Height-reducing variants and selection for short stature in Sardinia

Magdalena Zoledziewska1,19, Carlo Sidore1,2,19, Charleston W K Chiang3,19, Serena Sanna1,19, Antonella Mulas1,4, Maristella Steri1, Fabio Busonero1, Joseph H Marcus5, Michele Marongiu1, Andrea Maschio1,2, Diego Ortega Del Vecchyo7, Matteo Floris1,4,8, Antonella Meloni9, Alessandro Delitala10, Maria Pina Concas1, Federico Murgia1, Ginevra Biino11, Simona Vaccargiu1, Ramaiah Nagaraja12, Kirk E Lohmueller3, UK10K Consortium13, Nicholas J Timpson14, Nicole Soranzo15,16, Ioanna Tachmazidou15, George Deoussis17, Eleftheria Zeggini15, The Understanding Society Scientific Group13, Sergio Uzzau4,18, Chris Jones8, Robert Lyons6, Andrea Angius1,8, Gonçalo R Abecasis2,20, John Novembre5,20, David Schlessinger12,20 & Francesco Cucca1,4,20

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 identify two variants with large effects. One variant, which introduces a stop codon in the GHR gene, is relatively frequent in Sardinia (0.87% versus <0.01% elsewhere) and in the homozygous state causes Laron syndrome involving short stature. We find that this variant reduces height in heterozygotes by an average of 4.2 cm (~0.64 s.d.). The other variant, in the imprinted KCNQ1 gene (minor allele frequency (MAF) = 7.7% in Sardinia versus <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 frequencies in Sardinians than would be expected by genetic drift. The findings are consistent with selection for shorter stature in Sardinia and a suggestive human example of the proposed ‘island effect’ reducing the size of large mammals.

Human height is a classical complex trait under tight genetic control, with heritability of 80–90% (refs. 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 of <0.3 cm)1,4. Little is known about the impact of rare and founder variants on stature at a population level and whether these variants 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 elsewhere2 and are not included on commercial genotyping arrays or multiple-population sequencing panels that are commonly used to characterize genetic variants through imputation3.

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 (ref. 7), are among the shortest European populations.

We used whole-genome sequencing (~4× coverage) 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 of the SardiNIA study from villages in the Lanusei valley in Sardinia. The GWAS found two signals strongly associated with stature, one located in the GHR gene (5p12).

Received 10 June; accepted 21 August; published online 14 September 2015; doi:10.1038/ng.3403

1Istituto di Ricerca Genetica e Biomedica, Consiglio Nazionale delle Ricerche (CNR), Monserrato, Cagliari, Italy. 2Center for Statistical Genetics, University of Michigan, Ann Arbor, Michigan, USA. 3Department of Ecology and Evolutionary Biology, University of California, Los Angeles, California, USA. 4Dipartimento di Scienze Biomediche, Università degli Studi di Sassari, Sassari, Italy. 5Department of Human Genetics, University of Chicago, Chicago, Illinois, USA. 6DNA Sequencing Core, University of Michigan, Ann Arbor, Michigan, USA. 7Bioinformatics Interdepartmental Program, University of California, Los Angeles, California, USA. 8Center for Advanced Studies, Research and Development in Sardinia (CRS4), Advanced Genomics Computing Technology (AGCT) Program, Parco Scientifico e Tecnologico della Sardegna, Pula, Italy. 9I Clinica Pediatria, Ospedale Microcitemico, Cagliari, Italy. 10Department of Clinical and Experimental Medicine, Azienda Ospedaliero Universitaria di Sassari, Sassari, Italy. 11Istituto di Genetica Molecolare, CNR, Pavia, Italy. 12Laboratory of Genetics, National Institute on Aging, US National Institutes of Health, Baltimore, Maryland, USA. 13A full list of members and affiliations appears in the Supplementary Note. 14Medical Research Council (MRC) Integrative Epidemiology Unit, University of Bristol, Bristol, UK. 15Department of Haematology, University of Cambridge, Cambridge, UK. 16Department of Dietetics-Nutrition, Harokopio University, Athens, Greece. 17Porto Conte Ricerche, Tramariglio, Alghero, Italy. 18Porto Conte Ricerche, Tramariglio, Alghero, Italy. 19These authors contributed equally to this work. 20These authors jointly supervised this work. Correspondence should be addressed to M.Z. (magdalena.zoledziewska@irgb.cnr.it), F.C. (fcucca@uniss.it) or D.S. (schlessingerd@mail.nih.gov).
and the other in the KCNQ1 gene (11p15.5), which encode the growth hormone receptor and a voltage-gated potassium channel, respectively (Supplementary Fig. 1c). Notably, the joint effect for these loci in the SardiNIA cohort was as large as the effect contributed jointly by the top ten height-associated alleles assessed in meta-analysis by the GIANT Consortium and by the top five variants when using the effect sizes observed in the replication set.

