Supplementary Material for:

Sex and age interaction with genetic association of atherogenic uric acid concentrations

Anita Brandstätter PhD[1], Claudia Lamina PhD[1], Stefan Kiechl MD[2], Steven C. Hunt PhD[3], Stefan Coassin MSc[1], Bernhard Paulweber MD[4], Felix Kramer PhD[5], Monika Summerer PhD[1], Johann Willeit MD[2], Lyudmyla Kedenko PhD[4], Ted D. Adams PhD[3], Florian Kronenberg MD[1]

1 Division of Genetic Epidemiology; Department of Medical Genetics, Molecular and Clinical Pharmacology; Innsbruck Medical University, Innsbruck, Austria
2 Department of Neurology, Innsbruck Medical University, Innsbruck, Austria
3 Cardiovascular Genetics Division, University of Utah School of Medicine, Salt Lake City, UT, USA
4 First Department of Internal Medicine, Paracelsus Medical University, Salzburg, Austria
5 Department of Mathematics, Leopold-Franzens University, Innsbruck, Austria

Address of correspondence:
Florian Kronenberg, MD
Division of Genetic Epidemiology
Department of Medical Genetics, Molecular and Clinical Pharmacology
Innsbruck Medical University
Schöpfstr. 41, A-6020 Innsbruck, AUSTRIA
Phone: (+43) 512 9003-70560
Fax: (+43) 512 9003-73560
E-mail: Florian.Kronenberg@i-med.ac.at

Study Populations

Bruneck Study. The Bruneck Study is a prospective population-based survey designed to investigate the epidemiology and pathogenesis of atherosclerosis [(1)]. Briefly, the study population was recruited as a sex- and age-stratified random sample of all inhabitants of Bruneck, Italy (125 women and 125 men in the 5^{th} to 8^{th} decades each, n=1000). At the 1990 baseline, 93.6% of recruited subjects participated, with data assessment completed in 919 subjects. Follow-up examinations were performed 1995, 2000 and 2005. Detailed information on prevalent and incident metabolic syndrome components, diabetes mellitus and cardiovascular events is available from all examinations. The present analysis focuses on the 1995 reexamination. In 1995, the study population still consisted of 826 subjects (96.5% of those alive). Sufficient DNA was available for 800 participants.

SAPHIR Study. The Salzburg Atherosclerosis Prevention Program in Subjects at High Individual Risk (SAPHIR) is an observational study conducted in the years 1999–2002 involving 1,770 healthy unrelated subjects: 663 females from 50 to 70 years of age and 1,107 males from 40 to 60 years of age. Study participants were recruited by health screening programs in large companies in and around the city of Salzburg. At baseline, all study participants were subjected to a comprehensive program [(2)]. DNA was available from 1,732 persons.

Utah obesity case-control Study. The study included 1960 individuals from two groups of subjects gathered in the region of Utah. The study population was composed of 1,106 subjects recruited for severe obesity ("severe obesity group" with a BMI between 33 and 92 kg/m^2) and a general population sample of 854 persons with the same ethnicity ("controls"). The two groups of subjects were described in detail elsewhere [(3,4)]. Briefly, the 1,106 subjects with severe obesity were either seeking gastric bypass surgery or were randomly chosen from a population-based sample of severely obese participants. The examination of patients undergoing gastric bypass surgery was done prior to the intervention. The control group consisted of 854 individuals from the same geographical region and was found to be representative of the Utah population spanning the entire BMI range.

Extended Information on Selected Loci

SLC2A9 is a facilitative glucose, fructose and high-capacity urate transporter [5] predominantly expressed in liver and kidney. Mutations in *SLC2A9* cause renal hypouricemia [6]. *ABCG2* is a urate efflux transporter located in the brush border membrane of kidney proximal tubule cells, where it mediates renal urate secretion [7]. *SLC17A3* codes for a sodium phosphate transporter. Interestingly, while the minor alleles of polymorphisms within *SLC2A9* were associated with lower UA levels especially in women [8,9], the minor alleles within *ABCG2* increased UA levels especially in men [10,11].

Extended Information on Statistical Analysis

In order to determine the best-fitting genetic model, an unconstrained inheritance model was assumed first. The choice of the genetic mode of inheritance was based on the quotient of the corresponding coefficients of the heterozygote and homozygote rare carriers β_{het}/β_{hom} [12]. A quotient of 0.5, for example, corresponds to an additive model, whereas a quotient of 0 indicates that a recessive model is most appropriate and a quotient of 1 supports a dominant model. This analysis, as shown in Supplementary Table S4 suggested an additive mode of inheritance to be most appropriate for most SNPs and samples. Therefore, all single SNP analyses were assumed to follow an additive genetic model.

