A meta-analysis of genome-wide data from five European isolates reveals an association of COL22A1, SYT1, and GABRR2 with serum creatinine level

Citation for published version:
Pattaro, C, De Grandi, A, Vitart, V, Hayward, C, Franke, A, Aulchenko, YS, Johansson, A, Wild, SH, Melville, SA, Isaacs, A, Polasek, O, Ellinghaus, D, Kolcic, I, Noethlings, U, Zgaga, L, Zemunik, T, Gnewuch, C, Schreiber, S, Campbell, S, Hastie, N, Boban, M, Meitinger, T, Oostra, BA, Riegler, P, Minelli, C, Wright, AF, Campbell, H, van Duijn, CM, Gyllensten, U, Wilson, JF, Krawczak, M, Rudan, I, Pramstaller, PP & EUROSPAN Consortium 2010, 'A meta-analysis of genome-wide data from five European isolates reveals an association of COL22A1, SYT1, and GABRR2 with serum creatinine level' BMC Medical Genetics, vol 11, no. 41, pp. -. DOI: 10.1186/1471-2350-11-41

Digital Object Identifier (DOI):
10.1186/1471-2350-11-41

Link:
Link to publication record in Edinburgh Research Explorer

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

Published In:
BMC Medical Genetics

Publisher Rights Statement:
© 2010 Pattaro et al; licensee BioMed Central Ltd.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/2.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

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

Take down policy
The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact openaccess@ed.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.
A meta-analysis of genome-wide data from five European isolates reveals an association of COL22A1, SYT1, and GABRR2 with serum creatinine level

Cristian Pattaro1*, Alessandro De Grandi1, Veronique Vitart2, Caroline Hayward2, Andre Franke3, Yuri S Aulchenko4, Asa Johansson5, Sarah H Wild6, Scott A Melville3, Aaron Isaacs4, Ozren Polasek7,8, David Ellinghaus3, Ivana Kolcic7, Ute Nöthlings9,10, Lina Zgaga7, Tatjana Zemunik11, Carsten Gnewuch12, Stefan Schreiber3, Susan Campbell2, Nick Hastie2, Mladen Boban11, Thomas Meitinger13,14, Ben A Oostra4, Peter Riegler15, Cosetta Minelli1, Alan F Wright2, Harry Campbell6, Cornelia M van Duijn6, Ulf Gyllensten5, James F Wilson6, Michael Krawczak9,16, Igor Rudan6,8,11, Peter P Pramstaller1,17,18*, the EUROSPAN consortium

Abstract

Background: Serum creatinine (S\textsubscript{CR}) is the most important biomarker for a quick and non-invasive assessment of kidney function in population-based surveys. A substantial proportion of the inter-individual variability in S\textsubscript{CR} level is explicable by genetic factors.

Methods: We performed a meta-analysis of genome-wide association studies of S\textsubscript{CR} undertaken in five population isolates (‘discovery cohorts’), all of which are part of the European Special Population Network (EUROSPAN) project. Genes showing the strongest evidence for an association with S\textsubscript{CR} (candidate loci) were replicated in two additional population-based samples (‘replication cohorts’).

Results: After the discovery meta-analysis, 29 loci were selected for replication. Association between S\textsubscript{CR} level and polymorphisms in the collagen type XXII alpha 1 (COL22A1) gene, on chromosome 8, and in the synaptotagmin-1 (SYT1) gene, on chromosome 12, were successfully replicated in the replication cohorts (p value = 1.0 × 10\textsuperscript{-6} and 1.7 × 10\textsuperscript{-4}, respectively). Evidence of association was also found for polymorphisms in a locus including the gamma-aminobutyric acid receptor rho-2 (GABRR2) gene and the ubiquitin-conjugating enzyme E2-J1 (UBE2J1) gene (replication p value = 3.6 × 10\textsuperscript{-3}). Previously reported findings, associating glomerular filtration rate with SNPs in the uromodulin (UMOD) gene and in the schroom family member 3 (SCHROOM3) gene were also replicated.

Conclusions: While confirming earlier results, our study provides new insights in the understanding of the genetic basis of serum creatinine regulatory processes. In particular, the association with the genes SYT1 and GABRR2 corroborate previous findings that highlighted a possible role of the neurotransmitters GABA\textsubscript{A} receptors in the regulation of the glomerular basement membrane and a possible interaction between GABA\textsubscript{A} receptors and synaptotagmin-I at the podocyte level.

* Correspondence: cristian.pattaro@eurac.edu; peter.pramstaller@eurac.edu
1 Institute of Genetic Medicine, European Academy Bozen/Bolzano (EURAC), Bolzano, Italy - Affiliated Institute of the University Lübeck, Lübeck, Germany

© 2010 Pattaro et al; licensee BioMed Central Ltd. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/2.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.
Background
In epidemiological population-based surveys, serum creatinine (SCR) represents the most important biomarker for a quick and non-invasive assessment of kidney function, allowing estimation of the glomerular filtration rate (GFR) [1]. At the same time, an increased SCR level has also been recognized as a risk factor for adverse outcomes in patients hospitalized for cardiac surgery or heart failure [2]. A substantial proportion of the interindividual variability in SCR level is explicable by genetic factors. Twin [3] and pedigree-based studies [4-7] yielded heritability estimates ranging from 0.19 to 0.53, with higher values observed in subjects not treated for hypertension [4]. A large number of genetic loci have emerged from genome-wide linkage analyses as being related to variation in SCR level or in SCR-based estimates of both GFR and creatinine clearance [4-15].

Focusing upon quantitative renal phenotypes, a recent GWA study [16] and a subsequent replication study [17] identified variants in the 5,10-methenyltetrahydrofolate synthetase (MTHFS) gene to be associated with chronic kidney disease (CKD; defined as GFR < 60 ml/min/1.73 m²). More recently, variants in the uromodulin (UMOD) gene have been shown to be associated with GFR and CKD independent of age, sex, hypertension, and diabetic status [18]. In the same study, SNPs in the schroom family member 3 (SCHROOM3) gene, the glycine amidinotransferase (GATM)-spermatogenesis-associated 5-like 1 (SPATA5L1) locus, and in the jagged 1 (JAG1) gene were also found to be associated with GFR [18].