The first of these signals was rs121909358 (\(P = 1.07 \times 10^{-10}\); effect = −0.64 s.d., corresponding to −4.2 cm; Fig. 1a and Supplementary Fig. 2). The height-reducing T allele was found on a single haplotype (Supplementary Fig. 3). It creates a loss-of-function termination codon in GHR (encoding p.Arg61*). 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), as the variant is extremely rare outside of Sardinia (frequency <1/60,000, 2 mutated alleles detected on 121,388 chromosomes tested by the Exome Aggregation Consortium Browser; see URLs).

Homozgyosity for this stop-gain variant is one of several mutations in GHR known to cause Laron syndrome (Online Mendelian Inheritance in Man (OMIM), 262500), a rare autosomal recessive condition characterized by primary growth hormone insensitivity. Since the initial description of Laron syndrome, more than 250 cases have been reported (Orphanet; see URLs), with the majority of the muscular system and an abnormal degree of humerus rotation (Supplementary Table 2). Interestingly, among the 2,120 sequenced Sardinians, we also found instances of 2 additional rare variants described to cause Laron syndrome in southern European and South American populations (Supplementary Table 3 and Supplementary Note); however, these variants were at frequencies

![Figure 1](https://example.com/figure1)

**Figure 1** Regional plots showing association with height for the GHR and KCNQ1 loci. In each plot, the y axis shows the association strength (−\(\log_{10}(P\) value)) versus the genomic position (hg19/GRCh37) of the most significant SNP (purple) on the x axis. Other SNPs are colored to reflect their LD with the top SNP. Symbols reflect genomic functional annotations. (a) Regional plot at the GHR locus. Genes and the positions of exons are shown below. (b) Regional plots at the KCNQ1 locus for paternal and maternal effects. The position of GWAS catalog SNPs (see URLs) with the corresponding traits and the position of exons in the KCNQ1 region are shown below. TFBS, transcription factor binding site; pQTLs, protein quantitative trait loci; AB peptides, plasma amyloid β peptide concentrations; ECG, electrocardiographic traits.

