Height-reducing variants and selection for short stature in Sardinia

16 July 2024

Published in final edited form as:
Nat Genet. 2015 November ; 47(11): 1352–1356. doi:10.1038/ng.3403.

Height-reducing variants and selection for short stature in Sardinia

Magdalena Zoledziewska[#1], **Carlo Sidore**[#1,2], **Charleston W. K. Chiang**[#3], **Serena Sanna**[#1], **Antonella Mulas**[1,4], **Maristella Steri**[1], **Fabio Busonero**[1], **Joseph H. Marcus**[5], **Michele Marongiu**[1], **Andrea Maschio**[1,2,6], **Diego Ortega Del Vecchyo**[7], **Matteo Floris**[1,4,8], **Antonella Meloni**[9], **Alessandro Delitala**[10], **Maria Pina Concas**[1], **Federico Murgia**[1], **Ginevra Biino**[11], **Simona Vaccargiu**[1], **Ramaiah Nagaraja**[12], **Kirk E. Lohmueller**[3], **UK10K consortium**[13], **Nicholas J. Timpson**[14], **Nicole Soranzo**[15,16], **Ioanna Tachmazidou**[15], **George Dedoussis**[17], **Eleftheria Zeggini**[15], **The Understanding Society Scientific Group**[13], **Sergio Uzzau**[4,18], **Chris Jones**[8], **Robert Lyons**[6], **Andrea Angius**[1,8], **Gonçalo R. Abecasis**[2,20], **John Novembre**[5,20], **David Schlessinger**[12,20], and **Francesco Cucca**[1,4,20]

[1]Istituto di Ricerca Genetica e Biomedica, CNR, Monserrato, Cagliari, Italy

[2]Center for Statistical Genetics, Ann Arbor, University of Michigan, MI, USA

[3]Department of Ecology and Evolutionary Biology, University of California, Los Angeles, CA, USA

[4]Università degli Studi di Sassari, Sassari, Italy

[5]Department of Human Genetics, University of Chicago, IL, USA

[6]University of Michigan, DNA Sequencing Core, Ann Arbor, MI, USA

[7]Bioinformatics Interdepartmental Program, University of California, Los Angeles, CA, USA

[8]Center for Advanced Studies, Research, and Development in Sardinia (CRS4), AGCT Program, Parco Scientifico e Tecnologico della Sardegna, Pula, Italy

[9]II Clinica Pediatrica, Ospedale Microcitemico, Cagliari, Italy

[10]Department of Clinical and Experimental Medicine, Azienda Ospedaliero-Universitaria di Sassari, Sassari, Italy

Correspondence should be addressed to: Magdalena Zoledziewska magdalena.zoledziewska@irgb.cnr.itIstituto di Ricerca Genetica e Biomedica, Consiglio Nazionale delle RicercheMonserrato, CagliariItaly Francesco Cucca fcucca@uniss.itIstituto di Ricerca Genetica e Biomedica, Consiglio Nazionale delle RicercheMonserrato, CagliariItaly David Schlessinger SchlessingerD@mail.nih.govLaboratory of Genetics, National Institute on Aging, National Institutes of HealthBaltimore, MDUSA.

[20]These authors jointly supervised this work.

Author Contributions

D.S., F.C. M.Z, J.N. and G.R.A. conceived and supervised the study. M.Z., C.S., C.W.K.C., J.N., D.S., and F.C. drafted the manuscript. S.S., K.E.L. and G.R.A. revised the manuscript and wrote specific sections of it. A.A., C.J. and R.L. supervised sequencing experiments. F.B., A.Ma., performed sequencing experiments. C.S., M.S., M.M. and S.S. carried out genetic association analyses. C.S. analyzed DNA sequence data. M.Z., A.Mu., F.B., S.U., R.N. carried out SNP array genotyping. M.Z. and A.Mu. verified genotypes by Taqman genotyping. J.H.M., C.W.K.C., M.S., M.F., D.O.D.V., K.L., and J.N. performed polygenic score and related population genetic analyses. A.Me. and A.D. performed the clinical characterization of Laron carriers. S.V. provided DNAs for the Sardinian replication sample set, F.M., M.P.C., G.B., M.S., S.S. performed replication analysis. N.S., N.T., G.D., I.T., E.Z., and the UK10K group provided *KCNQ1* fine mapping data. All authors reviewed and approved the final manuscript.

Competing financial interests

The authors declare no competing financial interests.

Europe PMC Funders Author Manuscripts

[11]Istituto di Genetica Molecolare, CNR, Pavia, Italy

[12]Laboratory of Genetics, National Institute on Aging, National Institutes of Health, Baltimore, MD, USA

[13]A complete list of members and affiliations is provided in the Supplementary Note

[14]MRC Integrative Epidemiology Unit, University of Bristol, Bristol, UK

[15]Human Genetics, Wellcome Trust Sanger Institute, Genome Campus, Hinxton, CB10 1HH

[16]Department of Haematology, University of Cambridge, Hills Rd, Cambridge CB2 0AH

[17]Department of Dietetics-Nutrition, Harokopio University, 17671 Athens, Greece

[18]Porto Conte Ricerche srl, Tramariglio, Alghero, Italy

[#] These authors contributed equally to this work.

Europe PMC Funders Author Manuscripts

Abstract

We report sequencing-based whole-genome association analyses to evaluate the impact of rare and founder variants on stature in 6,307 individuals on the island of Sardinia. We identified two variants with large effects. One is a stop codon in the *GHR* gene, relatively frequent in Sardinia (0.87% vs <0.01% elsewhere), which in homozygosity causes the short stature Laron syndrome. We find that it reduces height in heterozygotes by an average of 4.2 cm (−0.64 s.d). The other variant, in the imprinted *KCNQ1* gene (MAF = 7.7% vs <1% elsewhere) reduces height by an average of 1.83 cm (−0.31 s.d.) when maternally inherited. Additionally, polygenic scores indicate that known height-decreasing alleles are at systematically higher frequency in Sardinians than would be expected by genetic drift. The findings are consistent with selection toward shorter stature in Sardinia and a suggestive human example of the proposed “island effect” reducing the size of large mammals.