We have reported linkage between SCR level and a region at 22q13 containing the myosin heavy chain 9 non-muscle (MYH9) gene [4], a locus that had been associated with non-diabetic end-stage renal disease (ESRD) [19,20] and glomerulosclerosis [21] before. This linkage was detected in families from three isolated European populations participating in the European Special Population Research Network (EUROSPAN). Here, we have performed a genome-wide association analysis of SCR level combining data from all five EUROSPAN populations, with SCR re-measured using an enzymatic method in one and the same central laboratory. We have performed population-specific GWA studies and subjected the results to an inverse-variance, fixed effects meta-analysis. Selected candidate regions were then tested in two additional, population-based samples from Europe.

Methods
Study samples
For all EUROSPAN and replication studies, written informed consent was obtained from all participants and all protocols were approved by the institutional ethical review committees of the participating centres.

Eurospan
The EUROSPAN project http://homepages.ed.ac.uk/s0565445/index.html was initiated in 2006 and involves five population isolates from Italy, Croatia, Scotland, Sweden, and the Netherlands. The project aims at assessing the genetic structure of European isolates and at identifying genes underlying common traits, taking advantage of the genetic and environmental homogeneity that usually characterizes population isolates. In the current context, according to Neel [22], with population isolates we mean “secondary isolates”, i.e. groups that, for some reasons, detached or were detached from larger populations. In particular, EUROSPAN cohorts were derived from small population samples which have grown slowly, with little recruitment from outside the groups.

The ERF study is a family-based project including over 3000 participants that originated from 22 couples living in the Rucphen region of the Netherlands in the 19th century [23-25]. All descendants of these people were invited to visit a clinical research center in the region, where they were examined in person and where blood was taken after fasting. Height and weight were measured for each participant. All participants filled out a questionnaire on risk factors.

The MICROS study is part of the genomic health care program ‘GenNova’ and was carried out in three villages of the Val Venosta, South Tyrol (Italy), in 2001-2003. It comprised members of the populations of Stelvio, Valletina and Martello. A detailed description of the MICROS study can be found elsewhere [26]. Briefly, study participants were volunteers from three isolated villages located in the Italian Alps, in a German-speaking region bordering upon Austria and Switzerland. Owing to geographical, historical and political reasons, the entire region experienced a prolonged period of isolation from surrounding populations. Information on the participants’ health status was collected through a standardized questionnaire. Laboratory data were obtained from standard blood analyses. Genotyping was performed on >1400 participants, with 1334 of them suitable for analysis after data cleaning.

The Northern Swedish Population Health Study (NSPHS) is a family-based study including a comprehensive health assessment and the collection of data on family structure, lifestyle, diet, medical history and of samples for laboratory analyses [27,28]. Participants came from the northern part of the Swedish mountain region (County of Norrbotten, Parish of Karesuando). Historic population accounts show that little migration...
or population changes have occurred in this area over the last 200 years.

The Orkney Complex Disease Study (ORCADES) is an ongoing, family-based and cross-sectional study in the isolated Scottish archipelago of Orkney [29]. Genetic diversity in this population is reduced in comparison to Mainland Scotland, consistent with high levels of historical endogamy. Participants were aged 18-100 years and came from a subgroup of ten islands. Fasting blood samples were collected and over 200 health-related phenotypes and environmental exposures were measured in each individual.

The Vis study includes 986 unselected Croatians, aged 18-93 years, who were recruited during 2003 and 2004 from the villages of Vis and Komiza on the Dalmatian island of Vis [30,31]. The settlements on Vis island have a unique history and remained isolated from other villages and the outside world for centuries. Participants were phenotyped for 450 disease-related quantitative traits. Biochemical and physiological measurements were performed, detailed genealogies reconstructed, questionnaires on lifestyle and environmental exposures collected, and blood samples and lymphocytes extracted and stored for further analyses.

All DNA samples were genotyped on Illumina Infinium HumanHap300 v2 SNP bead microarrays according to the manufacturer’s instructions, except for samples from Vis for which version 1 was used (the Vis samples had 311,398 SNPs genotyped in common with the other populations).

For all five studies, serum or plasma creatinine was measured at the Institute for Clinical Chemistry and Laboratory Medicine, Regensburg University Medical Center, Germany, using an enzymatic photometric assay on an ADVIA1650 clinical chemistry analyzer (Siemens Healthcare Diagnostics GmbH, Eschborn, Germany) [32]. The number of individuals with available creatinine, sex, and age information is reported for each study in Table 1.

**Replication Cohorts**

Data on German healthy control individuals were obtained from the popgen biobank [33]. Genotyping constituted an essential part of the GWAS initiative of the German National Genome Research Network (NGFN) and was performed at an Affymetrix service facility (South San Francisco, CA, USA) using the Affymetrix Genome-Wide Human SNP Array 6.0 (1000 k) (Santa Clara, CA, USA). Genotype calling was carried out using Affymetrix’ Birdseed v2 algorithm with default quality thresholds. Samples with more than 5% missing genotypes, showing excess genetic dissimilarity to the remaining subjects, or with evidence for a cryptic relatedness to other study participants were removed. These quality control measures left 1213 control samples for inclusion in the replication cohort. All sex assignments could be verified by reference to the proportion of heterozygous SNPs on the X chromosome. Serum creatinine was available for 1140 individuals and was measured at the Institute for Clinical Chemistry in Kiel, Germany, using an enzymatic *in vitro* assay (CREAplus, Cobas®, Roche Diagnostics, Indianapolis, IN).

