

1 **Title**

2 Systematic mediation and interaction analyses of kidney function genetic loci in a general  
3 population study.

4

5 **Authors**

6 Dariush Ghasemi-Semeskandeh<sup>1,2,3\*</sup>, David Emmert<sup>1</sup>, Eva König<sup>1</sup>, Luisa Foco<sup>1</sup>, Martin Gögele<sup>1</sup>,  
7 Laura Barin<sup>1</sup>, Ryosuke Fujii<sup>1,4,5</sup>, Christian Fuchsberger<sup>1</sup>, Dorien J.M. Peters<sup>2</sup>, Peter P.  
8 Pramstaller<sup>1</sup>, Cristian Pattaro<sup>1\*</sup>

9

10 **Affiliations**

11 <sup>1</sup> Institute for Biomedicine, Eurac Research, Bolzano, Italy

12 <sup>2</sup> Department of Human Genetics, Leiden University Medical Center, Leiden, The Netherlands

13 <sup>3</sup> Health Data Science Centre, Human Technopole, Milan, Italy

14 <sup>4</sup> Department of Preventive Medical Sciences, Fujita Health University School of Medical  
15 Sciences, Toyoake, Japan

16 <sup>5</sup> Department of Preventive Medicine, Nagoya University Graduate School of Medicine, Nagoya,  
17 Japan

18

19 **\* Corresponding authors**

20 Dariush Ghasemi-Semeskandeh  
21 Human Technopole, Health Data Science Centre  
22 V.le Rita Levi-Montalcini, 1, Milan 20157, Italy.  
23 email: [ghasemi.dariush@yahoo.com](mailto:ghasemi.dariush@yahoo.com)

24

25 Cristian Pattaro  
26 Eurac Research Institute for Biomedicine,  
27 Via Volta 21, 39100 Bolzano, Italy  
28 Tel. +39 0471 055 527  
29 email: [cristian.pattaro@eurac.edu](mailto:cristian.pattaro@eurac.edu)

30

31 **Abstract**

32 Chronic kidney disease (CKD) is a complex disease affecting >10% of the global population,  
33 with large between- and within-continent variability reflecting major environmental effects. To  
34 identify molecular targets for treatment, genome-wide association study meta-analyses  
35 (GWAMAs) of CKD-defining traits have identified hundreds of genetic loci in aggregated  
36 populations. However, while GWAMAs estimate the average allelic effect across studies, single  
37 population studies may be relevant to unravel specific mechanisms. To assess whether an  
38 individual study from a specific population could extend existing knowledge on kidney function  
39 genetics, we selected 147 kidney function relevant loci identified by a large GWAMA, assessing  
40 their association with the glomerular filtration rate estimated from serum creatinine (eGFR<sub>crea</sub>)  
41 in the Cooperative Health Research In South Tyrol (CHRIS) study (n=10,146), conducted in an  
42 Alpine region where thyroid dysfunction is common. We replicated associations with single  
43 nucleotide polymorphisms (SNPs) at 11 loci, showing up-to-5.4 times larger effect sizes than in  
44 the corresponding GWAMA, not explainable by minor allele frequency differences. Systematic  
45 mediation analysis across 70 quantitative traits identified serum magnesium and the activated  
46 partial thromboplastin time as partial mediators of the eGFR<sub>crea</sub> associations at *SHROOM3*  
47 and *SLC34A1*, respectively. Given that free triiodothyronine and thyroxine were effect modifiers  
48 across all loci, we conducted SNP-by-thyroid stimulating hormone (TSH) interaction analyses,  
49 identifying significant interactions at *STC1*: SNPs had larger effects on eGFR<sub>crea</sub> at higher TSH  
50 levels, possibly reflecting stanniocalcin-1 autocrine and paracrine role. Individual population  
51 studies can help characterize genetic associations. The interplay between phenotypes at  
52 *SHROOM3* and *SLC34A1* and the role of thyroid function as a genetic effect modifier warrant  
53 further investigations.

