Genome-Wide Association of Serum Uric Acid Concentration: Replication of Sequence Variants in an Island Population of the Adriatic Coast of Croatia

Rebekah Karns1,†, Ge Zhang2,†, Guangyun Sun1, Subba Rao Indugula1, Hong Cheng1, Dubravka Havas-Augustin3, Natalija Novokmet3, Dusko Rudan3, Zijad Durakovic3, Sasa Missoni3, Ranjit Chakraborty4, Pavao Rudan3 and Ranjan Deka1,*

1Center for Genome Information, Department of Environmental Health, University of Cincinnati, Cincinnati, OH, USA
2Human Genetics Division, Cincinnati Children’s Hospital, Cincinnati, OH, USA
3Institute for Anthropological Research, Zagreb, Croatia
4Center for Computational Genomics, University of North Texas Health Science Center, Forth Worth, TX, USA

Summary

A genome-wide association study of serum uric acid (SUA) levels was performed in a relatively isolated population of European descent from an island of the Adriatic coast of Croatia. The study sample included 532 unrelated and 768 related individuals from 235 pedigrees. Inflation due to relatedness was controlled by using genomic control. Genetic association was assessed with 2,241,249 single nucleotide polymorphisms (SNPs) in 1300 samples after adjusting for age and gender. Our study replicated four previously reported SUA loci (SLC2A9, ABCG2, RREB1, and SLC22A12). The strongest association was found with a SNP in SLC2A9 (rs13129697, P = 2.33 × 10^{-19}), which exhibited significant gender-specific effects, 35.76 μmol/L (P = 2.11 × 10^{-19}) in females and 19.58 μmol/L (P = 5.40 × 10^{-5}) in males. Within this region of high linkage disequilibrium, we also detected a strong association with a nonsynonymous SNP, rs16890979 (P = 2.24 × 10^{-17}), a putative causal variant for SUA variation. In addition, we identified several novel loci suggestive of association with uric acid levels (SEMA5A, TMEM18, SLC28A2, and ODZ2), although the P-values (P < 5 × 10^{-6}) did not reach the threshold of genome-wide significance. Together, these findings provide further confirmation of previously reported uric-acid-related genetic variants and highlight suggestive new loci for additional investigation.

Introduction

Genome-wide association studies (GWAS) have uncovered over 30 common sequence variants influencing serum uric acid (SUA) concentration and gout (Hindorff et al., 2011). Among these, the most significant findings are the single-nucleotide polymorphisms (SNPs) located within the solute carrier family 2 member 9 (SLC2A9) gene on chromosome 4, which have been consistently replicated across multiple populations (Charles et al., 2011; Dehghan et al., 2008; Döring et al., 2008; Kolz et al., 2009; Li et al., 2007; McArdle et al., 2008; Vitart et al., 2008; Wallace et al., 2008; Yang et al., 2010). Additional GWAS and meta-analyses have identified variants in several other genes including PDZK1, GCKR, ABCG2, RREB1, LRRC16A, SLC17A1, SLC17A3, SLC22A11, and SLC22A12 that have reached genome-wide significance levels (Dehghan et al., 2008; Kolz et al., 2009; Yang et al., 2010). We conducted a GWAS of metabolic traits, including SUA concentration, in a relatively isolated population from the Adriatic coast of Croatia. This study strongly replicated the SLC2A9 findings and identified several suggestive novel loci that may represent genuine effects. In addition, our study also replicated associations of SNPs in ABCG2, RREB1, and SLC22A12, although the signals did not reach genome-wide significance.
Materials and Methods