### Table 1 Parent-of-origin effects at KCNQ1

| rsID          | Chr-position | Minor/other allele | MAF | n    | Both parents | Maternal | Paternal | Effect (SE) | P value | Effect (SE) | P value | Effect (SE) | P value | Heterogeneity | P value |
|---------------|---------------|--------------------|-----|------|-------------|----------|----------|-------------|---------|-------------|---------|-------------|---------|---------------|---------|
| rs150199504   | 11:2814960    | G/C                | 0.083 | 5,059 | −0.168 (0.039) | 1.84 \(\times 10^{-5}\) | −0.315 (0.054) | 5.56 \(\times 10^{-9}\) | 0.0032 | (0.050) | 0.9488 | 2.46 \(\times 10^{-5}\) |
| rs143840904   | 11:2813322    | T/C                | 0.094 | 5,041 | −0.152 (0.038) | 4.58 \(\times 10^{-5}\) | −0.274 (0.050) | 3.92 \(\times 10^{-8}\) | 0.0021 | (0.049) | 0.9054 | 7.55 \(\times 10^{-5}\) |
| rs2075870     | 11:2790091    | A/G                | 0.094 | 5,044 | −0.158 (0.038) | 2.65 \(\times 10^{-5}\) | −0.273 (0.051) | 6.97 \(\times 10^{-8}\) | 0.0172 | (0.048) | 0.7931 | 0.0002 |
| rs149658560   | 11:2767262    | A/G                | 0.076 | 5,050 | −0.161 (0.042) | 1.01 \(\times 10^{-4}\) | −0.297 (0.058) | 2.93 \(\times 10^{-7}\) | 0.0121 | (0.052) | 0.8183 | 0.0003 |
| rs12790610    | 11:2749989    | G/A                | 0.095 | 5,014 | −0.165 (0.037) | 1.02 \(\times 10^{-5}\) | −0.258 (0.051) | 4.73 \(\times 10^{-7}\) | 0.0444 | (0.048) | 0.3531 | 0.0023 |
| rs67004488    | 11:2788704    | Q/A                | 0.104 | 5,026 | −0.157 (0.036) | 1.2 \(\times 10^{-6}\) | −0.244 (0.049) | 5.21 \(\times 10^{-7}\) | 0.0400 | (0.047) | 0.3875 | 0.0024 |

The table summarizes the strongest results for the parent-of-origin association testing at the KCNQ1 locus (defined by association at \(P < 1 \times 10^{-5}\) in either the maternal or paternal test for the assessed 500-kb region). For each SNP, we report the number of informative transmissions used (n; Online Methods) and the association parameters obtained evaluating the minor allele (i) without considering parent of origin, (ii) when the allele was maternally inherited and (iii) when the allele was paternally inherited. The last column reports the \(P\) value for heterogeneity between estimated paternal and maternal effects.
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 was maternally inherited (Fig. 1 and Table 1), with the strongest maternal effect at rs150199504 (MAF = 7.7%; \( P = 5.6 \times 10^{-6} \); maternal effect = −0.315 s.d., corresponding to −1.83 cm) and no significant paternal effect (\( P = 0.95 \)) (Table 1 and 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 \( P < 10^{-6} \), we confirmed also showed a modest, albeit significant association with decreased height in ~90,000 individuals of European origin \( P < 10^{-6} \), we confirmed that differences in allele frequency and LD patterns among the variants in Sardinia in comparison to other populations provided a route to prioritize the list for the responsible locus. Furthermore, the 6 variants showed no significant association with any of the 193 traits measured in the SardiNIA study participants. However, we found that differences in allele frequency and LD pattern among the variants in Sardinia in comparison to other populations provided a route to prioritize the list for the responsible locus. The maternal effect we observed for KCNQ1 on height is consistent with the established monoallelic expression of the maternal alleles at this imprinted locus \( \text{ENCODE}; \text{see URLs} \), hinting at a possible effect on expression. The association signal spanned 48 kb, encompassing rs2075870 and four additional variants in linkage disequilibrium (LD) with rs150199504 (\( P < 10^{-6} \); \( r^2 > 0.7 \)) (Fig. 1 and Table 1), making it difficult to identify the causal variant(s).

To further assess candidacy, we directly tested the 6 core associated variants in 19,053 individuals from 6 European GWAS cohorts, among which we expected more resolving power than in Sardinians owing to lower LD in the region (Fig. 2b,c). Among the five variants that passed quality checks, rs150199504 was again the most significantly associated and had the strongest effect in these samples as well (\( P = 2.82 \times 10^{-4} \); effect = −0.243 s.d.), even though it was the rarest of the five variants (MAF = 0.89%). To a lesser extent, significant association was also seen for rs143840904 (\( P = 1.23 \times 10^{-3} \); effect = −0.145 s.d.) but was not observed for the three 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, there was residual association 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 (Encyclopedia of DNA Elements (ENCODE); see 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 the maternal alleles at this imprinted locus \( \text{ENCODE}; \text{see URLs} \), hinting at a possible effect on expression. The association signal spanned 48 kb, encompassing rs2075870 and four additional variants in linkage disequilibrium (LD) with rs150199504 (\( P < 10^{-6} \); \( r^2 > 0.7 \)) (Fig. 1 and Table 1), making it difficult to identify the causal variant(s).