The *Korcula* study included 944 unselected 18-98 year old Croatians, recruited into the study during 2007 from

---

**Table 1 Characteristics of studies and study participants.**

| Study name        | ERF          | MICRO       | NPSHS       | ORCADES     | VIS                 | popgen       | Korcula  |
|-------------------|--------------|-------------|-------------|-------------|---------------------|--------------|----------|
| Nationality       | The Netherlands | Italy       | Sweden      | UK          | Croatia             | Germany      | Croatia  |
| Population type   | isolated     | isolated    | isolated    | isolated    | isolated            | general      | isolated |
| Genotyping platform | Illumina 318 K | Illumina 318 K | Illumina 318 K | Illumina 318 K | Illumina 318 K | Affymetrix 1000 K | Illumina 370 K |
| Sample size*      | 775          | 1086        | 653         | 718         | 774                 | 1140         | 895      |
| Females: n (%)    | 472 (61%)    | 615 (57%)   | 345 (53%)   | 385 (54%)   | 454 (59%)           | 534 (47%)    | 572 (64%) |
| Age: mean (sd)    | 53 (15)      | 45 (16)     | 47 (21)     | 54 (16)     | 57 (15)             | 54 (15)      | 56 (14)  |
| Diabetes: n (%)   | 36 (4.9%)    | 39 (3.6%)   | 44 (6.7%)   | 21 (3.0%)   | 72 (9.4%)           | 18 (1.6%)    | 93 (10.3%)|
| AHT*: n (%)       | 190 (24.5%)  | 85 (7.8%)   | 124 (19.0%) | 152 (21.2%) | 192 (25.2%)         | Not available | 197 (22.0%) |
| \(S_{\text{Cr}}\) mg/dl: mean (sd) | 1.01 (0.20) | 0.96 (0.14) | 0.94 (0.20) | 0.97 (0.15) | 1.01 (0.31) | 0.93 (0.15) | 0.92 (0.15) |
| Females           | 0.85 (0.21)  | 0.78 (0.12) | 0.75 (0.14) | 0.77 (0.19) | 0.79 (0.27) | 0.74 (0.12) | 0.75 (0.12) |
| eGFR† ml/min/1.73 m² | 81.1 (19.0) | 87.8 (15.1) | 92.2 (22.1) | 84.3 (16.0) | 82.0 (19.4) | 88.0 (17.5) | 77.1 (14.9) |
| Females           | 75.8 (20.5)  | 83.4 (16.8) | 88.7 (19.2) | 81.5 (18.5) | 81.0 (19.8) | 85.5 (17.2) | 72.4 (14.1) |

* Number of individuals with available creatinine, sex, and age information.

† AHT: Anti-hypertensive treatment

† Estimated using the updated abbreviated MDRD study equation [58]
the island of Korcula [34]. The settlements on Korcula included the Eastern region of the island, which has a unique population history and has maintained a high level of isolation from other mainland populations and from the Western part of the island. Participants were phenotyped for >400 disease-related quantitative traits. Biochemical and physiological measurements were performed, genealogies reconstructed, questionnaires on lifestyle and environmental exposures collected, and blood samples and lymphocytes extracted and stored for further analyses. DNA was genotyped using the Illumina Infinium HumanCNV370v1 SNP bead microarrays. Serum creatinine as measured by the Jaffé rate method was available for 895 individuals.

**Statistical analysis**

To ensure normality within centers and comparability of $S_{\text{CR}}$ values across centers, we applied a quantile normalization in all discovery and replication studies, which involves ranking all $S_{\text{CR}}$ values and converting them to $z$-scores according to a standard normal distribution.

**Discovery stage**

To account for inter-individual relatedness within the five EUROSPAN cohorts, genome-wide association (GWA) analysis was carried out following a two stage approach. In the first stage, a sex- and age-adjusted linear model was fitted to the normalized $S_{\text{CR}}$ in order to estimate the residuals. A polygenic model was then fitted to estimate the inverse of the variance-covariance matrix, which accounts for the inter-individual relatedness and is based upon a genomic kinship matrix as described in Amin et al. [35]. The association between SNPs and residuals was assessed by means of an approximate score test statistic [36], assuming an additive model, as implemented in the GenABEL package [37]. The results from the five EUROSPAN cohorts were then combined into a fixed-effects meta-analysis with inverse-variance weighting, using MetABEL http://mga.bionet.nsc.ru/~yurii/ABEL/. Only SNPs that had a call rate $\geq 0.95$, a Hardy-Weinberg equilibrium (HWE) test $p$ value $> 10^{-6}$ and a minor allele frequency (MAF) $\geq 0.01$ were included in the analyses. In total, 322,498 SNPs were tested for association with $S_{\text{CR}}$. The threshold for genome-wide statistical significance, according to Bonferroni adjustment for multiple testing, was set to $1.55 \times 10^{-7}$. Between-study heterogeneity was quantified using the $I^2$ statistic, i.e. the percentage of total variation explained by heterogeneity rather than sampling error [38].

**Replication stage**

In the absence of clear methodological guidance on what may be the best strategy for passing SNPs to a replication stage, we selected SNPs for replication based on a tradeoff that enabled us to include our best findings from the discovery analysis, whilst avoiding the risk of an excessively long list. This was achieved by including all SNPs with a $p$ value $\leq 10^{-5}$, but also allowing for SNPs with higher $p$ values to get into the replication list in the presence of additional evidence for association provided by other SNPs within 100 kb. The following criteria were applied: (i) at least one $p$ value $\leq 10^{-3}$; (ii) at least one $p$ value $\leq 10^{-4}$ and at least one additional SNP with $p$ value $\leq 10^{-3}$ within 100 kb; (iii) at least three SNPs with $p$ value $\leq 10^{-3}$ within 100 kb. For each group of candidate SNPs, we selected a candidate gene (all SNPs included in the gene were considered for replication) or region (all SNPs included within $\pm$ 100 kb of the candidate SNPs were considered for replication) to test in the independent cohorts of Korcula and popgen. The complete workflow of the test procedure is depicted in Figure 1.

In popgen, the association between $S_{\text{CR}}$ and each SNP within a candidate region was assessed by means of a sex- and age-adjusted linear regression model assuming additive genetic effects, using PLINK version 1.05 http://pngu.mgh.harvard.edu/purcell/plink/ [39]. For Korcula, given that it is a family-based study, we used the same approach as described above for the EUROSPAN cohorts.