Subjects

The study population has been described previously (Karns et al., 2011; Zhang et al., 2010). Briefly, participants were derived from the middle Dalmatian island of Hvar on the eastern Adriatic coast of Croatia. The population is primarily of Slavic descent, which had emigrated from the mainland before the 18th century and remained relatively isolated since that time (Rudan et al., 1992). Phenotypic measures and blood samples were collected in two field surveys conducted in May 2007 and May 2008, with no consideration of disease status or medication. Blood samples were collected following an overnight fast, and SUA levels were measured using the enzymatic color method. In total, 1395 related and unrelated subjects aged >20 years with SUA measures were included in the current study. Descriptive statistics of quantitative traits (age, body mass index, fasting plasma glucose, and blood pressure) and prevalence of four metabolic disorders (type 2 diabetes, hypertension, gout, and metabolic syndrome) are provided in Supplementary Table S1. Data on type 2 diabetes, gout, and hypertension were collected through self-reports, medical review, and clinical diagnostic measures. The study was approved by the Ethics Committee of the Institute for Anthropological Research in Zagreb, Croatia, and the Institutional Review Board of the University of Cincinnati.

Genotyping

Genome-wide SNP genotyping was performed using the Affymetrix Human SNP Array 5.0 following the manufacturer’s protocol. Genotype calls were determined using the CRLMM algorithm (Carvalho et al., 2007, 2010) among chips that passed the prescribed Dynamic Model genotyping QC call rate (>0.86). Following further QC filtering of the genotype data (MAF > 0.02, HWE P > 0.0001, call rate >95%) using the check.marker function implemented in GenABEL (Aulchenko et al., 2007), we obtained a cleaned data set of 344,512 SNPs in 1300 samples (563 males and 737 females). From this cleaned data set, we performed genotype imputation using MACH (Li et al., 2009) and the reference haplotype data from the Phase II CEU HapMap (International HapMap Consortium, 2007). The same QC procedures were performed on the imputed data, yielding a final genotype data set of 2,241,249 SNPs in 1300 samples.

Statistical Analysis

All statistical analyses were performed in R v2.11; genome-wide association analysis was performed using the GenABEL package (v.1.6). Single-locus tests adjusted for age and sex were conducted using the gtscore routine. Since our samples included 532 unrelated as well as 768 related individuals from 235 families, genomic control (GC) was applied to correct for inflation due to inclusion of related individuals (Devlin & Roeder, 1999; Devlin et al., 2001). The inflation factor (λ) was estimated using the median method (Bacanu et al., 2002), and P-values based on the adjusted test statistics (1 d.f. assuming additive effects) were reported. Association signals of significant regions were plotted using LocusZoom (Pruim et al., 2010).

Results

SUA levels were normally distributed in both males (N = 563) and females (N = 737) and the mean levels were significantly higher in males (361.0±79.19 μmol/L) than females (265.4±77.65 μmol/L), though Bartlett’s and Fligner’s tests revealed no significant gender-based differences in SUA variance. Regression analysis indicated that SUA levels were significantly associated with age in both genders and the association was more significant in females. SUA change per year in females was 1.797 μmol/L (P = 2.2×10⁻²⁴, r² = 13.1%) and in males was 0.786 μmol/L (P = 0.00019, r² = 2.28%) (Supplementary Fig. S1).

As anticipated from the relatedness among the samples, the test statistics were inflated compared to the null distribution with an estimated inflation factor λ = 1.20. As shown in the quantile–quantile (QQ) plot (Supplementary Fig. S2), after adjustment for this inflation factor and exclusion of significant SNPs in the SLC2A9 region, the test statistics fit well with the expected values, indicating appropriate control of false positive rate.