Human height is a canonical complex trait, under tight genetic control with heritability of 80-90% (1,2). Although rare variants with strong effects have been reported in families with monogenic forms of dwarfism or gigantism, the ~700 reported variants affecting height - which explain only about 16% of the observed heritability - are typically common alleles with modest effect sizes (average <0.3 cm) ([3,4]). Little is known about the impact of rare and founder variants on stature at a population level and whether they contribute to variation in height between populations. The founder Sardinian population is especially suitable to assess the impact of such variants. Although most of the common genetic variants present elsewhere in Europe also exist in Sardinia, the isolated island population is enriched for numerous variants that are very rare or absent elsewhere ([5]) and were not included in the commercial genotyping arrays or multi-population sequencing panels that are commonly used to characterize genetic variants through imputation ([6]).

We therefore used whole genome sequencing to investigate height in a large sample of Sardinians, who, with an average male stature of 168.5 cm ([7]), are among the shortest European populations.

We used whole genome sequencing (~4x) of 2,120 Sardinians to construct a reference panel of ~17.6 million SNPs (Supplementary Fig. 1a,b) and carry out a genome wide association study (GWAS) for height. After stringent quality controls and imputation using a scaffold of 890,542 genotyped SNPs, 11,826,948 SNPs were assessed in 6,307 participants in the SardiNIA study, from villages in the Lanusei valley ([1]). The GWAS found two signals strongly associated with stature, one located in the *GHR* (5p12) and the other in the *KCNQ1* (11p15.5) genes, which encode the growth hormone receptor and a voltage-gated potassium channel, respectively (Supplementary Fig. 1c). Notably, their joint effect in the SardiNIA cohort is as large as that contributed jointly by the top 10 height associated alleles assessed in the GIANT meta-analysis[[4]] and by the top 5 when using the effect sizes observed in the replication set.

The first of these signals is rs121909358 ($p=1.07\times10^{-10}$, effect −0.64 s.d. corresponding to −4.2 cm, Fig.1a, Supplementary Fig. 2). The height-reducing T allele is found on a single haplotype (Supplementary Fig. 3). It creates a loss of function termination codon (R61X) in *GHR*. The variant and its association with height would not have been detected without imputation from the Sardinian sequencing panel (imputation accuracy, RSQR=0.94, validated by direct genotyping) ([6]), as the variant is extremely rare outside of Sardinia (frequency <1/60,000, ExAC Browser, URLs).

Homozygosity for this stop codon variant is one of several mutations in *GHR* known to cause Laron syndrome (LS) (OMIM#262500); a rare autosomal recessive condition characterized by primary growth hormone insensitivity. Since the initial description ([8]), more than 250 LS cases have been reported (Orphanet, URLs), with the majority of patients identified in Maghrebi-Sephardic Jewish groups ([9]) and an isolated population of Spanish descent in Ecuador ([10]). The global estimated prevalence of LS is 1-9 per million (Orphanet, URLs) suggesting world-wide carrier frequencies of less than 0.01%. In contrast, we observed an unexpectedly high frequency of 0.87% for the R61X variant among 1,481 unrelated individuals from the SardiNIA cohort. Consistent with this frequency, 1 homozygous affected LS individual has been observed among the 10,721 inhabitants of the 4 villages in the Lanusei valley. The association of R61X with height was replicated in an independent Sardinian cohort of 5,314 individuals from an additional 6 villages (Supplementary Note), though its frequency and the effect size are estimated to be smaller (MAF= 0.46% in 857 unrelated individuals, $p_{one\text{-}tail}$ =0.015, effect −0.31 s.d., corresponding to −1.89 cm).

Our results extend to the general population the evidence that *GHR* mutations affect height of heterozygous carriers (Supplementary Table 1, [11,12]). In addition, 30% of the carriers from the SardiNIA study also showed limited elbow extension, which is very rare in

URLs
HGDP: http://www.hagsc.org/hgdp/index.html
OMIM: http://www.omim.org/
ExAC Browser: http://exac.broadinstitute.org
SardiNIA project home page: https://sardinia.irp.nia.nih.gov/
EPACTS: http://genome.sph.umich.edu/wiki/EPACTS
ENCODE: https://www.encodeproject.org/
GWAS Catalog: https://www.ebi.ac.uk/gwas/

unaffected individuals but characteristic of LS patients due to underdevelopment of the muscular system and an abnormal degree of humerus rotation (Supplementary Table 2, [8]). Interestingly, among 2,120 sequenced Sardinians, we also found instances of two additional rare variants described to cause LS in Southern European and South American populations (Supplementary Note, Supplementary Table 3); however those variants were at frequencies too low in the SardiNIA cohort (MAF<0.003) to assess phenotypic effects in heterozygotes.

The second GWAS signal in *KCNQ1* (Fig. 1b) is complicated by the fact that it falls in a known tissue-specific imprinted gene cluster. Indeed, we found striking evidence that the association with short stature is maternally inherited (Fig. 1, Table 1), with the strongest maternal effects at rs150199504 (MAF= 7.7%, $p=5.6\times10^{-9}$, maternal effect −0.315 s.d., corresponding to −1.83 cm), and no significant paternal effect (p=0.95) (Table 1, Supplementary Fig. 2). By directly typing one of the top associated variants, rs2075870, which also showed a modest albeit significant association with decreased height in ~90,000 individuals of European origin ([13]), we confirmed the association in the independent Sardinian cohort ($p=3.6\times10^{-4}$ for the maternal effect −0.22 s.d., corresponding to −1.17 cm and p=0.1 for paternal effect). The association signal spans 48Kb encompassing rs2075870 and 4 additional variants in LD with rs150199504 (pvalue $<1\times10^{-6}$, $r^2>0.7$) (Fig. 1, Table 1) making it difficult to identify the causal variant (s).