The percentage of explained variance per SNP was obtained for each sex as the difference of R^2 of a full linear model (regressing BMI, age and each SNP separately on the standardized UA levels) to the R^2 of a reduced linear model ignoring the respective SNP. The differences of these sex-specific explained variances were then tested by means of a z-test based on the corresponding asymptotic variances [13].

Bioinformatic analysis

The potential functional effects of rs2231142 and rs1165205 were evaluated using several bioinformatic tools. First of all, some general information about the genomic loci of *SLC17A3* and *ABCG2* was retrieved, including information on the occurrence of copy number variations (CNVs) (Database of Genomic Variants, http://projects.tcag.ca/variation/project.html), ESPERR regulatory potential [14] of the SNPs' sequence context (UCSC Genomic Browser) and tissue-specific expression patterns (BioGPS, https://biogps.gnf.org/).

The position of rs2231142 (*ABCG2*: Gln141Lys) relative to functional protein domains and amino acid residues was investigated in the BioSapiens DASTY tool

(http://www.biosapiens.info/page.php?page=dasty) and possible effects were predicted using Polyphen [15], SIFT [16] and VisualSNP (http://genepipe.ngc.sinica.edu.tw/visualsnp/c_input.do). For *SLC17A3*, the effects of the intronic SNP rs1165205 on transcription factor binding sites were evaluated using the Genomatix Software Suite (www.genomatix.de). For both SNPs also the correlation with other known amino acid exchanges was evaluated.

Finally, to investigate the role of *SLC17A3* and *ABCG2* in uric acid metabolism and to predict possible interaction partners, we performed data mining using STRING 8 [17].

Supplementary Table S1: Basic characteristics of the analyzed SNPs

Population	Call rate [%]	HWE p-value	AA [%]	Aa [%]	aa [%]	MAF [%]
rs1165205 (*SLC17A3*)						
Bruneck	100.0	0.493	30.0	48.4	21.7	45.9
Saphir	98.2	0.644	29.4	49.1	21.5	46.0
Utah	97.6	0.520	20.9	48.9	30.2	54.6
rs2231142 (*ABCG2*)						
Bruneck	98.8	0.634	76.5	22.3	1.3	12.4
Saphir	99.0	0.069	77.4	20.7	2.0	12.3
Utah	98.0	0.486	80.2	18.5	1.2	10.5
rs12510549 (*SLC2A9*)						
Bruneck	100.0	0.367	64.2	31.2	4.6	20.2
Saphir	98.7	0.051	64.1	30.9	5.0	20.4
Utah	98.0	0.898	63.6	32.2	4.1	20.3
rs6449213 (*SLC2A9*)						
Bruneck	100.0	0.749	65.6	30.6	3.9	19.1
Saphir	99.4	0.330	64.3	31.3	4.4	20.1
Utah	97.8	0.330	65.4	31.4	3.2	18.9
rs6855911 (*SLC2A9*)						
Bruneck	100.0	0.897	58.3	36.3	5.4	23.6
Saphir	99.7	0.813	54.0	38.8	7.2	26.6
Utah	97.8	0.397	54.6	39.1	6.3	25.8
rs7442295 (*SLC2A9*)						
Bruneck	100.0	1.000	62.5	33.1	4.3	20.9
Saphir	99.4	0.577	59.9	34.7	5.4	22.7
Utah	97.8	0.249	59.9	35.6	4.5	22.3

Supplementary Table S2: Results on the linear mixed effect model for the analyzed SNPs within *SLC17A3*, *ABCG2* and *SLC2A9* showing combined p-values and ß-estimates for uric acid levels. The model was adjusted for age, BMI and sex in the combined analyses, and for age and BMI in the sex-stratified analyses. Percentages of explained variance were obtained from linear models on standardized uric acid levels.

		All			Men			Women			P-value testing sex-specific differences of explained variance
Gene	**SNP**	**ß**	**p-value**	**% of variance explained**	**ß**	**p-value**	**% of variance explained**	**ß**	**p-value**	**% of variance explained**	
SLC17A3	rs1165205	0.099	2.00e-04	0.219	0.122	2.50e-03	0.169	0.078	2.44e-02	0.156	0.9567
ABCG2	rs2231142	0.340	1.10e-16	1.641	0.345	1.44e-08	1.605	0.334	1.37e-09	1.666	0.9359
SLC2A9	rs12510549	-0.302	1.86e-20	1.928	-0.236	1.14e-06	1.109	-0.368	2.02e-17	2.992	0.0237
SLC2A9	rs6449213	-0.362	2.17e-27	2.341	-0.285	1.14e-08	1.116	-0.431	2.62e-22	3.608	0.0049
SLC2A9	rs6855911	-0.382	2.36e-36	3.248	-0.337	1.17e-13	2.213	-0.425	3.73e-26	4.267	0.0478
SLC2A9	rs7442295	-0.362	1.18e-29	2.606	-0.312	7.57e-11	1.745	-0.409	7.20e-22	3.398	0.0762
	All SNPs			4.882			3.904			6.075	0.0956

Supplementary Table S3: Associations between SNPs within *ABCG2*, *SLC2A9* and *SLC17A3* with uric acid levels in the investigated populations.