To assess the significance in the replication cohorts of associations with different SNPs from those genotyped in the discovery cohorts, we defined region-specific $p$ values as follows: for a genomic region containing $N$ SNPs, we counted the number $n$ of SNPs that achieved $p$ value $< 0.05$. Assuming that $p$ values are uniformly distributed between 0 and 1 under the null hypothesis of no association, the distribution function of $n$, $F(n)$, is that of a binomial with parameters $N$ and 0.05. The region-specific $p$ value then equals $1-F_{N,0.05}(n-1)$. Given that in small genomic regions the assumption of independence is rarely met, the estimated $p$ value should be considered as conservative since the effective number of independent tests can only be lower than the total number of SNPs tested. Meta-analysis of region-specific $p$ values from popgen and Korcula was finally performed using the Fisher’s combined probability test [40], which is suitable for combining tests performed in independent samples. We further evaluated rejection of the null hypothesis of no association based on a false discovery rate (FDR) of 0.05 [41], where $p$ values are sorted in ascending order and the first $k$ tests with $p$ value $\leq i/m \times \alpha$ are considered significant (in our case, $i = 1..29$, $m = 29$, $\alpha = 0.05$).

**Replication of previous findings**

We finally assessed whether any of the four loci reported to be associated with eGFR by Köttgen et al. [18] showed evidence for an association with $S_{\text{CR}}$ in our discovery meta-analysis. Köttgen et al. reported four
Figure 1 Workflow of the discovery and replication stages of the study (for details, see the Statistical Analysis section)
SNPs to be associated with GFR: rs17319721 (\textit{SHROOM3}), rs2467853 (\textit{SPATA5L1-GATM}), rs12917707 (\textit{UMOD}), and rs6040055 (\textit{JAG1}). For each of the four SNPs, we defined LD blocks around them, following the method by Gabriel et al. [42] and using SNPs with MAF \(\geq 1\%\) from the HapMap-CEU database, Phase III/Release 2. Given that the genes were in four independent loci, the threshold for claiming significant replication, according to Bonferroni, was set to 0.05/4 = 0.0125. To compare our results with the commonly used method of testing replication at the same SNP or at a proxy SNP, we repeated the test procedure at one SNP per locus, using the same threshold for statistical significance.

If not specified otherwise, all data management, data analysis, programming, and the creation of graphs were performed using R 2.8.0 [43].

\section*{Results}
Study-specific characteristics of the participants are reported in Table 1. A total of 4006 individuals were included in the meta-analysis whilst the replication cohorts comprised 2035 individuals. All EUROSPAN cohorts comprised a higher percentage of females (between 53\% and 61\%) than males. The mean age ranged from 45 years in MICROS to 57 years in VIS. The replication samples had a similar age range, but a smaller percentage of females than males included in popgen.

The results of the GWA meta-analysis are depicted in Figure 2. The genomic control factor (\(\lambda\)), as assessed by the quantile-quantile plot included in Figure 2, was 1.004 (SE < 10\(^{-5}\)), indicating that no cryptic relatedness or gross population structure affected our results [44]. The smallest p value in the meta-analysis was 1.5 \(\times\) 10\(^{-6}\) and was obtained for SNP rs2396463 in the collagen type IV alpha-3 (\textit{COL4A3}) gene on chromosome 2q. In the absence of hits of genome-wide significance, we selected a set of promising regions for follow-up in two independent replication samples, namely Korcula (an island population from Croatia) and popgen (a random population sample from the most northern part of

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{figure2.png}
\caption{Manhattan plot of -log\(_{10}\) (p values) from the meta-analysis of five GWA studies. GWA results were combined across all studies by fixed-effects meta-analysis using inverse variance weighting. Square dots indicate SNPs satisfying criteria for selection of candidate regions. SNPs within replicated regions are colored in dark red and SNPs within non-replicated regions in pink. The genomic control factor of \(\lambda = 1.004\), assessed by the quantile-quantile plot in the upper-right panel, indicates that cryptic relatedness and population structure have been modeled appropriately.}
\end{figure}
Germany). Twenty-nine genomic regions were considered for replication. The full list of regions and the results of the discovery meta-analysis are provided in the Additional file 1 - Table S1. Since the SNP panels genotyped in the replication samples differed from one another and from that used in the discovery GWA studies, all markers present in the candidate regions were included in the replication phase. A total of 2224 SNPs were analyzed in the popgen samples and 1136 SNPs were analyzed in the Korcula samples. Results of the replication analysis are fully reported in the Additional file 1 - Table S2, with loci sorted by significance.

Two candidate regions from the discovery GWA studies also showed significant evidence for an association with SCR level in the popgen and Korcula samples after Bonferroni correction for multiple testing of the combined p value (significance threshold for 29 independent tests: $\alpha = 0.0017$). When controlling for the FDR, i.e. for the probability of reporting a false positive result, a third region could be added to the list of replicated findings (Figure 3).

The first replicated region was defined by four SNPs (rs4075073, rs4588898, rs9324496, and rs2873682) in the collagen type XXII alpha 1 (COL22A1) gene, all of which yielded p values $\leq 9.79 \times 10^{-4}$ in the meta-analysis of the discovery stage, with homogeneous effects observed across populations: $I^2 = 0\%$ for all four SNPs. A forest plot of the marker with the strongest evidence for association, SNP rs4588898, is provided in Figure 3A. Six other SNPs in the gene (out of 72) had p values between $3.42 \times 10^{-3}$ and 0.0414 (Table 2). As it can be inferred from Figure 5, some 19 of 140 COL22A1 SNPs genotyped in popgen were found to be significantly associated with SCR (red squares), as were 11 out of 71 SNPs in Korcula (green squares). Fisher’s combined probability test p value was $1.0 \times 10^{-6}$. Nevertheless, the gene regions with the strongest SCR association in EUROSPAN and either popgen or Korcula failed
to overlap and were found to be located in distinct LD blocks. Interestingly, however, the smallest $p$ values for popgen and Korcula fell into the same LD block.

