffmann W, Hoffmann P, Holmen O, Houwing-Duistermaat JJ, Illig T, Isaacs A, James AL, Jeff J, Johansen B, Johansson Åsa, Jolley J, Juliusdottir T, Junttila J, Kho AN, Kinnunen L, Klopp N, Kocher T, Kratzer W, Lichtner P, Lind L, Lindström J, Lobbens S, Lorentzon M, Lu Y, Lyssenko V, Magnusson PKE, Mahajan A, Maillard M, McArdle WL, McKenzie CA, McLachlan S, McLaren PJ, Menni C, Merger S, Milani L, Moayyeri A, Monda KL, Morken MA, Müller G, Müller-Nurasyid M, Musk AW, Narisu N, Nauck M, Nolte IM, Nöthen MM, Oozageer L, Pilz S, Rayner NW, Renstrom F, Robertson NR, Rose LM, Roussel R, Sanna S, Scharnagl H, Scholtens S, Schumacher FR, Schunkert H, Scott RA, Sehmi J, Seufferlein T, Shi J, Silventoinen K, Smit JH, Smith AV, Smolonska J, Stanton AV, Stirrups K, Stott DJ, Stringham HM, Sundström J, Swertz MA, Syvänen A-C, Tayo BO, Thorleifsson G, Tyrer JP, van Dijk S, van Schoor NM, van der Velde N, van Heemst D, van Oort FVA, Vermeulen SH, Verweij N, Vonk JM, Waite LL, Waldenberger M, Wennauer R, Wilkens LR, Willenborg C, Wilsgaard T, Wojczynski MK, Wong A, Wright AF, Zhang Q, Arveiler D, Bakker SJL, Beilby J, Bergman RN, Bergmans S, Biffar R, Blangero J, Boomsma DI, Bornstein SR, Bovet P, Brambilla P, Brown MJ, Campbell H, Caulfield MJ, Chakravarti A, Collins R, Collins FS, Crawford DC, Cupples LA, Danesh J, de Faire U, den Ruijter HM, Erbel R, Erdmann J, Eriksson JG, Farrall M, Ferrannini E, Ferrières J, Ford I, Forouhi NG, Forrester T, Gansevoort RT, Gejman PV, Gieger C, Golay A, Gottesman O, Gudnason V, Gyllensten U, Haas DW, Hall AS, Harris TB, Hattersley AT, Heath AC, Hengstenberg C, Hicks AA, Hindorff LA, Hingorani AD, Hofman A, Hovingh GK, Humphries SE, Hunt SC, Hypponen E, Jacobs KB, Jarvelin M-R, Jousilahti P, Jula AM, Kaprio J, Kastelein JJP, Kayser M, Kee F, Keinanen-Kiukaanniemi SM, Kiemeney LA, Kooner JS, Kooperberg C, Koskinen S, Kovacs P, Kraja AT, Kumari M, Kuusisto J, Lakka TA, Langenberg C, Le Marchand L, Lehtimäki T, Lupoli S, Madden PAF, Männistö S, Manunta P, Marette A, Matise TC, McKnight B, Meitinger T, Moll FL, Montgomery GW, Morris AD, Morris AP, Murray JC, Nelis M, Ohlsson C, Oldehinkel AJ, Ong KK, Ouwehand WH, Pasterkamp G, Peters A, Pramstaller PP, Price JF, Qi L, Raitakari OT, Rankinen T, Rao DC, Rice TK, Ritchie M, Rudan I, Salomaa V, Samani NJ, Saramies J, Sarzynski MA, Schwarz PEH, Sebert S, Sever P, Shuldiner AR, Sinisalo J, Steinthorsdottir V, Stolk RP, Tardif J-C, Tönjes A, Tremblay A, Tremoli E, Virtamo J, Vohl M-C, Amouyel P, Asselbergs FW, Assimes TL, Bochud M, Boehm BO, Boerwinkle E, Bottinger EP, Bouchard C, Cauchi S, Chambers

JC, Chanock SJ, Cooper RS, de Bakker PIW, Dedoussis G, Ferrucci L, Franks PW, Froguel P, Groop LC, Haiman CA, Hamsten A, Hayes MG, Hui J, Hunter DJ, Hveem K, Jukema JW, Kaplan RC, Kivimaki M, Kuh D, Laakso M, Liu Y, Martin NG, März W, Melbye M, Moebus S, Munroe PB, Njølstad I, Oostra BA, Palmer CNA, Pedersen NL, Perola M, Pérusse L, Peters U, Powell JE, Power C, Quertermous T, Rauramaa R, Reinmaa E, Ridker PM, Rivadeneira F, Rotter JI, Saaristo TE, Saleheen D, Schlessinger D, Slagboom PE, Snieder H, Spector TD, Strauch K, Stumvoll M, Tuomilehto J, Uusitupa M, van der Harst P, Völzke H, Walker M, Wareham NJ, Watkins H, Wichmann H-E, Wilson JF, Zanen P, Deloukas P, Heid IM, Lindgren CM, Mohlke KL, Speliotes EK, Thorsteinsdottir U, Barroso I, Fox CS, North KE, Strachan DP, Beckmann JS, Berndt SI, Boehnke M, Borecki IB, McCarthy MI, Metspalu A, Stefansson K, Uitterlinden AG, van Duijn CM, Franke L, Willer CJ, Price AL, Lettre G, Loos RJF, Weedon MN, Ingelsson E, O'Connell JR, Abecasis GR, Chasman DI, Goddard ME, Visscher PM, Hirschhorn JN, Frayling TM, Electronic Medical Records and Genomics (eMEMERGEGE) Consortium, MIGen Consortium, PAGEGE Consortium, LifeLines Cohort Study. Defining the role of common variation in the genomic and biological architecture of adult human height. *Nat Genet* 2014;46:1173–86.