## 54 Introduction

55 Chronic kidney disease (CKD) is a common complex disease that increases the risk of kidney  
56 failure, cardiovascular mortality, and all-cause mortality<sup>1</sup>. CKD is predicted to become the 5<sup>th</sup>  
57 leading cause of death by 2040<sup>2</sup>. While affecting >10% of the population globally, CKD  
58 prevalence shows marked variability across countries, both between and within continents<sup>3</sup>.  
59 Between-continent variability is likely guided by combinations of environmental and genetic  
60 differences, such as is the case with sickle cell trait and albuminuria<sup>4</sup>. Within-continent variability  
61 could be more related to environmental differences, including lifestyle, public health policies,  
62 and assessment methods<sup>4,5</sup>.

63 To unravel the genetic basis of CKD, several genome-wide association study meta-  
64 analyses (GWAMAs) were conducted that analyzed the CKD-defining marker glomerular  
65 filtration rate (GFR) estimated from serum creatinine (eGFR<sub>crea</sub>)<sup>6</sup>. More than 400 associated  
66 loci have been identified to date<sup>7</sup>. While GWAMAs estimate the average allelic effect across  
67 studies, evidence shows that individual population studies may be extremely relevant to  
68 advance discoveries not only in the field of rare diseases<sup>8</sup> but also with regard to markers of  
69 common chronic diseases such as triglycerides and low density lipoprotein (LDL) cholesterol<sup>9</sup>,  
70 age at diabetes onset<sup>10</sup>, and mood disorders<sup>11</sup>. The evidence accumulated so far suggests that  
71 there is no general rule and that the contribution of an individual study to the dissection the  
72 biological basis of common phenotypes can only be assessed on a case-by-case basis and  
73 depends on study-specific characteristics.

74 Here we analyzed genetic associations with eGFR<sub>crea</sub> in the Cooperative Health  
75 Research In South Tyrol (CHRIS) study, a general population study conducted in a  
76 mountainous region of central Europe<sup>12</sup>. A sample from this same population was previously  
77 instrumental to identify genetic linkage of serum creatinine with the *APOL1-MYH9* locus<sup>13</sup> and to  
78 replicate common variants associated with eGFR<sub>crea</sub><sup>14</sup>. Characteristics of the geographical  
79 region where the study was conducted include the demographic stability over time<sup>15</sup>, the rural  
80 environment, and a high frequency of hypothyroidism, typical of mountainous regions<sup>16</sup>. Aim of  
81 our analysis was to assess whether this particular study sample could provide information that  
82 extend current knowledge on kidney function genetics, by leveraging on local specific  
83 characteristics.

84 For a given single population study, with a relatively limited sample size, to help expand  
85 knowledge on loci already identified by a large consortium, it is necessary and appropriate to  
86 limit investigation to the subset of loci associated with the phenotype of interest in the given

87 study. Thus, we first identified which, among 147 loci found to be associated with kidney  
88 function in a large GWAMA from the CKDGen Consortium<sup>17</sup>, were associated with eGFR<sub>crea</sub> in  
89 the CHRIS study. We take it for granted that the identified loci are relevant for kidney function.  
90 This was demonstrated by Wuttke et al.<sup>17</sup> through the combined analysis of eGFR<sub>crea</sub> and  
91 blood urea nitrogen (BUN).

92 CKDGen loci were tested for replication in the CHRIS study. The purpose of this  
93 replication analysis was to identify those loci for which a study of limited size can try to  
94 contribute new information. The replication analysis should not be understood as a  
95 demonstration of the validity of the analyzed loci: in fact, they should already be considered as  
96 validated, as they were identified through a large meta-analysis on more than 760,000  
97 individuals and replicated in an independent study on more than 280,000 individuals<sup>17</sup>.

98 The features of the replicated loci were then assessed against corresponding European  
99 ancestry CKDGen GWAMA results. Given that several replicated loci showed larger effects in  
100 CHRIS than in the European ancestry CKDGen GWAMA, we examined the possible reasons of  
101 such larger effects and conducted extensive mediation analyses across 70 quantitative  
102 biochemical and anthropometric traits, reflecting multiple health conditions. Further, after  
103 observing a systematic modification of the genetic association with eGFR<sub>crea</sub> when adjusting  
104 for thyroid-related traits, we conducted SNP-by-thyroid function interaction analyses. The aims  
105 of the mediation and the interaction analyses were to identify the presence of intermediate  
106 phenotypes and effect modifiers, respectively, that were specific of the study sample.