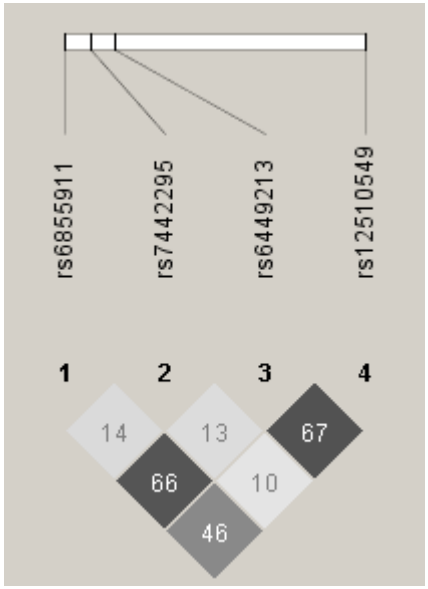
	Uric acid [mg/dl]					
	AA	**Aa**	**aa**	**$ß_{het}$ / $ß_{hom}$**	**$ß_{add}$**	**p-value$_{add}$**
rs1165205 (*SLC17A3*)						
Bruneck	4.53±0.07	4.71±0.06	4.67±0.09	1.26	0.080	0.1600
Saphir	5.48±0.05	5.63±0.04	5.69±0.06	0.72	0.109	0.0057
Utah controls	5.41±0.09	5.43±0.06	5.60±0.07	0.12	0.101	0.0759
Utah obese	6.45±0.10	6.66±0.07	6.66±0.09	1.01	0.097	0.1165
rs2231142 (*ABCG2*)						
Bruneck	4.55±0.05	4.95±0.09	5.13±0.36	0.69	0.376	1.90E-05
Saphir	5.52±0.03	5.84±0.06	6.25±0.20	0.45	0.339	8.90E-09
Utah controls	5.40±0.05	5.79±0.09	5.82±0.41	0.91	0.352	0.0002
Utah obese	6.57±0.06	6.88±0.11	7.09±0.36	0.62	0.294	0.0033
rs12510549 (*SLC2A9*)						
Bruneck	4.80±0.05	4.42±0.07	4.15±0.19	0.59	-0.356	4.81E-07
Saphir	5.71±0.04	5.45±0.05	5.06±0.13	0.41	-0.292	1.36E-09
Utah controls	5.58±0.05	5.35±0.07	5.07±0.20	0.46	-0.243	0.0008
Utah obese	6.75±0.06	6.53±0.08	5.78±0.21	0.23	-0.332	1.03E-05
rs6449213 (*SLC2A9*)						
Bruneck	4.81±0.05	4.41±0.07	3.84±0.20	0.42	-0.435	1.20E-09
Saphir	5.74±0.04	5.43±0.05	4.93±0.13	0.38	-0.347	1.15E-12
Utah controls	5.58±0.05	5.31±0.07	5.00±0.26	0.46	-0.272	0.0004
Utah obese	6.74±0.06	6.46±0.09	5.66±0.22	0.26	-0.384	5.99E-07
rs6855911 (*SLC2A9*)						
Bruneck	4.84±0.05	4.47±0.07	3.91±0.17	0.40	-0.415	5.91E-10
Saphir	5.80±0.04	5.46±0.05	4.86±0.10	0.36	-0.406	2.29E-20
Utah controls	5.61±0.05	5.33±0.07	5.14±0.17	0.59	-0.259	0.0001
Utah obese	6.82±0.07	6.50±0.08	5.84±0.17	0.33	-0.411	4.05E-09
rs7442295 (*SLC2A9*)						
Bruneck	4.82±0.05	4.44±0.07	3.90±0.19	0.41	-0.412	4.50E-09
Saphir	5.77±0.04	5.42±0.05	4.89±0.12	0.39	-0.389	5.18E-17
Utah controls	5.59±0.05	5.31±0.07	5.14±0.21	0.61	-0.257	0.0004
Utah obese	6.77±0.07	6.50±0.08	5.82±0.20	0.28	-0.362	9.12E-07

The GLM was adjusted for age, sex and BMI. For calculating expected means, an inheritance-free model was used. The factor $ß_{het}/ß_{hom}$ was included as indicator for the underlying genetic model. The p-value was obtained from an additive model. A quotient of 0.5, for example, corresponds to an additive model, whereas a quotient of 0 indicates that a recessive model is most appropriate and a quotient of 1 supports a dominant model.

