ORIGINAL ARTICLE

Multitrait genome association analysis identifies new susceptibility genes for human anthropometric variation in the GCAT cohort

Iván Galván-Femenía,[1] Mireia Obón-Santacana,[1,2] David Piñeyro,[3] Marta Guindo-Martinez,[4] Xavier Duran,[1] Anna Carreras,[1] Raquel Pluvinet,[3] Juan Velasco,[1] Laia Ramos,[3] Susanna Aussó,[3] J M Mercader,[5,6] Lluis Puig,[7] Manuel Perucho,[8] David Torrents,[4,9] Victor Moreno,[2,10] Lauro Sumoy,[3] Rafael de Cid[1]

► Additional material is published online only. To view, please visit the journal online (http://dx.doi.org/10.1136/jmedgenet-2018-105437).

For numbered affiliations see end of article.

Correspondence to
Dr Rafael de Cid, GCAT lab Group, Program of Predictive and Personalized Medicine of Cancer (PMPPC), Germans Trias i Pujol Research Institute (IGTP), Crta. de Can Ruti, Badalona 08916, Spain; rdecid@igtp.cat

Received 24 April 2018
Revised 19 July 2018
Accepted 21 July 2018
Published Online First 30 August 2018

To cite: Galván-Femenía I, Obón-Santacana M, Piñeyro D, *et al*. *J Med Genet* 2018;**55**:765–778.

ABSTRACT

Background Heritability estimates have revealed an important contribution of SNP variants for most common traits; however, SNP analysis by single-trait genome-wide association studies (GWAS) has failed to uncover their impact. In this study, we applied a multitrait GWAS approach to discover additional factor of the missing heritability of human anthropometric variation.

Methods We analysed 205 traits, including diseases identified at baseline in the GCAT cohort (Genomes For Life- Cohort study of the Genomes of Catalonia) (n=4988), a Mediterranean adult population-based cohort study from the south of Europe. We estimated SNP heritability contribution and single-trait GWAS for all traits from 15 million SNP variants. Then, we applied a multitrait-related approach to study genome-wide association to anthropometric measures in a two-stage meta-analysis with the UK Biobank cohort (n=336 107).

Results Heritability estimates (eg, skin colour, alcohol consumption, smoking habit, body mass index, educational level or height) revealed an important contribution of SNP variants, ranging from 18% to 77%. Single-trait analysis identified 1785 SNPs with genome-wide significance threshold. From these, several previously reported single-trait hits were confirmed in our sample with *LINC01432* ($p=1.9\times10^{-9}$) variants associated with male baldness, *LDLR* variants with hyperlipidaemia (ICD-9:272) ($p=9.4\times10^{-10}$) and variants in *IRF4* ($p=2.8\times10^{-57}$), *SLC45A2* ($p=2.2\times10^{-130}$), *HERC2* ($p=2.8\times10^{-176}$), *OCA2* ($p=2.4\times10^{-121}$) and *MC1R* ($p=7.7\times10^{-22}$) associated with hair, eye and skin colour, freckling, tanning capacity and sun burning sensitivity and the Fitzpatrick phototype score, all highly correlated cross-phenotypes. Multitrait meta-analysis of anthropometric variation validated 27 loci in a two-stage meta-analysis with a large British ancestry cohort, six of which are newly reported here (p value threshold $<5\times10^{-9}$) at *ZRANB2-AS2*, *PIK3R1*, *EPHA7*, *MAD1L1*, *CACUL1* and *MAP3K9*.

Conclusion Considering multiple-related genetic phenotypes improve associated genome signal detection. These results indicate the potential value of data-driven multivariate phenotyping for genetic studies in large population-based cohorts to contribute to knowledge of complex traits.

INTRODUCTION

Common disorders cause 85% of deaths in the European Union (EU).[1] The increasing incidence and prevalence of cancer, cardiovascular diseases, chronic respiratory diseases, diabetes and mental illness represent a challenge that leads to extra costs for the healthcare system. Moreover, as European population is getting older, this scenario will be heightened in the next few years. Like complex traits, many common diseases are complex inherited conditions with genetic and environmental determinants. Advancing in their understanding requires the use of multifaceted and long-term prospective approaches. Cohort analyses provide an exceptional tool for dissecting the architecture of complex diseases by contributing knowledge for evidence-based prevention, as exemplified by the Framingham Heart Study[2] or the European Prospective Investigation into Cancer and Nutrition cohort study.[3]

In the last decades, high performance DNA genotyping technology has fuelled genomic research in large cohorts, having been the most promising line in research on the aetiology of most common diseases. Genome-wide association studies (GWAS) have provided valuable information for many single conditions.[4] Despite the perception of the limitations of the GWAS analyses, efforts combining massive data deriving from whole-genome sequencing at population scale with novel conceptual and methodological analysis frameworks have been set forth to explore the last frontier of the missing heritability issue,[5] driving the field of genomic research on complex diseases to a new age.[6] Pritchard and colleagues recently proposed the breakthrough idea of the *omnigenic* character of genetic architecture of diseases and complex traits.[7] They suggested that beyond a handful of driver genes (ie, core genes) directly connected to an illness, the missing heritability could be accounted for by multiple genes (ie, peripheral genes) not clustered in functional pathways, but dispersed along the genome, explaining the pleiotropy frequently seen in most complex traits. Core genes have been already outlined by the GWAS approach, but most of the possible contributing genes have been disregarded based on methodological issues such as p value or lower minor

allele frequency (MAF). Pathway disturbances have also been a landmark in the search for genetic associations,[8] but not always appear to the root of the mechanism of inheritance of complex diseases, at least for peripheral genes.[7] With this challenging vision, a multitrait genome association analysis of the whole phenome[9] becomes a more appropriate way to detect peripheral gene variation effects and new network disturbances affecting core genes. Multitrait analysis approaches are developed for research of genetically complex conditions using raw or summary-level data statistics from GWAS in order to explain the largest possible amount of the covariation between SNPs and traits.[10–15]

The contribution of total genetic variation, known as heritability (broad-sense heritability, h^2), is estimated now from genome-wide studies in large cohorts directly from SNP data (known as h^2SNP). However, even if most disease conditions have a strong genetic basis, it is well known that our capacity to find genetic effects depends on the overall genetic contribution of the trait. Overall estimations differed depending on the ancestry, sample ascertainment, gender and age of the population under study. Recently, data from the UK Biobank determined genetic contributions with a phenome-based approach[16] and identified a shared familial environment as a significant important factor besides genetic *heritability* values in 12 common diseases analysed.[17]

In this study, we present new data on phenotype-wide estimation of the heritability of 205 complex traits (including diseases) and new insights into the genetics of anthropometric traits in a Mediterranean Caucasian population using a two-stage meta-analysis approach with multiple-related phenotypes (MRPs).

MATERIALS AND METHODS

Population

The methodology of the GCAT study has been previously described.[18] Briefly, the subjects of the present study are part of the GCAT project, a prospective study that includes a cohort of a total of 19 267 participants recruited from the general population of Catalonia, a western Mediterranean region in the Northeast of Spain. Healthy general population volunteers between 40 and 65 years with the sole condition of being users of the Spanish National Health Service were invited to be part of the study mostly through the Blood and Tissue Bank, a public agency of the Catalan Department of Health. All eligible participants signed an informed consent agreement form and answered a comprehensive epidemiological questionnaire. Anthropometric measures and blood samples were also collected at baseline by trained healthcare personnel. The GCAT study was approved by the local ethics committee (Germans Trias University Hospital) in 2013 and started on 2014.

Study participants

This study analyses the GCATcore data, a subset of 5459 participants (3066 women) with genotype data belonging to the interim GCATdataset, August 2017 (see the URLs section). GCATcore participants were randomly selected from whole cohort based on overall demographic distribution (ie, gender, age, residence). In this study, in order to increase the robustness of heritability estimates, only Caucasian participants with a Spanish origin (based on principal component analysis (PCA) analysis, see later in this section) and with available genetic data were finally included: 4988 GCAT participants (2777 women). All samples passed genotyping quality control (QC) (see later in this section).

Phenome

Baseline variables were obtained from a self-reported epidemiological questionnaire and included biological traits, medical diagnoses, drug use, lifestyle habits and sociodemographic and socioeconomic variables.[18] Description of GCAT variables dataset is available at GCAT (see the URLs section). To keep as many as possible of the genotyped samples in the study, we imputed anthropometric missing values (<1%) from the overall distribution values using statistical approaches. Missing values (<1%) for biological and anthropometric measures (height, weight, waist and hip circumference, systolic and diastolic blood pressure and heart rate) were imputed by stratifying the whole GCAT cohort by gender and age and using multiple imputation by the fully conditional specification method, implemented in the R mice package.[19] For GWAS analysis, we retained all variables with at least five observations (n=205). For heritability estimates, only variables with at least 500 individuals per class were retained (n=96) for robustness. The description of the traits and measures included in this study is summarised in online supplementary table S1.

Genotyping, relatedness and population structure

Genotyping of the 5459 GCAT participants (GCATcore) was done using the Infinium Expanded Multi-Ethnic Genotyping Array (MEGA^{Ex}) (ILLUMINA, San Diego, California, USA). A customised cluster file was produced from the entire sample dataset and used for joint calling. We applied PCA to detect any hidden substructure and the method of moments for the estimation of identity by descent probabilities to exclude cases with cryptic relatedness. The extensive QC protocol used for cluster analysis and call filtering is accessible at GCAT (see the URLs section) and presented as supplementary material (online supplementary file S1). Briefly, GCAT participants were excluded from the analysis for different reasons, including poor call rate <0.94 (n=61), gender mismatch (n=19), duplicates (n=8), family relatedness up to second degree (n=88) and excess or loss of heterozygosity (n=52). Non-Caucasian individuals detected as outliers in the PCA plot of the European populations from the 1000 Genomes Project (n=96) and born outside of Spain (n=147) were also excluded from the study. After QC and filtering, 4988 GCAT participants and 1 652 023 genetic variants were included. Genotyping was performed at the PMPPC-IGTP High Content Genomics and Bioinformatics Unit.

Multipanel imputation

For imputation analysis, 665 592 SNPs were included (40%). Sexual and mitochondrial chromosomes were discarded as well as autosomal chromosome variants with MAF <0.01 and AT-CG sites. We followed a two-stage imputation procedure, which consists of prephasing the genotypes into whole chromosome haplotypes followed by imputation itself.[20] The prephasing was performed using SHAPEIT2, and genotype imputation was performed with IMPUTE2. As reference panels for genotype imputation, we used the 1000 Genomes Project phase 3,[21] the Genome of the Netherlands,[22] UK10K[23] and the Haplotype Reference Consortium.[24] All variants with IMPUTE2 *info* <0.7 were removed. After imputing the genotypes using each reference panel separately, we combined the results selecting the variants with a higher *info* score when they were present in more than one reference panel. The SNP dosage from IMPUTE2 was transformed to binary PLINK format by using the '-hard-call-threshold 0.1' flag from PLINK. The final core set had approximately 15 million variants with MAF>0.001 and 9.5 million

variants with MAF>0.01. Imputation was performed at the Barcelona Supercomputing Center.

Heritability

Trait SNP heritability (h^2_{SNP}) was estimated from SNP/INDEL array/imputed data with the GREML-LDMS method implemented in the GCTA software.[25] Since this method is relatively unbiased regarding MAF and linkage disequilibrium (LD) parameters, we considered autosomal variants with MAF>0.001 (15 060 719 SNPs) to avoid under/overestimation of heritability due to the relatively small sample analysed in the core study. Cryptic relatedness of distant relatives was also considered, and individuals whose relatedness in the genetic relationship matrix was >0.025 were discarded (n=4717). Population stratification was controlled in the linear mixed model using the first 20 principal components of the PCA derived from population genetic structure analysis of the GCAT. Gender and age were also included as covariates in the model. The h^2_{SNP}CIs were calculated by using FIESTA.[26]

Single-trait genome-wide association analysis

We performed independent GWAs analyses for 205 selected traits (61 continuous and 144 binary). A total of 9 499 600 SNPs with MAF>0.01 were considered for this purpose. Linear regression models for continuous traits were assessed with PLINK.[27] For binary traits, given the unbalanced design of most of the traits considered, we used a scoring test with saddle point approximation included in the *SPAtest* R package.[28] This approach compensates a slight loss of power with the inclusion of uncommon and rare conditions, without affecting robustness. All the models included the first 20 PCAs, age and gender as covariates. A PCA-mixed analysis was applied to approximate the number of independent traits[29] (online supplementary figure S1). Based on these figures, Bonferroni correction for multiple traits was defined at $p<5\times10^{-10}$accounting for 100 independent traits explaining 80% of the phenome variability.

