From
the Institute of Medical Informatics, Biometry and Epidemiology,
Ludwig-Maximilians-University Munich
Chair of Epidemiology: Prof. Dr. Dr. H.-Erich Wichmann (emeritus)
and
the Institute of Genetic Epidemiology,
Helmholtz Zentrum München –
German Research Center for Environmental Health
Director: Prof. Dr. Konstantin Strauch

Genetic and metabolic components in the regulation of serum urate levels in humans

Thesis

submitted for a doctoral degree in natural sciences at the Faculty of Medicine,

Ludwig-Maximilians-University Munich, Germany

by

Eva Albrecht

from

Munich, Germany

2013

With approval of the Faculty of Medicine of the
Ludwig-Maximilians-University Munich

Supervisor / Examiner:	Prof. Dr. Dr. H.-Erich Wichmann
Co-Examiners:	Prof. Dr. Thomas Illig
Co-Supervisor:	Dr. Christian Gieger
Dean:	Prof. Dr. med. Dr. h.c. M. Reiser, FACR, FRCR
Date of oral examination:	13.02.2014

Contents

Abbreviations and definitions i
List of figures v
List of tables v
Summary vi
Zusammenfassung viii

1 Introduction 1
1.1 Serum urate 1
1.2 Genome-wide association studies (GWAS) 2
1.3 History of serum urate GWAS 6
1.4 Biological networks 7

2 Outline of the thesis 9

3 Serum urate GWAS within ENGAGE 10
3.1 Material and methods 10
3.1.1 Participating studies 10
3.1.2 Statistical analysis 10
3.2 Results 12
3.3 Discussion 17

4 Serum urate GWAS within GUGC 20
4.1 Material and methods 20
4.1.1 Participating studies 20
4.1.2 Statistical analysis 21
4.2 Results 25
4.3 Discussion 32

5 Serum urate genetics in different ancestries 36
5.1 Comparison of different ancestries within GUGC 36
5.2 Serum urate GWAS performed in non-Europeans 40

6 Serum urate GGM 44
6.1 Materials and methods 44
6.2 Results 45
6.3 Discussion 50

7 Conclusion and outlook 52

References 55
Supplementary information 64
Acknowledgement 102
Curriculum vitae 103
Publications 104
Eidesstattliche Versicherung 114

Abbreviations and definitions

AGEN — **Asian Genetic Epidemiology Network**: Consortium of genetic epidemiological studies among Asian populations.

Allele — One of the alternative forms of a → *DNA* sequence that can exist at a single → *locus*.

Base pair — Building block of → *DNA*: **A**denine – **T**hymine (A–T) or **C**ytosine – **G**uanine (C–G).

chr — **Chromosome**: Structures build of → *DNA* storing the genetic information.

CR — **Call rate**: Measure of → *SNP* quality indicating the percentage of successfully called individuals for this variant.

DNA — **Deoxyribonucleic acid**: Nucleic acid carrying the genetic information in the cell; composed of a sequence of nucleotide bases.

EAF — **Effect allele frequency**: Frequency of the → *allele* for which the genetic effect is estimated.

ENGAGE — **European Network for Genetic and Genomic Epidemiology**: Consortium of genetic epidemiological studies among European populations.

Gene — A segment of → *DNA* coding for transcription.

Genome — Entire hereditary information of an individual encoded in the → *DNA* including → *genes* and non-coding sequence.

Genotype — The actual → *alleles* present in a certain individual.

GGM — **Gaussian graphical model**: Network graph based on partial correlations.

GRAIL — **Gene Relationships Across Implicated Loci**: Tool to examine relationships between → *genes.* Can be used to pick candidate → *genes* for given → *SNPs.*

GUGC — **Global Urate Genetics Consortium**: Consortium of genetic epidemiological studies focusing on the genetics of urate.

GWAS	**Genome-wide association study**: Examination of genetic variants across the whole → *genome* to test their associations with a trait of interest.
HapMap	**Haplotype Mapping** Project: International consortium providing haplotype maps of the human → *genome* which are used for → *imputation*.
HWE	**Hardy-Weinberg equilibrium**: Principle describing the distribution of → *genotypes* in an ideal population.
I^2	Measure of heterogeneity in → *meta-analyses*.
Imputation	Process of replacing missing data. In a genetic context it is the estimation of unmeasured → *genotypes*.
Intron	**Intervening region**: Non-coding part within a → *gene*.
kb	**Kilobase**: Measurement of the length of → *DNA* which counts → *base pairs*. 1kb = 1,000 → *base pairs*.
KORA	**Kooperative Gesundheitsforschung in der Region Augsburg** (Cooperative Health Research in the Region of Augsburg): Series of population-based epidemiological studies in the South of Germany.
λ	Genetic inflation factor describing the deviation of the observed distribution of test statistics from its expected distribution.
LD	**Linkage disequilibrium**: Non-random association of → *alleles* at different → *loci*.
Locus	Specific location on a chromosome.
MAF	**Minor allele frequency**: Frequency of the → *allele* which is less frequent in the population of interest.
Mb	**Megabase**: Measurement of the length of → *DNA* which counts → *base pairs*. 1Mb = 1,000,000 → *base pairs*.
Meta-analysis	Statistical approach to combine the effect estimates from independent studies.
Metabolite	Small intermediate molecule of metabolic processes.

NCBI	**National Center for Biotechnology Information**: American center which provides databases relevant to biotechnology and biomedicine.
PPI	**Protein-protein interaction**: Physical connection between two proteins.
Protein	Large molecule composed of amino acids, encoded by the sequence of a → *gene*.
r^2	Squared correlation coefficient used as a measure for → *LD* between two → *SNPs*.
Recombination hotspot	Region in the → *genome* exhibiting an elevated → *recombination rate*.
Recombination rate	Probability that a new combination of → *alleles* constitutes which is different from either parental combination.
SNP	**Single nucleotide polymorphism**: Variation of a single → *base pair* in → *DNA*.

List of figures

1	Purine metabolism	1
2	Simplified representation of DNA	3
3	Schematic representation of a SNP	3
4	Manhattan plots of ENGAGE GWAS	13
5	Regional association plots of 9 ENGAGE loci	14
6	Serum urate – *SLC16A9* – metabolite triangle	16
7	Strategy of systematic network approach	25
8	Manhattan plots of GUGC GWAS	27
9	Quantile-quantile plot comparing network SNPs	31
10	Regional association plot of the *ALDH16A1* region in GUGC	31
11	X chromosomal Manhattan plot	32
12	Serum urate SNP effects between ancestries	39
13	*MAF* locus in Asians and Europeans	41
14	*LRP2* locus in Asians and Europeans	42
15	*SGK1*/*SLC2A12* locus in African Americans and Europeans	43
16	Serum urate GGM	47
17	Medication and sex effects on urate, xanthine, and hypoxanthine	48
18	Manhattan plots with increasing power	53
S1	Regional association plots of 44 GUGC loci	90

List of tables

2 SNP coding 4
3 Comparison of Pearson and partial correlation 8
4 List of ENGAGE studies 10
5 Association results within ENGAGE 15
6 List of GUGC discovery studies 20
7 List of GUGC replication studies 22
8 Association results within GUGC 28
9 List of GUGC studies of non-European ancestry 36
10 Association results between different ancestries 37
11 Partial correlation coefficients in serum urate GGM 45
12 Medication and sex effects in serum urate GGM 49
13 Comparison of sample sizes and findings between serum urate GWAS 52
S1 Genotyping information of ENGAGE studies 65
S2 Study descriptions of GUGC studies 67
S3 Genotyping information of GUGC studies 85

Summary

Uric acid is the final breakdown product of purine metabolism in humans and present in the blood as urate. Elevated serum urate levels can cause gout, a painful inflammatory arthritis, and are implicated in a number of common diseases such as cardiovascular disease, metabolic syndrome, and type 2 diabetes. The regulation of serum urate levels is assumed to result from a complex interplay between genetic, environmental, and lifestyle factors. The underlying functional biological processes are still not completely understood.

The present thesis aimed to identify genetic and metabolic factors in the regulation of serum urate levels. Therefore, two different hypothesis-free approaches were applied. First, two genome-wide association studies were performed in order to identify genetic loci that are involved in the regulation of serum urate levels within the framework of two huge international consortia. The impact of identified genetic loci was compared between different ancestries. Second, a metabolic network analysis within a population-based study was performed aiming to describe the metabolic vicinity of serum urate.

By combining data of approximately 28,000 individuals in a genome-wide association study, nine genetic loci were identified to be involved in the regulation of serum urate levels. The increase of the sample size to a total of approximately 140,000 individuals within a world-wide consortium, combined with a systematic protein-protein interaction network approach, raised the number of detected genetic loci to 28. Although serum urate shows distinct sex differences, an investigation of the X chromosome did not provide additional findings.

Whereas the first identified genes were predominantly involved in urate transport, none of the later identified genes are obviously involved in its transport but underline the importance of the metabolic control of its production and excretion. A comparison between results from different ancestries showed that several of the loci found in Europeans do also play a role in non-Europeans. However, results from one ancestry cannot directly be transferred to other ancestries as the genetic architecture at certain loci can vary between ancestries.

In the metabolite network analysis, serum urate was not only connected to the well-known purine metabolism, but also to a group of essential amino acids and a group of several steroids. Furthermore, association with uricostatic medication intake was not only confined to purine metabolism but seen for nine metabolites within the network. The findings highlight pathways that are important in the regulation of serum urate and suggest that amino acids as well as steroid hormones play a role in its regulation.

The results of both approaches help to better understand the complexity of serum urate regulation in humans, and may help to advance drug development for the treatment and prevention of hyperuricemia and gout.

Zusammenfassung

Harnsäure ist im menschlichen Stoffwechsel das Endprodukt aus dem Abbau von Purinen. Ein erhöhter Harnsäurespiegel kann zu Gicht, einer schmerzhaften Gelenksentzündung, führen und spielt bei einer Vielzahl häufiger Erkrankungen, wie zum Beispiel kardiovaskulären Erkrankungen, dem metabolischen Syndrom und Typ 2 Diabetes eine Rolle. Man geht davon aus, dass der Harnsäurespiegel in einem komplexen Zusammenspiel von genetischen Komponenten, Umwelteinflüssen und Lebensstil reguliert wird. Die zugrundeliegenden biologischen Prozesse sind jedoch bisher nicht vollständig bekannt.

Ziel der vorliegenden Doktorarbeit war es, sowohl genetische Faktoren als auch Stoffwechselkomponenten zu identifizieren, die Einfluss auf die Regulation des Harnsäurespiegels haben. Dazu wurden zwei hypothesenfreie Ansätze gewählt. Zum einen wurden in großen internationalen Konsortien zwei genomweite Assoziationsstudien durchgeführt, um genetische Einflussfaktoren auf die Regulation des Harnsäurespiegels zu finden. Der Einfluss der identifizierten Regionen wurde daraufhin zwischen verschiedenen Ethnizitäten verglichen. Des Weiteren wurden in einer Netzwerkanalyse in einer populationsbasierten Studie metabolische Einflussfaktoren untersucht.

In der ersten genomweiten Assoziationsstudie konnten mit einer Datengrundlage von mehr als 28 000 Individuen neun genetische Regionen mit Einfluss auf die Regulation des Harnsäurespiegels identifiziert werden. Durch eine Erhöhung der Fallzahl auf insgesamt etwa 140 000 innerhalb einer weltweiten Kooperation in Kombination mit einem systematischen Netzwerkansatz basierend auf Protein-Protein-Interaktionen, konnte die Zahl der identifizierten genetischen Regionen auf 28 erhöht werden. Trotz großer Geschlechtsunterschiede im Harnsäurespiegel führte eine Analyse des X-Chromosoms zu keinen zusätzlichen Erkenntnissen.

Während die anfangs gefundenen Gene hauptsächlich am Transport der Harnsäure beteiligt sind, kann keines der später gefundenen Gene mit deren Transport in Verbindung gebracht werden; vielmehr deuten diese auf Stoffwechselkomponenten die bei deren Bildung und Ausscheidung eine Rolle zu spielen scheinen. Ein Vergleich der Ergebnisse zwischen Populationen unterschiedlicher Herkunft zeigt, dass Ergebnisse für Europäer häufig auch in anderen Populationen Gültigkeit besitzen, allerdings aufgrund unterschiedlicher genetischer Architektur nicht unbedingt direkt übertragen werden können.

In der Metabolitennetzwerkanalyse wurde die Harnsäure nicht nur mit dem Purinstoffwechsel verknüpft, sondern auch mit verschiedenen Aminosäuren und Steroiden. Auch die Wirkung von harnsäuresenkenden Medikamenten war nicht auf die Purine beschränkt, sondern zeigte Auswirkung auf neun Metaboliten im Netzwerk. Die Ergebnisse zeigen Zusammenhänge in der Regulation des Harnsäurespiegels und legen nahe, dass sowohl Aminosäuren als auch Steroide eine Rolle spielen.

Die Ergebnisse beider Ansätze helfen, die komplexen Mechanismen in der Regulation der Harnsäure im Menschen besser zu verstehen und werden möglicherweise in der Entwicklung von Medikamenten zur Behandlung und Prävention von Gicht Anwendung finden.

1 Introduction

1.1 Serum urate

Uric acid results from the oxidation of purines. In humans, it is the final breakdown product of purine metabolism, due to an inactivation of the urate oxidase gene during evolution[59]. It is renally excreted. Figure 1 illustrates the pathway via guanine, xanthosine, or hypoxanthine to xanthine and uric acid.

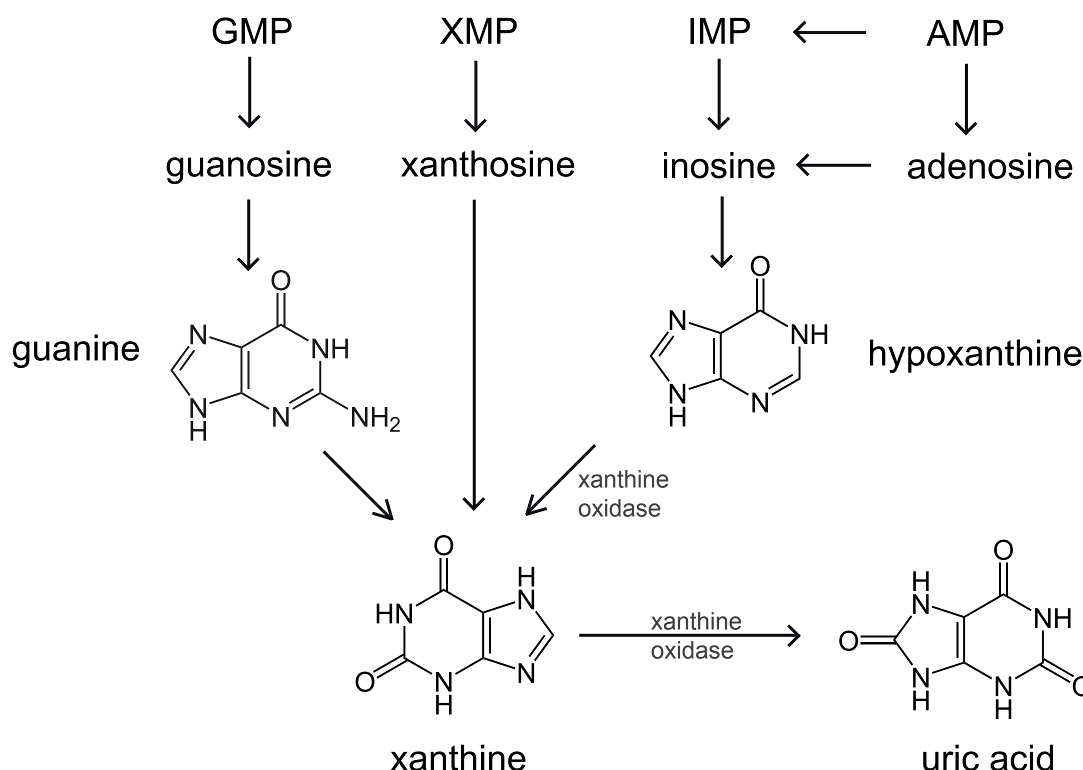


Figure 1: **Purine metabolism.** The figure illustrates fundamental pathways resulting in uric acid as an end product of purine catabolism in humans from guanosine monophosphate (GMP), xanthosine monophosphate (XMP), inosine monophosphate (IMP), and adenosine monophosphate (AMP).

In the blood, 98% of uric acid is present as urate. Serum urate concentrations are determined by a balance between its production and its disposal, regulated by a complex interplay between genetic, environmental, and lifestyle factors such as diet and alcohol consumption[31]. Nevertheless, the underlying functional biological processes of its regulation are still not completely understood.

Serum urate levels are known to substantially differ between sexes with higher levels in men compared to women. In the 14 studies of European ancestry which were analysed in chapter 3 mean levels ranged from 2.69 mg/dl to 5.48 mg/dl in females and from 3.44 mg/dl to 6.33 mg/dl in males[41].

Hyperuricemia, the presence of unusually high serum urate levels, leads to tissue depositions of urate crystals causing gout[67], a painful inflammatory arthritis. In

developed countries, the prevalence of gout is remarkably high with estimated 3.9% among US adults (8.3 million individuals)[87], predominantly men. Besides, hyperuricemia is known to be implicated in cardiovascular disease and mortality and a wide range of cardiovascular risk factors, including hypertension, coronary artery disease, and kidney disease[22,23]. Moreover, elevated serum urate levels have been shown to be associated with obesity and insulin resistance, and consequently with metabolic syndrome and type 2 diabetes[16,23]. On the other hand, extreme low serum urate levels have been observed in multiple sclerosis patients[74] and patients with Parkinson's disease[10].

The heritability of serum urate is estimated to be 40–70%[58,78,84], proving the importance of its genetic determinants. Until now, several studies have aimed to identify genes that are involved in the regulation of serum urate. Whereas a preceding linkage study on serum urate concentrations revealed no significant findings[84], the hypothesis-free approach of genome-wide association studies (GWAS) was very successful. An overview of all GWAS undertaken until now can be found in section 1.3.

1.2 Genome-wide association studies (GWAS)

In the human genome, approximately 3.3 billion single nucleotide pairs of deoxyribonucleic acid (DNA) are distributed across 22 autosomal chromosome pairs and one pair of sex chromosomes. DNA is composed of four nucleotide bases: adenine (A), cytosine (C), guanine (G) and thymine (T). The two strands of DNA are connected by hydrogen bonds between adenine and thymine, or cytosine and guanine, forming the characteristic double helix structure. A simplified representation of DNA is shown in Figure 2.

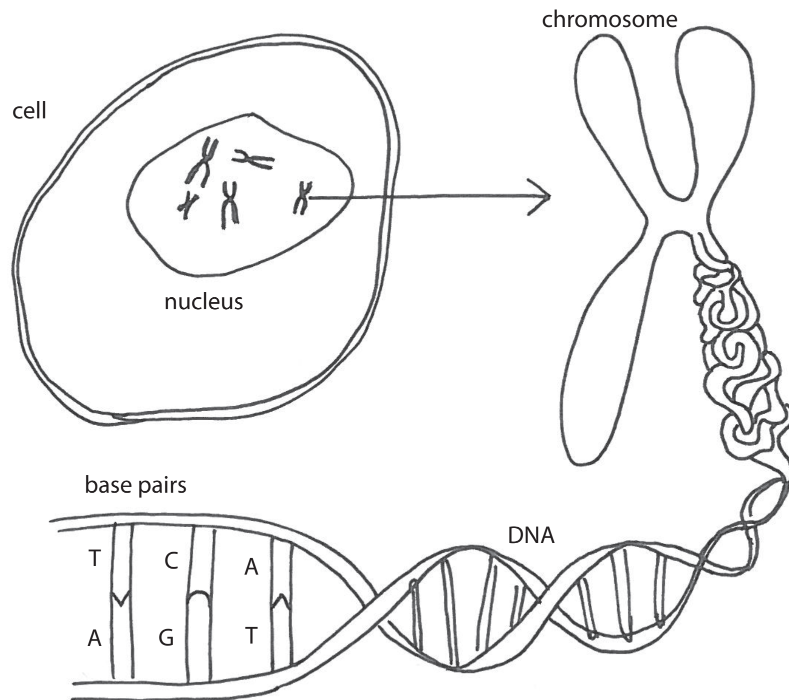


Figure 2: **Simplified representation of DNA** showing its characteristic double helix structure and composition of the four nucleotide bases adenine (A), cytosine (C), guanine (G), and thymine (T).

The genetic information of each individual is stored in the order of its base pairs. For any two human beings, this linear sequence is identical for approximately 99.9% of nucleotide pairs. Nevertheless, there are positions where the base pairs can vary between humans, making every human being unique. Variations at one single base pair are called *single nucleotide polymorphisms (SNPs)* and are illustrated in Figure 3. 90% of the genetic variation can be explained by SNPs. Variants

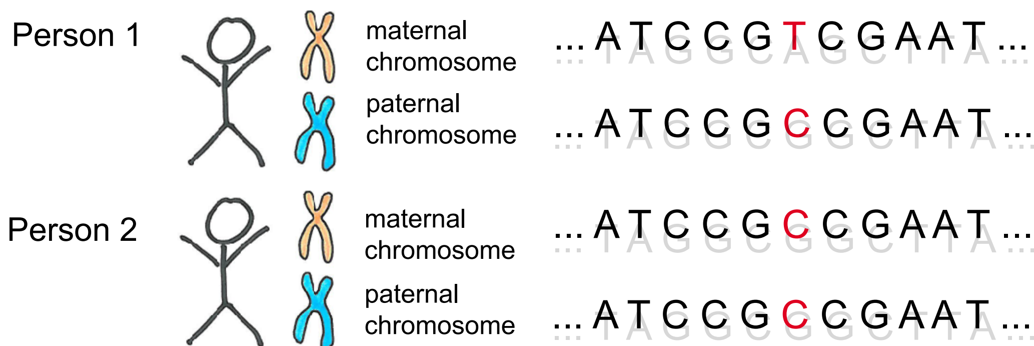


Figure 3: **Schematic representation of a SNP**.

at a *locus*, a specific location on a chromosome, are called *alleles*. An individual can either be *homozygous* at a DNA locus, meaning he/she carries the same allele on both chromosome copies (as Person 2 in Figure 3), or *heterozygous*, if the two

alleles are different (as for Person 1 in Figure 3). The frequency of the less frequent allele across a population is called *minor allele frequency* (MAF). Alleles at two neighbouring loci are often associated, resulting in a complex structure of *linkage disequilibrium* (LD) between SNPs. The measurement for LD between two SNPs used in this thesis is the squared correlation coefficient (r^2).

Over the past few decades, a series of molecular biological laboratory techniques have been developed to sequence DNA segments as well as to genotype SNPs. Micro-array based platforms, provided by companies like Affymetrix and Illumina, facilitate high-throughput genotyping for hundreds of thousands of SNPs, and enable us to systematically investigate SNPs in the entire human genome. In GWAS, disease phenotypes of interest are tested for their association with all available SNPs in order to detect genomic regions being associated with the phenotype. *A priori*, no hypothesis about the location of associated regions is made. Detected SNPs may not be causally involved but associated to the phenotype due to their vicinity to functional variants. Ideally, detected variants are located within or close to genes, and provide therefore insights into biological processes influencing the phenotype.

The association between a quantitative phenotype Y and each SNP is typically tested by a linear regression model with an optional adjustment for n covariates such as sex, age, and other factors which are known to influence the phenotype as well.

$$Y = \beta_0 + \beta_1 Cov_1 + ... + \beta_n Cov_n + \beta_{SNP} SNP$$

Within this linear regression model the SNP can be coded in three different ways as shown in Table 2.

Table 2: **SNP coding:** Different coding possibilities, exemplarily for an A/C SNP with respect to C (typically the minor allele). In this case C is named *coded allele.*

A/A	A/C	C/C	
0	1	1	dominant coding
0	0	1	recessive coding
0	1	2	additive coding

The dominant coding assumes that the presence of one (minor) allele has the same effect on the phenotype as the presence of two (minor) alleles. The recessive coding assumes that only the presence of two (minor) alleles show an effect whereas no effect is seen for the presence of one allele in comparison to the presence of two other (major) alleles. The additive coding assumes that the presence of two (minor) alleles has a two-fold effect in comparison to the presence of one (minor) allele. The additive coding is the most commonly used SNP coding as the experience in GWAS has shown that this assumption is valid in most cases.

In case of a dichotomous phenotype like a diseases status, the linear model is replaced by a logistic regression model.

In order to avoid false positive results due to bad genotyping quality SNPs are typically checked for a number of quality criteria: They should show a high *call rate*, meaning they could successfully be measured in all or a large number of individuals of the study. Furthermore, they are typically checked for the *Hardy-Weinberg Equilibrium (HWE)*, which describes the theoretical distribution of genotypes in an ideal population. Deviations from the HWE may hint to problems during genotype calling. SNPs with bad quality are typically excluded from the analysis.

As the hypothesis-free approach of GWAS considers several hundreds of thousands or even millions of SNPs, and one statistical test is conducted for each SNP, it is necessary to correct the statistical significance level for multiple testing. The simplest and most conservative method to correct for multiple testing is the so-called Bonferroni correction which divides the statistical significance level by the number of statistical tests performed. Due to the complex LD structure of the human genome, it is thought to be sufficient to correct for only one million independent tests[19,61] even if approximately 2.5 million tests are performed (see below). This results in a genome-wide significance level of 5×10^{-8} using a Bonferroni correction at a 5% level. The application of this significance level is the common approach in GWAS so far. In order to reach this significance level, huge sample sizes are necessary to detect associations of moderate effect sizes.

To investigate possible inflation of the test statistics, it is suggested to calculate the inflation factor λ in a GWAS[3]. The λ value describes the inflation of the observed distribution of test statistics across all analysed SNPs with respect to the expected χ^2_1 distribution. Such inflation is possibly observed in case of population stratification in one study. A value of one corresponds to no inflation. For values above one the standard errors of all SNP effect estimates are typically corrected for this inflation in order to avoid an influence on the test statistics due to population stratification[3].

In most studies, several thousand genotyped individuals are available. Sample sizes at such range are only sufficient to detect very large effects. Therefore, data of several studies can be combined in meta-analyses in order to increase sample sizes and power. As several studies usually use different genotyping platforms, the list of genotyped SNPs differs between studies. In order to generate the same data basis across studies, missing genotypes are imputed by use of the known LD structure between SNPs[48,53]. Until now, mainly data from the HapMap project (`http://www.hapmap.org`) was used as a reference for imputation, containing approximately 2.5 million SNPs. Lately, data from the 1000 Genomes Project (`http://www.1000genomes.org`) containing approximately 30 million SNPs got available and will be used for imputation in future GWAS. Commonly used imputation software are IMPUTE[36,54] and MACH[49].

All imputation programs provide measures for the imputation quality of each imputed SNP scaled from zero (bad) to one (high). It is recommended to exclude SNPs with bad imputation quality from the analysis[63]. Commonly used cut-offs are 0.3 for the quality measure provided by IMPUTE and 0.4 for the quality measure provided by MACH. Furthermore, the HWE can also be considered as a

quality criterion for imputed SNPs, and SNPs with deviations from the HWE can be excluded.

GWAS results from all single studies are meta-analysed meaning that effect estimates for each analysed SNP are combined between studies. Herein, the effects are either weighted by the sample sizes of each study or by the inverse variance of each estimate[14]. The inverse variance method gives more weight to studies with more accurate effect estimates and therefore indirectly accounts for the study size. Estimates can be combined with fixed or random effects. Fixed effect are only recommended if effects between studies do not show heterogeneity. As a measure of heterogeneity the I^2 measure is commonly used[32]. It describes the percentage of total variation across all studies due to heterogeneity rather than chance. For a formal statistical testing Cochran's heterogeneity test can be applied[12].

GWAS have proven to be a powerful tool in the detection of genetic loci that are associated with complex diseases. According to the National Human Genome Research Institute (NHGRI) GWA Catalog (`http://www.genome.gov/gwastudies`)[34], 1,467 publications have reported 8,123 SNPs to be associated with various traits until December 2012.

1.3 History of serum urate GWAS

The first GWAS on serum urate levels were performed in 2007 and 2008 by Li *et al.* in a Sardinian population[47], Wallace *et al.* in an English population[77], Vitart *et al.* in a Croatian isolate[76], and Döring *et al.* in the German Cooperative Health Research in the Region of Augsburg (KORA) F3 study[18]. (See section 6.1 for details about the KORA study.) All four GWAS reported on variants in the *SLC2A9* gene to be associated with serum urate and provided evidence for their findings by replication in independent studies. The proportion of variance explained by the reported variants was remarkably high with about 1.2% in men and 6% in women[18]. To this time, the encoded protein SLC2A9 (GLUT9) was thought to be a glucose transporter. Following up the GWAS approaches, functional characterisations initiated by the GWAS showed that SLC2A9 also acts as a urate transporter[8,76].

Also in 2008, Dehghan *et al.*[15] performed a larger GWAS, combining American and Dutch samples, and identified two additional genomic regions at *ABCG2* and *SLC17A3*. The identified variants also showed direction consistency in their association with gout.

In 2009, we performed a GWAS within the European Network for Genetic and Genomic Epidemiology (ENGAGE) consortium as published in Kolz *et al.*[41]. This analysis is one part of this doctoral thesis. It is described in detail in chapter 3. We combined data of 14 independent studies, totalling 28,141 individuals of European descent. We were able to identify nine independent variants in or near *SLC2A9*, *ABCG2*, *SLC17A1*, *SLC22A11*, *SLC22A12*, *SLC16A9*, *GCKR*, *LRRC16A*, and *PDZK1* as described in chapter 3.

In parallel to the ENGAGE analysis, mainly US American studies, organized in the Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE)

consortium performed a GWAS on serum urate levels published in Yang *et al.* in 2010[85]. They were able to replicate *SLC2A9*, *ABCG2*, *SLC17A1*, *SLC22A11*, *GCKR*, and *PDZK1* of the regions published in Kolz *et al.* and additionally reported on two novel regions in our near *INHBC* and *RREB1*.

After the publication of the results within ENGAGE and CHARGE independent from each other, we organized the fusion of both consortia forming the Global Urate Genetics Consortium (GUGC) and additionally invited new joiners to participate in this big effort. With a discovery sample size of more than 110,000 individuals of European descent, we were able to identify multiple novel regions as published in Köttgen *et al.* in 2012[45]. This analysis represents the second part of this doctoral thesis and is described in detail in chapter 4.

In the meantime, Sulem *et al.* published the first whole-genome sequencing analysis investigating on serum urate in 2011[71]. They imputed 15,506 Icelanders with serum urate measurement based on 457 whole-genome sequenced individuals. Whereas genome-wide association studies until now analysed up to approximately 2.5 million genotyped and imputed SNPs, Sulem et al. could test 16 million sequenced SNPs for their association with serum urate and gout. In addition to the replication of the previously known loci *SLC2A9*, *ABCG2*, *SLC17A1*, *SLC22A11*, *GCKR*, *INHBC*, *RREB1*, and *SLC16A9*, they detected one low-frequency variant in *ALDH16A1*, which was not shown to be associated with serum urate before.

The studies mentioned here are based on individuals of European descent. Several additional studies conducted serum urate GWAS in individuals of African American or Asian ancestry. A comparison between findings in Europeans and non-Europeans is given in chapter 5.

1.4 Biological networks

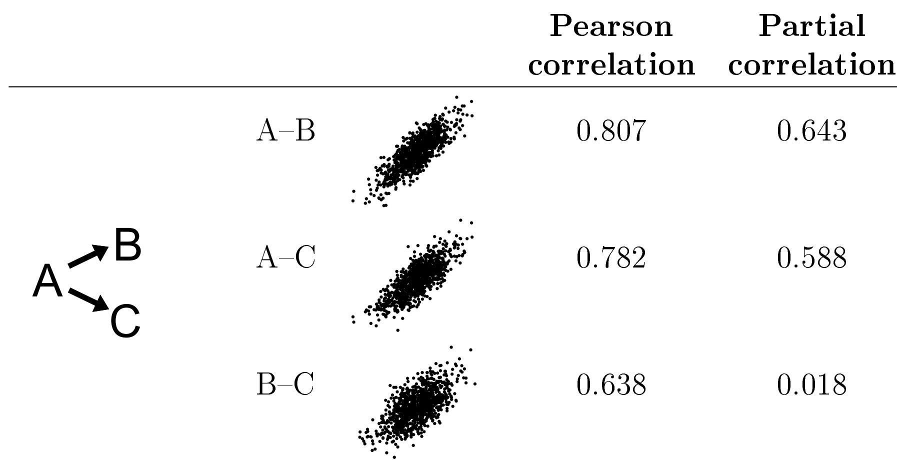
Networks are systems whose structure can be illustrated graphically and are named graphs in a mathematical context. In other words, networks are used to represent linked data structures. Topologically, the basic elements of a network are *nodes* and relationships between nodes are represented by *edges*. Networks are often used to describe and visualize biological systems in an abstract way. In biological networks nodes may be genes (see chapter 4), proteins, metabolites (see chapter 6), or any other elements whose interactions shall be illustrated. Those interactions can be known *a priori*, for instance stored in data bases, or can be calculated from data as for instance in *Gaussian Graphical Models (GGMs)*. GGMs have lately been suggested as a means to reconstruct pathways in metabolomics data[44].

Gaussian graphical models (GGMs)

In the present thesis, a GGM was constructed to describe underlying interconnections in a set of measured metabolites. The most intuitive way to assess statistical association between metabolites would be to calculate Pearson's correlation coefficient for all pairs of metabolites. However, this has the drawback that it cannot distinguish between direct and indirect correlations. The idea in the GGM is to

construct networks based on partial correlation coefficients instead. This means that the correlation between two metabolites is estimated by conditioning on all other available metabolites. Table 3 illustrates the advantage of using partial correlation in a situation where B and C are both correlated with A and therefore indirectly correlated with each other.

Table 3: **Comparison of Pearson and partial correlation** in a situation where B and C are both correlated with A and therefore indirectly correlated (simulated data).



		Pearson correlation	**Partial correlation**
	A–B	0.807	0.643
A→B, A→C	A–C	0.782	0.588
	B–C	0.638	0.018

In the example shown in Table 3 the Pearson correlation coefficient detects high correlation between B and C because of the shared influence of A. On the other hand, the partial correlation coefficient gives no hint of a correlation between B and C as it is able to differentiate between direct and indirect associations in the data. Networks constructed by GGMs on metabolomics data have been shown to be able to reconstruct metabolic reaction pathways[43,44].

2 Outline of the thesis

This thesis aimed to identify novel genetic determinants playing a role in the regulation of serum urate levels, as well as to describe the metabolic vicinity of serum urate in a metabolite network. The thesis contains four parts:

Genetic determinants in the regulation of serum urate levels were investigated by GWAS (see section 1.2) within large consortia. Chapter 3 contains a meta-analysis of GWAS on serum urate levels performed in a total of 28,141 participants of European descent. In order to account for potential sex differences, the analysis was additionally stratified by sex.

To increase power a second meta-analysis of GWAS on serum urate levels was performed in a larger consortium and is described in chapter 4. The study contains a total of 110,347 individuals in the discovery step and 32,813 individuals in the replication step. In order to select genomic regions for replication which did not reach the significance level in the discovery step a systematic network approach was invented. To address the pronounced sex differences a sex-stratified analysis was performed as well as an analysis of the X chromosome.

Both GWAS described in chapter 3 and chapter 4 were conducted in individuals of European ancestry and those findings may not be relevant to individuals of non-European ancestry. Therefore, chapter 5 provides a comparison of findings between different ancestries. First, all genomic regions found in the above-mentioned GWAS were investigated for their impact in samples of Indian, African American, and Japanese ancestry. Second, findings from GWAS on serum urate levels performed in non-Europeans were compared to the results within our European study.

Despite the success of the GWAS approaches in describing the genetic background of serum urate, a detailed functional understanding of the underlying biological processes in the regulation of serum urate levels is still lacking. The detection and functional characterization of such pathways is crucial to improve the management and treatment of patients with hyperuricemia and gout. Chapter 6 contains a study that aimed to describe the metabolic vicinity of serum urate. We examined the metabolic neighbourhood of serum urate by constructing a GGM (see section 1.4) around serum urate. Additionally, we analysed sex differences and the influence of urate lowering medication for all metabolites within the generated network.

3 Serum urate GWAS within ENGAGE

3.1 Material and methods

3.1.1 Participating studies

In this GWAS on serum urate, 14 discovery studies with a combined sample size of 28,141 individuals (thereof 12,328 men and 15,813 women) were involved. All study participants were of European descent. A list of all participating studies is given in Table 4 together with sample sizes and distributions of age and serum urate levels.

Table 4: **List of ENGAGE studies.** Serum urate levels are given in mg/dl.

Study	**Sample size**	**Females %**	**Age mean (range)**	**Serum urate mean (sd)**
BRIGHT	1,743	60.4	56.9 (21–85)	5.39 (1.44)
CoLaus	5,411	53.0	53.4 (35–75)	5.27 (1.42)
CROATIA	774	58.7	56.6 (18–93)	5.27 (1.53)
Health 2000	2,212	50.8	50.4 (30–75)	3.06 (0.77)
KORA F3	1,644	50.6	62.5 (35–79)	5.21 (1.36)
KORA S4	1,814	51.3	56.4 (45–69)	5.40 (1.44)
MICROS	1,086	56.6	45.3 (18–88)	5.33 (1.53)
NSPHS	655	52.8	47.0 (14–91)	5.38 (1.42)
ORCADES	715	53.7	53.6 (17–98)	5.29 (1.34)
PROCARDIS	1,203	19.0	61.1 (37–82)	6.17 (1.44)
SardiNIA	4,305	56.2	43.6 (14–101)	4.32 (1.48)
SHIP	4,087	50.7	49.8 (21–80)	4.39 (1.43)
SSAGA	379	100	46.7 (30–83)	4.82 (1.25)
TwinsUK	2,113	100	47.2 (18–79)	4.49 (1.07)

For each study, genotypes from a genome-wide SNP chip were available and imputed to up to approximately 2.5 million SNPs using HapMap II CEU (build 35 or 36) as a reference panel. Quality control before imputation was conducted in each study separately. Imputation was performed using Impute[54] or MACH[49] in all studies. Study-specific details on genotyping platforms, imputation methods, and quality control are given in supplementary Table S1. Due to different quality control and imputation strategies at a study level, specific sample sizes vary slightly per SNP.

Informed consent was obtained from all participants and the studies were approved by the local ethics committees. A detailed description of study designs is provided as a supplementary text (Text S1) in Kolz *et al.*[41]

3.1.2 Statistical analysis

In each study, SNP associations with serum urate levels were computed by a study-specific analyst and result files of each study were shared in order to enable

the meta-analysis of the results afterwards. On a study level, serum urate levels were adjusted for age and sex and transformed to Z-scores in order to account for variabilities in serum urate levels between studies (see Table 4). Associations between Z-transformed serum urate levels and all 2.5 million SNPs were tested by linear regression models. SNP effects were assumed to be additive. Additionally, the analysis was performed stratified between men and women. Z-scores were calculated in each stratum separately.

The analysis in KORA F4 was performed using SNPTEST[54]. SNPTEST can consider the probabilities of imputed genotypes and therefore allows to account for the uncertainty introduced by the genotype imputation. Details of the software used for all study-specific association analyses are given in supplementary Table S1.

Prior to the meta-analysis, rare SNPs (MAF ≤ 0.01) and SNPs with low imputation quality (≤ 0.4 for MACH or ≤ 0.3 for IMPUTE) were excluded from the datasets of all studies in addition to the study-specific quality control filters to ensure high-quality data and to filter out false positive results. Furthermore, the inflation factor λ was calculated in all studies as provided in supplementary Table S1. All study-specific standard errors were genomic control corrected for their inflation factor λ (if $\lambda > 1$).

The meta-analyses of all genome-wide scans were performed using an inverse-variance weighted fixed effects model in the software metal[79]. In total, 2,493,963 SNPs, distributed across the 22 autosomes, were analysed. The overall inflation factor after the meta-analysis was $\lambda = 1.028$. As this value is close to one which gives no hint of inflated test statistics, no further correction of the meta-analysed results was applied.

SNPs reaching a significance level of $5{\times}10^{-8}$ were considered to be significant. All regions reaching the genome-wide significance level of $5{\times}10^{-8}$ were visualized in regional association plots (see Figure 5), showing $-\log_{10}$ p-values for all SNPs in the region. The LD between the SNP with the smallest p-value (index SNP) and all other SNPs in the region is shown according to the indicated color scheme. Recombination rates in the region are given by the light blue line. Peaks indicate recombination hotspots. Locations of genes are given by green arrows, pointing into the direction of transcription.

Regions where SNPs with noticeable low LD to the index SNP showed genome-wide significant associations as well were tested for independent signals. Therefore, all SNPs with the lowest p-value (index SNPs) of possible independent loci were combined in a multiple SNP model and resulting beta estimates were compared to those of the single SNP models.

Sex differences were tested by means of the test statistic $(\beta_w - \beta_m)/\sqrt{se_w^2 + se_m^2}$ which approximately follows a standard normal distribution.

The percentage of the variance of serum urate that is explained by one SNP was calculated based on its effect estimates (β) and its effect allele frequencies (EAF) by $2 \times \text{EAF} \times (1 - \text{EAF}) \times (\beta^2/var)$. The phenotypic variance var is equal to 1 as the analysis was performed using Z-transformed serum urate levels.

To further characterize the identified variants, we analysed their association with a panel of 163 metabolites measured in 2,020 randomly selected individuals of the KORA F4 survey. The metabolite panel was measured by the AbsoluteIDQ kit (Biocrates Life Sciences AG, Innsbruck, Austria), containing 14 amino acids, one sugar, 41 acylcarnitines, 15 sphingolipids, and 92 glycerophospholipids. Detailed information about metabolite measurements is provided in Kolz *et al.*[41]. Genotype information was available for 1,814 of these individuals. Associations between the nine identified index SNPs and all available metabolite concentrations in μM were tested using the same linear regression models as in the genome-wide scan, adjusting for age and sex. Associations were considered significant for p-values below 3.4×10^{-4}, corresponding to a Bonferroni correction for $(9 \times 163 =)$ 1,467 independent tests. Metabolites showing a significant association with one of the SNPs were tested for their association with serum urate levels in mg/dl by a univariate regression model without further transformation or adjustment.

For the creation of Manhattan plots and regional association plots, as well as for the testing for sex differences, the calculation of explained variances, and the metabolite regression analysis, the statistical software R (`www.r-project.org`) was used.

3.2 Results

In a meta-analysis of 14 genome-wide scans, totalling 28,141 individuals of European ancestry, 954 SNPs reached the genome-wide significance level of 5×10^{-8}. Those SNPs were distributed across seven genetic regions. Figure 4A shows the association results for all 2,493,963 analysed SNPs. At two of the regions two independent signals could be identified, resulting in a total of nine independent loci. Figure 5 shows regional association plots for all nine identified loci ordered by their chromosomal position. The strongest association was found for rs734553 ($p = 5.2\times10^{-201}$), which is an intronic SNP in *SLC2A9* (see Figure 5C). The *SLC2A9* locus had been identified in previous GWAS[15,18,47,76,77]. Also, the second strongest signal in *ABCG2* (rs2231142, $p = 3.1\times10^{-26}$, see Figure 5D) had been shown before in Dehghan *et al.*[15] In the same study, one signal for *SCL17A3* had been found. Our analysis identified the same genomic region (see Figure 5F). However, the SNP with the smallest p-value (index SNP) in this region is intronic to the *SLC17A1* gene (rs1183201, $p = 3.0\times10^{-14}$), which is harbouring *SCL17A3*. The index SNP identified in this analysis, rs1183201, is in high LD with the previously reported rs1165205 ($r^2 = 0.97$). Furthermore, an independent signal was found within an intron of *LRRC16A* (rs742132, $p = 8.5\times10^{-9}$), about 200 kb away from *SLC17A1* (see Figure 5E). Novel identified loci are in or near *SLC22A11* (rs17300741, $p = 6.7\times10^{-14}$, see Figure 5H), *SCL22A12* (rs505802, $p = 2.0\times10^{-9}$, see Figure 5I), *GCKR* (rs780094, $p = 1.4\times10^{-9}$, see Figure 5B), *PDZK1* (rs12129861, $p = 2.7\times10^{-9}$, see Figure 5A), and *SLC16A9* (rs12356193, $p = 1.1\times10^{-8}$, see Figure 5G). The highest amount of variance in serum urate levels was explained by *SLC2A9* with 3.53%, followed by *ABCG2* with 0.57%. Taken together, all nine loci explained 5.22% of the variance in serum urate levels in our data.

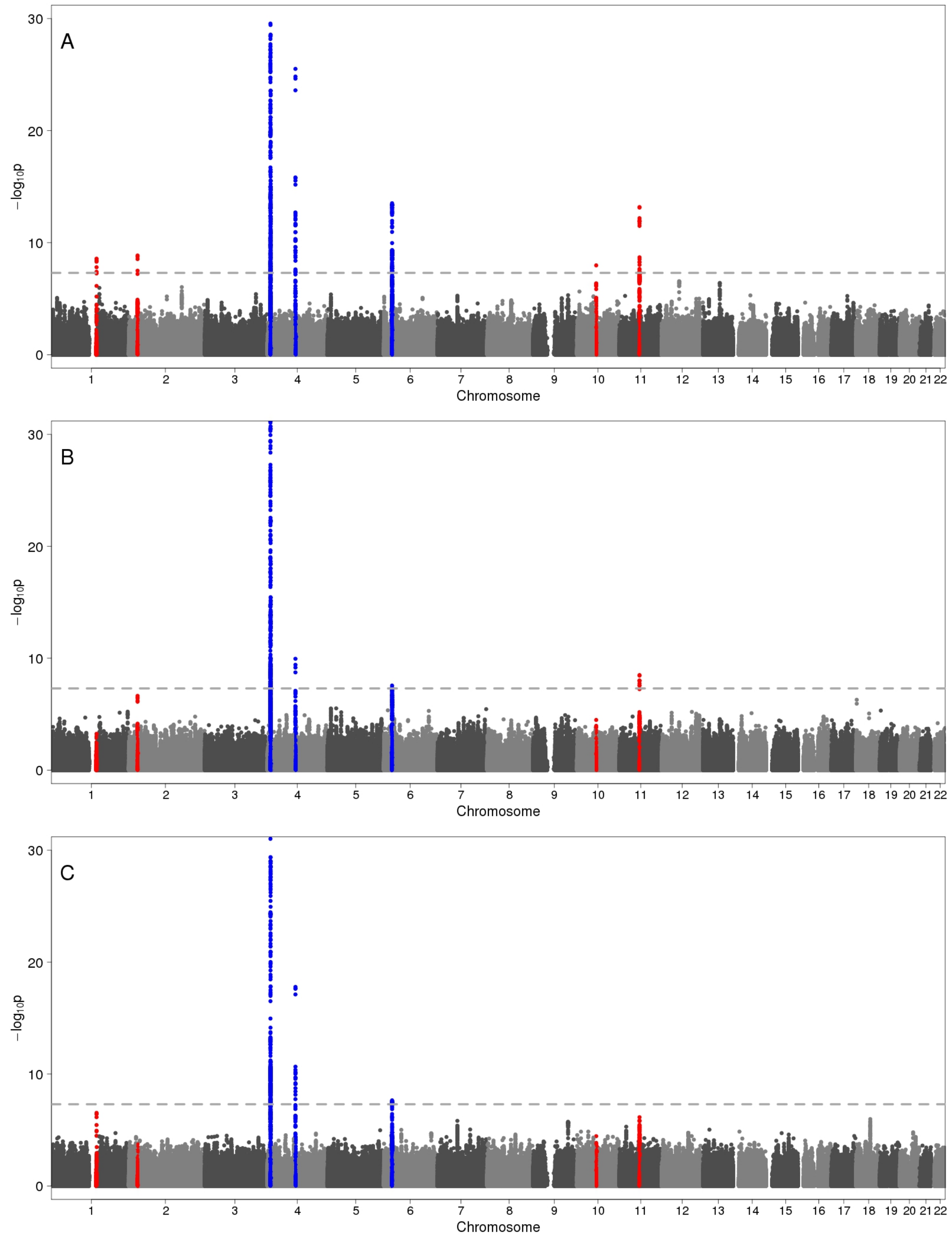


Figure 4: **Manhattan plots of ENGAGE GWAS.** Results of the meta-analysis are plotted as $-\log_{10}$ p-values, ordered by their chromosomal position in A) the overall analysis, B) women only, and C) men only. Previously known loci are coloured in blue whereas all novel identified loci are coloured in red. The grey dashed line indicates the significance level at 5×10^{-8}. The plots are truncated at $p=1\times10^{-30}$. The smallest p-values are 5.2×10^{-201} in the overall analysis, 2.4×10^{-196} in women, and 1.1×10^{-41} in men.

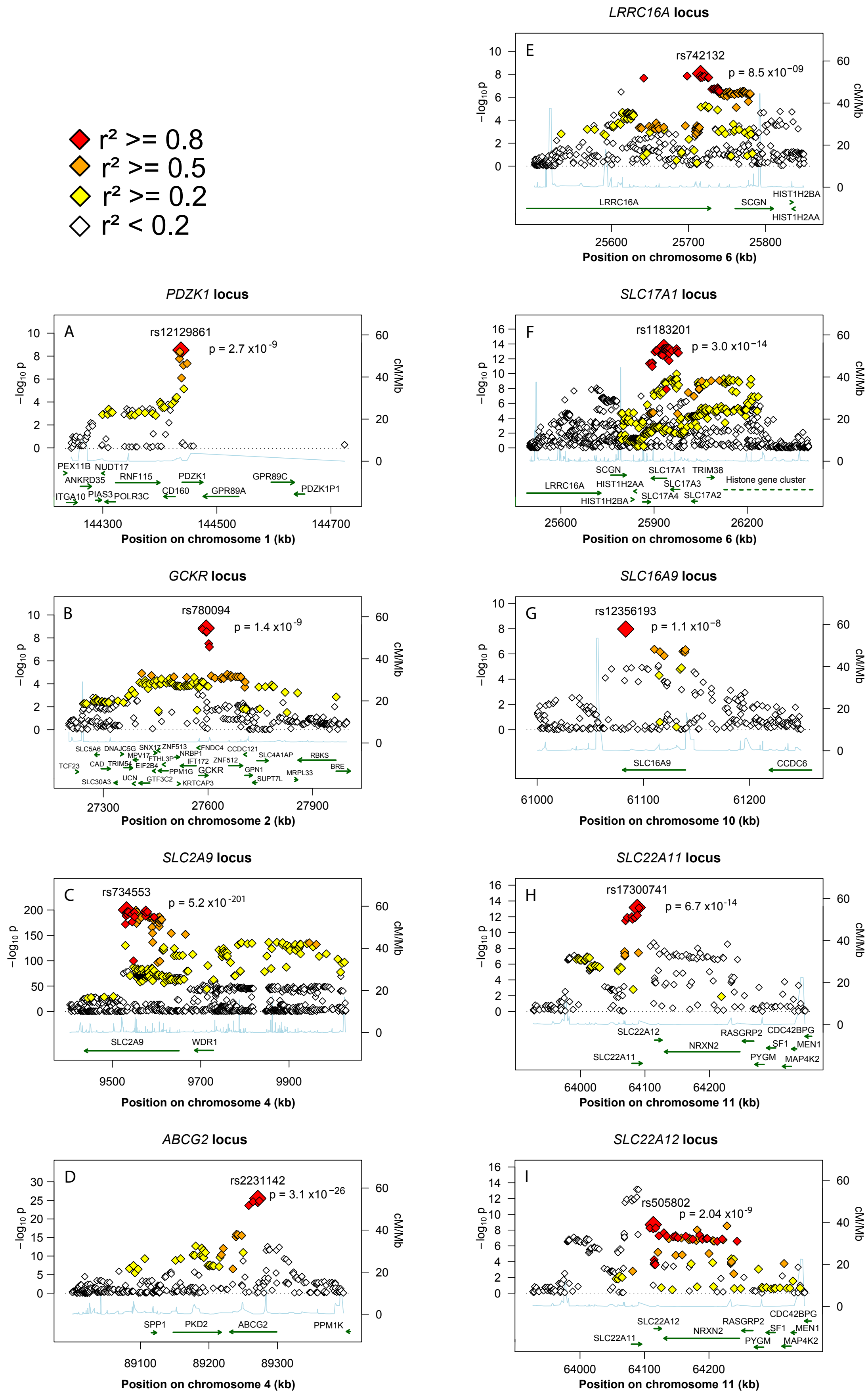


Figure 5: **Regional association plots of nine ENGAGE loci.**

In the sex stratified analysis among 15,813 women and 12,328 men, which is shown in Figure 4B for women and Figure 4C for men, no further regions reached the genome-wide significance level. Table 5 shows the association results for all nine index SNPs in the overall analysis as well as in the sex stratified analysis.