42 Kichaev G, Bhatia G, Loh P-R, Gazal S, Burch K, Freund MK, Schoech A, Pasaniuc B, Price AL. Leveraging polygenic functional enrichment to improve GWAS power. *Am J Hum Genet* 2019;104:65–75.

43 Wong JCY, Alon N, Mckerlie C, Huang JR, Meyn MS, Buchwald M. Targeted disruption of exons 1 to 6 of the Fanconi anemia group a gene leads to growth retardation, strain-specific microphthalmia, meiotic defects and primordial germ cell hypoplasia. *Hum Mol Genet* 2003;12:2063–76.

44 Parmar K, D'Andrea A, Niedernhofer LJ. Mouse models of Fanconi anemia. *Mutat Res* 2009;668:133–40.

45 Wang P-Y, Ma W, Park J-Y, Celi FS, Arena R, Choi JW, Ali QA, Tripodi DJ, Zhuang J, Lago CU, Strong LC, Talagala SL, Balaban RS, Kang J-G, Hwang PM. Increased oxidative metabolism in the Li-Fraumeni syndrome. *N Engl J Med* 2013;368:1027–32.

46 Wang P-Y, Li J, Walcott FL, Kang J-G, Starost MF, Talagala SL, Zhuang J, Park J-H, Huffstutler RD, Bryla CM, Mai PL, Pollak M, Annunziata CM, Savage SA, Fojo AT, Hwang PM. Inhibiting mitochondrial respiration prevents cancer in a mouse model of Li-Fraumeni syndrome. *J Clin Invest* 2017;127:132–6.

47 Wheaton WW, Weinberg SE, Hamanaka RB, Soberanes S, Sullivan LB, Anso E, Glasauer A, Dufour E, Mutlu GM, Budigner GS, Chandel NS. Metformin inhibits mitochondrial complex I of cancer cells to reduce tumorigenesis. *Elife* 2014;3:e02242.

48 Stacey SN, Sulem P, Jonasdottir A, Masson G, Gudmundsson J, Gudbjartsson DF, Magnusson OT, Gudjonsson SA, Sigurgeirsson B, Thorisdottir K, Ragnarsson R, Benediktsdottir KR, Nexø BA, Tjønneland A, Overvad K, Rudnai P, Gurzau E, Koppova K, Hemminki K, Corredera C, Fuentelsaz V, Grasa P, Navarrete S, Fuertes F, García-Prats MD, Sanambrosio E, Panadero A, De Juan A, Garcia A, Rivera F, Planelles D, Soriano V, Requena C, Aben KK, van Rossum MM, Cremers RGHM, van Oort IM, van Spronsen D-J, Schalken JA, Peters WHM, Helfand BT, Donovan JL, Hamdy FC, Badescu D, Codreanu O, Jinga M, Csiki IE, Constantinescu V, Badea P, Mates IN, Dinu DE, Constantin A, Mates D, Kristjansdottir S, Agnarsson BA, Jonsson E, Barkardottir RB, Einarsson GV, Sigurdsson F, Moller PH, Stefansson T, Valdimarsson T, Johannsson OT, Sigurdsson H, Jonsson T, Jonasson JG, Tryggvadottir L, Rice T, Hansen HM, Xiao Y, Lachance DH, O Neill BP, Kosel ML, Decker PA, Thorleifsson G, Johannsdottir H, Helgadottir HT, Sigurdsson A, Steinthorsdottir V, Lindblom A, Sandler RS, Keku TO, Banasik K, Jørgensen T, Witte DR, Hansen T, Pedersen O, Jinga V, Neal DE, Catalona WJ, Wrensch M, Wiencke J, Jenkins RB, Nagore E, Vogel U, Kiemeney LA, Kumar R, Mayordomo JI, Olafsson JH, Kong A, Thorsteinsdottir U, Rafnar T, Stefansson K, Swedish Low-risk Colorectal Cancer Study Group. A germline variant in the TP53 polyadenylation signal confers cancer susceptibility. *Nat Genet* 2011;43:1098–103.

49 Haworth S, Shapland CY, Hayward C, Prins BP, Felix JF, Medina-Gomez C, Rivadeneira F, Wang C, Ahluwalia TS, Vrijheid M, Guxens M, Sunyer J, Tachmazidou I, Walter K, Iotchkova V, Jackson A, Cleal L, Huffmann J, Min JL, Sass L, Timmers PRHJ, Davey Smith G, Fisher SE, Wilson JF, Cole TJ, Fernandez-Orth D, Bønnelykke K, Bisgaard H, Pennell CE, Jaddoe VWV, Dedoussis G, Timpson N, Zeggini E, Vitart V, St Pourcain B, UK10K consortium. Low-Frequency variation in TP53 has large effects on head circumference and intracranial volume. *Nat Commun* 2019;10:357.

50 Li X, Liang L, Feng Y-CA, De Vivo I, Giovannucci E, Tang JY, Han J, Vivo D I, Height HJ. Height, height-related SNPs, and risk of non-melanoma skin cancer. *Br J Cancer* 2017;116:134–40.