Multitrait meta-analysis for correlated traits

We applied a multitrait approach for the analysis of anthropometric traits (weight, height, body mass index (BMI) and waist and hip circumference) in a two-stage association study using individuals of British ancestry from the UK Biobank cohort (N=336 107).[30] Waist-to-hip ratio was excluded from this analysis due to its unavailability from the UK Biobank resource. UK Biobank summary-level statistics was calculated using linear regression models with the inferred gender and the first 10 PCAs as covariates, similarly to the model applied on GCAT data (see the URLs section). All SNPs with suggestive association $p<1x10^{-5}$ for any trait were retained from the GCAT GWAS analysis. Then, only SNPs intersecting with the UK Biobank resource were used for multitrait meta-analysis association testing in both samples, and $p<5x10^{-9}$was considered significant. The multitrait association testing was based on the distribution of the sum of squares of the z scores which is insensitive to the direction of the scores.[31] Briefly, let $Z = (z_1, z_2, \ldots, z_k)$ be the z scores for a given SNP for k phenotypes. The sum of squares of the z scores, $S_{sq} = \sum_{i=1}^{k} z_i^2$, can be approximated by the χ^2 distribution (χ^2). Let Σ be the covariance matrix of the genome-wide z scores from the phenotypes under analysis. And let c_i be the eigenvalues of Σ, the distribution of S_{sq} is well approximated by $a\chi^2_d + b$, where a, b and d depend on c_i. Then, we calculated the p value as: $p\left(\chi^2_d > \left(S_{sq} - b\right)/a\right)$. To estimate the covariance matrix of the correlated traits, we selected independent SNPs (LD pruning in PLINK "--indep-pairwise 50 5 0.2") and filtered out SNPs with |z scores|>1.96 to avoid possible bias in the estimation of Σ because of the difference in sample size and association p values in the GCAT-UK Biobank. A summary flow chart of the methods applied in this study is shown in figure 1.

Polygenic risk score

Genetic architecture was analysed by the polygenic risk score (PRS). Polygenic risk score software (PRSice)[32] was used to predict the genetic variability of the identified loci for a given trait. PRSice plots the percentage of variance explained for a trait by using SNPs with different p value thresholds (P_T) (online supplementary figure S2). Here, we considered $P_T=0.05$.

URLs

GCAT study, http://genomesforlife.com;

National Human Genome Research Institute GWAS Catalog, http://www.genome.gov/gwastudies/ (gwas_catalog_v1.0-associations_e91_r2018-02-06);

1000 Genomes Project http://www.internationalgenome.org/ (phase 3, v5a.20130502);

Genome of Netherland http://www.nlgenome.nl/ (Release 5.4);

UK10K https://www.uk10k.org/ (Release 2012-06-02, updated on 15 Feb 2016) ;

Haplotype Reference Consortium http://www.haplotype-reference-consortium.org/(Release 1.1);

UKBiobank GWAS Results; https://sites.google.com/broadinstitute.org/ukbbgwasresults/home?authuser=0, (Manifest20170915);

GTExportal, https://www.gtexportal.org/home/. (last data accession, Release V.7, dbGaP accession phs000424. v7. P2);

RESULTS

Heritability estimates

SNP heritability estimation (h^2_{SNP}) in the GCATcore study showed values ranging from 77% to 18%, with height being the trait showing the strongest SNP contribution. The h^2_{SNP} SE for most traits was high (near 10%), with wide CIs, as expected by sample size. However, robustness of the analysis is supported by similar values to those reported elsewhere (see wide summary in Genome-wide complex trait analysis, Wikipedia. *The Free Encyclopedia*, 2018). Statistically significant h^2_{SNP} estimations for continuous and binary traits (cases >500) are shown in table 1. In particular, values for height: $h^2_{SNP}=0.77$, 95% CI0.56 to 0.94 and BMI: $h^2_{SNP}=0.38$, 95% CI0.20 to 0.59 were identical to the maxima achieved in other European populations, using comparable genomic approaches. Besides the anthropometric traits, the Fitzpatrick's phototype score, a numerical classification schema for human skin colour to measure the response of different types of skin to ultraviolet light, had a high genetic consistency in our sample ($h^2_{SNP}=0.63$, 95% CI 0.4 to 0.8), and concordantly all related categories (eye colour, hair colour, freckling and skin sensitivity) showed high heritability ($h^2_{SNP}>0.3$). It is worth noting that skin colour had the lowest value ($h^2_{SNP}=0.18$, 95% CI 0.02 to 0.38), which is in concordance with the blurred genetic architecture of skin colour.[33] Interestingly, other non-biological traits showed relatively high values in our study. Educational level showed the third highest heritability value ($h^2_{SNP}=0.54$, 95% CI 0.35 to 0.74). Lower estimates have been observed in other Caucasian populations, but this could be explained by the fact that this estimate is for educational level as a categorical variable and not as binary (higher/lower).

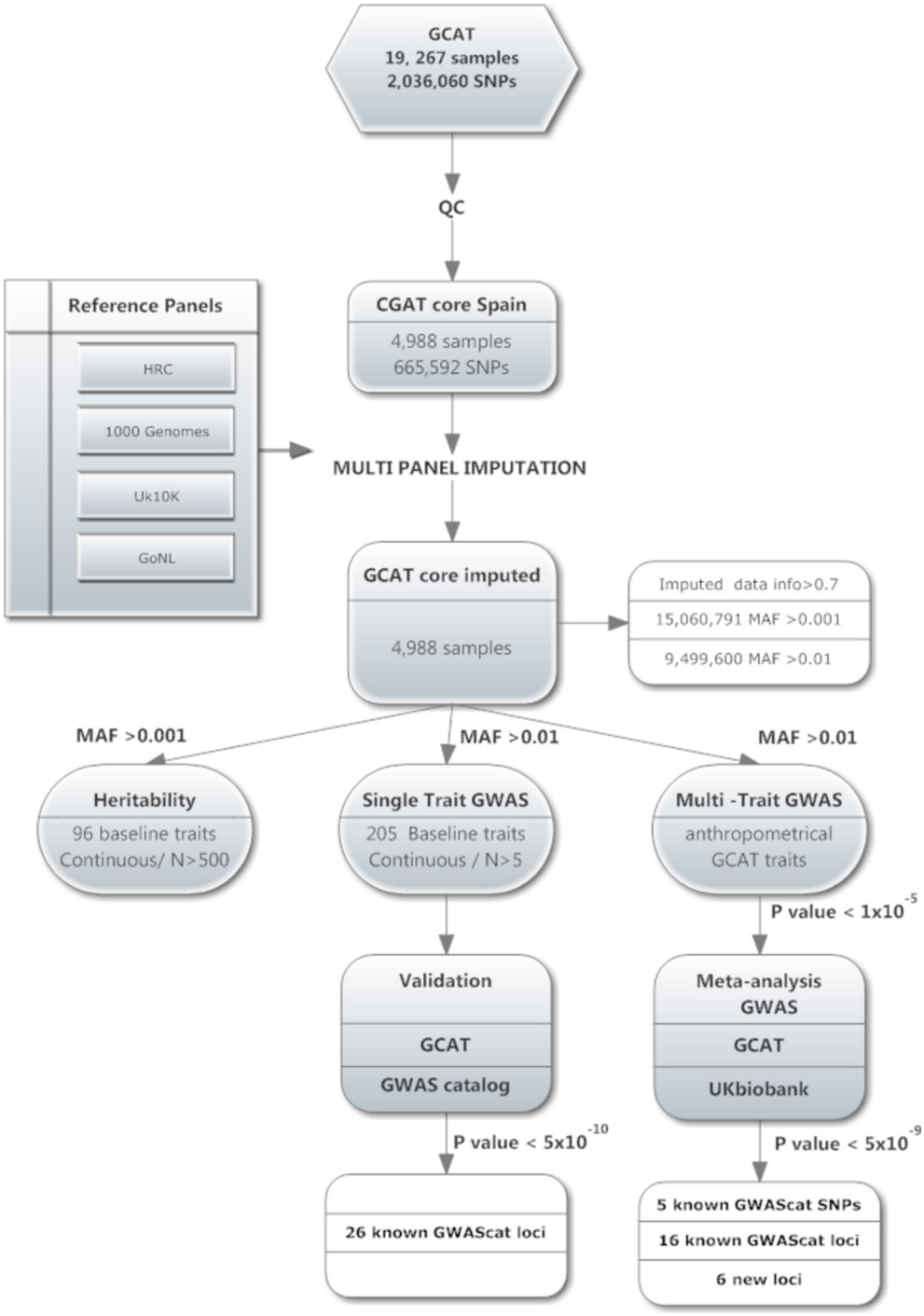

Figure 1 Flow chart of the methods and criteria used in this study. GCAT, Genomes For Life- Cohort Study of the Genomes of Catalonia; GWAS, genome-wide association studies; MAF, minor allele frequency; QC, quality control.

Self-perceived health was similar to h^2_{SNP} from recent data from a larger UK Biobank study,[16] with values around 20% (h^2_{SNP}=0.22, 95% CI 0.04 to 0.43).

Phenome analysis

GWAS identified 6820 associations in 1785 SNPs with genome-wide significance threshold $p<5\times10^{-8}$ and 29 343 associations with a suggestive association $p<1\times10^{-5}$. Here, we report 26 genome-wide association hits identified in our study which confirm results previously identified in other European ancestry samples (GWAS Catalog database (release V.1.0, e90, 27 September 2017)).[4] In table 2, we show the SNP associations with the minimum p value for each locus, the remaining SNPs are shown in online Supplementary file 5. Five genes associated with pigmentary traits were identified in the analysis with highly significant SNP associations: *SLC45A2* (rs16891982, β=−0.546, SE=0.021, $p=2.2\times10^{-130}$), *IRF4* (rs12203592, β=1.915, SE=0.118, $p=2.8\times10^{-57}$), *HERC2* (rs1667394, β=−0.608, SE=0.02, $p=2.8\times10^{-176}$), *OCA2* (rs11855019, β=−0.548, SE=0.022, $p=2.4\times10^{-121}$) and *MC1R* (rs1805007, β=3.615, SE=0.326, $p=7.7\times10^{-22}$) (online supplementary figure S3). These genes are involved in the regulation and distribution of melanin pigmentation or enzymes involved in melanogenesis itself within the melanocyte cells present in the skin, hair and eyes in Caucasian populations.[33–35] Pigmentary traits (mainly

Table 1 h^2_{SNP} of the analysed traits with h^2_{SNP}>0, SE <0.12, p<0.05 and n_b>500