Supplementary Table S4: Results of the general linear regression models on standardized uric acid levels for the analyzed SNPs within *SLC17A3*, *ABCG2* and *SLC2A9* showing combined p-values and ß-estimates. The analysis is stratified for sex and age quartiles and the analysis within each quartile is adjusted for age and BMI.

		Age quartile 1 (≤45 yrs)			Age quartile 2 (46-52 yrs)			Age quartile 3 (53-58 yrs)			Age quartile 4 (≥59 yrs)		
Gene	SNP	ß	p-value	% of variance explained	ß	p-value	% of variance explained	ß	p-value	% of variance explained	ß	p-value	% of variance explained
Women													
SLC17A3	rs1165205	0.053	0.31	0.35	0.063	0.26	-0.22	-0.021	0.72	-0.04	0.143	0.006	1.72
ABCG2	rs2231142	0.133	0.12	0.30	0.437	2.1e-6	3.71	0.333	2.0e-4	2.27	0.265	0.001	1.37
SLC2A9	rs12510549	-0.260	5.4e-5	2.50	-0.236	0.001	1.71	-0.314	1.1e-5	3.16	-0.357	8.8e-8	4.94
SLC2A9	rs6449213	-0.261	8.3e-5	2.30	-0.303	3.7e-5	3.13	-0.380	2.4e-7	4.39	-0.418	1.6e-9	5.57
SLC2A9	rs6855911	-0.255	1.9e-5	2.85	-0.295	1.2e-5	2.92	-0.383	1.1e-8	5.08	-0.413	4.1e-11	6.65
SLC2A9	rs7442295	-0.239	1.7e-4	2.32	-0.303	1.5e-5	2.99	-0.360	4.7e-7	3.88	-0.398	1.6e-9	5.67
Men													
SLC17A3	rs1165205	0.094	0.12	-0.17	0.097	0.07	0.39	-0.012	0.85	0.01	0.177	0.009	2.24
ABCG2	rs2231142	0.169	0.05	0.57	0.380	4.4e-6	3.10	0.197	0.06	0.86	0.201	0.05	0.95
SLC2A9	rs12510549	-0.200	0.004	1.69	-0.261	6.1e-5	2.29	-0.018	0.82	-0.07	-0.178	0.03	0.95
SLC2A9	rs6449213	-0.257	3.4e-4	1.59	-0.290	1.8e-5	2.70	-0.057	0.49	0.05	-0.204	0.02	1.14
SLC2A9	rs6855911	-0.338	1.5e-7	4.98	-0.288	3.5e-6	3.21	-0.191	0.01	1.27	-0.126	0.11	0.28
SLC2A9	rs7442295	-0.320	2.4e-6	4.04	-0.273	2.5e-5	2.43	-0.164	0.04	0.77	-0.116	0.15	0.41

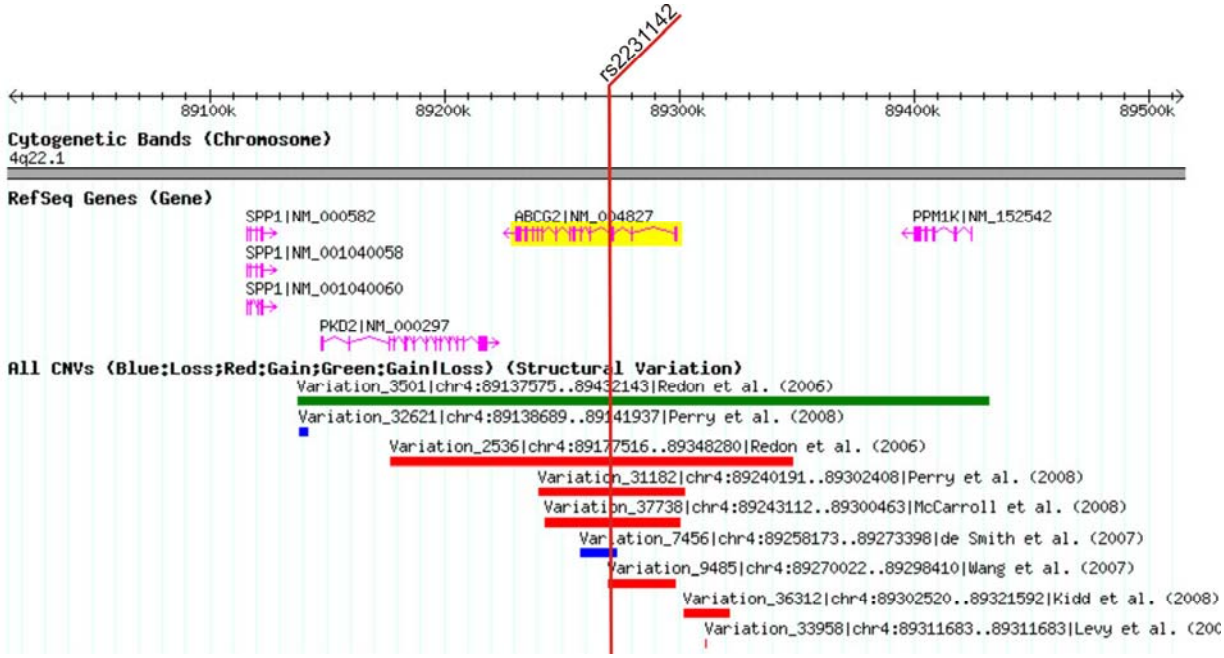
Supplementary Figure S1 Linkage disequilibrium plot in *SLC2A9.* Each diamond contains a pairwise r^2 value between two SNPs, with a dark shade representing higher correlation. The locations of the SNPs are marked on the top panel.