107

108

## 109 **Methods**

### 110 *Study sample*

111 The present analyses were based on 10,146 individuals with complete genotype data who  
112 participated to the CHRIS study, a population-based study conducted in South Tyrol, Italy,  
113 between 2011 and 2018, which has been extensively described elsewhere<sup>12,18</sup>. Briefly,  
114 participants underwent blood drawing, urine collection, anthropometric measurements, and  
115 clinical assessments early in the morning, after an overnight fast. Medical history was  
116 reconstructed through interviewer- and self-administered standardized questionnaires. Drug  
117 treatment was identified via barcode scan of the drug containers that participants were

118 instructed to bring at the study center visit, with the drug code linked to an official drug  
119 databank<sup>12</sup>.

120

### 121 *eGFRcrea estimation*

122 Serum creatinine (SCr) was measured with a colorimetric assay on Roche Modular PPE  
123 (n=4,176) and Abbott Diagnostic Architect c16000 (n=5,970) instrumentations. According to  
124 previous analyses<sup>18</sup>, SCr was normalized by instrument (fixed effect) and participation period  
125 (random effect) using a linear mixed model implemented in the 'lme4' R package<sup>19</sup>. To reflect  
126 the same methods used in the CKDGen Consortium meta-analysis<sup>17</sup>, eGFRcrea was estimated  
127 using the 2009 CKD-EPI equation<sup>20</sup> implemented in the R package 'nephro' v1.2<sup>21</sup> and  
128 winsorized at 15 and 200 ml/min/1.73m<sup>2</sup>. Finally, the natural logarithm transformation was  
129 applied.

130

### 131 *Genotype imputation*

132 Genotyping was performed in two batches using genotyping array chips based on the Illumina  
133 Human Omni Microarray platform. Following quality control analysis, the two data batches were  
134 merged, phased with SHAPEIT2 v2.r837, using the duoHMM method (--duohmm -W 5) with 800  
135 states and 30 rounds<sup>22</sup>, and imputed based on the TOPMed r2 standard reference panel on the  
136 Michigan Imputation Server<sup>23</sup>. We obtained 34,084,280 SNPs with minimum imputation quality  
137 index Rsq  $\geq 0.3$  and minor allele count (MAC)  $\geq 1$ , aligned with human genome assembly  
138 GRCh38. A genetic relatedness matrix (GRM) was obtained from the autosomal genotyped  
139 SNPs using EFACTS v3.2.6 modified for compatibility with assembly GRCh38. Genetic principal  
140 components (PCs) were estimated based on the genotyped variants data with minor allele  
141 frequency (MAF)  $> 0.05$  using GCTA<sup>24</sup>.

142

### 143 *GWAS and replication of CKDGen results*

144 We conducted a GWAS of age- and sex-adjusted residuals of  $\ln(\text{eGFRcrea})$  using the EMMAX  
145 method<sup>25</sup> implemented in EFACTS v3.2.6, including the GRM to model the sample structure.  
146 After the analysis, we removed variants with minor allele frequency (MAF)  $< 0.005$ , leaving  
147 10,158,100 SNPs for further characterization. Genomic inflation was assessed on this set of  
148 SNPs by estimating the genomic control factor  $\lambda$ <sup>26</sup>.

149 Summary statistics from the CKDGen GWAMA<sup>17</sup> were downloaded from the publicly  
150 available repository at <https://ckdgen.imbi.uni-freiburg.de>. Given a subset of 4661 samples from  
151 the CHRIS study was previously included in the CKDGen GWAMA, we removed CHRIS data  
152 from the CKDGen GWAMA summary statistics using MetaSubtract v1.60<sup>27</sup>. After this process,  
153 the 147 variants remained genome-wide significant in CKDGen. We lifted the CKDGen genomic  
154 positions from the GRCh37 to the GRCh38 map, using CrossMap v0.5.3<sup>28</sup>.