A Manhattan plot of the genome-wide association signals (Fig. 1) shows the strongest association around the SLC2A9 gene, a well-established uric-acid-associated gene. The significant region spans roughly 650 kb and covers the SLC2A9 and the WDR1 genes (Fig. 2). This region is delimited by two recombination hot spots with local recombination rate >25 cM/Mb. One-hundred sixteen SNPs with P-value less than the genome-wide significant level (5×10⁻⁸) were identified within this region. The strongest signal was found on the imputed SNP rs13129697 (P = 2.33×10⁻¹⁹); the minor allele was associated with an average SUA decrease of 28.99 μmol/L. Consistent with previous studies, the effect size estimate showed substantial gender difference with 35.76 μmol/L (P = 2.11×10⁻¹⁹) in females and 19.58 μmol/L (P = 5.40×10⁻⁵) in males. This pattern was similar across all of the 116 SNPs that reached genome-wide significance. Of interest is a nonsynonymous SNP, rs16890979 (P = 2.24×10⁻¹⁷), also reported previously with...
Figure 1  Manhattan plot of GWA single-locus $P$-values. The two horizontal dash lines indicate significant thresholds at $5 \times 10^{-8}$ and $5 \times 10^{-6}$. Six regions that reach suggestive genome-wide significance ($P < 5 \times 10^{-6}$) are highlighted with names of nearby genes. Gene names in black are previously reported uric acid associated genes.

Figure 2  LocusZoom plot of the SLC2A9 region. GC-adjusted single-locus $P$-values are plotted against SNP physical positions (NCBI build 36). Pairwise linkage disequilibrium ($r^2$) from the most significant SNP (rs13129697) is color-coded. Size of each dot indicates whether the SNP is a genotyped (large) or imputed (small). The light blue curve shows the local recombination rate based on HapMap Phase II data. rs16890979 is the nonsynonymous SNP and rs874432 is the most significant genotyped SNP.

In addition to the well-established SLC2A9 region, five other potentially significant regions (Table 1) were identified with at least one SNP above a threshold of $P < 5 \times 10^{-6}$. The most salient of these is a suggestive novel locus in an intergenic region on chromosome 5; the most significant variant, rs200113 ($P = 7.02 \times 10^{-8}$), is located $\sim 400$ kb genome-wide significance (Dehghan et al., 2008; McArdle et al., 2008). Reanalysis of the region, conditional on either rs13129697 or rs16890979, failed to completely abolish the signals of the other SNPs with the smallest conditional $P \sim 1.7 \times 10^{-3}$ (data not shown), suggesting the possibility of multiple functional variants within the region.
downstream of the **SEMA5A** gene. The remaining four regions are located within or near the genes **TMEM118**, **SLC28A2**, **ODZ2**, and **ABCG2**, respectively. **ABCG2** is a confirmed uric-acid-associated gene (Dehghan et al., 2008; Kolz et al., 2009; Yang et al., 2010) and the variant, showing the highest signal \( (P = 5.14 \times 10^{-6}) \) is a nonsynonymous SNP (rs2231142, NP_004818.2, Gln141Lys) and the mutant allele associated with an average increase of 27.4 μmol/L (Supplementary Fig. S3). This SNP showed significant gender-specific effects, with 31.11 μmol/L in males compared to 22.97 μmol/L in females. Reanalysis conditional on this SNP explained all the association across the region, which is highly suggestive of this missense variant being the functional SNP at the **ABCG2** locus.

To compare our findings with previous GWA studies, we analyzed 30 serum urate- or uric-acid-associated SNPs listed in the GWA catalog (six of the 36 reported SNPs were missing from our imputed and cleaned data set) (Table 2). In addition to the aforementioned **SLC2A9** and **ABCG2** loci, we replicated associations with nominal significance \( (P < 0.05) \) at two additional loci, **RREB1** (rs675209, \( P = 0.0032 \)) and **SLC22A12** (rs17300741, \( P = 0.0034 \)). The effects of all significant SNPs were in the same direction as those reported in previous GWA studies.

Discussion

We present the results of a GWAS of SUA in an isolated island population based on 2,241,249 imputed and genotyped SNPs in 1300 samples. Our purpose was to replicate previously reported loci and uncover novel SUA-related loci, taking advantage of population attributes of limited admixture and homogeneous environmental exposures. We have used GC to provide correction for inflation due to relatedness while maximizing power to detect associations by including all samples. Previous study indicated that GC is a valid and powerful method for the analysis of pedigree-based quantitative trait loci (Amin et al., 2007).