However, we found that differences in allele frequency and LD pattern among the variants in Sardinia in comparison 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, rs149658560, rs12790610 and rs6704488 from being causal on the basis of their frequencies, LD patterns and results from GWAS in other populations. In particular, these variants are common (MAF ~10%) and in LD with each other (\( r^2 > 0.5 \)) in South Asia, and yet no association of rs2075870 with height has been observed there \( \text{ENCODE}; \text{see URLs} \). By contrast, among our core associated SNPs, the top variants rs150199504 and rs143840904 were in lower LD with rs2075870 and were much rarer in South Asia (\( r^2 < 0.3 \); MAF <1.2% and <2.6%, respectively) (Fig. 2d), and association with height thus could be missed if the associated variants are not directly typed in very large sample sets. Hence, rs143840904 and especially our lead variant rs150199504 are plausible candidate causal variants.

![Figure 2](image-url)
To evaluate the overall impact of known variants on the average short stature observed in Sardinians 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-modifying alleles in a population, weighting each allele by its effect size. A general north-to-south gradient for height in Europe due to directional selection has been reported, with Sardinians representing a significant outlier among the Human Genome Diversity Panel (HGDP) European populations (see URLs). Consistent with these studies, we observed a significantly lower polygenic height score in Sardinians in comparison to other European populations examined in the 1000 Genomes Project, including southern European Tuscans and Spaniards (Fig. 3 and Supplementary Fig. 4). Adding our KCNQ1 and GHR variants to the 691 previously described alleles, the polygenic score for Sardinians decreased by 3.8%. Overall, the Sardinian scores were lower than would be expected in comparison to the scores for other European populations (P = 1.62 × 10^{-6}), −5.9 cm relative to the CEU population (Utah residents of Northern and Western European ancestry); 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 had 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 difference 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 as a result of founder effects. However, the impact of increased levels of homozygosity among Sardinians on height appears to be small (0.129 s.d.) relative to the effect predicted by the polygenic score (0.910 s.d.) (Supplementary Note).

An example of a low-frequency allele affecting height was recently reported in the Icelandic population. However, our findings demonstrate for the first time, to our knowledge, that part of the missing heritability for human height can be attributable to rare variants involved in monogenic disorders, as evidenced by GHR as well as by variants common in isolated populations but rare elsewhere, as exemplified by KCNQ1. Indeed, a shift toward higher frequencies for variants with large effect sizes observed in Sardinia 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. The extinct Sardinian mammoth (Mammuthus lamarmorae) and deer (Megaloceros cazioti) are two examples of this phenomenon. One complication in assessing this in humans is that selection for decreased height likely began before the peopling of Sardinia, among the early European farmer lineage that is thought to have initially colonized the island, and Sardinians might have simply retained the short stature that evolved earlier. However, we observed lower polygenic height scores in Sardinia even in comparisons to other populations with high proportions of early European Neolithic ancestry (Tuscans and Spaniards). Thus, selection for decreased height likely continued and was particularly strong in the lineage leading to present-day 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 Sardinia remains an open and interesting question.
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 cohort). A full list of the investigators who contributed to the generation of the data is available from http://www.UK10K.org/. Funding for UK10K was provided by the Wellcome Trust under award WT091310.

We thank J. Berg for scripts and suggestions on the polygenic score analysis.