Questionnaire—section	Description	Trait name	h2SNP	SE	95% CI	P values	n	n_b	NA
Anthropometric and blood pressure	Height	height_c	0.77	0.11	0.56 to 0.94	2×10^{-12}	4717	–	0
Other habits	Phototype score	phototype_ score	0.63	0.11	0.4 to 0.8	3.7×10^{-9}	4664	–	56
Demographic and socioeconomic	Educational level	education	0.54	0.10	0.35 to 0.74	1.1×10^{-8}	4698	–	19
Other habits	Fitzpatrick phototype score	phototype_score categorical	0.52	0.11	0.29 to 0.74	6.0×10^{-7}	4664	–	56
Other habits	Eye colour phototype score	eye_color_phototype_score	0.48	0.11	0.27 to 0.68	7.1×10^{-6}	4716	–	1
Other habits	Freckling (has freckles)	freckling_binary	0.47	0.11	0.26 to 0.68	8.1×10^{-6}	4713	590	4
Other habits	Hair colour phototype score	hair_color_phototype_score	0.46	0.11	0.26 to 0.68	6.7×10^{-6}	4709	–	9
Other habits	Eye colour	eye_color	0.44	0.11	0.24 to 0.65	3.4×10^{-5}	4716	–	1
Other habits	Hair colour	hair_color	0.41	0.11	0.21 to 0.63	4.1×10^{-5}	4709	–	9
Other habits	Hair colour (black)	hair_color_black	0.39	0.11	0.22 to 0.59	0.00018	4709	952	9
Anthropometric and blood pressure	BMI (kg/m^2)	bmi	0.38	0.11	0.2 to 0.59	0.00013	4717	–	0
Anthropometric and blood pressure	Weight	weight_c	0.37	0.11	0.19 to 0.57	0.00016	4717	–	0
Tobacco consumption	Smoking habit	smoking_habit	0.36	0.11	0.19 to 0.58	0.00037	4717	–	0
Tobacco consumption	Smoking packs per day	smoking_packs	0.35	0.11	0.17 to 0.55	0.00082	4717	–	0
Other habits	Skin sensitivity to sun	skin_sensitivity_to_sun	0.33	0.11	0.15 to 0.52	0.0011	4714	–	3
Anthropometric and blood pressure	Hip circumference	hip_c	0.31	0.11	0.15 to 0.51	0.0011	4717	–	0
Occupation	Working status (active)	working_status_active	0.31	0.11	0.13 to 0.54	0.0014	4696	1570	23
Other habits	Skin sensitivity to sun phototype score	skin_sensitivity_to_sun_ phototype_score	0.30	0.11	0.12 to 0.51	0.0022	4714	–	3
Anthropometric and blood pressure	BMI obesity	bmi_who_obesity	0.29	0.11	0.12 to 0.51	0.0031	4717	1388	0
Physical activity	Sleep duration	sleep_duration	0.29	0.11	0.1 to 0.49	0.0033	4645	–	79
Other habits	Freckling	freckling	0.28	0.11	0.11 to 0.5	0.0043	4713	–	4
Medical history	Mental health (MHI-5)	sadness	0.26	0.11	0.09 to 0.48	0.0053	4717	504	0
Occupation	Working last year	working_last_year	0.26	0.11	0.09 to 0.47	0.0065	4685	1190	32
Other habits	Freckling phototype score	freckling_phototype_score	0.26	0.11	0.09 to 0.46	0.0076	4713	–	4
Other habits	Eye colour (dark)	eye_color_dark	0.25	0.11	0.07 to 0.47	0.012	4716	1192	1
Other habits	Hair colour (brown)	hair_color_brown	0.24	0.11	0.07 to 0.45	0.012	4709	1229	9
Anthropometric and blood pressure	Waist circumference	waist_c	0.24	0.11	0.06 to 0.44	0.01	4717	–	0
Anthropometric and blood pressure	Waist-to-hip ratio WHO categories	whr_who	0.23	0.11	0.05 to 0.45	0.016	4717	–	0
Medical history	Self-perceived health	self_perceived_health	0.22	0.11	0.04 to 0.43	0.024	4715	–	2
Tobacco consumption	Smoking status (ever smoked)	smoking_status	0.21	0.11	0.02 to 0.42	0.026	4522	1828	204
Alcohol consumption	Current alcohol consumption	alcohol_actual	0.20	0.11	0.03 to 0.4	0.031	4713	3670	4
Diet	Predimed score	predimed_score	0.20	0.11	0.03 to 0.41	0.031	4627	–	95
Women's health	No of female children	offspring_female	0.19	0.11	0.02 to 0.4	0.028	4717	–	0
Anthropometric and blood pressure	Waist-to-hip ratio obesity	whr_who_obesity	0.19	0.11	0.04 to 0.39	0.036	4717	1512	0
Women's health	No of male children	offspring_male	0.19	0.11	0.02 to 0.41	0.036	4717	–	0
Medical history	Self-perceived health (bad)	self_perceived_health_binary	0.18	0.11	0.02 to 0.4	0.047	4715	629	2
Medical history	Certain adverse effects not classified elsewhere	icd9_code3_995	0.18	0.11	0.01 to 0.37	0.042	4717	775	0
Demographic and socioeconomic	Civil status (ever been married)	civil_status_ever_married	0.18	0.11	0.01 to 0.38	0.04	4703	523	15
Other habits	Skin colour phototype score	skin_color_phototype_score	0.18	0.11	0.02 to 0.38	0.047	4714	–	3

BMI, body mass index; h^2_{SNP}, SNP heritability estimation; MHI-5, Mental Health Inventory 5-item questionnaire; n_b,sample size of the minor category in binary traits; _c for Weight_c, height_c, hip_c and waist_c mean calculated-imputed variable.

the red hair colour phenotype) are related to the defensive capacity of the skin in response to sun exposure (UV-induced skin tanning or sun burning), and it has been established as a risk factor for sun-induced cancers (both melanoma and non-melanocytic skin cancers).[36] Other GWAS hits from the phenome-wide analysis validated previously reported findings in *CCDC141-LOC105373766* (rs79146658, β=2.359, SE=0.374, $p=3.4\times10^{-10}$), *SMARCA4-LDLR* (rs10412048, β=−0.5, SE=0.079, $p=3.2\times10^{-10}$; rs6511720, β=−0.493, SE=0.08, $p=9.4\times10^{-10}$) and *LINC01432* (rs1160312, β=0.193, SE=0.03, $p=1.9\times10^{-9}$) loci, related with cardiovascular risk (heart_rate), hyperlipidaemia (icd9_code3_272) and male pattern baldness (hair_loss_40), respectively (see table 2).

Table 2 Twenty-six genome-wide associated loci with GCAT traits and reported in the GWAS Catalog

Gene	SNP	Chr:position*	Imputed	Info	GWAS Catalog traits†	Studies	Published year	GCAT trait	β	SE	P values
CCDC141	rs151041685	2:179725237	Yes	0.998	Resting heart rate	1	2016	heart_rate_c	2.06	0.361	1.2×10^{-8}**
CCDC141, LOC105373766	rs79146658	2:179786068	Yes	0.971	Diastolic blood pressure	1	2017	heart_rate_c	2.359	0.3749	3.4×10^{-10}
SLC45A2	rs16891982	5:33951693	Yes	0.985	Hair colour, eye colour, black versus non-black hair colour, skin sensitivity to sun, squamous cell carcinoma, melanoma, monobrow	6	2010, 2015, 2016, 2017	skin_color	−0.546	0.021	2.2×10^{-130}
DUSP22, IRF4	rs7773324	6:382559	Yes	0.986	Crohn's disease, inflammatory bowel disease	1	2015	freckling_phototype_score	0.281	0.045	6.5×10^{-10}**
IRF4	rs12203592	6:396321	No	–	Black versus blond hair colour, black versus red hair colour, hair colour, eye colour, freckling, progressive supranuclear palsy, non-melanoma skin cancer, tanning, sunburns, facial pigmentation, skin colour saturation, cutaneous squamous cell carcinoma, squamous cell carcinoma, basal cell carcinoma	9	2008, 2010, 2011, 2013, 2015, 2016	hair_color_phototype_score	1.915	0.118	2.8×10^{-57}
IRF4, LOC105374875	rs62389424	6:422631	Yes	0.882	Blond versus non-blond hair colour, brown versus non-brown hair colour, light versus dark hair colour, lung cancer in ever smokers	2	2015, 2017	freckling_phototype_score	−0.926	0.073	1.6×10^{-35}
LOC105374875	rs12210050	6:475489	No	–	Tanning, basal cell carcinoma, schizophrenia	4	2009, 2011, 2012, 2016	hair_phototype_score	1.025	0.123	1.7×10^{-16}
RNU2-47P, TYRP1	rs1408799	9:12672097	No	–	Blue versus green eyes, eye colour	2	2008, 2013	eye_phototype_score	0.453	0.071	2.2×10^{-10}
BNC2-LOC105375983		9:16884586	Yes	0.991	Cutaneous squamous cell carcinoma, basal cell carcinoma	2	2		−0.089	0.016	3.4×10^{-8}**
LOC107984363, TYR	rs1126809	11:89017961	Yes	0.993	Tanning, sunburns, cutaneous squamous cell carcinoma, squamous cell carcinoma, basal cell carcinoma	4	2013, 2016	skin_color	−1.672	0.282	3.5×10^{-9}**
LOC105370627	rs12896399	14:92773663	No	–	Blond versus brown hair colour, blue versus green eyes, black versus blond hair colour, hair colour, eye colour	4	2007, 2008, 2010, 2013	phototype_score	0.093	0.016	1.9×2.5^{-8}**
OCA2	rs11855019	15:28335820	No	–	Black versus blond hair colour, black versus red hair colour	1	2008	hair_color	−0.548	0.022	2.4×10^{-121}
HERC2	rs1667394	15:28530182	No	–	Blond versus brown hair colour, blue versus green eyes, blue versus brown eyes, eye colour	2	2007, 2012	eye_color	−0.608	0.02	2.8×10^{-176}
SPG7, RPL13	rs67689854	16:89625227	Yes	0.902	Stromal cell-derived factor 1 alpha levels	1	2016	eye_color	2.284	0.278	7.9×10^{-11}
SPATA33	rs35063026	16:89736157	Yes	0.987	Facial pigmentation, squamous cell carcinoma	2	2015, 2016	hair_color_red	3.112	0.309	5.4×10^{-17}
CDK10	rs258322	16:89755903	No	–	Black versus red hair colour, melanoma	5	2008, 2009, 2011, 2014, 2017	hair_color_red	2.431	0.267	1.2×10^{-13}
FANCA	rs12931267	16:89818732	Yes	0.989	Hair colour, freckling, skin sensitivity to sun	2	2015, 2017	hair_color_red	3.218	0.311	3.9×10^{-18}
MC1R	rs1805007	16:89986117	No	–	Freckles, blond versus brown hair colour, red versus non-red hair colour, skin sensitivity to sun, basal cell carcinoma, tanning, hair colour, sunburns, non-melanoma skin cancer, perceived skin darkness, cutaneous squamous cell, melanoma	7	2007, 2011, 2013, 2015, 2016, 2017	hair_color_red	3.615	0.326	7.7×10^{-22}
DEF8	rs146972365	16:90022693	Yes	0.974	Red versus non-red hair colour, light versus dark hair colour, brown versus non-brown hair colour	1	2015	hair_color	0.442	0.052	3×10^{-17}
AFG3L1P	rs8063160	16:90054709	Yes	0.988	Brown versus non-brown hair colour, light versus dark hair colour, red versus non-red hair colour	1	2015	hair_color_red	2.577	0.277	8.9×10^{-15}
TSPAN10	rs9747347	17:79606820	Yes	0.982	Myopia	1	2016	hair_color_phototype_score	−0.526	0.087	1.8×10^{-9}**
HMGN1P31-CDH20		18:58840518	Yes	0.953	Deep ovarian and/or rectovaginal disease with dense	1	2017	handedness	0.045	0.008	3.9×10^{-8}**

Continued

Table 2 Continued

Gene	SNP	Chr:position*	Imputed	Info	GWAS Catalog traits†	Studies	Published year	GCAT trait	β	SE	P values
SMARCA4, LDLR		19:11193949	Yes	0.999	Cholesterol, total	1	2017	icd9_code3_272	−0.501	0.079	3.2×10^{-10}
LDLR	rs6511720	19:11202306	No	–	LDL cholesterol, carotid intima media thickness, cardiovascular disease risk factors, lipoprotein-associated phospholipase A2 activity and mass, cholesterol, total, metabolite levels, lipid metabolism phenotypes, Abdominal aortic aneurysm	12	2008, 2009, 2010, 2011, 2012, 2013	icd9_code3_272	−0.493	0.081	9.4×10^{-10}**
RPL41P1-LINC01432		20:22000281	No	–	Male-pattern baldness	1	2016	hair_loss_40	0.19	0.032	6.2×10^{-9}**
LINC01432	rs1160312	20:22050503	No	–	Male-pattern baldness	1	2008	hair_loss_40	0.193	0.032	1.9×10^{-9}**

*Chr:position based on hg19.
†GWAS Catalog traits based on GWAS Catalog database (release V.1.0, e90, 27 September 2017).
5×10^{-8} threshold for univariate GWAS and 5×10^{-10} threshold accounting for multiple phenotypes.
GWAS, genome-wide association studies; LDL, low-density lipoprotein.
_c for heart-rate_c, means calculated-imputed variable.

Multitrait meta-analysis of anthropometric traits

Anthropometric traits had a high heritability in our sample (height=77%, BMI=38%, weight=37%, hip circumference=31% and waist circumference=24%), and all were highly correlated (online supplementary figure S1). In the first stage, from single-trait GWAS, we retained 606 SNPs with suggestive association ($p<1\times10^{-5}$) (see figure 2). None of them reached the genome-wide significance threshold. In the second stage, we analysed those 476 SNPs that intersected with the UK Biobank cohort dataset. Multitrait meta-analysis identified 111 SNPs in 27 independent loci with $p<5\times10^{-9}$ (online Supplementary file 7). Table 3 shows the SNPs with the highest significance for each independent *loci* and the univariate summary statistics of the anthropometric traits in both cohorts.