However, we found that differences in allele frequencies and LD patterns among the variants in Sardinia compared to other populations provided a route to prioritize the list for the responsible variant(s) (Fig. 2). Remarkably, among the SNPs in LD in Sardinia, we could exclude rs2075870, rs67004488, rs149658560 and rs12790610 as causal based on their frequencies, LD patterns and results from GWAS in other populations. In particular, these variants are common (MAF ~10%), in LD with each other ($r^2>0.3$) in South Asia, and yet no association of rs2075870 with height has been observed there ([13]). By contrast, among our core associated SNPs, the top variants rs150199504 and rs143840904 are in lower LD with rs2075870 and much rarer in South Asia ($r^2<0.3$ and MAF <1.2 % and <2.6% respectively) (Fig. 2d) and thus association with height could be missed if they are not directly typed in very large sample sets. Hence, rs143840904 and especially our lead variant rs150199504 are plausible causal candidates.

To further assess their candidacy, we directly tested the 6 core associated variants in 19,053 individuals from 6 GWAS European cohorts, among which we expect more resolving power than in Sardinia due to lower LD in the region (Fig. 2b, 2c). Among the 5 variants that passed quality checks, rs150199504 was again the most significantly associated and had the strongest effect in these samples as well ($p=2.82\times10^{-4}$, effect −0.243 s.d) – even though it was the rarest of the five (MAF = 0.89 %). To a lesser extent significant association was also seen for rs143840904 ($p=1.23\times10^{-3}$, effect −0.145 s.d.), but was not observed for the 3 other variants (Supplementary Table 4). Interestingly, in a reciprocal conditional analysis, the effect of rs143840904 was completely accounted for by rs150199504 (p=0.24, effect −0.06 s.d.). By contrast residual association remained at rs150199504 after conditioning on rs143840904 (p=0.06, effect −0.172 s.d.). This further genetic evidence supports rs150199504 as the main driver of the association with decreased height at this locus.

Suggestively, rs150199504 (and rs143840904) fall in a differentially methylated region (ENCODE, URLs), hinting at a possible effect on expression.

The maternal effect we observed for *KCNQ1* on height is consistent with the established monoallelic expression of maternal alleles at this imprinted locus ([14]). Furthermore, the observation that translocations and inversions disrupting the function of *KCNQ1* result in Beckwith-Wiedemann gigantism ([15]) suggests that, by inference, the short stature alleles reported here result in a gain of function.

KCNQ1 variation has been implicated in several other traits, including platelet aggregation, electrocardiographic measures and type 2 diabetes, with the latter also influenced by parent of origin effects ([16–20]). Those associations were, however, all completely independent of any of the 6 top *KCNQ1* associated variants considered here ($r^2<0.08$). Furthermore, the 6 variants showed no significant association with any of 193 traits measured in the SardiNIA study participants (data not shown)([1,21]).

To evaluate the overall impact of known variants on the average short stature observed in Sardinia relative to other populations and to test the possibility that short stature might be selected for in this island population, we used polygenic height scores. These scores measure the total frequency of height-changing alleles in a population, weighing each allele by its effect size. A general North-to-South gradient for height in Europe due to directional selection has been reported ([22,23]) with Sardinia as a significant outlier among the Human Genome Diversity Panel European populations (URLs). Consistent with these studies, we observed a significantly lower polygenic height score in Sardinia compared to other European populations examined in the 1000 Genomes project, including the Southern European Tuscans and Spanish (Fig. 3). Adding our *KCNQ1* and *GHR* variants to the previously described 691 alleles ([4]), the polygenic score of Sardinians decreased by 3.8%. Overall, Sardinian scores are lower than would be expected compared to other European populations ($p=1.62\times10^{-6}$, −5.9cm relative to CEU, 1.6% average increase in frequency for height decreasing alleles, Supplementary Fig.5), even when calibrating for genome-wide patterns of differentiation due to genetic drift, suggesting that selection has played a role in decreasing height in Sardinia. The differences in height explained by the polygenic score are in accord with the observed ~10 cm of phenotypic differences between Sardinians and the other European populations.

We have also considered the possibility that Sardinians might have an additional contribution of reduced height due to the expression of recessively acting height-decreasing alleles exposed due to founder effects. However, the impact of elevated homozygosity among Sardinians on height appears to be small (0.129 s.d.) relative to the effects predicted by the polygenic score (0.910 s.d.) (Supplementary Note).

An example of low frequency allele affecting height was recently reported from the Icelandic population ([24]). However, our findings demonstrate for the first time that part of the missing heritability of human height can be attributable to rare variants involved in monogenic disorders, as shown by *GHR*, as well as by variants common in isolated populations but rare elsewhere, as exemplified by *KCNQ1*. Indeed, a shift toward higher

Europe PMC Funders Author Manuscripts

frequencies for variants with large size effects observed in Sardinia ([6,25]) – and in this case the powerful height-decreasing variants -- allowed us to detect, in a cohort of thousands of participants, associations that were missed in GWAS and meta-analyses of hundreds of thousands of individuals.

Intriguingly, the increased frequencies of height-decreasing alleles at *GHR* and *KCNQ1,* and especially the polygenic height scores in this population, are also consistent with the long-standing observation of an “island effect” in which many large animals become adaptively smaller on islands relative to their mainland counterparts ([26]). The extinct Sardinian mammoth (*Mammuthus lamarmorae*) and deer (*Megaloceros cazioti*) are two examples ([27]). One complication to assess this in in humans is that selection for decreased height likely began prior to the peopling of Sardinia among the early European farmer lineage ([28]) that is thought to have initially colonized the island([29]), and Sardinians might have simply retained short stature that evolved earlier. However, we observe lower polygenic height scores in Sardinia even when compared with other populations with high proportions of early European Neolithic ancestry (Tuscans and Spanish)([30]). Thus, selection for decreased height likely continued and was particularly strong in the lineage leading to modern Sardinians. One conjecture is that crop yields or other nutritional sources were limited in the restricted island environment, but exactly why selection for decreased height was acting among the Neolithic ancestors of the Sardinians, and likely intensified after the occupation of the island, remains an open and interesting question.