AUTHOR CONTRIBUTIONS
M.Z., G.R.A., J.N., D.S. and F.C. 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. and A. Maschio 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. Mulas, F.B., S.U. and R.N. carried out SNP array genotyping. M.Z. and A. Mulas verified genotypes by TaqMan genotyping. J.H.M., C.W.K.C., M.S., F.D.O.D.V., K.E.L. and J.N. performed polygenic score and related population genetic analyses. A. Meloni and A.D. performed clinical characterization of Laron carriers. S.V. provided DNA for the Sardinian replication sample set. F.M., M.P.C., G.B., M.S. and S.S. performed replication analysis. N.S., N.J.T., G.D., I.T., E.Z. and the UK10K Consortium, derived from samples from UK10K_COHORTS_TWINSUK, UK10K_COHORT_ALSPAC (the TwinsUK cohort) and UK10K_COHORTS_TWINSUK, performed polygenic score and related population genetic analyses. The authors declare no competing financial interests.

COMPETING FINANCIAL INTERESTS
The authors declare no competing financial interests.

Reprints and permissions information is available online at http://www.nature.com/reprints/index.html.

1. Pilia, G. et al. Heritability of cardiovascular and personality traits in 6,148 Sardinians. PLoS Genet. 2, e132 (2006).
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 90, 627–630 (2000).
3. Lango Allen, H. et al. Hundreds of variants clustered in genomic loci and biological pathways affect adult human height. Nature 467, 832–838 (2010).
4. Wood, A.R. et al. Defining the role of common variation in the genomic and biological architecture of adult human height. Nat. Genet. 46, 1173–1186 (2014).
5. Francalacci, P. et al. Low-pass DNA sequencing of 1200 Sardinians reconstructs European Y-chromosome phylogeny. Science 341, 565–569 (2013).
6. Sidore, C. et al. Genome sequencing elucidates Sardinian genetic architecture and augments association analyses for lipid and blood inflammatory markers. Nat. Genet. doi:10.1038/ng.3368 (14 September 2015).
7. Arcaleni, L. Secular trend and regional differences in the stature of Italians. J. Anthropol. Sci. 90, 233–237 (2012).
8. Laron, Z. Laron Syndrome—From Man to Mouse (Springer-Verlag, 2011).
9. Laron, Z. The syndrome of familial dwarfism and high plasma immunoreactive human growth hormone. Birth Defects Orig. Artic. Ser. 10, 231–238 (1974).
10. Rosenberg, A.L., Guerera Aguirre, J., Rosenfeld, R.G. & Fielder, P.J. The little women of Loja—growth hormone–receptor deficiency in an inbred population of southern Ecuador. N. Engl. J. Med. 323, 1367–1374 (1990).
11. Laron, Z., Klinger, B., Erster, B. & Silbengeld, A. Serum GH binding protein activities identifies the heterozygous carriers for Laron type dwarfism. Acta Endocrinol. (Copenh.) 121, 603–608 (1989).
12. Guerera-Aguirre, J. et al. Effects of heterozygosity for the 180 splice mutation causing growth hormone receptor deficiency in Ecuador on IGF-I, IGFBP-3, and stature. Growth Horm. IGF Res. 17, 261–264 (2007).
13. Lanktree, M.B. et al. Meta-analysis of dense gene-centric association studies reveals common and uncommon variants associated with height. Am. J. Hum. Genet. 88, 6–18 (2011).
14. Lee, M.P. 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. USA 96, 5203–5208 (1999).
15. Soejima, H. & Higashimoto, K. Epigenetic and genetic alterations of the imprinting disorder Beckwith-Wiedemann syndrome and related disorders. J. Hum. Genet. 58, 402–409 (2013).
16. Horki, M. et al. New loci associated with birth weight identify genetic links between intrauterine growth and adult height and metabolism. Nat. Genet. 45, 76–82 (2013).
17. Johnson, A.D. et al. Genome-wide meta-analyses identifies seven loci associated with platelet aggregation in response to agonists. Nat. Genet. 42, 608–613 (2010).
18. Newton-Cheh, C. et al. Common variants at ten loci influence QT interval duration in the QTGEN Study. Nat. Genet. 41, 399–406 (2009).
19. Voight, B.F. et al. Twelve type 2 diabetes susceptibility loci identified through large-scale association analysis. Nat. Genet. 42, 579–589 (2010).
20. Kong, A. et al. Parental origin of sequence variants associated with complex diseases. Nature 462, 868–874 (2009).
21. Drü, V. et al. Genetic variants regulating immune cell levels in health and disease. Cell 155, 242–256 (2013).
22. Turchin, M.C. et al. Evidence of widespread selection on standing variation in Europe at height-associated SNPs. Nat. Genet. 44, 1015–1019 (2012).
23. Berg, J.J. & Coop, G. A population genetic signal of polygenic adaptation. PLoS Genet. 10, e1004412 (2014).
24. Steinhorsdottir, V. et al. Identification of low-frequency and rare sequence variants associated with elevated or reduced risk of type 2 diabetes. Nat. Genet. 46, 294–298 (2014).
25. Danjou, F. et al. Genome-wide association analyses based on whole-genome sequencing in Sardinia provide insights into regulation of hemoglobin levels. Nat. Genet. doi:10.1038/nrg3307 (14 September 2015).
26. Millen, V. Morphological evolution is accelerated among island mammals. PLoS Biol. 4, e321 (2006).
27. van der Geer, A., Lyras, G., de Vos, J. & Dermitzakis, M. in Genome sequencing elucidates Sardinian genetic architecture and variants associated with elevated or reduced risk of type 2 diabetes. Nat. Genet. 46, 294–298 (2014).
28. Lazaridis, I. et al. Ancient human genomes support three ancestral populations for present-day Europeans. Nature 513, 409–413 (2014).
29. Haak, W. et al. Massive migration from the steppe was a source for Indo-European languages in Europe. Nature 522, 207–211 (2015).
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; see URLs). The study involves four villages—Lanusei, Ilbono, Elini end Arzana—located in the Lanusei valley (21-23). Previously, 6,148 volunteers have been described (24), and an additional 773 individuals have been enrolled during the follow-up stage of the project (25). Of these, 6,602 individuals had complete genotyping data. For analyses, we only included measurements for individuals >20 years of age and discarded 4 subjects with Morquio syndrome (MIM *607939), leading to a total of 6,307 samples.

