Genome-wide association study of anthropometric traits in Korcula Island, Croatia

Citation for published version:

Digital Object Identifier (DOI):
10.3325/cmj.2009.50.7

Link:
Link to publication record in Edinburgh Research Explorer

Document Version:
Publisher's PDF, also known as Version of record

Published In:
Croatian Medical Journal

Publisher Rights Statement:
Copyright © 2009 by the Croatian Medical Journal. All rights reserved. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

General rights
Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy
The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact openaccess@ed.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.
Genome-wide Association Study of Anthropometric Traits in Korčula Island, Croatia

Aim To identify genetic variants underlying six anthropometric traits: body height, body weight, body mass index, brachial circumference, waist circumference, and hip circumference, using a genome-wide association study.

Methods The study was carried out in the isolated population of the island of Korčula, Croatia, with 898 adult examinees who participated in the larger DNA-based genetic epidemiological study in 2007. Anthropometric measurements followed standard internationally accepted procedures. Examinees were genotyped using HumanHap 370CNV chip by Illumina, with a genome-wide scan containing 316,730 single nucleotide polymorphisms (SNP).

Results A total of 11 SNPs were associated with the investigated traits at the level of \( P < 10^{-5} \), with one SNP (rs7792939 in gene zinc finger protein 498, \( \text{ZNF498} \)) associated with body weight, hip circumference, and brachial circumference \( \left( P = 3.59 \times 10^{-5} \right) \), and another one (rs157350 in gene delta-sarcoglycan, \( \text{SGCD} \)) with both brachial and hip circumference \( \left( P = 3.70 \times 10^{-5} \right) \). Variants in \( \text{CRIM1} \), a gene regulating delivery of bone morphogenetic proteins to the cell surface, and \( \text{ITGA1} \), involved in the regulation of mesenchymal stem cell proliferation and cartilage production, were also associated with brachial circumference \( \left( P = 7.82 \times 10^{-6} \right) \) and \( \left( P = 9.68 \times 10^{-6} \right) \), respectively and represent interesting functional candidates. Other associations involved those between genes \( \text{SEZ6L2} \) and \( \text{MAX} \) and waist circumference, \( \text{XTP6} \) and brachial circumference, and \( \text{AMPK} \) and height.

Conclusion Although the study was underpowered for the reported associations to reach formal threshold of genome-wide significance under the assumption of independent multiple testing, the consistency of association between the 2 variants and a set of anthropometric traits makes \( \text{CRIM1} \) and \( \text{ITGA1} \) highly interesting for further replication and functional follow-up. Increased linkage disequilibrium between the used markers in an isolated population makes the formal significance threshold overly stringent, and changed allele frequencies in isolate population may contribute to identifying variants that would not be easily identified in large outbred populations.

Received: December 18, 2008
Accepted: January 19, 2009

Correspondence to: Igor Rudan, Croatian Centre for Global Health University of Split Medical School Šoltanska 2 21000 Split, Croatia irudan@hotmail.com
Interest in anthropometric traits has a long history, ever since the ancient philosophical debates and pioneering attempts to understand human anatomy (1). Modern research on these traits has largely shifted from fundamental morphological anatomy toward the understanding of environmental and genetic factors which determine, affect, and modify these traits. While environmental factors are known to have an important effect on some of these traits, primarily weight and other weight-related traits (2), genetic factors remained elusive to researchers for a long time. Population genetics theory hypothesized, based on the pre-genomic knowledge and comparative studies in plants and animals (3), that majority of these traits will be complex and have highly polygenic background (2,4,5). Additionally, some of these traits have been reported to have very high heritability, with that of height being used as the example of an extremely heritable trait (generally in the range of 0.90 to 0.95) (2). However, the search for the genetic variants underlying height has long been futile, casting doubt on the fundamental assumption that highly heritable traits will be good candidates for gene mapping studies. It is only lately that very large consortia were formed and achieved the needed statistical power to identify the first candidate loci using genome-wide association studies and expanded our understanding of the determinants of human stature (6-11).

The growing world-wide epidemic of obesity is one of the major issues in modern public health. Search for genes underlying obesity has also been rather unsuccessful, due to an even greater degree of complexity than for height, with large environmental effects and variation over time (12). The search for obesity genes is further complicated by the differences in the central and peripheral type of obesity (13,14). The genetic background of skinfolds has largely been under-investigated (15), with only a few genes implied in their regulation (15,16), some of which do seem to show strong interaction with the environment (17,18). Skinfold measurements may also show variation in time (19), making this research area highly complex and usually considered secondary to the more classical estimates of the amount of fat tissue, including body mass index and waist circumference or waist-to-hip ratio.