Online Methods

Research subjects

All individuals included in the study were of Sardinian origin and participate in a longitudinal study of age-related quantitative traits on the island (SardiNIA, URLs). The study involves four villages: Lanusei, Ilbono, Elini end Arzana, located in the Lanusei Valley([1,21,31]). 6,148 volunteers have been described before ([1]) and an additional 773 individuals have been enrolled during the follow up stage of the project ([6]). 6,602 individuals had complete genotyping data. For analyses, we only included measurements for individuals at age >20 years, and also discarded 4 subjects with Morquio Syndrome (OMIM *607939), leading to a total of 6,307 samples.

All participants provided informed consent and studies were approved by the Local Research Ethic Committees (No 2009/0016600).

Genotyping methods, low-pass sample sequencing, variant calling, genotype imputation and GWAS analysis

All SardiNIA individuals were typed with four Illumina Infinium arrays. Low pass sequencing, variant calling, genotype imputation and GWAS analysis was conducted as previously described ([31]).

GWA analysis

For our GWAS we tested association for the 11,826,948 imputed or genotyped variants that passed quality control filters [MACH r^2>0.3 for MAF>=0.01, r^2>=0.6 for MAF<0.01 ([31])], assuming an additive model of inheritance and adjusting for age, age squared and gender as covariates and applying the inverse normal transformation to the residuals. Association was performed using EMMAX ([32]) as implemented in the software EPACTS (URLs), which accounts for relatedness and population structure using an empirical kinship matrix derived from genotype data. The genomic control inflation factor was λ=0.989, indicating no inflation of results.

Validation of imputation results by genotyping

GWAS identified three loci significantly associated with stature: the *GHR* gene, with top variant rs121909358; the *KCNQ1* gene, with 6 variants in LD (Table 1); and the *SMURF2* gene, with top variant rs143051029.

We validated imputation of rs121909358 genotypes by directly genotyping 2,818 samples with a TaqMan assay. Concordance between imputation and validation was 99.89%. At *KCNQ1*, two leading variants, rs67004488 and rs2075870, were present on the Cardio-Metabo Illumina chip, so that validation was not necessary. The third association at rs143051029 was evaluated with standard Sanger sequencing. We selected 96 samples for sequencing, including 4 imputed homozygotes, 22 imputed heterozygotes with uncertain allele dosages and 70 randomly selected samples. The variant, located in a complex region, did not pass validation due to the high mismatch rate (34.4%) between imputed genotypes and those validated by Sanger sequencing and was not further considered in analyses.

Conditional analysis

We conducted standard conditional analyses using EPACTS software for the two identified regions by including the top variants as covariates. We examined the 1Mb region around the top SNPs (rs121909358 for *GHR* and rs150199504 for *KCNQ1*). In both cases, the top variant completely explained the association at the two loci; none of the SNPs in the region passed the significance threshold after Bonferroni correction. The variant chr5:43229441, 540Kb away, from rs121909358, was fully explained by the effect of rs121909358 (p after conditional = 0.1).

Replication cohorts

We replicated findings in an independent cohort of 5,314 Sardinians and 19,053 non-Sardinian European samples. Details on genotyping and analyses are described in Supplementary.

Characterization of the associated region on chromosome 5

To visualize the haplotypes carrying the Laron variant (Supplementary Fig. 3), we interrogated ±3Mb surrounding chr5:42689036 in 11 sequenced unrelated carriers of rs121909358. The analysis was performed using SelScan ([33]) and included 9,526 SNPs with MAF >5% in Sardinia.

Europe PMC Funders Author Manuscripts

Parent-of-origin effects

For SNPs in the *KCNQ1* locus, we estimated parental origin of alleles for all individuals using Merlin (--*best* option)([34]). We then considered two separate variables, one for the maternal (G_m) and one for the paternal (G_p) allele, coded as 1 if the corresponding transmitted allele was the minor allele at the SNP, and 0 otherwise. Missing values were assigned to founders and other individuals for whom parental origin could not be defined unambiguously. Of consequence variables G_m and G_p were non-missing for 5,026 SardiNIA individuals and 4,666 OGP individuals. Two linear models were then used:

$$\mathrm{Y} \sim \beta_0 + \beta_1 \mathrm{G_m} + \beta^T \mathrm{C}$$
$$\mathrm{Y} \sim \beta_0 + \beta_2 \mathrm{G_p} + \beta^T \mathrm{C}$$

where Y denotes trait and C, other covariates. As both the SardiNIA and OGP studies consists of large families, the transmissions evaluated by G_p and G_m are not independent. We therefore tested the null hypothesis $\beta_1 \neq 0$ (for model 1) and $\beta_2 \neq 0$ (for model 2) by fitting a mixed linear regression model that accounts for familiar relatedness (*lmekin()* and *kinship()* functions in the coxme and kinship R packages). In the models, we used the same covariates and trait normalization procedure as in the GWAS analysis. We then assessed the hypothesis of heterogeneity of effects, $\beta_1 \neq \beta_2$, using Cochran's Q statistic. The test was carried out for all SNPs in the *KCNQ1* gene, and on SNP rs2075870 in the OGP cohort.

Population-level height polygenic score calculation and evaluation

In the population genetic analyses, we focused on a subset of 1,081 unrelated sequenced individuals (Supplementary Note).

To investigate whether height-decreasing loci have been under selection in Sardinia, for each population *m*, we calculated the polygenic height score as

$$Z_m = 2\sum_{l=1}^{L} \beta_l p_{ml}$$

where β_l is the effect size of the height-increasing allele *l* and p_{ml} is the frequency of allele *l* in population *m*. To avoid biases and to ensure uniformity of the source of effect size estimates, we used the effect size estimates from the Sardinian dataset regardless of whether the variant is significantly associated with height in this dataset. We first calculated the polygenic height score (Z_m) based on the 691 height loci identified by the GIANT consortium ([4]) with effect sizes estimated in the Sardinian dataset and then added the two top variants reported, totaling 693 height alleles. To test if there were a signature of polygenic adaptation on height in Sardinia, we adopted a framework developed by Berg and Coop ([23]), which builds a multivariate normal model based on matched, presumably neutral variants, to account for relationships among populations (Fig. 3). Populations with extreme polygenic scores relative to the expectation (pvalue = 0.01) are likely to have undergone selection. To construct a null distribution of frequencies needed for the multivariate normal framework, we obtained for each of the height loci all variants in the 1000 Genomes phase 3

European data with minor allele count +/− 10 counts (~ 1% in frequency), B score ([35]) +/− 50 units, and local recombination rates +/− 0.5 cM/Mb. A random subset of 509,386 SNPs, representing 10% of the union of the matched SNPs, were then used as a set of matched SNPs for the analysis. Of note, we also repeated the calculation using effect sizes estimated by the GIANT consortium as well as using only a subset of 162 SNPs that are not subject to population stratification ([22]) (Supplementary Fig. 4).