Table 5: **Association results within ENGAGE.** Association results are shown for the SNP with the smallest p-value (index SNP) in the overall analysis, even if there was a SNP with a smaller p-value in the sex stratified analysis at this locus. The first allele is the effect allele. Positions are given for NCBI build 36.

SNP Gene	Chr Position	Sample	Alleles	EAF	N	Beta	p-value
rs12129861	chr 1	overall	G/A	0.54	25,627	0.062	2.68×10^{-9}
PDZK1	144437046	women			13,739	0.047	9.10×10^{-4}
		men			11,888	0.080	3.68×10^{-7}
rs780094	chr 2	overall	T/C	0.42	27,991	0.052	1.40×10^{-9}
GCKR	27594741	women			15,736	0.055	3.11×10^{-7}
		men			12,255	0.050	3.05×10^{-4}
rs734553	chr 4	overall	T/G	0.77	27,817	0.315	5.22×10^{-201}
SLC2A9	9532102	women			15,639	0.397	1.05×10^{-192}
		men			12,178	0.220	1.13×10^{-41}
rs2231142	chr 4	overall	T/G	0.11	23,622	0.173	3.10×10^{-26}
ABCG2	89271347	women			13298	0.138	1.13×10^{-10}
		men			10,324	0.221	2.25×10^{-18}
rs742132	chr 6	overall	A/G	0.70	27,923	0.054	8.50×10^{-9}
LRRC16A	25715550	women			15,688	0.048	8.14×10^{-5}
		men			12,235	0.062	2.68×10^{-5}
rs1183201	chr 6	overall	T/A	0.52	27,908	0.062	3.04×10^{-14}
SLC17A1	25931423	women			15,702	0.055	4.48×10^{-8}
		men			12,206	0.076	2.52×10^{-8}
rs12356193	chr 10	overall	A/G	0.83	23,559	0.078	1.07×10^{-8}
SLC16A9	61083359	women			13,244	0.073	3.29×10^{-5}
		men			10,315	0.089	3.57×10^{-5}
rs17300741	chr 11	overall	A/G	0.51	27,727	0.062	6.68×10^{-14}
SLC22A11	64088038	women			15607	0.060	3.60×10^{-9}
		men			12,120	0.066	1.50×10^{-6}
rs505802	chr 11	overall	C/T	0.30	27,967	0.056	2.04×10^{-9}
SLC22A12	64113648	women			15,735	0.047	1.02×10^{-4}
		men			12,232	0.073	7.22×10^{-7}

In a comparison of men- and women-specific estimates, the effect at *SLC2A9* showed a significant sex difference after a Bonferroni correction for nine independent tests ($p = 3.8\times10^{-17}$), whereas *ABCG2* showed a sex difference which is only significant at a nominal significance level of 0.05 ($p = 0.013$). *SLC2A9* showed a stronger effect in women, whereas *ABCG2* showed a stronger effect in men. For the remaining seven loci no sex differences were found.

By analysing the association of the nine identified variants with a panel of 163 metabolites, we found rs12356193 within *SLC16A9* to be significantly associated with DL-carnitine concentrations ($\beta = 23.58$, $p = 4.0 \times 10^{-26}$) and propionyl-L-carnitine concentrations ($\beta = 20.06$, $p = 5.0 \times 10^{-8}$) after correction for multiple testing. In turn, DL-carnitine concentrations ($\beta = 0.06$, $p = 1.4 \times 10^{-57}$) as well as propionyl-L-carnitine concentrations ($\beta = 1.78$, $p = 8.1 \times 10^{-54}$) were associated with serum urate levels, forming a triangle between SNP, metabolites, and serum urate levels. The relationships are illustrated in Figure 6.

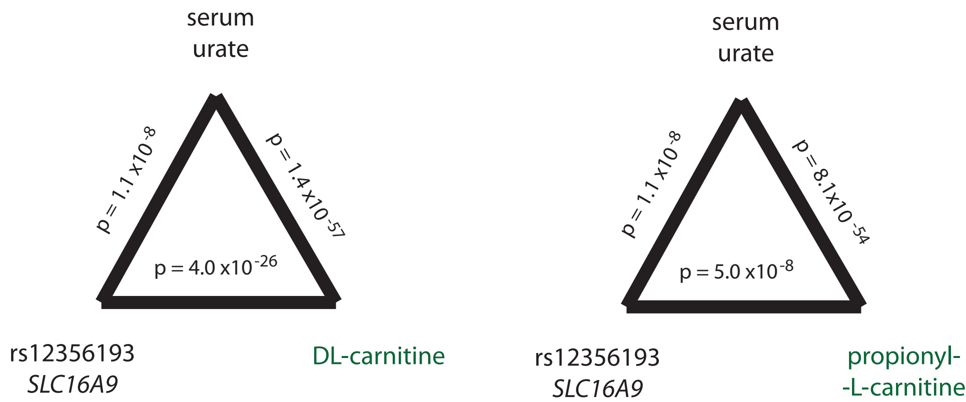


Figure 6: **Serum urate –** ***SLC16A9*** **– metabolite triangle** illustrating the relationship between serum urate, *SLC16A9*, and DL-carnitine and propionyl-L-carnitine respectively.

3.3 Discussion

We identified nine independent genetic loci being significantly associated with serum urate levels in a GWAS containing 28,141 individuals. Five of these loci were novel at this time (*SLC22A11*, *SCL22A12*, *GCKR*, *PDZK1*, *SLC16A9*).

Many of the identified loci harbour genes that encode urate transporters: The *SLC22A12* gene encodes URAT1 which was the first characterized urate transporter[20]. *SLC2A9* encodes SLC2A9 (GLUT9) which is involved in renal urate re-absorption[8,76] and was characterized as a urate transporter after the first series of serum urate GWAS. Later it was shown that homozygous loss-of-function mutations of GLUT9 cause a total defect of uric acid absorption, leading to severe renal hypouricemia[17]. The ABC transporter ABCG2, encoded by *ABCG2*, has been shown to operate as a urate transporter contributing to the excretion of urate via the kidney, and several mutations of ABCG2 were shown to reduce its function[55,80,81]. *SLC17A1* encodes NPT1 which can transport urate as well and is likely involved in urate excretion[39]. *SLC22A11* encodes for OAT4 which operates as a urate transporter as well[30]. *PDZK1* is known to influence urate transport indirectly as the urate transporters URAT1, NPT1, and OAT4 are known to bind to PDZK1[2,57]. Those findings strongly support the hypothesis that transport proteins are playing a major role in the regulation of serum urate levels.

The index SNP rs12356193 on chromosome 10 is located within *SLC16A9* which encodes for MCT9. We saw an association between rs12356193 and two of the investigated metabolites, namely DL-carnitine and propionyl-L-carnitine, which in turn were associated with serum urate levels, forming a triangle between SNP, metabolites, and serum urate levels. After the publication of these results, the predicted function of SLC16A9 (MTC9) as a carnitine efflux transporter was experimentally validated[70].

For the remaining genes the function is less clear. The independent signal 200 kb away from *SLC17A1* on chromosome 6 contains *LRRC16A* and *SCGN*. The closest gene to rs780094, the SNP with the strongest association signal on chromosome 2, is *GCKR*. However the region contains a number of genes and the GWAS approach cannot distinguish which of the genes in the region is functionally involved in the regulation of serum urate levels.

The relationship between urate and other metabolites is more precisely investigated in chapter 6 where a metabolite network was constructed based on a panel of metabolites measured by Metabolon Inc.[21] The Metabolon panel covers more broadly several different classes of metabolites as described in chapter 6 whereas the Biocrates panel which was used in this chapter's analysis mainly contains glycerophospholipids.

Two of the detected regions show gender differences. This observation is line with previous findings[18] and not surprising as serum urate levels are known to vary between men and women (see section 1.1).

During the analysis of a GWAS meta-analysis, a number of decisions concerning statistical model, adjustment for covariates, quality control and filtering of single

study results, and the statistical approach in the meta-analysis have to be made. It is important to carefully choose the most appropriate approaches in each situation in order to avoid false positive results. The results of the serum urate GWAS presented herein were published without a formal replication which is usually expected in the field. Nevertheless the later publications of Yang *et al.*[85] and Köttgen *et al.*[45] confirmed that all findings were no false positive results, underpinning the validity of the approach used here.

Taken together we identified nine genetic loci highlighting biological pathways that are involved in the regulation of serum urate levels. The findings suggest that transport proteins are playing a major role and point towards novel potential targets for the treatment and prevention of hyperuricemia and gout.

With 5.22% of the variance, we explain a comparably large proportion of phenotypic variation in comparison to GWAS of other traits. Nevertheless, the estimated heritability of about 40–70%[58,78,84] for serum urate levels suggests that additional loci remain to be identified. The extension of this analysis to a sample size of more than 110,000 individuals of European descent in the discovery step is described in chapter 4.

My work presented in this chapter is also published in

Kolz M, Johnson T, Sanna S, Teumer A, Vitart V, Perola M, Mangino M, **Albrecht E**, Wallace C, Farrall M, Johansson A, Nyholt DR, Aulchenko Y, Beckmann JS, Bergmann S, Bochud M, Brown M, Campbell H; EUROSPAN Consortium, Connell J, Dominiczak A, Homuth G, Lamina C, McCarthy MI; ENGAGE Consortium, Meitinger T, Mooser V, Munroe P, Nauck M, Peden J, Prokisch H, Salo P, Salomaa V, Samani NJ, Schlessinger D, Uda M, Völker U, Waeber G, Waterworth D, Wang-Sattler R, Wright AF, Adamski J, Whitfield JB, Gyllensten U, Wilson JF, Rudan I, Pramstaller P, Watkins H; PROCARDIS Consortium, Doering A, Wichmann HE; KORA Study, Spector TD, Peltonen L, Völzke H, Nagaraja R, Vollenweider P, Caulfield M; WTCCC, Illig T, Gieger C. **Meta-analysis of 28,141 individuals identifies common variants within five new loci that influence uric acid concentrations**. *PLoS Genet.* 2009 Jun;5(6):e1000504. Epub 2009 Jun 5.

My detailed contribution in this project was as follows:

I was the analyst of the KORA F4 study performing the GWAS and independency analysis at a study level. Furthermore, I was the analyst conducting the meta-analysis of all genome-wide scans.

4 Serum urate GWAS within GUGC

4.1 Material and methods

4.1.1 Participating studies

This GWAS on serum urate involved 48 discovery studies, totalling 110,347 individuals (49,825 men and 60,522 women) of European descent. In the replication stage, we included 12 studies with *in silico* genotype information, totalling 18,821 individuals (8,993 men and 9,828 women) and three studies with *de novo* genotyping, providing 13,992 individuals (6,153 men and 7,839 women). Specific sample sizes varied slightly per SNP. Sample sizes and study characteristics in terms of sex, age, and serum urate distributions are provided in Table 6 for all discovery studies and in Table 7 for all replication studies. Detailed information about individual study designs, population details, and serum urate measurements can be found in supplementary Table S2.

Table 6: **List of GUGC discovery studies**. Serum urate levels are given in mg/dl.

Study	Sample size	Females %	Age mean (sd)	Serum urate mean (sd)
AGES Reykjavik Study	3,219	58.0	76.4 (5.5)	5.97 (1.62)
Amish	1,139	48.1	49.8 (16.8)	4.20 (1.10)
ARIC	9,049	52.9	54.3 (5.7)	5.94 (1.50)
ASPS	845	43.2	65.2 (8.0)	5.32 (1.42)
AUSTWIN	11,520	59.5	39.2 (17.2)	4.85 (1.32)
BLSA	521	47.8	70.6 (14.1)	5.21 (1.49)
BRIGHT	1,743	60.4	56.9 (10.3)	5.39 (1.44)
CARDIA	1,713	53.4	25.5 (3.3)	5.33 (1.37)
CHS	3,252	60.9	72.3 (5.4)	5.50 (1.44)
CoLaus	5,409	52.9	53.4 (10.7)	5.27 (1.42)
CROATIA-KORCULA	895	63.9	56.2 (14.0)	4.92 (1.29)
CROATIA-SPLIT	490	57.9	49.0 (14.6)	4.80 (1.38)
CROATIA-VIS	912	57.7	56.4 (15.5)	5.23 (1.59)
DESIR	716	75.1	50.2 (8.2)	3.86 (0.92)
EPIC-Norfolk cohort	1,835	54.3	59.3 (9.0)	4.99 (1.37)
ERF	889	60.7	49.6 (15.2)	5.52 (1.56)
Estonian Biobank	931	50.8	39.4 (15.6)	4.95 (1.36)
Family Heart Study (FamHS)	3,837	52.4	52.1 (13.7)	5.42 (1.47)
FHS	7,699	53.1	37.9 (9.4)	5.34 (1.51)
Health 2000	2,069	50.9	50.6 (11.0)	5.17 (1.29)
InCHIANTI	1,205	55.5	68.2 (15.5)	5.08 (1.43)
INCIPE	940	52.6	61.2 (11.5)	5.50 (1.45)
INGI-Carlantino	432	61.3	49.9 (16.5)	4.90 (1.41)
INGI-CILENTO	859	55.2	52.5 (19.4)	4.60 (1.59)
INGI-FVG	1,018	61.6	48.2 (19.7)	5.55 (1.57)
INGI-Val Borbera	1,658	55.8	54.7 (18.3)	5.02 (1.28)

KORA F3	1,643	50.5	62.5 (10.1)	5.21 (1.36)
KORA F4	1,814	51.3	60.9 (8.9)	5.37 (1.45)
LBC1936	769	47.7	72.5 (0.7)	5.51 (1.47)
LifeLines	3,343	59.2	55.5 (9.9)	5.04 (1.25)
LOLIPOP-EW-A	587	12.9	54.3 (10.4)	5.62 (1.32)
LOLIPOP-EW-P	650	0	55.7 (9.1)	5.75 (1.35)
LOLIPOP-EW610	924	26.9	55.9 (9.8)	5.30 (1.45)
LURIC	963	27.8	62.0 (10.6)	5.04 (1.68)
MICROS	1,236	56.8	45.0 (16.8)	5.31 (1.42)
NESDA	1,731	67.9	42.3 (12.5)	4.37 (1.17)
NSPHS	655	52.8	47.0 (20.7)	5.38 (1.42)
ORCADES	888	54.6	53.5 (15.7)	4.98 (1.20)
PREVEND	3,785	48.9	49.6 (12.4)	5.12 (1.34)
PROCARDIS	3,742	24.4	62.1 (7.0)	6.10 (1.46)
RS-I	4,274	61.6	70.1 (9.0)	5.41 (1.37)
RS-II	2,123	54.4	64.8 (8.0)	5.25 (1.28)
SardiNIA	4,694	56.3	43.3 (17.6)	4.32 (1.48)
SHIP	4,067	50.7	49.7 (16.3)	4.90 (1.40)
SOCCS	1,105	50.0	51.0 (5.7)	4.60 (1.25)
Sorbs	896	59.6	47.2 (16.3)	5.80 (1.64)
TwinsUK	3,640	100	48.1 (12.9)	4.49 (1.07)
Young Finns Study	2,023	54.7	37.7 (5.0)	4.74 (1.27)

In each discovery study, genotypes from a standard genotyping platform were available and imputed to up to approximately 2.5 million SNPs using HapMap II CEU as a reference. Study-specific details on genotyping platforms, imputation methods, and quality control for discovery and *in silico* studies are provided in supplementary Table S3.

4.1.2 Statistical analysis

Analysis at the study level

SNP associations in each study were computed by a study-specific analyst and result files of each study were shared for subsequent meta-analysis. In each discovery study, SNP associations with serum urate levels were analysed by linear regression models, assuming the SNP effect to be additive. Unlike the analysis in the ENGAGE GWAS, each cohort used the raw serum urate values, measured in mg/dl. Adjustment for age and sex was made within the linear model, as well as for study-specific covariates where appropriate, such as adjustment for study center, population stratification, or family structure. The analysis was done for the overall sample as well as stratified by sex. All analyses accounted for the uncertainty introduced by the genotype imputation by using the expected allele dosages. In the replication studies, the same models were applied as in the discovery studies. Details of the software used for study-specific association analyses are given in supplementary Table S3.

Table 7: **List of GUGC replication studies.** Serum urate levels are given in mg/dl.

Study	Sample size	Females %	Age mean (sd)	Serum urate mean (sd)
***in silico* replication**				
EPIC cases	793	58.4	59.3 (8.8)	5.51 (1.53)
GHS I	2,995	48.5	55.9 (10.9)	4.83 (1.47)
GHS II	1,179	50.0	55.1 (10.9)	4.75 (1.47)
GSK cases	819	66.3	50.9 (13.7)	5.00 (1.40)
GSK controls	851	67.7	51.9 (13.2)	4.99 (1.34)
Hunter Community Study	1,088	51.1	65.9 (7.4)	5.37 (1.34)
Lifelines replication	5,031	56.1	43.2 (9.1)	4.85 (1.21)
LURIC replication GZ	804	34.0	59.0 (12.0)	5.00 (1.70)
LURIC replication HD	1,156	29.8	64.9 (9.2)	5.20 (1.70)
MARS cases	636	52.7	48.4 (14.0)	5.19 (1.29)
OGP-Talana	1,039	55.9	50.9 (19.1)	4.59 (1.64)
SAPALDIA asthmatics	570	52.5	51.3 (11.3)	5.31 (1.49)
SAPALDIA non-asthmatics	874	50.1	52.9 (11.1)	5.25 (1.46)
SHIP-Trend	986	56.2	50.1 (13.7)	4.79 (1.25)
***de novo* replication**				
HYPEST	751	63.5	57.8 (9.8)	5.79 (1.50)
KORA S2	3,685	49.0	49.6 (14.1)	4.93 (1.47)
OGP (Ogliastra)	9,556	56.1	49.6 (17.9)	4.36 (1.48)

Quality control and meta-analysis

Before meta-analysis, all study-specific genome-wide association results underwent extensive and standardized quality control procedures. All files were checked for completeness and all variables were checked for plausibility of their descriptive statistics, supported by the gwasqc function of the GWAtoolbox package[26] in R (`www.r-project.org`). As a positive control, the effect direction and size, as well as the p-value of rs16890979 in *SLC2A9* was compared between all studies, as the known influence of *SLC2A9* is strong enough to be seen in moderate sample sizes. In addition to the study-specific quality control filters, we excluded SNPs being monomorphic in the respective study and corrected study-specific results by their inflation factor λ if $\lambda > 1$.

The meta-analysis of all genome-wide scans was performed using an inverse-variance weighted fixed effects model in metal[79], for the overall, men-, and women-specific scans separately. All meta-analyses were double-checked by an independent analyst. After the meta-analysis we removed all SNPs that were not available in at least 75% of the overall sample size, which lead to a total number of 2,450,547 analysed genotyped or imputed autosomal SNPs. The genomic inflation factor λ was 1.07 in the overall analysis, 1.08 in the women-specific analysis, and 1.03 in the men-specific analysis. We applied a second genomic control correction. In line with the ENGAGE GWAS, the commonly used genome-wide significance level of

5×10^{-8} was applied. Additionally, SNPs reaching p-values below 1×10^{-6} were followed-up in the replication analysis. To investigate the detected regions for potential independent signals, an independency analysis was performed as described in Köttgen *et al.*[45] The heterogeneity between studies was investigated by the I^2 measure[32]. The calculation of explained variances is described in Köttgen *et al.*[45]

Replication

In the replication analysis, results of all *in silico* and *de novo* studies were meta-analysed in metal[79] applying inverse-variance weighting and fixed effect modelling as for the discovery cohorts. Afterwards, the results from the discovery step and from the replication step were meta-analysed using the same approach. A SNP was considered as replicated if its q-value[69] was below 0.05 in the replication step. Unlike the well-known p-value, the q-value is a measure of significance in terms of the false discovery rate rather than the false positive rate. Given the limited power in the replication step, we decided to apply this more liberal approach. SNPs that had not reached the genome-wide significance level in the discovery phase, additionally had to reach the genome-wide significance level after combining discovery and replication results.

Metabolite lookup

All detected loci were checked for their associations with a set of 276 metabolites and 37,179 metabolite ratios in 1,768 individuals of the KORA F4 study. Association results of the metabolite panel with approximately 2.5 million HapMap II SNPs is provided at `http://metabolomics.helmholtz-muenchen.de/gwa/` and described elsewhere[70]. We searched the database for all associations between the index SNPs and one of the metabolites or metabolite ratios with p-values below 5×10^{-6}.

X chromosome analysis

In addition to the analysis of all autosomal chromosomes, we analysed the X chromosome in a subset of studies. Imputed data was available in 19 of the discovery studies (AGES, ARIC, CHS, CoLaus, EPIC-Norfolk cohort, ERF, Estonian Biobank, FHS, INCIPE, INGI-CILENTO, INGI380 Val Borbera, KORA F3, KORA F4, LBC1936, NESDA, RS-I, RS-II, SardiNIA, SHIP). Six additional studies contributed data for genotyped SNPs only (Amish, AUSTWIN, BLSA, InCHIANTI, INGI-Carlantino, INGI-FVG). Imputation was performed using MACH[49] or IMPUTE[54] with HapMapII as a reference panel. Linear regression models for the SNP associations were calculated in the same way as for autosomal SNPs, with men coded as homozygous in the non-pseudoautosomal region. Quality checks and meta-analysis were performed analogous to the autosomal analysis. After the meta-analysis, we additionally applied a MAF filter of 5%, as well as a p-value filter of 5×10^{-8} for Cochran's heterogeneity test due to the higher observed heterogeneity compared to the autosomal SNPs. In total, 54,926 SNPs were analysed

in up to 72,026 individuals in the overall analysis, 52,897 SNPs in 39,212 women, and 55,210 SNPs in 31,086 men.

Network analysis

In order to trace possible underlying biological processes, as well as to detect additionally involved genes in a systematic approach, we performed a network analysis following up the GWAS analysis. In a first step, all independent index SNPs, not only at a genome-wide significance level, but with p-values below the suggestive significance level of 1×10^{-6} in the overall GWAS, were assigned to a gene. This step was done using GRAIL[64]. For rs1493664, no gene could be assigned by GRAIL, thus, we selected *LUZP2* which is the closest gene to this locus. For the resulting list of 37 genes, which we named "seed genes" in the network, we searched for known protein-protein interactions (PPI) in the STRING database[72]. We allowed PPIs which were verified in databases or experiments but did not use text mining as an additional option. For 31 of the genes, PPIs with other genes were known. No PPIs could be found for *TMEM171*, *MUSTN1*, *C17orf82*, *STC1*, *SLC16A9*, and *LUZP2*. All known interactions in a 1-, 2-, and 3-neighbourhood were assigned to a network graph. The complete graphs are available at `http://www.gwas.eu/gugc`. Sub-networks are visualized in Supplementary Figures 8 and 9 of Köttgen *et al.*[45]

For the systematic approach, we further investigated all genes assigned to the 1-neighbourhood network. The 1-neighbourhood network included 814 genes in addition to the 31 seed genes. For each of the 814 genes identified by the network, we defined a window extending in 110 kb upstream and 40 kb downstream direction of the gene as suggested before[66] and searched for the SNP with the smallest p-value in the results of the serum urate GWAS within this region. The resulting SNP list contained 735 SNPs since 33 windows contained no such SNP and because some of the SNPs were selected for multiple windows. Of those 735 SNPs, 27 passed a Bonferroni corrected significance level of $\alpha = 0.05/735 = 6.80\times10^{-5}$. Of those, we removed SNPs located within a 1 Mb-wide window around each seed gene (or 2.5 Mb-wide for the *HLA* locus) and SNPs in high LD with other SNPs in the list. This resulted in a list of 17 independent SNPs. These 17 SNPs were followed up in the replication step in addition to the SNPs discovered by the GWAS directly. The strategy of the systematic network approach is summarised in Figure 7.

In order to address the question how the network approach performs in comparison to a random selection of SNPs, we compared the 17 network SNPs to randomly selected SNP sets of 17 SNPs within the same p-value range. There were 2,210 SNPs with p-values between 6.8×10^{-5} and 1×10^{-6} in our screen. After exclusion of the already investigated regions, these could be grouped into 115 independent loci using the same clumping algorithm as for the genome-wide significant SNPs as described in Köttgen *et al.*[45] From the 115 independent loci, the SNPs with the smallest p-value per locus were used to randomly select 100,000 sets of 17 SNPs. Afterwards the rank-sums of the p-values from the 100,000 SNP sets were compared to the rank-sum of the p-values of the 17 network SNPs. For this comparison p-values for the 115 SNPs were taken from the meta-analysis results combining all

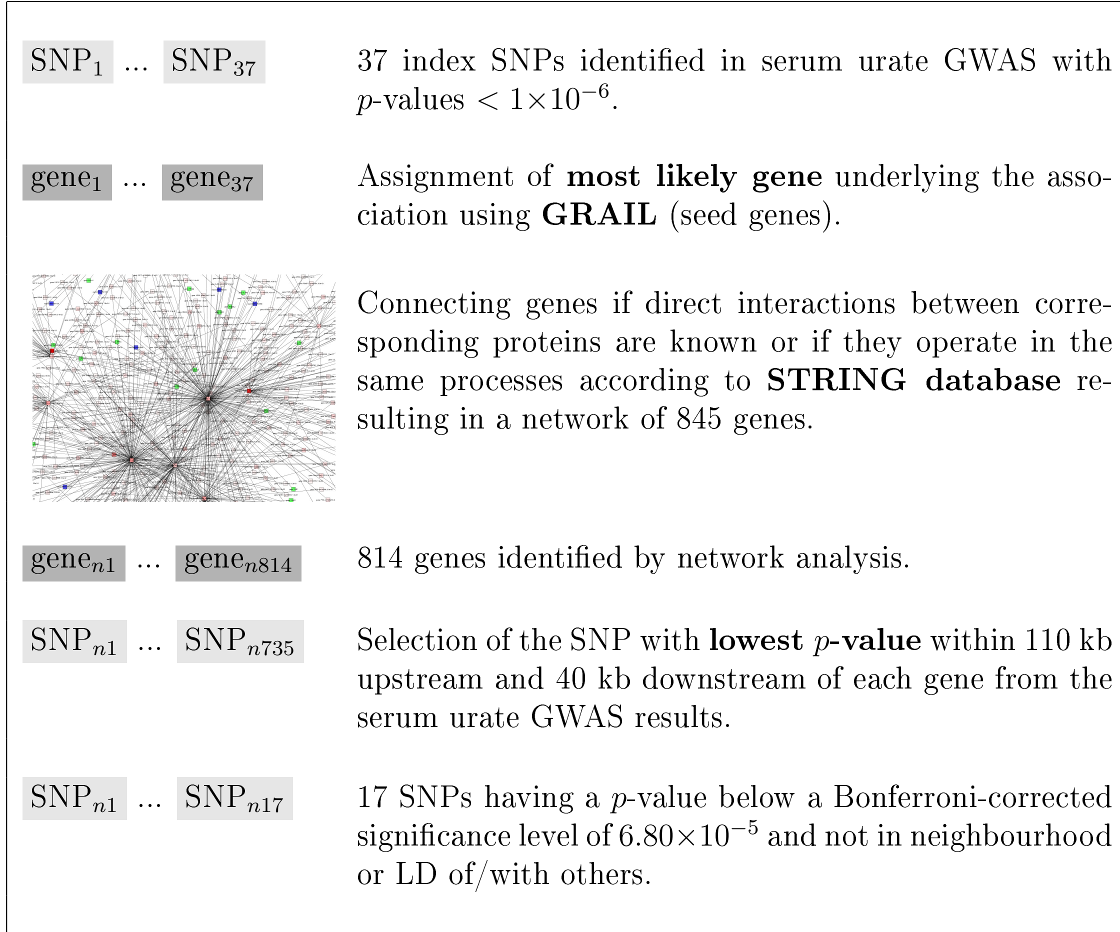


Figure 7: **Strategy of systematic network approach**.

in silico replication studies.

4.2 Results

In a meta-analysis of 48 genome-wide scans, totalling 110,347 individuals of European descent, 2,201 SNPs showed an association with serum urate concentrations at the genome-wide significance level of $5{\times}10^{-8}$. Those SNPs are distributed across 26 independent loci and located in or near *SLC2A9*, *ABCG2*, *SLC17A1*, *GCKR*, *SLC22A11*, *PDZK1*, *SLC16A9*, *INHBC*, *RREB1*, *HNF4G*, *SFMBT1*, *TRIM46*, *OVOL1*, *IGF1R*, *VEGFA*, *A1CF*, *BAZ1B*, *UBE2Q2*, *ATXN2*, *NRXN2*, *TMEM171*, *HLF*, *BCAS3*, *ORC4L*, *INHBB*, and *NFAT5*. Of those 26 loci *SLC22A11* and *NRXN2* are two independent signals within the same region. 380 further SNPs reached a suggestive significance level of $1{\times}10^{-6}$, comprising eleven additional independent genetic loci near *STC1*, *MAF*, *ADPGK*, *INSR*, *USP2*, *DACH1*, *QRICH2*, *FGF5*, *B4GALT1*, *LUZP2*, and *PRKAG2*. In the sex stratified analysis, no additional region reached the genome-wide significance level. However, five additional regions reached the suggestive significance level for women near *HNF1A*, *DAB2*, *MC4R*, *FRK*, and *ANKRD55*, and one additional region for men near *HLA-DRB5*. The GWAS results of the overall, as well as of the sex

stratified analyses are visualized as truncated Manhattan plots in Figure 8. The 37 regions from the overall analysis as well as the six regions from the sex-stratified analysis which reached at least a suggestive significance level of 1×10^{-6}, are shown as regional association plots in supplementary Figure S1. Association results for the best SNP of each region (index SNP) are shown in Table 8.

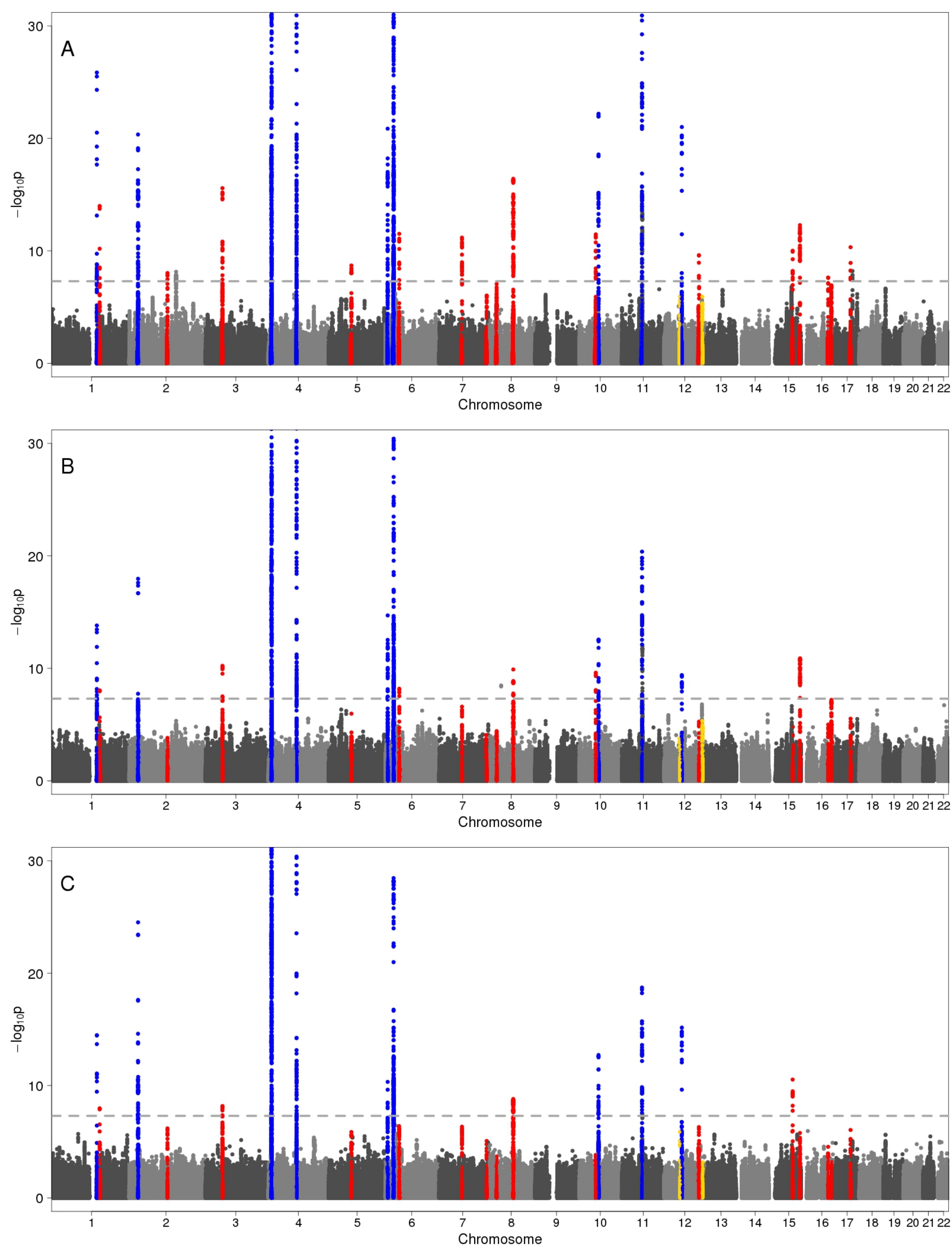


Figure 8: **Manhattan plots of GUGC GWAS.** Results of the meta-analysis are plotted as $-\log_{10}$ p-values ordered by their chromosomal position in A) the overall analysis, B) women only, and C) men only. Previously known loci are coloured in blue whereas all novel identified loci are coloured in red. Replicated loci identified by the network analysis are coloured in orange. The grey dashed line indicates the significance level at 5×10^{-8}. The plots are truncated at 1×10^{-30}.

Table 8: **Association results within GUGC.** The first allele is the effect allele. The given gene is the closest gene to the index SNP and can differ from the GRAIL gene used for the network analysis. *The p-value was $<1\times10^{-700}$. For replicated SNPs rs number and gene name are in bold font.

						Discovery			Replication				Combined	
SNP	**Chr**	**Position**	**Gene**	**Alleles**	**EAF**	**N**	**Beta**	p**-value**	**N**	**Beta**	p**-value**	q**-value**	**Beta**	p**-value**
Overall GWAS														
rs1471633	1	144435096	***PDZK1***	A/C	0.46	108616	0.061	1.4×10^{-26}	17229	0.047	6.3×10^{-5}	2.3×10^{-4}	0.059	1.2×10^{-29}
rs11264341	1	153418117	***TRIM46***	T/C	0.43	105030	-0.048	1.0×10^{-14}	16616	-0.060	4.7×10^{-6}	2.2×10^{-5}	-0.050	6.2×10^{-19}
rs1260326	2	27584444	***GCKR***	T/C	0.41	110195	0.077	1.3×10^{-40}	16856	0.060	8.7×10^{-7}	4.6×10^{-6}	0.074	1.2×10^{-44}
rs17050272	2	121022910	***INHBB***	A/G	0.43	103514	0.037	9.4×10^{-9}	24752	0.031	2.4×10^{-3}	4.4×10^{-3}	0.035	1.6×10^{-10}
rs2307394	2	148432898	*ORC4L*	T/C	0.68	109642	-0.035	7.3×10^{-9}	28960	-0.011	1.4×10^{-1}	1.2×10^{-1}	-0.029	2.2×10^{-8}
rs6770152	3	53075254	***SFMBT1***	T/G	0.58	109889	-0.048	2.7×10^{-16}	17675	-0.025	2.9×10^{-2}	3.7×10^{-2}	-0.044	2.6×10^{-16}
rs12498742	4	9553150	***SLC2A9***	A/G	0.77	110041	0.379	0*	18786	0.343	8.2×10^{-102}	3.1×10^{-100}	0.373	0*
rs11099098	4	81388936	*FGF5*	T/G	0.29	103351	-0.033	7.6×10^{-7}	26334	-0.015	1.1×10^{-1}	1.0×10^{-1}	-0.029	6.5×10^{-7}
rs2231142	4	89271347	***ABCG2***	T/G	0.11	110093	0.221	4.4×10^{-116}	16958	0.197	1.1×10^{-20}	2.0×10^{-19}	0.217	1.0×10^{-134}
rs17632159	5	72467238	***TMEM171***	C/G	0.31	108153	-0.038	2.0×10^{-9}	22010	-0.039	1.8×10^{-3}	3.5×10^{-3}	-0.039	3.5×10^{-11}
rs675209	6	7047083	***RREB1***	T/C	0.27	109893	0.063	1.4×10^{-21}	16860	0.051	1.2×10^{-3}	2.6×10^{-3}	0.061	1.3×10^{-23}
rs1165151	6	25929595	***SLC17A1***	T/G	0.47	109945	-0.093	4.5×10^{-60}	17914	-0.087	6.6×10^{-13}	8.3×10^{-12}	-0.091	7.0×10^{-70}
rs729761	6	43912549	***VEGFA***	T/G	0.30	108706	-0.046	3.1×10^{-12}	26930	-0.049	2.3×10^{-5}	9.0×10^{-5}	-0.047	8.0×10^{-16}
rs1178977	7	72494985	***BAZ1B***	A/G	0.81	109469	0.050	6.7×10^{-12}	22170	0.032	1.6×10^{-2}	2.3×10^{-2}	0.047	1.2×10^{-12}
rs10480300	7	151036938	***PRKAG2***	T/C	0.28	108251	0.032	9.4×10^{-7}	18188	0.050	2.9×10^{-4}	7.4×10^{-4}	0.035	4.1×10^{-9}
rs17786744	8	23832951	***STC1***	A/G	0.58	109956	-0.031	8.8×10^{-8}	26014	-0.022	2.4×10^{-2}	3.3×10^{-2}	-0.029	1.4×10^{-8}
rs2941484	8	76641323	***HNF4G***	T/C	0.44	109903	0.049	3.9×10^{-17}	21120	0.022	2.7×10^{-2}	3.6×10^{-2}	0.044	4.4×10^{-17}
rs10813960	9	33170362	*B4GALT1*	T/C	0.29	102248	-0.033	7.9×10^{-7}	25607	-0.019	5.0×10^{-2}	5.6×10^{-2}	-0.030	3.5×10^{-7}
rs10821905	10	52316099	***A1CF***	A/G	0.18	107809	0.053	3.4×10^{-12}	25777	0.072	7.7×10^{-7}	4.6×10^{-6}	0.057	7.4×10^{-17}
rs1171614	10	61139544	***SLC16A9***	T/C	0.22	103697	-0.074	6.5×10^{-23}	10695	-0.117	4.6×10^{-8}	3.5×10^{-7}	-0.079	2.3×10^{-28}
rs1493664	11	25657565	*LUZP2*	T/C	0.44	105525	-0.029	8.3×10^{-7}	30797	0.000	5.2×10^{-1}	3.0×10^{-1}	-0.021	2.3×10^{-5}
rs2078267	11	64090690	***SLC22A11***	T/C	0.51	97905	-0.078	8.7×10^{-36}	18000	-0.050	9.5×10^{-5}	3.0×10^{-4}	-0.073	9.4×10^{-38}
rs478607	11	64234639	***NRXN2***	A/G	0.84	109621	-0.049	5.3×10^{-10}	17969	-0.039	1.0×10^{-2}	1.5×10^{-2}	-0.047	4.4×10^{-11}
rs642803	11	65317196	*OVOL1*	T/C	0.46	109895	-0.043	4.5×10^{-14}	31054	-0.016	4.6×10^{-2}	5.3×10^{-2}	-0.036	2.9×10^{-13}
rs2195525	11	118740614	*USP2*	T/C	0.52	110105	-0.031	2.6×10^{-7}	26906	-0.004	3.5×10^{-1}	2.1×10^{-1}	-0.025	2.4×10^{-6}
rs3741414	12	56130316	***INHBC***	T/C	0.24	102628	-0.071	9.8×10^{-22}	11770	-0.081	1.3×10^{-5}	5.5×10^{-5}	-0.072	2.2×10^{-25}
rs653178	12	110492139	***ATXN2***	T/C	0.51	110119	-0.036	2.4×10^{-10}	27530	-0.028	4.1×10^{-3}	6.5×10^{-3}	-0.035	7.2×10^{-12}
rs584480	13	71243506	*DACH1*	T/C	0.40	109721	-0.030	2.9×10^{-7}	27460	-0.001	4.6×10^{-1}	3.0×10^{-1}	-0.023	8.9×10^{-6}
rs4777542	15	70869419	*ADPGK*	T/C	0.32	105438	-0.033	1.5×10^{-7}	22289	-0.014	1.2×10^{-1}	1.0×10^{-1}	-0.029	2.2×10^{-7}
rs1394125	15	73946038	***UBE2Q2***	A/G	0.34	105463	0.043	9.8×10^{-11}	21367	0.045	2.6×10^{-4}	7.2×10^{-4}	0.043	2.5×10^{-13}
rs6598541	15	97088658	***IGF1R***	A/G	0.36	109224	0.044	5.2×10^{-13}	16661	0.043	1.1×10^{-3}	2.5×10^{-3}	0.043	4.8×10^{-15}

						Discovery			Replication				Combined	
SNP	**Chr**	**Position**	**Gene**	**Alleles**	**EAF**	**N**	**Beta**	***p*-value**	**N**	**Beta**	***p*-value**	***q*-value**	**Beta**	***p*-value**
rs7193778	16	68121391	***NFAT5***	T/C	0.86	109915	-0.047	2.4×10^{-8}	27295	-0.044	4.9×10^{-3}	7.5×10^{-3}	-0.046	8.2×10^{-10}
rs7188445	16	78292488	***MAF***	A/G	0.33	109775	-0.032	1.2×10^{-7}	27348	-0.032	2.0×10^{-3}	3.8×10^{-3}	-0.032	1.6×10^{-9}
rs7224610	17	50719787	***HLF***	A/C	0.58	110044	-0.038	4.7×10^{-11}	27544	-0.055	3.7×10^{-8}	3.5×10^{-7}	-0.042	5.4×10^{-17}
rs2079742	17	56820479	*BCAS3*	T/C	0.85	107523	0.051	6.2×10^{-9}	21418	0.019	9.7×10^{-2}	9.3×10^{-2}	0.043	1.2×10^{-8}
rs164009	17	71795264	*QRICH2*	A/G	0.61	109669	0.029	7.1×10^{-7}	22156	0.023	3.1×10^{-2}	3.9×10^{-2}	0.028	1.6×10^{-7}
rs1035942	19	7150803	*INSR*	A/G	0.27	109818	0.033	2.2×10^{-7}	31704	0.018	4.4×10^{-2}	5.3×10^{-2}	0.029	1.2×10^{-7}
Candidate gene approach														
rs4149178	6	43380166	*SLC22A7*	A/G	0.84	109910	-0.032	1.9×10^{-5}	26250	-0.039	3.4×10^{-3}	6.8×10^{-3}	-0.034	1.2×10^{-6}
GWAS in men														
rs11757159	6	32628250	*HLA-DRB5*	T/C	0.34	49757	-0.048	3.6×10^{-7}	7496	0.021	8.2×10^{-1}	3.6×10^{-1}	-0.038	1.5×10^{-5}
GWAS in women														
rs11954519	5	39938122	*DAB2*	A/T	0.73	56627	-0.041	4.6×10^{-7}	14383	-0.006	3.4×10^{-1}	6.1×10^{-1}	-0.033	3.9×10^{-6}
rs456867	5	55846849	*ANKRD55*	T/C	0.17	60491	-0.046	6.1×10^{-7}	14401	-0.010	2.9×10^{-1}	5.8×10^{-1}	-0.038	2.9×10^{-6}
rs1933737	6	116416980	*FRK*	T/C	0.69	57561	0.039	5.8×10^{-7}	14615	0.017	1.2×10^{-1}	2.4×10^{-1}	0.034	6.5×10^{-7}
rs2244608	12	119901371	*HNF1A*	A/G	0.68	59938	0.040	1.6×10^{-7}	17116	0.022	5.7×10^{-2}	1.1×10^{-1}	0.036	8.1×10^{-8}
rs12955983	18	56023969	*MC4R*	A/G	0.73	56340	-0.042	5.5×10^{-7}	12423	-0.010	2.9×10^{-1}	5.7×10^{-1}	-0.036	1.5×10^{-6}
Network analysis														
rs884080	1	2016609	*PRKCZ*	A/G	0.55	91054	-0.026	5.9×10^{-5}	20489	-0.001	4.8×10^{-1}	9.6×10^{-1}	-0.020	4.5×10^{-4}
rs4970988	1	149216686	*ARNT*	A/G	0.36	109878	-0.026	7.0×10^{-6}	22089	-0.034	1.5×10^{-3}	3.0×10^{-3}	-0.028	1.0×10^{-7}
rs10489401	1	185009570	*PTGS2*	A/G	0.67	110063	-0.026	4.7×10^{-5}	22388	-0.010	1.9×10^{-1}	3.9×10^{-1}	-0.022	6.3×10^{-5}
rs6707470	2	158054816	*ACVR1C*	A/G	0.98	97848	-0.092	6.6×10^{-5}	20926	-0.054	1.4×10^{-1}	2.8×10^{-1}	-0.085	4.5×10^{-5}
rs4972801	2	176591512	*HOXD12*	C/G	0.19	109272	0.031	5.6×10^{-5}	12492	0.031	6.8×10^{-2}	1.4×10^{-1}	0.031	1.7×10^{-5}
rs12468226	2	203045246	*BMPR2*	A/G	0.88	103117	-0.041	4.8×10^{-6}	15198	-0.027	1.1×10^{-1}	2.1×10^{-1}	-0.039	2.9×10^{-6}
rs300915	4	144493606	*GAB1*	A/G	0.38	109831	0.026	8.9×10^{-6}	21980	0.010	1.9×10^{-1}	3.8×10^{-1}	0.023	1.3×10^{-5}
rs4073745	5	176665947	*SLC34A1*	A/G	0.69	108617	-0.028	1.3×10^{-5}	20686	-0.012	1.8×10^{-1}	3.5×10^{-1}	-0.025	1.3×10^{-5}
rs7944548	11	492217	*HRAS*	T/C	0.31	97080	-0.027	5.3×10^{-5}	19152	-0.008	2.7×10^{-1}	5.4×10^{-1}	-0.023	9.8×10^{-5}
rs11056399	12	15256777	*PTPRO*	T/C	0.33	102778	0.026	3.9×10^{-5}	13700	0.023	7.4×10^{-2}	1.5×10^{-1}	0.025	1.3×10^{-5}
rs4760636	12	46459619	*HDAC7*	T/C	0.23	110120	0.030	8.1×10^{-6}	22165	0.016	1.1×10^{-1}	2.2×10^{-1}	0.027	5.7×10^{-6}
rs7976059	12	50537539	***ACVR1B***	T/G	0.35	108516	0.030	1.2×10^{-6}	22341	0.041	1.6×10^{-4}	3.1×10^{-4}	0.032	1.9×10^{-9}
rs7953704	12	121191945	***B3GNT4***	A/G	0.47	109200	-0.028	1.2×10^{-6}	23078	-0.031	2.6×10^{-3}	5.2×10^{-3}	-0.029	2.6×10^{-8}
rs11624421	14	75861575	*ESRRB*	C/G	0.16	109975	0.033	2.1×10^{-5}	22368	-0.022	9.2×10^{-1}	1.6×10^{-1}	0.022	1.7×10^{-3}
rs3751043	15	61635858	*USP3*	C/G	0.12	109520	0.038	2.4×10^{-5}	17269	-0.011	7.0×10^{-1}	7.0×10^{-1}	0.030	2.5×10^{-4}
rs2472297	15	72814933	*CSK*	T/C	0.24	107229	-0.033	1.3×10^{-5}	15538	-0.010	2.7×10^{-1}	5.3×10^{-1}	-0.029	2.5×10^{-5}
rs11574736	20	42472951	*HNF4A*	C/G	0.17	101159	0.039	4.8×10^{-5}	14091	0.018	2.0×10^{-1}	4.1×10^{-1}	0.035	4.9×10^{-5}

Within the network analysis, 17 additional loci showing p-values between 1×10^{-6} and 6.8×10^{-5} were selected (see section 4.1.2) and are presented in Table 8 as well. Together with one SNP identified in a candidate gene approach, as described in Köttgen *et al.*[45] and shown in Table 8, the 37 SNPs from the overall analysis, the 6 SNPs from the sex stratified analysis and the 17 SNPs from the network analysis were followed up in the replication analysis including up to 32,813 individuals of European descent. Of those 61 SNPs, 28 could be replicated, including the ten previously known regions near *PDZK1*, *GCKR*, *SLC2A9*, *ABCG2*, *RREB1*, *SLC17A1*, *SLC16A9*, *SLC22A11*, *NRXN2*, and *INHBC*, as well as 16 novel regions from the overall analysis near *TRIM46*, *INHBB*, *SFMBT1*, *TMEM171*, *VEGFA*, *BAZ1B*, *PRKAG2*, *STC1*, *HNF4G*, *A1CF*, *ATXN2*, *UBE2Q2*, *IGF1R*, *NFAT5*, *MAF*, and *HLF*, and two regions of the network analysis near *ACVR1B/ACVRL1*, and *B3GNT4*. All replicated loci are coloured in Figure 8 and marked in Table 8.

From the network approach, two genes out of 17 were replicated. One can raise the question if 2 out of 17 is better than chance. From the classical approach 26 out of 44 SNPs were replicated, but from those 26 which were replicated, 23 were already genome-wide significant in the discovery step. To address this question we compared the ranksum of the 17 network SNPs to the ranksums of randomly selected sets of 17 SNPs within the same p-value range. A comparison of those ranksums in the *in silico* data showed that 8% of the random SNP sets had a lower rank-sum compared to the network SNP set. This gives the hint that the network approach performs well in the selection of SNPs. Figure 9 shows the p-value distribution of the selected network SNPs in comparison to all SNPs within the same p-value range.