We estimated the covariance matrix (Σ) for each dataset (GCAT, UK Biobank and GCAT +UK Biobank). Then, as described in the Materials and methods section, we selected those independent SNPs with |z scores|<1.96, resulting in 765 646, 630 890 and 535 860 being considered for the Σ estimation. Eigenvalues of Σ showed d=1.36, 1.4 and 2.72 values. Covariance matrices were similar in both GCAT and UK Biobank (online supplementary tables S4 and S5). One degree of freedom (GCAT and UK Biobank) and three (GCAT +UK Biobank) of the ^{2} distribution were considered for multitrait analysis. We identified 27 independent multitrait loci associated in GCAT and UK Biobank (table 3). We intersected these SNPs with the GWAS Catalog, and we found that 5 SNPs had previously been reported in multiple GWAS, 16 loci were reported considering a ±250 000 base pair window from the identified SNP and 6 were new loci involving the following genes/SNPs: *MAD1L1* (rs62444886, $p=2.3\times10^{-15}$), *PIK3R1* (rs12657050, $p=2.8\times10^{-13}$; rs695166, $p=8.4\times10^{-15}$), *ZRANB2-AS2* (rs11205277, $p=1.4\times10^{-9}$), *EPHA7* (rs143547391, $p=6.5\times10^{-10}$), *CACUL1* (rs12414412, $p=4\times10^{-13}$) and *MAP3K9* (rs7151024, $p=5.7\times10^{-10}$). Regarding *DPYD*, *DPYD-IT1* (rs140281723), *GABRG3-AS1* and *GABRG3* (rs184405367) genes/SNPs, we did not replicate association in UK Biobank samples (UKmulti p=0.035 and 1, respectively). The risk allele, frequency and functional annotation using the Variant Effect Predictor tool[37] of identified variants are shown in online Supplementary file 9.

Polygenic risk score

The skin phototype association analysis identified five loci accounting for a high predictive value (PRS of 15.6%) suggesting few main genes (oligogenic architecture) contributing to the phenotype (online supplementary figure S2). However, for anthropometric traits, 27 loci were identified in our cohort but with a lower PRS (2.3%) suggesting a polygenic architecture with multiple genes and a high environmental impact. The newly identified loci only increased PRS slightly over the corresponding single-trait analysis (2.2% to 2.5%, 2.3% to 3.3%, 2.2% to 3.5%, 2.5% to 3.7% and 1.5% to 2.6% for height, weight, BMI and hip and waist circumference, respectively) pointing towards the multitrait approach as an effective screening strategy to identify new biomarkers.

DISCUSSION

Dissecting the architecture of common diseases should incorporate multitrait approaches to understand the phenome and its genetic aetiology, including pleiotropy and the co-occurrence of multiple morbidities, correlated traits and the diseasome as targets for genomic analysis.[38] In this study, we used the GCAT study, a South-European Mediterranean population prospective

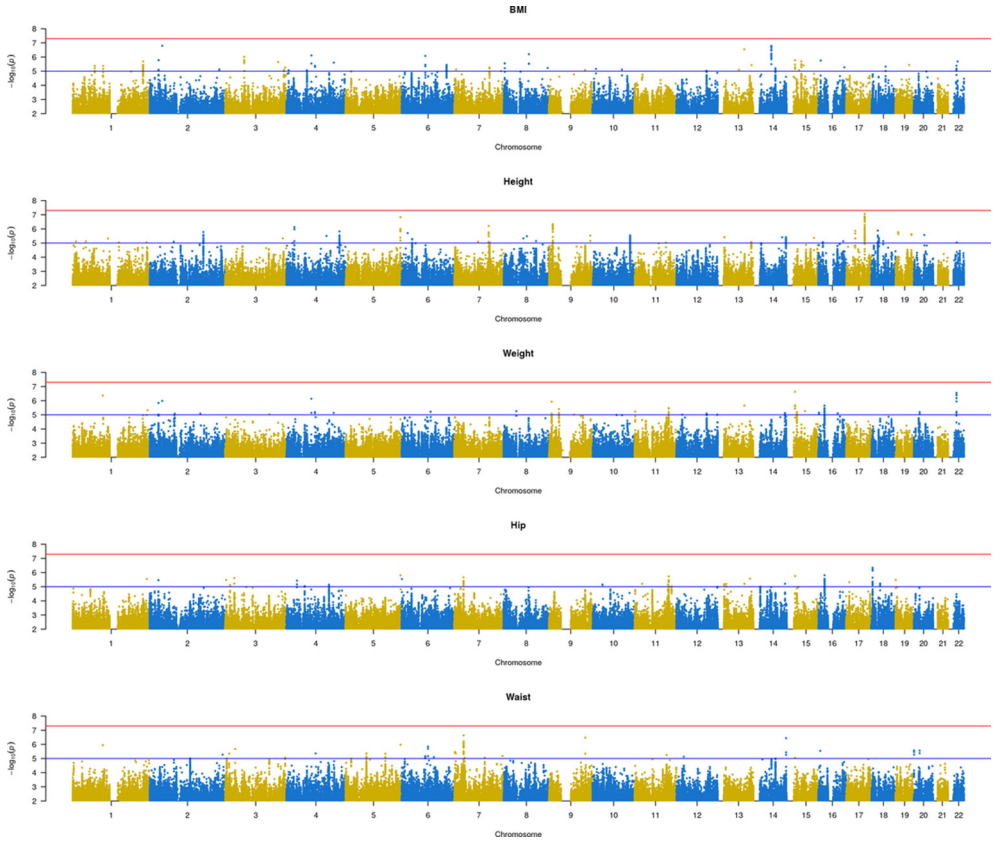

Figure 2 Manhattan plot of the anthropometric traits (BMI, height, weight and hip and waist circumference) from the GCAT. BMI, body mass index.

cohort to analyse the phenotypic variation attributable to genotype variability for 205 selected human traits (including diseases as well as biological, anthropometric and social features). Our results show that by considering genetic covariance matrices for interrelated traits, we increased the number of detected *loci* from six new *loci* for anthropometric traits, pointing to multitrait analysis as an effective strategy to gain statistical power to identify genetic association.

The relative importance of genetic and non-genetic factors varies across populations. Moreover, this is not constant in a population and changes with age.[16] Here, we have reported heritability estimates on an adult population based on SNP data. In the present study, h^2_{SNP} values move in a wide range from 18% to 77%, being anthropometric traits (height) and skin colour-related traits (Fitzpatrick's phototype score) the traits with the highest genetic determination. In our cohort, heritability of anthropometric traits, such as height and BMI, was likely estimated as a maximum, with negligible missed heritability when comparing with other reported estimates in similar populations[39] and in the same way being the observed genetic variance only a small part of their complete variance (around 3%). In the case of skin colour-related traits, the portion of the explained variance was larger, in accordance with a less complex polygenic nature of this trait, and fewer genes baring stronger predictive value (*IRF4*, *HERC2*, *OCA2*, *MC1R* and *SLC45A2*) (PRS=15.6%). The variants identified in these loci associated with skin colour-related traits are functional and have been reported elsewhere in several studies. These differences in heritability and prediction values indicate a different genomic architecture, suggesting an exposure variation, the exposome,[3] as a main actor for many polygenic traits. Higher estimates in self-perceived health heritability, and probably some other reported traits such as 'smoking_habits', 'smoking_packs', or 'sadness' (item from the Mental-Health Inventory 5-item questionnaire), reflect a pleiotropic effect[40] with multiple associated loci. In this sense, a recent meta-analysis on subjective well-being revealed new loci accounting for a polygenic model of well-being status.[41]

Single-trait GWAS analysis identified a number of genetic variants associated with skin colour-related traits (online supplementary figure S3) and other complex traits (heart rate, hyperlipidaemia or male pattern baldness); whereas failed to identify specific variants associated with any single anthropometric trait (at the $p<5\times10^{-8}$ threshold cut-off). However, we should observe that gender differences were not considered in this analysis even though it has been shown that genetic effects have a gender bias.[42] Applying multitrait analyses of anthropometric traits, we identified 27 loci, six of which had not been reported previously; *CALCUL1*, *ZRANB2-AS2*, *MAD1L1*, *EPHA7*, *PIK3R1* and *MAP3K9*. Owing to LD and the occurrence of all identified variants in non-coding regions (see online Supplementary file 9), we cannot be certain about the genes involved. Two out of six of the identified associated variants, in *CALCUL1* and *MAP3K9*, are putative expression quantitative trait loci (eQTL) (see the URLs section). Three of the variants (*ZRANB2-AS2*chr1:71702511, *EPHA7*chr6:94075927 and *MAP3K9*chr14:71268446) are specific of the GCAT sample ($p<5\times10^{-9}$) (online Supplementary files 10,11, S,12) probably due to genetic background differences between populations (ie, LD patterns) or as an expression of a particular genetic contribution of the Mediterranean populations to these polygenic traits. Identified variants implicate genes with diverse functions, involved in several pathways and processes. Some of them are involved in growth, developmental or metabolic processes.

Table 3 Loci associated with anthropometric traits in GCAT and UK Biobank cohorts