Three closely linked SNPs (rs10506807, rs12300068, and rs11112829) on chromosome 12q21, yielded $p$ values $\leq 7.55 \times 10^{-5}$ in the EUROSPAN discovery meta-analysis. $I^2$ for heterogeneity was 21.88%, 24.11%, and 19.94%, respectively, none of which was statistically significant. A forest plot for SNP rs12300068 is provided in Figure 4B. Depending upon the annotation of the 5'-UTR sequence, all SNPs are located either $>100$ kb upstream (NCBI) or directly inside (Ensembl and UCSC) the synaptotagmin-1 ($\text{SYT1}$) gene. In the popgen sample, nine SNPs located less than 50 kb upstream of the three discovery SNPs were significantly associated with SCR, with $p$ values between $2.0 \times 10^{-5}$ and $8.5 \times 10^{-3}$ (Table 3). The Fisher’s combined probability test $p$ value for the whole region was as low as $1.7 \times 10^{-4}$.

The third locus was defined by the four SNPs rs3777514 ($p$ value $= 4.5 \times 10^{-4}$), rs2064831 ($p$ value $= 7.2 \times 10^{-5}$), rs1998576 ($p$ value $= 8.6 \times 10^{-5}$), and rs7744005 ($p$ value $= 3.6 \times 10^{-4}$). The effect of the three SNPs upon $S_{CR}$ with $p$ values between $2.0 \times 10^{-5}$ and $8.5 \times 10^{-3}$ (Table 3). The Fisher’s combined probability test $p$ value for the whole region was as low as $1.7 \times 10^{-4}$.

The pooling plot of the SNP with the strongest evidence for association (rs2064831), is reported in Figure 4C. The SNPs in question are located in a region on chromosome 6q15 containing the gamma-aminobutyric acid receptor rho-2 ($\text{GABRR2}$) and the ubiquitin-conjugating enzyme E2-J1 ($\text{UBE2J1}$) genes (Figure 6). Five additional SNPs (out of a total of 28 SNPs tested) in this region had $p$ values between $2.2 \times 10^{-3}$ and 0.0280 (Table 4). Of the 37 SNPs genotyped in the popgen sample, eight were associated with $S_{CR}$ with a $p$ value $\leq 0.05$; seven of them overlapped with the $\text{GABRR2}$ and $\text{UBE2J1}$ genes (see the blue squares in Figure 6). The most consistent association signals in EUROSPAN and popgen were found near the $\text{GABRR2}$ gene and its promoter, and the respective SNPs were in linkage disequilibrium ($r^2 > 0.59$, Figure 6, middle panel). SNP rs12195070 was genotyped in both EUROSPAN and popgen: effect direction was discordant but in both cases association was very significant with $p$ values of $4.5 \times 10^{-3}$ in EUROSPAN and $p$ value $= 7.2 \times 10^{-3}$ in popgen, respectively. In the Korcula samples, the smallest $p$ value in the region was 0.0616 so that we cannot formally claim replication in this population. However, Fisher’s combined probability test $p$ value equaled $3.6 \times 10^{-3}$. This value does not meet the Bonferroni threshold for multiple testing, but it does meet the FDR threshold of 0.0052, which controls the probability of reporting false positives at a 0.05 level (see Additional file 1 - Table S2 and Figure 3). Detailed results for this locus are reported in Table 4.

Although consistently observed in all EUROSPAN populations (Figure 4D), the association noted between $S_{CR}$ and the $\text{COL4A3}$ gene could not be verified in the
replication cohorts (Fisher's combined probability test p value = 0.216).

Unfortunately, the number of people with diabetes included in the EUROSPAN populations was too small to allow any meaningful subgroup analysis, adjusting for disease state. However, when we repeated our GWA meta-analysis in non-diabetic subjects only, the p values for the three replicated gene regions changed so little that any strong interaction of the respective loci with diabetes status appears rather unlikely.

Table 2 SNPs showing the highest evidence of association in the COL22A1 gene (chromosome 8).[^1]

| study     | name      | position       | Genotype (Ref. All.) | Ref. All. Freq. | n     | Beta* | SE  | p value** |
|-----------|-----------|----------------|----------------------|----------------|-------|-------|-----|-----------|
| EUROSPLAN | rs4075073 | 139808143      | A/C (A)              | 0.37           | 3996  | 0.0671| 0.0203| 9.8 × 10⁻⁴|
| EUROSPLAN | rs4588898 | 139815390      | A/G (A)              | 0.30           | 3999  | 0.0786| 0.0212| 2.1 × 10⁻⁴|
| EUROSPLAN | rs13255079| 139828786      | C/T (T)              | 0.16           | 4001  | 0.0769| 0.0263| 3.4 × 10⁻⁴|
| EUROSPLAN | rs9324496 | 139833578      | C/T (T)              | 0.34           | 3995  | 0.0716| 0.0205| 4.6 × 10⁻⁴|
| EUROSPLAN | rs7824025 | 139839088      | A/C (A)              | 0.07           | 3998  | 0.0762| 0.0374| 0.0414    |
| EUROSPLAN | rs2873682 | 139840693      | A/G (A)              | 0.30           | 2610  | 0.0915| 0.0264| 5.2 × 10⁻⁴|
| EUROSPLAN | rs11166844| 139848062      | C/T (C)              | 0.39           | 3993  | 0.0485| 0.0202| 0.0166    |
| EUROSPLAN | rs1574172 | 139941802      | A/G (A)              | 0.14           | 3998  | -0.0713| 0.0284| 0.0122    |
| EUROSPLAN | rs4736078 | 139944298      | C/T (C)              | 0.22           | 3996  | -0.1603| 0.0342| 0.0003    |
| EUROSPLAN | rs6596041 | 139988449      | C/T (C)              | 0.35           | 3936  | -0.1032| 0.0318| 0.0003    |

[^1]: association models were adjusted for age and sex; *in standard deviations; **p value: for EUROSPLAN, the p value refers to the meta-analysis of the five individual GWA studies, for popgen and Korcula, the p value is from the linear regression analysis (for details, see Methods).