All participants provided informed consent, and the studies were approved by local research ethic committees: Comitato Etiico di Azienda Sanitaria Locale 8, Lanusei (2009/0016600) and Comitato Etiico di Azienda Sanitaria Locale 1, Sassari (2171/CE).

Genotyping methods, low-pass sample sequencing, variant calling, genotype imputation and GWAS analysis. All individuals in the SardiNIA cohort were typed with four Illumina Infinium arrays. Low-pass sequencing, variant calling, genotype imputation and GWAS analysis were conducted as previously described (26).

GWAS 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, $r^2 \geq 0.6$ for MAF $<0.01$ (ref. 31)), assuming an additive model of inheritance, adjusting for age, age$^2$ and sex as covariates and applying the inverse normal transformation to the residuals. Association analysis was performed using EMMAX (27) as implemented in the software EPACTS (see URLs), which accounts for relatedness and population structure using an empirical kinship matrix derived from genotype data. The genomic control inflation factor was $\lambda = 0.989$, indicating no inflation of the results.

Validation of imputation results by genotyping. GWAS analysis identified three loci significantly associated with stature: the $GHR$ gene, with top variant rs121909358; the KCNQ1 gene, with six variants in LD (Table 1); and the SMURF2 gene, with top variant rs143051029.

We validated imputation of the rs121909358 genotypes by directly genotyping 2,818 samples with a TaqMan assay. The concordance between the imputation and validation genotypes was 99.89%. At KCNQ1, two leading variants, rs67004488 and rs2075870, were present on the Illumina Cardiio-Metabochip, such 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 owing to the high mismatch rate (34.4%) between the imputed genotypes and those generated by Sanger sequencing and was not further considered in the 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 1-Mb region centered on each top SNP (rs121909358 for $GHR$ and rs150199504 for KCNQ1). In both cases, the top variant completely explained the association at the locus; none of the other SNPs in each region passed the significance threshold after Bonferroni correction. The variant chr5:43229441, 540 kb away from rs121909358, was fully explained by the effect of rs121909358 ($P$ value after conditional analysis $=0.1$).