In this article, we report on a comprehensive genome-wide association study of height, weight, body mass index, and three circumference measurements – brachial, waist, and hip – in the isolated population of the island of Korčula in Croatia. The study is a part of a larger genetic epidemiology research program in Croatian island isolates, “10,001 Dalmatians.” The genetic epidemiology research program in Croatian island isolates began in 1999 (20,21), then expanded to study human genetic variation and effects of isolation and inbreeding (22-29), and finally entered the phase of focusing on diseases and gene mapping studies (8,30-35). By now, the research project has included more than 3000 examinees from isolated populations, and eventually it aims to reach 10,001 examinees.

**SUBJECTS AND METHODS**

This study was carried out in the adult population of the island of Korčula, Croatia. The field work was performed in 2007 in the eastern parts of the island, targeting healthy volunteers from the town of Korčula and villages Lumbarđa, Žrnovo, and Račišće (Figure 1). Participants were invited by mail, posters, radio, and personal contacts. The sampling scheme for this study was convenient sampling, as the study aim was to include approximately 1000 of adult island inhabitants for the purpose of the genome-wide association analysis, regardless on the sample representativeness and demographic structure.

All examinees were aged 18 and over and had signed informed consent before entering the study, which was approved by the Ethical Committee of the Medical School, University of Zagreb. A total of 944 examinees were included in the study between March and December 2007. In all examinees, a large number of quantitative phenotypic traits were measured, including height, weight (and derived body mass index), and brachial, waist, and hip circumference. For these anthropometric measurements, standard methods were used (36).

DNA extraction was performed using Nucleon kits (Tepnel, Manchester, UK) and a total of 944 samples were genotyped in Institute of Human Genetics, Helmholtz Zentrum München, Germany. Genotyping was performed using Illumina HumanHap 370CNV (Illumina, San Diego, CA, USA), with a total of 346,027 single nucleotide polymorphism (SNP) markers. Quality control of the genotype...
data, excluding markers with a call rate <98%, with minor allele frequency <2%, or out of Hardy-Weinberg equilibrium ($P < 10^{-10}$), left 316 730 SNPs in the analysis. Exclusion of individuals carrying markers with a call rate of <98% and those sampled twice in the study, left 898 people for inclusion in the analysis.

Relatedness between examinees was estimated from their whole genome data using the sharing of genome identical by descent (IBD) in PLINK (37). This method is robust to pedigree information errors, undeclared relationships, and samples swap, and gives realized sharing rather than expected one given by the pedigree information (for the same pedigree-based relationship, sharing varies within a range due to segregation and recombination stochastic events). Using this function, the sample could be grouped into 154 parent-child pairs, 107 sibling pairs, avuncular or equivalent 137, first cousins or equivalent 363, first cousins once removed or equivalent 1861, second cousins or equivalent 8962. The mean IBD sharing between all possible pairs of individuals was 0.003 (min = 0, max = 0.61).

Genome-wide associations between anthropometric phenotypes (adjusted for age and sex) and SNP markers were analyzed using the “mmscore” function of the GenABEL R statistical package (38), using an additive model. This score test for family-based association accounts for pedigree structure and allows unbiased estimations of SNP allelic effect (39). No correction for multiple testing was applied. The relationship matrix used in this analysis was generated by the identity-by-state (ibs) function of GenABEL, which used ibs genotype sharing to determine the realized pair-wise kinship coefficient similarly to the PLINK genome function. Minor allele frequencies were reported for the identified SNPs. Weighted genomic inbreeding coefficient was used in the inbreeding estimation (23). All identified SNPs that reached significance at the level of $P < 10^{-5}$ were visualized using HaploView software (MIT/Harvard Broad Institute, MA, USA).

**TABLE 1.** Descriptive statistics (mean ± standard deviation, SD) of the investigated traits of the examinees (n = 898) from the Korcula island.