Supplementary Figure S2 CNVs in the gene region of *ABCG2*

This figure shows CNVs currently reported in the Database of Genomic Variants (DGV) in the genetic region of *ABCG2* (+/- approx. 200kb). The red line marks the position of rs2231142 relative to the coding region of *ABCG2* and known CNVs in this region (represented by coloured horizontal boxes). The table below gives details about the CNVs harboring rs2231142, including their frequency in the population investigated in the respective reports (according to the Database of Genomic Variants).

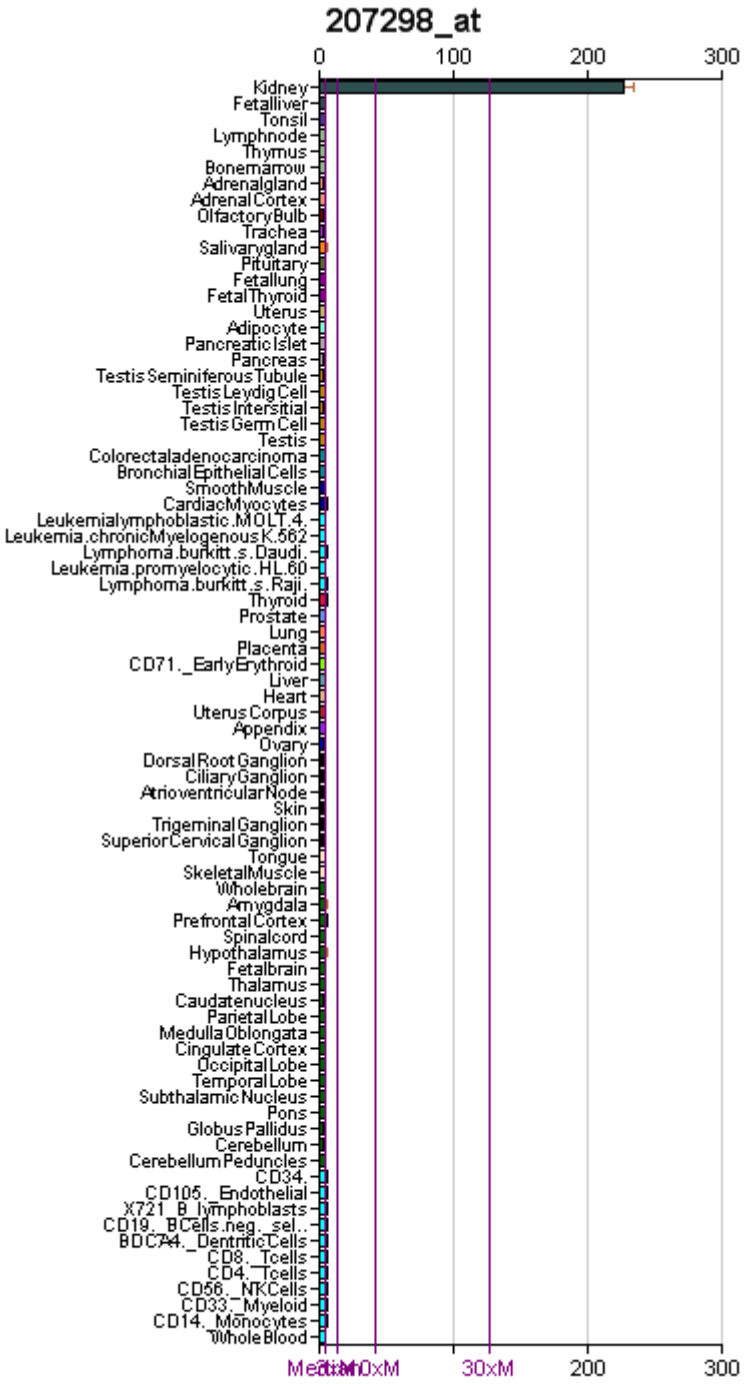
Essentially, rs2231142 lies on 6 know CNVs, thus indicating a region of high genomic variability.