The *ALDH16A1* region, which was previously reported to be associated with serum urate levels in a whole-genome sequencing analysis[71] is shown in Figure 10. The smallest p-value in a ±250 kb window around *ALDH16A1* was observed for rs2288481 ($p = 5.8\times10^{-3}$).

In the analysis of the X chromosome which was performed in a subset of studies totalling up to 72,026 individuals, none of the SNPs reached the genome-wide significance level of 5×10^{-8} or the suggestive significance level of 1×10^{-6}. The same remained true when the analysis was stratified by sex. Within the two candidate regions of *PRPS1*[65] and *HPRT1*[83] the smallest p-value within a ±250 kb window around *PRPS1*, was nominally significant with $p = 2.9\times10^{-2}$ (rs5962404). The smallest p-value within a ±250 kb window around *HPRT1* was $p = 7.3\times10^{-2}$ for rs4830303. Results are shown in Figure 11.

In the investigation of all index SNPs for their association with a panel of serum glutamyl metabolites, not surprisingly, rs12498742 within *SLC2A9* showed a significant association with serum urate ($p = 1.6\times10^{-17}$) which was one of the measured metabolites. Due to the reduced sample size compared to the GWAS meta-analysis no other SNP showed a significant association with serum urate in this dataset. None of the additionally measured metabolites was associated with one of the SNPs directly. Nevertheless two SNPs showed associations with metabolite ratios: rs729761 (*VEGFA*) was associated with gamma-glutamylglutamate/ pyroglutamylglycine ($p = 4.2\times10^{-7}$), and rs653178 (*ATXN2*) was associated with gamma-

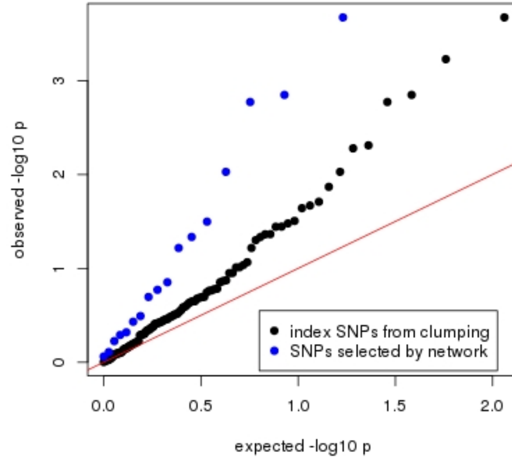


Figure 9: **Quantile-quantile plot comparing network SNPs** to SNPs in same p-value range. Observed p-values are plotted against p-values expected by chance. The 17 network SNPs are coloured in blue whereas the 115 SNPs with the lowest p-value of all independent regions within the same p-value range are coloured in black. Observed p-values were taken from the meta-analysis combining all *in silico* studies.

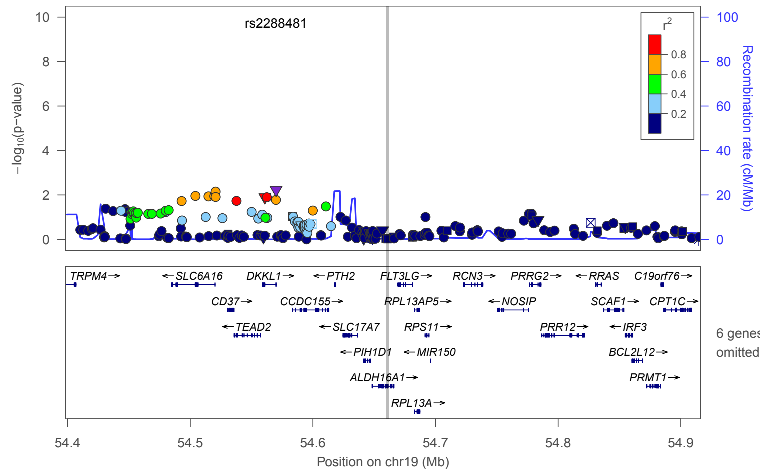


Figure 10: **Regional association plot of the *ALDH16A1* region in GUGC.** Results of the meta-analysis are plotted as $-\log_{10}$ p-values ordered by their chromosomal position in a ±250 kb window around *ALDH16A1*. Positions are given for NCBI build 36. The grey vertical line corresponds to the position of the rare variant identified by Sulem *et al.*[71]: chr19:54660818.

glutamylleucine/ valine ($p = 4.1 \times 10^{-8}$), as well as with gamma-glutamylleucine/ glucose ($p = 8.0 \times 10^{-7}$).

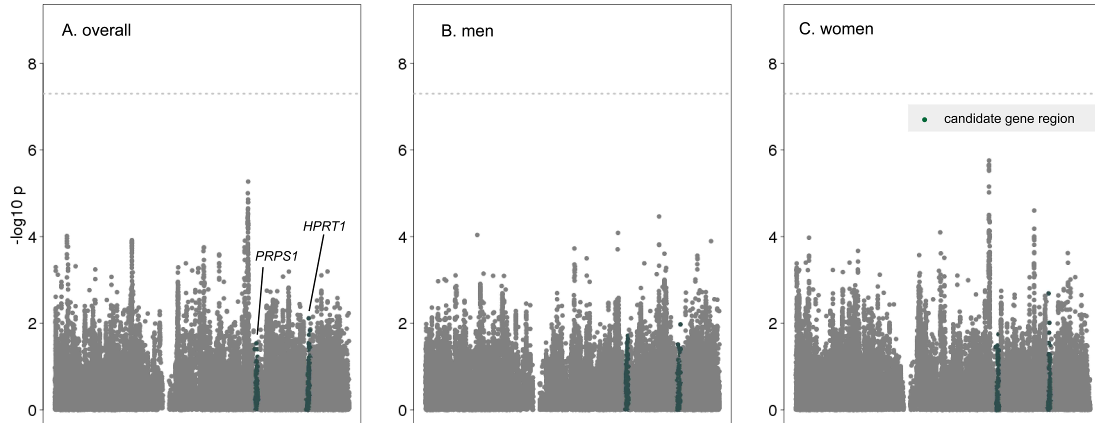


Figure 11: **X chromosomal Manhattan plot** showing $-\log_{10}$ p-values for all SNPs analysed on the X chromosome ordered by their chromosomal position. Results are shown for A) the overall analysis, B) men only, and C) women only. The locations of the two candidate gene regions *PRPS1* (left) and *HPRT1* (right) are coloured in green. The plot is taken from Köttgen *et al.*[45]

4.3 Discussion

With a total sample size of more than 110,000 individuals in the discovery GWAS, and a replication analysis including approximately 30,000 individuals, we have conducted the largest GWAS on serum urate levels until now. All regions known from previous GWAS could be replicated in this analysis. Within the ENGAGE GWAS, as described in chapter 3, *LRRC16A17* was found to be an independent signal from *SLC17A1*. This independency could not be confirmed within the GUGC data. Nevertheless two independent signals could be confirmed in the *SLC22A11* region, however, the closest gene to the second signal is *NRXN2* in the GUGC analysis instead of *SLC22A12* in the ENGAGE analysis.

Furthermore, we identified 18 novel regions, 16 of which in the common GWAS approach, and two additional in a systematic network approach. In total, 28 independent loci were identified and replicated. In relation to the discovery step, the sample size in the replication step is modest. Therefore, the failed replication of some of the genes, such as *ORC4L*, *OVOL1*, or *BCAS3*, which show genome-wide significant associations in the discovery step, could be due to a lack of power.

While previously identified genes are mainly involved in urate transport, none of the novel identified genes is known to act as a urate transporter or to be involved in urate transport indirectly. Several of the novel genes are encoding for transcription and growth factors and are therefore more likely connected to metabolic control of serum urate production and excretion. Interestingly, five of the replicated loci (*GCKR*, *VEGFA*, *STC1*, *ATXN2*, and *UBE2Q2*) and two of the loci which could not be replicated (*DACH1* and *BCAS3*), have been shown to be associated with chronic kidney disease before[46]. Two of them, *GCKR* and *ATXN2*, are known for their pleiotropic effect, as they showed associations in a broad number of GWAS

with various traits (`http://www.genome.gov/gwastudies`)[34].

Recently, Sulem *et al.*[71] reported in a whole-genome sequencing analysis, that rare variants within the *ALDH16A1* region are associated with serum urate levels. In our analysis, no signal was found within this region although the region contains several analysed SNPs. The conducted GWAS was based on HapMap II data mainly focusing on common variants, and may therefore fail to detect the influence of rare variants. This could also be the reason why we could not find any strong signal on the X chromosome despite two known candidate regions.

We applied a systematic network approach to select 17 genes within a p-value range between 1×10^{-6} and 6.80×10^{-5}. Two of the genes could be verified in the replication analysis and would have been missed in the GWAS approach alone. The systematic network approach provides a helpful strategy to select additional loci for replication which did not reach genome-wide significance in the discovery step. Furthermore, one clear advantage of the network approach is that it helps bringing the detected genes into a biological context as described in the text and in Supplementary Figures 8 and 9 of Köttgen *et al.*[45] The novel idea of selecting loci with p-values below 5×10^{-8} for replication by means of the systematic strategy of the network approach can be applied in future investigations. While the GWAS approach does not use any prior knowledge, the network approach incorporates prior biological knowledge in a systematic way and eases the biological interpretation of the findings.

In the related publication, the detected loci from the GWAS approach described in this thesis were followed up by a detailed characterization[45]. It could be shown that all urate-increasing alleles were positively associated with the risk of gout by investigating the effects in an additionally performed GWAS on gout. For *SLC2A9*, *GCKR*, and *IGF1R* associations with the fractional excretion of uric acid were shown. In pathway approaches, the findings were linked to glucose-metabolism, and the network analyses implicate the inhibins-activins signalling pathways to be involved in urate control.

Taken together, we identified 28 loci that are associated to serum urate levels in the largest GWAS on this topic to date. The findings are biologically plausible and highlight the importance of metabolic control in serum urate production and excretion.

My work presented in this chapter is also published in

Köttgen A, **Albrecht E**, Teumer A, Vitart V, Krumsiek J, Hundertmark C, Pistis G, Ruggiero D, O'Seaghdha CM, Haller T, Yang Q, Tanaka T, Johnson AD, Kutalik Z, Smith AV, Shi J, Struchalin M, Middelberg RP, Brown MJ, Gaffo AL, Pirastu N, Li G, Hayward C, Zemunik T, Huffman J, Yengo L, Zhao JH, Demirkan A, Feitosa MF, Liu X, Malerba G, Lopez LM, van der Harst P, Li X, Kleber ME, Hicks AA, Nolte IM, Johansson A, Murgia F, Wild SH, Bakker SJ, Peden JF, Dehghan A, Steri M, Tenesa A, Lagou V, Salo P, Mangino M, Rose LM, Lehtimäki T, Woodward OM, Okada Y, Tin A, Müller C, Oldmeadow C, Putku M, Czamara D, Kraft P, Frogheri L, Thun GA, Grotevendt A, Gislason GK, Harris TB, Launer LJ, McArdle P, Shuldiner AR, Boerwinkle E, Coresh J, Schmidt H, Schallert M, Martin NG, Montgomery GW, Kubo M, Nakamura Y, Tanaka T, Munroe PB, Samani NJ, Jacobs DR Jr, Liu K, D'Adamo P, Ulivi S, Rotter JI, Psaty BM, Vollenweider P, Waeber G, Campbell S, Devuyst O, Navarro P, Kolcic I, Hastie N, Balkau B, Froguel P, Esko T, Salumets A, Khaw KT, Langenberg C, Wareham NJ, Isaacs A, Kraja A, Zhang Q, Wild PS, Scott RJ, Holliday EG, Org E, Viigimaa M, Bandinelli S, Metter JE, Lupo A, Trabetti E, Sorice R, Döring A, Lattka E, Strauch K, Theis F, Waldenberger M, Wichmann HE, Davies G, Gow AJ, Bruinenberg M; LifeLines Cohort Study, Stolk RP, Kooner JS, Zhang W, Winkelmann BR, Boehm BO, Lucae S, Penninx BW, Smit JH, Curhan G, Mudgal P, Plenge RM, Portas L, Persico I, Kirin M, Wilson JF, Leach IM, van Gilst WH, Goel A, Ongen H, Hofman A, Rivadeneira F, Uitterlinden AG, Imboden M, von Eckardstein A, Cucca F, Nagaraja R, Piras MG, Nauck M, Schurmann C, Budde K, Ernst F, Farrington SM, Theodoratou E, Prokopenko I, Stumvoll M, Jula A, Perola M, Salomaa V, Shin SY, Spector TD, Sala C, Ridker PM, Kähönen M, Viikari J, Hengstenberg C, Nelson CP; CARDIoGRAM Consortium; DIAGRAM Consortium; ICBP Consortium; MAGIC Consortium, Meschia JF, Nalls MA, Sharma P, Singleton AB, Kamatani N, Zeller T, Burnier M, Attia J, Laan M, Klopp N, Hillege HL, Kloiber S, Choi H, Pirastu M, Tore S, Probst-Hensch NM, Völzke H, Gudnason V, Parsa A, Schmidt R, Whitfield JB, Fornage M, Gasparini P, Siscovick DS, Polašek O, Campbell H, Rudan I, Bouatia-Naji N, Metspalu A, Loos RJ, van Duijn CM, Borecki IB, Ferrucci L, Gambaro G, Deary IJ, Wolffenbuttel BH, Chambers JC, März W, Pramstaller PP, Snieder H, Gyllensten U, Wright AF, Navis G, Watkins H, Witteman JC, Sanna S, Schipf S, Dunlop MG, Tönjes A, Ripatti S, Soranzo N, Toniolo D, Chasman DI, Raitakari O, Kao WH, Ciullo M, Fox CS, Caulfield M, Bochud M, Gieger C. **Genome-wide association analyses identify 18 new loci associated with serum urate concentrations**. *Nat Genet.* 2013 Feb;45(2):145-54. doi: 10.1038/ng.2500. Epub 2012 Dec 23.

In this publication, the first authorship is shared between Anna Köttgen, myself, Alexander Teumer, Veronique Vitart, and Jan Krumsiek. My detailed contribution in this project was as follows:

Anna Köttgen, Christian Gieger, and I designed the analysis plan of this GWAS and organized the participation of all discovery and replication studies. I was part of the analysis group which discussed all aspects of the analysis in weekly telephone conferences.

At the study level, I performed the GWAS on serum urate and gout (not shown here) within KORA F3 and KORA F4, as well as the X chromosome analysis, the conditional analysis, and the risk score analysis (KORA F4, not shown here). Furthermore I performed the GWAS on serum urate and the conditional analysis in the LURIC discovery sample and the replication analysis in KORA S2.

Within the analysis team we shared the work of quality control and plausibility checks of all genome-wide scans. Afterwards I performed the meta-analysis for all overall and sex-stratified serum urate GWAS in parallel with Anna Köttgen, including filtering and sensitivity analyses.

In the replication step, I designed the iPlexes for *de novo* genotyping together with Norman Klopp and performed the quality control for all genotyped SNPs within the studies genotyped at Helmholtz Zentrum München (Ogliastra Genetic Park and KORA S2). In parallel with Anna Köttgen I performed the meta-analysis of all replication studies, as well as the meta-analysis of discovery and replication step.

I planned and organized the X chromosome analysis within all studies, made the quality control of all study specific result files, and conducted the meta-analysis in parallel with Claudia Hundertmark.

I had a major share in the development of the strategy in the systematic network approach. The network itself was created by Jan Krumsiek.

Furthermore, I looked up the associations with other traits, and the association between detected SNPs and metabolites.

Last but not least, I contributed to the interpretation of the results and the paper writing within the writing group.

In this thesis, I focus on my own contributions to all performed analyses within this huge collaboration.

5 Serum urate genetics in different ancestries

Both GWAS, performed in ENGAGE and GUGC, as described in chapter 3 and chapter 4, included samples of European ancestry only. As the genetic structure varies between different ancestries, it is not possible to draw direct conclusions from findings in Europeans to their impact in non-European ancestries. Within GUGC several data sets of non-European studies were available and we compared the estimates of all detected loci in Europeans to those of non-European ancestries (section 5.1). Furthermore, other consortia performed serum urate GWAS within studies of other ancestries. An overview is given in section 5.2 together with a comparison to the GUGC findings.

5.1 Comparison of different ancestries within GUGC

Materials and Methods

Within GUGC, seven studies of non-European ancestry were available as listed in Table 9. Detailed information about study design and genotyping is provided in supplementary Tables S2 and S3.

Table 9: **List of GUGC studies of non-European ancestry.** Serum urate levels are given in mg/dl.

Study	Sample size	Females %	Age mean (sd)	Serum urate mean (sd)
LOLIPOP-IA317 Indian	2,139	0	48.3 (10.5)	5.6 (1.3)
LOLIPOP-IA610 Indian	5,589	18.4	56.9 (10.0)	5.6 (1.5)
LOLIPOP-IA-P Indian	612	0	51.1 (8.3)	5.5 (1.3)
ARIC African American	2,749	62.9	53.0 (5.8)	6.3 (1.7)
CARDIA African American	937	60.9	24.0 (3.8)	5.1 (1.3)
JHS African American	2,134	60.8	50.0 (12.1)	5.5 (1.7)
BioBank Japan	15,288	43.6	63.0 (11.7)	5.4 (1.5)

GWAS on serum urate levels were performed at the study level as described for the European studies in section 4.1.2, and meta-analysed between three datasets of Indian ancestry (LOLIPOP), totalling 8,340 individuals, and the three studies of African American ancestry (ARIC, CARDIA, and JHS), totalling 5,820 individuals. Furthermore, 15,288 individuals of the BioBank Japan were available. The index SNPs of all replicated loci of the GUGC overall analysis in the European sample were looked up in the Indian, African American, and Japanese association results. Some SNPs were not available in all ancestry panels due to low MAF or quality issues. Therefore the number of performed tests was 74, leading to a Bonferroni corrected significance level of 6.7×10^{-4}.

Results and Discussion

Table 10 shows effect estimates together with p-values and allele frequencies in the samples of African American, Indian, and Japanese ancestry. Among 5,820 African Americans, *SLC2A9* ($p=8.6\times10^{-29}$) as well as *SLC22A11* ($p=2.2\times10^{-4}$) showed a significant association, which had already been shown before[9,73]. Among 8,340 individuals of Indian ancestry eight of the loci, namely, *SLC2A9* ($p=4.3\times10^{-21}$), *ABCG2* ($p=2.1\times10^{-16}$), *SLC22A11* ($p=1.3\times10^{-15}$), *GCKR* ($p=1.7\times10^{-10}$), *SLC17A1* ($p=5.0\times10^{-8}$), *RREB1*

Table 10: **Association results between different ancestries.** The first allele is the effect allele. The given gene is the closest gene to the index SNP.

					African American			Indian			Japanese		
SNP	**Chr**	**Position**	**Gene**	**Alleles**	**EAF**	**Beta**	***p*-value**	**EAF**	**Beta**	***p*-value**	**EAF**	**Beta**	***p*-value**
rs1471633	1	144435096	*PDZK1*	A/C	0.71	0.041	1.7×10^{-1}	0.62	0.076	3.4×10^{-4}	0.91	0.054	5.2×10^{-2}
rs11264341	1	153418117	*TRIM46*	T/C	0.25	-0.037	3.0×10^{-1}	0.47	-0.069	7.4×10^{-4}	0.76	-0.074	9.5×10^{-4}
rs1260326	2	27584444	*GCKR*	T/C	0.16	0.014	7.2×10^{-1}	0.26	0.140	1.7×10^{-10}	0.56	0.043	6.9×10^{-3}
rs17050272	2	121022910	*INHBB*	A/G	0.10	0.025	6.4×10^{-1}	0.31	0.063	6.9×10^{-3}	0.49	0.033	4.4×10^{-2}
rs6770152	3	53075254	*SFMBT1*	T/G	0.70	-0.047	1.1×10^{-1}	0.56	-0.039	4.7×10^{-2}	0.46	-0.016	3.1×10^{-1}
rs12498742	4	9553150	*SLC2A9*	A/G	0.48	0.298	8.6×10^{-29}	0.73	0.199	4.3×10^{-21}	0.99	0.230	7.2×10^{-3}
rs2231142	4	89271347	*ABCG2*	T/G	NA	NA	NA	0.08	0.391	2.1×10^{-16}	0.30	0.164	2.8×10^{-21}
rs17632159	5	72467238	*TMEM171*	C/G	0.19	-0.018	6.0×10^{-1}	0.34	-0.078	3.6×10^{-4}	0.28	-0.045	1.1×10^{-2}
rs675209	6	7047083	*RREB1*	T/C	0.48	0.057	4.9×10^{-2}	0.54	0.081	2.5×10^{-5}	0.93	0.018	5.7×10^{-1}
rs1165151	6	25929595	*SLC17A1*	T/G	0.13	-0.074	7.0×10^{-2}	0.49	-0.104	5.0×10^{-8}	0.17	-0.067	1.5×10^{-3}
rs729761	6	43912549	*VEGFA*	T/G	0.12	-0.029	4.8×10^{-1}	0.23	0.010	6.8×10^{-1}	0.07	0.002	9.7×10^{-1}
rs1178977	7	72494985	*BAZ1B*	A/G	0.73	-0.004	9.1×10^{-1}	0.87	0.025	4.0×10^{-1}	0.89	0.058	2.5×10^{-2}
rs10480300	7	151036938	*PRKAG2*	T/C	0.23	0.030	4.0×10^{-1}	NA	NA	NA	NA	NA	NA
rs17786744	8	23832951	*STC1*	A/G	0.67	-0.002	9.6×10^{-1}	0.84	-0.039	1.4×10^{-1}	0.75	-0.018	3.2×10^{-1}
rs2941484	8	76641323	*HNF4G*	T/C	0.71	0.038	2.2×10^{-1}	0.38	0.029	1.5×10^{-1}	0.43	0.050	1.8×10^{-3}
rs10821905	10	52316099	*A1CF*	A/G	0.28	0.026	4.2×10^{-1}	0.23	-0.017	4.8×10^{-1}	0.05	0.075	4.2×10^{-2}
rs1171614	10	61139544	*SLC16A9*	T/C	0.26	-0.054	1.5×10^{-1}	0.16	-0.085	1.7×10^{-3}	0.01	-0.062	7.2×10^{-1}
rs2078267	11	64090690	*SLC22A11*	T/C	0.15	-0.150	2.2×10^{-4}	0.41	-0.153	1.3×10^{-15}	0.01	0.043	6.4×10^{-1}
rs478607	11	64234639	*NRXN2*	A/G	0.45	-0.089	1.5×10^{-3}	0.87	-0.012	6.7×10^{-1}	0.78	-0.112	4.6×10^{-9}
rs3741414	12	56130316	*INHBC*	T/C	0.10	-0.159	1.1×10^{-3}	0.10	-0.107	1.6×10^{-3}	0.08	-0.020	4.9×10^{-1}
rs653178	12	110492139	*ATXN2*	T/C	0.92	-0.061	2.6×10^{-1}	NA	NA	NA	1.00	0.760	2.7×10^{-1}
rs1394125	15	73946038	*UBE2Q2*	A/G	0.35	-0.012	7.0×10^{-1}	0.26	0.014	5.3×10^{-1}	0.08	0.021	4.8×10^{-1}
rs6598541	15	97088658	*IGF1R*	A/G	0.52	0.069	1.4×10^{-2}	0.48	-0.003	9.0×10^{-1}	0.50	0.033	3.8×10^{-2}
rs7193778	16	68121391	*NFAT5*	T/C	0.98	0.049	5.7×10^{-1}	0.88	-0.069	8.9×10^{-2}	0.90	-0.053	4.8×10^{-2}
rs7188445	16	78292488	*MAF*	A/G	0.27	-0.065	3.4×10^{-2}	0.14	0.051	8.5×10^{-2}	0.31	-0.060	4.5×10^{-4}
rs7224610	17	50719787	*HLF*	A/C	0.90	0.056	2.4×10^{-1}	0.69	-0.006	7.8×10^{-1}	0.82	-0.004	8.3×10^{-1}

($p = 2.5 \times 10^{-5}$), *PDZK1* ($p = 3.4 \times 10^{-4}$), and *TMEM171* ($p = 3.6 \times 10^{-4}$), were significantly associated with serum urate. Among 15,288 samples of the BioBank Japan, *ABCG2* ($p = 2.8 \times 10^{-21}$), *NRXN2* ($p = 4.6 \times 10^{-9}$), and *MAF* ($p = 4.5 \times 10^{-4}$) showed a significant association. In the region of *NRXN2*/*SLC22A12* the W258X mutation is known to cause hypouricemia in Japanese[42] and was seen genome-wide significantly associated with serum urate in a Japanese GWAS before[40]. Also, *ABCG2* and *SLC2A9* were reported to be associated with serum urate in Japanese[40]. However, the index SNP in *SLC2A9* found in the European sample, rs12498742, showed a p-value of 7.2×10^{-3} in the analysed Japanese sample and did therefore not meet the significance level of 6.7×10^{-4}.

Although only few SNPs show a significant association after correcting for multiple testing, most effects are consistent in their direction with the estimates in the European sample. A comparison between the effect estimates of all investigated SNPs between all four ancestries, together with corresponding allele frequencies, is shown in Figure 12. The figure indicates that respective allele frequencies vary considerably between ancestries; nevertheless, the effect estimates in all four ancestries are of identical direction and comparable in their effect size for the majority of SNPs. This might indicate that the genetic loci identified in Europeans do also play a role in other ancestries. Bigger sample sizes will be needed to confirm their impact in individuals of non-European ancestry. One must also consider the possibility that other variants in the same region might show stronger effects than the index SNPs discovered in Europeans.

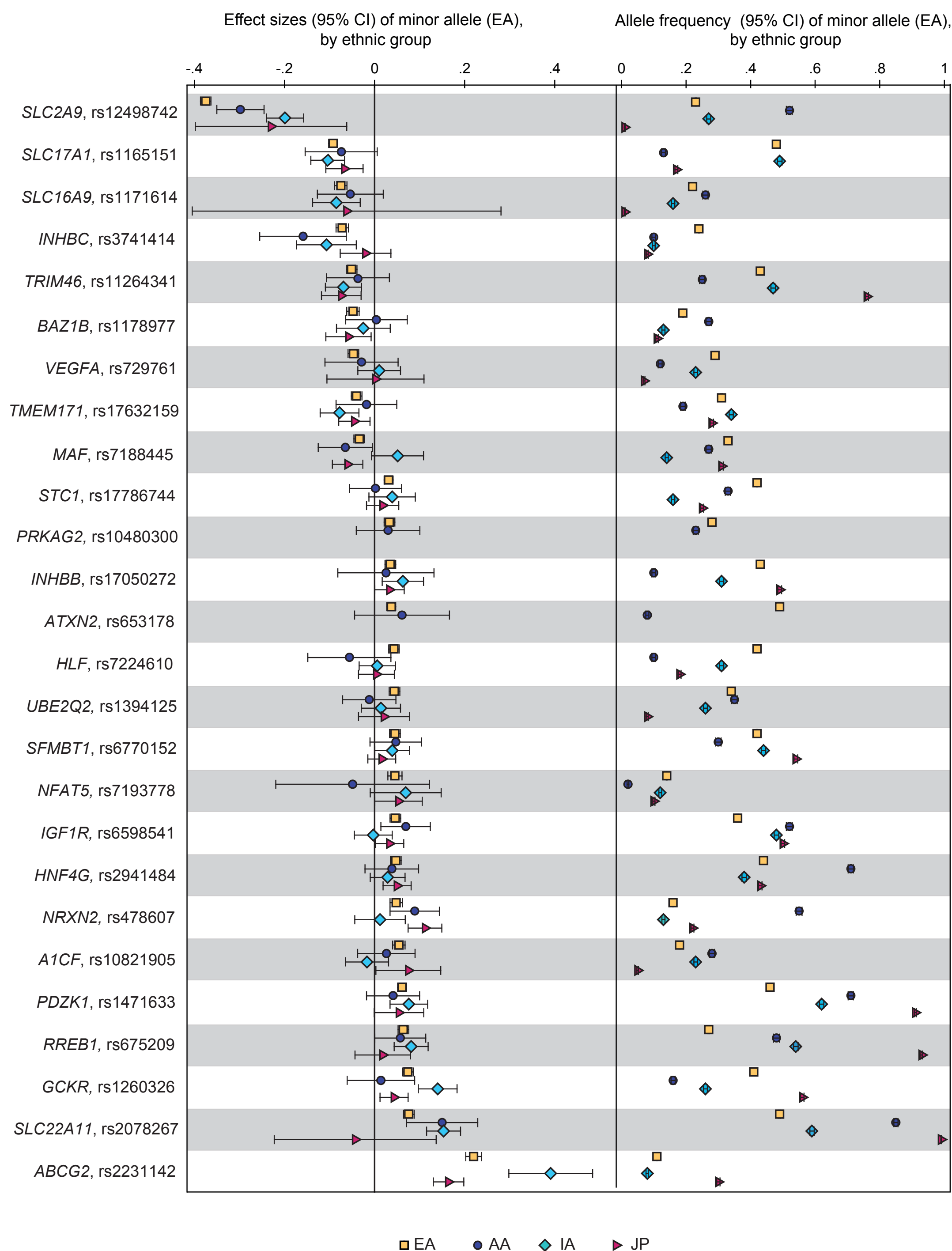


Figure 12: **Serum urate SNP effects between ancestries** comparing effect sizes and allele frequencies between individuals of European ancestry (EA), African American ancestry (AA), Indian ancestry (IA), and Japanese ancestry (JP) ordered by the effect size in Europeans. Effects and frequencies are shown for the minor allele in Europeans.

5.2 Serum urate GWAS performed in non-Europeans

In 2012, the Asian Genetic Epidemiology Network (AGEN) consortium performed a GWAS on serum urate levels as published by Okada *et al.*[60] With a total sample size of 33,074 individuals of Asian ancestry, they report variants in or near *SLC2A9* (rs3775948, $p = 1.6\times10^{-65}$), *SLC22A12* (rs504915, $p = 3.3\times10^{-63}$), *ABCG2* (rs504915, $p = 4.2\times10^{-30}$), and *MAF* (rs889472, $p = 1.1\times10^{-9}$) to be genome-wide significantly associated with serum urate levels. As our work within GUGC was not published by then, the *MAF* locus was a novel finding. We performed a lookup of those four SNPs in the GWAS results of the European GUGC analysis (see Supplementary Table 5 of Okada *et al.*[60]) Not surprisingly, in the European sample, the index SNPs of the Asian analysis also showed strong associations at *SLC2A9* (rs3775948, $p < 1\times10^{-600}$), *ABCG2* (rs2725220, $p = 7.6\times10^{-3}$), and *SLC22A11* (rs504915, $p = 1.6\times10^{-23}$). However, despite both scans showing a signal in the region of *MAF*, the index SNP in the Asian scan, rs889472, is not significantly associated in Europeans ($p = 0.23$). Figure 13 shows the results of both consortia in the region. The association signal observed in the Asian sample is about 100 kb closer to *MAF* than the signal observed in the European sample. The two regions showing the signal in Europeans or Asians are separated by a recombination hotspot.

A comparable situation is seen in the *LRP2* region. In 2010, Kamatani *et al.*[40] reported rs2544390 in *LRP2* to be genome-wide significantly associated with serum urate in 8,868 Japanese individuals. Within the GUGC results, rs2544390 is not associated with serum urate ($p = 0.218$). Nevertheless, one other SNP within *LRP2*, rs3815574, shows an association at $p = 1.3\times10^{-5}$ ($\beta = 0.0245$ for the A allele with EAF= 50.3%). According to the HapMap II CEU sample, rs2544390 and rs3815574 are not in LD ($r^2 = 0.002$). The association results for the whole *LRP2* region within the analysis of Kamatani *et al.* as well as within GUGC are shown in Figure 14. In the AGEN analysis, which also mainly includes individuals of Japanese ancestry, the region does not reach genome-wide significance though the sample size is much higher. The best SNP in this region within AGEN is rs2673172 with $p = 8.1\times10^{-5}$.

Furthermore, in 2011 Tin *et al.*[73] performed a GWAS on serum urate levels in 5,820 African Americans. This is the same African American dataset as available in GUGC and which was analysed in section 5.1. In their publication, they report rs9321453 close to *SGK1*/*SLC2A12* to be genome-wide significantly associated with serum urate. A lookup of the region in the GUGC results showed no evidence of a signal in Europeans (see Figure 15).

The examples show that results of an association analysis conducted in one ancestry may not directly be transferred to other ancestries as the genetic architecture varies between ancestries. Nevertheless, association results are frequently found in the same genetic region, even if the associated variants may not be the same but localized several base pairs apart.

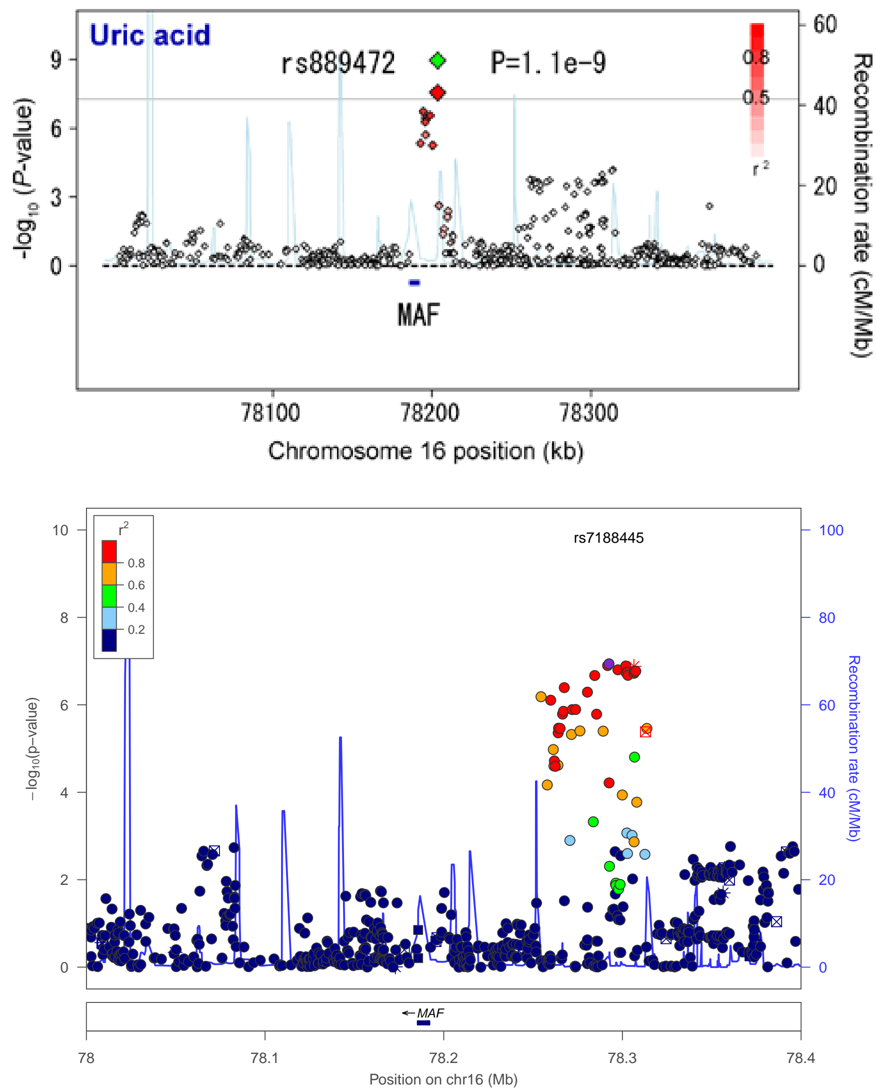


Figure 13: ***MAF* locus in Asians (top) and Europeans (bottom)**. The regional association plot showing the results in Asians is taken from Okada *et al.*[60] The regional association plot showing the results in Europeans is the result of the GUGC analysis.

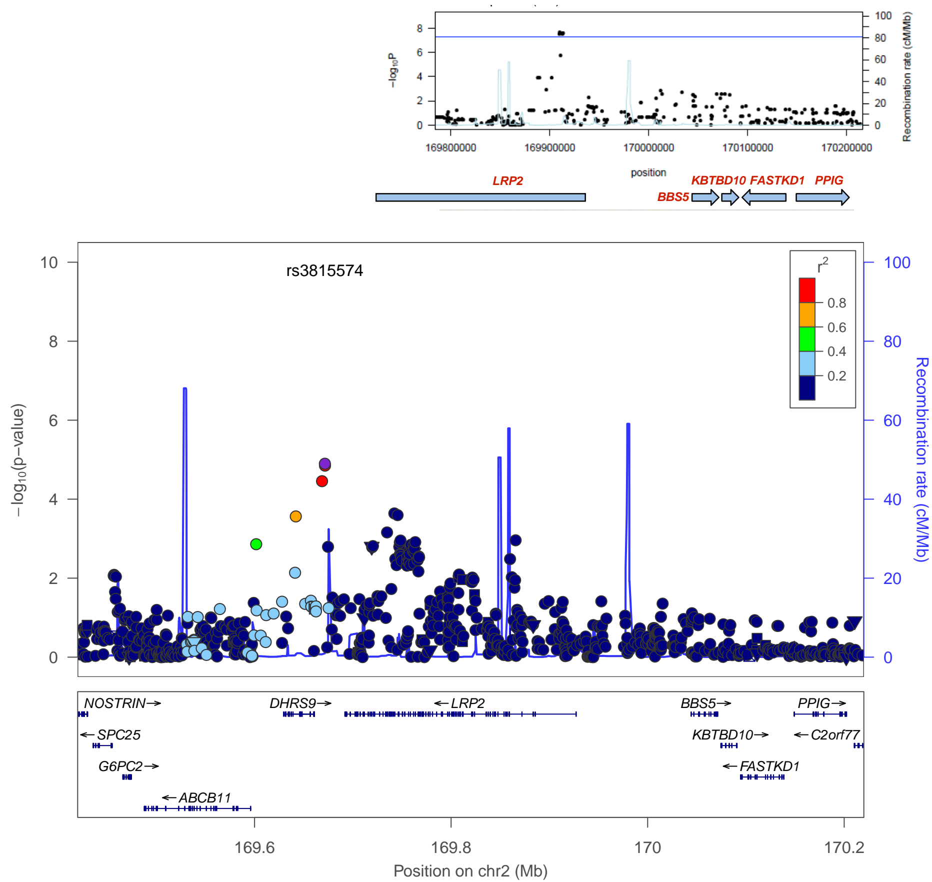


Figure 14: ***LRP2* locus in Asians (top) and Europeans (bottom)**. The regional association plot showing the results in Asians is taken from Kamatani *et al.*[40] The regional association plot showing the results in Europeans is the result of the GUGC analysis.

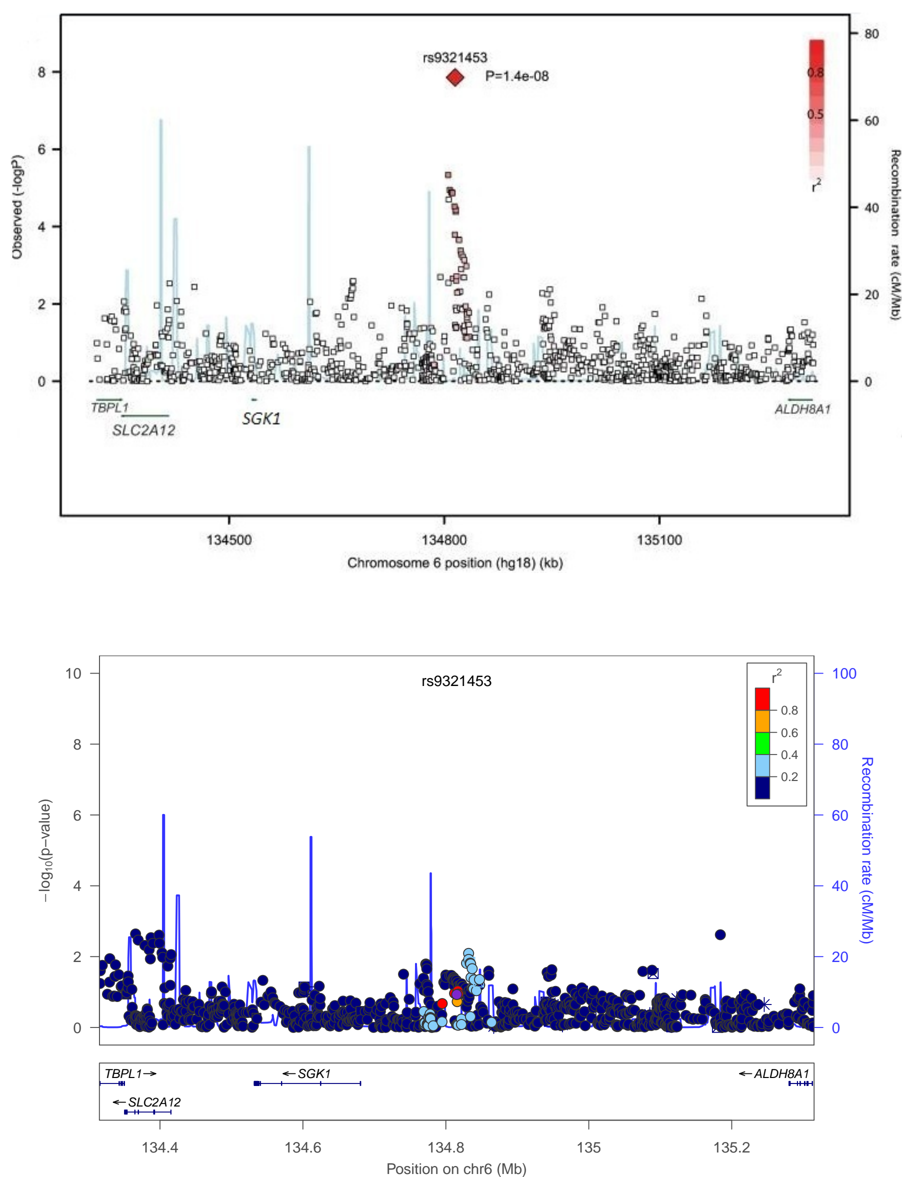


Figure 15: ***SGK1*/*SLC2A12* locus in African Americans (top) and Europeans (bottom)**. The regional association plot showing the results in African Americans is taken from Tin *et al.*[73] The regional association plot showing the results in Europeans is the result of the GUGC analysis.

6 Serum urate GGM

In chapter 3 to chapter 5 genetic components in the regulation of serum urate are investigated. Therefore, data of many studies organized in large consortia was combined. In the present chapter, metabolic components in the regulation of serum urate shall be investigated. Although the pathways in purine metabolism are well known (see Figure 1) not all underlying biological processes in the regulation of serum urate levels are completely understood. Therefore, we analysed the metabolic vicinity of serum urate within a metabolite network. In contrast to the consortia effort of the GWAS analyses, this analysis was conducted in one single study: the KORA F4 study.

6.1 Materials and methods

Study population

The KORA studies are a series of population-based studies from the South of Germany[35] which started as part of the MONICA (monitoring trends and determinants in cardiovascular disease) project. The first KORA survey (KORA S1) was conducted in 1984/85 including 4,022 participants from the general population living in the region of Augsburg, followed by KORA S2 in 1989/90 ($n = 4{,}940$), KORA S3 in 1994/95 ($n = 4{,}856$), and KORA S4 in 1999/2001 ($n = 4{,}261$). Participants of KORA S3 were invited to the follow-up examination KORA F3 in 2004/2005 ($n = 3{,}184$), and KORA S4 was followed up in 2006/08 by KORA F4 ($n = 3{,}080$). Written informed consent has been given by all participants and the study has been approved by the local ethics committee. Data of KORA S2, KORA F3, KORA S4, and KORA F4 contributed to the analysis in chapter 3 and/or chapter 4. Here, data of KORA F4 underlies the analysis, comprising 1,764 individuals (908 females and 856 males) in an age range of 32–81 years (mean: 60.86 years).

Metabolomics measurements

Metabolites were measured in serum by Metabolon Inc., as described in detail elsewhere[21,70]. The panel includes 517 metabolites, spanning several metabolic classes (amino acids, acylcarnitines, sphingomyelins, glycerophospholipids, carbohydrates, vitamins, lipids, nucleotides, peptides, xenobiotics, and steroids). The quantified metabolites can be distinguished into chemically identified metabolites, and unidentified, or “unknown” metabolites. Nine of those unknown metabolites have recently been identified by Krumsiek *et al.*[43] Urate is one of the measured metabolites on the panel. Metabolite levels are given as normalised ion counts.

Medication ascertainment

All KORA F4 participants were asked to bring their medications taken in the 7 days preceding the examination to the interview. Medication data was obtained online using the IDOM program (online drug-database led medication assessment). The medications were categorized according to the Anatomical Therapeutical Chemical (ATC) classification index. Of the 1,764 individuals in this study 83 were treated by urate lowering medication. All 83 were treated by allopurinol (uricostatic drug) and four of them additionally by benzbromaron (uricosuric drug).

Statistical analysis

After excluding metabolites with more than 20% missing values and samples with more than 10% missing values, the metabolite data matrix contained 1,764 samples and 355 metabolites (245 known and 110 unknown). Missing values were imputed using the 'mice' R package[75]. All normalized ion counts were transformed by natural logarithm. For each metabolite pair, partial correlations were calculated conditioning on age, sex, all other metabolites, and 473 SNPs which showed a significant association with at least one of the 355 known or unknown metabolites[44]. Partial correlations between two metabolites were considered to be significant with a p-value below 4.61×10^{-5}, which corresponds to a significance level of 0.05 and a correction for multiple testing by the false discovery rate (FDR)[4,5]. The resulting network is referred to as a GGM (see section 1.4). Within the GGM each node presents a metabolite and nodes are connected by an edge if their partial correlation is significant. The network was visualized in a 3-neighbourhood around urate, which means that metabolites were assigned to the network graph if they were connected to urate by a maximum of three edges. All metabolites within the 3-neighbourhood were further tested for associations with sex and urate lowering medication by means of a linear model which was additionally adjusted for age. Effects were considered to be significant below a threshold of 6.9×10^{-4}, which corresponds to a Bonferroni correction for 72 independent tests at a significance level of 0.05.

6.2 Results

By creating a 3-neighbourhood GGM around serum urate based on partial correlations, 36 metabolites were assigned to a network, containing 22 known as well as 14 unknown metabolites. The network is visualized in Figure 16. Table 11 shows the corresponding partial correlation coefficients and p-values for each of the edges within the network.

Table 11: **Partial correlation coefficients in serum urate GGM** for all significant associations within a 3-neighbourhood of serum urate.

Metabolite 1 - Metabolite 2	**Partial correlation coefficient**	p**-value**
urate - histidine	-0.228	1.90×10^{-12}
urate - methionine	0.181	2.70×10^{-8}
urate - N-[3-(2-oxopyrrolidin-1-yl)propyl]acetamide	0.142	1.30×10^{-5}
urate - androstene disulfate	0.154	2.40×10^{-6}
urate - X-11422 (xanthine)	-0.220	1.10×10^{-11}
histidine - methionine	0.189	6.10×10^{-9}
methionine - tyrosine	0.140	1.70×10^{-5}
androstene disulfate - dehydroepiandrosterone sulfate	0.352	1.30×10^{-28}
androstene disulfate - epiandrosterone sulfate	-0.173	1.00×10^{-7}
androstene disulfate - X-18601	0.145	8.80×10^{-6}
androstene disulfate - X-11440	0.350	2.60×10^{-28}
androstene disulfate - X-11443	0.510	4.40×10^{-63}
androstene disulfate - X-11450	0.156	1.60×10^{-6}
X-11422 (xanthine) - hypoxanthine	0.164	4.30×10^{-7}
X-11422 (xanthine) - xanthine	0.478	1.80×10^{-54}
tyrosine - 2-hydroxybutyrate	-0.138	2.40×10^{-5}
tyrosine - 3-(4-hydroxyphenyl)lactate	0.322	5.50×10^{-24}

Metabolite 1 - Metabolite 2	Partial correlation coefficient	p-value
tyrosine - caffeine	0.133	$4.30{\times}10^{-5}$
tyrosine - citrate	-0.138	$2.20{\times}10^{-5}$
tyrosine - gamma-glutamyltyrosine	0.464	$4.90{\times}10^{-51}$
tyrosine - phenylalanine	0.199	$9.20{\times}10^{-10}$
tyrosine - tryptophan	0.271	$3.30{\times}10^{-17}$
dehydroepiandrosterone sulfate - epiandrosterone sulfate	0.294	$4.40{\times}10^{-20}$
dehydroepiandrosterone sulfate - X-18601	0.575	$3.10{\times}10^{-83}$
dehydroepiandrosterone sulfate - X-11315	0.140	$1.80{\times}10^{-5}$
dehydroepiandrosterone sulfate - X-11443	-0.469	$2.50{\times}10^{-52}$
dehydroepiandrosterone sulfate - X-11450	0.391	$1.60{\times}10^{-35}$
dehydroepiandrosterone sulfate - X-12063	-0.194	$2.20{\times}10^{-9}$
dehydroepiandrosterone sulfate - X-12844	0.138	$2.10{\times}10^{-5}$
epiandrosterone sulfate - androsterone sulfate	0.755	$1.90{\times}10^{-173}$
epiandrosterone sulfate - X-09789	0.133	$4.50{\times}10^{-5}$
epiandrosterone sulfate - X-11440	-0.173	$1.10{\times}10^{-7}$
epiandrosterone sulfate - X-11443	0.408	$6.70{\times}10^{-39}$
epiandrosterone sulfate - X-12844	0.156	$1.60{\times}10^{-6}$
X-18601 - taurolithocholate 3-sulfate	-0.133	$4.60{\times}10^{-5}$
X-18601 - X-12063	0.209	$1.20{\times}10^{-10}$
X-18601 - X-12844	-0.162	$6.90{\times}10^{-7}$
X-11440 - X-11445	0.313	$9.50{\times}10^{-23}$
X-11440 - X-11450	0.150	$3.80{\times}10^{-6}$
X-11440 - X-11470	0.140	$1.60{\times}10^{-5}$
X-11440 - X-12844	0.193	$2.60{\times}10^{-9}$
X-11443 - X-11450	0.212	$5.30{\times}10^{-11}$
X-11443 - X-12844	-0.155	$2.00{\times}10^{-6}$
hypoxanthine - arginine	0.136	$2.90{\times}10^{-5}$
hypoxanthine - inosine	0.254	$3.50{\times}10^{-15}$
hypoxanthine - lactate	0.134	$4.20{\times}10^{-5}$
hypoxanthine - uridine	0.151	$3.70{\times}10^{-6}$
hypoxanthine - X-10810	0.164	$4.70{\times}10^{-7}$
hypoxanthine - X-12442	-0.136	$3.10{\times}10^{-5}$
2-hydroxybutyrate - lactate	0.160	$9.50{\times}10^{-7}$
3-(4-hydroxyphenyl)lactate - citrate	0.143	$1.20{\times}10^{-5}$

The general structure of the serum urate network clusters into three parts of connected metabolites. The first cluster contains mainly nucleotides, namely xanthine, hypoxanthine, inosine, and uridine, as well as arginine, lactate, and three unknown metabolites. It includes the well-known pathway from inosine via hypoxanthine and xanthine to urate (see Figure 1), though xanthine is not directly connected to urate and hypoxanthine but via the unknown metabolite X-11422. The central position of X-11422 in this well-known pathway induces speculations about its chemical identity.