				Single-trait analysis														Multitrait analysis	
				Weight (kg)			Height (cm)			BMI (kg/m2)			Waist circumference (cm)			Hip circumference (cm)				
Loci*	Chr:position†	SNP	Cohort	β	SE	P values	β	SE	P values	β	SE	P values	β	SE	P values	β	SE	P values	P values	GWAS Catalog‡
SF3B4, SV2A	1:149892872	rs11205277	GCAT	−0.064	0.093	0.49	0.55	0.12	9.1×10^{-6}	0.29	0.27	0.27	0.2	0.24	0.4	0.2	0.19	0.29	0.00092	Reported SNP
			UK Biobank	−0.0017	0.0024	0.47	0.034	0.0017	1.1×10^{-85}	0.017	0.0021	4.3×10^{-16}	0.009	0.0022	3.3×10^{-5}	0.019	0.0024	5.1×10^{-15}	3.8×10^{-53}	
			GCAT-UK Biobank																1.5×10^{-53}	
ZRANB2-AS2	1:71702511	rs115213730	GCAT	1.7	0.37	4.2×10^{-6}	−0.77	0.49	0.12	3.8	1.1	0.00033	3.8	0.95	6.5×10^{-5}	2.8	0.76	0.0002	1.6×10^{-8}	New loci
			UK Biobank	0.021	0.0068	0.0026	0.0044	0.0049	0.37	0.019	0.006	0.0016	0.017	0.0061	0.0063	0.016	0.0068	0.017	0.00015	
			GCAT-UK Biobank																1.4×10^{-9}	
DPYD, DPYD-IT1	1:97884058	rs140281723	GCAT	1.9	0.45	2×10^{-5}	1.1	0.6	0.071	6.5	1.3	4.3×10^{-7}	5.6	1.2	1.1×10^{-6}	3.8	0.93	4.7×10^{-5}	8.6×10^{-11}	No association
			UK Biobank	0.011	0.0086	0.21	−0.015	0.0062	0.012	0.0004	0.0076	0.96	0.013	0.0077	0.1	0.0023	0.0086	0.79	0.035	
			GCAT-UK Biobank																1.9×10^{-9}	
PRELID1, RAB24, MXD3	5:176735612	rs111251222	GCAT	0.06	0.12	0.62	0.78	0.16	1×10^{-6}	0.84	0.34	0.014	0.33	0.31	0.28	0.13	0.25	0.59	0.0001	Reported loci
			UK Biobank	−0.0084	0.0028	0.0024	0.034	0.002	2.7×10^{-67}	0.012	0.0024	1.2×10^{-6}	0.0084	0.0025	0.00063	0.0007	0.0028	0.8	3.5×10^{-35}	
			GCAT-UK Biobank																2.2×10^{-36}	
LMAN2, AC146507.1	5:176772736	rs4976686	GCAT	0.049	0.1	0.63	0.71	0.13	1.5×10^{-7}	0.79	0.29	0.0071	0.29	0.26	0.26	0.14	0.21	0.5	2.8×10^{-5}	Reported loci
			UK Biobank	−0.0056	0.0026	0.034	0.028	0.0019	9.8×10^{-49}	0.01	0.0023	8.6×10^{-6}	0.0042	0.0023	0.073	0.0021	0.0026	0.42	2.7×10^{-25}	
			GCAT-UK Biobank																4.9×10^{-27}	
PIK3R1	5:67579576	rs12657050	GCAT	−0.29	0.11	0.0083	−0.33	0.15	0.022	−1	0.32	0.0011	−1.3	0.28	8.9×10^{-6}	−0.59	0.23	0.009	1.1×10^{-6}	Unreported locus
			UK Biobank	−0.0043	0.0028	0.13	−0.014	0.002	4.9×10^{-12}	−0.011	0.0025	7.7×10^{-6}	−0.009	0.0025	0.00035	−0.0067	0.0028	0.017	4.1×10^{-10}	
			GCAT-UK Biobank																2.8×10^{-13}	
	5:67604628	rs695166	GCAT	−0.27	0.1	0.011	−0.41	0.14	0.0029	−1.1	0.3	0.00043	−1.2	0.27	7.4×10^{-6}	−0.69	0.21	0.0012	1.2×10^{-7}	
			UK Biobank	−0.0031	0.0027	0.24	−0.015	0.0019	2.3×10^{-14}	−0.01	0.0024	1×10^{-5}	−0.0083	0.0024	0.00051	−0.0054	0.0027	0.044	8.7×10^{-11}	
			GCAT-UK Biobank																8.4×10^{-15}	
GMDS	6:1944345	rs62391629	GCAT	0.54	0.17	0.0017	0.52	0.23	0.023	2	0.49	7.9×10^{-5}	1.4	0.44	0.0022	1.7	0.35	2.9×10^{-6}	4.9×10^{-8}	Reported locus
			UK Biobank	0.013	0.0051	0.0085	0.0085	0.0036	0.02	0.016	0.0045	0.00045	0.014	0.0045	0.0014	0.017	0.0051	0.00068	6.4×10^{-6}	
			GCAT-UK Biobank																2×10^{-10}	
ID4, AL022068.1, –	6:19839415	rs41271299	GCAT	−0.11	0.22	0.62	1.4	0.3	2×10^{-6}	0.96	0.64	0.13	0.081	0.58	0.89	0.35	0.46	0.45	0.00048	Reported loci
			UK Biobank	−0.0032	0.0054	0.55	0.094	0.0039	1.4×10^{-129}	0.049	0.0048	1.8×10^{-24}	0.027	0.0049	1.9×10^{-8}	0.041	0.0054	6.8×10^{-14}	2.7×10^{-77}	
			GCAT-UK Biobank																4.5×10^{-78}	
GRM4, HMGA1	6:34199092	rs2780226	GCAT	0.029	0.15	0.85	0.91	0.2	5.6×10^{-6}	0.89	0.44	0.042	0.14	0.39	0.73	0.34	0.31	0.28	0.00043	Reported SNP
			UK Biobank	0.00064	0.0042	0.88	0.067	0.003	7.4×10^{-109}	0.037	0.0037	7.9×10^{-23}	0.033	0.0038	8.3×10–19	0.02	0.0042	2.7×10^{-6}	1.6×10^{-68}	
			GCAT-UK Biobank																2.7×10^{-69}	
HMGA1, SMIM29, AL354740.1	6:34214322	rs1150781	GCAT	0.029	0.15	0.85	0.9	0.2	9.8×10^{-6}	0.86	0.44	0.049	0.08	0.39	0.84	0.37	0.31	0.24	0.00059	Reported SNP
			UK Biobank	0.0023	0.0042	0.59	0.066	0.003	2.4×10^{-106}	0.037	0.0037	5.7×10^{-24}	0.034	0.0038	7.2×10^{-20}	0.021	0.0042	5.1×10^{-7}	1×10^{-68}	
			GCAT-UK Biobank																2.3×10^{-69}	
EPHA7	6:94075927	rs143547391	GCAT	1.2	0.4	0.0019	1.7	0.53	0.0013	5.2	1.1	6×10^{-6}	3.5	1	0.00063	2.8	0.82	0.00082	3.4×10^{-8}	New locus
			UK Biobank	−0.022	0.0088	0.014	−0.01	0.0063	0.1	−0.025	0.0078	0.0014	−0.027	0.0079	0.00077	−0.027	0.0088	0.0025	3.3×10^{-5}	
			GCAT-UK Biobank																6.5×10^{-10}	
AOC1, KCNH	7:150599205	rs10216051	GCAT	−0.44	0.099	9.9×10^{-6}	0.089	0.13	0.5	−1	0.28	0.00022	−1.1	0.25	2.6×10^{-5}	−0.84	0.2	3.5×10^{-5}	9.5×10^{-9}	R eported loci
			UK Biobank	0.0073	0.0025	0.0039	0.012	0.0018	2.7×10^{-11}	0.012	0.0022	2.5×10^{-8}	0.0082	0.0023	0.00029	0.0087	0.0025	0.00062	4.2×10^{-12}	
			GCAT-UK Biobank																4.8×10^{-17}	
MAD1L1, –	7:2068330	rs62444886	GCAT	−0.76	0.18	3.9×10^{-5}	−0.23	0.25	0.34	−2.3	0.53	1.6×10	−2.2	0.47	4×10^{-6}	−1.6	0.38	3.6×10^{-5}	1.9×10^{-9}	Unreported locus
			UK Biobank	−0.026	0.0052	7.3×10^{-7}	0.0051	0.0037	0.17	−0.019	0.0046	−5 4.7×10^{-5}	−0.021	0.0047	1×10^{-5}	−0.025	0.0052	2.5×10^{-6}	9.9×10^{-10}	
			GCAT-UK Biobank																2.3×10^{-15}	

Continued

Table 3 Continued

				Single-trait analysis														Multitrait analysis	
				Weight (kg)			Height (cm)			BMI (kg/m2)			Waist circumference (cm)			Hip circumference (cm)				
Loci*	Chr:position†	SNP	Cohort	β	SE	P values	β	SE	P values	β	SE	P values	β	SE	P values	β	SE	P values	P values	GWAS Catalog‡
FUBP3	9:133482006	rs11792294	GCAT	−0.075	0.099	0.45	−0.59	0.13	7.1×10^{-6}	−0.69	0.29	0.016	−0.079	0.26	0.76	−0.2	0.21	0.32	0.00029	Reported locus
			UK Biobank	0.0021	0.0025	0.39	−0.02	0.0018	2×10^{-29}	−0.0092	0.0022	2.5×10^{-5}	−0.0024	0.0022	0.27	−0.0032	0.0025	0.2	5.5×10^{-16}	
			GCAT-UK Biobank																6.3×10^{-17}	
CALCUL1	10:120465796	rs12414412	GCAT	0.054	0.17	0.75	1	0.23	5.9×10^{-6}	1.1	0.5	0.029	0.15	0.44	0.73	0.45	0.35	0.2	0.00033	Unreported locus
			UK Biobank	0.022	0.0043	4.3×10^{-7}	0.0074	0.0031	0.017	0.022	0.038	6.4×10^{-9}	0.015	0.0039	6.6×10^{-5}	0.023	0.0043	1×10^{-7}	3.7×10^{-12}	
			GCAT-UK Biobank																4×10^{-13}	
INS-IGF2, IGF2-AS, –	11:2172830	rs7948458	GCAT	−0.37	0.11	0.00044	−0.42	0.14	0.0027	−1.4	0.31	5.8×10^{-6}	−0.91	0.27	0.00091	−0.94	0.22	1.7×10^{-5}	4.5×10^{-9}	Reported loci
			UK Biobank	−0.002	0.0031	0.5	−0.022	0.0022	4.2×10^{-24}	-0.014	0.0027	3.6×10^{-7}	−0.0044	0.0027	0.1	−0.014	0.0031	5.8×10^{-6}	2.3×10^{-16}	
			GCAT-UK Biobank																1.5×10^{-21}	
MAP3K9	14:71268446	rs7151024	GCAT	−0.49	0.11	6.3×10^{-6}	0.34	0.14	0.017	−1	0.31	0.0013	−1.1	0.28	4.4×10^{-5}	−0.93	0.22	3.1×10^{-5}	5.7×10^{-9}	New locus
			UK Biobank	−0.0097	0.0029	0.00073	0.0054	0.0021	0.0084	−0.0051	0.0025	0.042	−0.0065	0.0026	0.012	−0.0058	0.0029	0.044	0.00015	
			GCAT-UK Biobank																5.7×10^{-10}	
GABRG3-AS1, GABRG3	15:27398499	rs184405367	GCAT	1.5	0.32	1.7×10^{-6}	0.67	0.42	0.11	4.7	0.91	2.4×10^{-7}	3.6	0.81	9.1×10^{-6}	3.1	0.65	1.8×10^{-6}	1.3×10^{-11}	No association
			UK Biobank	−0.0027	0.016	0.86	−0.0015	0.011	0.89	−0.0041	0.014	0.77	−0.005	0.014	0.72	0.0024	0.016	0.88	1	
			GCAT-UK Biobank																3.5×10^{-9}	
SEMA6D	15:47923520	rs10220751	GCAT	−0.44	0.093	2×10^{-6}	0.21	0.12	0.086	−1	0.27	0.00015	−0.8	0.24	0.00086	−0.59	0.19	0.0022	7.1×10^{-8}	Reported locus
			UK Biobank	−0.011	0.0024	2.7×10^{-6}	0.0043	0.0018	0.014	-0.0071	0.0022	0.001	−0.0045	0.0022	0.039	−0.011	0.0024	1.1×10^{-5}	1.6×10^{-7}	
			GCAT-UK Biobank																8×10^{-12}	
GPRC5B-GPR139	16:19988852	rs9940317	GCAT	0.43	0.12	0.00033	0.42	0.16	0.0085	1.6	0.35	6.3×10^{-6}	1.4	0.31	1.2×10^{-5}	1.2	0.25	2.7×10^{-6}	3.7×10^{-10}	Reported loci
			UK Biobank	0.012	0.0029	3.8×10^{-5}	0.0068	0.0021	0.00096	0.013	0.0025	1.4×10^{-7}	0.0079	0.0026	0.0022	0.011	0.0029	0.00012	3×10^{-9}	
			GCAT-UK Biobank																1.6×10^{-15}	
GPR139	16:20046115	rs2045457	GCAT	0.38	0.1	0.00016	0.24	0.13	0.069	1.3	0.29	1.4×10^{-5}	1.2	0.26	3.3×10^{-6}	0.93	0.21	8.6×10^{-6}	9×10^{-10}	Reported locus
			UK Biobank	0.013	0.0026	7.6×10^{-7}	0.0057	0.0019	0.0024	0.014	0.0023	2×10^{-9}	0.0068	0.0023	0.0038	0.011	0.0026	5.4×10^{-5}	1.1×10^{-10}	
			GCAT-UK Biobank																1.4×10^{-16}	
ECI1, AC009065.8	16:2296197	rs77407216	GCAT	−0.39	0.13	0.0036	−0.7	0.18	9.6×10^{-5}	−1.7	0.39	6.9×10^{-6}	−0.82	0.35	0.018	−0.97	0.28	0.00053	5.4×10^{-8}	Reported loci
			UK Biobank	−0.0021	0.0035	0.55	−0.016	0.0025	4.4×10^{-10}	−0.01	0.0031	0.0011	−0.0076	0.0031	0.016	−0.0068	0.0031	0.051	3.1×10^{-7}	
			GCAT-UK Biobank																1.2×10^{-11}	
ATAD5, AC130324.2	17:29165934	rs9890032	GCAT	−0.098	0.095	0.3	−0.61	0.13	1.4×10^{-6}	−0.83	0.27	0.0023	−0.49	0.24	0.044	−0.41	0.2	0.035	7×10^{-6}	Reported SNP
			UK Biobank	-6.7×10^{-5}	0.0025	0.98	−0.032	0.0018	1.8×10^{-71}	-0.017	0.0022	1.6×10^{-15}	−0.011	0.0022	2.2×10^{-6}	−0.013	0.0025	3.8×10^{-7}	1.4×10^{-43}	
			GCAT-UK Biobank																8.2×10^{-46}	
TBX2	17:59498052	rs7214743	GCAT	0.084	0.095	0.37	0.66	0.13	2×10^{-7}	0.82	0.27	0.0028	0.43	0.24	0.082	0.47	0.2	0.016	3×10^{-6}	Reported locus
			UK Biobank	−0.0093	0.0026	0.00027	0.034	0.0018	1.3×10^{-78}	0.011	0.0023	3.8×10^{-7}	0.0064	0.0023	0.0053	0.00026	0.0026	0.92	1.9×10^{-40}	
			GCAT-UK Biobank																5.1×10^{-43}	
CABLES1	18:20758310	rs34302357	GCAT	0.22	0.12	0.071	−0.75	0.16	2.9×10^{-6}	−0.13	0.35	0.7	0.21	0.31	0.49	−0.037	0.25	0.88	0.00048	Reported locus
			UK Biobank	−0.0024	0.0031	0.43	−0.042	0.0022	3.3×10^{-80}	−0.024	0.0027	3.8×10^{-19}	−0.018	0.0028	2.7×10^{-10}	−0.017	0.0031	3.7×10^{-8}	2.9×10^{-51}	
			GCAT-UK Biobank																6.4×10^{-52}	
RIOK3, Y RNA	18:21039393	rs9954741	GCAT	0.2	0.1	0.05	−0.6	0.14	8.7×10^{-6}	−0.025	0.29	0.93	0.14	0.26	0.58	−0.032	0.21	0.88	0.00076	Reported loci
			UK Biobank	−0.01	0.0025	7.1×10^{-5}	−0.013	0.0018	1.1×10^{-13}	−0.016	0.0022	2.2×10^{-12}	−0.015	0.0023	1.1×10^{-11}	−0.014	0.0025	2.8×10^{-8}	8.9×10^{-21}	
			GCAT-UK Biobank																2.7×10^{-21}	
ADAMTS10	19:8670147	rs62621197	GCAT	0.48	0.19	0.011	−1.2	0.25	2.3×10^{-6}	0.29	0.55	0.59	0.37	0.49	0.45	0.2	0.39	0.61	0.00016	Reported locus
			UK Biobank	0.014	0.0067	0.036	−0.11	0.0048	5.4×10^{-121}	−0.05	0.0059	2.6×10^{-17}	−0.036	0.006	2.4×10^{-9}	−0.042	0.0067	4×10^{-10}	1.9×10^{-69}	
			GCAT-UK Biobank																1.3×10^{-70}	