The most significant associations emerged from multiple SNPs in and around **SLC2A9** on chromosome 4, a widely replicated SUA-associated region. The SNP with the strongest signal, rs13129697, is located in intron 7 of the gene. Of particular interest, however, was the association of rs16890979 (Val253Ile), a nonsynonymous imputed SNP that has been reported in previous GWAS (Dehghan et al., 2008; McArdle et al., 2008).

We performed a comparative analysis of previously reported per-allele effect sizes of the significant SNPs in **SLC2A9** and found that, in general, our effect sizes are somewhat higher than those reported in previous GWAS (Dehghan et al., 2008; Döring et al., 2008; Kolz et al., 2009; McArdle et al., 2008; Yang et al., 2010; Zemunik et al., 2009) (Supplementary Fig. S4). Across studies, **SLC2A9** variant effect sizes in females are markedly elevated compared to males. In our population, we found that males had significantly higher mean SUA concentrations, though female SUA concentration was more strongly associated with age. Sex-specific effects of **SLC2A9** variants were more extensively examined by Döring et al. (2008), who showed that in addition to genotypic effects, **SLC2A9** expression levels were stronger in females. Together these observations suggest that the **SLC2A9** variants may play a more significant role influencing uric acid concentrations in females, which could be due to physiological and vascular differences between males and females and due to decreased uricosuric-related estrogen action following

---

Table 1 Summary of the significant SNPs and their regions.

<table>
<thead>
<tr>
<th>Chr</th>
<th>Start</th>
<th>End</th>
<th>Gene(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>9401384</td>
<td>10050458</td>
<td>SLC2A9, WDR1</td>
</tr>
<tr>
<td>5</td>
<td>8654931</td>
<td>8870372</td>
<td>SEMA5A</td>
</tr>
<tr>
<td>2</td>
<td>628483</td>
<td>730659</td>
<td>TMEM18</td>
</tr>
<tr>
<td>15</td>
<td>43262339</td>
<td>43621220</td>
<td>SLC28A2</td>
</tr>
<tr>
<td>5</td>
<td>167385703</td>
<td>167478088</td>
<td>ODZ2</td>
</tr>
<tr>
<td>4</td>
<td>89151747</td>
<td>89313430</td>
<td>ABCG2</td>
</tr>
</tbody>
</table>

**Note:** Position is based on NCBI Genome Build 36; A1 and A2 are the major and minor alleles, respectively; Eff-All, Eff-M, and Eff-F are effect sizes (μmol/L) of the minor alleles in total, male and female samples. A significant region was selected starting with an “index” SNP (rs2231142, 5.14 × 10−10), a nonsynonymous variant, which is highly suggestive of this missense variant being the functional SNP at the **ABCG2** locus.
Table 2  Genome-wide replication of previously reported GWA SNPs.