Following the ideas for unknown identification in Krumsiek et al.[43], we defined possible candidates for X-11422 by considering its direct neighbours in the GGM, its mass, and its fragmentation spectrum: alloxanthine, which is the active agent of allopurinol, or xanthine itself displaying altered chromatographic characteristics. A co-elution spiking

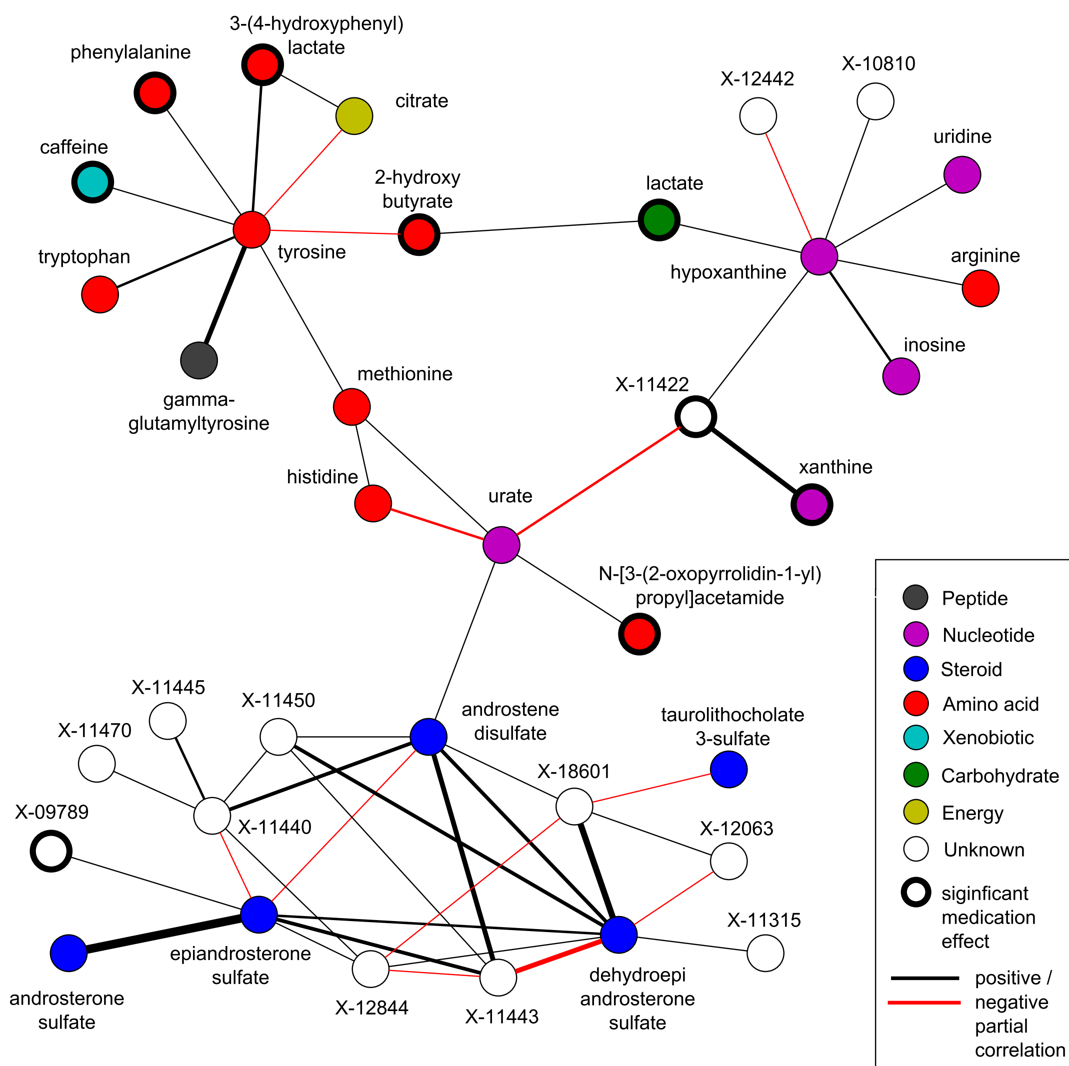


Figure 16: **Serum urate GGM** representing all significant associations within a 3-neighbourhood of serum urate. The thickness of each edge corresponds to the strength of partial correlation. Positive associations are marked as black lines whereas negative correlations are represented by red lines. Metabolites are coloured according to their biological pathways.

experiment demonstrated that xanthine, not alloxanthine, co-eluted perfectly in this matrix with the peak identified as X-11422, and therefore X-11422 represented an alternate measurement of xanthine.

A second cluster, in which several amino acids, namely histidine, methionine, tyrosine, tryptophan, phenylalanine, 3-(4-hydroxyphenyl)lactate, and 2-hydroxybutyrate group together with citrate, caffeine, and gamma-glutamyltyrosine, is connected via tyrosine and methionine to urate. In a third cluster, the steroids androsterone sulfate, epiandrosterone sulfate, and dehydroepiandrosterone sulfate are grouped with a number of unknown metabolites and are connected to urate via androstene disulfate. Furthermore, N-[3-(2-oxopyrrolidin-1-yl)propyl]acetamide is connected to urate without further connections to other metabolites within our panel.

Within our sample of 1,764 subjects, 83 were treated with urate lowering medication. For all metabolites within the network we tested the influence of sex and urate lowering medication within a linear model. Table 12 shows the corresponding effect estimates.
25 of the 36 metabolites show strong differences between men and women ($7.5\times10^{-5} \geq p \geq 8.1\times10^{-196}$). Nine of the metabolites show a significant influence of urate lowering medication. According to our data, the medication shows the strongest influence on the alternate measurement of xanthine X-11422 ($p = 7.1\times10^{-157}$) and the original measurement of xanthine ($p = 1.1\times10^{-85}$). Furthermore, the medication shows a significant influence on N-[3-(2-oxopyrrolidin-1-yl)propyl]acetamide ($p = 3.5\times10^{-12}$), phenylalanine ($p = 1.9\times10^{-7}$), caffeine ($p = 2.2\times10^{-7}$), 3-(4-hydroxyphenyl)lactate ($p = 4.1\times10^{-6}$), lactate ($p = 1.2\times10^{-4}$), 2-hydroxybutyrate ($p = 1.7\times10^{-4}$), and the unknown metabolite X-09789 ($p = 1.4\times10^{-5}$). Urate levels themselves do not show differences between medicated and medication-free individuals ($p = 0.18$) and also for hypoxanthine the influence of medication is much weaker than on xanthine and not significant after correcting for multiple testing ($p = 2.7\times10^{-3}$). Figure 17 visualizes the medication and sex effects for urate, both xanthine measurements, and hypoxanthine.

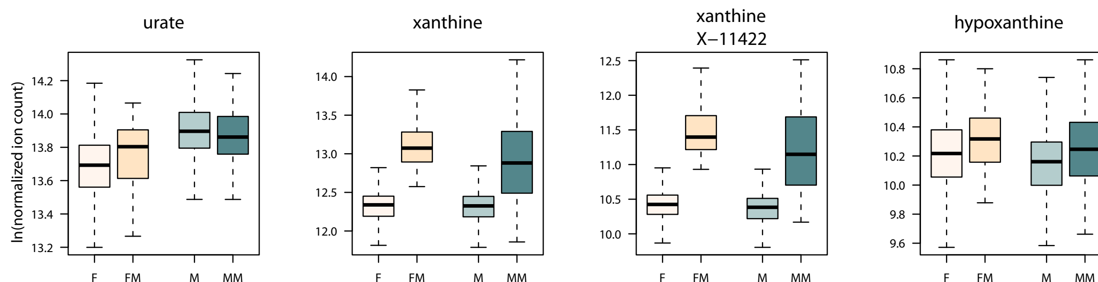


Figure 17: **Medication and sex effects on urate, xanthine, and hypoxanthine**: Levels of urate, xanthine, and hypoxanthine stratified by sex and medication; F=medication-free females ($n = 891$), FM=medicated females ($n = 17$), M=medication-free males ($n = 790$), and MM=medicated males ($n = 66$).

Table 12: **Medication and sex effects in serum urate GGM.** Influence of sex and urate lowering medication on levels of all metabolites within the 3-neighbourhood of serum urate. The linear model was additionally adjusted for age (effects not shown). For the sex effect, men were coded "0", whereas women were coded "1". Medication intake was coded with "1" compared to no medication intake "0".

	Sex		**Medication**	
Metabolite	**Beta**	***p*-value**	**Beta**	***p*-value**
2-hydroxybutyrate	-0.108	1.20×10^{-8}	0.169	1.70×10^{-4}
3-(4-hydroxyphenyl)lactate	-0.340	1.80×10^{-90}	0.175	4.10×10^{-6}
androstene disulfate	-0.817	8.80×10^{-103}	0.226	7.40×10^{-3}
androsterone sulfate	-0.360	7.90×10^{-23}	-0.092	2.80×10^{-1}
arginine	0.019	1.00×10^{-1}	-0.056	4.00×10^{-2}
caffeine	-0.074	1.30×10^{-1}	0.605	2.20×10^{-7}
citrate	0.029	2.10×10^{-2}	0.055	7.30×10^{-2}
dehydroepiandrosterone sulfate	-0.436	1.90×10^{-49}	-0.195	4.10×10^{-3}
epiandrosterone sulfate	-0.518	6.30×10^{-59}	-0.141	5.40×10^{-2}
gamma-glutamyltyrosine	-0.091	2.70×10^{-16}	0.076	3.50×10^{-3}
histidine	0.032	5.30×10^{-7}	-0.006	6.80×10^{-1}
hypoxanthine	0.056	7.50×10^{-5}	0.100	2.70×10^{-3}
inosine	0.226	4.70×10^{-9}	-0.034	7.10×10^{-1}
lactate	-0.072	4.80×10^{-10}	0.105	1.20×10^{-4}
methionine	-0.110	3.90×10^{-52}	-0.054	1.40×10^{-3}
N-[3-(2-oxopyrrolidin-1-yl) propyl]acetamide	-0.026	1.10×10^{-1}	0.275	3.50×10^{-12}
phenylalanine	-0.056	1.70×10^{-19}	0.076	1.90×10^{-7}
taurolithocholate 3-sulfate	-0.002	9.60×10^{-1}	0.249	2.90×10^{-3}
tryptophan	-0.080	3.00×10^{-32}	0.036	2.20×10^{-2}
tyrosine	-0.060	3.40×10^{-12}	0.041	4.10×10^{-2}
urate	-0.206	1.20×10^{-112}	-0.027	1.80×10^{-1}
uridine	0.009	3.40×10^{-1}	-0.054	2.00×10^{-2}
xanthine	0.010	4.20×10^{-1}	0.635	1.10×10^{-85}
X-09789	-0.089	4.90×10^{-3}	0.328	1.40×10^{-5}
X-10810	-0.061	3.60×10^{-3}	-0.058	2.40×10^{-1}
X-11315	0.151	1.40×10^{-13}	-0.048	3.20×10^{-1}
X-11422 (xanthine)	0.059	3.70×10^{-6}	0.896	7.10×10^{-157}
X-11440	-0.600	3.00×10^{-90}	0.069	3.00×10^{-1}
X-11443	-1.247	8.10×10^{-196}	0.166	5.60×10^{-2}
X-11445	-0.081	1.30×10^{-2}	-0.019	8.10×10^{-1}
X-11450	-0.512	1.60×10^{-88}	0.049	4.00×10^{-1}
X-11470	-0.158	1.30×10^{-17}	-0.102	1.90×10^{-2}
X-12063	-0.207	1.10×10^{-12}	0.207	2.70×10^{-3}
X-12442	0.126	8.00×10^{-8}	0.093	9.20×10^{-2}
X-12844	0.030	1.00×10^{-1}	-0.087	4.40×10^{-2}
X-18601	-0.550	2.10×10^{-71}	-0.134	5.60×10^{-2}

6.3 Discussion

The underlying functional biological processes in the regulation of serum urate are still not completely understood. They are assumed to result from a complex interplay between genetic, environmental, and lifestyle factors. The detection and functional characterization of such pathways is crucial to improve the management and treatment of patients with hyperuricemia and gout. In previous studies, GGMs have been demonstrated to reconstruct metabolic pathways from large-scale metabolomics data[43,44]. In order to describe the metabolic vicinity of serum urate, we analysed 355 metabolites in 1,764 individuals of the KORA F4 study and constructed a data-driven metabolite network around serum urate using GGMs. This hypothesis-free approach highlights pathways that may be important in the regulation of serum urate concentrations.

Not surprisingly, serum urate was linked to purine metabolism. The conversion of hypoxanthine to xanthine and xanthine to urate (see Figure 1) is catalysed by the rate-limiting enzyme xanthine oxidase, the only enzyme capable of catalysing the formation of urate in human[62]. The complex mechanism by which xanthine oxidase catalyses hypoxanthine and xanthine conversion has been described previously[6,33]. Xanthine oxidase is significantly elevated in a variety of cardiovascular conditions such as coronary artery disease and heart failure[27]. There is a large variability in human xanthine oxidase expression, which can be up to three-fold and on average 20% higher in men than in women[29]. Although basal expression of xanthine oxidase is low in humans, hypoxias, IL-1, IL-6, TNF-α, lipo-polysaccharides as well as steroid treatment have been shown to up-regulate transcription[6].

The second cluster correlated with serum urate is composed of several essential amino acids. Interestingly, histidine, tryptophan, and tyrosine are amino acids which are especially sensitive to hydroxyl radical exposure[13]. The amino acid cluster is connected via methionine and histidine to serum urate in our network. Methionine enriched diet is known to decrease urate levels in chickens and ducks, whereas only a few small studies have analysed the effect in humans[82,86]. Furthermore, methionine can be demethylated to homocysteine. Elevated homocysteine levels, as well as elevated urate levels, have been shown to be a risk factor for atherosclerosis, coronary heart disease, and chronic kidney disease[24,37,50]. Significant associations between serum urate and homocysteine have been shown in plasma and serum[51,52].

The third cluster correlated with serum urate is composed of steroids and several unknowns. The different concentrations of serum urate in both sexes and the higher incidence of gout in men compared to women, suggest a hormonal influence on the pathogenesis of gout[28]. Excretion of urinary dehydroepiandrosterone and androsterone has been reported to be significantly lower in subjects with gout[68]. A small study investigating the hormonal urinary excretion reported that patients previously treated with allopurinol showed slightly higher values of androsterone and dehydroepiandrosterone, and slightly lower values of 11-hydroxyandrosterone in comparison to normal subjects, suggesting different hormonal patterns between individuals with and without gout[28].

Furthermore urate is connected to N-[3-(2-oxopyrrolidin-1-yl)propyl]acetamide (acisoga), a metabolite of spermidine. Our metabolite network does not provide the first link between urate and spermidine. Spermidine and spermine were previously found to bind the organic anion transporter OAT1 in mice, and to be putative novel endogenous substrates of OAT1[1] which is also known to be a urate transporter[38].

Within our analysed dataset, not only serum urate shows differences between men and

women, but 25 out of 36 metabolites, which underlines the important sex effect on metabolite profiles reported before[56].

Allopurinol intake inhibits the enzyme xanthine oxidase which is responsible for the successive oxidation of hypoxanthine to xanthine and xanthine to urate. While hypoxanthine is not significantly elevated by allopurinol intake in our data, xanthine displays differential concentration levels between medicated and medication-free individuals. As expected, urate levels of medicated individuals lie in a normal range compared to untreated individuals (see Figure 17). Furthermore, our data showed an effect of allopurinol intake on caffeine levels. Several epidemiological studies found that coffee consumption is inversely associated with serum urate levels[11] and an influence of allopurinol medication on caffeine has been described[7,25]. The strongest influence of allopurinol intake was observed on the unknown metabolite X-11422, which we could identify to be xanthine in spiking experiments. While the association between allopurinol intake and xanthine was expected, we additionally observed influence on phenylalanine, 3-(4-hydroxyphenyl)lactate, lactate, and 2-hydroxybutyrate.

In the present study, data-driven GGMs on metabolomics profiles were used to reconstruct pathways of biochemically related metabolites in a hypothesis-free approach. Three main clusters were grouped around urate, including purines, amino acids, and steroids and strong sex-specific differences were observed for 25 out of 36 metabolites. Furthermore, we observed an effect of allopurinol intake not only on purine metabolism but on metabolites in each of the three clusters.

7 Conclusion and outlook

The present thesis includes two large GWAS on serum urate levels in order to detect genes that are involved in the regulation of serum urate levels as well as a metabolite network approach to describe the metabolic vicinity of serum urate. Both ideas are hypothesis-free, thus all results are data driven only. Both the GWAS approach and the metabolite network approach provide new insight into additional pathways that are involved in the regulation of serum urate levels. Those may point towards novel potential targets for pharmacological intervention for the treatment or prevention of hyperuricemia and related diseases as gout, cardiovascular disease, and type 2 diabetes.

In the field of GWAS, the detection of genes gets the more successful, the more the sample size and therefore power can be increased. Before the meta-analysis performed within ENGAGE, only the three genes *SLC2A9*, *ABCG2*, and *SLC17A3* were known to be associated with serum urate. By combining the data of many European studies within the ENGAGE and GUGC consortia, we could identify a total of 28 genes playing a role in the regulation of serum urate levels. Table 13 and Figure 18 illustrate how the increasing sample size increases the number of findings. Table 13 compares sample sizes, *p*-values for *SLC2A9*, and number of detected loci within the published serum urate GWAS conducted in individuals of European ancestry. Figure 18 shows Manhattan plots of serum urate GWAS in KORA F3, KORA F3 and KORA F4 combined, the combination of all ENGAGE studies, as well as the combination of all GUGC studies.

Table 13: **Comparison of sample sizes and findings between serum urate GWAS** in samples of European ancestry indicating how the number of findings increases with increasing sample size.

Study	**Publication**	**Sample size**	***p*-value *SLC2A9***	**Number of genome-wide significant loci**
KORA F3	Döring *et al.* (2008)[18]	1,644	1.6×10^{-12}	1
ENGAGE	Kolz *et al.* (2009)[41]	28,141	5.2×10^{-201}	11
CHARGE	Yang *et al.* (2010)[85]	28,283	1.5×10^{-242}	8
GUGC	Köttgen *et al.* (2012)[45]	110,347	$< 1\times10^{-700}$	26

Figure 18 as well as Table 13 prove that “sample size matters”. Although the combination of multiple studies requires an extended quality control and is accompanied by increasing heterogeneity, the increased sample size increases power and a the-bigger-the-better practice in the field of GWAS is justified.

There is a lot of discussion going on about the “missing heritability”, as most GWAS only explain a small proportion of the estimated heritability of the respective phenotype. In the present case we only explain about 7.0% of the variance of serum urate levels, whereas the heritability of serum urate levels is estimated to be about 40–70%[58,78,84]. Further increasing the sample size could be one strategy to detect additional variants playing a role. Lately, several consortia started to combine data of different ancestries in trans-ethnic GWAS to increase sample sizes. Even though one extremely increases the heterogeneity in this approach, the increase in power may still lead to additional findings.

On the other hand, the GWAS approach focuses on the analysis of common variants as most rare variants are not tagged by GWAS chips. However, common variants are only

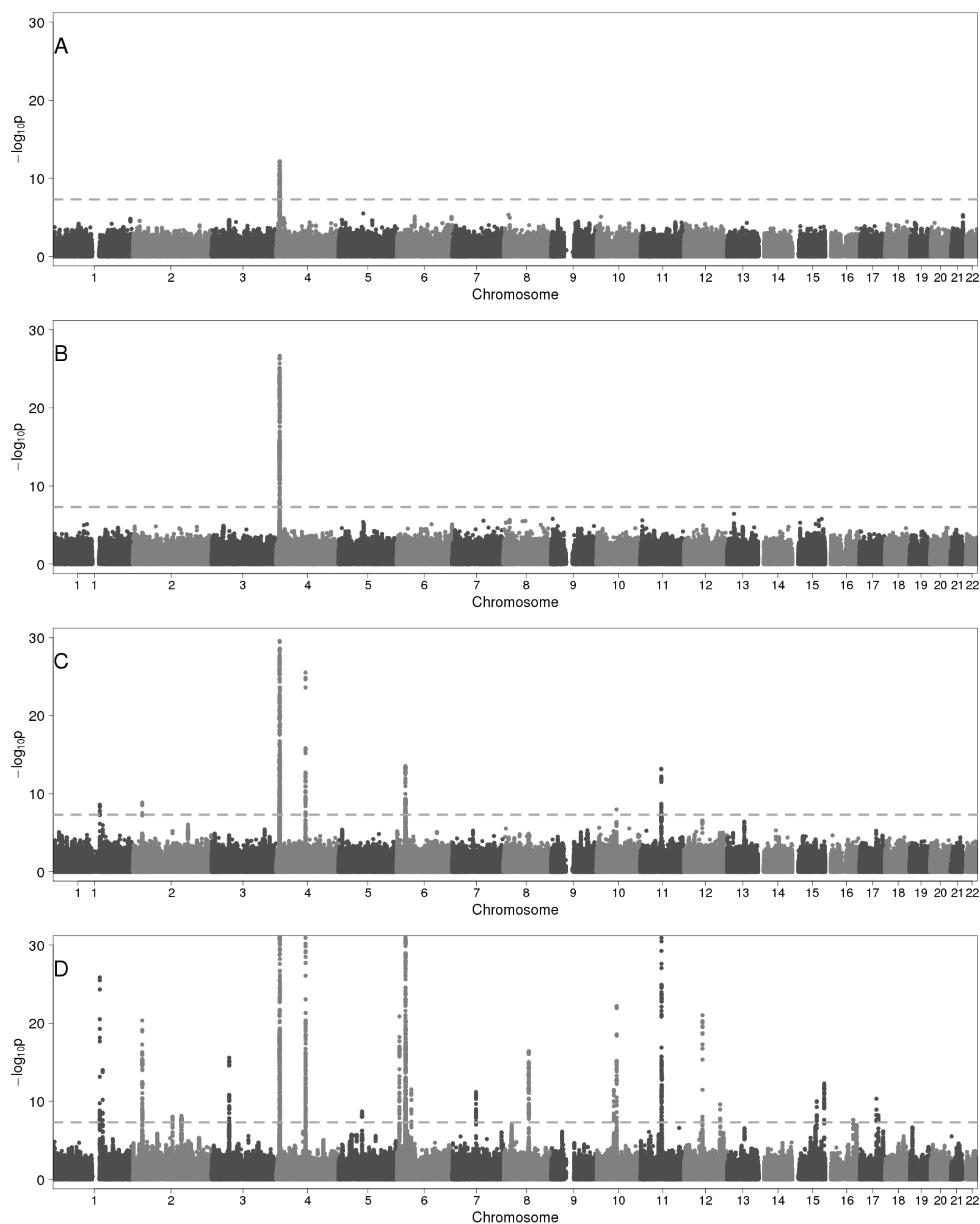


Figure 18: **Manhattan plots with increasing power** showing serum urate GWAS in A) KORA F3, B) the meta-analysis of KORA F3 and KORA F4, C) the meta-analysis of all ENGAGE studies as presented in chapter 3, and D) the meta-analysis of all GUGC discovery studies as presented in chapter 4. For A) a MAF filter of 5% was applied. C) and D) are truncated at 1×10^{-30}.

"the peak of an iceberg" in the diversity of the human genome and rare variants might have a major impact and might rather be functional. In the case of serum urate, the whole-genome sequencing analysis conducted by Sulem *et al.*[71] demonstrates how the

GWAS approach fails to detect regions where rare variants are of influence (see Figure 10).

Imputation with the lately available 1000g reference panel (http://www.1000genomes.org) will possibly be able to cover a larger number of rare variants. Within GUGC a meta-analysis of 1000g imputed GWAS is planned, as well as a meta-analysis of Exomechip based GWAS.

Future investigations planned within GUGC include an additional characterisation of the genetic loci by Mendelian randomisation projects and the translation of the findings in terms of direct impact on gout. Furthermore, gene-gene or gene-environment interaction analysis, for example with urate lowering medication, might provide additional insights.

In the quickly developing "omics" field systems epidemiology approaches provide possibilities to incorporate information at multiple levels: genomics, epigenomics, transcriptomics, proteomics, metabolomics, microbiomics. This thesis focused on genomics and metabolomics of serum urate. The KORA sudies provide a broad data basis for future projects, especially concerning epigenomics and transcriptomics. The incorporation of multiple levels will further improve our understanding of the biological mechanisms underlying the regulation of serum urate levels in humans.

References

[1] S.-Y. Ahn, N. Jamshidi, M. L. Mo, W. Wu, S. A. Eraly, A. Dnyanmote, K. T. Bush, T. F. Gallegos, D. H. Sweet, B. . Palsson, and S. K. Nigam. Linkage of organic anion transporter-1 to metabolic pathways through integrated "omics"-driven network and functional analysis. *J Biol Chem*, 286(36):31522–31531, Sep 2011.

[2] N. Anzai, H. Miyazaki, R. Noshiro, S. Khamdang, A. Chairoungdua, H.-J. Shin, A. Enomoto, S. Sakamoto, T. Hirata, K. Tomita, Y. Kanai, and H. Endou. The multivalent PDZ domain-containing protein PDZK1 regulates transport activity of renal urate-anion exchanger URAT1 via its C terminus. *J Biol Chem*, 279(44):45942–45950, Oct 2004.

[3] S. A. Bacanu, B. Devlin, and K. Roeder. The power of genomic control. *Am J Hum Genet*, 66(6):1933–1944, Jun 2000.

[4] Y. Benjamini and Y. Hochberg. Controlling the False Discovery Rate: a Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society. Series B*, 57(1):289–300, 1995.

[5] Y. Benjamini and D. Yekutieli. The control of the false discovery rate in multiple testing under dependency. *The Annals of Statistics*, 29(4):1165–1188, 2001.

[6] C. E. Berry and J. M. Hare. Xanthine oxidoreductase and cardiovascular disease: molecular mechanisms and pathophysiological implications. *J Physiol*, 555(Pt 3):589–606, Mar 2004.

[7] D. J. Birkett, J. O. Miners, L. Valente, K. J. Lillywhite, and R. O. Day. 1-Methylxanthine derived from caffeine as a pharmacodynamic probe of oxypurinol effect. *Br J Clin Pharmacol*, 43(2):197–200, Feb 1997.

[8] M. J. Caulfield, P. B. Munroe, D. O'Neill, K. Witkowska, F. J. Charchar, M. Doblado, S. Evans, S. Eyheramendy, A. Onipinla, P. Howard, S. Shaw-Hawkins, R. J. Dobson, C. Wallace, S. J. Newhouse, M. Brown, J. M. Connell, A. Dominiczak, M. Farrall, G. M. Lathrop, N. J. Samani, M. Kumari, M. Marmot, E. Brunner, J. Chambers, P. Elliott, J. Kooner, M. Laan, E. Org, G. Veldre, M. Viigimaa, F. P. Cappuccio, C. Ji, R. Iacone, P. Strazzullo, K. H. Moley, and C. Cheeseman. SLC2A9 is a high-capacity urate transporter in humans. *PLoS Med*, 5(10):e197, Oct 2008.

[9] B. A. Charles, D. Shriner, A. Doumatey, G. Chen, J. Zhou, H. Huang, A. Herbert, N. P. Gerry, M. F. Christman, A. Adeyemo, and C. N. Rotimi. A genome-wide association study of serum uric acid in african americans. *BMC Med Genomics*, 4:17, 2011.

[10] X. Chen, G. Wu, and M. A. Schwarzschild. Urate in Parkinson's disease: more than a biomarker? *Curr Neurol Neurosci Rep*, 12(4):367–375, Aug 2012.

[11] H. K. Choi and G. Curhan. Coffee, tea, and caffeine consumption and serum uric acid level: the third national health and nutrition examination survey. *Arthritis Rheum*, 57(5):816–821, Jun 2007.

[12] W. G. Cochran. The combination of estimates from different experiments. *BIOMETRICS*, 10:101–129, 1954.

[13] K. J. Davies, M. E. Delsignore, and S. W. Lin. Protein damage and degradation by oxygen radicals. II. Modification of amino acids. *J Biol Chem*, 262(20):9902–9907, Jul 1987.

[14] P. I. W. de Bakker, M. A. R. Ferreira, X. Jia, B. M. Neale, S. Raychaudhuri, and B. F. Voight. Practical aspects of imputation-driven meta-analysis of genome-wide association studies. *Hum Mol Genet*, 17(R2):R122–R128, Oct 2008.

[15] A. Dehghan, A. Köttgen, Q. Yang, S.-J. Hwang, W. L. Kao, F. Rivadeneira, E. Boerwinkle, D. Levy, A. Hofman, B. C. Astor, E. J. Benjamin, C. M. van Duijn, J. C. Witteman, J. Coresh, and C. S. Fox. Association of three genetic loci with uric acid concentration and risk of gout: a genome-wide association study. *Lancet*, 372(9654):1953–1961, Dec 2008.

[16] A. Dehghan, M. van Hoek, E. J. G. Sijbrands, A. Hofman, and J. C. M. Witteman. High serum uric acid as a novel risk factor for type 2 diabetes. *Diabetes Care*, 31(2):361–362, Feb 2008.

[17] D. Dinour, N. K. Gray, S. Campbell, X. Shu, L. Sawyer, W. Richardson, G. Rechavi, N. Amariglio, L. Ganon, B.-A. Sela, H. Bahat, M. Goldman, J. Weissgarten, M. R. Millar, A. F. Wright, and E. J. Holtzman. Homozygous slc2a9 mutations cause severe renal hypouricemia. *J Am Soc Nephrol*, 21(1):64–72, Jan 2010.

[18] A. Döring, C. Gieger, D. Mehta, H. Gohlke, H. Prokisch, S. Coassin, G. Fischer, K. Henke, N. Klopp, F. Kronenberg, B. Paulweber, A. Pfeufer, D. Rosskopf, H. Völzke, T. Illig, T. Meitinger, H.-E. Wichmann, and C. Meisinger. SLC2A9 influences uric acid concentrations with pronounced sex-specific effects. *Nat Genet*, 40(4):430–436, Apr 2008.

[19] F. Dudbridge and A. Gusnanto. Estimation of significance thresholds for genomewide association scans. *Genet Epidemiol*, 32(3):227–234, Apr 2008.

[20] A. Enomoto, H. Kimura, A. Chairoungdua, Y. Shigeta, P. Jutabha, S. H. Cha, M. Hosoyamada, M. Takeda, T. Sekine, T. Igarashi, H. Matsuo, Y. Kikuchi, T. Oda, K. Ichida, T. Hosoya, K. Shimokata, T. Niwa, Y. Kanai, and H. Endou. Molecular identification of a renal urate anion exchanger that regulates blood urate levels. *Nature*, 417(6887):447–452, May 2002.

[21] A. M. Evans, C. D. DeHaven, T. Barrett, M. Mitchell, and E. Milgram. Integrated, nontargeted ultrahigh performance liquid chromatography/electrospray ionization tandem mass spectrometry platform for the identification and relative quantification of the small-molecule complement of biological systems. *Anal Chem*, 81(16):6656–6667, Aug 2009.

[22] J. Fang and M. H. Alderman. Serum uric acid and cardiovascular mortality the nhanes i epidemiologic follow-up study, 1971-1992. national health and nutrition examination survey. *JAMA*, 283(18):2404–2410, May 2000.

[23] D. I. Feig, D.-H. Kang, and R. J. Johnson. Uric acid and cardiovascular risk. *N Engl J Med*, 359(17):1811–1821, Oct 2008.

[24] M. E. Francis, P. W. Eggers, T. H. Hostetter, and J. P. Briggs. Association between serum homocysteine and markers of impaired kidney function in adults in the United States. *Kidney Int*, 66(1):303–312, Jul 2004.

[25] P. Fuchs, W. E. Haefeli, H. R. Ledermann, and M. Wenk. Xanthine oxidase inhibition by allopurinol affects the reliability of urinary caffeine metabolic ratios as markers for N-acetyltransferase 2 and CYP1A2 activities. *Eur J Clin Pharmacol*, 54(11):869–876, Jan 1999.

[26] C. Fuchsberger, D. Taliun, P. P. Pramstaller, C. Pattaro, and C. K. D. G. consortium. GWAtoolbox: an R package for fast quality control and handling of genome-wide association studies meta-analysis data. *Bioinformatics*, 28(3):444–445, Feb 2012.

[27] J. George and A. D. Struthers. Role of urate, xanthine oxidase and the effects of allopurinol in vascular oxidative stress. *Vasc Health Risk Manag*, 5(1):265–272, 2009.

[28] L. Gregolini, S. Ferrari, R. Marcolongo, M. F. Aleo, E. Marinello, and E. Bianchi. Hormonal aspects of human gout–excretion of adrenal hormone derivatives in gouty patients. *Clin Chim Acta*, 130(3):269–277, Jun 1983.

[29] R. Guerciolini, C. Szumlanski, and R. M. Weinshilboum. Human liver xanthine oxidase: nature and extent of individual variation. *Clin Pharmacol Ther*, 50(6):663–672, Dec 1991.

[30] Y. Hagos, D. Stein, B. Ugele, G. Burckhardt, and A. Bahn. Human renal organic anion transporter 4 operates as an asymmetric urate transporter. *J Am Soc Nephrol*, 18(2):430–439, Feb 2007.

[31] M. A. Hediger, R. J. Johnson, H. Miyazaki, and H. Endou. Molecular physiology of urate transport. *Physiology (Bethesda)*, 20:125–133, Apr 2005.

[32] J. P. T. Higgins, S. G. Thompson, J. J. Deeks, and D. G. Altman. Measuring inconsistency in meta-analyses. *BMJ*, 327(7414):557–560, Sep 2003.

[33] R. Hille and V. Massey. Studies on the oxidative half-reaction of xanthine oxidase. *J Biol Chem*, 256(17):9090–9095, Sep 1981.

[34] L. A. Hindorff, J. MacArthur, J. Morales, H. A. Junkins, P. N. Hall, A. K. Klemm, and T. A. Manolio. A Catalog of Published Genome-Wide Association Studies. *Available at: www.genome.gov/gwastudies*, Accessed 28 Dec. 2012.

[35] R. Holle, M. Happich, H. Löwel, H. E. Wichmann, and M. O. N. I. C. A. O. R. A. S. Group. KORA–a research platform for population based health research. *Gesundheitswesen*, 67 Suppl 1:S19–S25, Aug 2005.

[36] B. N. Howie, P. Donnelly, and J. Marchini. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLoS Genet*, 5(6):e1000529, Jun 2009.

[37] L. L. Humphrey, R. Fu, K. Rogers, M. Freeman, and M. Helfand. Homocysteine level and coronary heart disease incidence: a systematic review and meta-analysis. *Mayo Clin Proc*, 83(11):1203–1212, Nov 2008.

[38] K. Ichida, M. Hosoyamada, H. Kimura, M. Takeda, Y. Utsunomiya, T. Hosoya, and H. Endou. Urate transport via human pah transporter hoat1 and its gene structure. *Kidney Int*, 63(1):143–155, Jan 2003.

[39] M. Iharada, T. Miyaji, T. Fujimoto, M. Hiasa, N. Anzai, H. Omote, and Y. Moriyama. Type 1 sodium-dependent phosphate transporter (SLC17A1 Protein) is a Cl(-)-dependent urate exporter. *J Biol Chem*, 285(34):26107–26113, Aug 2010.

[40] Y. Kamatani, K. Matsuda, Y. Okada, M. Kubo, N. Hosono, Y. Daigo, Y. Nakamura, and N. Kamatani. Genome-wide association study of hematological and biochemical traits in a Japanese population. *Nat Genet*, 42(3):210–215, Mar 2010.

[41] M. Kolz, T. Johnson, S. Sanna, A. Teumer, V. Vitart, M. Perola, M. Mangino, E. Albrecht, C. Wallace, M. Farrall, A. Johansson, D. R. Nyholt, Y. Aulchenko, J. S. Beckmann, S. Bergmann, M. Bochud, M. Brown, H. Campbell, E. U. R. O. S. P. A. N. Consortium, J. Connell, A. Dominiczak, G. Homuth, C. Lamina, M. I. McCarthy, E. N. G. A. G. E. Consortium, T. Meitinger, V. Mooser, P. Munroe, M. Nauck, J. Peden, H. Prokisch, P. Salo, V. Salomaa, N. J. Samani, D. Schlessinger, M. Uda, U. Völker, G. Waeber, D. Waterworth, R. Wang-Sattler, A. F. Wright, J. Adamski, J. B. Whitfield, U. Gyllensten, J. F. Wilson, I. Rudan, P. Pramstaller, H. Watkins, P. R. O. C. A. R. D. I. S. Consortium, A. Doering, H.-E. Wichmann, K. O. R. A. Study, T. D. Spector, L. Peltonen, H. Völzke, R. Nagaraja, P. Vollenweider, M. Caulfield, W. T. C. C. C., T. Illig, and C. Gieger. Meta-analysis of 28,141 individuals identifies common variants within five new loci that influence uric acid concentrations. *PLoS Genet*, 5(6):e1000504, Jun 2009.

[42] F. Komoda, T. Sekine, J. Inatomi, A. Enomoto, H. Endou, T. Ota, T. Matsuyama, T. Ogata, M. Ikeda, M. Awazu, K. Muroya, I. Kamimaki, and T. Igarashi. The W258X mutation in SLC22A12 is the predominant cause of Japanese renal hypouricemia. *Pediatr Nephrol*, 19(7):728–733, Jul 2004.

[43] J. Krumsiek, K. Suhre, A. M. Evans, M. W. Mitchell, R. P. Mohney, M. V. Milburn, B. Wägele, W. Römisch-Margl, T. Illig, J. Adamski, C. Gieger, F. J. Theis, and G. Kastenmüller. Mining the unknown: a systems approach to metabolite identification combining genetic and metabolic information. *PLoS Genet*, 8(10):e1003005, Oct 2012.

[44] J. Krumsiek, K. Suhre, T. Illig, J. Adamski, and F. J. Theis. Gaussian graphical modeling reconstructs pathway reactions from high-throughput metabolomics data. *BMC Syst Biol*, 5:21, 2011.

[45] A. Köttgen, E. Albrecht, A. Teumer, V. Vitart, J. Krumsiek, C. Hundertmark, G. Pistis, D. Ruggiero, C. M. O'Seaghdha, T. Haller, Q. Yang, T. Tanaka, A. D. Johnson, Z. Kutalik, A. V. Smith, J. Shi, M. Struchalin, R. P. S. Middelberg, M. J. Brown, A. L. Gaffo, N. Pirastu, G. Li, C. Hayward, T. Zemunik, J. Huffman, L. Yengo, J. H. Zhao, A. Demirkan, M. F. Feitosa, X. Liu, G. Malerba, L. M. Lopez, P. van der Harst, X. Li, M. E. Kleber, A. A. Hicks, I. M. Nolte, A. Johansson, F. Murgia, S. H. Wild, S. J. L. Bakker, J. F. Peden, A. Dehghan, M. Steri, A. Tenesa, V. Lagou, P. Salo, M. Mangino, L. M. Rose, T. Lehtimäki, O. M. Woodward, Y. Okada, A. Tin, C. Müller, C. Oldmeadow, M. Putku, D. Czamara,

P. Kraft, L. Frogheri, G. A. Thun, A. Grotevendt, G. K. Gislason, T. B. Harris, L. J. Launer, P. McArdle, A. R. Shuldiner, E. Boerwinkle, J. Coresh, H. Schmidt, M. Schallert, N. G. Martin, G. W. Montgomery, M. Kubo, Y. Nakamura, T. Tanaka, P. B. Munroe, N. J. Samani, D. R. Jacobs, Jr, K. Liu, P. D'Adamo, S. Ulivi, J. I. Rotter, B. M. Psaty, P. Vollenweider, G. Waeber, S. Campbell, O. Devuyst, P. Navarro, I. Kolcic, N. Hastie, B. Balkau, P. Froguel, T. Esko, A. Salumets, K. T. Khaw, C. Langenberg, N. J. Wareham, A. Isaacs, A. Kraja, Q. Zhang, P. S. Wild, R. J. Scott, E. G. Holliday, E. Org, M. Viigimaa, S. Bandinelli, J. E. Metter, A. Lupo, E. Trabetti, R. Sorice, A. Döring, E. Lattka, K. Strauch, F. Theis, M. Waldenberger, H.-E. Wichmann, G. Davies, A. J. Gow, M. Bruinenberg, L. C. S. , R. P. Stolk, J. S. Kooner, W. Zhang, B. R. Winkelmann, B. O. Boehm, S. Lucae, B. W. Penninx, J. H. Smit, G. Curhan, P. Mudgal, R. M. Plenge, L. Portas, I. Persico, M. Kirin, J. F. Wilson, I. M. Leach, W. H. van Gilst, A. Goel, H. Ongen, A. Hofman, F. Rivadeneira, A. G. Uitterlinden, M. Imboden, A. von Eckardstein, F. Cucca, R. Nagaraja, M. G. Piras, M. Nauck, C. Schurmann, K. Budde, F. Ernst, S. M. Farrington, E. Theodoratou, I. Prokopenko, M. Stumvoll, A. Jula, M. Perola, V. Salomaa, S.-Y. Shin, T. D. Spector, C. Sala, P. M. Ridker, M. Kähönen, J. Viikari, C. Hengstenberg, C. P. Nelson, C. A. R. D. I. R. A. M. C. , D. I. A. G. R. A. M. C. , I. C. B. P. C. , M. A. G. I. C. C. , J. F. Meschia, M. A. Nalls, P. Sharma, A. B. Singleton, N. Kamatani, T. Zeller, M. Burnier, J. Attia, M. Laan, N. Klopp, H. L. Hillege, S. Kloiber, H. Choi, M. Pirastu, S. Tore, N. M. Probst-Hensch, H. Völzke, V. Gudnason, A. Parsa, R. Schmidt, J. B. Whitfield, M. Fornage, P. Gasparini, D. S. Siscovick, O. Polašek, H. Campbell, I. Rudan, N. Bouatia-Naji, A. Metspalu, R. J. F. Loos, C. M. van Duijn, I. B. Borecki, L. Ferrucci, G. Gambaro, I. J. Deary, B. H. R. Wolffenbuttel, J. C. Chambers, W. März, P. P. Pramstaller, H. Snieder, U. Gyllensten, A. F. Wright, G. Navis, H. Watkins, J. C. M. Witteman, S. Sanna, S. Schipf, M. G. Dunlop, A. Tönjes, S. Ripatti, N. Soranzo, D. Toniolo, D. I. Chasman, O. Raitakari, W. H. L. Kao, M. Ciullo, C. S. Fox, M. Caulfield, M. Bochud, and C. Gieger. Genome-wide association analyses identify 18 new loci associated with serum urate concentrations. *Nat Genet*, Dec 2012.

[46] A. Köttgen, C. Pattaro, C. A. Böger, C. Fuchsberger, M. Olden, N. L. Glazer, A. Parsa, X. Gao, Q. Yang, A. V. Smith, J. R. O'Connell, M. Li, H. Schmidt, T. Tanaka, A. Isaacs, S. Ketkar, S.-J. Hwang, A. D. Johnson, A. Dehghan, A. Teumer, G. Paré, E. J. Atkinson, T. Zeller, K. Lohman, M. C. Cornelis, N. M. Probst-Hensch, F. Kronenberg, A. Tönjes, C. Hayward, T. Aspelund, G. Eiriksdottir, L. J. Launer, T. B. Harris, E. Rampersaud, B. D. Mitchell, D. E. Arking, E. Boerwinkle, M. Struchalin, M. Cavalieri, A. Singleton, F. Giallauria, J. Metter, I. H. de Boer, T. Haritunians, T. Lumley, D. Siscovick, B. M. Psaty, M. C. Zillikens, B. A. Oostra, M. Feitosa, M. Province, M. de Andrade, S. T. Turner, A. Schillert, A. Ziegler, P. S. Wild, R. B. Schnabel, S. Wilde, T. F. Munzel, T. S. Leak, T. Illig, N. Klopp, C. Meisinger, H.-E. Wichmann, W. Koenig, L. Zgaga, T. Zemunik, I. Kolcic, C. Minelli, F. B. Hu, A. Johansson, W. Igl, G. Zaboli, S. H. Wild, A. F. Wright, H. Campbell, D. Ellinghaus, S. Schreiber, Y. S. Aulchenko, J. F. Felix, F. Rivadeneira, A. G. Uitterlinden, A. Hofman, M. Imboden, D. Nitsch, A. Brandstätter, B. Kollerits, L. Kedenko, R. Mägi, M. Stumvoll, P. Kovacs, M. Boban, S. Campbell, K. Endlich, H. Völzke, H. K. Kroemer, M. Nauck, U. Völker, O. Polasek, V. Vitart, S. Badola, A. N. Parker, P. M. Ridker, S. L. R. Kardia, S. Blankenberg, Y. Liu, G. C. Curhan, A. Franke, T. Rochat, B. Paulweber, I. Prokopenko, W. Wang, V. Gud-

nason, A. R. Shuldiner, J. Coresh, R. Schmidt, L. Ferrucci, M. G. Shlipak, C. M. van Duijn, I. Borecki, B. K. Krämer, I. Rudan, U. Gyllensten, J. F. Wilson, J. C. Witteman, P. P. Pramstaller, R. Rettig, N. Hastie, D. I. Chasman, W. H. Kao, I. M. Heid, and C. S. Fox. New loci associated with kidney function and chronic kidney disease. *Nat Genet*, 42(5):376–384, May 2010.

[47] S. Li, S. Sanna, A. Maschio, F. Busonero, G. Usala, A. Mulas, S. Lai, M. Dei, M. Orrù, G. Albai, S. Bandinelli, D. Schlessinger, E. Lakatta, A. Scuteri, S. S. Najjar, J. Guralnik, S. Naitza, L. Crisponi, A. Cao, G. Abecasis, L. Ferrucci, M. Uda, W.-M. Chen, and R. Nagaraja. The GLUT9 gene is associated with serum uric acid levels in Sardinia and Chianti cohorts. *PLoS Genet*, 3(11):e194, Nov 2007.

[48] Y. Li, C. Willer, S. Sanna, and G. Abecasis. Genotype imputation. *Annu Rev Genomics Hum Genet*, 10:387–406, 2009.

[49] Y. Li, C. J. Willer, J. Ding, P. Scheet, and G. R. Abecasis. MaCH: using sequence and genotype data to estimate haplotypes and unobserved genotypes. *Genet Epidemiol*, 34(8):816–834, Dec 2010.

[50] M. Lubomirova, A. Tzoncheva, J. Petrova, and B. Kiperova. Homocystein and carotid atherosclerosis in chronic renal failure. *Hippokratia*, 11(4):205–209, Oct 2007.

[51] S. Lussier-Cacan, M. Xhignesse, A. Piolot, J. Selhub, J. Davignon, and J. Genest. Plasma total homocysteine in healthy subjects: sex-specific relation with biological traits. *Am J Clin Nutr*, 64(4):587–593, Oct 1996.

[52] M. R. Malinow, J. Levenson, P. Giral, F. J. Nieto, M. Razavian, P. Segond, and A. Simon. Role of blood pressure, uric acid, and hemorheological parameters on plasma homocyst(e)ine concentration. *Atherosclerosis*, 114(2):175–183, Apr 1995.

[53] J. Marchini and B. Howie. Genotype imputation for genome-wide association studies. *Nat Rev Genet*, 11(7):499–511, Jul 2010.

[54] J. Marchini, B. Howie, S. Myers, G. McVean, and P. Donnelly. A new multipoint method for genome-wide association studies by imputation of genotypes. *Nat Genet*, 39(7):906–913, Jul 2007.

[55] H. Matsuo, T. Takada, K. Ichida, T. Nakamura, A. Nakayama, Y. Ikebuchi, K. Ito, Y. Kusanagi, T. Chiba, S. Tadokoro, Y. Takada, Y. Oikawa, H. Inoue, K. Suzuki, R. Okada, J. Nishiyama, H. Domoto, S. Watanabe, M. Fujita, Y. Morimoto, M. Naito, K. Nishio, A. Hishida, K. Wakai, Y. Asai, K. Niwa, K. Kamakura, S. Nonoyama, Y. Sakurai, T. Hosoya, Y. Kanai, H. Suzuki, N. Hamajima, and N. Shinomiya. Common defects of ABCG2, a high-capacity urate exporter, cause gout: a function-based genetic analysis in a Japanese population. *Sci Transl Med*, 1(5):5ra11, Nov 2009.

[56] K. Mittelstrass, J. S. Ried, Z. Yu, J. Krumsiek, C. Gieger, C. Prehn, W. Roemisch-Margl, A. Polonikov, A. Peters, F. J. Theis, T. Meitinger, F. Kronenberg, S. Weidinger, H. E. Wichmann, K. Suhre, R. Wang-Sattler, J. Adamski, and T. Illig. Discovery of sexual dimorphisms in metabolic and genetic biomarkers. *PLoS Genet*, 7(8):e1002215, Aug 2011.

[57] H. Miyazaki, N. Anzai, S. Ekaratanawong, T. Sakata, H. J. Shin, P. Jutabha, T. Hirata, X. He, H. Nonoguchi, K. Tomita, Y. Kanai, and H. Endou. Modulation of renal apical organic anion transporter 4 function by two PDZ domain-containing proteins. *J Am Soc Nephrol*, 16(12):3498–3506, Dec 2005.

[58] S. D. Nath, V. S. Voruganti, N. H. Arar, F. Thameem, J. C. Lopez-Alvarenga, R. Bauer, J. Blangero, J. W. MacCluer, A. G. Comuzzie, and H. E. Abboud. Genome scan for determinants of serum uric acid variability. *J Am Soc Nephrol*, 18(12):3156–3163, Dec 2007.

[59] M. Oda, Y. Satta, O. Takenaka, and N. Takahata. Loss of urate oxidase activity in hominoids and its evolutionary implications. *Mol Biol Evol*, 19(5):640–653, May 2002.

[60] Y. Okada, X. Sim, M. J. Go, J.-Y. Wu, D. Gu, F. Takeuchi, A. Takahashi, S. Maeda, T. Tsunoda, P. Chen, S.-C. Lim, T.-Y. Wong, J. Liu, T. L. Young, T. Aung, M. Seielstad, Y.-Y. Teo, Y. J. Kim, J.-Y. Lee, B.-G. Han, D. Kang, C.-H. Chen, F.-J. Tsai, L.-C. Chang, S.-J. C. Fann, H. Mei, D. C. Rao, J. E. Hixson, S. Chen, T. Katsuya, M. Isono, T. Ogihara, J. C. Chambers, W. Zhang, J. S. Kooner, T. K. Consortium, T. C. Consortium, E. Albrecht, T. G. consortium, K. Yamamoto, M. Kubo, Y. Nakamura, N. Kamatani, N. Kato, J. He, Y.-T. Chen, Y. S. Cho, E.-S. Tai, and T. Tanaka. Meta-analysis identifies multiple loci associated with kidney function-related traits in east Asian populations. *Nat Genet*, 44(8):904–909, 2012.