<table>
<thead>
<tr>
<th>Trait</th>
<th>Mean ± SD</th>
<th>Range</th>
</tr>
</thead>
<tbody>
<tr>
<td>Height (mm)</td>
<td>1680.9 ± 90.8</td>
<td>1410.0-1970.0</td>
</tr>
<tr>
<td>Weight (kg)</td>
<td>79.1 ± 14.2</td>
<td>49.5-166.6</td>
</tr>
<tr>
<td>Body mass index (kg/m²)</td>
<td>27.9 ± 4.1</td>
<td>16.6-53.8</td>
</tr>
<tr>
<td>Brachial circumference (mm)</td>
<td>333.4 ± 45.0</td>
<td>200.0-944.0</td>
</tr>
<tr>
<td>Hip circumference (mm)</td>
<td>1040.6 ± 78.6</td>
<td>748.0-1563.0</td>
</tr>
<tr>
<td>Waist circumference (mm)</td>
<td>941.4 ± 121.5</td>
<td>620.0-1385.0</td>
</tr>
</tbody>
</table>

Genome-wide association study of height using Haploview software, showing peaks on chromosomes 1 and 5 reaching genome-wide significance level of $P < 10^{-5}$. 

Figure 2.
RESULTS

Genetic variants associated with height, weight, and body mass index

A total of 898 examinees were genotyped and included in the analysis. Descriptive statistics for the investigated traits are presented in Table 1. The average inbreeding coefficient was 0.013 ± 0.025. Initial genome-wide association study revealed two SNPs associated with body height at the level of $P < 10^{-5}$ on chromosomes 1 and 5. The variant at chromosome 1 ($P = 4.81 \times 10^{-5}$) did not have a gene in proximity (±100 kilobases). The variant on chromosome 5 ($P = 9.77 \times 10^{-6}$) implicated a gene ionotropic glutamate receptor AMPA1 (GRIA1) (Table 2, Figure 2).

Two further SNPs were associated with weight, one belonging to hypothetical protein LOC122509 (FAM14B) gene on chromosome 14, while the other was within zinc finger protein 498 gene (ZNF498) on chromosome 7 (Table 2, Figure 3). The $P$ value for the former variant was $3.13 \times 10^{-6}$ and $3.59 \times 10^{-6}$ for the latter.

Body mass index was associated with a single SNP on the chromosome 2 that was not associated with any gene (Table 2, Figure 4), with a $P$ value of $5.76 \times 10^{-6}$.

Genetic variants associated with brachial, hip, and waist circumference

Brachial circumference showed association with 6 SNPs at chromosomes 2, 5, 7, and 13. Two of the SNPs (rs157350 and