Replication cohorts. We replicated the findings in an independent cohort of 5,314 Sardinian and 19,053 non-Sardinian European samples. Details on genotyping and analyses are provided in the Supplementary Note.

Characterization of the associated region on chromosome 5. To visualize the haplotypes carrying the variant causing Laron syndrome (Supplementary Fig. 3), we interrogated the 6-Mb region centered on chr5:42689036 in 11 sequenced unrelated carriers of rs121909358. The analysis was performed using SelScan (32) and included 9,526 SNPs with MAF $>5\%$ in Sardinia.

Parent-of-origin effects. For SNPs in the KCNQ1 locus, we estimated the parental origin of alleles for all individuals using Merlin (—best option) (33). We then considered two separate variables, one for the maternal allele ($G_m$) and one for the paternal allele ($G_p$), 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 the parental origin of alleles could not be defined unambiguously. Of consequence, the $G_m$ and $G_p$ variables were non-missing for 5,026 individuals in the SardiNIA cohort and 4,666 individuals in the OGP cohort. Two linear models were then used

\[ y - \beta_0 + \beta_G G_m + \beta^T c \]

\[ y - \beta_0 + \beta_G G_p + \beta^T c \]

where $y$ denotes the trait, $\beta_1$ ($\beta_2$) is the effect size for each copy of the minor allele inherited from the mother (father), and $c$ denotes other covariates with effect sizes estimated by $\beta^T$. As both the SardiNIA and OGP study cohorts consist of large families, the transmissions evaluated by $G_m$ and $G_p$ were not independent. We therefore tested the null hypotheses $\beta_1 = 0$ (for model 1) and $\beta_2 = 0$ (for model 2) by fitting a mixed linear regression model that accounts for family relatedness (using the 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

\[ Z_m = \sum_{i=1}^{l} \hat{\beta}_P p_{ml} \]

where $\hat{\beta}_P$ 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 in the source of effect size estimates, we used the effect size estimates from the Sardinian data set regardless of whether the variant was significantly associated with height in this data set. We first calculated the polygenic height score ($Z_m$) on the basis of the 691 height-associated loci identified by the GIANT Consortium (4) with the effect sizes estimated in the Sardinian data set and then added the top 2 variants reported, including 693 height-associated alleles in total. To test whether there was a signature of polygenic adaptation on height in Sardinia, we adopted a framework developed by Berg and Coop (29), 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 ($P = 0.01$) 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-associated loci all variants in the 1000 Genomes Project phase 3 European data with $\pm 10$ counts for the minor allele (frequency $\sim 1\%$), a B score $\leq 50$ units and a local recombination rate $\pm 0.5 \text{ cM/Mb}$. A random subset of 509,386 SNPs, representing 10% of the union of the matched SNPs, was 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 (30) (Supplementary Fig. 4).

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. 23, 975–983 (2015).
32. Kang, H.M. et al. Variance component model to account for sample structure in genome-wide association studies. Nat. Genet. 42, 348–354 (2010).
33. Szpiech, Z.A. & Hernandez, R.D. selscan: an efficient multithreaded program to perform EHH-based scans for positive selection. Mol. Biol. Evol. 31, 2824–2827 (2014).
34. Abecasis, G.R., Cherny, S.S., Cookson, W.O. & Cardon, L.R. Merlin—rapid analysis of dense genetic maps using sparse gene flow trees. Nat. Genet. 30, 97–101 (2002).
35. McVicker, G., Gordon, D., Davis, C. & Green, P. Widespread genomic signatures of natural selection in hominid evolution. PLoS Genet. 5, e1000471 (2009).