<table>
<thead>
<tr>
<th>SNP</th>
<th>Chr</th>
<th>Position</th>
<th>A1</th>
<th>A2</th>
<th>MAF</th>
<th>P-value</th>
<th>Eff</th>
<th>Gene</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs1967017</td>
<td>1</td>
<td>144435002</td>
<td>C</td>
<td>T</td>
<td>0.48</td>
<td>0.451555</td>
<td>2.46</td>
<td>PDZK1</td>
</tr>
<tr>
<td>rs12129861</td>
<td>1</td>
<td>144437046</td>
<td>A</td>
<td>G</td>
<td>0.49</td>
<td>0.405643</td>
<td>2.75</td>
<td>GCKR</td>
</tr>
<tr>
<td>rs780094</td>
<td>2</td>
<td>27594741</td>
<td>C</td>
<td>T</td>
<td>0.49</td>
<td>0.492119</td>
<td>2.27</td>
<td></td>
</tr>
<tr>
<td>rs780093</td>
<td>2</td>
<td>27596107</td>
<td>C</td>
<td>T</td>
<td>0.49</td>
<td>0.492119</td>
<td>2.27</td>
<td></td>
</tr>
<tr>
<td>rs6442522</td>
<td>3</td>
<td>15415560</td>
<td>C</td>
<td>T</td>
<td>0.47</td>
<td>0.462695</td>
<td>2.38</td>
<td>NR</td>
</tr>
<tr>
<td>rs16890979</td>
<td>4</td>
<td>9531265</td>
<td>C</td>
<td>T</td>
<td>0.45</td>
<td>2.24E-17</td>
<td>28.29</td>
<td>SLC2A9, WDR1</td>
</tr>
<tr>
<td>rs734553</td>
<td>4</td>
<td>9532102</td>
<td>T</td>
<td>G</td>
<td>0.41</td>
<td>0.410E-17</td>
<td>28.21</td>
<td></td>
</tr>
<tr>
<td>rs13129697</td>
<td>4</td>
<td>9536065</td>
<td>T</td>
<td>G</td>
<td>0.36</td>
<td>0.492119</td>
<td>2.27</td>
<td>GCKR</td>
</tr>
<tr>
<td>rs6442522</td>
<td>3</td>
<td>15415560</td>
<td>C</td>
<td>T</td>
<td>0.47</td>
<td>0.462695</td>
<td>2.38</td>
<td>NR</td>
</tr>
<tr>
<td>rs16890979</td>
<td>4</td>
<td>9531265</td>
<td>C</td>
<td>T</td>
<td>0.45</td>
<td>2.24E-17</td>
<td>28.29</td>
<td>SLC2A9, WDR1</td>
</tr>
<tr>
<td>rs734553</td>
<td>4</td>
<td>9532102</td>
<td>T</td>
<td>G</td>
<td>0.41</td>
<td>0.410E-17</td>
<td>28.21</td>
<td></td>
</tr>
<tr>
<td>rs13129697</td>
<td>4</td>
<td>9536065</td>
<td>T</td>
<td>G</td>
<td>0.36</td>
<td>0.492119</td>
<td>2.27</td>
<td>GCKR</td>
</tr>
<tr>
<td>rs6442522</td>
<td>3</td>
<td>15415560</td>
<td>C</td>
<td>T</td>
<td>0.47</td>
<td>0.462695</td>
<td>2.38</td>
<td>NR</td>
</tr>
<tr>
<td>rs16890979</td>
<td>4</td>
<td>9531265</td>
<td>C</td>
<td>T</td>
<td>0.45</td>
<td>2.24E-17</td>
<td>28.29</td>
<td>SLC2A9, WDR1</td>
</tr>
<tr>
<td>rs734553</td>
<td>4</td>
<td>9532102</td>
<td>T</td>
<td>G</td>
<td>0.41</td>
<td>0.410E-17</td>
<td>28.21</td>
<td></td>
</tr>
<tr>
<td>rs13129697</td>
<td>4</td>
<td>9536065</td>
<td>T</td>
<td>G</td>
<td>0.36</td>
<td>0.492119</td>
<td>2.27</td>
<td>GCKR</td>
</tr>
<tr>
<td>rs6442522</td>
<td>3</td>
<td>15415560</td>
<td>C</td>
<td>T</td>
<td>0.47</td>
<td>0.462695</td>
<td>2.38</td>
<td>NR</td>
</tr>
<tr>
<td>rs16890979</td>
<td>4</td>
<td>9531265</td>
<td>C</td>
<td>T</td>
<td>0.45</td>
<td>2.24E-17</td>
<td>28.29</td>
<td>SLC2A9, WDR1</td>
</tr>
<tr>
<td>rs734553</td>
<td>4</td>
<td>9532102</td>
<td>T</td>
<td>G</td>
<td>0.41</td>
<td>0.410E-17</td>
<td>28.21</td>
<td></td>
</tr>
<tr>
<td>rs13129697</td>
<td>4</td>
<td>9536065</td>
<td>T</td>
<td>G</td>
<td>0.36</td>
<td>0.492119</td>
<td>2.27</td>
<td>GCKR</td>
</tr>
<tr>
<td>rs6442522</td>
<td>3</td>
<td>15415560</td>
<td>C</td>
<td>T</td>
<td>0.47</td>
<td>0.462695</td>
<td>2.38</td>
<td>NR</td>
</tr>
</tbody>
</table>