Table 2 SNPs showing the highest evidence of association in the COL22A1 gene (chromosome 8).[^1]

Results of the replication analysis at the four loci reported to be associated with eGFR by Köttgen et al [18] are reported in the Additional file 1 - Table S3. We identified three SNPs in the UMOD gene locus:
rs4293393 was significantly associated with $S_{CR}$ level ($p$ value $= 3.9 \times 10^{-5}$) and weaker signals were observed for rs11647727 ($p$ value $= 0.011$) and rs4506906 ($p$ value $= 0.028$). The region-specific $p$ value for this locus was $1.3 \times 10^{-3}$. Of the five SNPs identified in the SPA-TASL1-GATM locus, rs1153860, rs1346268, and rs1719247 were associated with $S_{CR}$ at $p$ values of $5.0 \times 10^{-4}$, $1.9 \times 10^{-3}$, and $4.6 \times 10^{-3}$, respectively (region-specific $p$ value $= 1.2 \times 10^{-3}$). At the SCHROOM3 gene locus, 53 SNPs were tested: three of them resulted in a $p$ value $\leq 0.05$ (region-specific $p$ value $= 0.50$). The smallest $p$ value of $1.7 \times 10^{-3}$ was observed for SNP
rs4859682 (this SNP was in strong LD with the rs17319721 reported in the paper, \(r^2 = 0.94\)). None of the 8 polymorphisms tested in the \(JAG1\) gene was significantly associated with \(S_{CR}\) level, including the rs6400455 which was included in our database as well. Our conclusions would have been slightly different had we attempted replication at the level of the individual variants rather than at the locus level. In this case, also the SCHROOM3 locus would have been replicated.

**Discussion**

The main result of our study was the identification of three novel genetic associations with \(S_{CR}\) level. These were detected in a meta-analysis of five European isolated population samples and replicated in two independent population samples, one from the general Central-European population (popgen), and one from another isolate (Korcula). The replicated loci included the \(COL22A1\) gene on chromosome 8 and the \(SYT1\) gene on chromosome 12. The third locus, including the \(GABRR2\) and \(UBE2J1\) genes on chromosome 6, showed substantial evidence of association as well. In addition, the association of two loci (\(UMOD\) and \(SPATASLI\) and one SNP in the \(SCHROOM3\) locus, previously reported with GFR, was also found to apply to \(S_{CR}\). This confirmation also supports our choice of using \(S_{CR}\) adjusted by age and sex which is equivalent to study GFR calculated from \(S_{CR}\) weighted by age and sex [1] (black or white ethnicity, which should also be accounted for, was not relevant in our study involving individuals of European origin).

The two SNPs with the strongest evidence for association in the discovery stage of our study were located in the promoter region of the \(COL4A3\) gene, which is expressed in the glomerular basement membrane (GBM) of the kidney, and which has been associated with Alport syndrome, Goodpasture syndrome [45] and benign hematuria [46]. Despite making biological sense, however, the association between \(COL4A3\) and \(S_{CR}\) as observed in our meta-analysis could not be verified in the replication stage. On the one hand, this could be interpreted to indicate that no common variants of that gene are involved in the regulation of kidney function. On the other hand, the lack of replication could be due either to the involvement of the gene in regulatory pathways, implying a strong interaction with other genes, or to a particularly prominent interaction with population-specific environmental conditions [47].

In any case, a slightly less significant association between three other genes and \(S_{CR}\) from the discovery meta-analysis could be replicated in either the popgen or the Korcula data. Different SNPs in the \(COL22A1\) gene were found to be associated with \(S_{CR}\) in all discovery and replication cohorts. This association may reflect the biological relationship between muscle mass formation and creatinine levels. In fact, *in situ* hybridization of myotendinous junctions has revealed that muscle cells produce collagen XXII, and functional tests have shown that collagen XXII acts as a cell adhesion ligand for skin epithelial cells and fibroblasts [48].

The discovery stage association of \(S_{CR}\) with the \(SYT1\) gene was well replicated in the popgen sample. Even though the annotation of gene reference sequences is not uniform across different databases (NCBI, UCSC, Ensembl), the associated SNPs are located in the fourth and last intron of the 5' \(SYT1\) un-translated region.
Figure 6 Genomic structure and association results at the GABRR2-UBE2J1 locus. **Upper panel**: log10(p values) are plotted by physical position for the EUROSPAN discovery meta-analysis (red squares), the popgen (blue squares) and the Korcula (green squares) replication cohorts. **Middle panel**: linkage disequilibrium (LD) as quantified by $r^2$ (the higher the LD, the darker the color, with black indicating perfect LD), based upon the HapMap-CEU database, Phase III/Release 2, (NCBI build 36). **Lower panel**: genes located in the plotted region, with coding exons indicated by black rectangles and orientation.
and neurite outgrowth [51] and is thought to interact with neurotransmitters such as the GABA (gamma-aminobutyric acid) receptor, expressed in the GBM. It also appears to participate in the regulation of podocyte homeostasis [50]. Synaptotagmin-I plays a role in neurotransmitter release and neurite outgrowth [50]. Synaptotagmin-I is a member of the rho subunit family of the transmembrane receptors for gamma-aminobutyric acid (GABA_A), which are linked to potassium channels via G-proteins. GABA_A receptors have been shown to be expressed in the kidney of multiple species, with subunits β1 and β2 localized in the proximal tubules [54]. More recently, several subtypes of GABA receptors, including the GABBR2, have been shown to be transcribed in normal human glomeruli [50] and it plays a role in retinal neurotransmission [55]. The gene is located close to the GABBR1 gene in the same transcriptional orientation, suggesting a similar expression and regulatory pattern. Interestingly, the locus falls into the supporting interval for the SCR reported in Mexican Americans [7], and to creatinine clearance as reported in Caucasians [9].

Our discovery analysis was performed in five isolated communities from Southern, Central, and Northern Europe. As has been demonstrated empirically for South Tyrolean villages [56], population isolates provide a reduced within-study heterogeneity of environmental factors. The same locus becomes of particular interest in the present context is expression of the SYT1 gene product, synaptotagmin-I, in renal podocytes. Glomerular podocytes possess structures resembling synaptic vesicles and contain glutamate, they co-express Rab3A (a GTPase restricted to cell types that regulate exocytosis), synapsin, synaptophysin, and synaptotagmin-I, and undergo spontaneous and stimulated exocytosis and recycling, with glutamate release [50]. Synaptotagmin-I plays a role in neurotransmitter release and neurite outgrowth [51] and is thought to interact with neurotransmitters such as the GABA receptor, expressed in the GBM. It also appears to participate in the regulation of podocyte homeostasis [50].