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank all the volunteers who generously participated in this study and made this research possible. All participants provided informed consent and studies were approved by the Local Research Ethic Committees (No 2009/0016600). This study was funded in part by the National Institutes of Health (National Institute on Aging, National Heart Lung and Blood Institute, and National Human Genome Research Institute). This research was supported by National Human Genome Research Institute grants HG005581, HG005552, HG006513, HG007089, HG007022, and HG007089; by National Heart Lung and Blood Institute grant HL117626; by the Intramural Research Program of the NIH, National Institute on Aging, with contracts N01-AG-1-2109 and HHSN271201100005C; by Sardinian Autonomous Region (L.R. no. 7/2009) grant cRP3-154; by grant FaReBio2011 “Farmaci e Reti Biotecnologiche di Qualità”; PB05 InterOmics MIUR Flagship Project; by NIH NRSA postdoctoral fellowship (F32GM106656) to C.W.K.C; by UC MEXUS-CONACYT doctoral fellowship 213627 to D.O.D.V; by Italian Ministry of Education, University and Research (MIUR) no.5571/DSPAR/2002. The HELIC study was funded by the Wellcome Trust (098051) and the European Research Council (ERC-2011-StG 280559-SEPI). The TEENAGE study has been supported by the Wellcome Trust (098051), European Union (European Social Fund—ESF) and Greek national funds through the Operational Programme ‘Education and Lifelong Learning’ of the National Strategic Reference Framework (NSRF)—Research Funding Programme: Heracleitus II, investing in knowledge society through the European Social Fund. The UK Household Longitudinal Study is led by the Institute for Social and Economic Research at the University of Essex and funded by the Economic and Social Research Council. Information on how to access the data can be found on the Understanding Society website https://www.understandingsociety.ac.uk/. This study makes use of data generated by the UK10K Consortium, derived from samples from UK10K_COHORTS_TWINSUK (The TwinsUK Cohort) and UK10K_COHORT_ALSPAC (the Avon Longitudinal Study of Parents and Children). A full list of the investigators who contributed to the generation of the data is available from www.UK10K.org. Funding for UK10K was provided by the Wellcome Trust under award WT091310. We thank Jeremy Berg for scripts and suggestions on the polygenic score analysis.

REFERENCES

1. Pilia G, et al. Heritability of cardiovascular and personality traits in 6,148 Sardinians. PLoS Genet. 2006; 2:e132. [PubMed: 16934002]
2. Silventoinen K, Kaprio J, Lahelma E, Koskenvuo M. Relative effect of genetic and environmental factors on body height: differences across birth cohorts among Finnish men and women. Am. J. Public Health. 2000; 90:627–630. [PubMed: 10754982]
3. Lango Allen H, et al. Hundreds of variants clustered in genomic loci and biological pathways affect human height. Nature. 2010; 467:832–8. [PubMed: 20881960]
4. Wood AR, et al. Defining the role of common variation in the genomic and biological architecture of adult human height. Nat. Genet. 2014; 46:1173–1186. [PubMed: 25282103]
5. Francalacci P, et al. Low-pass DNA sequencing of 1200 Sardinians reconstructs European Y-chromosome phylogeny. Science. 2013; 341:565–569. [PubMed: 23908240]
6. Sidore C, et al. Genome sequencing elucidates Sardinian genetic architecture and augments association analyses for lipid and blood inflammatory markers. Nature Genetics. 2015
7. Arcaleni E. Secular trend and regional differences in the stature of Italians. J. Anthropol. Sci. Riv. Antropol. JASS Ist. Ital. Antropol. 2012; 90:233–237.