Continued

Table 3 Continued

Loci*	Chr:position†	SNP	Cohort	Single-trait analysis												Multitrait analysis	
				Weight (kg)			Height (cm)			BMI (kg/m2)			Waist circumference (cm)			Hip circumference (cm)				
				β	SE	P values	β	SE	P values	β	SE	P values	β	SE	P values	β	SE	P values	P values	GWAS Catalog‡
GDF5, GDF5OS	20:34025756	rs143384	GCAT	0.072	0.094	0.45	0.59	0.13	2.7×10^{-6}	0.74	0.27	0.0065	0.31	0.24	0.21	0.55	0.19	0.0052	1.4×10^{-5}	Reported SNP
			UK Biobank	−0.0014	0.0024	0.58	0.064	0.0018	8.8×10^{-282}	0.033	0.0022	1.9×10^{-53}	0.0071	0.0022	0.0013	0.028	0.0024	1.6×10^{-30}	8.3×10^{-168}	
			GCAT-UK Biobank																1.3×10^{-170}	
HORMAN, LIF	22:30610546	rs9608851	GCAT	0.45	0.095	2×10^{-6}	−0.029	0.13	0.022	1	0.27	0.00027	0.82	0.24	0.00086	0.66	0.2	0.00075	3.2×10^{-8}	Reported loci
			UK Biobank	0.005	0.0024	0.038	0.0062	0.0017	0.00037	0.0077	0.0021	0.00032	0.0058	0.0022	0.007	0.0054	0.0024	0.025	1.7×10^{-5}	
			GCAT-UK Biobank																3.3×10^{-10}	

*Loci, a locus was considered as the ±250000 base pair window flanking the identified SNP.
†Chr:position, coordinates on hg19.
‡GWAS Catalog traits, data from GWAS Catalog database (release V1.0, e90, 27 September 2017).
Single-trait and multi trait results are presented. Concordant significant results are marked in violet
BMI, body mass index; GWAS, genome-wide association studies.

MAP3K9, mitogen-activated protein kinase 9, has been associated to some rare cancers (ie, retroperitoneum carcinoma and retroperitoneum neuroblastoma), and GWAS studies have identified variants associated with reasoning ability.[43] Based on GTEx database (see URL section) we identified *rs7151024* as an eQTL, expressed in subcutaneous adipose tissue ($p=1.4\times10^{-8}$, eQTL effect size (es)=−0.38) that may affect fat distribution and anthropometric traits. *ZRANB2-AS2* is a non-coding RNA, and GWAS studies have identified variants in *ZRANB2-AS2* associated with facial morphology,[44] and also with general cognitive function,[45] traits which are genetically correlated with a wide range of physical variables. *EPHA7* belongs to the ephrin receptor subfamily of protein-tyrosine kinase, implicated in mediating developmental events, particularly in the nervous system. *EPHA7* has been implicated in neurodevelopment processes[46] as well being as a tumour suppressor gene in cancer.[47]*CACUL1*, CDK2-associated cullin domain 1, is a cell cycle-dependent kinase binding protein capable of promoting cell progression. In the GWAS Catalog, any of the anthropometric traits analysed here have been associated with variants in *CACUL1* (online Supplementary file 13). However, the associated rs12414412, reported as an eQTL expressed in skeletal muscle ($p=1.4\times10^{-7}$, eQTL es=−0.31), may affect body constitution. *CACUL1* suppresses androgen receptor (AR) transcriptional activity, impairing LSD-mediated activation of the AR,[48] whose genetic variation is associated with longitudinal height in young boys.[49]*MAD1L1,* mitotic arrest deficient 1-like protein 1, is a component of the mitotic spindle-assembly checkpoint, and some cancers (prostate and gastric) have been associated to *MAD1L1* dysfunction.[50] Our study identified BMI, weight and hip and waist circumference single-trait association ($p<10^{-5}$) with the intronic variant *rs62444886* in the *MAD1L1* locus, as well as a significant multitrait association in meta-analysis (table 3, online Supplementary file 14). GWAS analysis identified *MAD1L1* as a susceptibility gene for bipolar disorder and schizophrenia, involved in reward system functions in healthy adults,[51] but until now, no other study has identified it as a genetic contributor to weight. The higher prevalence of obesity and related disorders such as diabetes in schizophrenia patients could reflect a possible underlying common genetic contribution. In this sense, we observed also GWAS significant signals in *INS-IGF2* (GCAT-UKmulti $p=1.5\times10^{-21}$), an analogue of the *INS* gene (previously associated with diabetes type I and type II disorders).[52] Additionally, epigenome-wide association studies in adults[53] and children[54] support a role for *MAD1L1* in BMI–methylation association, with differentially methylated CpG patterns in CD4+ and CD8+ T cells between obese and non-obese women. *PIK3R1,* phosphoinositide-3-kinase regulatory subunit 1, plays a role in the metabolic actions of insulin, and a mutation in this gene has been associated with insulin resistance. Moreover, common variants are associated with lower body fat percentage as well as the control of peripheral adipose tissue mobilisation.[55] Genetic variation in the GWAS Catalog is also associated with cartilage thickness[56] and mineral bone density,[57] both related to anthropometric traits. Diseases associated with *PIK3R1* include SHORT syndrome,[58] characterised by individuals with short stature and a restricted intrauterine growth, in addition to multiple anomalies. Our study identified the intronic variant (*rs695166*) associated with waist circumference association in single-trait analysis ($p<10^{-6}$), but not in the UKdataset, which associates with height ($p=2.3\times10^{-14}$). However, analysis of the UKBiobank data supported a similar peak profile overlapping the

gene region (see online Supplementary file 12) and multitrait analysis association (GCAT-UK multi $p=8.4\times10^{-15}$) (table 3).

Multiple approaches for multitrait analysis using GWAS data have been successfully applied in the research of genetically complex conditions using raw data or summary-level data statistics. Using raw data, Ferreira and Purcell[11] used a test based on the Wilk's lambda derived from a canonical correlation analysis. Korte *et al*[13] implemented a mixed-model approach accounting for correlation structure and the kinship relatedness matrix. O'Reilly *et al*[14] proposed an inverted regression model for each SNP as the response and all the traits as covariates. Regarding the use of GWAS summary-level data statistics, Cotsapas *et al*[10] developed a statistic for cross-phenotype analysis based on an asymptotic ^{2} distribution derived from p values of the SNP associations. Zhu *et al*[15] implemented CPASSOC that accounts for the genetic correlation structure of the traits and the sample size for each cohort. Kim *et al*[12] proposed an adaptive association test for multiple traits that uses Monte Carlo simulations to approximate its null distribution. Recently, Bayes factor approaches[59] have been proposed for studying multitrait genetic associations. Here, for meta-analysis purposes, we chose the multitrait analysis described by Yang and Wang.[31] This test, based on the ^{2} distribution with 'd' df, depends on the genetic covariance structure of the traits and considers the distribution of the sum square of the z scores which is insensitive to the heterogeneous effect of the SNP. Nevertheless, this approach doesn't allow allele effect estimation. In this sense, maximum likelihood methods have been recently proposed to deal with this limitation[41] by accounting for different measures of the same phenotypic trait with different levels of heritability.

In complex diseases research, MRPs are the common observation in genome-wide association analysis of large cohorts, and over simplification of extreme phenotypes or the use of standardised phenotypes for meta-analysis reduces the power to detect the underlying genetic contribution to complex traits. As an alternative, multitrait analyses help to detect additional loci that are missing by applying a conventional meta-analysis. Our results highlight the potential value of data-driven multivariate phenotyping for genetic studies in large complex cohorts.

Author affiliations
[1]GenomesForLife-GCAT Lab Group, Program of Predictive and Personalized Medicine of Cancer (PMPPC), Germans Trias i Pujol Research Institute (IGTP), Crta. de Can Ruti, Badalona, Catalunya, Spain
[2]Unit of Biomarkers and Susceptibility, Cancer Prevention and Control Program, Catalan Institute of Oncology (ICO), IDIBELL and CIBERESP, Barcelona, Spain
[3]High Content Genomics and Bioinformatics Unit, Program of Predictive and Personalized Medicine of Cancer (PMPPC), Germans Trias i Pujol Research Institute (IGTP), Badalona, Catalunya, Spain
[4]Life Sciences - Computational Genomics, Barcelona Supercomputing Center (BSC-CNS), Joint BSC-CRG-IRB Research Program in Computational Biology, Barcelona, Spain
[5]Programs in Metabolism and Medical & Population Genetics, Broad Institute of Harvard and MIT, Cambridge, Massachusetts, US
[6]Diabetes Unit and Center for Human Genetic Research, Massachusetts General Hospital, Boston, Massachusetts, US
[7]Blood Division, Banc de Sang i Teixits, Barcelona, Spain
[8]Cancer Genetics and Epigenetics Group, Program of Predictive and Personalized Medicine of Cancer (PMPPC), Germans Trias i Pujol Research Institute (IGTP), Badalona, Catalunya, Spain
[9]ICREA, Catalan Institution for Research and Advanced Studies, Barcelona, Catalunya, Spain
[10]Department of Clinical Sciences, Faculty of Medicine, University of Barcelona, Barcelona, Spain

Correction notice This article has been corrected since it was published online first. JMM has been added to the authors list and to the 'Contributors' section.

Acknowledgements The authors thank all the GCAT participants and all BST members for generously helping with this research.

Contributors All authors contributed to the feedback of the manuscript and played an important role in implementing the study. IG-F, MP, VM and RdC conceived the study. IG-F and RdC planned the study. LP coordinated the cohort recruitment. AC, JV and XD prepared the samples. MO-S and XD curated the epidemiological data variables. DP, RP, LR, SA and LS conducted the genotyping. IG-F, DP and LS analysed the clustering analysis. IG-F, MG-M, JMM and DT conducted the imputation analysis. IG-F and RdC conducted and supervised the genetic analysis. IG-F, MO-S and RdC wrote the manuscript. RdC submitted and supervised the study.

Funding This work was supported in part by the Spanish Ministerio de Economía y Competitividad (MINECO) project ADE 10/00026, by the Catalan Departament de Salut and by the Departament d'Empresa i Coneixement de la Generalitat de Catalunya, the Agència de Gestió d'Estudis Universitaris i de Recerca (AGAUR) (SGR 1269, SGR 1589 and SGR 647). RdC is the recipient of a Ramon y Cajal grant (RYC-2011-07822). The Project GCAT is coordinated by the Germans Trias i Pujol Research Institute (IGTP), in collaboration with the Catalan Institute of Oncology (ICO), and in partnership with the Blood and Tissue Bank of Catalonia (BST). IGTP is part of the CERCA Programme/Generalitat de Catalunya.

Competing interests None declared.

Patient consent Obtained.

Ethics approval http://www.ceicgermanstrias.cat/.

Provenance and peer review Not commissioned; externally peer reviewed.