Note: Replicated regions reaching nominal significance are shown in bold.

menopause (Adamopoulos et al., 1977; Puig et al., 1991). In addition, they suggest a potential gene-environment interaction that may be related to the gender-specific effects of the SLC2A9 variants.

In addition to reconfirming the SLC2A9 locus, we provide replications for three previously reported GWAS loci (ABCG2, RREB1, and SLC22A12), though the P-values do not reach strict GWAS significance. We report significant gender-specific effects of a nonsynonymous variant in ABCG2, similar to those previously reported by Kolz et al. (2009). In addition to the replicated regions, we observed suggestive association signals (\(P < 5 \times 10^{-6}\)) at several novel loci. The most significant was a SNP (rs200113) located downstream of SEMA5A, which encodes the semaphorin-5A protein. SNPs in its vicinity were significantly associated with Parkinson’s disease and autism in separate GWAS (Maraganore et al., 2005; Weiss et al., 2009). While elevated uric acid is correlated with lower risk of developing Parkinson’s disease (Alonso et al., 2007; Davis et al., 1996; de Lau et al., 2005; Weiskopf et al., 2007), apart from a hyperuricosuric subtype of autism (Page & Coleman, 2000), no link between autism and uric acid has been reported.

In summary, our study replicated four previously reported SUA associated loci (SLC2A9, ABCG2, RREB1, and SLC22A12) with different levels of significance, and detected suggestive associations at several novel loci (SEMA4A, TMEM18, SLC22A12, and ODZ2) that did not reach the threshold of genome-wide significance. However, due to the moderate sample size and the lack of a replication cohort, the observed associations at these novel loci are preliminary and require further exploration and confirmation in other populations.

Authors’ Contributions

The study was conceived and designed by RD, PR, and RC. Recruitment of subjects, data and sample collection, and data cleaning was conducted by DH-A, NN, DR, SM, ZD, and...
PR. Genotyping was performed by GS, SRI, and HC. Statistical analysis was conducted by GZ and RK. The draft manuscript was prepared by RK, GZ, PR, RC, and RD. RK and GZ have contributed equally. All authors read and approved the manuscript.

Acknowledgements

The study was supported by grants from the National Institutes of Health, USA (R01 DK069845 and P30 ES006096) and from the Croatian Ministry of Science, Education and Sports (196-1962766-2751, 196-1962766-2747, and 196-0342282-0291). RK was supported by a training grant fellowship from the National Institutes of Environmental Health Sciences, USA (T32 ES010957).

Conflict of Interest

The authors declare no conflict of interest.

References


**Supporting Information**

Additional supporting information may be found in the online version of this article:

**Figure S1** Correlation between serum uric acid and age in males and females.

**Figure S2** Quantile–quantile plot of GC-correlated test statistics.

**Figure S3** LocusZoom plot of the ABCC2 region.

**Figure S4** Comparative plot of effect sizes of our most significant genotyped SNP (rs874332) with previous GWA SNPs within the SLC2A9 region.

**Table S1** Demographic and descriptive statistics of the study population, including quantitative traits (mean ± standard deviation) and metabolic disease prevalence (%) in males, females, and the combined sample.

As a service to our authors and readers, this journal provides supporting information supplied by the authors. Such materials are peer-reviewed and may be re-organised for online delivery, but are not copy-edited or typeset. Technical support issues arising from supporting information (other than missing files) should be addressed to the authors.

*Received: 27 September 2011  Accepted: 7 December 2011*