8. Laron, Z. Laron Syndrome - From Man to Mouse - Lessons from Clinical and Experimental Experience. Springer-Verlag; 2011. at <http://www.springer.com/us/book/9783642111822>
9. Laron Z. The syndrome of familial dwarfism and high plasma immunoreactive human growth hormone. Birth Defects Orig. Artic. Ser. 1974; 10:231–238. [PubMed: 4470894]
10. Rosenbloom AL, Guevara Aguirre J, Rosenfeld RG, Fielder PJ. The little women of Loja--growth hormone-receptor deficiency in an inbred population of southern Ecuador. N. Engl. J. Med. 1990; 323:1367–1374. [PubMed: 2233903]
11. Laron Z, Klinger B, Erster B, Silbergeld A. Serum GH binding protein activities identifies the heterozygous carriers for Laron type dwarfism. Acta Endocrinol. (Copenh.). 1989; 121:603–608. [PubMed: 2800930]
12. Guevara-Aguirre J, et al. Effects of heterozygosity for the E180 splice mutation causing growth hormone receptor deficiency in Ecuador on IGF-I, IGFBP-3, and stature. Growth Horm. IGF Res. Off. J. Growth Horm. Res. Soc. Int. IGF Res. Soc. 2007; 17:261–264.
13. Lanktree MB, et al. Meta-analysis of Dense Genecentric Association Studies Reveals Common and Uncommon Variants Associated with Height. Am. J. Hum. Genet. 2011; 88:6–18. [PubMed: 21194676]
14. Lee MP, et al. Loss of imprinting of a paternally expressed transcript, with antisense orientation to KVLQT1, occurs frequently in Beckwith-Wiedemann syndrome and is independent of insulin-like growth factor II imprinting. Proc. Natl. Acad. Sci. U. S. A. 1999; 96:5203–5208. [PubMed: 10220444]
15. Soejima H, Higashimoto K. Epigenetic and genetic alterations of the imprinting disorder Beckwith-Wiedemann syndrome and related disorders. J. Hum. Genet. 2013; 58:402–409. [PubMed: 23719190]
16. Horikoshi M, et al. New loci associated with birth weight identify genetic links between intrauterine growth and adult height and metabolism. Nat. Genet. 2013; 45:76–82. [PubMed: 23202124]
17. Johnson AD, et al. Genome-wide meta-analyses identifies seven loci associated with platelet aggregation in response to agonists. Nat. Genet. 2010; 42:608–613. [PubMed: 20526338]
18. Newton-Cheh C, et al. Common variants at ten loci influence QT interval duration in the QTGEN Study. Nat. Genet. 2009; 41:399–406. [PubMed: 19305408]
19. Voight BF, et al. Twelve type 2 diabetes susceptibility loci identified through large-scale association analysis. Nat. Genet. 2010; 42:579–589. [PubMed: 20581827]
20. Kong A, et al. Parental origin of sequence variants associated with complex diseases. Nature. 2009; 462:868–874. [PubMed: 20016592]
21. Orru V, et al. Genetic variants regulating immune cell levels in health and disease. Cell. 2013; 155:242–56. [PubMed: 24074872]
22. Turchin MC, et al. Evidence of widespread selection on standing variation in Europe at height-associated SNPs. Nat. Genet. 2012; 44:1015–1019. [PubMed: 22902787]
23. Berg JJ, Coop G. A population genetic signal of polygenic adaptation. PLoS Genet. 2014; 10:e1004412. [PubMed: 25102153]
24. Steinthorsdottir V, et al. Identification of low-frequency and rare sequence variants associated with elevated or reduced risk of type 2 diabetes. Nat. Genet. 2014; 46:294–298. [PubMed: 24464100]
25. Danjou F, et al. WHOLE GENOME SEQUENCING-BASED GWAS IN SARDINIA EXPLICATES GENETIC REGULATION OF HEMOGLOBIN LEVELS AND CLINICAL CONSEQUENCES. Nature Genetics. 2015
26. Millien V. Morphological evolution is accelerated among island mammals. PLoS Biol. 2006; 4:e321. [PubMed: 16968136]
27. van der Geer, A.; Lyras, G.; de Vos, J.; Dermitzakis, M. Evolution of Island Mammals. Wiley-Blackwell; 2010. p. 103-130.at <http://onlinelibrary.wiley.com/doi/10.1002/9781444323986.ch9/summary>
28. Mathieson I, et al. Eight thousand years of natural selection in Europe. bioRxiv. 2015:016477. doi: 10.1101/016477.
29. Lazaridis I, et al. Ancient human genomes suggest three ancestral populations for present-day Europeans. Nature. 2014; 513:409–413. [PubMed: 25230663]

30. Haak W, et al. Massive migration from the steppe was a source for Indo-European languages in Europe. Nature. 2015 advance online publication.

Methods references

31. Pistis G, et al. Rare variant genotype imputation with thousands of study-specific whole-genome sequences: implications for cost-effective study designs. Eur. J. Hum. Genet. 2014 doi:10.1038/ejhg.2014.216.

32. Kang HM, et al. Variance component model to account for sample structure in genome-wide association studies. Nat Genet. 2009; 42:348–54. [PubMed: 20208533]

33. Szpiech ZA, Hernandez RD. selscan: an efficient multithreaded program to perform EHH-based scans for positive selection. Mol. Biol. Evol. 2014; 31:2824–2827. [PubMed: 25015648]

34. Abecasis GR, Cherny SS, Cookson WO, Cardon LR. Merlin--rapid analysis of dense genetic maps using sparse gene flow trees. Nat. Genet. 2002; 30:97–101. [PubMed: 11731797]

35. McVicker G, Gordon D, Davis C, Green P. Widespread genomic signatures of natural selection in hominid evolution. PLoS Genet. 2009; 5:e1000471. [PubMed: 19424416]