[61] I. Pe'er, R. Yelensky, D. Altshuler, and M. J. Daly. Estimation of the multiple testing burden for genomewide association studies of nearly all common variants. *Genet Epidemiol*, 32(4):381–385, May 2008.

[62] K. D. Pfeffer, T. P. Huecksteadt, and J. R. Hoidal. Xanthine dehydrogenase and xanthine oxidase activity and gene expression in renal epithelial cells. Cytokine and steroid regulation. *J Immunol*, 153(4):1789–1797, Aug 1994.

[63] S. Purcell, B. Neale, K. Todd-Brown, L. Thomas, M. A. R. Ferreira, D. Bender, J. Maller, P. Sklar, P. I. W. de Bakker, M. J. Daly, and P. C. Sham. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet*, 81(3):559–575, Sep 2007.

[64] S. Raychaudhuri, R. M. Plenge, E. J. Rossin, A. C. Y. Ng, I. S. Consortium, S. M. Purcell, P. Sklar, E. M. Scolnick, R. J. Xavier, D. Altshuler, and M. J. Daly. Identifying relationships among genomic disease regions: predicting genes at pathogenic SNP associations and rare deletions. *PLoS Genet*, 5(6):e1000534, Jun 2009.

[65] B. J. Roessler, J. M. Nosal, P. R. Smith, S. A. Heidler, T. D. Palella, R. L. Switzer, and M. A. Becker. Human X-linked Phosphoribosylpyrophosphate Synthetase Superactivity is associated with distinct point mutations in the PRPSl gene. *J Biol*, 268:26476–26481, 1993.

[66] A. V. Segrè, D. I. A. G. R. A. M. C. , M. A. G. I. C. i. , L. Groop, V. K. Mootha, M. J. Daly, and D. Altshuler. Common inherited variation in mitochondrial genes is not enriched for associations with type 2 diabetes or related glycemic traits. *PLoS Genet*, 6(8), Aug 2010.

[67] A. So and B. Thorens. Uric acid transport and disease. *J Clin Invest*, 120(6):1791–1799, Jun 2010.

[68] M. Sparagana and G. Phillips. Dehydroepiandrosterone (3 -hydroxy-5-androsten-17-one) metabolism in gout. *Steroids*, 19(4):477–491, Apr 1972.

[69] J. D. Storey and R. Tibshirani. Statistical significance for genomewide studies. *Proc Natl Acad Sci U S A*, 100(16):9440–9445, Aug 2003.

[70] K. Suhre, S.-Y. Shin, A.-K. Petersen, R. P. Mohney, D. Meredith, B. Wägele, E. Altmaier, C. A. R. D. I. R. A. M., P. Deloukas, J. Erdmann, E. Grundberg, C. J. Hammond, M. H. de Angelis, G. Kastenmüller, A. Köttgen, F. Kronenberg, M. Mangino, C. Meisinger, T. Meitinger, H.-W. Mewes, M. V. Milburn, C. Prehn, J. Raffler, J. S. Ried, W. Römisch-Margl, N. J. Samani, K. S. Small, H.-E. Wichmann, G. Zhai, T. Illig, T. D. Spector, J. Adamski, N. Soranzo, and C. Gieger. Human metabolic individuality in biomedical and pharmaceutical research. *Nature*, 477(7362):54–60, Sep 2011.

[71] P. Sulem, D. F. Gudbjartsson, G. B. Walters, H. T. Helgadottir, A. Helgason, S. A. Gudjonsson, C. Zanon, S. Besenbacher, G. Bjornsdottir, O. T. Magnusson, G. Magnusson, E. Hjartarson, J. Saemundsdottir, A. Gylfason, A. Jonasdottir, H. Holm, A. Karason, T. Rafnar, H. Stefansson, O. A. Andreassen, J. H. Pedersen, A. I. Pack, M. C. H. de Visser, L. A. Kiemeney, A. J. Geirsson, G. I. Eyjolfsson, I. Olafsson, A. Kong, G. Masson, H. Jonsson, U. Thorsteinsdottir, I. Jonsdottir, and K. Stefansson. Identification of low-frequency variants associated with gout and serum uric acid levels. *Nat Genet*, 43(11):1127–1130, Nov 2011.

[72] D. Szklarczyk, A. Franceschini, M. Kuhn, M. Simonovic, A. Roth, P. Minguez, T. Doerks, M. Stark, J. Muller, P. Bork, L. J. Jensen, and C. von Mering. The STRING database in 2011: functional interaction networks of proteins, globally integrated and scored. *Nucleic Acids Res*, 39(Database issue):D561–D568, Jan 2011.

[73] A. Tin, O. M. Woodward, W. H. L. Kao, C.-T. Liu, X. Lu, M. A. Nalls, D. Shriner, M. Semmo, E. L. Akylbekova, S. B. Wyatt, S.-J. Hwang, Q. Yang, A. B. Zonderman, A. A. Adeyemo, C. Palmer, Y. Meng, M. Reilly, M. G. Shlipak, D. Siscovick, M. K. Evans, C. N. Rotimi, M. F. Flessner, M. Köttgen, L. A. Cupples, C. S. Fox, A. Köttgen, C. A. R. , and C. H. A. R. G. E. Consortia. Genome-wide association study for serum urate concentrations and gout among African Americans identifies genomic risk loci and a novel URAT1 loss-of-function allele. *Hum Mol Genet*, 20(20):4056–4068, Oct 2011.

[74] G. Toncev, B. Milicic, S. Toncev, and G. Samardzic. Serum uric acid levels in multiple sclerosis patients correlate with activity of disease and blood-brain barrier dysfunction. *Eur J Neurol*, 9(3):221–226, May 2002.

[75] S. van Buuren and K. Groothuis-Oudshoorn. mice: Multivariate Imputation by Chained Equations in R. *Journal of Statistical Software*, 45(3):1–67, 2011.

[76] V. Vitart, I. Rudan, C. Hayward, N. K. Gray, J. Floyd, C. N. A. Palmer, S. A. Knott, I. Kolcic, O. Polasek, J. Graessler, J. F. Wilson, A. Marinaki, P. L. Riches, X. Shu, B. Janicijevic, N. Smolej-Narancic, B. Gorgoni, J. Morgan, S. Campbell, Z. Biloglav, L. Barac-Lauc, M. Pericic, I. M. Klaric, L. Zgaga, T. Skaric-Juric, S. H.

Wild, W. A. Richardson, P. Hohenstein, C. H. Kimber, A. Tenesa, L. A. Donnelly, L. D. Fairbanks, M. Aringer, P. M. McKeigue, S. H. Ralston, A. D. Morris, P. Rudan, N. D. Hastie, H. Campbell, and A. F. Wright. SLC2A9 is a newly identified urate transporter influencing serum urate concentration, urate excretion and gout. *Nat Genet*, 40(4):437–442, Apr 2008.

[77] C. Wallace, S. J. Newhouse, P. Braund, F. Zhang, M. Tobin, M. Falchi, K. Ahmadi, R. J. Dobson, A. C. B. Marçano, C. Hajat, P. Burton, P. Deloukas, M. Brown, J. M. Connell, A. Dominiczak, G. M. Lathrop, J. Webster, M. Farrall, T. Spector, N. J. Samani, M. J. Caulfield, and P. B. Munroe. Genome-wide association study identifies genes for biomarkers of cardiovascular disease: serum urate and dyslipidemia. *Am J Hum Genet*, 82(1):139–149, Jan 2008.

[78] J. B. Whitfield and N. G. Martin. Inheritance and alcohol as factors influencing plasma uric acid levels. *Acta Genet Med Gemellol (Roma)*, 32(2):117–126, 1983.

[79] C. J. Willer, Y. Li, and G. R. Abecasis. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics*, 26(17):2190–2191, Sep 2010.

[80] O. M. Woodward, A. Köttgen, J. Coresh, E. Boerwinkle, W. B. Guggino, and M. Köttgen. Identification of a urate transporter, ABCG2, with a common functional polymorphism causing gout. *Proc Natl Acad Sci U S A*, 106(25):10338–10342, Jun 2009.

[81] O. M. Woodward, A. Köttgen, and M. Köttgen. ABCG transporters and disease. *FEBS J*, 278:3215–3225, 2011.

[82] M. Xie, S. S. Hou, W. Huang, L. Zhao, J. Y. Yu, W. Y. Li, and Y. Y. Wu. Interrelationship between methionine and cystine of early Peking ducklings. *Poult Sci*, 83(10):1703–1708, Oct 2004.

[83] Y. Yamada, K. Yamada, N. Nomura, A. Yamano, R. Kimura, S. Tomida, M. Naiki, and N. Wakamatsu. Molecular analysis of two enzyme genes, HPRT1 and PRPS1, causing X-linked inborn errors of purine metabolism. *Nucleosides Nucleotides Nucleic Acids*, 29(4-6):291–294, Jun 2010.

[84] Q. Yang, C.-Y. Guo, L. A. Cupples, D. Levy, P. W. F. Wilson, and C. S. Fox. Genome-wide search for genes affecting serum uric acid levels: the Framingham Heart Study. *Metabolism*, 54(11):1435–1441, Nov 2005.

[85] Q. Yang, A. Köttgen, A. Dehghan, A. V. Smith, N. L. Glazer, M.-H. Chen, D. I. Chasman, T. Aspelund, G. Eiriksdottir, T. B. Harris, L. Launer, M. Nalls, D. Hernandez, D. E. Arking, E. Boerwinkle, M. L. Grove, M. Li, W. H. L. Kao, M. Chonchol, T. Haritunians, G. Li, T. Lumley, B. M. Psaty, M. Shlipak, S.-J. Hwang, M. G. Larson, C. J. O'Donnell, A. Upadhyay, C. M. van Duijn, A. Hofman, F. Rivadeneira, B. Stricker, A. G. Uitterlinden, G. Paré, A. N. Parker, P. M. Ridker, D. S. Siscovick, V. Gudnason, J. C. Witteman, C. S. Fox, and J. Coresh. Multiple genetic loci influence serum urate levels and their relationship with gout and cardiovascular disease risk factors. *Circ Cardiovasc Genet*, 3(6):523–530, Dec 2010.

[86] X. A. Zhan, J. X. Li, Z. R. Xu, and R. Q. Zhao. Effects of methionine and betaine supplementation on growth performance, carcase composition and metabolism of lipids in male broilers. *Br Poult Sci*, 47(5):576–580, Oct 2006.

[87] Y. Zhu, B. J. Pandya, and H. K. Choi. Prevalence of gout and hyperuricemia in the US general population: the National Health and Nutrition Examination Survey 2007-2008. *Arthritis Rheum*, 63(10):3136–3141, Oct 2011.

Supplementary information

Table S1: **Genotyping information of ENGAGE studies**.

Study	Genotyping platform	Calling algorithm	Quality filters before imputation	SNPs before imputation	NCBI buid	Imputation software	Statistical analysis	Inflation factor λ
BRIGHT	Affymetrix 500K	CHIAMO	CR<95%(MAF>5%), CR<99%(MAF<5%), pHWE<5.7E-7	490,032	35	IMPUTE	SNPTEST	1.007
CoLaus	Affymetrix 500K	BRLMM	CR<70%, pHWE<1E-7	390,631	35	IMPUTE v0.2	custom C++	1.025
CROATIA	Illumina HumanHap300v1	BeadStudio	CR<98%,MAF<1%, pHWE<1E-6	305,068	36	MACH v1.0.15	ProABEL	1.013
Health 2000	Illumina 610K	GenCall	CR<95%,MAF <2%, pHWE<1E-4	598,203	35	MACH v1.0.10	PLINK v1.04	1.024
KORA F3	Affymetrix 500K	BRLMM	CR<90%	490,032	35	MACH v1.0.9	MACH2QTL v1.0.4	1.006
KORA F4	Affymetrix 6.0	Birdseed2	CR<90%	909,622	36	IMPUTE v0.4.2	SNPTEST v1.1.5	1.016
ORCADES	Illumina HumanHap300v2	BeadStudio	CR<98%,MAF<1%, pHWE<1E-6	306,207	36	MACH v1.0.15	ProABEL	1.182
PROCARDIS	Illumina 1M	BeadStudio	CR<95%, pHWE<1E-3	882,598	36	IMPUTE v0.3.2	SNPTEST	1.022
NSPHS	Illumina Infinium HumanHap300v2	BeadStudio	CR<90%,MAF<1%, pHWE<1E-6	315,315	36	MACH v1.0	ProbABEL	1.253
SardiNIA	Affymetrix 500K	BRLMM	CR<90%,MAF<5%, pHWE<1E-6	356,359	35	MACH v1.0.9	Merlin –fastassoc	1.090
SHIP	Affymetrix 6.0	Birdseed2		869,224	36	IMPUTE v0.5.0	SNPTEST v1.1.5	1.031
SSAGA	Illumina HumanHap300	BeadStudio		312,431	35	MACH v1.0.16	PLINK v1.04	1.022
MICROS	Illumina HumanHap300v2	BeadStudio	CR<98%,MAF<1%, pHWE<1E-6	306,207	36	MACH v1.0.15	ProABEL	1.246
TwinsUK	Illumina 317K	Illumina protocol	CR<95%,MAF<1%, pHWE<1E-4	279,801	36	IMPUTE v0.4.2	GenABEL	1.122

Table S2: **Study descriptions of GUGC studies**.

Study name	Study design	Total genotyped sample size	Exclusion criteria for study participation or disease enrichment	Exclusions	Population stratification	Serum Urate measurement and QC	Key study references
Discovery studies							
AGES Reykjavik Study	Prospective, population-based	3,219 of European ancestry	none	none	All individuals from Iceland, with no significant stratification within the population.	Serum urate was measured at the Icelandic Heart Association using the Roche-Hitachi P-Module instrument with Roche uricase method. The coefficient of variation for the urate assay was 4.3%.	Harris et al. (2007)[1]
Amish Studies	Founder "healthy" population based[2].	European ancestry	none	none	NA	Serum uric acid (UA) levels drawn at the screening exam were assayed by Quest Diagnostics (Baltimore, MD) and measured to the nearest 0.1 mg/dl[3].	Mitchell et al. (2008)[2], McArdle et al. (2008)[3]
Atherosclerosis Risk in Communities (ARIC) Study	Prospective, population-based[4]	9,713 of European ancestry	none	Of the 9713 genotyped individuals of European ancestry, we excluded 658 individuals based on discrepancies with previous genotypes, disagreement between reported and genotypic sex, one randomly selected member of a pair of first-degree relatives, or outlier based on measures of average DST or more than 8 SD away on any of the first 10 principal components.	Two principal components were associated with uric acid measurements and included as covariates in the regression.	UA was measured using the uricase method[5] at study visit 1. Repeated measurements of UA in 40 individuals, taken at least one week apart, yielded a reliability coefficient of 0.91, and the coefficient of variation was 7.2%[6].	ARIC (1989)[4], Iribarren et al. (1996)[5], Eckfeldt et al. (1994)[6]
Austrian Stroke Prevention Study (ASPS)	Prospective, population-based	923 genotyped Caucasians living in the city of Graz, Austria	no history or signs of stroke and dementia	Of the 923 genotyped individuals we excluded 67 subjects based on excess autosomal heterozygosity, mismatch between called and phenotypic gender, or by being outliers identified by the IBD analysis. The final population for genetic analysis comprised 856 subjects. Additionally serum urate was not available in 3 cases.	Age and sex were included as covariates in the regression	UA was measured using the uricase method on a Hitachi 917 chemical analyzer at study visit 1. Reproducibility was assessed in 21 subjects and revealed a variation coefficient of 1.7%	Schmidt et al. (1994)[7]
Australian Twin-Family Study (AUSTWIN)	Population-based, twin-pairs and their families	11,520 of European ancestry	none	Samples were excluded for less than 95% of SNPs successfully typed, sex or Mendelian errors, Non-European ancestry	Two principal components were included as covariates in the regression.	Serum uric acid was measured with the uricase method on a Roche 917 or Modular P analyser.	Whitfield et al. (2002)[8], Middelberg et al. (2007)[9], Benyamin et al. (2009)[10]
Baltimore Longitudinal Study of Aging (BLSA)	Prospective, population-based	1,230	none	Of the 1230 genotyped subjects, genetic relatedness was assessed using PCA analysis using the HapMap population as reference. Out of the 857 subjects of European ancestry, 5 subjects were excluded for low genotyping (< 98.5%), 4 subjects were	Use top two principle components included as covariates in the regression model.	UA was measured using the uricase method (Johnson and Johnson, VITROS chemistry system).	Shock et al. (1984)[11]

				removed for sex misspecification. From the 848 subjects with European ancestry passing quality control, 718 subjects with uric acid data was used for this study.			
The BRItish Genetics of HyperTension (BRIGHT) study	Hypertensive cases from the BRIGHT study resource4.	1,743	Control exclusion criteria included BMI>35, diabetes, secondary hypertension or a co-existing illness. Blood pressure was measured using the OMRON-705CP blood pressure monitor.	Of 2000 cases typed, we excluded 257 people with poor genotype quality.	NA	Non-fasting blood samples were obtained from study participants and UA analyses was carried out on frozen serum stored at −20°C. UA concentrations were measured using an uricase method on a Hitachi auto-analyser).	Caulfield et al. (2003)[12]
Coronary Artery Disease Risk in Young Adults (CARDIA)	Prospective, population-based	1,725 of European Ancestry	none	1 sex mismatch; 3 outliers in PCA; 1 discordant genotype	4 principal components included as covariates; none associated with UA	Serum uric acid was measured by the uricase method at multiple visits. The coefficient of variation of uric acid was 2.6%; the split sample technical error was 4.6%.	Friedman et al. (1988)[13], The data collection forms used at each exam as well as the CARDIA protocols are available from the CARDIA website: http://www.cardia.dopm.uab.edu/em_dacf.htm
The Cardiovascular Health Study (CHS)	Prospective, population-based	3,329 CHS Caucasian participants	1908 persons were excluded due coronary heart disease, congestive heart failure, peripheral vascular disease, valvular heart disease, stroke or transient ischemic attack.	The present report is based upon genotyping results from 3,329 CHS Caucasian participants, who were free of clinical cardiovascular disease at baseline, consented to genetic testing, and had DNA available for genotyping. Genotypes were called using the Illumina BeadStudio software. Genotyping was successful in 3,291 persons.	Study sites (clinic sites) were included as covariates in the regression to account for population stratification.	Serum uric acid concentrations were measured at the baseline visit using the Kodak Ektachem 700 Analyzer with reagents (Eastman Kodak, Rochester, NY). The final study sample with available genotype and phenotype data consisted of 3,252 individuals for the analyses of uric acid.	Fried et al. (1991)[14]
Cohorte Lausannoise (CoLaus) Study	Population based	5,636 of European ancestry	none	Individuals with call rate below 90% were excluded. The younger of 1st/2nd degree related pairs were removed from the analysis.	First two ancestry principal components were used as covariates.	Serum uric acid was measured by uricase-PAP (1.0% - 0.5% maximum inter and intra-batch coefficients of variation).	Firmann et al. (2008)[15]
CROATIA-KORCULA	Cross-sectional, population-based	971	none	898 individuals left after QC based on genotyping quality, sex and ancestry check	None of the first 3 principal components strongly associated with uric acid; relatedness of participants taken into account using a mixed linear model with the polygenic effect set as random effect.	UA was measured using the uricase UV photometry method in "Labor Centar" biochemical lab, Bukovcev trg 3, 10000 Zagreb Croatia (www.laborcentar.hr).	Zemunik et al. (2009)[16]
CROATIA-SPLIT	Cross-sectional, population-	535	none	499 individuals left after QC based on genotyping quality, sex and	None of the first 3 principal	UA was measured using the uricase UV photometry method	Rudan et al. (2009)[17]

	based			ancestry check.	components strongly associated with uric acid; relatedness of participants taken into account using a mixed linear model with the polygenic effect set as random effect.	in "Labor Centar" biochemical lab, Bukovcev trg 3, 10000 Zagreb Croatia (www.laborcentar.hr).	
CROATIA-VIS	Cross-sectional, population-based	991	none	924 individuals left after QC based on genotyping quality, sex and ancestry check	None of the first 3 principal components strongly associated with uric acid; relatedness of participants taken into account using a mixed linear model with the polygenic effect set as random effect.	UA was measured using the uricase UV photometry method in "Labor Centar" biochemical lab, Bukovcev trg 3, 10000 Zagreb Croatia (www.laborcentar.hr). A subset of 774 samples had also been measured independently in the Institute for Clinical Chemistry and Laboratory Medicine, University Hospital Regensburg, Germany. Pearson correlation between the two urate measurements was 94%.	Vitart et al. (2006)[18]
Data from the Epidemiological Study on the Insulin Resistance Data from the Epidemiological Study on the Insulin Resistance Syndrome (DESIR) Study	Controls for the study of T2D and obesity selected from a population-based study.	716 of European ancestry	none	Using the STRUCTURE software, we identified 4 individuals of non-European ancestry. In order to minimize admixture bias in the rest of the DESIR participants, we excluded these individuals before analyses	none	UA was measured using the uricase method [19] at study visit 1. Repeated measurements of UA in 40 individuals, taken at least one week apart, yielded a reliability coefficient of 0.91, and the coefficient of variation was 7.2% [20].	Balkau et al. (1997)[19], Vernay et al. (2004)[20]
European Prospective Investigation of Cancer (EPIC) Norfolk Study	Prospective, population-based, case-cohort design consisting of a random sample (cohort) of 2566 participants at baseline and 1284 obese cases [21],*	3,850 of European ancestry	none	We excluded individuals who were duplicated samples DNA concordance >99%, cryptically related, related individuals DNA concordance >70% and <99%, ethnic outliers, and heterozygosity <23% or >30%. In the discovery analysis only controls were used. Obese cases were used for replication.	The 3552 individuals who were used for GWAS repeatedly showed no evidence of population stratification. Consequently, we have not adjusted for population stratification.	Out of these individuals 2856 had uric acid measured, marked as serum L:89 H:1785 umol/L Olympus AU640.	Day et al. (1999)[21], *http://www.srl.cam.ac.uk/epic/about/
Erasmus Ruchphen Family (ERF) Study	Family based	2,385	none	none	Score test for association in related people implemented in R package GenABEL	UA concentrations were measured using an uricase/peroxidase method (DVIA1650-Autoanalyzer, Siemens Healthcare	Pardo et al. (2005)[22]

					was used to control for family related ness.	Diagnostics)	
Estonian Genome Center of University of Tartu (EGCUT)	Prospective, population-based	931 of European ancestry	none	Low genotyping quality (call rate <98%, MAF <1%, HWE p-value 10E-6); disagreement between reported and genotypic sex, one randomly selected member of a pair of first-degree relatives	Three principal components were associated with uric acid measurements and included as covariates in the regression.	UA was measured using the uricase method	Nelis et al. (2009)[23], Metspalu et al. (2004)[24]
Family Heart Study (FamHS)	Population family-based [25]	4,135 of European ancestry		Quality control was performed before imputation. To assess Mendelian errors, we ran LOKI on our family data and removed 5,035 SNPs with Mendelian errors. We also removed 2 individuals that had an unaccepted number of Mendelian errors. As a final familial QC check, we used GRR software to check familial relationships based on IBS. Quality control procedures for SNPs included cleaning SNPs reported by Illumina as uninformative and unavailable on successive arrays (n=13,844), removing SNPs due to deviations from Hardy-Weinberg equilibrium (p<1E-06) or SNPs with minor allele frequency <1% or >99% (n=22,088), and removing SNPs that are available in our data but not in HapMap (n=1,509). Additionally, 21 SNPs were designated as ambiguous and removed. After these quality control procedures, genotypes are available for 4,135 European American (EA) subjects with imputed genotypes for ~2.5 million SNPs.	Ten principal components (EIGENSTRAT) were estimated using the genotype data of the largest sample of independent subjects (N= 753) and then applied to the family members. These principal components were included in the adjustment procedure of uric acid using stepwise regression analysis and held if they were significant at 5% level.	Uric acid was measured by a thin film adaptation of an uricase enzymatic method using the Vitros analyzer (Johnson & Johnson Clinical Diagnostics, Inc. Rochester NY 14650).	Higgins et al. (1996)[25], Neogi et al.. (2011)[26], Neogi et al. (2009)[27], Tang et al. (2006)[28], Tang et al. (2003)[29], Wilk et al (2000)[30]
The Framingham Heart Study	Prospective, family based	9,274	none	Individuals with a sample call rate <97%, or heterozygosity > ±5 SD from the mean are excluded from association analyses.	Principal components of the genotypes of 550K SNPs were computed using the Eigenstrat software 5, and none of the first 10 components were found association with either urate levels or gout using a Bonferroni correction on alpha	Serum urate was measured at the first examination cycle of each cohort using an autoanalyzer with a phosphotungstic acid reagent	Dawber et al. (1963)[31], Crowley et al. (1964)[32]

					of 0.05, which indicated that there is little population admixture for these two traits and therefore no need to adjust for admixture in GWAS.		
Health 2000	Population-based	2,123 Finns	none	Samples with discrepancy between reported and genotypic sex were excluded. For pairs with pi_hat > 0.2 one of the pairs was excluded. Individuals with 0.05 < pi_hat < 0.2 to many other individuals were excluded.	NA	Uricase method, a colorimetric enzymatic method (Thermo Fisher Scientific, Vantaa, Helsinki).	http://www.terveys2000.fi/doc/methodologyrep.pdf
InCHIANTI study	Prospective, population-based	1,230 European ancestry	none	Of the 1231 genotyped subjects, 22 subjects were removed based on genotyping completeness (<97%), low heterozygosity (<0.3), or sex misspecification. 1205 subjects with uric acid data was used for the analysis.	Genomic Control	Plasma UA (mg/dl) was measured using an enzymatic-colorimetric method (Roche Diagnostics, GmbH, Germany). The lower limits of detection were 0.2 mg/dl, range 0.2–25.0 mg/dl, intra- assay and inter assay coefficients of variation (CV) were 0.5 and 1.7%, respectively.	Ferrucci et al. (2000)[33]
INCIPE	Randomly chosen from the lists of patients of 62 randomly selected general practitioners (GPs) based in four geographical areas in the Veneto region, Northern Italy.	942 from Northern Italy	none	992 genotyped individuals (then 50 removed). Disagreement between reported and genotypic sex, one randomly selected member of a pair of first-degree relatives	From same geographical area	UA was measured using the UV uricase method; the between series CV is 1.5%	Gambaro et al. (2010)[34]
INGI-Carlantino	Population-Based	659	none	Removed people with call rate <0.95 or too high IBS or heterozigosity. Removed people that did not pass sex chromosome checks or were < 18 years of age.	Corrected using mixed model regression analysis.	UA was measured with the colorimetric method using Targa 3000 from Biotecnica Instruments.	Tepper et al. (2008)[35]
INGI-CILENTO	Population-Based study with pedigree information	859	none	Of the 859 participants who underwent genotyping, none was excluded	none	UA was measured using an enzymatic method.	Ciullo et al. (2006)[36], Colonna et al. (2007)[37], Ciullo et al. (2008)[38], Sala et al. (2008)[39], Traglia et al. (2009)[40], Heid et al. (2009)[41], Colonna et al. (2009)[42], Bedin et al. (2009)[43], Siervo et al. (2010)[44]

INGI-FVG	Population-Based	1,471	none	Removed people with call rate <0.95 or too high IBS or heterozigosity. Removed people that did not pass sex chromosome checks or were < 18 years of age.	Corrected using mixed model regression analysis.	UA was measured with the colorimetric method using Targa 3000 from Biotecnica Instruments	Girotto et al (2011)[45]
INGI-Val Borbera	Family Population-based	1,665	none	Of the 1665 participants who underwent genotyping, we made the following exclusions: sample call rate <95% (n=1)	NA	UA was measured using HITACHI 917 ROCHE and Unicel Dx-C 800 BECKMAN	Traglia et al. (2009)[40]
KORA F3	Population-based	1,644	none	Only subjects with overall genotyping efficiencies of at least 93% were included. In addition the called gender had to agree with the gender in the KORA study database.	none	Non-fasting blood samples were obtained from study participants. Serum urate analyses were carried out on fresh samples. Serum urate concentrations were measured using an uricase method (URCA Flex, Dade Behring).	Wichmann et al. (2005)[46]
KORA F4	Population-based	1,814	none	Only subjects with overall genotyping efficiencies of at least 93% were included. In addition the called gender had to agree with the gender in the KORA study database.	none	Fasting blood samples were obtained from study participants. Serum urate analyses were carried out on fresh samples. Serum urate concentrations were measured using an uricase method (URCA Flex, Dade Behring).	Wichmann et al. (2005)[46]
LBC1936	Retrospective and prospective community-based cohort study[47]	1,005 of European ancestry	none	Individuals with a disagreement between genetic and reported gender were removed (n=12). Relatedness between subjects was investigated and, for any related pair of individuals, one was removed (PI_HAT (proportion of IBD) > 0.25, n=8). Samples with a call rate ≤ 0.95 (n=16), and those showing evidence of non-European descent by multidimensional scaling, were also removed (n=1).	None of the four extracted principal components were associated with uric acid measurements so were not included in the model.	Serum uric acid was determined using the VITROS URIC DT slide method performed using the VITROS URIC DT slide and the VITROS Chemistry products DT Calibrator Kit on VITROS DT60/DT60 II Chemistry systems (VITROS). This was performed at the Combined Biochemistry and Haematology Labs, Western General Hospital, Edinburgh.	Deary et al. (2007)[47], Houlihan et al. (2010)[48]
LifeLines Cohort Study	Prospective, population-based	3,367 of European ancestry	none	Of the 3900 genotyped individuals, we excluded 533 individuals based on discrepancies with previous genotypes, disagreement between reported and genotypic sex, one randomly selected member of a pair of first-degree relatives, and non-European ancestry	NA	Uric acid was measured on a Roche/Hitachi Modular System (Roche Diagnostics GmbH), by the uricase/peroxida enzymatic method	Stolk et al. (2008)[49]
London Life Sciences Population (LOLIPOP) study, LOLIPOP_EW610	Prospective, population-based	945		Duplicates, gender discrepancy, contaminated samples, relatedness	The first ten principal components were used as covariates in the regression.	Venous blood was collected into 5.0ml BD Vacutainer SST II Advance tube. Serum urate measurements were measured using the uricase method on Roche/Hitachi Cobas C 501 systems (USA).	
London Life Sciences	Prospective, population-	878		Duplicates, contaminated samples, relatedness, samples already in	The first ten principal	Venous blood was collected into 5.0ml BD Vacutainer SST	Yuan et al. (2008)[50]

Population (LOLIPOP) study, LOLIPOP_EW_A	based			EW610	components were used as covariates in the regression.	II Advance tube. Serum urate measurements were measured using the uricase method on Roche/Hitachi Cobas C 501 systems (USA).	
London Life Sciences Population (LOLIPOP) study, LOLIPOP_EW_P	Prospective, population-based	1,006		Duplicates, contaminated samples, samples already in EW610 and EW_A	The first ten principal components were used as covariates in the regression.	Venous blood was collected into 5.0ml BD Vacutainer SST II Advance tube. Serum urate measurements were measured using the uricase method on Roche/Hitachi Cobas C 501 systems (USA).	Kooner et al. (2008)[51]
Ludwigshafen Risk and Cardiovascular Health Study (LURIC)	Prospective, case-control (CAD)	963	any acute illness other than ACSs, any chronic disease where non-cardiac disease predominated a history of malignancy within the past five years	Individuals with genotyping call rates below 0.96 were removed.	none	UA was measured using a photometric colour test (Harnsäure Farb-Reagenz, Greiner, Germany) on a Hitachi 717 at study entry.	Winkelmann et al. (2001)[52]
MICROS	Cross-sectional, population-based	1,345	none	1,268 individuals left after QC based on genotyping quality, sex and ancestry check.	None of the first 3 principal components strongly associated with uric acid but village of origin kept as cofactor; relatedness of participants taken into account using a mixed linear model with the polygenic effect set as random effect.	UA was measured using the uricase /peroxidase method.	Pattaro et al. (2007)[53]
Netherlands Study of Depression and Anxiety (NESDA)	Longitudinal cohort study of individuals with depressive and/or anxiety disorder	1,862 of western-European ancestry	Individuals were almost all cases with major depression or anxiety disorder (n=1705)	Ethnic outliers, XO and XXY samples, and samples with a call rate <95%, high genome-wide homo- or heterozygosity, excess IBS were excluded	none	UA was measured by enzymatic colorimetric test (uricase method, Roche Modular system). The coefficients of variation, over the complete measurement period, were 1.6% at a level of 0.25 mmol/l and 1.2% at a level of 0.55 mmol/l.	Penninx et al. (2008)[54], Sullivan et al. (2009)[55]
NSPHS	Cross-sectional, population-based	700	none	656 individuals left after QC based on genotyping quality, sex and ancestry check	None of the first 3 principal components strongly associated with uric acid; relatedness of participants taken into account using a mixed linear model with the polygenic effect set	UA was measured using the uricase /peroxidase method.	Igl et al. (2010)[56]

					as random effect.		
ORCADES	Cross-sectional, population-based	920	of non-orcadian ancestry	889 individuals left after QC based on genotyping quality, sex and ancestry check.	None of the first 3 principal components strongly associated with uric acid; relatedness of participants taken into account using a mixed linear model with the polygenic effect set as random effect.	UA was measured using the uricase /peroxidase method in the Balfour Hospital, Kirkwall, UK. A subset of 718 samples had also been measured independently in the Institute for Clinical Chemistry and Laboratory Medicine, University Hospital Regensburg, Germany. Pearson correlation between the two urate measurements was 99%.	McQuillan et al. (2008)[57]
PREVEND	Prospective, population-based [58]	4,016 of European ancestry	none	Of the 4,016 genotyped individuals, we excluded 148 individuals based on discrepancies with previous genotypes, disagreement between reported and genotypic sex, first-degree relatives, or outlier based PCA.	none	Uric acid was measured in plasma and urine with the uricase PAP method as described previously (MEGA, Merck, Darmstadt, Germany). [59]	Hillege et al. (2002)[58]
Procardis	Case-Control study of CAD	3,742	none	Dataset was prefiltered for individuals with success rate <95%, ancestory outliers on PCA, heterozygosity, IBC	Country of Origin was added as a covariate, population stratification was checked using PCA but was not adjusted for beyond Country of Origin.	Measured using uricase method in hospital clinical lab	Broadbent et al. (2008)[60]
RS-I	Prospective, population based	5,974	none		none	Serum urate was measured at the baseline visit using a Kone Diagnostica reagent kit and autoanalyzer.	Hofman et al. (1991)[61], Hofman et al. (2009)[62]
RS-II	Prospective, population based	2,157	none		none	Serum urate was measured at the baseline visit using a Kone Diagnostica reagent kit and autoanalyzer.	Hofman et al. (1991)[61], Hofman et al. (2009)[62]
SardiNIA Study	Population-based study in Sardinia. The SardiNIA study consists of 6,148 individuals, males and females, ages 14-102 y, that were recruited from a cluster of four towns in the Lanusei Valley of	4,694	none	none	none	During physical examination, a blood sample was collected in the morning after the participants had been fasting for at least 12 h and after sitting for 15 min and divided into two aliquots. One was used for genomic DNA extraction and the second aliquot to characterize several blood phenotypes, including evaluation of serum UA. UA (mg/dl) was measured using enzymatic–colorimetric methods (Bayer) The lower	Pilia et al. (2006)[63], Li et al. (2007)[64], Sanna et al. (2008)[65]

	Sardinia. Samples have been characterized for several quantitative traits and medical conditions, including serum urate.					limits of detection were 0.2 mg/dl, range 0.2–25.0 mg/dl, intra-assay and inter assay coefficients of variation were equal to 0.5% and 1.7%, respectively.	
Study of Health in Pomerania (SHIP)	Population-based	4,081 of European ancestry	none	24 individuals identified as duplicated or with reported/genotyped gender mismatch	none	Uricase method, a colorimetric enzymatic method (Uric acid PAP, Boehringer) from non-fasting, fresh serum	John et al. (2001)[66], Völzke et al. (2011)[67]
SOCCS	Colorectal cancer case control study, population-based	2,024	none	1,984 individuals after QC, 1,105 of whom had uric acid phenotypes.	No PCs of ancestry included in analysis	UA was measured using the uricase /peroxidase method.	Tenesa et al. (2008)[68]
Sorbs	Population-based	1,020	46 individuals excluded as they were on a medication that lowers serum uric acid	ethnic outliers, duplicates, and gender mismatches	Estimation of kinship matrix to take account of relatedness.	enzymatic color test (Roche Diagnostics, Inc)	Tönjes et al. (2009)[69], Tönjes et al. (2010)[70], Veeramah et al. (2011)[71]
TwinsUK	Twins	5,654 of European ancestry	none	Samples: Exclusion criteria were: (i) sample call rate <98%, (ii) heterozygosity across all SNPs >2 s.d. from the sample mean; (iii) evidence of non-European ancestry as assessed by PCA comparison with HapMap3 populations; (iv) observed pairwise IBD probabilities suggestive of sample identity errors; (v). We corrected misclassified monozygotic and dizygotic twins based on IBD probabilities.	Estimation of kinship matrix to take account of relatedness.	Ektachem/Vitros system, Johnson & Johnson Clinical Diagnostics	Moayyeri et al. (2012)[72]
Young Finns Study	Birth cohort follow-up	2,443 Finns	none	Samples with discrepancy between reported and genotypic sex were excluded. For pairs with pi_hat > 0.2 one of the pairs was excluded. Individuals with 0.05 < pi_hat < 0.2 to many other individuals were excluded.	none	Uricase method, a colorimetric enzymatic method (Thermo Fisher Scientific, Vantaa, Helsinki).	Raitakari et al. (2008)[73]
In silico replication studies							
EPIC - cases	See description above.						
GSK cases/controls	Case-control study for unipolar depressive	819 cases/ 851 controls of European ancestry	GSK cases: patients with unipolar recurrent depression,	MDS-analysis revealed no outliers (more than 8SD away on any of the first 10 principal components): after QC 819 cases/851 controls.	No principal component was associated with uric acid so none was	UA was measured using the uricase method (Roche/Hitachi cobas c system, UA ver.2).	Lucae et al. (2006)[74], Kloiber et al. (2010)[75], Kohli et al. (2011)[76]

	disorder		exclusion criteria: presence of manic or hypomanic episodes, mood incongruent psychotic symptoms, lifetime diagnosis of drug abuse and depressive symptoms secondary to alcohol or substance abuse or dependence or to a medical illness or medication GSK controls: exclusion criteria: anxiety and affective disorders.		included as covariates. No principal component was associated with uric acid so none was included as covariate.		
Gutenberg Health Study (GHS I + II)	Population-based	4860 (3422 (GHS I) + 1438 (GHS II))	age below 35 and above 74	Of the 4860 we excluded 685 (426 + 259) based on a call rate less than 97 %, a rate of heterorzygosity 3 standard deviations away from the mean, disagreement between reported and genotypic sex, estimated IBD > 0.25, IBS based principal components.	none	UA was measured using the uricase method at study visit during routine meausrements. Intra coefficient of variation (CV) was 0% at a mean value of 4.9 mg/dL and 0.44% at a mean value of 9.52 mg/dL, the inter CV% was 2.25% at a mean value of 4.9 mg/dL, and 0.97% at a mean value of 9.4 mg/dL.	Zeller et al. (2010)[77], Wild et al. (2010)[78], Wild et al. (2011)[79]
Hunter Community Study (HCS)	Prospective, population-based	1,230 of European ancestry	none	Individuals were excluded for genotype call rate <95%, discrepancies between clinical and inferred gender, one randomly selected member of a pair of first- or second-degree relatives or clear evidence of non-European ancestry in Eigenstrat PCA.	No principal components were associated with uric acid, and were not included as covariates.	The HAPS pathology service did the urate measurements. They are a NATA accredited lab and meet national standards for quality assurance.	McEvoy et al. (2010)[80]
LifeLines Cohort Study	Prospective, population-based	5,031 of European ancestry	none	none	NA	Uric acid was measured on a Roche/Hitachi Modular System (Roche Diagnostics GmbH), by the uricase/peroxidase enzymatic method.	Stolk et al. (2008)[49]
Ludwigshafen Risk and Cardiovascular Health Study (LURIC)	Prospective, case-control (CAD)	1,960	any acute illness other than ACSs, any chronic disease where non-cardiac disease predominated a history of malignancy within the past five years	Individuals which were part of the discovery analysis were removed. Samples were also removed because of gender discrepancy, relatedness or low call rate (<90%).	Sample HD (Heidelberg) n=1156 and GZ (Graz) n=804 were analyzed separately.	UA was measured using a photometric colour test (Harnsäure Farb-Reagenz, Greiner, Germany) on a Hitachi 717 at study entry.	Winkelmann et al. (2001)[52]

MARS cases	Case-control study for depressive disorder	643 cases of European ancestry	MARS cases: patients with depressive episode, exclusion criteria: depressive disorders caused by a medical or neurologic condition and alcohol or substance dependence.	MDS-analysis revealed 7 outliers (more than 8SD away on any of the first 10 principal components): after QC: 636 cases.	No principal component was associated with uric acid so none was included as covariates. No principal component was associated with uric acid so none was included as covariate.	UA was measured using the uricase method (Roche/Hitachi cobas c system, UA ver.2).	Kohli et al. (2011)[76]
Ogliastra Genetic Park - Talana	Population-based study with pedigree information	860	none	none	none	Uric acid levels were measured using the uricase method with an automated TARGA BT-3000 Chemistry Analyser	Portas et al. (2010)[81], Biino et al. (2010)[82], Tore et al. (2011)[83]
Study of Health in Pomerania - Trend (SHIP-Trend)	Population-based	986 of European ancestry	none	array call rate < 94%, individuals identified as duplicated or with reported/genotyped gender mismatch	none	UA was measured from non-fasting, fresh serum. An Uricase methode was used on a Dimension Vista® System (SIEMENS, Eschborn, Germany). The coefficient of variation was 1.92% at low level of control material (mean value = 291 µmol/L).	John et al. (2001)[66], Völzke et al. (2011)[67]
Swiss Cohort Study on **A**ir **P**ollution **A**nd **L**ung and Heart **Di**seases in **A**dults	Prospective, population-based	1,640	asthmatics and non-asthmatics separate	28 failed genotyping, 35 low call rate (<97%), 17 non-European descent, 64 cryptic relatedness, 26 overlap with ECRHS, 12 males with high X-heterozygosity, 1 sex inconsistency, 13 missing UA levels (=1444 included in this analysis)	Two principal components were included as covariates in the regression.	Uric acid concentrations were determined by a colorimetric uricase/peroxidase method using reagents and the Modular P autoanalyser from Roche diagnostics (Rotkreuz, Switzerland). At concentrations of 203 micromol/l and 355 micromol/l the inter assay imprecision was 1% or less.	Martin et al. (1997)[84], Ackermann-Liebrich et al. (2005)[85]
De novo replication studies							
HYPertension in ESTonia (HYPEST)	Hypertensive cases recruited at the clinics and population-based controls [86-88]	758 of European (Estonian) ancestry	none	none	none	The venous blood for serum biomarker analysis was drawn in the morning after an overnight fast [86,88]. UA was measured by standardized assays (Cobas Integra 800Ò analytical platform, Roche Diagnostics, Inc.) at the United Laboratories, Tartu University Clinics or at the Diagnostics Division Laboratory, the North Estonia Medical Centre [88]. EURACHEM guidelines were applied to estimate measurement uncertainty (9.7%).	Ong et al. (2011)[86], Ong et al.(2009)[87], Juhanson et al.(2008)[88]
KORA S2	Population-	3,685	none	Only subjects with overall	none	Non-fasting blood samples	Wichmann et al.

	based			genotyping efficiencies of at least 93% were included.		were obtained from study participants. Urate analyses were carried out on fresh samples. Urate concentrations were measured using an uricase method (Technicon, SMAC AutoAnalyzer).	(2005)[46]
Ogliastra Genetic Park	Population-based study with pedigree information	9,704 of Sardinian ancestry	none	Individuals with a call rate <0.9 in de novo genotyping were excluded.	Study center was included as covariate in the regression	UA was measured in MG/DL units using TARGA 3000 with enzymatic colorimetric uricase method.	Portas et al. (2010)[81], Biino et al. (2009)[82], Pistis et al. (2009)[89]
Study samples of non-European ancestry							
London Life Sciences Population (LOLIPOP) study, LOLIPOP_IA317	Prospective, population-based	2,694	none	Duplicates, gender discrepancy, contaminated samples, relatedness, samples already in IA610	The first ten principal components were used as covariates in the regression.	Venous blood was collected into 5.0ml BD Vacutainer SST II Advance tube. Serum urate measurements were measured using the uricase method on Roche/Hitachi Cobas C 501 systems (USA).	Chambers et al. (2008)[90]
London Life Sciences Population (LOLIPOP) study, LOLIPOP_IA610	Prospective, population-based	7,032	none	Duplicates, gender discrepancy, contaminated samples, relatedness	The first ten principal components were used as covariates in the regression.	Venous blood was collected into 5.0ml BD Vacutainer SST II Advance tube. Serum urate measurements were measured using the uricase method on Roche/Hitachi Cobas C 501 systems (USA).	Chambers et al. (2008)[90]
London Life Sciences Population (LOLIPOP) study, LOLIPOP_IA_P	Prospective, population-based	1,005	none	Duplicates, contaminated samples, samples already in IA610 and IA317	The first ten principal components were used as covariates in the regression.	Venous blood was collected into 5.0ml BD Vacutainer SST II Advance tube. Serum urate measurements were measured using the uricase method on Roche/Hitachi Cobas C 501 systems (USA).	Kooner et al. (2008)[51]
ARIC	Population-based	2,749	not self-identified as black	Data cleaning conducted centrally at the Broad Institute	Adjustment for the first 10 principle components	Serum urate concentrations were measured with the uricase method at visit 1	Iribarren et al. (1996)[5]
CARDIA	Population-based	937	none	Data cleaning conducted centrally at the Broad Institute	Adjustment for the first 10 principle components	Serum urate was measured at baseline using the uricase method.	Friedman et al. (1988)[13]
JHS	Population-based	3,443	none	Data cleaning conducted centrally at the Broad Institute	Adjustment for the first 10 principle components	Baseline serum urate was measured using the uricase method.	Taylor et al. (2005)[91], Fuqua eta al. (2005)[92]
The BioBank Japan Project	Disease patients cohort [93]	15,288 of Japanese disease patients affected with each of the 21 diseases [93-95].	none	The following subjects were excluded.(i) low call rate (<98%), (ii) in 1st or 2nd kinships, (iii) outliers from East-Asian clusters in the result of principal component analysis (PCA) performed with HapMap Phase II populations, (iv) serum urate, sex, or age were not available, (v) age <18, age > 85, with dialysis treatment or with kidney failure.	Subjects who were determined to be of non-Japanese origin by self-report or by PCA were excluded. No principal component was included as covariate in the regression.	UA levels were obtained from medical records of the medical institutes which participated in the BioBank Japan Projects.	Nakamura et al. (2007)[93], Kamatani et al. (2010)[94], Okada et al. (2011)[95]

References Table S2

1. Harris, T.B. *et al.* Age, Gene/Environment Susceptibility-Reykjavik Study: multidisciplinary applied phenomics. *American journal of epidemiology* **165**, 1076-87 (2007).
2. Mitchell, B.D. *et al.* The genetic response to short-term interventions affecting cardiovascular function: rationale and design of the Heredity and Phenotype Intervention (HAPI) Heart Study. *American heart journal* **155**, 823-8 (2008).
3. McArdle, P.F. *et al.* Association of a common nonsynonymous variant in GLUT9 with serum uric acid levels in old order amish. *Arthritis and rheumatism* **58**, 2874-81 (2008).
4. The Atherosclerosis Risk in Communities (ARIC) Study: design and objectives. The ARIC investigators. *American journal of epidemiology* **129**, 687-702 (1989).
5. Iribarren, C., Folsom, A.R., Eckfeldt, J.H., McGovern, P.G. & Nieto, F.J. Correlates of uric acid and its association with asymptomatic carotid atherosclerosis: the ARIC Study. Atherosclerosis Risk in Communities. *Annals of epidemiology* **6**, 331-40 (1996).
6. Eckfeldt, J.H., Chambless, L.E. & Shen, Y.L. Short-term, within-person variability in clinical chemistry test results. Experience from the Atherosclerosis Risk in Communities Study. *Archives of pathology & laboratory medicine* **118**, 496-500 (1994).
7. Schmidt, R. *et al.* Assessment of cerebrovascular risk profiles in healthy persons: definition of research goals and the Austrian Stroke Prevention Study (ASPS). *Neuroepidemiology* **13**, 308-13 (1994).
8. Whitfield, J.B., Zhu, G., Nestler, J.E., Heath, A.C. & Martin, N.G. Genetic covariation between serum gamma-glutamyltransferase activity and cardiovascular risk factors. *Clinical chemistry* **48**, 1426-31 (2002).
9. Middelberg, R.P., Medland, S.E., Martin, N.G. & Whitfield, J.B. A longitudinal genetic study of uric acid and liver enzymes in adolescent twins. *Twin research and human genetics : the official journal of the International Society for Twin Studies* **10**, 757-64 (2007).
10. Benyamin, B. *et al.* Common variants in TMPRSS6 are associated with iron status and erythrocyte volume. *Nature genetics* **41**, 1173-5 (2009).
11. Shock, N.W. *et al.* Normal Human Aging: The Baltimore Study of Aging. *U. S. Department of Health and Human Services* (1984).
12. Caulfield, M. *et al.* Genome-wide mapping of human loci for essential hypertension. *Lancet* **361**, 2118-23 (2003).
13. Friedman, G.D. *et al.* CARDIA: study design, recruitment, and some characteristics of the examined subjects. *Journal of clinical epidemiology* **41**, 1105-16 (1988).
14. Fried, L.P. *et al.* The Cardiovascular Health Study: design and rationale. *Annals of epidemiology* **1**, 263-76 (1991).
15. Firmann, M. *et al.* The CoLaus study: a population-based study to investigate the epidemiology and genetic determinants of cardiovascular risk factors and metabolic syndrome. *BMC cardiovascular disorders* **8**, 6 (2008).
16. Zemunik, T. *et al.* Genome-wide association study of biochemical traits in Korcula Island, Croatia. *Croat Med J* **50**, 23-33 (2009).
17. Rudan, I. *et al.* "10001 Dalmatians:" Croatia launches its national biobank. *Croat Med J* **50**, 4-6 (2009).
18. Vitart, V. *et al.* 3000 years of solitude: extreme differentiation in the island isolates of Dalmatia, Croatia. *Eur J Hum Genet* **14**, 478-87 (2006).
19. Balkau, B., Eschwege, E., Tichet, J. & Marre, M. Proposed criteria for the diagnosis of diabetes: evidence from a French epidemiological study (D.E.S.I.R.). *Diabetes & metabolism* **23**, 428-34 (1997).
20. Vernay, M. *et al.* Alcohol consumption and insulin resistance syndrome parameters: associations and evolutions in a longitudinal analysis of the French DESIR cohort. *Annals of epidemiology* **14**, 209-14 (2004).
21. Day, N. *et al.* EPIC-Norfolk: study design and characteristics of the cohort. European Prospective Investigation of Cancer. *British journal of cancer* **80 Suppl 1**, 95-103 (1999).