REFERENCES

1 Eurostat Statistics Explained. Mortality and life expectancy statistics, 2016. http://ec.europa.eu/eurostat/statistics-explained/index.php/Mortality_and_life_expectancy_statistics
2 Dawber TR, Meadors GF, Moore FE. Epidemiological approaches to heart disease: the Framingham Study. *Am J Public Health Nations Health* 1951;41:279–86.
3 Riboli E, Hunt KJ, Slimani N, Ferrari P, Norat T, Fahey M, Charrondière UR, Hémon B, Casagrande C, Vignat J, Overvad K, Tjønneland A, Clavel-Chapelon F, Thiébaut A, Wahrendorf J, Boeing H, Trichopoulos D, Trichopoulou A, Vineis P, Palli D, Bueno-De-Mesquita HB, Peeters PH, Lund E, Engeset D, González CA, Barricarte A, Berglund G, Hallmans G, Day NE, Key TJ, Kaaks R, Saracci R. European Prospective Investigation into Cancer and Nutrition (EPIC): study populations and data collection. *Public Health Nutr* 2002;5:1113–24.
4 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.
5 Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorff LA, Hunter DJ, McCarthy MI, Ramos EM, Cardon LR, Chakravarti A, Cho JH, Guttmacher AE, Kong A, Kruglyak L, Mardis E, Rotimi CN, Slatkin M, Valle D, Whittemore AS, Boehnke M, Clark AG, Eichler EE, Gibson G, Haines JL, Mackay TF, McCarroll SA, Visscher PM. Finding the missing heritability of complex diseases. *Nature* 2009;461:747–53.
6 Visscher PM, Wray NR, Zhang Q, Sklar P, McCarthy MI, Brown MA, Yang J. 10 Years of GWAS Discovery: Biology, Function, and Translation. *Am J Hum Genet* 2017;101:5–22.
7 Boyle EA, Li YI, Pritchard JK. An Expanded View of Complex Traits: From Polygenic to Omnigenic. *Cell* 2017;169:1177–86.
8 Chakravarti A, Turner TN. Revealing rate-limiting steps in complex disease biology: The crucial importance of studying rare, extreme-phenotype families. *Bioessays* 2016;38:578–86.
9 Freimer N, Sabatti C. The human phenome project. *Nat Genet* 2003;34:15–21.
10 Cotsapas C, Voight BF, Rossin E, Lage K, Neale BM, Wallace C, Abecasis GR, Barrett JC, Behrens T, Cho J, De Jager PL, Elder JT, Graham RR, Gregersen P, Klareskog L, Siminovitch KA, van Heel DA, Wijmenga C, Worthington J, Todd JA, Hafler DA, Rich SS, Daly MJ. FOCiS Network of Consortia. Pervasive sharing of genetic effects in autoimmune disease. *PLoS Genet* 2011;7:e1002254.
11 Ferreira MAR, Purcell SM. A multivariate test of association. *Bioinformatics* 2009;25:132–3.
12 Kim J, Bai Y, Pan W. An Adaptive Association Test for Multiple Phenotypes with GWAS Summary Statistics. *Genet Epidemiol* 2015;39:651–63.
13 Korte A, Vilhjálmsson BJ, Segura V, Platt A, Long Q, Nordborg M. A mixed-model approach for genome-wide association studies of correlated traits in structured populations. *Nat Genet* 2012;44:1066–71.

14 O'Reilly PF, Hoggart CJ, Pomyen Y, Calboli FC, Elliott P, Jarvelin MR, Coin LJ. MultiPhen: joint model of multiple phenotypes can increase discovery in GWAS. *PLoS One* 2012;7:e34861.
15 Zhu X, Feng T, Tayo BO, Liang J, Young JH, Franceschini N, Smith JA, Yanek LR, Sun YV, Edwards TL, Chen W, Nalls M, Fox E, Sale M, Bottinger E, Rotimi C, Liu Y, McKnight B, Liu K, Arnett DK, Chakravati A, Cooper RS, Redline S; COGENT BP Consortium. Meta-analysis of correlated traits via summary statistics from GWASs with an application in hypertension. *Am J Hum Genet* 2015;96:21–36.
16 Ge T, Chen CY, Neale BM, Sabuncu MR, Smoller JW. Phenome-wide heritability analysis of the UK Biobank. *PLoS Genet* 2017;13:e1006711.
17 Muñoz M, Pong-Wong R, Canela-Xandri O, Rawlik K, Haley CS, Tenesa A. Evaluating the contribution of genetics and familial shared environment to common disease using the UK Biobank. *Nat Genet* 2016;48:980–3.
18 Obón-Santacana M, Vilardell M, Carreras A, Duran X, Velasco J, Galván-Femenía I, Alonso T, Puig L, Sumoy L, Duell EJ, Perucho M, Moreno V, de Cid R. GCAT|Genomes for life: a prospective cohort study of the genomes of Catalonia. *BMJ Open* 2018;8:e018324.
19 Liu Y, De A. Multiple Imputation by Fully Conditional Specification for Dealing with Missing Data in a Large Epidemiologic Study. *Int J Stat Med Res* 2015;4:287–95.
20 Howie B, Fuchsberger C, Stephens M, Marchini J, Abecasis GR. Fast and accurate genotype imputation in genome-wide association studies through pre-phasing. *Nat Genet* 2012;44:955–9.
21 Auton A, Brooks LD, Durbin RM, Garrison EP, Kang HM, Korbel JO, Marchini JL, McCarthy S, McVean GA, Abecasis GR; 1000 Genomes Project Consortium. A global reference for human genetic variation. *Nature* 2015;526:68–74.
22 Deelen P, Menelaou A, van Leeuwen EM, Kanterakis A, van Dijk F, Medina-Gomez C, Francioli LC, Hottenga JJ, Karssen LC, Estrada K, Kreiner-Møller E, Rivadeneira F, van Setten J, Gutierrez-Achury J, Westra HJ, Franke L, van Enckevort D, Dijkstra M, Byelas H, van Duijn CM, de Bakker PI, Wijmenga C, Swertz MA; Genome of Netherlands Consortium. Improved imputation quality of low-frequency and rare variants in European samples using the 'Genome of The Netherlands'. *Eur J Hum Genet* 2014;22:1321–6.
23 Huang J, Howie B, McCarthy S, Memari Y, Walter K, Min JL, Danecek P, Malerba G, Trabetti E, Zheng HF, Gambaro G, Richards JB, Durbin R, Timpson NJ, Marchini J, Soranzo N, Turki SA, Amuzu A, Anderson CA, Anney R, Antony D, Artigas MS, Ayub M, Bala S, Barrett JC, Barroso I, Beales P, Benn M, Bentham J, Bhattacharya S, Birney E, Blackwood D, Bobrow M, Bochukova E, Bolton PF, Bounds R, Boustred C, Breen G, Calissano M, Carss K, Casas JP, Chambers JC, Charlton R, Chatterjee K, Chen L, Ciampi A, Cirak S, Clapham P, Clement G, Coates G, Cocca M, Collier DA, Cosgrove C, Cox T, Craddock N, Crooks L, Curran S, Curtis D, Daly A, Inm D, Day-Williams A, Dedoussis G, Down T, Du Y, van DCM, Dunham I, Edkins S, Ekong R, Ellis P, Evans DM, Farooqi IS, Fitzpatrick DR, Flicek P, Floyd J, Foley AR, Franklin CS, Futema M, Gallagher L, Gasparini P, Gaunt TR, Geihs M, Geschwind D, Greenwood C, Griffin H, Grozeva D, Guo X, Guo X, Gurling H, Hart D, Hendricks AE, Holmans P, Huang L, Hubbard T, Humphries SE, Hurles ME, Hysi P, Iotchkova V, Isaacs A, Jackson DK, Jamshidi Y, Johnson J, Joyce C, Karczewski KJ, Kaye J, Keane T, Kemp JP, Kennedy K, Kent A, Keogh J, Khawaja F, Kleber ME, van KM, Kolb-Kokocinski A, Kooner JS, Lachance G, Langenberg C, Langford C, Lawson D, Lee I, van LEM, Lek M, Li R, Li Y, Liang J, Lin H, Liu R, Lönnqvist J, Lopes LR, Lopes M, Luan J, MacArthur DG, Mangino M, Marenne G, März W, Maslen J, Matchan A, Mathieson I, McGuffin P, McIntosh AM, McKechanie AG, McQuillin A, Metrustry S, Migone N, Mitchison HM, Moayyeri A, Morris J, Morris R, Muddyman D, Muntoni F, Nordestgaard BG, Northstone K, O'Donovan MC, O'Rahilly S, Onoufriadis A, Oualkacha K, Owen MJ, Palotie A, Panoutsopoulou K, Parker V, Parr JR, Paternoster L, Paunio T, Payne F, Payne SJ, Perry JRB, Pietilainen O, Plagnol V, Pollitt RC, Povey S, Quail MA, Quaye L, Raymond L, Rehnström K, Ridout CK, Ring S, Ritchie GRS, Roberts N, Robinson RL, Savage DB, Scambler P, Schiffels S, Schmidts M, Schoenmakers N, Scott RH, Scott RA, Semple RK, Serra E, Sharp SI, Shaw A, Shihab HA, Shin S-Y, Skuse D, Small KS, Smee C, Smith GD, Southam L, Spasic-Boskovic O, Spector TD, Clair DS, Pourcain BS, Stalker J, Stevens E, Sun J, Surdulescu G, Suvisaari J, Syrris P, Tachmazidou I, Taylor R, Tian J, Tobin MD, Toniolo D, Traglia M, Tybjaerg-Hansen A, Valdes AM, Vandersteen AM, Varbo A, Vijayarangakannan P, Visscher PM, Wain LV, Walters JTR, Wang G, Wang J, Wang Y, Ward K, Wheeler E, Whincup P, Whyte T, Williams HJ, Williamson KA, Wilson C, Wilson SG, Wong K, Xu C, Yang J, Zaza G, Zeggini E, Zhang F, Zhang P, Zhang W; UK10K Consortium. Improved imputation of low-frequency and rare variants using the UK10K haplotype reference panel. *Nat Commun* 2015;6:8111.
24 McCarthy S, Das S, Kretzschmar W, Delaneau O, Wood AR, Teumer A, Kang HM, Fuchsberger C, Danecek P, Sharp K, Luo Y, Sidore C, Kwong A, Timpson N, Koskinen S, Vrieze S, Scott LJ, Zhang H, Mahajan A, Veldink J, Peters U, Pato C, van Duijn CM, Gillies CE, Gandin I, Mezzavilla M, Gilly A, Cocca M, Traglia M, Angius A, Barrett JC, Boomsma D, Branham K, Breen G, Brummett CM, Busonero F, Campbell H, Chan A, Chen S, Chew E, Collins FS, Corbin LJ, Smith GD, Dedoussis G, Dorr M, Farmaki AE, Ferrucci L, Forer L, Fraser RM, Gabriel S, Levy S, Groop L, Harrison T, Hattersley A, Holmen OL, Hveem K, Kretzler M, Lee JC, McGue M, Meitinger T, Melzer D, Min JL, Mohlke KL, Vincent JB, Nauck M, Nickerson D, Palotie A, Pato M, Pirastu N, McInnis M, Richards JB, Sala C, Salomaa V, Schlessinger D, Schoenherr S, Slagboom PE, Small K, Spector T, Stambolian D, Tuke M, Tuomilehto J, Van den Berg LH, Van Rheenen W, Volker U, Wijmenga C, Toniolo D, Zeggini E, Gasparini P, Sampson MG, Wilson JF, Frayling T, de Bakker PI, Swertz MA, McCarroll S, Kooperberg C, Dekker A, Altshuler D, Willer C, Iacono W, Ripatti S, Soranzo N, Walter K, Swaroop A, Cucca F, Anderson CA, Myers RM, Boehnke M, McCarthy MI, Durbin R; Haplotype Reference Consortium. A reference panel of 64,976 haplotypes for genotype imputation. *Nat Genet* 2016;48:1279–83.
25 Yang J, Lee SH, Goddard ME, Visscher PM. GCTA: a tool for genome-wide complex trait analysis. *Am J Hum Genet* 2011;88:76–82.
26 Schweiger R, Fisher E, Rahmani E, Shenhav L, Rosset S, Halperin E. Using Stochastic Approximation Techniques to Efficiently Construct Confidence Intervals for Heritability: In. *Research in Computational Molecular Biology*. Cham: Springer, 2017:241–56.
27 Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience* 2015;4:7.
28 Dey R, Schmidt EM, Abecasis GR, Lee S. A Fast and Accurate Algorithm to Test for Binary Phenotypes and Its Application to PheWAS. *Am J Hum Genet* 2017;101:37–49.
29 Chavent M, Kuentz-Simonet V, Labenne A, Saracco J. *Multivariate analysis of mixed type data: The PCAmixdata R package*, 2014.
30 Sudlow C, Gallacher J, Allen N, Beral V, Burton P, Danesh J, Downey P, Elliott P, Green J, Landray M, Liu B, Matthews P, Ong G, Pell J, Silman A, Young A, Sprosen T, Peakman T, Collins R. UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS Med* 2015;12:e1001779.
31 Yang Q, Wang Y. Methods for Analyzing Multivariate Phenotypes in Genetic Association Studies. *J Probab Stat* 2012;2012:1–13.
32 Euesden J, Lewis CM, O'Reilly PF. PRSice: Polygenic Risk Score software. *Bioinformatics* 2015;31:1466–8.
33 McEvoy B, Beleza S, Shriver MD. The genetic architecture of normal variation in human pigmentation: an evolutionary perspective and model. *Hum Mol Genet* 2006;15:R176–81.
34 Liu F, Visser M, Duffy DL, Hysi PG, Jacobs LC, Lao O, Zhong K, Walsh S, Chaitanya L, Wollstein A, Zhu G, Montgomery GW, Henders AK, Mangino M, Glass D, Bataille V, Sturm RA, Rivadeneira F, Hofman A, van IJcken WF, Uitterlinden AG, Palstra RJ, Spector TD, Martin NG, Nijsten TE, Kayser M. Genetics of skin color variation in Europeans: genome-wide association studies with functional follow-up. *Hum Genet* 2015;134:823–35.
35 Robles-Espinoza CD, Roberts ND, Chen S, Leacy FP, Alexandrov LB, Pornputtapong N, Halaban R, Krauthammer M, Cui R, Timothy Bishop D, Adams DJ. Germline MC1R status influences somatic mutation burden in melanoma. *Nat Commun* 2016;7:12064.
36 Sturm RA. Skin colour and skin cancer - MC1R, the genetic link. *Melanoma Res* 2002;12:405–16.
37 McLaren W, Gil L, Hunt SE, Riat HS, Ritchie GR, Thormann A, Flicek P, Cunningham F. The Ensembl Variant Effect Predictor. *Genome Biol* 2016;17:122.
38 Wysocki K, Ritter L. Diseasome: an approach to understanding gene-disease interactions. *Annu Rev Nurs Res* 2011;29:55–72.
39 Yang J, Bakshi A, Zhu Z, Hemani G, Vinkhuyzen AA, Lee SH, Robinson MR, Perry JR, Nolte IM, van Vliet-Ostaptchouk JV, Snieder H, Esko T, Milani L, Mägi R, Metspalu A, Hamsten A, Magnusson PK, Pedersen NL, Ingelsson E, Soranzo N, Keller MC, Wray NR, Goddard ME, Visscher PM; LifeLines Cohort Study. Genetic variance estimation with imputed variants finds negligible missing heritability for human height and body mass index. *Nat Genet* 2015;47:1114–20.
40 Krapohl E, Rimfeld K, Shakeshaft NG, Trzaskowski M, McMillan A, Pingault JB, Asbury K, Harlaar N, Kovas Y, Dale PS, Plomin R. The high heritability of educational achievement reflects many genetically influenced traits, not just intelligence. *Proc Natl Acad Sci U S A* 2014;111:15273–8.
41 Turley P, Walters RK, Maghzian O, Okbay A, Lee JJ, Fontana MA, Nguyen-Viet TA, Wedow R, Zacher M, Furlotte NA, Magnusson P, Oskarsson S, Johannesson M, Visscher PM, Laibson D, Cesarini D, Neale BM, Benjamin DJ; 23andMe Research Team, Social Science Genetic Association Consortium. Multi-trait analysis of genome-wide association summary statistics using MTAG. *Nat Genet* 2018;50:229–37.
42 Weedon MN, Lango H, Lindgren CM, Wallace C, Evans DM, Mangino M, Freathy RM, Perry JR, Stevens S, Hall AS, Samani NJ, Shields B, Prokopenko I, Farrall M, Dominiczak A, Johnson T, Bergmann S, Beckmann JS, Vollenweider P, Waterworth DM, Mooser V, Palmer CN, Morris AD, Ouwehand WH, Zhao JH, Li S, Loos RJ, Barroso I, Deloukas P, Sandhu MS, Wheeler E, Soranzo N, Inouye M, Wareham NJ, Caulfield M, Munroe PB, Hattersley AT, McCarthy MI, Frayling TM; Diabetes Genetics Initiative, Wellcome Trust Case Control Consortium, Cambridge GEM Consortium. Genome-wide association analysis identifies 20 loci that influence adult height. *Nat Genet* 2008;40:575–83.
43 McClay JL, Adkins DE, Åberg K, Bukszár J, Khachane AN, Keefe RSE, Perkins DO, McEvoy JP, Stroup TS, Vann RE, Beardsley PM, Lieberman JA, Sullivan PF, van den Oord EJCG. Genome-wide pharmacogenomic study of neurocognition as an indicator of antipsychotic treatment response in schizophrenia. *Neuropsychopharmacology* 2011;36:616–26.
44 Lee MK, Shaffer JR, Leslie EJ, Orlova E, Carlson JC, Feingold E, Marazita ML, Weinberg SM. Genome-wide association study of facial morphology reveals novel associations with FREM1 and PARK2. *PLoS One* 2017;12:e0176566.
45 Hill WD, Marioni RE, Maghzian O, Ritchie SJ, Hagenaars SP, McIntosh AM, Gale CR, Davies G, Deary IJ. A combined analysis of genetically correlated traits identifies 187 loci and a role for neurogenesis and myelination in intelligence. *Mol Psychiatry*;15.