Europe PMC Funders Author Manuscripts

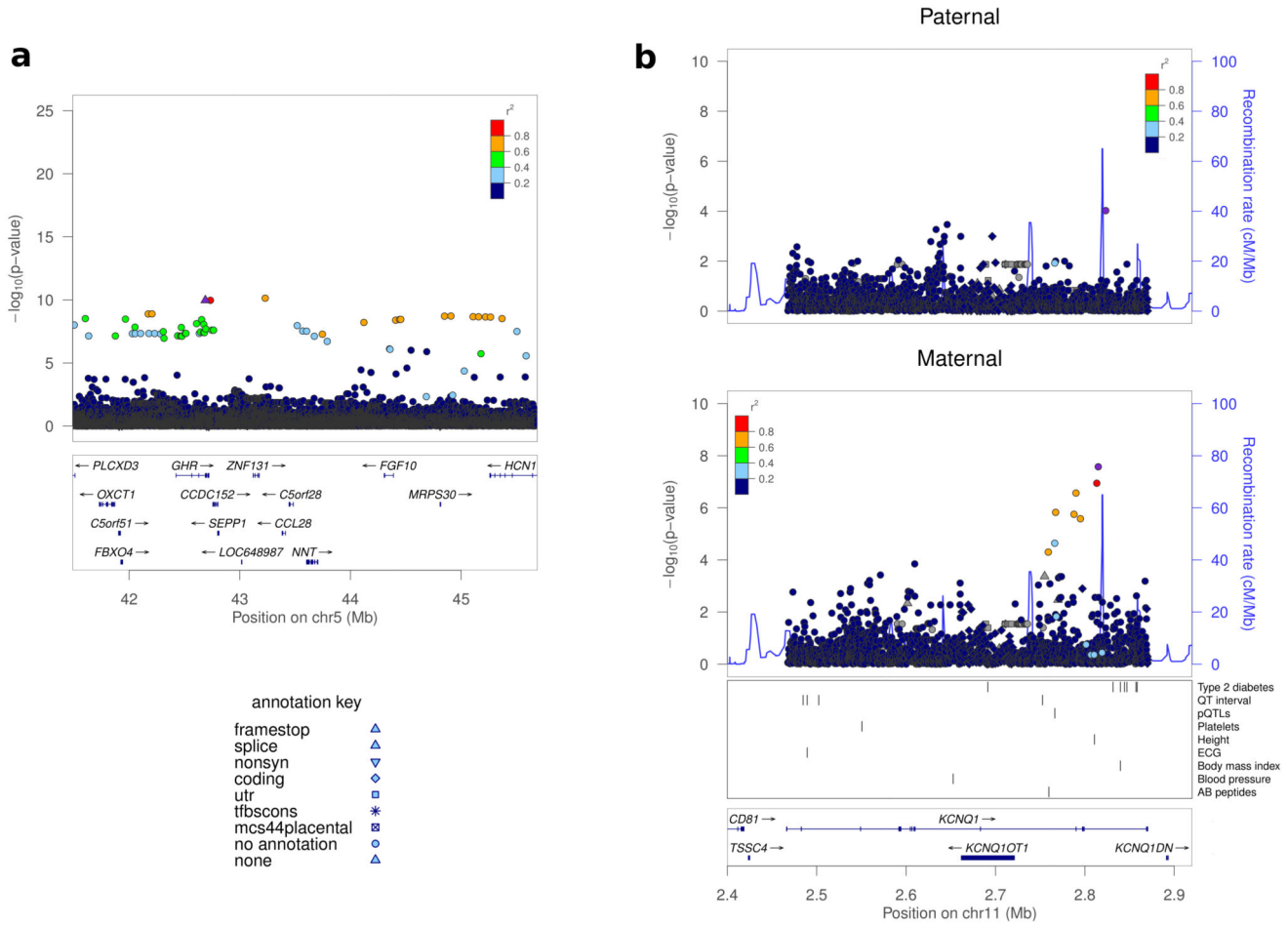

Figure 1. Regional association plots for *GHR* and *KCNQ1* locus

a) *GHR* locus, the Y axis shows the association strength (−log10 pvalue) versus the genomic positions (hg19/GRCh37) around the most significant SNP (purple). Other SNPs in the region are color-coded to reflect their LD with the top SNP. Symbols reflect genomic functional annotation. Genes and the position of exons are shown below. b) Regional plot at the *KCNQ1* locus for the paternal and maternal effects respectively. The position of GWAS catalog SNPs (URLs) with the corresponding traits and the position of exons in the *KCNQ1* region are indicated below.

Europe PMC Funders Author Manuscripts

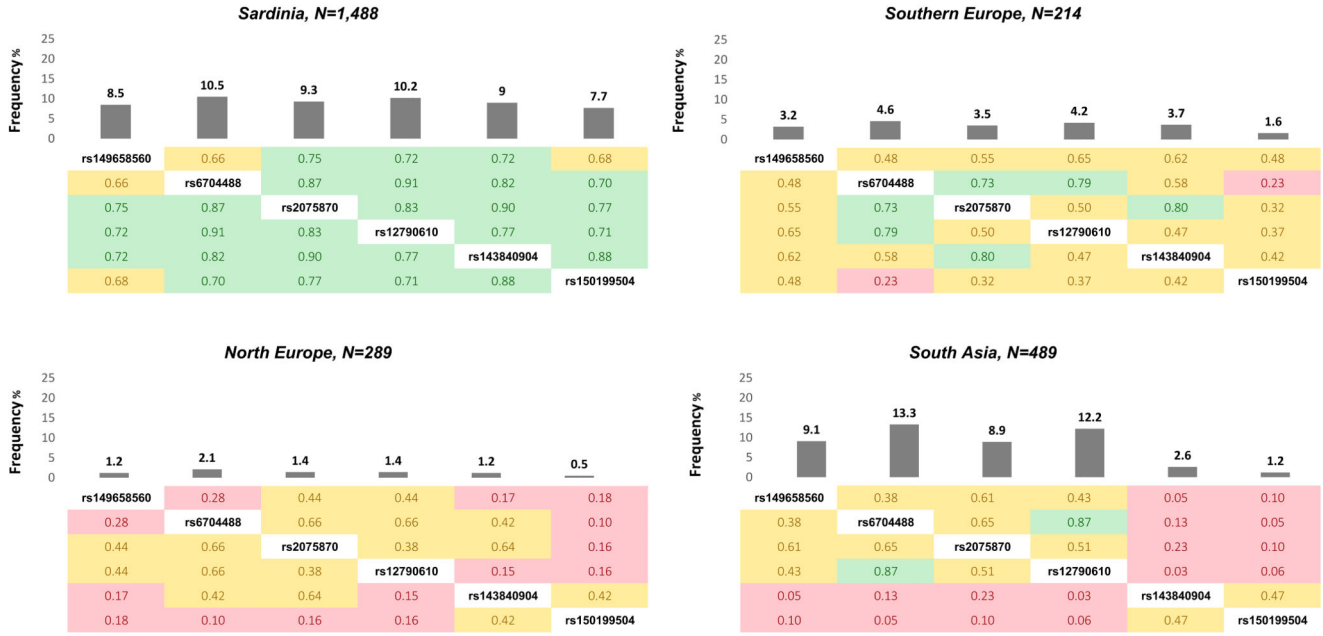

Figure 2. Worldwide frequency and LD pattern for the six top *KCNQ1* SNPs

The figure illustrates the frequency (upper panel) and the pairwise LD matrix (lower panel) for the six top SNPs associated in Sardinia at the *KCNQ1* locus. Data are presented for 4 populations: a) Sardinia, b) Southern Europe, c) Northern Europe, d) East Asia. Matrix cells are colored according to the LD value: green if r^2>=0.7; yellow if 0.3<= r^2 <0.7 ; red if r^2<0.3.