CNV Name	Sample Size	gain/loss	times observed	Citation
Variation_9485	112 control samples (HapMap)	gain	2	*(18)*
Variation_7456	50 control samples (French)	loss	1	*(19)*
Variation_37738	270 control samples (HapMap)	gain	2	*(20)*
Variation_31182	30 control samples (HapMap)	gain	1	*(21)*
Variation_3501	270 control samples (HapMap)	n.a. / both	2	*(22)*
Variation_2536	270 control samples (HapMap)	gain	1	*(22)*

Supplementary Figure S3 Expression level of *SLC17A3* in 79 different tissues as reported by BioGPS.

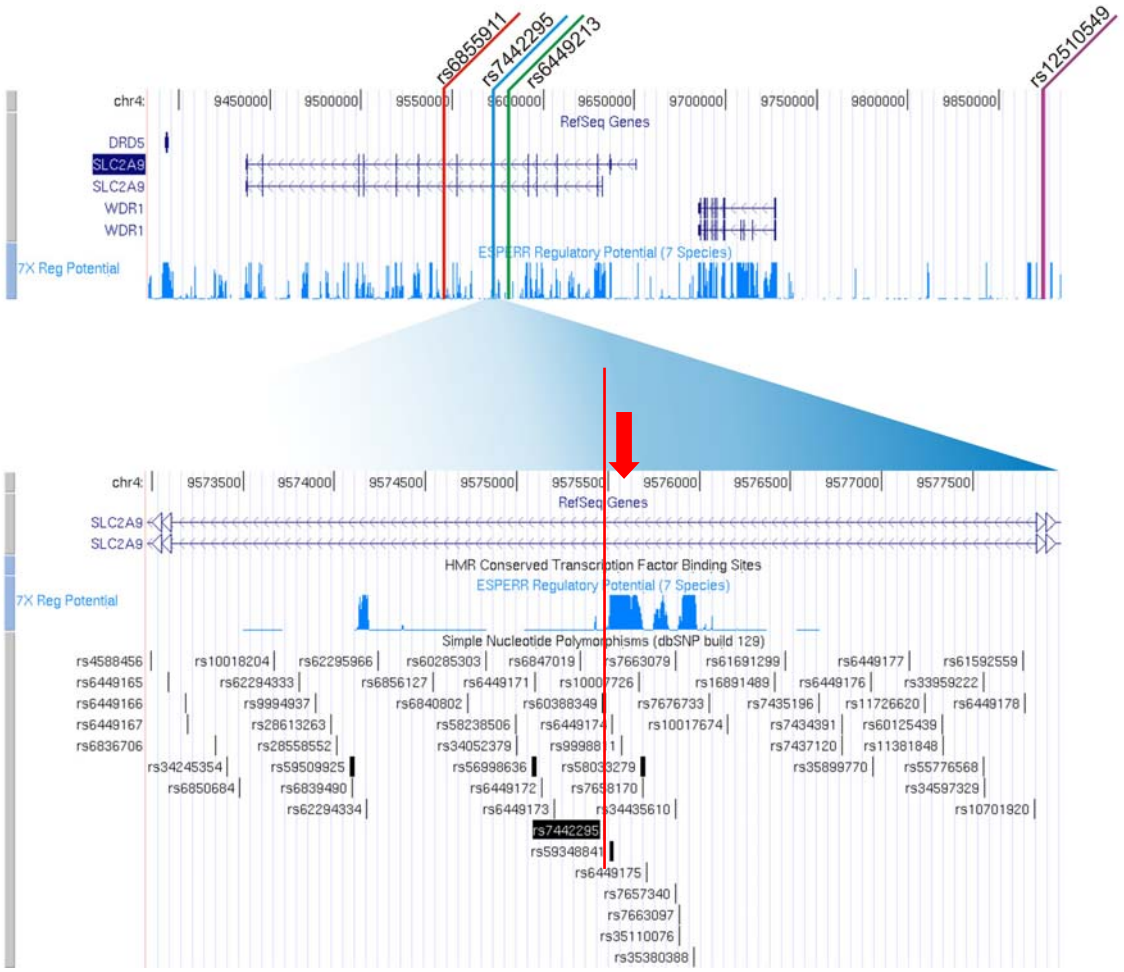
The expression levels were measured using Affymetrix Chip technology and normalized using GC-RMA. *SLC17A3* is reported to be ubiquitously expressed in all tissues to a very low extend, with exception of the kidney tissues, where *SLC17A3* shows an expression level corresponding to more than the 30-fold of the median expression level (marked as 30xM in the picture below).