22. Pardo, L.M., MacKay, I., Oostra, B., van Duijn, C.M. & Aulchenko, Y.S. The effect of genetic drift in a young genetically isolated population. *Annals of human genetics* **69**, 288-95 (2005).
23. Nelis, M. *et al.* Genetic structure of Europeans: a view from the North-East. *PloS one* **4**, e5472 (2009).
24. Metspalu, A., Kohler, F., Laschinski, G., Ganten, D. & Roots, I. [The Estonian Genome Project in the context of European genome research]. *Deutsche medizinische Wochenschrift* **129 Suppl 1**, S25-8 (2004).
25. Higgins, M. *et al.* NHLBI Family Heart Study: objectives and design. *American journal of epidemiology* **143**, 1219-28 (1996).
26. Neogi, T., Terkeltaub, R., Ellison, R.C., Hunt, S. & Zhang, Y. Serum Urate Is Not Associated with Coronary Artery Calcification: The NHLBI Family Heart Study. *The Journal of rheumatology* **38**, 111-7 (2011).
27. Neogi, T. *et al.* Serum uric acid is associated with carotid plaques: the National Heart, Lung, and Blood Institute Family Heart Study. *The Journal of rheumatology* **36**, 378-84 (2009).
28. Tang, W. *et al.* Familial clustering for features of the metabolic syndrome: the National Heart, Lung, and Blood Institute (NHLBI) Family Heart Study. *Diabetes care* **29**, 631-6 (2006).
29. Tang, W. *et al.* Linkage analysis of a composite factor for the multiple metabolic syndrome: the National Heart, Lung, and Blood Institute Family Heart Study. *Diabetes* **52**, 2840-7 (2003).
30. Wilk, J.B. *et al.* Segregation analysis of serum uric acid in the NHLBI Family Heart Study. *Human genetics* **106**, 355-9 (2000).
31. Dawber, T.R., Kannel, W.B. & Lyell, L.P. An approach to longitudinal studies in a community: the Framingham Study. *Annals of the New York Academy of Sciences* **107**, 539-56 (1963).
32. Crowley, L.V. Determination of Uric Acid: An Automated Analysis Based on a Carbonate Method. *Clinical chemistry* **10**, 838-44 (1964).
33. Ferrucci, L. *et al.* Subsystems contributing to the decline in ability to walk: bridging the gap between epidemiology and geriatric practice in the InCHIANTI study. *Journal of the American Geriatrics Society* **48**, 1618-25 (2000).
34. Gambaro, G. *et al.* Prevalence of CKD in northeastern Italy: results of the INCIPE study and comparison with NHANES. *Clinical journal of the American Society of Nephrology : CJASN* **5**, 1946-53 (2010).
35. Tepper, B.J. *et al.* Variation in the bitter-taste receptor gene TAS2R38, and adiposity in a genetically isolated population in Southern Italy. *Obesity* **16**, 2289-95 (2008).
36. Ciullo, M. *et al.* New susceptibility locus for hypertension on chromosome 8q by efficient pedigree-breaking in an Italian isolate. *Human molecular genetics* **15**, 1735-43 (2006).
37. Colonna, V. *et al.* Campora: a young genetic isolate in South Italy. *Human heredity* **64**, 123-35 (2007).
38. Ciullo, M. *et al.* Identification and replication of a novel obesity locus on chromosome 1q24 in isolated populations of Cilento. *Diabetes* **57**, 783-90 (2008).
39. Sala, C. *et al.* Variation of hemoglobin levels in normal Italian populations from genetic isolates. *Haematologica* **93**, 1372-5 (2008).
40. Traglia, M. *et al.* Heritability and demographic analyses in the large isolated population of Val Borbera suggest advantages in mapping complex traits genes. *PloS one* **4**, e7554 (2009).
41. Heid, I.M. *et al.* Meta-analysis of the INSIG2 association with obesity including 74,345 individuals: does heterogeneity of estimates relate to study design? *PLoS genetics* **5**, e1000694 (2009).
42. Colonna, V. *et al.* Comparing population structure as inferred from genealogical versus genetic information. *European journal of human genetics : EJHG* **17**, 1635-41 (2009).
43. Bedin, E. *et al.* Age-related hearing loss in four Italian genetic isolates: an epidemiological study. *International journal of audiology* **48**, 465-72 (2009).
44. Siervo, M. *et al.* Angiogenesis and biomarkers of cardiovascular risk in adults with metabolic syndrome. *Journal of internal medicine* **268**, 338-47 (2010).
45. Girotto, G. *et al.* Hearing function and thresholds: a genome-wide association study in European isolated populations identifies new loci and pathways. *Journal of medical genetics* **48**, 369-74 (2011).

46. Wichmann, H.E., Gieger, C. & Illig, T. KORA-gen--resource for population genetics, controls and a broad spectrum of disease phenotypes. *Gesundheitswesen* **67 Suppl 1**, S26-30 (2005).
47. Deary, I.J. *et al.* The Lothian Birth Cohort 1936: a study to examine influences on cognitive ageing from age 11 to age 70 and beyond. *BMC geriatrics* **7**, 28 (2007).
48. Houlihan, L.M. *et al.* Common variants of large effect in F12, KNG1, and HRG are associated with activated partial thromboplastin time. *American journal of human genetics* **86**, 626-31 (2010).
49. Stolk, R.P. *et al.* Universal risk factors for multifactorial diseases: LifeLines: a three-generation population-based study. *European journal of epidemiology* **23**, 67-74 (2008).
50. Yuan, X. *et al.* Population-based genome-wide association studies reveal six loci influencing plasma levels of liver enzymes. *American journal of human genetics* **83**, 520-8 (2008).
51. Kooner, J.S. *et al.* Genome-wide scan identifies variation in MLXIPL associated with plasma triglycerides. *Nature genetics* **40**, 149-51 (2008).
52. Winkelmann, B.R. *et al.* Rationale and design of the LURIC study--a resource for functional genomics, pharmacogenomics and long-term prognosis of cardiovascular disease. *Pharmacogenomics* **2**, S1-73 (2001).
53. Pattaro, C. *et al.* The genetic study of three population microisolates in South Tyrol (MICROS): study design and epidemiological perspectives. *BMC Med Genet* **8**, 29 (2007).
54. Penninx, B.W. *et al.* The Netherlands Study of Depression and Anxiety (NESDA): rationale, objectives and methods. *International journal of methods in psychiatric research* **17**, 121-40 (2008).
55. Sullivan, P.F. *et al.* Genome-wide association for major depressive disorder: a possible role for the presynaptic protein piccolo. *Molecular psychiatry* **14**, 359-75 (2009).
56. Igl, W., Johansson, A. & Gyllensten, U. The Northern Swedish Population Health Study (NSPHS)--a paradigmatic study in a rural population combining community health and basic research. *Rural Remote Health* **10**, 1363 (2010).
57. McQuillan, R. *et al.* Runs of homozygosity in European populations. *Am J Hum Genet* **83**, 359-72 (2008).
58. Hillege, H.L. *et al.* Urinary albumin excretion predicts cardiovascular and noncardiovascular mortality in general population. *Circulation* **106**, 1777-82 (2002).
59. Mostert, J.P. *et al.* Serum uric acid levels and leukocyte nitric oxide production in multiple sclerosis patients outside relapses. *Journal of the neurological sciences* **231**, 41-4 (2005).
60. Broadbent, H.M. *et al.* Susceptibility to coronary artery disease and diabetes is encoded by distinct, tightly linked SNPs in the ANRIL locus on chromosome 9p. *Human molecular genetics* **17**, 806-14 (2008).
61. Hofman, A., Grobbee, D.E., de Jong, P.T. & van den Ouweland, F.A. Determinants of disease and disability in the elderly: the Rotterdam Elderly Study. *European journal of epidemiology* **7**, 403-22 (1991).
62. Hofman, A. *et al.* The Rotterdam Study: 2010 objectives and design update. *European journal of epidemiology* **24**, 553-72 (2009).
63. Pilia, G. *et al.* Heritability of cardiovascular and personality traits in 6,148 Sardinians. *PLoS genetics* **2**, e132 (2006).
64. Li, S. *et al.* The GLUT9 gene is associated with serum uric acid levels in Sardinia and Chianti cohorts. *PLoS genetics* **3**, e194 (2007).
65. Sanna, S. *et al.* Common variants in the GDF5-UQCC region are associated with variation in human height. *Nature genetics* **40**, 198-203 (2008).
66. John, U. *et al.* Study of Health In Pomerania (SHIP): a health examination survey in an east German region: objectives and design. *Sozial- und Praventivmedizin* **46**, 186-94 (2001).
67. Volzke, H. *et al.* Cohort Profile: The Study of Health in Pomerania. *International journal of epidemiology* **40**, 294-307 (2011).
68. Tenesa, A. *et al.* Genome-wide association scan identifies a colorectal cancer susceptibility locus on 11q23 and replicates risk loci at 8q24 and 18q21. *Nat Genet* **40**, 631-7 (2008).

69. Tonjes, A. *et al.* Genetic variation in GPR133 is associated with height: genome wide association study in the self-contained population of Sorbs. *Human molecular genetics* **18**, 4662-8 (2009).
70. Tonjes, A. *et al.* Association of FTO variants with BMI and fat mass in the self-contained population of Sorbs in Germany. *European journal of human genetics : EJHG* **18**, 104-10 (2010).
71. Veeramah, K.R. *et al.* Genetic variation in the Sorbs of eastern Germany in the context of broader European genetic diversity. *Eur J Hum Genet* **19**, 995-1001 (2011).
72. Moayyeri, A., Hammond, C.J., Valdes, A.M. & Spector, T.D. Cohort Profile: TwinsUK and Healthy Ageing Twin Study. *International journal of epidemiology* (2012).
73. Raitakari, O.T. *et al.* Cohort profile: the cardiovascular risk in Young Finns Study. *International journal of epidemiology* **37**, 1220-6 (2008).
74. Lucae, S. *et al.* P2RX7, a gene coding for a purinergic ligand-gated ion channel, is associated with major depressive disorder. *Hum Mol Genet* **15**, 2438-45 (2006).
75. Kloiber, S. *et al.* Variations in tryptophan hydroxylase 2 linked to decreased serotonergic activity are associated with elevated risk for metabolic syndrome in depression. *Mol Psychiatry* **15**, 736-47 (2010).
76. Kohli, M.A. *et al.* The neuronal transporter gene SLC6A15 confers risk to major depression. *Neuron* **70**, 252-65 (2011).
77. Zeller, T. *et al.* Genetics and beyond--the transcriptome of human monocytes and disease susceptibility. *PLoS One* **5**, e10693 (2010).
78. Wild, P.S. *et al.* Distribution and categorization of left ventricular measurements in the general population: results from the population-based Gutenberg Heart Study. *Circulation. Cardiovascular imaging* **3**, 604-13 (2010).
79. Wild, P.S. *et al.* A genome-wide association study identifies LIPA as a susceptibility gene for coronary artery disease. *Circulation. Cardiovascular genetics* **4**, 403-12 (2011).
80. McEvoy, M. *et al.* Cohort profile: The Hunter Community Study. *International journal of epidemiology* **39**, 1452-63 (2010).
81. Portas, L. *et al.* History, geography and population structure influence the distribution and heritability of blood and anthropometric quantitative traits in nine Sardinian genetic isolates. *Genetics research* **92**, 199-208 (2010).
82. Biino, G. *et al.* Genetic architecture of hand quantitative ultrasound measures: a population-based study in a Sardinian genetic isolate. *Bone* **46**, 1197-203 (2010).
83. Tore, S. *et al.* Application of a new method for GWAS in a related case/control sample with known pedigree structure: identification of new loci for nephrolithiasis. *PLoS genetics* **7**, e1001281 (2011).
84. Martin, B.W. *et al.* SAPALDIA: methods and participation in the cross-sectional part of the Swiss Study on Air Pollution and Lung Diseases in Adults. *Sozial- und Praventivmedizin* **42**, 67-84 (1997).
85. Ackermann-Liebrich, U. *et al.* Follow-up of the Swiss Cohort Study on Air Pollution and Lung Diseases in Adults (SAPALDIA 2) 1991-2003: methods and characterization of participants. *Sozial- und Praventivmedizin* **50**, 245-63 (2005).
86. Org, E. *et al.* HYPEST study: profile of hypertensive patients in Estonia. *BMC cardiovascular disorders* **11**, 55 (2011).
87. Org, E. *et al.* Genome-wide scan identifies CDH13 as a novel susceptibility locus contributing to blood pressure determination in two European populations. *Hum Mol Genet* **18**, 2288-96 (2009).
88. Juhanson, P. *et al.* N-acetyltransferase 8, a positional candidate for blood pressure and renal regulation: resequencing, association and in silico study. *BMC Med Genet* **9**, 25 (2008).
89. Pistis, G. *et al.* High differentiation among eight villages in a secluded area of Sardinia revealed by genome-wide high density SNPs analysis. *PLoS One* **4**, e4654 (2009).
90. Chambers, J.C. *et al.* Common genetic variation near MC4R is associated with waist circumference and insulin resistance. *Nature genetics* **40**, 716-8 (2008).

91. Taylor, H.A., Jr. *et al.* Toward resolution of cardiovascular health disparities in African Americans: design and methods of the Jackson Heart Study. *Ethn Dis* **15**, S6-4-17 (2005).
92. Fuqua, S.R. *et al.* Recruiting African-American research participation in the Jackson Heart Study: methods, response rates, and sample description. *Ethn Dis* **15**, S6-18-29 (2005).
93. Nakamura, Y. The BioBank Japan Project. *Clinical advances in hematology & oncology : H&O* **5**, 696-7 (2007).
94. Kamatani, Y. *et al.* Genome-wide association study of hematological and biochemical traits in a Japanese population. *Nature genetics* **42**, 210-5 (2010).
95. Okada, Y. *et al.* Genome-wide association study for C-reactive protein levels identified pleiotropic associations in the IL6 locus. *Human molecular genetics* **20**, 1224-31 (2011).

Table S3: **Genotyping information of GUGC studies**.

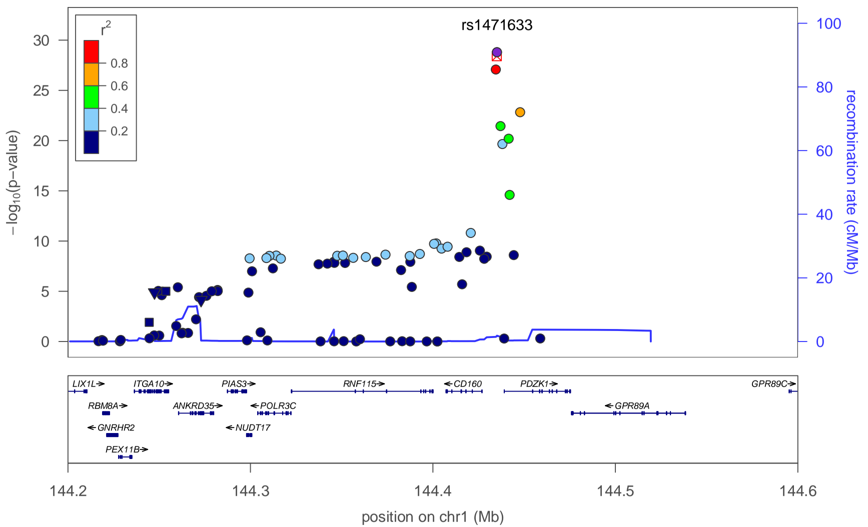
Study	Genotyping platform	Calling algorithm	Quality filters before imputation	SNPs before imputation	Imputation software	Imputation backbone (NCBI build)	Filtering of imputed genotypes	Data management and statistical analysis	inflation factor urate overall
Discovery studies									
AGES Reykjavik Study	Illumina Hu370CNV	Illumina BeadStudio	call rate <97%, MAF<1%, pHWE <10E-6	308,340	MACH v1.0.16	HapMap release 22 (build 36)	none	ProbABEL, PLINK, R	1.05
Amish	Affymetrix 500K, Affymetrix 6.0	BRLMM	call rate <95%, MAF<1%, pHWE <10E-6	338,598	MACH v1.0.15	HapMap release 22 (build 36)	none	Measured genotype accounting for polygenic component	1.03
ARIC	Affymetrix 6.0	Birdseed	call rate <95%, MAF<1%, pHWE <10E-5	669,450	MACH v1.0.16	HapMap release 22 (build 36)	none	ProbABEL, PLINK, R	1.02
ASPS	Illumina Human 610-Quad BeadChip	Illuminus software	call rate <97.5%, MAF <1%, , pHWE <1E-6	550,635	MACH v1.0.15	HapMap release 22 (build 36)	none	SPSS, ProbABEL, R	1.01
AUSTWIN	Illumina 370, Illumina 610	BeadStudio-gencall v3.0	call rate <95%, MAF <1%, pHWE <10E-5	269,840	MACH v1.0.15	HapMap release 22 (build 36)	r2 ≥ 0.3	MERLIN v1.1.2, PLINK v1.07	1.04
BLSA	Illumina 550K	BeadStudio	call rate <99%, MAF <1%, pHWE < 10E-4	501,704	MACH v1.0.16	HapMap release 22 (build 36)	none	ProbABEL, Merlin	1.02
BRIGHT	Affymetrix 500K	CHIAMO	call rate <95% (MAF>0.05) and <99% (MAF<0.05), pHWE <5.7E-7	490,032	IMPUTE	HapMap release 21 (build 35)	none	SNPTEST	1.00
CARDIA	Affymetrix 6.0	BEAGLE, Birdseed	call rate <95%, MAF<3%, pHWE <10E-4	579,630	BEAGLE	HapMap release 22 (build 36)	Rsq<0.3, MAF<1%	ProbABEL, PLINK, R	1.01
CHS	Illumina 370 CNV	Illumina BeadStudio	Call rate <97%, heterozygotes=0, pHWE<1E-5, SNP not in HapMap	306,655	BimBam	HapMap CEU release 22 (build 36)	dosage variance < 0.01	Linear an logistic regression using R	1.01
CoLaus	Affymetrix 500K	BRLMM	call rate <70%, MAF <1%, pHWE <1E-7	390,631	IMPUTE v0.2.0	HapMap release 21 (build 35)	none	Matlab	1.02
CROATIA-KORCULA	370CNV-Quad	BeadStudio	call rate <98%, MAF <1%, pHWE <10E-6	300,233	MACH v1	HapMap release 22 (build 36)	none	ProbABEL,mmscore argument	0.98
CROATIA-SPLIT	370CNV-Quadv3	GenomeStudio	call rate <98%, MAF <1%, pHWE <10E-6	330,889	MACH v1	HapMap release 22 (build 36)	none	ProbABEL,mmscore argument	1.01
CROATIA-VIS	HumanHap 300v1	BeadStudio	call rate <97%, MAF <1%, pHWE <10E-6	283,073	MACH v1	HapMap release 22 (build 36)	none	ProbABEL,mmscore argument	1.00
DESIR	Infinium Human1 and Hap300 BeadArrays	Illumina genome studio	call rate <95%, MAF <1%, pHWE <10E-4	300,286	IMPUTE v1	HapMap release 22 (build 36)	none	PLINK, R, SNPTEST	1.00
EPIC-Norfolk cohort	Illumina 370CNV /OmniExpress	GenomeStudio	call rate <98%, MAF<1%, pHWE <10E-6	188,473	IMPUTE v1.0	HapMap release 22 (build 36)	none	PLINK, SNPTEST	1.01
ERF	Affymetrix 500K	BRLMM	call rate <90%, MAF<1%, pHWE <10E-6	382,037	IMPUTE v0.3.1	HapMap release 21 (build 35)	none	SAS, Stata, EIGENSTRAT, PLINK	1.03
Estonian Biobank	Illumina 6k, Illumina 318K, Illumina 370K and Affymetrix 250K	Illumina BeadStudio, BRLMM	call rate <98%, MAF <1%, pHWE <10E-6	450,877	MACH v1.0.16	HapMap release 22 CEU (Build 36)	none	ProbABEL, R, adjustment for family relatedness	1.02

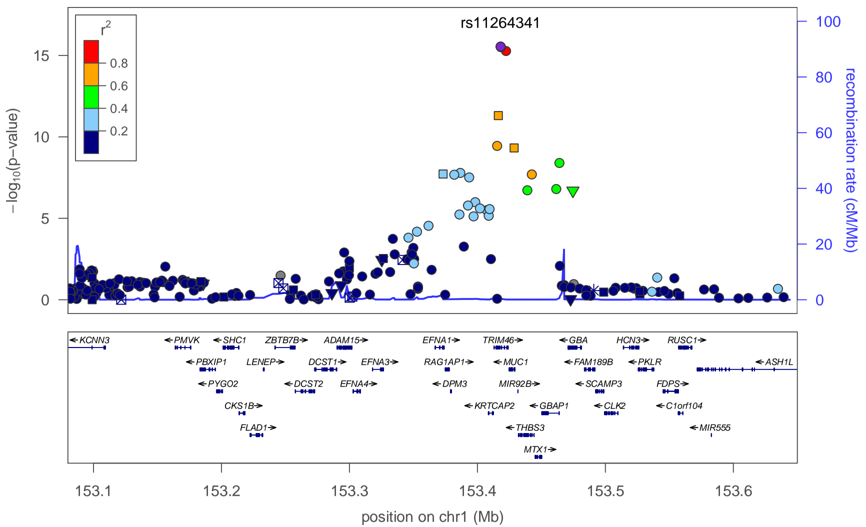
Family Heart Study (FamHS)	Illumina 550K, Illunina 610K, Illumina 1M	BeadStudio-gencall v3.0	MAF <1%, pHWE <1E-6	503,187	MACH v1.0.16	HapMap release 22 (build 36)	none	R	1.01
FHS	Affymetrix 500K, Affymetrix 50K supplemental	BRLMM	call rate <97%, pHWE <1E-6, Mishap p<1e-9, >100 Mendel errors, MAF<1%, strand mismatch with HapMap, not in HapMap	378,163	MACH v1.0.15	phased CEU haplotypes, HapMap release 22 (build 36)	none	kinship, GEE and GWAF packages in R gee() in GEE package in R	1.03
Health 2000	Illumina 610k	Illuminus	call rate <95%, MAF <1%, pHWE <10E-6	555,388	MACH v1.0.16	HapMap release 22 (build 36)	none	ProbABEL, PLINK, R	1.03
InCHIANTI	Illumina 550K	BeadStudio	call rate <98.5%, MAF <1%, pHWE <10E-4	484,115	MACH v1.0.16	HapMap release 22 (build 36)	none	ProbABEL, Merlin	1.02
INCIPE	Illumina,Infinium DNA Analysis Assay	Illuminus	call rate <97%, MAF <1%, pHWE <10E-5	635,654	IMPUTE v2	HapMap release 22 (build 36)	none	SNPTEST	1.00
INGI-Carlantino	Illumina 370CNV	BeadStudio	call rate <97%, MAF <1%, pHWE <10E-6	298,354	MACH	HapMap release 22 (build 36)	none	GenABEL, ProbABEL	1.00
INGI_CILENTO	370K Illumina	GenomeStudio (HumanCNV370-Quadv3_C.egt)	call rate <95%	299,456	MACH v1.0.16	HapMap release 22 (build 36)	none	GenABEL,ProbABEL,R	0.99
INGI-FVG	Illumina 370CNV	BeadStudio	call rate <97%, MAF <1%, pHWE <10E-6	306,478	MACH	HapMap release 22 (build 36)	none	GenABEL, ProbABEL	0.98
INGI-Val Borbera	Illumina SNP array 370K - HumanCNV370-Quadv3	BeadStudio	call rate <90%, MAF<1%, pHWE <10E-4	324,319	MACH	HapMap release 22 (build 36)	none	GenABEL, ProbABEL, R	0.99
KORA F3	Affymetrix 500K	BRLMM	individual call rate <93%	355,344	IMPUTE	HapMap release 21 (build 35)	none	SNPTEST v2.1.1	1.00
KORA F4	Affymetrix 6.0	Birdseed2	individual call rate <93%	909,622	IMPUTE	HapMap release 22 (build 36)	none	SNPTEST v2.1.1	1.01
LBC1936	Illumina Human 610_Quadv1	Illumina	call rate <98%, MAF <1%, pHWE <0.001	535,709	MACH v1.0.16	HapMap release 22 (build 36)	none	mach2qtl, R	0.99
LifeLines	Illumina CytoSNP12 V2	Illumina	call rate <95%, MAF<1%, pHWE <10E-5	257,581	BEAGLE v3.1.0	HapMap release 23a CEU	info<0.1, MAF<0.01	PLINK,Stata	1.02
LOLIPOP_EW_A	Affymetrix 500K	BRLMN	call rate <=90%, MAF <1%, pHWE <=10E-6	374,773	MACH	HapMap release 21 (build 35)	none	mach2qtl	1.00
LOLIPOP_EW_P	Perlegen custom	NA	call rate <=90%, MAF <1%, pHWE <=10E-6	184,469	MACH	HapMap release 21 (build 35)	none	mach2qtl	1.01
LOLIPOP_EW610	Illumina Human 610	BeadStudio	call rate <=90%, MAF<1%, pHWE <=10E-6	544,620	MACH	HapMap release 22 (build 36)	none	mach2qtl	0.99
LURIC	Affymetrix 6.0	Birdseed	individual call rate <96%	866,316	IMPUTE v0.4.2	HapMap release 22 (build 36)	none	SNPTEST v2.1.1	1.00
MICROS	HumanHap 300v2	BeadStudio	call rate <98%, MAF <1%, pHWE <10E-6	290,356	MACH v1	HapMap release 22 (build 36)	none	ProbABEL,mmscore argument	1.00
NESDA	Perlegen 600K	Perlegen	call rate <95%, MAF 1%, 5% genotype mismatches, 5% Mendelian errors	435,291	IMPUTE v0.3.2	HapMap release 22 (build 36)	Valid p-value	SNPTEST v2.1.1, R	1.02

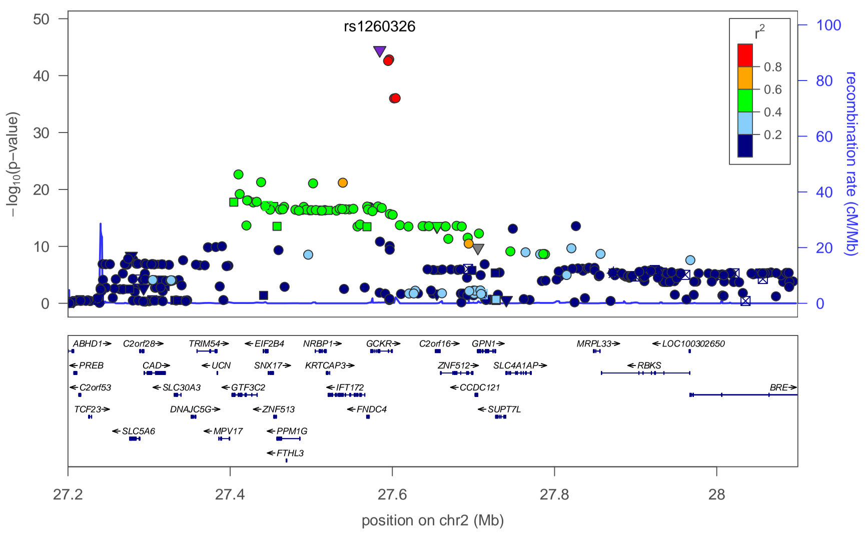
NSPHS	HumanHap 300v2	BeadStudio	call rate <98%, MAF<1%, pHWE <10E-6	292,220	MACH v1.0.15	HapMap release 22 (build 36)	none	ProbABEL,mmscore argument	1.00
ORCADES	most HumanHap 300v2, some 370CNV-Quad	BeadStudio	call rate <98%, MAF<1%, pHWE <10E-6	293,607	MACH v1	HapMap release 22 (build 36)	none	ProbABEL,mmscore argument	1.00
PREVEND	Illumina CytoSNP12 V2	Illumina	call rate <98%, MAF<1%, pHWE <10E-5	244,868	BEAGLE v3.1.0	HapMap release 23a CEU	info<0.1, MAF<0.01	PLINK,Stata	1.05
PROCARDIS	Illumina 610 & Illumina 1M	GenCall (BeadStudio)	call rate <95%, MAF<1%, pHWE <1E-6	487,783	MACH v1.0.16	HapMap release 22 (build 36)	none	STATA 10	1.03
RS-I	Version 3 Illumina Infinium II HumanHap550	BeadStudio	call rate <98%, MAF<1%, pHWE <10E-5	530,683	MACH v1.0.15	HapMap release 22 (build 36)	none	ProbABEL	1.02
RS-II	Version 3 Illumina Infinium II HumanHap550	BeadStudio	call rate <98%, MAF<1%, pHWE <10E-5	495,478	MACH v1.0.15	HapMap release 22 (build 36)	none	ProbABEL	1.00
SardiNIA	Affymetrix 10K, 500K, 6.0	BRLMM (10K/500K), Birdseed (6.0)	call rate <90% (10K/500K) and <95% (6.0), MAF<5% (10K/500K) and <1% (6.0), pHWE<10E-6	731,209	MACH v1.0.10	HapMap release 22 (build 36)	rsqr<0.3, MAF <1%, Excess Mendelian Errors	Merlin (fastAssoc), R	1.05
SHIP	Affymetrix 6.0	Birdseed v2	none	869,224	IMPUTE v0.5.0	HapMap release 22 (build 36)	none	Caché, InforSense, R, QUICKTEST	1.03
SOCCS	HumanHap 300v1 and 240S	BeadStudio	call rate <98%, MAF<1%, pHWE <10E-6	512,938	MACH v1	HapMap release 22 (build 36)	none	ProbABEL	1.01
Sorbs	Affymetrix 500K Affymetrix 6.0	BRLMM Birdseed	call rate <95%, MAF<1%, pHWE <10E-4	378,513	IMPUTE v1.0.0	HapMap release 21 (build 35)	none	GenABEL, ProbABEL	1.02
TwinsUK	Illumina 317K+610K+1M	Illumina protocol	call rate <95%, MAF<1%, pHWE <10E-4	NA	IMPUTE v2	HapMap release 24 (build 36)	none	GenABEL	1.00
Young Finns Study	Illumina 670k	Illuminus	call rate <95%, MAF<1%, pHWE <10E-6	546,677	MACH v1.0.16	HapMap release 22 (build 36)	none	ProbABEL, PLINK, R	1.02
In silico replication studies									
EPIC cases	Illumina 370CNV /OmniExpress	GenomeStudio	call rate <98%, MAF<1%, pHWE <10E-6	188,473	IMPUTE v1.0	HapMap release 22 (build 36)	none	PLINK, SNPTEST	
GHS I	Affymetrix 6.0	Birdseed	call rate <95%, MAF<1%, pHWE<10E-4	662,405	IMPUTE v2.1.0	HapMap release 24 (build 36)	none	MetABEL, R	
GHS II	Affymetrix 6.0	Birdseed	call rate <95%, MAF<1%, pHWE<10E-4	673,914	IMPUTE v2.1.0	HapMap release 24 (build 36)	none	MetABEL, R	
GSK cases/controls	Illumina 550K	BeadStudio	call rate <98%, pHWE <10E-5	517,946	IMPUTE v2	HapMap3 release #2 (Feb 2009) and 1000g data freeze Mar2010	none	PLINK, R, GTOOL	

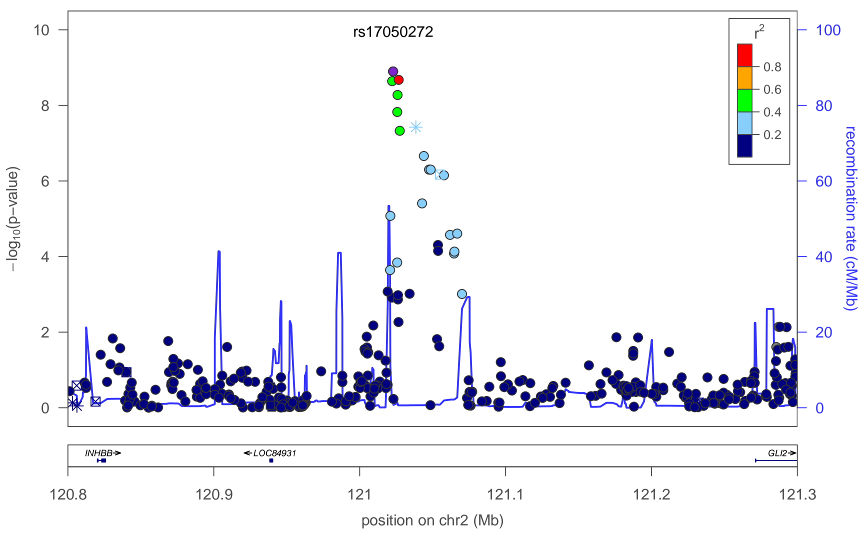
HPFS	Illumina Infinium Omni Express	BeadStudio	MAF<1%, pHWE<10E-4, genotyping rate <97%	553,716	MACH v1.0.16	HapMap release 22 (build 36)	none	ProbABEL, PLINK, R	
Hunter Community Study	Illumina 610K-Quad	Illumina	call rate <95%, MAF<1%, pHWE <10E-6	513,977	MACH v1.0.16	HapMap release 24 (build 36.1)	MAF<0.01, oevar_imp<0.3	PLINK, SAS,R	
Lifelines replication	Illumina CytoSNP12 V2	Illumina	call rate <95%, MAF<1%, pHWE <10E-5	257,581	BEAGLE v3.1.0	HapMap rel 22 CEU	info<0.1, MAF<0.01	PLINK,Stata	
LURIC_GZ	Affy 500k	BRLMM	call rate <95%, pHWE <10E-6	393,157	MACH	1000Genomes	none	SPSS, PLINK	
LURIC_HD	Affy 6.0	Birdseed v2	call rate <95%, pHWE <10E-6	893,909	MACH	1000Genomes	none	SPSS, PLINK	
MARS cases	Illumina 100k,300k,610k	BeadStudio	chipwise: call rate <98%, pHWE < 10E-05, over all chips: call rate <98%, MAF < 1%, pHWE < 10E-05	327,336	IMPUTE v2	HapMap3 release #2 (Feb 2009) and 1000g data freeze Mar2010	none	PLINK, R, GTOOL	
NHS	Illumina Infinium Omni Express	BeadStudio	MAF<1%, pHWE<10E-4,genotyping rate <97%	553,716	MACH v1.0.16	HapMap release 22 (build 36)	none	ProbABEL, PLINK, R	
OGP-Talana	Affymetrix 500k	BRLMM	call rate <93%, MAF <1%, pHWE <10E-5	329,122	MACH v1.0.16	HapMap release 22 (build 36)	Rsq<0.3	R, GenABEL, ProbABEL	
SAPALDIA	Illumina Human610-Quad BeadChip	Gencall	only autosomal (excl. sex-chromosomal, mitochondrial)	567,589	MACH v.1.0.16	HapMap release 22 (build 36)	none	ProbABEL, PLINK, Stata	
SHIP-Trend	Illumina Omni 2.5	GenomeStudio Genotyping Module v1.0	pHWE <= 0.0001, call rate <=90%, monomorphic SNPs	1,782,967	IMPUTE v2.1.2.3	HapMap release 22 (build 36)	duplicate RSID but different positions	Caché, InforSense, R, QUICKTEST	
Study samples of non-European ancestry									
LOLIPOP_ IA317	Illumina HumanHap300K	BeadStudio	call rate <= 90%, MAF < 1%, pHWE <= 10E-6	245,892	MACH	HapMap release 21 (build 35)	none	mach2qtl	
LOLIPOP_IA610	Illumina Human610	BeadStudio	call rate <= 90%, MAF < 1%, pHWE <= 10E-6	544,390	MACH	HapMap release 21 (build 35)	none	mach2qtl	
LOLIPOP_IA_P	Perlegen custom	Perlegen	call rate <= 90%, MAF < 1%, pHWE <= 10E-6	170,055	MACH	HapMap release 21 (build 35)	none	mach2qtl	
CARe Consortium (ARIC, CARDIA, JHS)	Affymetrix 6.0	Birdseed v1.33	all chip QC + pi_hat 0,05 for rate step	763,537 to 846,628	MACH, 2 rounds	combined CEU+YRI reference panel	MAF 1%, rsq_hat 0.3	plink −-dosage	
Japanese Biobank	Illumina HumanHap610-Quad Genotyping BeadChip	BEADSTUDIO-Genotyping Module v3.3.7	call rate <99%, MAF<1%, pHWE <10E-7	477,784	MACH v1.0.10	HapMap Phase II JPT+CHB individuals (release 24, build 36)	none	R v2.11.0	

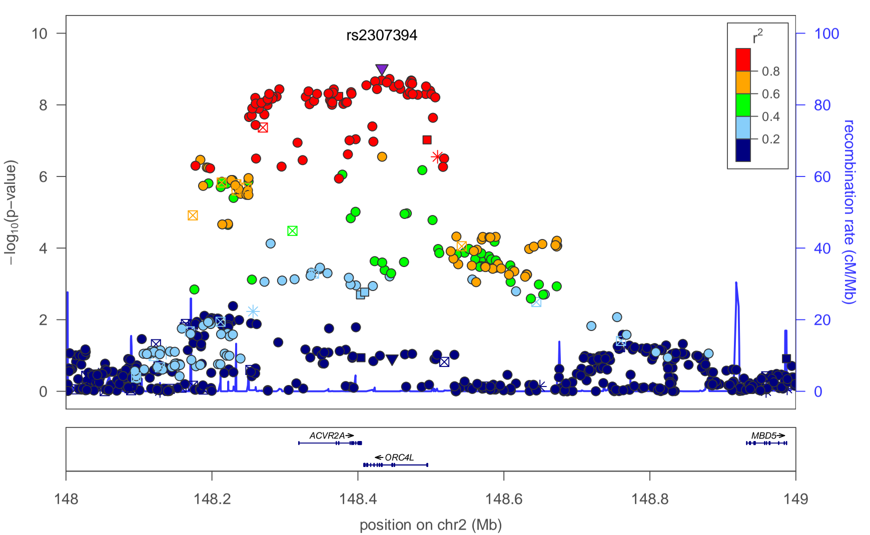
Figure S1: **Regional association plots of 44 GUGC loci** showing $-log_{10}$ p-values for all SNPs ordered by their chromosomal position within all regions reaching p-values $< 1{\times}10^{-6}$ in the discovery screen of the overall or sex-stratified serum urate GWAS as well as the candidate urate transporter gene region. For sex-specific loci, $-log_{10}$ p-value correspond to the respective sex-stratified urate GWAS. Each SNP is coloured according to its correlation with the index SNP within the region as specified in the colour scheme. Correlation structures correspond to HapMap II CEU r28. Gray colour indicates unknown correlation. Data point symbols correspond to nonsense, non-synonymous, coding, UTR, splice variants, transcription factor binding sites and multi-species conservation according to dbSNP or the 1000 Genomes Project (August 2009 release). Positions are given for NCBI build 36. Plots are taken from Köttgen *et al.*[45].

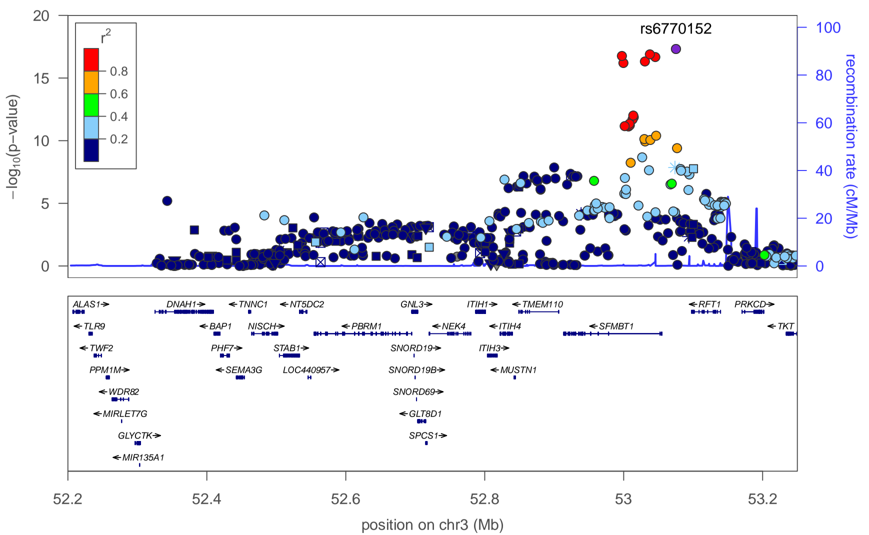

rs1471633
−log10(p−value)
recombination rate (cM/Mb)
position on chr1 (Mb)

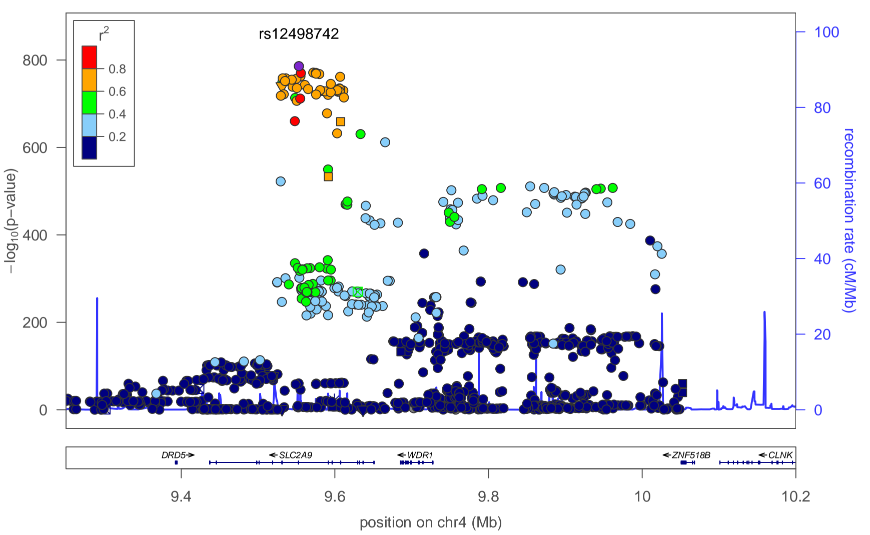

rs11264341
−log10(p−value)
recombination rate (cM/Mb)
position on chr1 (Mb)

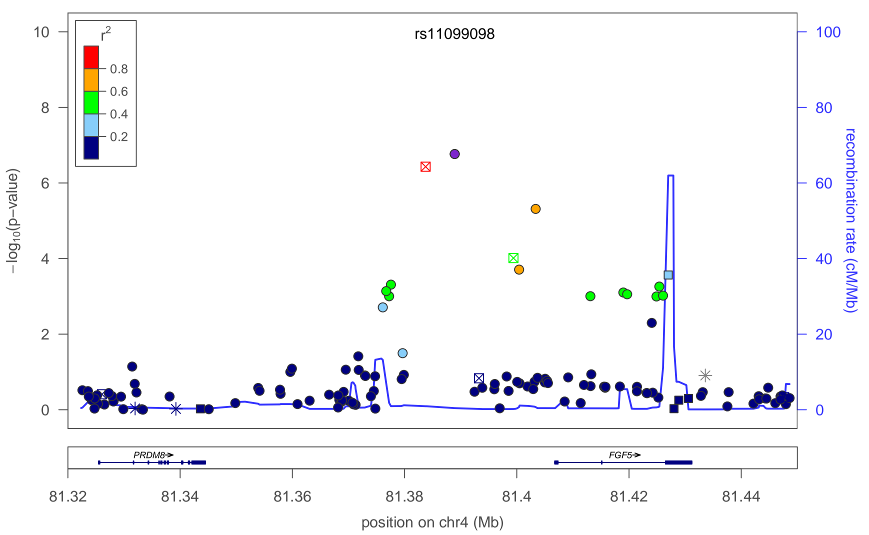

rs1260326
−log10(p−value)
recombination rate (cM/Mb)
position on chr2 (Mb)

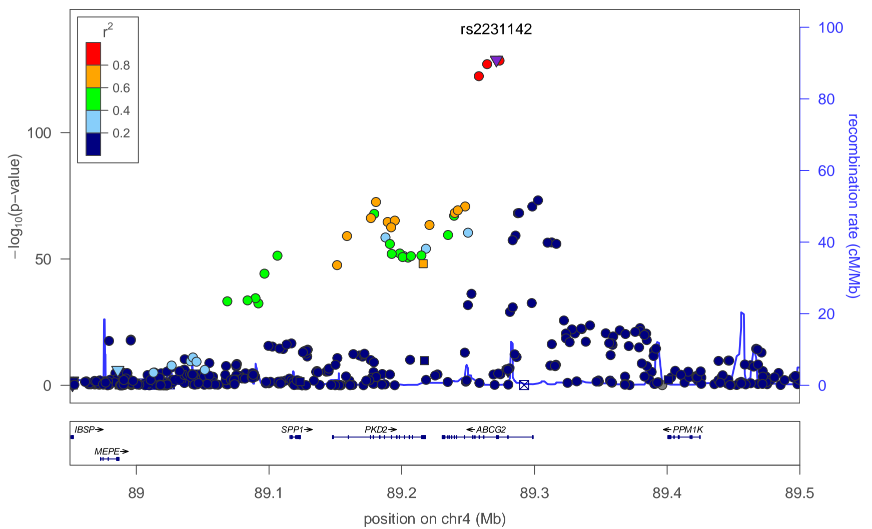

rs17050272
−log10(p−value)
recombination rate (cM/Mb)
position on chr2 (Mb)

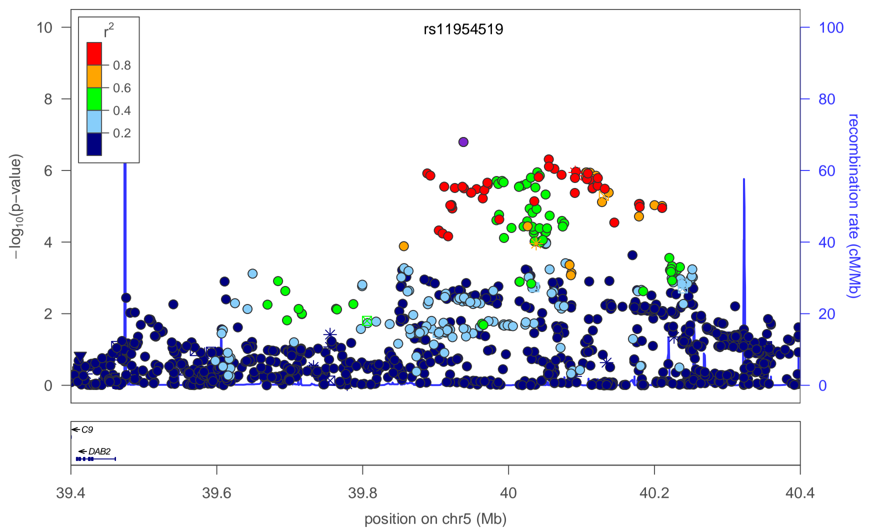

rs2307394
recombination rate (cM/Mb)
−log10(p−value)
position on chr2 (Mb)

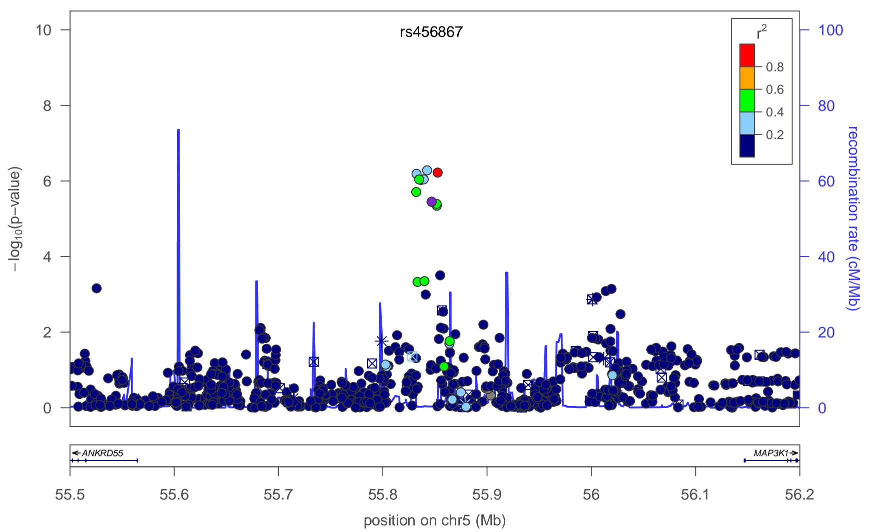

rs6770152
recombination rate (cM/Mb)
−log10(p−value)
position on chr3 (Mb)

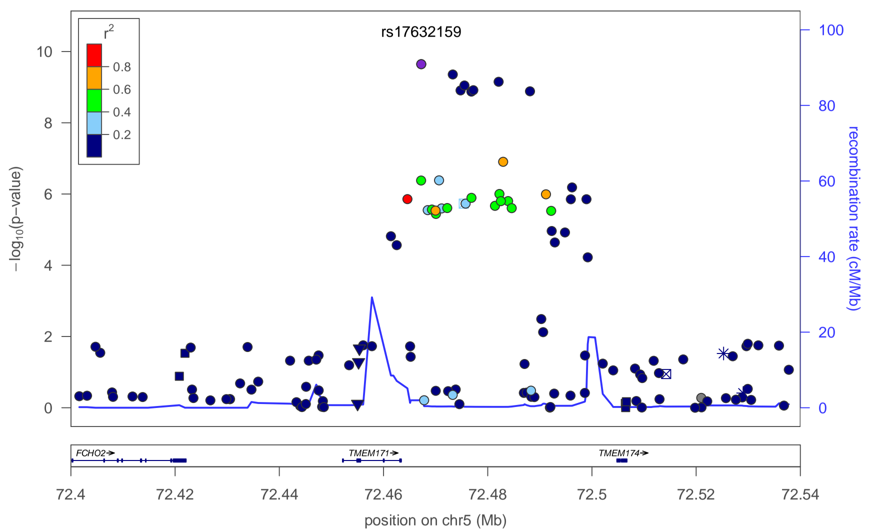

rs12498742
recombination rate (cM/Mb)
−log10(p−value)
position on chr4 (Mb)