155 Given small-sample bias can cause large effect estimate fluctuations and given it is not  
156 uncommon to observe small studies with opposite effect directions compared to the meta-  
157 analysis results in large GWAMAs, to ensure consistency of the effect direction, replication of  
158 the 147 loci in CHRIS was assessed based on a one-sided test, following common practice<sup>29</sup>.  
159 Statistical significance for replication was set at the Bonferroni-corrected level of 0.00034  
160 (0.05/147), for any variant in strong linkage disequilibrium (LD) ( $r^2 > 0.8$ ) with a lead CKDGen  
161 variant. LD was estimated using emeraLD v0.1<sup>30</sup>. The total number of SNPs within the 147 loci  
162 was 6337. Adjustment for 147 independent tests covers 98% of the SNP's cumulative variance  
163 (Suppl. Fig. S1).

164 We compared the variance of  $\ln(\text{eGFR}_{\text{crea}})$  explained by SNPs in CHRIS versus  
165 CKDGen European Ancestry analysis as  $b^2 \times 2 \times p \times (1-p) / \text{var}(y)$ ,<sup>17</sup> where  $y$  is the age- and  
166 sex-adjusted  $\ln(\text{eGFR}_{\text{crea}})$ ,  $b$  is the estimated SNP effect on  $y$ , and  $p$  is the SNP allele  
167 frequency. For comparison with CKDGen, we used a common estimate of  $\text{var}(y)$ .

168

### 169 *Mediation analysis*

170 Replicated  $\ln(\text{eGFR}_{\text{crea}})$ -SNP associations were submitted to mediation analysis across 70  
171 quantitative anthropometric, blood pressure and biochemical traits (listed in Supplementary  
172 Table S1), following the flowchart represented in Fig. 1. SCr was included as a positive control.  
173 To prevent potential measurement instrument effects<sup>18</sup>, we applied quantile normalization to  
174 each trait (Supplementary Methods). Mediation analysis was conducted for each *trait* following  
175 the established 4-step framework and fitting linear regression models throughout:

- 176 1.  $\ln(\text{eGFR}_{\text{crea}}) = \beta_1 \text{SNP} + \sum_i^K \theta_i X_i + \varepsilon_1$
- 177 2.  $\ln(\text{eGFR}_{\text{crea}}) = \beta_2 \text{SNP} + \gamma_2 \text{trait} + \sum_i^K \theta_i X_i + \varepsilon_2$
- 178 3. *trait* =  $\beta_3 \text{SNP} + \sum_i^K \theta_i X_i + \varepsilon_3$
- 179 4. *trait* =  $\beta_4 \text{SNP} + \delta_4 \ln(\text{eGFR}_{\text{crea}}) + \sum_i^K \theta_i X_i + \varepsilon_4$

180 where  $\varepsilon_1$  to  $\varepsilon_4$  refer to Gaussian error terms, and  $\sum_i^K \theta_i X_i$  indicate the inclusion of K covariates,  
181 namely age, sex, the first 10 genetic PCs and an intercept. Fitting PC-adjusted simple linear  
182 models came for practical reasons after observing substantial equivalence with kinship-adjusted  
183 linear mixed models obtained with EMMAX (Suppl. Fig. S2). Step 1 corresponds to the GWAS  
184 results: given we investigated the replicated variants, the estimate of  $\beta_1$  ( $b_1$ ) was always  
185 significant. We defined two criteria that must be satisfied in order for a trait to be considered at  
186 least a partial mediator of the  $\ln(\text{eGFRcrea})$ -SNP relation: (i) evidence of a substantial alteration  
187 of the SNP effect on  $\ln(\text{eGFRcrea})$  when adjusting for the trait (step 2) and (ii) the SNP is also  
188 associated with the trait (step 3) at a multiple-testing corrected level of  $6.5 \times 10^{-5}$  (corresponding  
189 to  $0.05 / (70 \text{ traits} \times 11 \text{ independent loci})$ ). Criterion (i) was assessed by analyzing the  
190 distribution of the estimate of  $\beta_2$  ( $b_2$ ): for a specific trait,  $b_2$  was classified as an outlier according  
191 to the rule  $b_2 < P_{10} - 1.5 (P_{75} - P_{25})$  or  $b_2 > P_{90} + 1.5 (P_{75} - P_{25})$ , where  $P_{10}$ ,  $P_{25}$ ,  $P_{75}$ , and  $P_{90}$   
192 indicate the 10<sup>th</sup>, 25<sup>th</sup>, 75<sup>th</sup>, and 90<sup>th</sup> percentile of the distribution of  $b_2$  across all traits. This is  
193 essentially a more stringent version of the Tukey's rule for outlier detection.