Supplementary Figure S4 Peaks of regulatory potential in *SLC2A9*

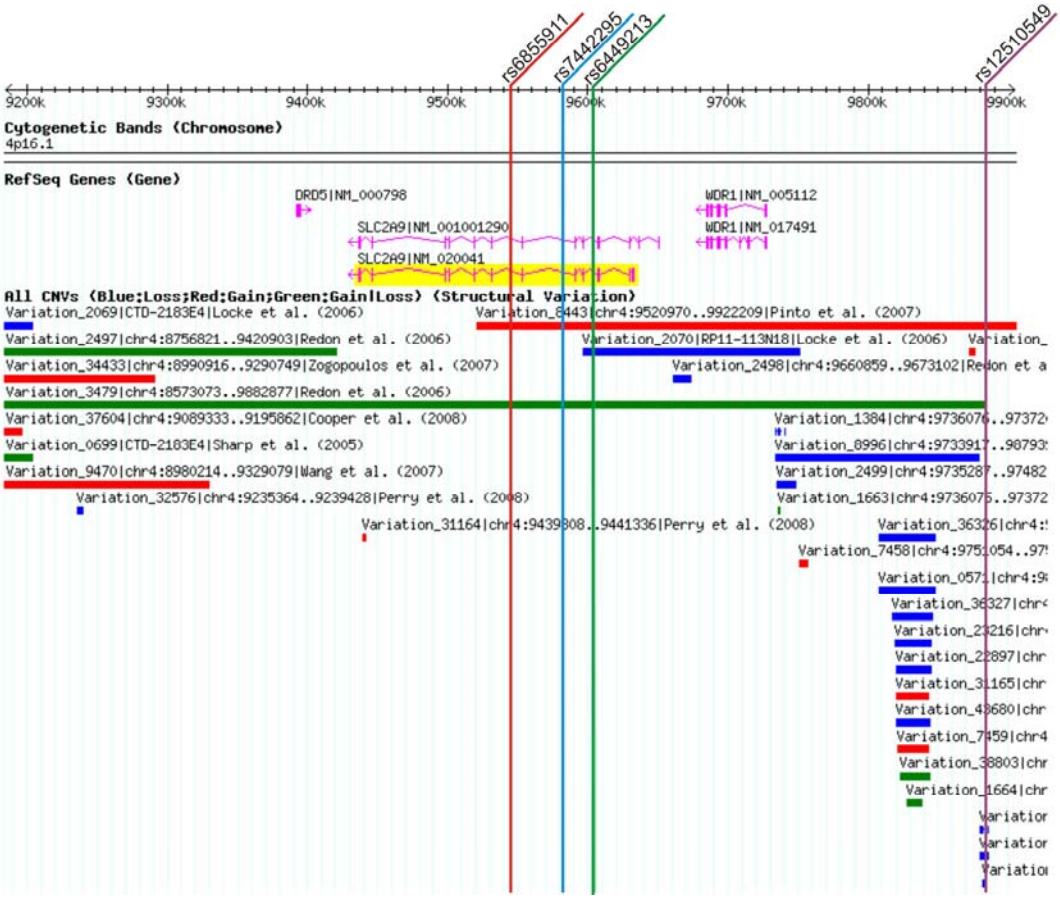
The upper panel shows the position of rs6855911, rs7442295, rs6449213 and rs12510549 relative to the gene region and the calculated ESPERR scores, while the lower panel shows the position of rs7442295 (+/- 2 kb; rs7442295 is marked by the red line).

The SNP rs7442295 is located immediately near a region with a strong peak of regulatory potential (marked by the red arrow)**, thus indicating a potential functional relevance.**



Supplementary Figure S5 CNVs in the gene region of *SLC2A9*

This figure shows CNVs currently reported in the Database of Genomic Variants (DGV) in the genetic region of *SLC2A9* (+/- approx. 200kb). The coloured lines mark the positions of the SNPs relative to the coding region of *SLC2A9* and the CNVs reported in this region (represented by coloured horizontal boxes). The table below gives details about the CNVs harbouring one of the four SNPs, including their frequency in the population investigated in the respective reports (according to the Database of Genomic Variants). **The SNPs rs6855911, rs7442295 and rs6449213 are located on two large CNVs, while rs12510549 lies on the edges of these CNVs.**



CNV Name	Related to SNP	Sample Size	Gain /loss	times seen	Citation
Variation_8443	rs6855911, rs6449213, rs7442295, rs12510549	776 control samples (506 Germans, 270 HapMap)	gain	1	(23)
Variation_2070	rs6449213	269 control samples (HapMap)	loss	10	(24)
Variation_3479	large CNVs affecting rs6855911 , rs6449213, rs7442295, rs12510549 as well as 7 other genes	270 control samples (HapMap)	gain/loss	222	(22)

References

1. Kiechl S, Willeit J, Mayr M, et al: Oxidized phospholipids, lipoprotein(a), lipoprotein-associated phospholipase A2 activity, and 10-year cardiovascular outcomes: prospective results from the Bruneck study. Arterioscler.Thromb.Vasc.Biol. 2007;27:1788-1795.