In light of the findings on the SYT1 gene, the association with the UBE2J1 and GABBR2 locus becomes of high interest. The association detected in the EUROSPAN data was corroborated by the popgen data. Moreover, SNPs associated with SCR in popgen enabled us to narrow down the associated region to the GABBR2 promoter. The same locus was not replicated in Korcula data. However, while the popgen genotype distribution of the SNPs in question was found to resemble that of the HapMap-CEU samples [52], indicating that popgen can be considered representative of the general Central to Northern European populations, this may not be true for the Korcula sample. In fact, Korcula is a small population located in the Croatian islands, where pronounced genetic isolation is known to have persisted until recently [53]. This could be the reason of the missed replication. The GABBR2 gene product is a member of the rho subunit family of the transmembrane receptors for gamma-aminobutyric acid (GABA_A), which are linked to potassium channels via G-proteins. GABA_A receptors have been shown to be expressed in the kidney of multiple species, with subunits β1 and β2 localized in the proximal tubules [54]. More recently, several subtypes of GABA receptors, including the GABBR2, have been shown to be transcribed in normal human glomeruli [50] and it plays a role in retinal neurotransmission [55]. The gene is located close to the GABBR1 gene in the same transcriptional orientation, suggesting a similar expression and regulatory pattern. Interestingly, the locus falls into the supporting interval for linkage to SCR reported in Mexican Americans [7], and to creatinine clearance as reported in Caucasians [9].

Our discovery analysis was performed in five isolated communities from Southern, Central, and Northern Europe. As has been demonstrated empirically for South Tyrolean villages [56], population isolates provide a reduced within-study heterogeneity of environmental factors.
factors, thereby facilitating the genetic dissection of complex traits. This internal homogeneity of environmental and life style factors could be counterbalanced by enhanced between-study heterogeneity. However, this appears not to have been the case in our study. For SNPs in 27 of the 29 regions selected for replication, the hypothesis of homogeneity of effects’ sizes was not rejected, even at a very stringent significance threshold (given the small number of studies involved in the meta-analysis, \( \alpha \) was set to 0.10). In particular, homogeneity of association signals was verified for all replicated loci and for the \( \text{COL4A3} \) gene locus. This finding strengthens our results in the sense that, despite the expected between-population heterogeneity, the results of the association analyses were found to be very consistent across studies.

Another advantage of our study has been that \( S_{\text{CR}} \) values of all EUROSPAN participants were measured in one and the same centralized laboratory using the enzymatic method. In this way, calibration difference could be excluded as a confounder of our genetic association analyses. While the enzymatic method was used to measure \( S_{\text{CR}} \) in EUROSPAN and popgen participants, the Jaffé rate method was used in the Korcula study. This method is known to be less precise than the enzymatic one. Additional noise introduced by increased measurement error would best explain the reduced evidence for association replication in the Korcula study. However, \( S_{\text{CR}} \) was standardized in all discovery and replication samples using a transformation based upon the ranks of a standard normal distribution. The standardization of the phenotype measures allowed us to combine data across studies avoiding bias due to technical inter-laboratory differences.

The validity of our analysis was also evidenced by the replication of earlier association findings [18]. Association of \( S_{\text{CR}} \) with \( \text{UMOD} \) and \( \text{SPATA5L1-GATM} \) was confirmed using the same replication approach used in our own analysis. With our method, however, we could not replicate the association over the \( \text{SCHROOM3} \) locus (the locus would be confirmed if using a classic replication based on the specific SNP or a proxy SNP). This finding could be an indication that our method of replication is more likely to be conservative and, so, less prone to false positive results. In the \( \text{COL22A1} \) and the \( \text{SYT1} \) genes the replicated signals were located in different LD blocks, highlighting the possibility that different polymorphisms at the same locus may be associated with the same phenotype in different populations. In the \( \text{GABRR2} - \text{UBE2J1} \) gene locus, association signals between discovery and replication studies were discordant but clearly overlapping. Whether the discordance of the effect alleles is an indication of a false positive result is very unlikely, given the size of the effects in EUROSPAN and in popgen. The presence of a flip-flop phenomenon [57] could be hypothesized and should be investigated further.

Conclusions

In conclusion, in addition to confirming earlier findings, our search for genes associated with \( S_{\text{CR}} \) variation led to the discovery of three novel genes that represent sensible candidates for further functional analysis. The recent hypothesis of an important role of neurotransmitters in the regulation of the GBM at the podocyte level [50] renders our findings particularly relevant for understanding the regulation of renal function, suggesting that the possible interaction between \( \text{SYT1} \) and \( \text{GABRR2} \) warrants further investigation.

Acknowledgements

We owe a debt of gratitude to all participants in the seven studies and their relatives. We are grateful to Prof. John Thompson for the very constructive discussion on statistical methodology. EUROSPAN (European Special Populations Research Network) was supported by European Commission FP6 STRP grant number 018947 (LSHG-CT-2006-01947). High-throughput genome-wide association analysis of the data was supported by joint grant from Netherlands Organization for Scientific Research and the Russian Foundation for Basic Research (NWO-RFBR 047.017.043). The ERF study was supported by grants from the NWO, Erasmus MC and the Centre for Medical Systems Biology (CMSB). We are grateful general practitioners and neurologists for their contributions and to P. Veraart for her help in genealogy, Jeannette Vergeer for the supervision of the laboratory work and P. Sniders for his help in data collection.

For the MICROs study, we thank the primary care practitioners Raffaela Stocker, Stefan Waldner, Toni Pizzecco, Josef Plangger, Ugo Marcadent and the personnel of the Hospital of Silandro (Department of Laboratory Medicine) for their participation and collaboration in the research project. In South Tyrol, the study was supported by the Ministry of Health and Department of Educational Assistance, University and Research of the Autonomous Province of Bolzano and the South Tyrolean Sparkasse Foundation.