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

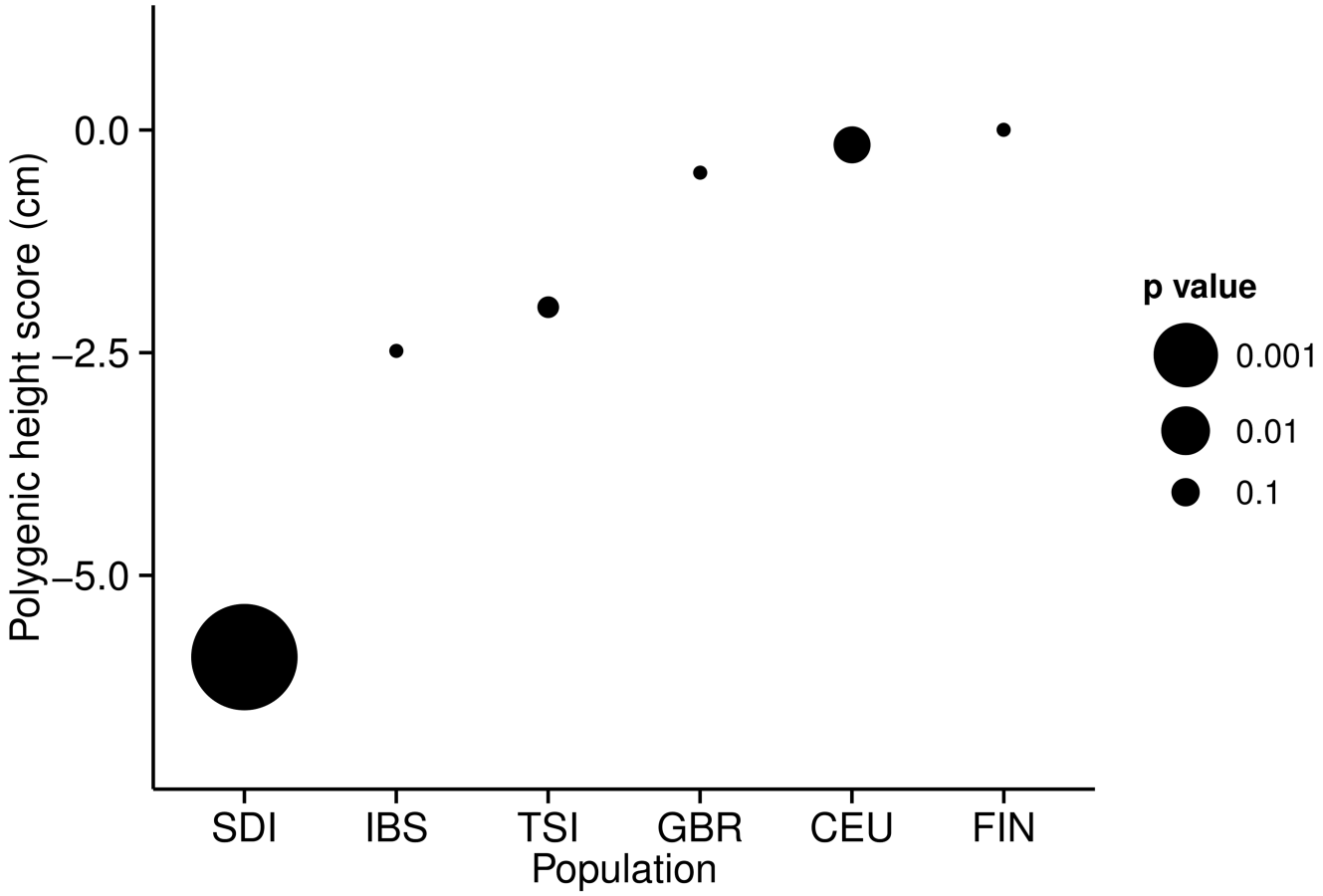

Figure 3. Polygenic score analysis for height

Polygenic score based on the 2 top associated variants (rs121909358 and rs150199504) and the 691 height loci from GIANT for which the effect size in Sardinia and allele frequencies in 1000 Genomes phase 3 data are available. The black circles indicate the scale for display of p-values according to circle size. Abbreviations: SDI: SardiNIA cohort; IBS, TSI, GBR, CEU, and FIN: 1000 Genomes populations.

Europe PMC Funders Author Manuscripts

Table 1

Parental of origin effects at *KCNQ1*

The table summarizes the strongest results for the parental of origin association test at the *KCNQ1* locus (defined as pvalue<1×10^{-6} in either the maternal or paternal tests for the assessed 500Kb region). At each SNP, we report in the column N the number of informative transmissions used (see Methods) and the association parameters obtained evaluating the minor allele i) without considering parent of origin, ii) when maternally inherited, and iii) paternally inherited. The last column reports the pvalue for heterogeneity between estimated paternal and maternal effects.

					Both		Maternal		Paternal		
rs ID	Chr:Position	Minor Allele/Other	MAF	N	Effect (StdErr)	Pvalue	Effect (StdErr)	pvalue	Effect (StdErr)	pvalue	Heterogenity pvalue
rs150199504	11:2814960	G/C	0.083	5059	−0.168 (0.039)	1.84×10^{-5}	−0.315 (0.054)	5.56×10^{-9}	−0.0032 (0.050)	0.9488	2.46×10^{-5}
rs143840904	11:2813322	T/C	0.094	5041	−0.152 (0.038)	4.58×10^{-5}	−0.274 (0.050)	3.92×10^{-8}	+0.0021 (0.049)	0.9653	7.55×10^{-5}
rs2075870	11:2790019	A/G	0.094	5044	−0.158 (0.038)	2.65×10^{-5}	−0.273 (0.051)	6.97×10^{-8}	−0.0172 (0.048)	0.793	0.0002
rs149658560	11:2767262	A/G	0.076	5050	−0.161 (0.042)	1.01×10^{-4}	−0.297 (0.058)	2.93×10^{-7}	−0.0121 (0.052)	0.8183	0.0003
rs12790610	11:2794998	G/A	0.095	5014	−0.165 (0.037)	1.02×10^{-5}	−0.258 (0.051)	4.73×10^{-7}	−0.044 (0.048)	0.3531	0.0023
rs67004488	11:2787804	G/A	0.104	5026	−0.157 (0.036)	1.2×10^{-6}	−0.244 (0.049)	5.21×10^{-7}	−0.040 (0.047)	0.3875	0.0024