2. Heid IM, Wagner SA, Gohlke H, et al: Genetic architecture of the *APM1* gene and its influence on adiponectin plasma levels and parameters of the metabolic syndrome in 1,727 healthy Caucasians. Diabetes 2006;55:375-384.

3. Schoenborn V, Heid IM, Vollmert C, et al: The *ATGL* gene is associated with free fatty acids, triglycerides, and type 2 diabetes. Diabetes 2006;55:1270-1275.

4. Pichler M, Kollerits B, Heid IM, et al: Association of the melanocortin-4 receptor V103I polymorphism with dietary intake in severely obese persons. Am.J.Clin.Nutr. 2008;88:797-800.

5. Caulfield MJ, Munroe PB, O'Neill D, et al: *SLC2A9* is a high-capacity urate transporter in humans. PLoS.Med. 2008;5:e197.

6. Matsuo H, Chiba T, Nagamori S, et al: Mutations in glucose transporter 9 gene *SLC2A9* cause renal hypouricemia. Am.J.Hum.Genet. 2008;83:744-751.

7. Woodward OM, Köttgen A, Coresh J, et al: Identification of a urate transporter, *ABCG2*, with a common functional polymorphism causing gout. Proc.Natl.Acad.Sci.U.S.A 2009;106:10338-10342.

8. Döring A, Gieger C, Mehta D, et al: *SLC2A9* influences uric acid concentrations with pronounced sex-specific effects. Nat.Genet. 2008;40:430-436.

9. Brandstätter A, Kiechl S, Kollerits B, et al: Sex-specific association of the putative fructose transporter *SLC2A9* variants with uric acid levels is modified by BMI. Diabetes Care 2008;31:1662-1667.

10. Dehghan A, Köttgen A, Yang Q, et al: Association of three genetic loci with uric acid concentration and risk of gout: a genome-wide association study. Lancet 2008;372:1953-1961.

11. Kolz M, Johnson T, Sanna S, et al: Meta-analysis of 28,141 individuals identifies common variants within five new loci that influence uric acid concentrations. PLoS.Genet. 2009;5:e1000504.

12. Minelli C, Thompson JR, Abrams KR, Thakkinstian A, Attia J: The choice of a genetic model in the meta-analysis of molecular association studies. Int.J.Epidemiol. 2005;34:1319-1328.

13. Olkin I, Finn JD: Correlations redux. Psychological Bulletin 1995;118:155-164.

14. Taylor J, Tyekucheva S, King DC, et al: ESPERR: learning strong and weak signals in genomic sequence alignments to identify functional elements. Genome Res. 2006;16:1596-1604.

15. Sunyaev S, Ramensky V, Koch I, et al: Prediction of deleterious human alleles. Hum.Mol.Genet. 2001;10:591-597.

16. Ng PC, Henikoff S: SIFT: Predicting amino acid changes that affect protein function. Nucleic Acids Res. 2003;31:3812-3814.

17. Jensen LJ, Kuhn M, Stark M, et al: STRING 8--a global view on proteins and their functional interactions in 630 organisms. Nucleic Acids Res. 2009;37:D412-D416.

18. Wang K, Li M, Hadley D, et al: PennCNV: an integrated hidden Markov model designed for high-resolution copy number variation detection in whole-genome SNP genotyping data. Genome Res. 2007;17:1665-1674.

19. de Smith AJ, Tsalenko A, Sampas N, et al: Array CGH analysis of copy number variation identifies 1284 new genes variant in healthy white males: implications for association studies of complex diseases. Hum.Mol.Genet 2007;16:2783-2794.

20. McCarroll SA, Kuruvilla FG, Korn JM, et al: Integrated detection and population-genetic analysis of SNPs and copy number variation. Nat.Genet. 2008;40:1166-1174.

21. Perry GH, Ben-Dor A, Tsalenko A, et al: The fine-scale and complex architecture of human copy-number variation. Am.J.Hum.Genet. 2008;82:685-695.

22. Redon R, Ishikawa S, Fitch KR, et al: Global variation in copy number in the human genome. Nature 2006;444:444-454.

23. Pinto D, Marshall C, Feuk L, Scherer SW: Copy-number variation in control population cohorts. Hum.Mol.Genet. 2007;16 Spec No. 2:R168-R173.

24. Locke DP, Sharp AJ, McCarroll SA, et al: Linkage disequilibrium and heritability of copy-number polymorphisms within duplicated regions of the human genome. Am.J.Hum.Genet. 2006;79:275-290.