<table>
<thead>
<tr>
<th>Trait</th>
<th>SNP</th>
<th>Chromosome</th>
<th>Position</th>
<th>$P$ value (× 10$^{-6}$)</th>
<th>Minor allele frequencies</th>
<th>Standard error (β)</th>
<th>Position</th>
<th>Gene</th>
</tr>
</thead>
<tbody>
<tr>
<td>Height</td>
<td>rs7513590</td>
<td>1</td>
<td>5070572</td>
<td>4.81</td>
<td>0.104</td>
<td>-0.316</td>
<td>0.069</td>
<td>no gene</td>
</tr>
<tr>
<td>Height</td>
<td>rs12658202</td>
<td>5</td>
<td>152959747</td>
<td>9.77</td>
<td>0.453</td>
<td>0.110</td>
<td>0.025</td>
<td>intronic glutamate receptor, ionotropic, AMPA 1 (GRIA1)</td>
</tr>
<tr>
<td>Weight</td>
<td>rs7792939</td>
<td>7</td>
<td>99045812</td>
<td>3.59</td>
<td>0.148</td>
<td>0.255</td>
<td>0.055</td>
<td>7kb 5' zinc finger protein 498 (ZNF498)</td>
</tr>
<tr>
<td>Weight</td>
<td>rs7157940</td>
<td>14</td>
<td>93632946</td>
<td>3.13</td>
<td>0.458</td>
<td>0.118</td>
<td>0.025</td>
<td>5' UTR hypothetical protein LOC122509 (FAM14B)</td>
</tr>
<tr>
<td>Body mass index</td>
<td>rs7590983</td>
<td>2</td>
<td>174070595</td>
<td>5.76</td>
<td>0.045</td>
<td>0.462</td>
<td>0.102</td>
<td>no gene</td>
</tr>
<tr>
<td>Brachial circumference</td>
<td>rs1863080</td>
<td>2</td>
<td>36379940</td>
<td>7.82</td>
<td>0.102</td>
<td>0.307</td>
<td>0.069</td>
<td>cysteine-rich motor neuron 1 (CRIM1)</td>
</tr>
<tr>
<td>Brachial circumference</td>
<td>rs7723398</td>
<td>5</td>
<td>52264862</td>
<td>9.68</td>
<td>0.222</td>
<td>0.191</td>
<td>0.043</td>
<td>intronic integrin alpha 1 precursor (ITGA1)</td>
</tr>
<tr>
<td>Brachial circumference</td>
<td>rs32056</td>
<td>5</td>
<td>156006717</td>
<td>9.07</td>
<td>0.093</td>
<td>-0.325</td>
<td>0.073</td>
<td>intronic delta-sarcoglycan (SGCD)</td>
</tr>
<tr>
<td>Brachial circumference</td>
<td>rs157350</td>
<td>5</td>
<td>156072147</td>
<td>3.70</td>
<td>0.105</td>
<td>-0.311</td>
<td>0.067</td>
<td>intronic delta-sarcoglycan (SGCD)</td>
</tr>
<tr>
<td>Brachial circumference</td>
<td>rs7792939</td>
<td>7</td>
<td>99045812</td>
<td>4.90</td>
<td>0.148</td>
<td>0.255</td>
<td>0.056</td>
<td>7kb 5' zinc finger protein 498 (ZNF498)</td>
</tr>
<tr>
<td>Brachial circumference</td>
<td>rs201789</td>
<td>13</td>
<td>49884119</td>
<td>6.20</td>
<td>0.279</td>
<td>0.170</td>
<td>0.038</td>
<td>XTP6</td>
</tr>
<tr>
<td>Hip circumference</td>
<td>rs157350</td>
<td>5</td>
<td>156072147</td>
<td>6.08</td>
<td>0.105</td>
<td>-0.306</td>
<td>0.068</td>
<td>intronic delta-sarcoglycan (SGCD)</td>
</tr>
<tr>
<td>Hip circumference</td>
<td>rs7792939</td>
<td>7</td>
<td>99045812</td>
<td>5.73</td>
<td>0.148</td>
<td>0.256</td>
<td>0.056</td>
<td>7kb 5' zinc finger protein 498 (ZNF498)</td>
</tr>
<tr>
<td>Waist circumference</td>
<td>rs7158173</td>
<td>14</td>
<td>64715003</td>
<td>3.93</td>
<td>0.471</td>
<td>-0.114</td>
<td>0.025</td>
<td>76kb 5' MAX</td>
</tr>
<tr>
<td>Waist circumference</td>
<td>rs4787483</td>
<td>16</td>
<td>29792948</td>
<td>2.10</td>
<td>0.343</td>
<td>0.164</td>
<td>0.035</td>
<td>intronic Seizure-related 6 homologue (mouse)-like 2 isoform (SEZ6L2)</td>
</tr>
</tbody>
</table>

*The table summarizes SNPs, their positions on the chromosomes, $P$ values, minor allele frequencies, effect size and direction (expressed as $\beta$ and standard error of $\beta$), effect allele, and implicated gene.
Figure 3.

Genome-wide association study of weight using Haploview software, showing peaks on chromosomes 7 and 14 reaching genome-wide significance level of $P < 10^{-5}$.

Figure 4.

Genome-wide association study of body mass index using Haploview software, showing a peak on chromosome 2 reaching genome-wide significance level of $P < 10^{-5}$.
Figure 5.

Genome-wide association study of brachial circumference using Haploview software, showing peaks on chromosomes 2, 5, 7, and 13 reaching genome-wide significance level of $P < 10^{-5}$.

Figure 6.

Genome-wide association study of hip circumference using Haploview software, showing peaks on chromosomes 5 and 7 reaching genome-wide significance level of $P < 10^{-5}$. 
rs7792939) implicated delta-sarcoglycan gene (SGCD). The SNP with the stronger effect (rs157350) of \( P = 3.70 \times 10^{-6} \) was also associated with the hip circumference \( (P = 6.08 \times 10^{-6}) \). Both brachial and hip circumference showed associations with the SNP rs7792939 in zinc finger protein 498 (ZNF498) on chromosome 7 (Table 2, Figure 5 and 6), with significance of \( P = 4.90 \times 10^{-6} \) and \( P = 5.73 \times 10^{-6} \), respectively. The same SNP showed strong association with body weight. Other variants associated with brachial circumference included rs1863080 in cysteine-rich motor neuron 1 (CRIM1) gene on chromosome 2 \( (P = 7.82 \times 10^{-6}) \), rs7723398 in integrin alpha 1 precursor (ITGA1) gene on chromosome 5 \( (P = 9.68 \times 10^{-6}) \), and rs201789 in XTP6 gene on chromosome 13 \( (P = 6.20 \times 10^{-6}) \).