The Northern Swedish Population Health Study was supported by grants from The Swedish Natural Sciences Research Council, The European Commission through EUROSPAN, The Foundation for Research and Development (SSF) and The Linneaus Centre for Bioinformatics (LCB). We are also grateful for the contribution of samples from the Medical Biobank in Umeå and for the contribution of the district nurse Svea Hennix in the Karesuando study. ORCADES was supported by the Scottish Executive Health Department and the Royal Society. DNA extractions were performed at the Wellcome Trust Clinical Research Facility in Edinburgh. We would like to acknowledge the invaluable contributions of Lorraine Anderson, the research nurses in Orkney, and the administrative team in Edinburgh.

The VIS study was supported through the grants from the Medical Research Council UK to HC, AFW and IR, and Ministry of Science, Education and Sport of the Republic of Croatia to IR (number 108-1080315-0302). The authors collectively thank a large number of individuals for their individual help in organizing, planning and carrying out the field work related to the project.
and data management: Professor Pavao Rudan and the staff of the Institute for Anthropological Research in Zagreb, Croatia (organization of the field work, anthropometric and physiological measurements, and DNA extraction); Professor Ariana Vorko-Jovic and the staff and medical students of the Andrija Stampar School of Public Health of the Faculty of Medicine, University of Zagreb, Croatia (questionnaires, genealogy reconstruction and data entry); Dr Branka Salzer from the biochemistry lab “Salzer”, Croatia (measurements of biochemical traits); local general practitioners and nurses (recruitment and communication with the study population); and the employees of several other Croatian institutions who participated in the field work, including but not limited to the University of Rijeka and Split, Croatia; Croatian Institute of Public Health; Institutes of Public Health in Split and Dubrovnik, Croatia. SNP Genotyping of the Vis samples was carried out by the Genetics Core Laboratory at the Wellcome Trust Clinical Research Facility, WGH, Edinburgh. The popgen study was supported by the German Ministry of Education and Research (BMBF) through the National Genome Research Network (NGFN). It is currently funded by the Ministry of Science, Commerce and Transportation of the State of Schleswig-Holstein. The project has also received infrastructure support through the DFG excellence cluster “Inflammation at Interfaces”.

Author details
1 Institute of Genetic Medicine, European Academy Bozen/Bolzano (EURAC), Bolzano, Italy - Affiliated Institute of the University of Lubeck, Lubeck, Germany.
2 MRC Human Genetics Unit, Institute of Genetics and Molecular Medicine, Edinburgh, UK.
3 Institute for Clinical Chemistry and Molecular Biology, Christian-Albrechts-University Kiel, Kiel, Germany.
4 Institute of Medical Genetics, University of Regensburg, Regensburg, Germany.
5 Institute of Human Genetics, Technical University of Munich, Munich, Germany.
6 Centre for Population Health Sciences, University of Edinburgh Medical School, Teviot Place, Edinburgh, Edinburgh EH8 9AG, UK.

Competing interests
The authors declare that they have no competing interests. The funders had no role in the study design, the data collection and analysis, the decision to publish, or the preparation of the manuscript.

Received: 30 September 2009 Accepted: 11 March 2010 Published: 11 March 2010

References
1. Stevens LA, Coren J, Greene T, Levy AS: Assessing kidney function—measured and estimated glomerular filtration rate. N Engl J Med 2006, 354(23):2473-2483.
2. Proutl MG, Shlipak MG: Are small changes in serum creatinine an important risk factor? Curr Opin Nephrol Hypertens 2005, 14(3):265-270.
3. Whitfield JB, Martin NG: The effects of inheritance on constituents of plasma: a twin study on some biochemical variables. Ann Clin Biochem 1984, 21(3):176-183.
4. Patterson C, Aulchenko YS, Isaacs A, Vitart V, Hayward C, Franklin CS, Polasek O, Kolcic I, Biloglav Z, Campbell R, Haste N, Lauk G, Geitner T, Oostra BA, Gyllensten U, Wilson JF, Pichler I, Hicks AA, Campbell H, Wright AF, Rudan I, van Duijn CM, Regler P, Marron F, Pramstaller PP: Genome-wide linkage analysis of serum creatinine in three isolated European populations. Kidney Int 2009, 76(3):297-306.
5. Fox CS, Yang Q, Cupples LA, Guo CY, Larson MG, Leip EP, Wilson PW, Leve D: Genomewide linkage analysis to serum creatinine, GFR, and creatinine clearance in a community-based population: the Framingham Heart Study. J Am Soc Nephrol 2004, 15(9):2457-2461.
55. Marcos I, Ruiz A, Blaschak CJ, Borrego S, Cutting GR, Antinolo G: Mutation analysis of GABRR1 and GABRR2 in autosomal recessive retinitis pigmentosa. J Med Genet 2000, 37(6):E5.

56. Marroni F, Grazio D, Pattaro C, Devoto M, Pramstaller P: Estimates of genetic and environmental contribution to 43 quantitative traits support sharing of a homogeneous environment in an isolated population from South Tyrol, Italy. Hum Hered 2008, 65(3):175-182.

57. Lin PI, Vance JM, Pericak-Vance MA, Martin ER: No gene is an island: the flip-flop phenomenon. Am J Hum Genet 2007, 80(3):531-538.

58. Levey AS, Coresh J, Greene T, Stevens LA, Zhang YL, Hendriksen S, Kusek JW, Van Lente F: Using standardized serum creatinine values in the modification of diet in renal disease study equation for estimating glomerular filtration rate. Ann Intern Med 2006, 145(4):247-254.

Pre-publication history
The pre-publication history for this paper can be accessed here:http://www.biomedcentral.com/1471-2350/11/41/prepub

doi:10.1186/1471-2350-11-41
Cite this article as: Pattaro et al: A meta-analysis of genome-wide data from five European isolates reveals an association of COL22A1, SYT1, and GABRR2 with serum creatinine level. BMC Medical Genetics 2010 11:41.