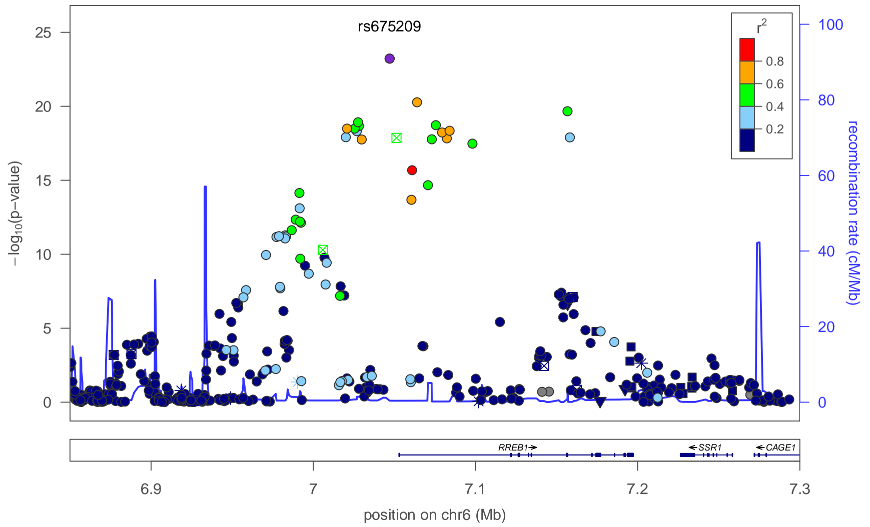

rs11099098
recombination rate (cM/Mb)
−log10(p−value)
position on chr4 (Mb)


rs2231142
r2
0.8
0.6
0.4
0.2
−log10(p−value)
recombination rate (cM/Mb)
IBSP
MEPE
SPP1
PKD2
ABCG2
PPM1K
89
89.1
89.2
89.3
89.4
89.5
position on chr4 (Mb)

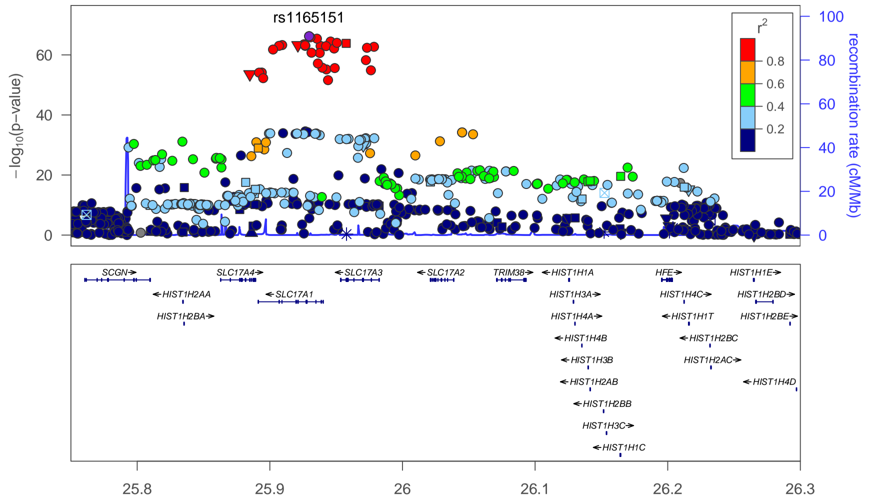

rs11954519
r2
0.8
0.6
0.4
0.2
−log10(p−value)
recombination rate (cM/Mb)
C9
DAB2
39.4
39.6
39.8
40
40.2
40.4
position on chr5 (Mb)


rs456867
r2
0.8
0.6
0.4
0.2
−log10(p−value)
recombination rate (cM/Mb)
ANKRD55
MAP3K1
55.5
55.6
55.7
55.8
55.9
56
56.1
56.2
position on chr5 (Mb)

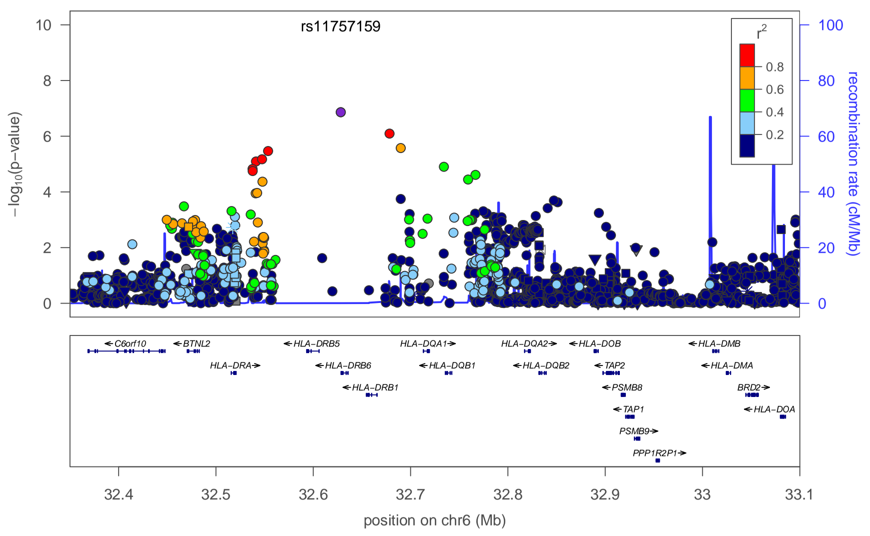

rs17632159
r2
0.8
0.6
0.4
0.2
−log10(p−value)
recombination rate (cM/Mb)
FCHO2
TMEM171
TMEM174
72.4
72.42
72.44
72.46
72.48
72.5
72.52
72.54
position on chr5 (Mb)



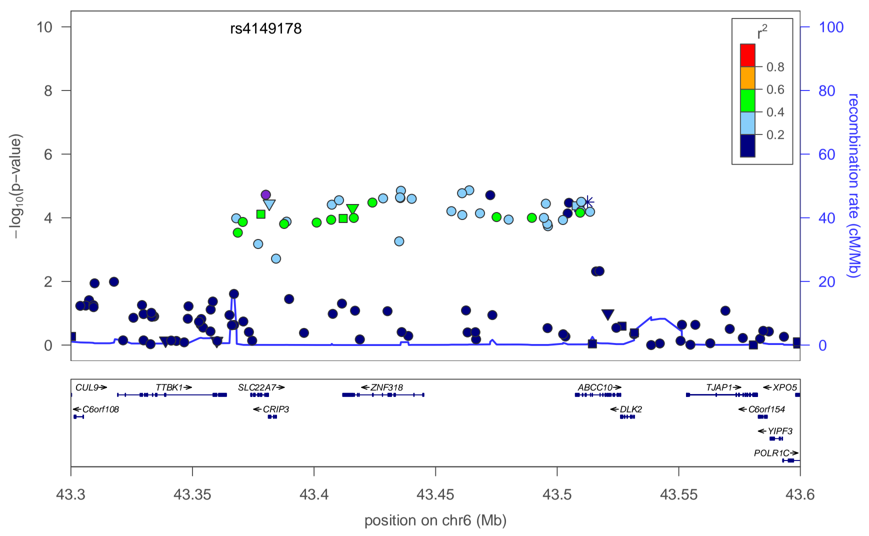
rs675209
−log₁₀(p−value)
25
20
15
10
5
0
r²
0.8
0.6
0.4
0.2
recombination rate (cM/Mb)
100
80
60
40
20
0
6.9
7
7.1
7.2
7.3
position on chr6 (Mb)



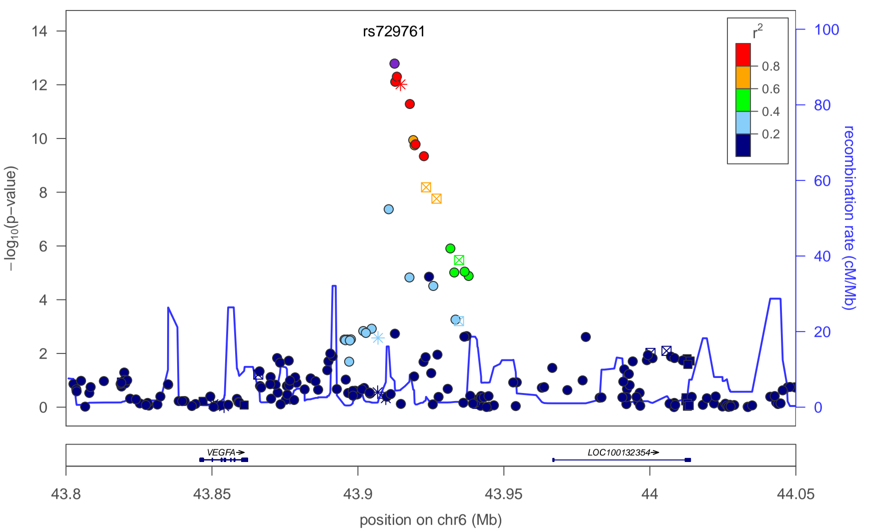
rs1165151
−log₁₀(p−value)
60
40
20
0
r²
0.8
0.6
0.4
0.2
recombination rate (cM/Mb)
100
80
60
40
20
0
25.8
25.9
26
26.1
26.2
26.3
position on chr6 (Mb)

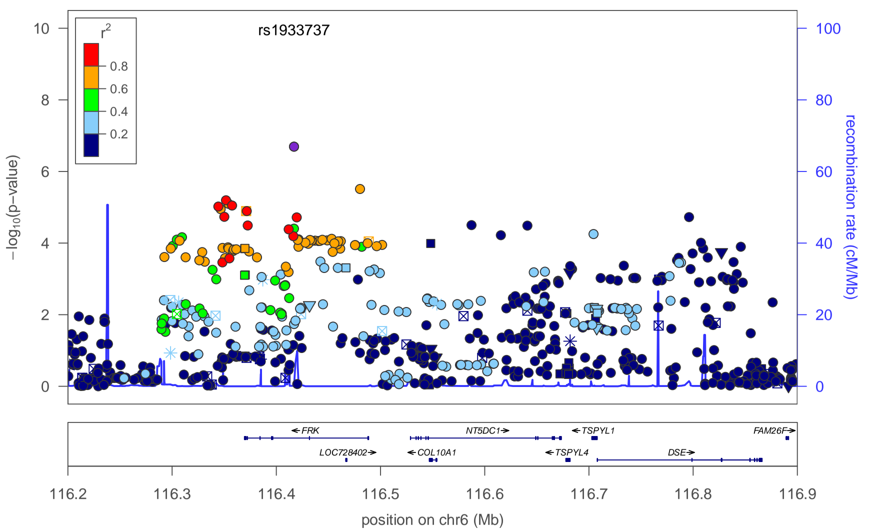


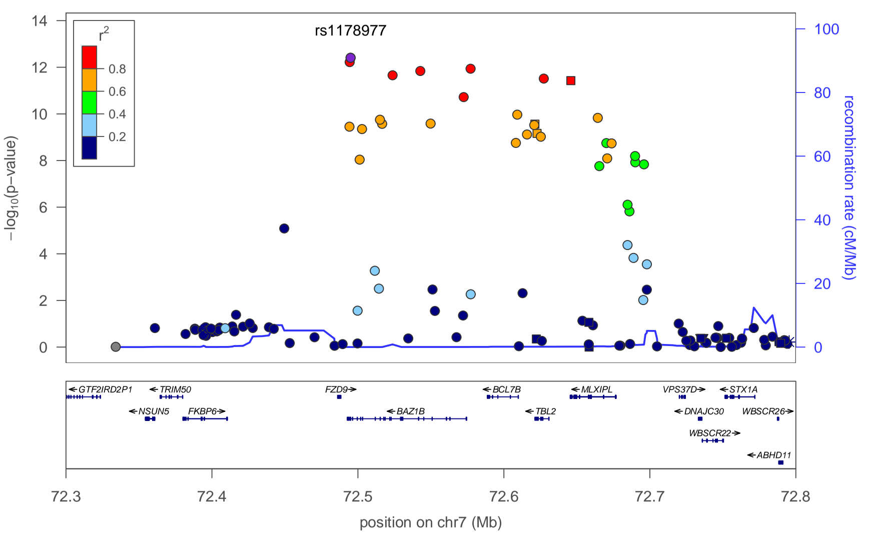
rs11757159
−log₁₀(p−value)
10
8
6
4
2
0
r²
0.8
0.6
0.4
0.2
recombination rate (cM/Mb)
100
80
60
40
20
0
32.4
32.5
32.6
32.7
32.8
32.9
33
33.1
position on chr6 (Mb)

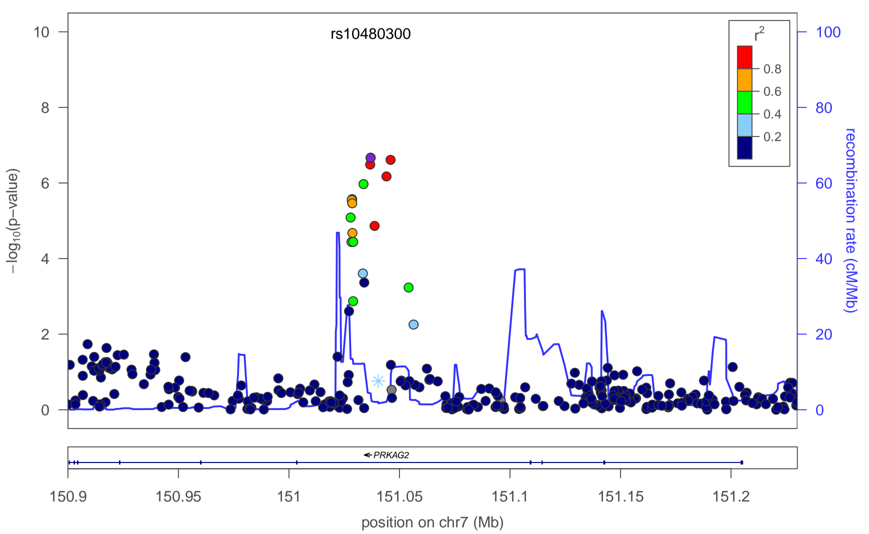


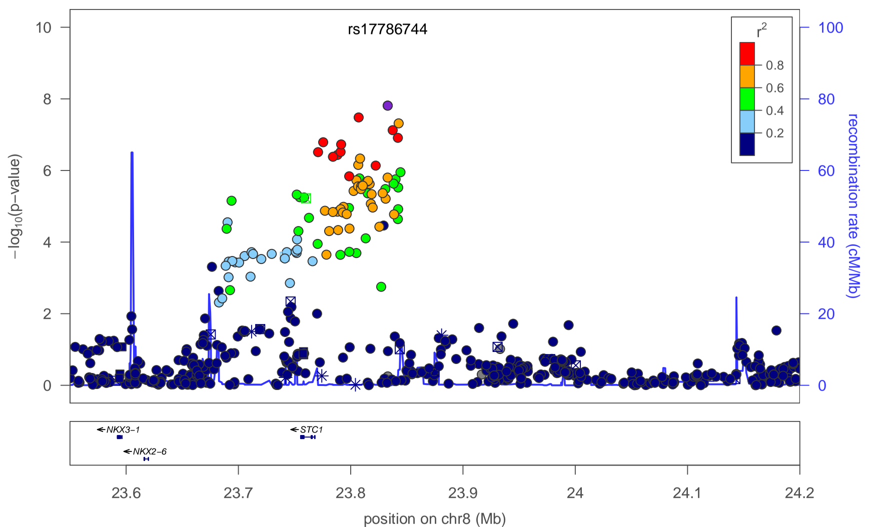
rs4149178
−log₁₀(p−value)
10
8
6
4
2
0
r²
0.8
0.6
0.4
0.2
recombination rate (cM/Mb)
100
80
60
40
20
0
43.3
43.35
43.4
43.45
43.5
43.55
43.6
position on chr6 (Mb)

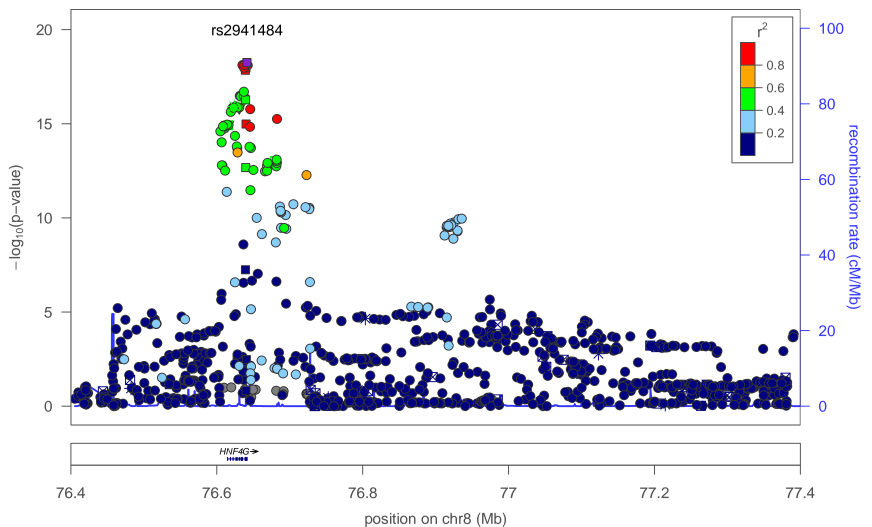

rs729761
r2
0.8
0.6
0.4
0.2
−log10(p−value)
recombination rate (cM/Mb)
VEGFA
LOC100132354
position on chr6 (Mb)

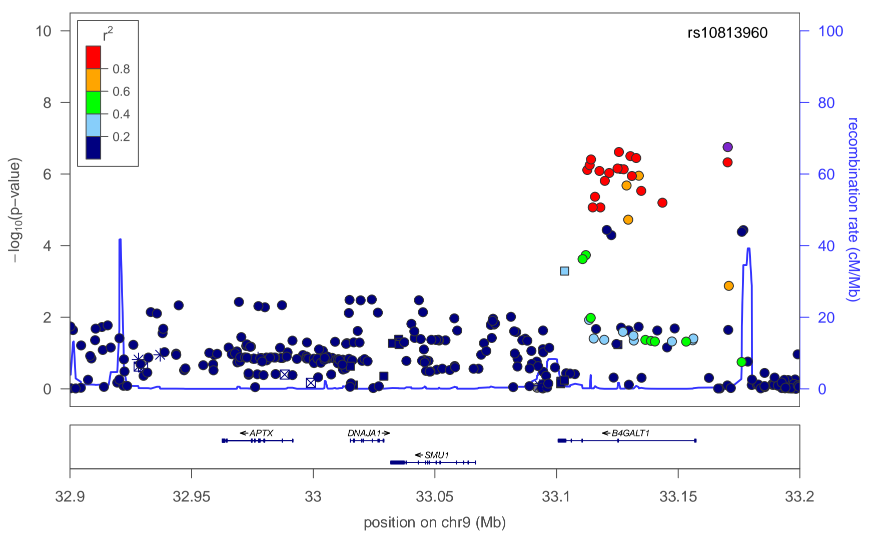

rs1933737
r2
0.8
0.6
0.4
0.2
−log10(p−value)
recombination rate (cM/Mb)
FRK
LOC728402
NT5DC1
COL10A1
TSPYL1
TSPYL4
DSE
FAM26F
position on chr6 (Mb)

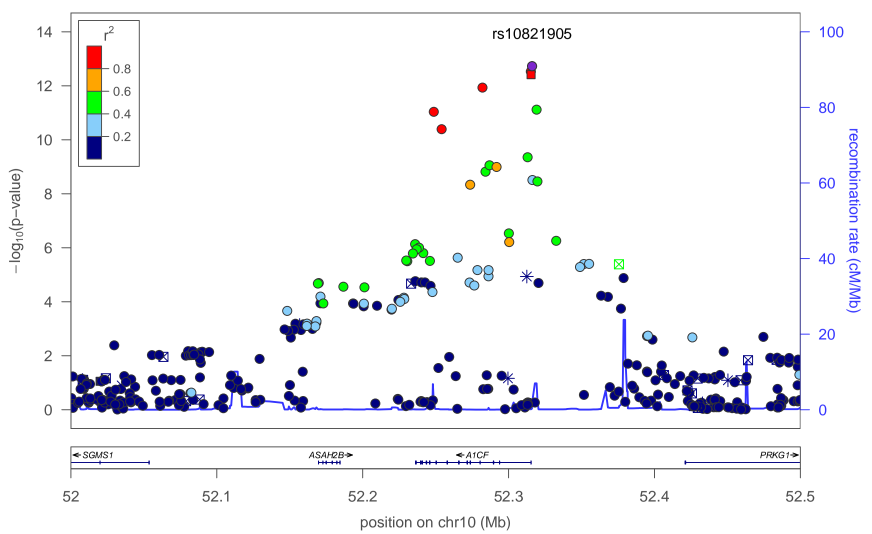

rs1178977
r2
0.8
0.6
0.4
0.2
−log10(p−value)
recombination rate (cM/Mb)
GTF2IRD2P1
TRIM50
NSUN5
FKBP6
FZD9
BAZ1B
BCL7B
TBL2
MLXIPL
VPS37D
DNAJC30
WBSCR22
STX1A
WBSCR26
ABHD11
position on chr7 (Mb)

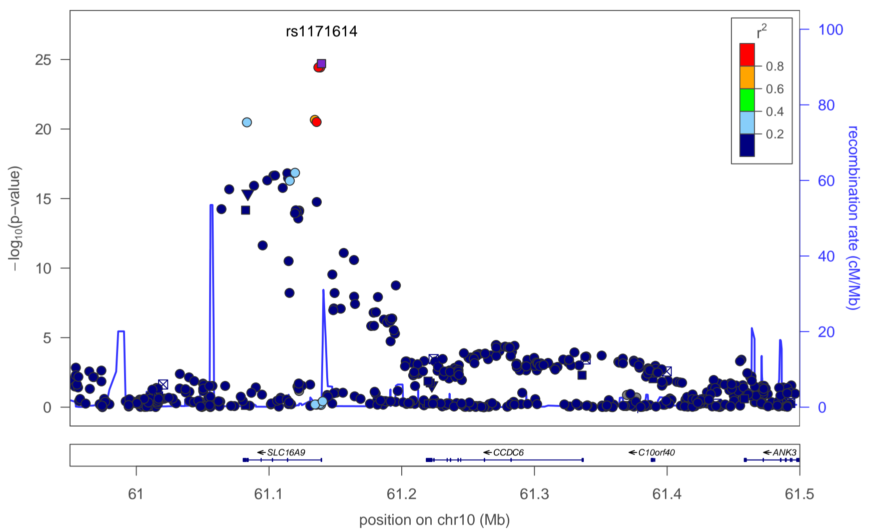

rs10480300
r2
0.8
0.6
0.4
0.2
−log10(p−value)
recombination rate (cM/Mb)
PRKAG2
position on chr7 (Mb)

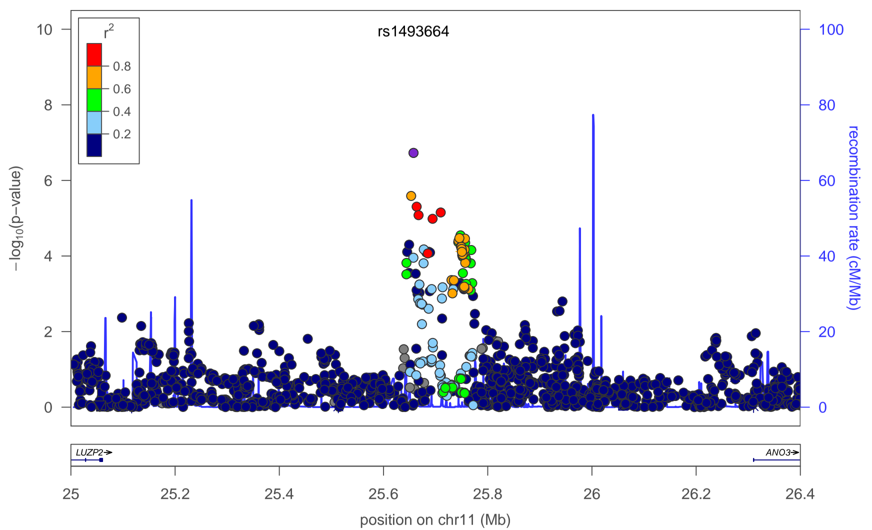

rs17786744
−log10(p−value)
recombination rate (cM/Mb)
position on chr8 (Mb)
NKX3-1
NKX2-6
STC1
r2
0.8
0.6
0.4
0.2

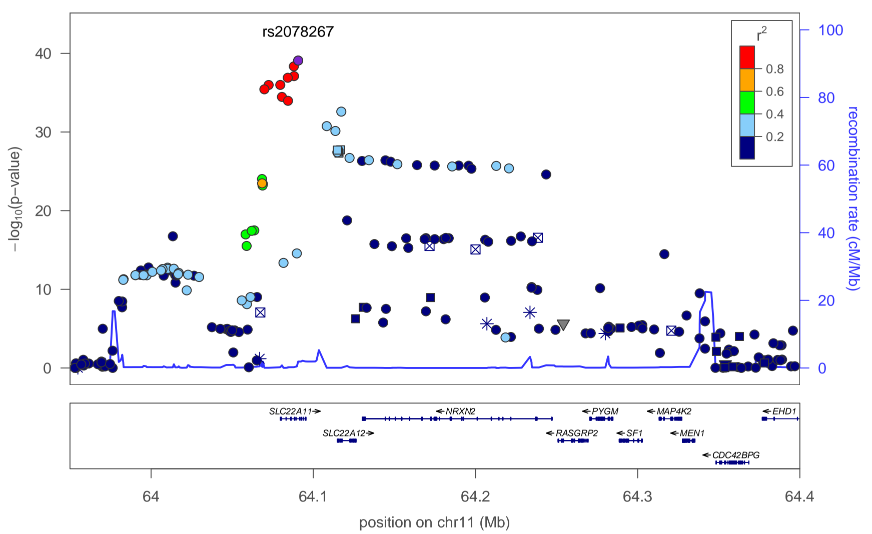

rs2941484
−log10(p−value)
recombination rate (cM/Mb)
position on chr8 (Mb)
HNF4G
r2
0.8
0.6
0.4
0.2

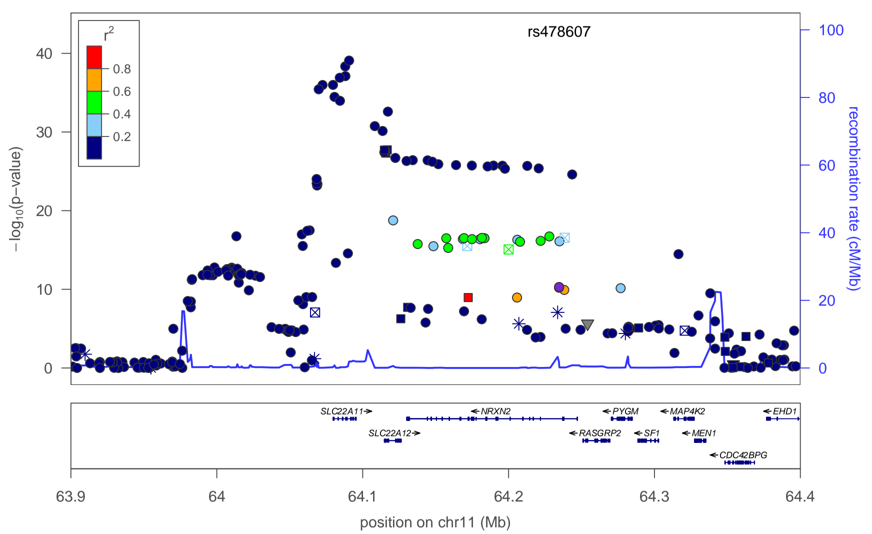

rs10813960
−log10(p−value)
recombination rate (cM/Mb)
position on chr9 (Mb)
APTX
DNAJA1
SMU1
B4GALT1
r2
0.8
0.6
0.4
0.2

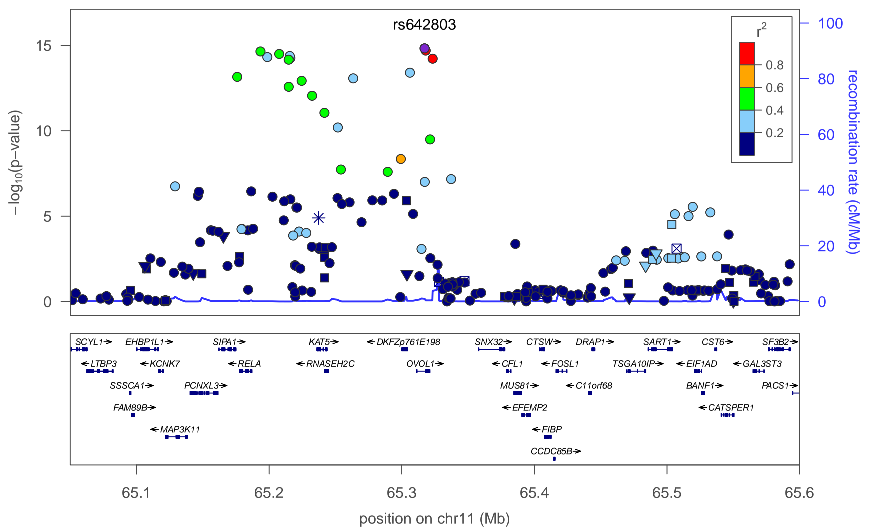

rs10821905
−log10(p−value)
recombination rate (cM/Mb)
position on chr10 (Mb)
SGMS1
ASAH2B
A1CF
PRKG1
r2
0.8
0.6
0.4
0.2

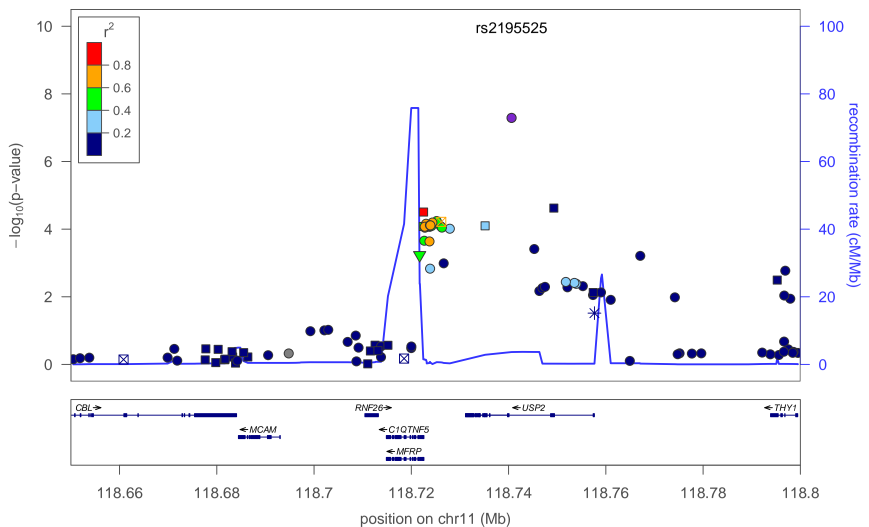

rs1171614
r2
0.8
0.6
0.4
0.2
−log10(p−value)
recombination rate (cM/Mb)
SLC16A9
CCDC6
C10orf40
ANK3
position on chr10 (Mb)

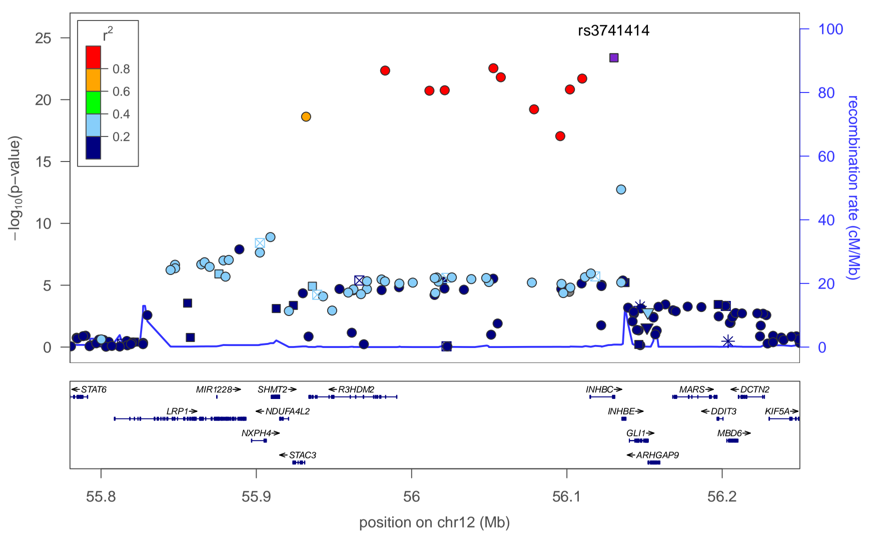

rs1493664
r2
0.8
0.6
0.4
0.2
−log10(p−value)
recombination rate (cM/Mb)
LUZP2
ANO3
position on chr11 (Mb)

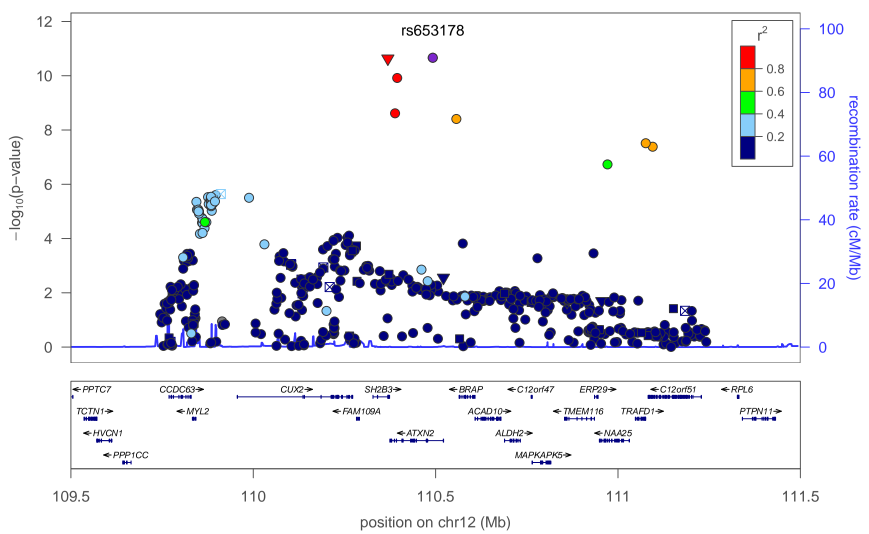

rs2078267
r2
0.8
0.6
0.4
0.2
−log10(p−value)
recombination rate (cM/Mb)
SLC22A11
SLC22A12
NRXN2
RASGRP2
PYGM
SF1
MAP4K2
MEN1
CDC42BPG
EHD1
position on chr11 (Mb)

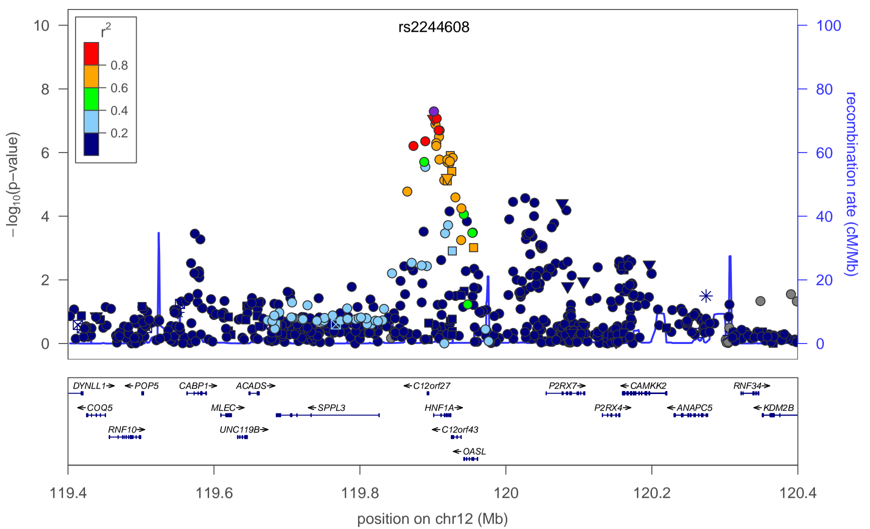

rs478607
r2
0.8
0.6
0.4
0.2
−log10(p−value)
recombination rate (cM/Mb)
SLC22A11
SLC22A12
NRXN2
RASGRP2
PYGM
SF1
MAP4K2
MEN1
CDC42BPG
EHD1
position on chr11 (Mb)

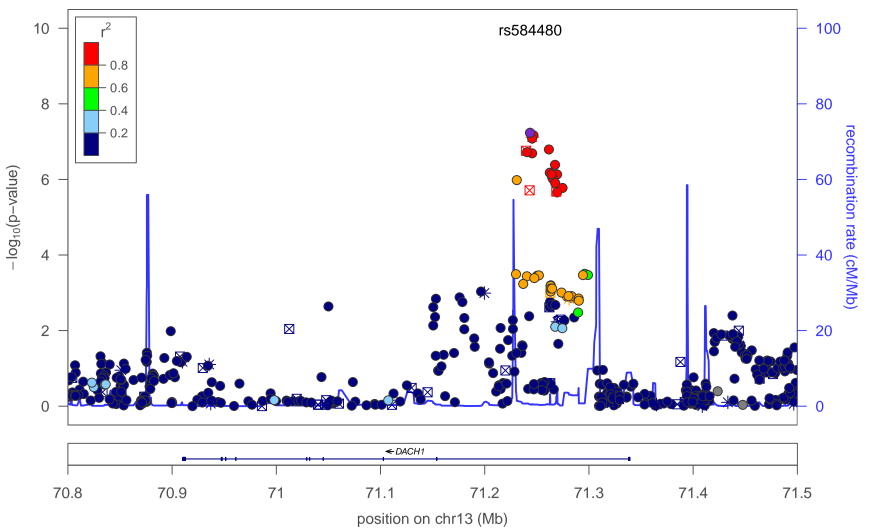

rs642803
−log10(p−value)
recombination rate (cM/Mb)
position on chr11 (Mb)

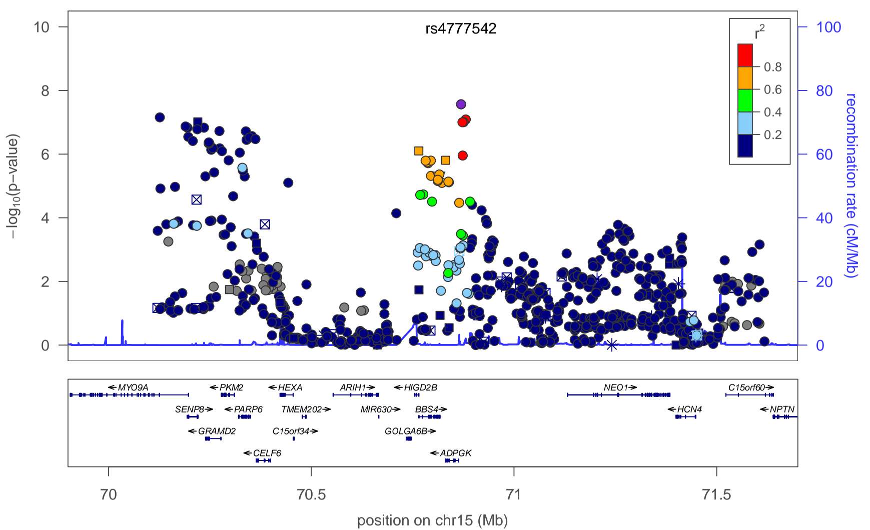

rs2195525
−log10(p−value)
recombination rate (cM/Mb)
position on chr11 (Mb)

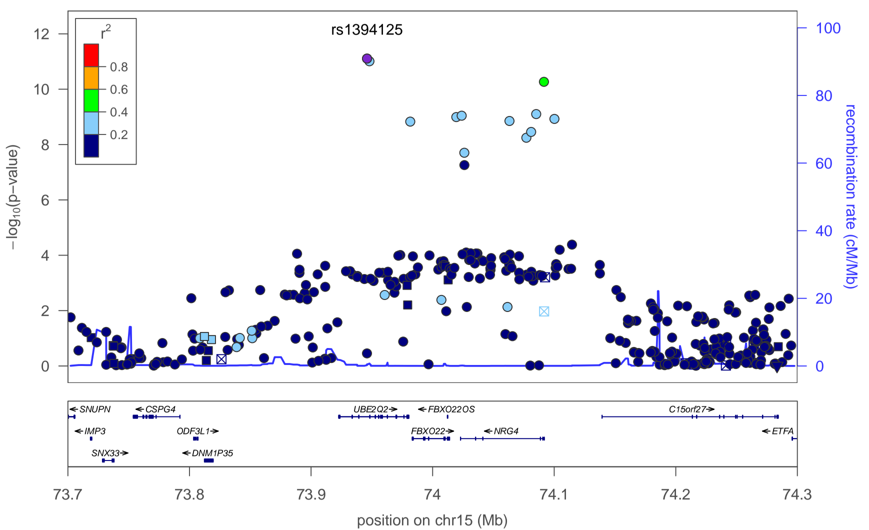

rs3741414
−log10(p−value)
recombination rate (cM/Mb)
position on chr12 (Mb)

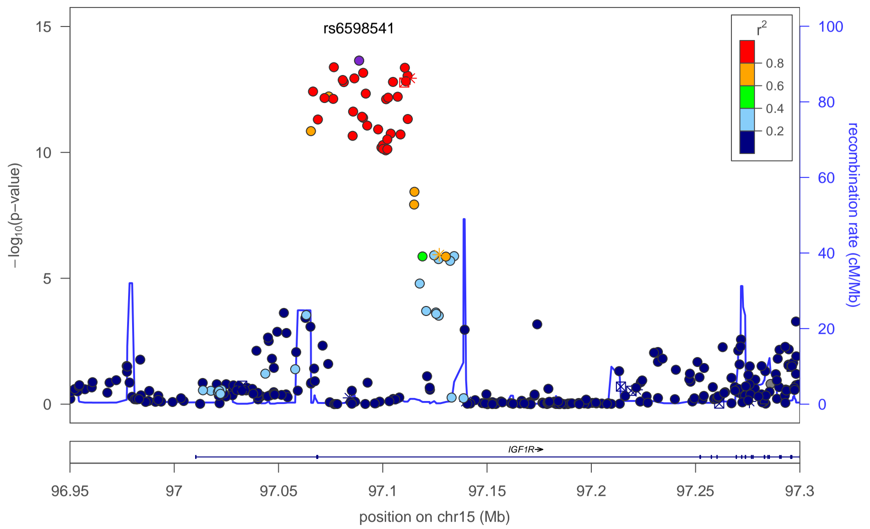

rs653178
−log10(p−value)
recombination rate (cM/Mb)
position on chr12 (Mb)

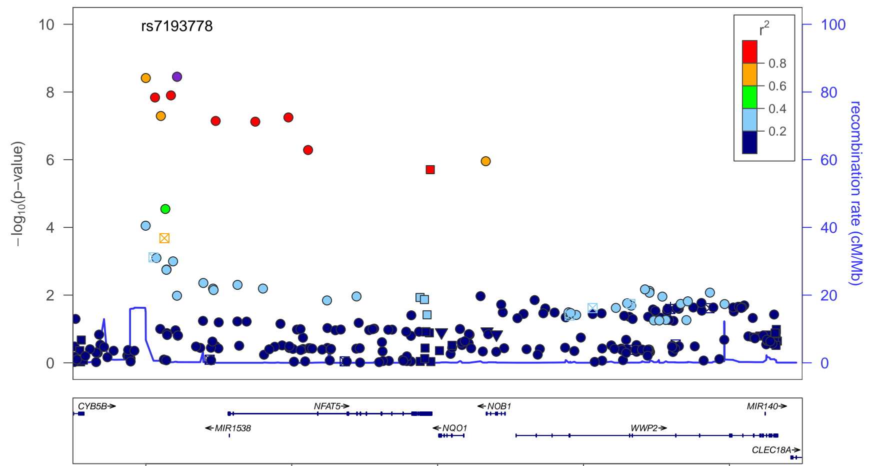

rs2244608
−log10(p−value)
recombination rate (cM/Mb)
r2
0.8
0.6
0.4
0.2
DYNLL1
POP5
CABP1
ACADS
C12orf27
P2RX7
CAMKK2
RNF34
COQ5
MLEC
SPPL3
HNF1A
P2RX4
ANAPC5
KDM2B
RNF10
UNC119B
C12orf43
OASL
119.4
119.6
119.8
120
120.2
120.4
position on chr12 (Mb)

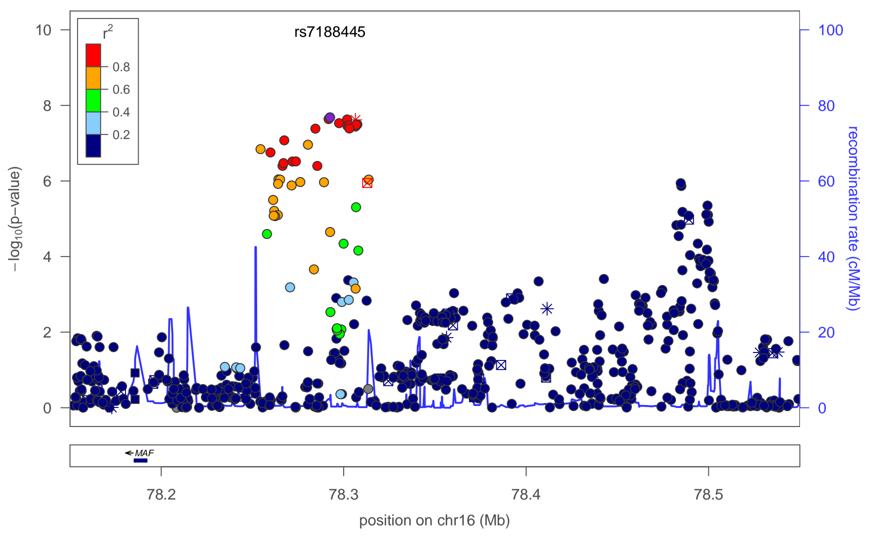

rs584480
−log10(p−value)
recombination rate (cM/Mb)
r2
0.8
0.6
0.4
0.2
DACH1
70.8
70.9
71
71.1
71.2
71.3
71.4
71.5
position on chr13 (Mb)

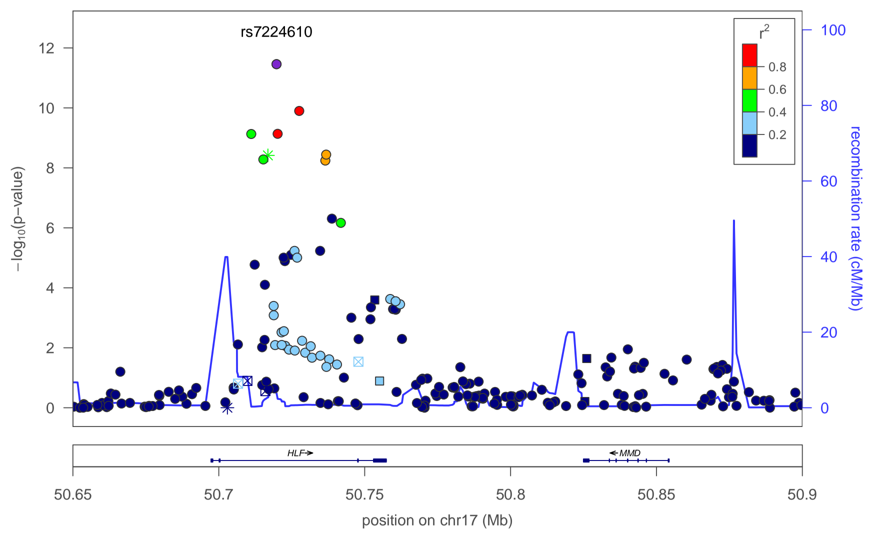

rs4777542
−log10(p−value)
recombination rate (cM/Mb)
r2
0.8
0.6
0.4
0.2
MYO9A
PKM2
HEXA
ARIH1
HIGD2B
NEO1
C15orf60
SENP8
PARP6
TMEM202
MIR630
BBS4
HCN4
NPTN
GRAMD2
C15orf34
GOLGA6B
CELF6
ADPGK
70
70.5
71
71.5
position on chr15 (Mb)

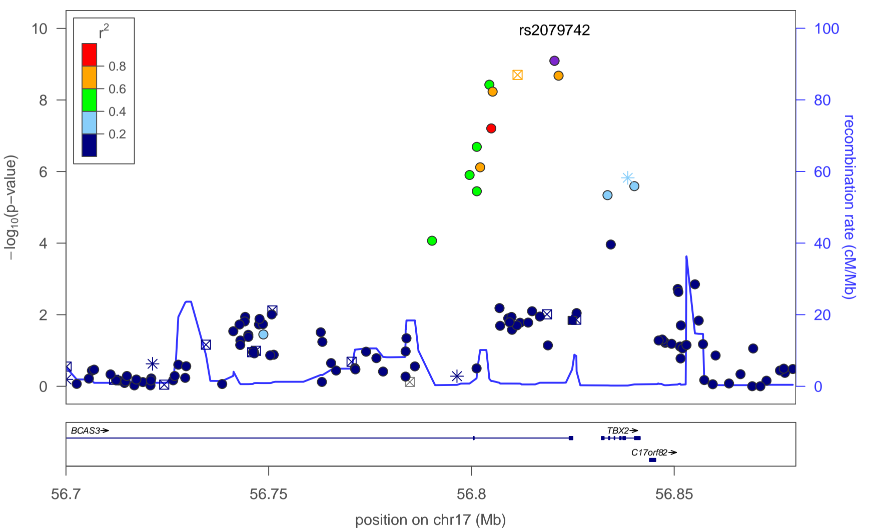

rs1394125
−log10(p−value)
recombination rate (cM/Mb)
r2
0.8
0.6
0.4
0.2
SNUPN
CSPG4
UBE2Q2
FBXO22OS
C15orf27
IMP3
ODF3L1
FBXO22
NRG4
ETFA
SNX33
DNM1P35
73.7
73.8
73.9
74
74.1
74.2
74.3
position on chr15 (Mb)

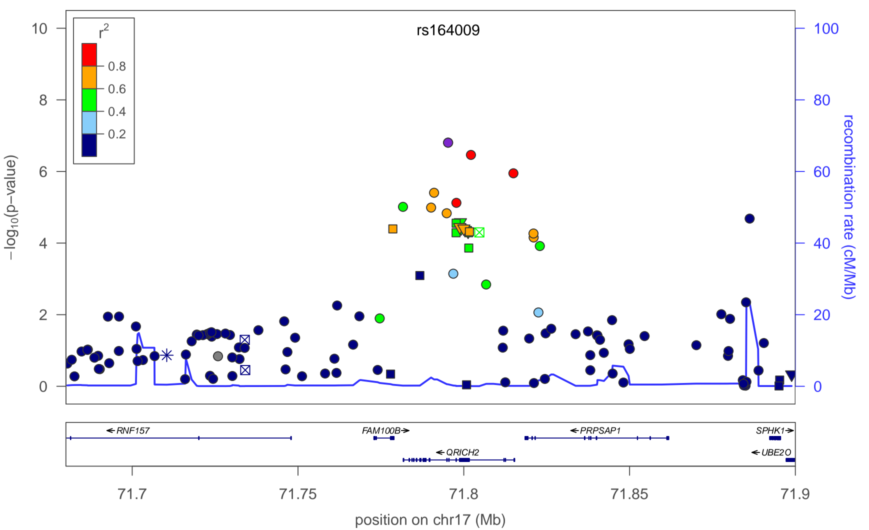

rs6598541
recombination rate (cM/Mb)
−log10(p−value)
IGF1R
position on chr15 (Mb)

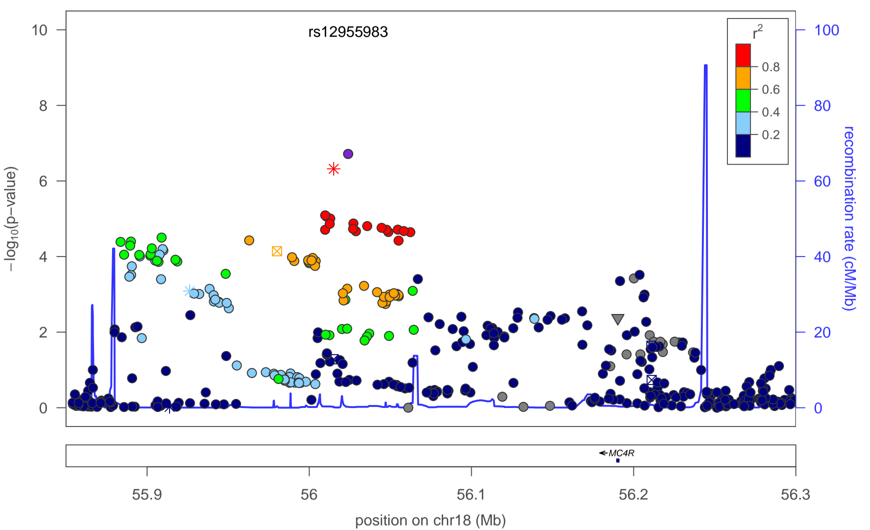

rs7193778
recombination rate (cM/Mb)
−log10(p−value)
CYB5B
NFAT5
NOB1
MIR140
MIR1538
NQO1
WWP2
CLEC18A
position on chr16 (Mb)

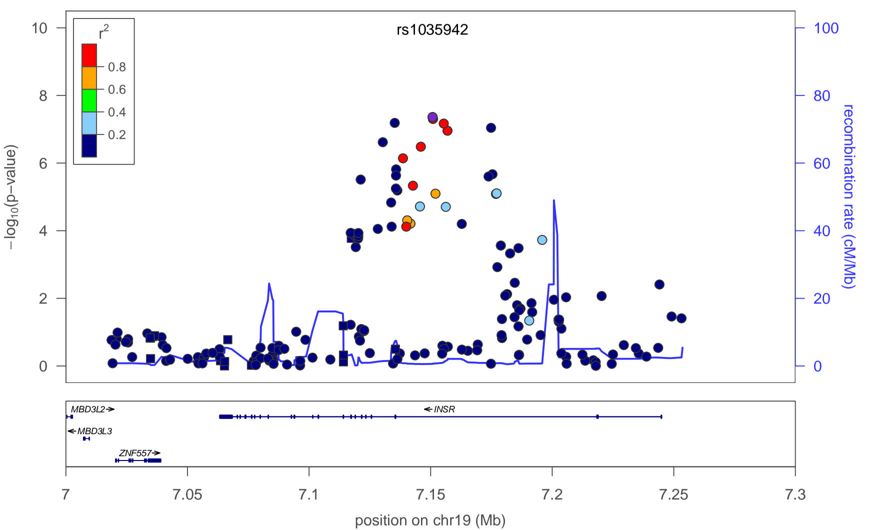

rs7188445
recombination rate (cM/Mb)
−log10(p−value)
MAF
position on chr16 (Mb)


rs7224610
recombination rate (cM/Mb)
−log10(p−value)
HLF
MMD
position on chr17 (Mb)


rs2079742
r^2
$-\log_{10}$(p-value)
recombination rate (cM/Mb)
BCAS3
TBX2
C17orf82
position on chr17 (Mb)


rs164009
r^2
$-\log_{10}$(p-value)
recombination rate (cM/Mb)
RNF157
FAM100B
QRICH2
PRPSAP1
SPHK1
UBE2O
position on chr17 (Mb)


rs12955983
r^2
$-\log_{10}$(p-value)
recombination rate (cM/Mb)
MC4R
position on chr18 (Mb)


rs1035942
r^2
$-\log_{10}$(p-value)
recombination rate (cM/Mb)
MBD3L2
MBD3L3
ZNF557
INSR
position on chr19 (Mb)

Acknowledgement

I want to thank all 140,000 participants of the individual studies who provided their genetic information as a data basis for this work and all field staff of those studies who made this data available.