Finally, waist circumference was associated with two SNPs, one on chromosome 14 (rs7158173) belonging to the MAX gene \( (P = 3.83 \times 10^{-6}) \) and the other on chromosome 16 (rs4787483) belonging to seizure-related 6 homologue (mouse)-like 2 isoform gene (SEZ6L2) \( (P = 2.10 \times 10^{-6}) \) (Table 2, Figure 7).

**DISCUSSION**

The results of this study suggest several potential candidate genes involved in determining height, weight, and body mass index, as well as the hip, waist, and brachial circumference. We identified novel genetic associations with 6 human anthropometric traits using an isolate population. In isolated populations, genome-wide association studies can be particularly useful because of several potential advantages. First, allele frequencies of some important variants may be skewed from those found in the general population due to a combination of population genetic effects such as founder effect, genetic drift, and inbreeding (2). This could dramatically increase the power of genome-wide association study to detect otherwise rare variants of reasonably large effects on phenotypes of interest, if their frequency is unusually increased in an isolate population, a suggestion that has been proposed for some isolated Croatian Adriatic islands (22). In addition, accepted methods of computation of formal threshold for genome-wide statistical significance that corrects for multiple testing performed for hundreds of thousands of genetic markers may be overly stringent in isolate populations because of their population genetic properties (22). Increased linkage disequilibrium in such populations makes many of the markers linked so that many of the performed tests are dependent on each other (linkage disequilibrium works against the assumption of the independence between all performed tests) (31), which in turn mathemati-
cally lowers the threshold for genome-wide significance for an unknown but substantial amount. Lowering the significance threshold means greater power for the studies performed in isolate populations. Due to the two mechanisms described above, it would not be unexpected that genetic variants underlying complex traits are identified in human isolate populations rather than in outbred population by using genome-wide scans of comparable density and much larger sample sizes (5,8).

In recent years, genome-wide association studies have proven their worth as an extremely powerful tool in identifying genetic variants underlying many human diseases and complex traits (40). Hundreds of genetic variants are presently being discovered by large consortia of scientists in pioneering efforts to use this reliable approach to provide hypothesis-free insights into the genetic architecture of complex human phenotypes (36). Our study succeeded in identifying 11 genetic loci that reached significance level of $P < 10^{-5}$. Although this does not make all of them necessarily genome-wide significant in formal terms, we believe that they should be carefully considered for replication in other populations and further functional studies due to combination of factors described above.

The gene AMPA1 (GRIA1) implicated in height in our study has also been implicated in stimulus-reward learning in mice (41,42) and possibly in bipolar disorder and schizophrenia in humans (43,44), but possible mechanisms remain obscure, and the finding has yet to be replicated. It has not been associated with human height to date. However, human height has been strongly inversely associated with the risk of schizophrenia (45), making this finding potentially interesting for further replication and functional follow-up and replication in other populations in order to establish their true significance in determining human body size and shape.

Unlike the previous candidate genes, no plausible information is currently available on possible roles and functions of XTP6, SEZ6L2, and MAX genes that would help us understand their possible role in determining human body size and shape.

The results presented here suggest that some identified genes were implicated in more than one trait, suggesting that they might be responsible for the genetic action on the common underlying property for these traits. Other genes, such as CRIA1 and ITGA1, represent very promising and biologically plausible functional candidates.

The shortcomings of this study primarily include the potential low statistical power, which is a consequence of the limited population size encountered in any genetic isolate. Studies performed in isolated populations should always seek replication to ensure that the findings are indeed representative of wider general human populations and not limited to specific circumstances of a particular isolate. Finally, variants identified in this study all require functional follow-up and replication in other populations in order to establish their true significance in determination of human body size and shape.

Acknowledgment
A. M. is the Editor in Chief, I. R. is the Editor for the International Health Issues, and Z. B. is the Cover Page Editor of the Croatian Medical Journal. S. J. is the Dean of the University of Split School of Medicine, one of the owners of the journal. To ensure that any possible conflict of interest has been addressed, this article was
reviewed according to best practice guidelines of international editorial organizations.

This work was supported by the grant 108-1080315-0302 to IR and several other grants from the Croatian Ministry for Science, Education and Sport to Croatian co-authors. It was also supported by the grants from the Medical Research Council UK to HC, AFW and IR; and European Commission FP6 STREP grant number 018947 (LSHG-CT-2006-01947) to HC. The authors collectively thank to very large number of individuals for their individual help in organizing, planning and carrying out the field work related to the project and data management.”

References


36 Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet. 2007;81:559-75. doi:10.1086/519795