46 Wang X, Sun J, Li C, Mao B. EphA7 modulates apical constriction of hindbrain neuroepithelium during neurulation in Xenopus. *Biochem Biophys Res Commun* 2016;479:759–65.
47 Prost G, Braun S, Hertwig F, Winkler M, Jagemann L, Nolbrant S, Leefa IV, Offen N, Miharada K, Lang S, Artner I, Nuber UA. The putative tumor suppressor gene EphA7 is a novel BMI-1 target. *Oncotarget* 2016;7:58203–17.
48 Choi H, Lee SH, Um SJ, Kim EJ. CACUL1 functions as a negative regulator of androgen receptor in prostate cancer cells. *Cancer Lett* 2016;376:360–6.
49 Voorhoeve PG, van Mechelen W, Uitterlinden AG, Delemarre-van de Waal HA, Lamberts SW. Androgen receptor gene CAG repeat polymorphism in longitudinal height and body composition in children and adolescents. *Clin Endocrinol* 2011;74:732–5.
50 Tsukasaki K, Miller CW, Greenspun E, Eshaghian S, Kawabata H, Fujimoto T, Tomonaga M, Sawyers C, Said JW, Koeffler HP. Mutations in the mitotic check point gene, MAD1L1, in human cancers. *Oncogene* 2001;20:3301–5.
51 Schizophrenia Psychiatric Genome-Wide Association Study (GWAS) Consortium. Genome-wide association study identifies five new schizophrenia loci. *Nat Genet* 2011;43:969–76.
52 Ng MC, Shriner D, Chen BH, Li J, Chen WM, Guo X, Liu J, Bielinski SJ, Yanek LR, Nalls MA, Comeau ME, Rasmussen-Torvik LJ, Jensen RA, Evans DS, Sun YV, An P, Patel SR, Lu Y, Long J, Armstrong LL, Wagenknecht L, Yang L, Snively BM, Palmer ND, Mudgal P, Langefeld CD, Keene KL, Freedman BI, Mychaleckyj JC, Nayak U, Raffel LJ, Goodarzi MO, Chen YD, Taylor HA, Correa A, Sims M, Couper D, Pankow JS, Boerwinkle E, Adeyemo A, Doumatey A, Chen G, Mathias RA, Vaidya D, Singleton AB, Zonderman AB, Igo RP, Sedor JR, Kabagambe EK, Siscovick DS, McKnight B, Rice K, Liu Y, Hsueh WC, Zhao W, Bielak LF, Kraja A, Province MA, Bottinger EP, Gottesman O, Cai Q, Zheng W, Blot WJ, Lowe WL, Pacheco JA, Crawford DC, Grundberg E, Rich SS, Hayes MG, Shu XO, Loos RJ, Borecki IB, Peyser PA, Cummings SR, Psaty BM, Fornage M, Iyengar SK, Evans MK, Becker DM, Kao WH, Wilson JG, Rotter JI, Sale MM, Liu S, Rotimi CN, Bowden DW. FIND Consortium eMERGE Consortium DIAGRAM Consortium MuTHER Consortium MEta-analysis of type 2 DIabetes in African Americans Consortium. Meta-analysis of genome-wide association studies in African Americans provides insights into the genetic architecture of type 2 diabetes. *PLoS Genet* 2014;10:e1004517.
53 Demerath EW, Guan W, Grove ML, Aslibekyan S, Mendelson M, Zhou YH, Hedman ÅK, Sandling JK, Li LA, Irvin MR, Zhi D, Deloukas P, Liang L, Liu C, Bressler J, Spector TD, North K, Li Y, Absher DM, Levy D, Arnett DK, Fornage M, Pankow JS, Boerwinkle E, ÅK H, Li L-A IMR. Epigenome-wide association study (EWAS) of BMI, BMI change and waist circumference in African American adults identifies multiple replicated loci. *Hum Mol Genet* 2015;24:4464–79.
54 Rzehak P, Covic M, Saffery R, Reischl E, Wahl S, Grote V, Weber M, Xhonneux A, Langhendries JP, Ferre N, Closa-Monasterolo R, Escribano J, Verduci E, Riva E, Socha P, Gruszfeld D, Koletzko B. DNA-Methylation and Body Composition in Preschool Children: Epigenome-Wide-Analysis in the European Childhood Obesity Project (CHOP)-Study. *Sci Rep* 2017;7:14349.
55 Lotta LA, Gulati P, Day FR, Payne F, Ongen H, van de Bunt M, Gaulton KJ, Eicher JD, Sharp SJ, Luan J, De Lucia Rolfe E, Stewart ID, Wheeler E, Willems SM, Adams C, Yaghootkar H, Forouhi NG, Khaw KT, Johnson AD, Semple RK, Frayling T, Perry JR, Dermitzakis E, McCarthy MI, Barroso I, Wareham NJ, Savage DB, Langenberg C, O'Rahilly S, Scott RA; EPIC-InterAct Consortium Cambridge FPLD1 Consortium. Integrative genomic analysis implicates limited peripheral adipose storage capacity in the pathogenesis of human insulin resistance. *Nat Genet* 2017;49:17–26.
56 Castaño-Betancourt MC, Evans DS, Ramos YF, Boer CG, Metrustry S, Liu Y, den Hollander W, van Rooij J, Kraus VB, Yau MS, Mitchell BD, Muir K, Hofman A, Doherty M, Doherty S, Zhang W, Kraaij R, Rivadeneira F, Barrett-Connor E, Maciewicz RA, Arden N, Nelissen RG, Kloppenburg M, Jordan JM, Nevitt MC, Slagboom EP, Hart DJ, Lafeber F, Styrkarsdottir U, Zeggini E, Evangelou E, Spector TD, Uitterlinden AG, Lane NE, Meulenbelt I, Valdes AM, van Meurs JB. Novel Genetic Variants for Cartilage Thickness and Hip Osteoarthritis. *PLoS Genet* 2016;12:e1006260.
57 Mullin BH, Walsh JP, Zheng HF, Brown SJ, Surdulescu GL, Curtis C, Breen G, Dudbridge F, Richards JB, Spector TD, Wilson SG. Genome-wide association study using family-based cohorts identifies the WLS and CCDC170/ESR1 loci as associated with bone mineral density. *BMC Genomics* 2016;17:136.
58 Dyment DA, Smith AC, Alcantara D, Schwartzentruber JA, Basel-Vanagaite L, Curry CJ, Temple IK, Reardon W, Mansour S, Haq MR, Gilbert R, Lehmann OJ, Vanstone MR, Beaulieu CL, Majewski J, Bulman DE, O'Driscoll M, Boycott KM, Innes AM; FORGE Canada Consortium. Mutations in PIK3R1 cause SHORT syndrome. *Am J Hum Genet* 2013;93:158–66.
59 Majumdar A, Haldar T, Bhattacharya S, Witte JS. An efficient Bayesian meta-analysis approach for studying cross-phenotype genetic associations. *PLoS Genet* 2018;14:e1007139.