Furthermore I want to thank all hundreds of authors of the related publications for conducting and managing the individual studies, genotpying, measuring serum urate levels, performing the statistical analysis at a study level, and especially for making this gigantic collaboration possible.

My special thank goes to all my colleagues who worked closely with me in those projects: the analysis- and writing group of the GUGC as well as my colleagues at Helmholtz Zentrum München, especially Anna Köttgen, Melanie Waldenberger, and Jan Krumsiek.

Last but not least I would like to sincerely thank Christian Gieger for his excellent support and advice and H.-Erich Wichmann for his supervision.

Publications

Rietveld CA, Medland SE, Derringer J, Yang J, Esko T, Martin NW, Westra HJ, Shakhbazov K, Abdellaoui A, Agrawal A, **Albrecht E**, Alizadeh BZ, Amin N, Barnard J, Baumeister SE, Benke KS, Bielak LF, Boatman JA, Boyle PA, Davies G, de Leeuw C, Eklund N, Evans DS, Ferhmann R, Fischer K, Gieger C, Gjessing HK, Hägg S, Harris JR, Hayward C, Holzapfel C, Ibrahim-Verbaas CA, Ingelsson E, Jacobsson B, Joshi PK, Jugessur A, Kaakinen M, Kanoni S, Karjalainen J, Kolcic I, Kristiansson K, Kutalik Z, Lahti J, Lee SH, Lin P, Lind PA, Liu Y, Lohman K, Loitfelder M, McMahon G, Vidal PM, Meirelles O, Milani L, Myhre R, Nuotio ML, Oldmeadow CJ, Petrovic KE, Peyrot WJ, Polasek O, Quaye L, Reinmaa E, Rice JP, Rizzi TS, Schmidt H, Schmidt R, Smith AV, Smith JA, Tanaka T, Terracciano A, van der Loos MJ, Vitart V, Völzke H, Wellmann J, Yu L, Zhao W, Allik J, Attia JR, Bandinelli S, Bastardot F, Beauchamp J, Bennett DA, Berger K, Bierut LJ, Boomsma DI, Bültmann U, Campbell H, Chabris CF, Cherkas L, Chung MK, Cucca F, de Andrade M, De Jager PL, De Neve JE, Deary IJ, Dedoussis GV, Deloukas P, Dimitriou M, Eiriksdottir G, Elderson MF, Eriksson JG, Evans DM, Faul JD, Ferrucci L, Garcia ME, Grönberg H, Gudnason V, Hall P, Harris JM, Harris TB, Hastie ND, Heath AC, Hernandez DG, Hoffmann W, Hofman A, Holle R, Holliday EG, Hottenga JJ, Iacono WG, Illig T, Järvelin MR, Kähönen M, Kaprio J, Kirkpatrick RM, Kowgier M, Latvala A, Launer LJ, Lawlor DA, Lehtimäki T, Li J, Lichtenstein P, Lichtner P, Liewald DC, Madden PA, Magnusson PK, Mäkinen TE, Masala M, McGue M, Metspalu A, Mielck A, Miller MB, Montgomery GW, Mukherjee S, Nyholt DR, Oostra BA, Palmer LJ, Palotie A, Penninx B, Perola M, Peyser PA, Preisig M, Räikkönen K, Raitakari OT, Realo A, Ring SM, Ripatti S, Rivadeneira F, Rudan I, Rustichini A, Salomaa V, Sarin AP, Schlessinger D, Scott RJ, Snieder H, St Pourcain B, Starr JM, Sul JH, Surakka I, Svento R, Teumer A; The LifeLines Cohort Study, Tiemeier H, van Rooij FJ, Van Wagoner DR, Vartiainen E, Viikari J, Vollenweider P, Vonk JM, Waeber G, Weir DR, Wichmann HE, Widen E, Willemsen G, Wilson JF, Wright AF, Conley D, Davey-Smith G, Franke L, Groenen PJ, Hofman A, Johannesson M, Kardia SL, Krueger RF, Laibson D, Martin NG, Meyer MN, Posthuma D, Thurik AR, Timpson NJ, Uitterlinden AG, van Duijn CM, Visscher PM, Benjamin DJ, Cesarini D, Koellinger PD. **GWAS of 126,559 Individuals Identifies Genetic Variants Associated with Educational Attainment**. *Science*. 2013 May 30. [Epub ahead of print]

Berndt SI, Gustafsson S, Mägi R, Ganna A, Wheeler E, Feitosa MF, Justice AE, Monda KL, Croteau-Chonka DC, Day FR, Esko T, Fall T, Ferreira T, Gentilini D, Jackson AU, Luan J, Randall JC, Vedantam S, Willer CJ, Winkler TW, Wood AR, Workalemahu T, Hu YJ, Lee SH, Liang L, Lin DY, Min JL, Neale BM, Thorleifsson G, Yang J, **Albrecht E**, Amin N, Bragg-Gresham JL, Cadby G, den Heijer M, Eklund N, Fischer K, Goel A, Hottenga JJ, Huffman JE, Jarick I, Johansson A, Johnson T, Kanoni S, Kleber ME, König IR, Kristiansson K, Kutalik Z, Lamina C, Lecoeur C, Li G, Mangino M, McArdle WL, Medina-Gomez C, Müller-Nurasyid M, Ngwa JS, Nolte IM, Paternoster L, Pechlivanis S, Perola M, Peters MJ, Preuss M, Rose LM, Shi J, Shungin D, Smith AV, Strawbridge RJ,

Surakka I, Teumer A, Trip MD, Tyrer J, Van Vliet-Ostaptchouk JV, Vandenput L, Waite LL, Zhao JH, Absher D, Asselbergs FW, Atalay M, Attwood AP, Balmforth AJ, Basart H, Beilby J, Bonnycastle LL, Brambilla P, Bruinenberg M, Campbell H, Chasman DI, Chines PS, Collins FS, Connell JM, Cookson WO, de Faire U, de Vegt F, Dei M, Dimitriou M, Edkins S, Estrada K, Evans DM, Farrall M, Ferrario MM, Ferrières J, Franke L, Frau F, Gejman PV, Grallert H, Grönberg H, Gudnason V, Hall AS, Hall P, Hartikainen AL, Hayward C, Heard-Costa NL, Heath AC, Hebebrand J, Homuth G, Hu FB, Hunt SE, Hyppönen E, Iribarren C, Jacobs KB, Jansson JO, Jula A, Kähönen M, Kathiresan S, Kee F, Khaw KT, Kivimäki M, Koenig W, Kraja AT, Kumari M, Kuulasmaa K, Kuusisto J, Laitinen JH, Lakka TA, Langenberg C, Launer LJ, Lind L, Lindström J, Liu J, Liuzzi A, Lokki ML, Lorentzon M, Madden PA, Magnusson PK, Manunta P, Marek D, März W, Leach IM, McKnight B, Medland SE, Mihailov E, Milani L, Montgomery GW, Mooser V, Mühleisen TW, Munroe PB, Musk AW, Narisu N, Navis G, Nicholson G, Nohr EA, Ong KK, Oostra BA, Palmer CN, Palotie A, Peden JF, Pedersen N, Peters A, Polasek O, Pouta A, Pramstaller PP, Prokopenko I, Pütter C, Radhakrishnan A, Raitakari O, Rendon A, Rivadeneira F, Rudan I, Saaristo TE, Sambrook JG, Sanders AR, Sanna S, Saramies J, Schipf S, Schreiber S, Schunkert H, Shin SY, Signorini S, Sinisalo J, Skrobek B, Soranzo N, Stančáková A, Stark K, Stephens JC, Stirrups K, Stolk RP, Stumvoll M, Swift AJ, Theodoraki EV, Thorand B, Tregouet DA, Tremoli E, Van der Klauw MM, van Meurs JB, Vermeulen SH, Viikari J, Virtamo J, Vitart V, Waeber G, Wang Z, Widén E, Wild SH, Willemsen G, Winkelmann BR, Witteman JC, Wolffenbuttel BH, Wong A, Wright AF, Zillikens MC, Amouyel P, Boehm BO, Boerwinkle E, Boomsma DI, Caulfield MJ, Chanock SJ, Cupples LA, Cusi D, Dedoussis GV, Erdmann J, Eriksson JG, Franks PW, Froguel P, Gieger C, Gyllensten U, Hamsten A, Harris TB, Hengstenberg C, Hicks AA, Hingorani A, Hinney A, Hofman A, Hovingh KG, Hveem K, Illig T, Jarvelin MR, Jöckel KH, Keinanen-Kiukaanniemi SM, Kiemeney LA, Kuh D, Laakso M, Lehtimäki T, Levinson DF, Martin NG, Metspalu A, Morris AD, Nieminen MS, Njølstad I, Ohlsson C, Oldehinkel AJ, Ouwehand WH, Palmer LJ, Penninx B, Power C, Province MA, Psaty BM, Qi L, Rauramaa R, Ridker PM, Ripatti S, Salomaa V, Samani NJ, Snieder H, Sørensen TI, Spector TD, Stefansson K, Tönjes A, Tuomilehto J, Uitterlinden AG, Uusitupa M, van der Harst P, Vollenweider P, Wallaschofski H, Wareham NJ, Watkins H, Wichmann HE, Wilson JF, Abecasis GR, Assimes TL, Barroso I, Boehnke M, Borecki IB, Deloukas P, Fox CS, Frayling T, Groop LC, Haritunian T, Heid IM, Hunter D, Kaplan RC, Karpe F, Moffatt MF, Mohlke KL, O'Connell JR, Pawitan Y, Schadt EE, Schlessinger D, Steinthorsdottir V, Strachan DP, Thorsteinsdottir U, van Duijn CM, Visscher PM, Di Blasio AM, Hirschhorn JN, Lindgren CM, Morris AP, Meyre D, Scherag A, McCarthy MI, Speliotes EK, North KE, Loos RJ, Ingelsson E. **Genome-wide meta-analysis identifies 11 new loci for anthropometric traits and provides insights into genetic architecture**. *Nat Genet.* 2013 Apr 26;45(5):501-12. doi: 10.1038/ng.2606. Epub 2013 Apr 7.

Fernández-Rhodes L, Demerath EW, Cousminer DL, Tao R, Dreyfus JG, Esko T, Smith AV, Gudnason V, Harris TB, Launer L, McArdle PF, Yerges-Armstrong LM, Elks CE, Strachan DP, Kutalik Z, Vollenweider P, Feenstra B, Boyd HA,

Metspalu A, Mihailov E, Broer L, Zillikens MC, Oostra B, van Duijn CM, Lunetta KL, Perry JR, Murray A, Koller DL, Lai D, Corre T, Toniolo D, **Albrecht E**, Stöckl D, Grallert H, Gieger C, Hayward C, Polasek O, Rudan I, Wilson JF, He C, Kraft P, Hu FB, Hunter DJ, Hottenga JJ, Willemsen G, Boomsma DI, Byrne EM, Martin NG, Montgomery GW, Warrington NM, Pennell CE, Stolk L, Visser JA, Hofman A, Uitterlinden AG, Rivadeneira F, Lin P, Fisher SL, Bierut LJ, Crisponi L, Porcu E, Mangino M, Zhai G, Spector TD, Buring JE, Rose LM, Ridker PM, Poole C, Hirschhorn JN, Murabito JM, Chasman DI, Widen E, North KE, Ong KK, Franceschini N. **Association of Adiposity Genetic Variants With Menarche Timing in 92,105 Women of European Descent**. *Am J Epidemiol.* 2013 Apr 4. [Epub ahead of print]

Codd V, Nelson CP, **Albrecht E**, Mangino M, Deelen J, Buxton JL, Hottenga JJ, Fischer K, Esko T, Surakka I, Broer L, Nyholt DR, Mateo Leach I, Salo P, Hägg S, Matthews MK, Palmen J, Norata GD, O'Reilly PF, Saleheen D, Amin N, Balmforth AJ, Beekman M, de Boer RA, Böhringer S, Braund PS, Burton PR, Craen AJ, Denniff M, Dong Y, Douroudis K, Dubinina E, Eriksson JG, Garlaschelli K, Guo D, Hartikainen AL, Henders AK, Houwing-Duistermaat JJ, Kananen L, Karssen LC, Kettunen J, Klopp N, Lagou V, van Leeuwen EM, Madden PA, Mägi R, Magnusson PK, Männistö S, McCarthy MI, Medland SE, Mihailov E, Montgomery GW, Oostra BA, Palotie A, Peters A, Pollard H, Pouta A, Prokopenko I, Ripatti S, Salomaa V, Suchiman HE, Valdes AM, Verweij N, Viñuela A, Wang X, Wichmann HE, Widen E, Willemsen G, Wright MJ, Xia K, Xiao X, van Veldhuisen DJ, Catapano AL, Tobin MD, Hall AS, Blakemore AI, van Gilst WH, Zhu H, Consortium C, Erdmann J, Reilly MP, Kathiresan S, Schunkert H, Talmud PJ, Pedersen NL, Perola M, Ouwehand W, Kaprio J, Martin NG, van Duijn CM, Hovatta I, Gieger C, Metspalu A, Boomsma DI, Jarvelin MR, Slagboom PE, Thompson JR, Spector TD, van der Harst P, Samani NJ. **Identification of seven loci affecting mean telomere length and their association with disease**. *Nat Genet.* 2013 Apr;45(4):422-7. doi: 10.1038/ng.2528. (shared first authorship between Codd V, Nelson CP, Albrecht E, and Mangino M)

Broer L, Codd V, Nyholt DR, Deelen J, Mangino M, Willemsen G, **Albrecht E**, Amin N, Beekman M, de Geus EJ, Henders A, Nelson CP, Steves CJ, Wright MJ, de Craen AJ, Isaacs A, Matthews M, Moayyeri A, Montgomery GW, Oostra BA, Vink JM, Spector TD, Slagboom PE, Martin NG, Samani NJ, van Duijn CM, Boomsma DI. **Meta-analysis of telomere length in 19713 subjects reveals high heritability, stronger maternal inheritance and a paternal age effect**. *Eur J Hum Genet.* 2013 Jan 16. doi: 10.1038/ejhg.2012.303. [Epub ahead of print]

Köttgen A, **Albrecht E**, Teumer A, Vitart V, Krumsiek J, Hundertmark C, Pistis G, Ruggiero D, O'Seaghdha CM, Haller T, Yang Q, Tanaka T, Johnson AD, Kutalik Z, Smith AV, Shi J, Struchalin M, Middelberg RP, Brown MJ, Gaffo AL, Pirastu N, Li G, Hayward C, Zemunik T, Huffman J, Yengo L, Zhao JH, Demirkan A, Feitosa MF, Liu X, Malerba G, Lopez LM, van der Harst P, Li X, Kleber ME, Hicks AA, Nolte IM, Johansson A, Murgia F, Wild SH, Bakker SJ, Peden JF, De-

hghan A, Steri M, Tenesa A, Lagou V, Salo P, Mangino M, Rose LM, Lehtimäki T, Woodward OM, Okada Y, Tin A, Müller C, Oldmeadow C, Putku M, Czamara D, Kraft P, Frogheri L, Thun GA, Grotevendt A, Gislason GK, Harris TB, Launer LJ, McArdle P, Shuldiner AR, Boerwinkle E, Coresh J, Schmidt H, Schallert M, Martin NG, Montgomery GW, Kubo M, Nakamura Y, Tanaka T, Munroe PB, Samani NJ, Jacobs DR Jr, Liu K, D'Adamo P, Ulivi S, Rotter JI, Psaty BM, Vollenweider P, Waeber G, Campbell S, Devuyst O, Navarro P, Kolcic I, Hastie N, Balkau B, Froguel P, Esko T, Salumets A, Khaw KT, Langenberg C, Wareham NJ, Isaacs A, Kraja A, Zhang Q, Wild PS, Scott RJ, Holliday EG, Org E, Viigimaa M, Bandinelli S, Metter JE, Lupo A, Trabetti E, Sorice R, Döring A, Lattka E, Strauch K, Theis F, Waldenberger M, Wichmann HE, Davies G, Gow AJ, Bruinenberg M; LifeLines Cohort Study, Stolk RP, Kooner JS, Zhang W, Winkelmann BR, Boehm BO, Lucae S, Penninx BW, Smit JH, Curhan G, Mudgal P, Plenge RM, Portas L, Persico I, Kirin M, Wilson JF, Leach IM, van Gilst WH, Goel A, Ongen H, Hofman A, Rivadeneira F, Uitterlinden AG, Imboden M, von Eckardstein A, Cucca F, Nagaraja R, Piras MG, Nauck M, Schurmann C, Budde K, Ernst F, Farrington SM, Theodoratou E, Prokopenko I, Stumvoll M, Jula A, Perola M, Salomaa V, Shin SY, Spector TD, Sala C, Ridker PM, Kähönen M, Viikari J, Hengstenberg C, Nelson CP; CARDIoGRAM Consortium; DIAGRAM Consortium; ICBP Consortium; MAGIC Consortium, Meschia JF, Nalls MA, Sharma P, Singleton AB, Kamatani N, Zeller T, Burnier M, Attia J, Laan M, Klopp N, Hillege HL, Kloiber S, Choi H, Pirastu M, Tore S, Probst-Hensch NM, Völzke H, Gudnason V, Parsa A, Schmidt R, Whitfield JB, Fornage M, Gasparini P, Siscovick DS, Polašek O, Campbell H, Rudan I, Bouatia-Naji N, Metspalu A, Loos RJ, van Duijn CM, Borecki IB, Ferrucci L, Gambaro G, Deary IJ, Wolffenbuttel BH, Chambers JC, März W, Pramstaller PP, Snieder H, Gyllensten U, Wright AF, Navis G, Watkins H, Witteman JC, Sanna S, Schipf S, Dunlop MG, Tönjes A, Ripatti S, Soranzo N, Toniolo D, Chasman DI, Raitakari O, Kao WH, Ciullo M, Fox CS, Caulfield M, Bochud M, Gieger C. **Genome-wide association analyses identify 18 new loci associated with serum urate concentrations**. *Nat Genet.* 2013 Feb;45(2):145-54. doi: 10.1038/ng.2500. Epub 2012 Dec 23. (shared first authorship between Köttgen A, Albrecht E, Teumer A, Vitart V, and Krumsiek J)

Buck D, **Albrecht E**, Aslam M, Goris A, Hauenstein N, Jochim A; International Multiple Sclerosis Genetics Consortium, Wellcome Trust Case Control Consortium, Cepok S, Grummel V, Dubois B, Berthele A, Lichtner P, Gieger C, Winkelmann J, Hemmer B. **Genetic variants in the immunoglobulin heavy chain locus are associated with the IgG index in multiple sclerosis**. *Ann Neurol.* 2013 Jan;73(1):86-94. doi: 10.1002/ana.23749. Epub 2012 Dec 7.

Boraska V, Jeroncic A, Colonna V, Southam L, Nyholt DR, William Rayner N, Perry JR, Toniolo D, **Albrecht E**, Ang W, Bandinelli S, Barbalic M, Barroso I, Beckmann JS, Biffar R, Boomsma D, Campbell H, Corre T, Erdmann J, Esko T, Fischer K, Franceschini N, Frayling TM, Girotto G, Gonzalez JR, Harris TB, Heath AC, Heid IM, Hoffmann W, Hofman A, Horikoshi M, Hua Zhao J, Jackson AU, Hottenga JJ, Jula A, Kähönen M, Khaw KT, Kiemeney LA, Klopp N, Kutalik

Z, Lagou V, Launer LJ, Lehtimäki T, Lemire M, Lokki ML, Loley C, Luan J, Mangino M, Mateo Leach I, Medland SE, Mihailov E, Montgomery GW, Navis G, Newnham J, Nieminen MS, Palotie A, Panoutsopoulou K, Peters A, Pirastu N, Polasek O, Rehnström K, Ripatti S, Ritchie GR, Rivadeneira F, Robino A, Samani NJ, Shin SY, Sinisalo J, Smit JH, Soranzo N, Stolk L, Swinkels DW, Tanaka T, Teumer A, Tönjes A, Traglia M, Tuomilehto J, Valsesia A, van Gilst WH, van Meurs JB, Smith AV, Viikari J, Vink JM, Waeber G, Warrington NM, Widen E, Willemsen G, Wright AF, Zanke BW, Zgaga L; Wellcome Trust Case Control Consortium (WTCCC), Boehnke M, d'Adamo AP, de Geus E, Demerath EW, den Heijer M, Eriksson JG, Ferrucci L, Gieger C, Gudnason V, Hayward C, Hengstenberg C, Hudson TJ, Järvelin MR, Kogevinas M, Loos RJ, Martin NG, Metspalu A, Pennell CE, Penninx BW, Perola M, Raitakari O, Salomaa V, Schreiber S, Schunkert H, Spector TD, Stumvoll M, Uitterlinden AG, Ulivi S, van der Harst P, Vollenweider P, Völzke H, Wareham NJ, Wichmann HE, Wilson JF, Rudan I, Xue Y, Zeggini E. **Genome-wide meta-analysis of common variant differences between men and women**. *Hum Mol Genet.* 2012 Nov 1;21(21):4805-15. doi: 10.1093/hmg/dds304. Epub 2012 Jul 27.

Okada Y, Sim X, Go MJ, Wu JY, Gu D, Takeuchi F, Takahashi A, Maeda S, Tsunoda T, Chen P, Lim SC, Wong TY, Liu J, Young TL, Aung T, Seielstad M, Teo YY, Kim YJ, Lee JY, Han BG, Kang D, Chen CH, Tsai FJ, Chang LC, Fann SJ, Mei H, Rao DC, Hixson JE, Chen S, Katsuya T, Isono M, Ogihara T, Chambers JC, Zhang W, Kooner JS; The KidneyGen Consortium, The CKDGen Consortium, **Albrecht E**, The GUGC consortium, Yamamoto K, Kubo M, Nakamura Y, Kamatani N, Kato N, He J, Chen YT, Cho YS, Tai ES, Tanaka T. **Meta-analysis identifies multiple loci associated with kidney function-related traits in east Asian populations**. *Nat Genet.* 2012 Jul 15;44(8):904-909. doi: 10.1038/ng.2352.

Granada M, Wilk JB, Tuzova M, Strachan DP, Weidinger S, **Albrecht E**, Gieger C, Heinrich J, Himes BE, Hunninghake GM, Celedón JC, Weiss ST, Cruikshank WW, Farrer LA, Center DM, O'Connor GT. **A genome-wide association study of plasma total IgE concentrations in the Framingham Heart Study**. *J Allergy Clin Immunol.* 2012 Mar;129(3):840-845. e21. Epub 2011 Nov 9.

Stolk L, Perry JR, Chasman DI, He C, Mangino M, Sulem P, Barbalic M, Broer L, Byrne EM, Ernst F, Esko T, Franceschini N, Gudbjartsson DF, Hottenga JJ, Kraft P, McArdle PF, Porcu E, Shin SY, Smith AV, van Wingerden S, Zhai G, Zhuang WV, **Albrecht E**, Alizadeh BZ, Aspelund T, Bandinelli S, Lauc LB, Beckmann JS, Boban M, Boerwinkle E, Broekmans FJ, Burri A, Campbell H, Chanock SJ, Chen C, Cornelis MC, Corre T, Coviello AD, d'Adamo P, Davies G, de Faire U, de Geus EJ, Deary IJ, Dedoussis GV, Deloukas P, Ebrahim S, Eiriksdottir G, Emilsson V, Eriksson JG, Fauser BC, Ferreli L, Ferrucci L, Fischer K, Folsom AR, Garcia ME, Gasparini P, Gieger C, Glazer N, Grobbee DE, Hall P, Haller T, Hankinson SE, Hass M, Hayward C, Heath AC, Hofman A, Ingelsson E, Janssens AC, Johnson AD, Karasik D, Kardia SL, Keyzer J, Kiel DP, Kolcic I, Kutalik Z, Lahti J, Lai S, Laisk T, Laven JS, Lawlor DA, Liu J, Lopez LM,

Louwers YV, Magnusson PK, Marongiu M, Martin NG, Klaric IM, Masciullo C, McKnight B, Medland SE, Melzer D, Mooser V, Navarro P, Newman AB, Nyholt DR, Onland-Moret NC, Palotie A, Paré G, Parker AN, Pedersen NL, Peeters PH, Pistis G, Plump AS, Polasek O, Pop VJ, Psaty BM, Räikkönen K, Rehnberg E, Rotter JI, Rudan I, Sala C, Salumets A, Scuteri A, Singleton A, Smith JA, Snieder H, Soranzo N, Stacey SN, Starr JM, Stathopoulou MG, Stirrups K, Stolk RP, Styrkarsdottir U, Sun YV, Tenesa A, Thorand B, Toniolo D, Tryggvadottir L, Tsui K, Ulivi S, van Dam RM, van der Schouw YT, van Gils CH, van Nierop P, Vink JM, Visscher PM, Voorhuis M, Waeber G, Wallaschofski H, Wichmann HE, Widen E, Wijnands-van Gent CJ, Willemsen G, Wilson JF, Wolffenbuttel BH, Wright AF, Yerges-Armstrong LM, Zemunik T, Zgaga L, Zillikens MC, Zygmunt M, Study TL, Arnold AM, Boomsma DI, Buring JE, Crisponi L, Demerath EW, Gudnason V, Harris TB, Hu FB, Hunter DJ, Launer LJ, Metspalu A, Montgomery GW, Oostra BA, Ridker PM, Sanna S, Schlessinger D, Spector TD, Stefansson K, Streeten EA, Thorsteinsdottir U, Uda M, Uitterlinden AG, van Duijn CM, Völzke H, Murray A, Murabito JM, Visser JA, Lunetta KL. **Meta-analyses identify 13 loci associated with age at menopause and highlight DNA repair and immune pathways**. *Nat Genet.* 2012 Jan 22;44(3):260-8. doi: 10.1038/ng.1051.

Boraska V, Day-Williams A, Franklin CS, Elliott KS, Panoutsopoulou K, Tachmazidou I, **Albrecht E**, Bandinelli S, Beilin LJ, Bochud M, Cadby G, Ernst F, Evans DM, Hayward C, Hicks AA, Huffman J, Huth C, James AL, Klopp N, Kolcic I, Kutalik Z, Lawlor DA, Musk AW, Pehlic M, Pennell CE, Perry JR, Peters A, Polasek O, St Pourcain B, Ring SM, Salvi E, Schipf S, Staessen JA, Teumer A, Timpson N, Vitart V, Warrington NM, Yaghootkar H, Zemunik T, Zgaga L, An P, Anttila V, Borecki IB, Holmen J, Ntalla I, Palotie A, Pietiläinen KH, Wedenoja J, Winsvold BS, Dedoussis GV, Kaprio J, Province MA, Zwart JA, Burnier M, Campbell H, Cusi D, Smith GD, Frayling TM, Gieger C, Palmer LJ, Pramstaller PP, Rudan I, Völzke H, Wichmann HE, Wright AF, Zeggini E. **Genome-wide association study to identify common variants associated with brachial circumference: a meta-analysis of 14 cohorts**. *PLoS One.* 2012;7(3):e31369. Epub 2012 Mar 29.

Paternoster L, Standl M, Chen CM, Ramasamy A, Bønnelykke K, Duijts L, Ferreira MA, Alves AC, Thyssen JP, **Albrecht E**, Baurecht H, Feenstra B, Sleiman PM, Hysi P, Warrington NM, Curjuric I, Myhre R, Curtin JA, Groen-Blokhuis MM, Kerkhof M, Sääf A, Franke A, Ellinghaus D, Fölster-Holst R, Dermitzakis E, Montgomery SB, Prokisch H, Heim K, Hartikainen AL, Pouta A, Pekkanen J, Blakemore AI, Buxton JL, Kaakinen M, Duffy DL, Madden PA, Heath AC, Montgomery GW, Thompson PJ, Matheson MC, Le Souëf P; Australian Asthma Genetics Consortium (AAGC), St Pourcain B, Smith GD, Henderson J, Kemp JP, Timpson NJ, Deloukas P, Ring SM, Wichmann HE, Müller-Nurasyid M, Novak N, Klopp N, Rodríguez E, McArdle W, Linneberg A, Menné T, Nohr EA, Hofman A, Uitterlinden AG, van Duijn CM, Rivadeneira F, de Jongste JC, van der Valk RJ, Wjst M, Jogi R, Geller F, Boyd HA, Murray JC, Kim C, Mentch F, March M, Mangino M, Spector TD, Bataille V, Pennell CE, Holt PG, Sly P, Tiesler CM, Thiering E, Illig T, Imboden M, Nystad W, Simpson A, Hottenga JJ, Postma D,

Koppelman GH, Smit HA, Söderhäll C, Chawes B, Kreiner-Møller E, Bisgaard H, Melén E, Boomsma DI, Custovic A, Jacobsson B, Probst-Hensch NM, Palmer LJ, Glass D, Hakonarson H, Melbye M, Jarvis DL, Jaddoe VW, Gieger C; Genetics of Overweight Young Adults (GOYA) Consortium, Strachan DP, Martin NG, Jarvelin MR, Heinrich J, Evans DM, Weidinger S; EArly Genetics & Lifecourse Epidemiology (EAGLE) Consortium. **Meta-analysis of genome-wide association studies identifies three new risk loci for atopic dermatitis**. *Nat Genet.* 2011 Dec 25;44(2):187-92. doi: 10.1038/ng.1017.

Soler Artigas M, Loth DW, Wain LV, Gharib SA, Obeidat M, Tang W, Zhai G, Zhao JH, Smith AV, Huffman JE, **Albrecht E**, Jackson CM, Evans DM, Cadby G, Fornage M, Manichaikul A, Lopez LM, Johnson T, Aldrich MC, Aspelund T, Barroso I, Campbell H, Cassano PA, Couper DJ, Eiriksdottir G, Franceschini N, Garcia M, Gieger C, Gislason GK, Grkovic I, Hammond CJ, Hancock DB, Harris TB, Ramasamy A, Heckbert SR, Heliövaara M, Homuth G, Hysi PG, James AL, Jankovic S, Joubert BR, Karrasch S, Klopp N, Koch B, Kritchevsky SB, Launer LJ, Liu Y, Loehr LR, Lohman K, Loos RJ, Lumley T, Al Balushi KA, Ang WQ, Barr RG, Beilby J, Blakey JD, Boban M, Boraska V, Brisman J, Britton JR, Brusselle GG, Cooper C, Curjuric I, Dahgam S, Deary IJ, Ebrahim S, Eijgelsheim M, Francks C, Gaysina D, Granell R, Gu X, Hankinson JL, Hardy R, Harris SE, Henderson J, Henry A, Hingorani AD, Hofman A, Holt PG, Hui J, Hunter ML, Imboden M, Jameson KA, Kerr SM, Kolcic I, Kronenberg F, Liu JZ, Marchini J, McKeever T, Morris AD, Olin AC, Porteous DJ, Postma DS, Rich SS, Ring SM, Rivadeneira F, Rochat T, Sayer AA, Sayers I, Sly PD, Smith GD, Sood A, Starr JM, Uitterlinden AG, Vonk JM, Wannamethee SG, Whincup PH, Wijmenga C, Williams OD, Wong A, Mangino M, Marciante KD, McArdle WL, Meibohm B, Morrison AC, North KE, Omenaas E, Palmer LJ, Pietiläinen KH, Pin I, Pola Sbreve Ek O, Pouta A, Psaty BM, Hartikainen AL, Rantanen T, Ripatti S, Rotter JI, Rudan I, Rudnicka AR, Schulz H, Shin SY, Spector TD, Surakka I, Vitart V, Völzke H, Wareham NJ, Warrington NM, Wichmann HE, Wild SH, Wilk JB, Wjst M, Wright AF, Zgaga L, Zemunik T, Pennell CE, Nyberg F, Kuh D, Holloway JW, Boezen HM, Lawlor DA, Morris RW, Probst-Hensch N; International Lung Cancer Consortium; GIANT consortium, Kaprio J, Wilson JF, Hayward C, Kähönen M, Heinrich J, Musk AW, Jarvis DL, Gläser S, Järvelin MR, Ch Stricker BH, Elliott P, O'Connor GT, Strachan DP, London SJ, Hall IP, Gudnason V, Tobin MD. **Genome-wide association and large-scale follow up identifies 16 new loci influencing lung function**. *Nat Genet.* 2011 Sep 25;43(11):1082-90. doi: 10.1038/ng.941.

Soler Artigas M, Wain LV, Repapi E, Obeidat M, Sayers I, Burton PR, Johnson T, Zhao JH, **Albrecht E**, Dominiczak AF, Kerr SM, Smith BH, Cadby G, Hui J, Palmer LJ, Hingorani AD, Wannamethee SG, Whincup PH, Ebrahim S, Smith GD, Barroso I, Loos RJ, Wareham NJ, Cooper C, Dennison E, Shaheen SO, Liu JZ, Marchini J; The NSHD Respiratory Study Team, Dahgam S, Naluai AT, Olin AC, Karrasch S, Heinrich J, Schulz H, McKeever TM, Pavord ID, Heliövaara M, Ripatti S, Surakka I, Blakey JD, Kähönen M, Britton JR, Nyberg F, Holloway JW, Lawlor DA, Morris RW, James AL, Jackson CM, Hall IP, Tobin MD; SpiroMeta

Consortium. **Effect of five genetic variants associated with lung function on the risk of chronic obstructive lung disease, and their joint effects on lung function**. *Am J Respir Crit Care Med.* 2011 Oct 1;184(7):786-95. doi: 10.1164/rccm.201102-0192OC.

Obeidat M, Wain LV, Shrine N, Kalsheker N, Soler Artigas M, Repapi E, Burton PR, Johnson T, Ramasamy A, Zhao JH, Zhai G, Huffman JE, Vitart V, **Albrecht E**, Igl W, Hartikainen AL, Pouta A, Cadby G, Hui J, Palmer LJ, Hadley D, McArdle WL, Rudnicka AR, Barroso I, Loos RJ, Wareham NJ, Mangino M, Soranzo N, Spector TD, Gläser S, Homuth G, Völzke H, Deloukas P, Granell R, Henderson J, Grkovic I, Jankovic S, Zgaga L, Polašek O, Rudan I, Wright AF, Campbell H, Wild SH, Wilson JF, Heinrich J, Imboden M, Probst-Hensch NM, Gyllensten U, Johansson A, Zaboli G, Mustelin L, Rantanen T, Surakka I, Kaprio J, Jarvelin MR, Hayward C, Evans DM, Koch B, Musk AW, Elliott P, Strachan DP, Tobin MD, Sayers I, Hall IP; SpiroMeta Consortium. **A comprehensive evaluation of potential lung function associated genes in the SpiroMeta general population sample**. *PLoS One.* 2011;6(5):e19382. Epub 2011 May 20.

Elks CE, Perry JR, Sulem P, Chasman DI, Franceschini N, He C, Lunetta KL, Visser JA, Byrne EM, Cousminer DL, Gudbjartsson DF, Esko T, Feenstra B, Hottenga JJ, Koller DL, Kutalik Z, Lin P, Mangino M, Marongiu M, McArdle PF, Smith AV, Stolk L, van Wingerden SH, Zhao JH, **Albrecht E**, Corre T, Ingelsson E, Hayward C, Magnusson PK, Smith EN, Ulivi S, Warrington NM, Zgaga L, Alavere H, Amin N, Aspelund T, Bandinelli S, Barroso I, Berenson GS, Bergmann S, Blackburn H, Boerwinkle E, Buring JE, Busonero F, Campbell H, Chanock SJ, Chen W, Cornelis MC, Couper D, Coviello AD, d'Adamo P, de Faire U, de Geus EJ, Deloukas P, Döring A, Smith GD, Easton DF, Eiriksdottir G, Emilsson V, Eriksson J, Ferrucci L, Folsom AR, Foroud T, Garcia M, Gasparini P, Geller F, Gieger C; GIANT Consortium, Gudnason V, Hall P, Hankinson SE, Ferreli L, Heath AC, Hernandez DG, Hofman A, Hu FB, Illig T, Järvelin MR, Johnson AD, Karasik D, Khaw KT, Kiel DP, Kilpeläinen TO, Kolcic I, Kraft P, Launer LJ, Laven JS, Li S, Liu J, Levy D, Martin NG, McArdle WL, Melbye M, Mooser V, Murray JC, Murray SS, Nalls MA, Navarro P, Nelis M, Ness AR, Northstone K, Oostra BA, Peacock M, Palmer LJ, Palotie A, Paré G, Parker AN, Pedersen NL, Peltonen L, Pennell CE, Pharoah P, Polasek O, Plump AS, Pouta A, Porcu E, Rafnar T, Rice JP, Ring SM, Rivadeneira F, Rudan I, Sala C, Salomaa V, Sanna S, Schlessinger D, Schork NJ, Scuteri A, Segrè AV, Shuldiner AR, Soranzo N, Sovio U, Srinivasan SR, Strachan DP, Tammesoo ML, Tikkanen E, Toniolo D, Tsui K, Tryggvadottir L, Tyrer J, Uda M, van Dam RM, van Meurs JB, Vollenweider P, Waeber G, Wareham NJ, Waterworth DM, Weedon MN, Wichmann HE, Willemsen G, Wilson JF, Wright AF, Young L, Zhai G, Zhuang WV, Bierut LJ, Boomsma DI, Boyd HA, Crisponi L, Demerath EW, van Duijn CM, Econs MJ, Harris TB, Hunter DJ, Loos RJ, Metspalu A, Montgomery GW, Ridker PM, Spector TD, Streeten EA, Stefansson K, Thorsteinsdottir U, Uitterlinden AG, Widen E, Murabito JM, Ong KK, Murray A. **Thirty new loci for age at menarche identified by a meta-analysis of genome-wide association studies**. *Nat Genet.* 2010 Dec;42(12):1077-85.

Marzi C, **Albrecht E**, Hysi PG, Lagou V, Waldenberger M, Tönjes A, Prokopenko I, Heim K, Blackburn H, Ried JS, Kleber ME, Mangino M, Thorand B, Peters A, Hammond CJ, Grallert H, Boehm BO, Kovacs P, Geistlinger L, Prokisch H, Winkelmann BR, Spector TD, Wichmann HE, Stumvoll M, Soranzo N, März W, Koenig W, Illig T, Gieger C. **Genome-wide association study identifies two novel regions at 11p15.5-p13 and 1p31 with major impact on acute-phase serum amyloid A**. *PLoS Genet.* 2010 Nov 18;6(11):e1001213. (shared first authorship between Marzi C and Albrecht E)

Lango Allen H, Estrada K, Lettre G, Berndt SI, Weedon MN, Rivadeneira F, Willer CJ, Jackson AU, Vedantam S, Raychaudhuri S, Ferreira T, Wood AR, Weyant RJ, Segrè AV, Speliotes EK, Wheeler E, Soranzo N, Park JH, Yang J, Gudbjartsson D, Heard-Costa NL, Randall JC, Qi L, Vernon Smith A, Mägi R, Pastinen T, Liang L, Heid IM, Luan J, Thorleifsson G, Winkler TW, Goddard ME, Sin Lo K, Palmer C, Workalemahu T, Aulchenko YS, Johansson A, Zillikens MC, Feitosa MF, Esko T, Johnson T, Ketkar S, Kraft P, Mangino M, Prokopenko I, Absher D, **Albrecht E**, Ernst F, Glazer NL, Hayward C, Hottenga JJ, Jacobs KB, Knowles JW, Kutalik Z, Monda KL, Polasek O, Preuss M, Rayner NW, Robertson NR, Steinthorsdottir V, Tyrer JP, Voight BF, Wiklund F, Xu J, Zhao JH, Nyholt DR, Pellikka N, Perola M, Perry JR, Surakka I, Tammesoo ML, Altmaier EL, Amin N, Aspelund T, Bhangale T, Boucher G, Chasman DI, Chen C, Coin L, Cooper MN, Dixon AL, Gibson Q, Grundberg E, Hao K, Juhani Junttila M, Kaplan LM, Kettunen J, König IR, Kwan T, Lawrence RW, Levinson DF, Lorentzon M, McKnight B, Morris AP, Müller M, Suh Ngwa J, Purcell S, Rafelt S, Salem RM, Salvi E, Sanna S, Shi J, Sovio U, Thompson JR, Turchin MC, Vandenput L, Verlaan DJ, Vitart V, White CC, Ziegler A, Almgren P, Balmforth AJ, Campbell H, Citterio L, De Grandi A, Dominiczak A, Duan J, Elliott P, Elosua R, Eriksson JG, Freimer NB, Geus EJ, Glorioso N, Haiqing S, Hartikainen AL, Havulinna AS, Hicks AA, Hui J, Igl W, Illig T, Jula A, Kajantie E, Kilpeläinen TO, Koiranen M, Kolcic I, Koskinen S, Kovacs P, Laitinen J, Liu J, Lokki ML, Marusic A, Maschio A, Meitinger T, Mulas A, Paré G, Parker AN, Peden JF, Petersmann A, Pichler I, Pietiläinen KH, Pouta A, Ridderstråle M, Rotter JI, Sambrook JG, Sanders AR, Schmidt CO, Sinisalo J, Smit JH, Stringham HM, Bragi Walters G, Widen E, Wild SH, Willemsen G, Zagato L, Zgaga L, Zitting P, Alavere H, Farrall M, McArdle WL, Nelis M, Peters MJ, Ripatti S, van Meurs JB, Aben KK, Ardlie KG, Beckmann JS, Beilby JP, Bergman RN, Bergmann S, Collins FS, Cusi D, den Heijer M, Eiriksdottir G, Gejman PV, Hall AS, Hamsten A, Huikuri HV, Iribarren C, Kähönen M, Kaprio J, Kathiresan S, Kiemeney L, Kocher T, Launer LJ, Lehtimäki T, Melander O, Mosley TH Jr, Musk AW, Nieminen MS, O'Donnell CJ, Ohlsson C, Oostra B, Palmer LJ, Raitakari O, Ridker PM, Rioux JD, Rissanen A, Rivolta C, Schunkert H, Shuldiner AR, Siscovick DS, Stumvoll M, Tönjes A, Tuomilehto J, van Ommen GJ, Viikari J, Heath AC, Martin NG, Montgomery GW, Province MA, Kayser M, Arnold AM, Atwood LD, Boerwinkle E, Chanock SJ, Deloukas P, Gieger C, Grönberg H, Hall P, Hattersley AT, Hengstenberg C, Hoffman W, Lathrop GM, Salomaa V, Schreiber S, Uda M, Waterworth D, Wright AF, Assimes TL, Barroso I, Hofman A, Mohlke KL, Boomsma DI, Caulfield MJ, Cupples LA, Erdmann J, Fox CS, Gudnason V, Gyllensten U, Harris TB, Hayes RB, Jarvelin MR, Mooser

V, Munroe PB, Ouwehand WH, Penninx BW, Pramstaller PP, Quertermous T, Rudan I, Samani NJ, Spector TD, Völzke H, Watkins H, Wilson JF, Groop LC, Haritunians T, Hu FB, Kaplan RC, Metspalu A, North KE, Schlessinger D, Wareham NJ, Hunter DJ, O'Connell JR, Strachan DP, Wichmann HE, Borecki IB, van Duijn CM, Schadt EE, Thorsteinsdottir U, Peltonen L, Uitterlinden AG, Visscher PM, Chatterjee N, Loos RJ, Boehnke M, McCarthy MI, Ingelsson E, Lindgren CM, Abecasis GR, Stefansson K, Frayling TM, Hirschhorn JN. **Hundreds of variants clustered in genomic loci and biological pathways affect human height**. *Nature.* 2010 Oct 14;467(7317):832-8. Epub 2010 Sep 29.

Repapi E, Sayers I, Wain LV, Burton PR, Johnson T, Obeidat M, Zhao JH, Ramasamy A, Zhai G, Vitart V, Huffman JE, Igl W, **Albrecht E**, Deloukas P, Henderson J, Granell R, McArdle WL, Rudnicka AR; Wellcome Trust Case Control Consortium, Barroso I, Loos RJ, Wareham NJ, Mustelin L, Rantanen T, Surakka I, Imboden M, Wichmann HE, Grkovic I, Jankovic S, Zgaga L, Hartikainen AL, Peltonen L, Gyllensten U, Johansson A, Zaboli G, Campbell H, Wild SH, Wilson JF, Gläser S, Homuth G, Völzke H, Mangino M, Soranzo N, Spector TD, Polasek O, Rudan I, Wright AF, Heliövaara M, Ripatti S, Pouta A, Naluai AT, Olin AC, Torén K, Cooper MN, James AL, Palmer LJ, Hingorani AD, Wannamethee SG, Whincup PH, Smith GD, Ebrahim S, McKeever TM, Pavord ID, MacLeod AK, Morris AD, Porteous DJ, Cooper C, Dennison E, Shaheen S, Karrasch S, Schnabel E, Schulz H, Grallert H, Bouatia-Naji N, Delplanque J, Froguel P, Blakey JD; NSHD Respiratory Study Team, Britton JR, Morris RW, Holloway JW, Lawlor DA, Hui J, Nyberg F, Jarvelin MR, Jackson C, Kähönen M, Kaprio J, Probst-Hensch NM, Koch B, Hayward C, Evans DM, Elliott P, Strachan DP, Hall IP, Tobin MD. **Genome-wide association study identifies five loci associated with lung function**. *Nat Genet.* 2010 Jan;42(1):36-44. Epub 2009 Dec 13.

Kolz M, Johnson T, Sanna S, Teumer A, Vitart V, Perola M, Mangino M, **Albrecht E**, Wallace C, Farrall M, Johansson A, Nyholt DR, Aulchenko Y, Beckmann JS, Bergmann S, Bochud M, Brown M, Campbell H; EUROSPAN Consortium, Connell J, Dominiczak A, Homuth G, Lamina C, McCarthy MI; ENGAGE Consortium, Meitinger T, Mooser V, Munroe P, Nauck M, Peden J, Prokisch H, Salo P, Salomaa V, Samani NJ, Schlessinger D, Uda M, Völker U, Waeber G, Waterworth D, Wang-Sattler R, Wright AF, Adamski J, Whitfield JB, Gyllensten U, Wilson JF, Rudan I, Pramstaller P, Watkins H; PROCARDIS Consortium, Doering A, Wichmann HE; KORA Study, Spector TD, Peltonen L, Völzke H, Nagaraja R, Vollenweider P, Caulfield M; WTCCC, Illig T, Gieger C. **Meta-analysis of 28,141 individuals identifies common variants within five new loci that influence uric acid concentrations**. *PLoS Genet.* 2009 Jun;5(6):e1000504. Epub 2009 Jun 5.

Eidesstattliche Versicherung

Albrecht, Eva

Name, Vorname

Ich erkläre hiermit an Eides statt,

dass ich die vorliegende Dissertation mit dem Thema

Genetic and metabolic components in the regulation of serum urate levels in humans

selbständig verfasst, mich außer der angegebenen keiner weiteren Hilfsmittel bedient und alle Erkenntnisse, die aus dem Schrifttum ganz oder annähernd übernommen sind, als solche kenntlich gemacht und nach ihrer Herkunft unter Bezeichnung der Fundstelle einzeln nachgewiesen habe.

Ich erkläre des Weiteren, dass die hier vorgelegte Dissertation nicht in gleicher oder in ähnlicher Form bei einer anderen Stelle zur Erlangung eines akademischen Grades eingereicht wurde.

München,

Ort, Datum

Unterschrift Doktorandin/Doktorand