194 To support associations in step 3 and assess whether lack of mediation evidence might  
195 have been due to lack of power in the CHRIS study, we interrogated the SNP-trait associations  
196 that passed the mediation analysis step 2 on the PhenoScanner v2<sup>31</sup>  
197 (<http://www.phenoscaner.medschl.cam.ac.uk/>, interrogated on 31-Jan-2023), GWAS Catalog<sup>32</sup>  
198 (<https://www.ebi.ac.uk/gwas/docs/api>; 24-Feb-2023), and the ThyroidOmics Consortium  
199 summary statistics<sup>33</sup> (<https://www.thyroidomics.com>), for associations at  $P < 5.0 \times 10^{-8}$ .

200

### 201 *Interaction with thyroid function phenotypes*

202 Free triiodothyronine (FT3) and thyroxine (FT4) did not qualify as mediators but caused a major  
203 departure of  $b_2$  from its expected distribution (see Results). These traits may reflect thyroid  
204 gland issues. Given the relevance of hypothyroidism in the region, we analyzed these traits in  
205 detail. Because FT3 and FT4 levels were measured only when the thyroid stimulating hormone  
206 (TSH) level was  $< 0.4$  or  $> 3.8$   $\mu\text{UI/mL}$ , we started with a series of sensitivity analyses to exclude  
207 the presence of underlying sample stratification. First, we tested whether the distribution of the  
208 genotypes was different between individuals with and without measured FT3 and FT4, using a  
209 two-sided Wilcoxon test at a significance level of 0.05. Second, to verify whether there was any  
210 geographical cluster of individuals with extreme TSH levels (and so measured FT3 and FT4

211 levels), which might have implicated population stratification, we further adjusted the  
212  $\ln(\text{eGFR}_{\text{crea}})$ -SNP association model in step 1 for municipality of residence.

213 After excluding the presence of sample stratification, we tested the interaction between  
214 the SNPs and TSH levels, which were measured in the whole sample. We excluded 416  
215 individuals reporting any condition among thyroid cancer (n=16), kidney cancer (n=1), goiter  
216 (n=277), having undergone surgery to the thyroid gland (n=312), or having both missing TSH  
217 measurement and therapy information (n=4), leaving 9730 individuals for the interaction  
218 analysis. Regression models included both the main and the interaction effect terms and were  
219 adjusted for the same covariates used in the step 1 model. In addition to quantitative TSH  
220 levels, we also tested interaction with broadly defined hyper- and hypothyroidism.  
221 Hyperthyroidism was defined as a TSH level of  $<0.4 \mu\text{UI/mL}$  or use of thiamazole (n=3) or  
222 propylthiouracil (n=1). Hypothyroidism was defined as TSH levels  $>3.8 \mu\text{UI/mL}$  or reported use  
223 of levothyroxine sodium (n=512). No other thyroid-related treatment was reported. The  
224 statistical significance level for interaction testing was set at  $0.0045=0.05/11$  loci.

225

#### 226 *Ethical statement*

227 The CHRIS study was approved by the Ethical Committee of the Healthcare System of the  
228 Autonomous Province of Bozen/Bolzano, protocol no. 21/2011 (19 Apr 2011). All participants  
229 provided written informed consent. All the methods were performed in strict accordance with the  
230 approved protocol.

231

232

#### 233 **Results**

234 In the study sample, the median age was 46.5 years (interquartile range, IQR: 32.8, 57.6),  
235 females were 55%, and the median  $\text{eGFR}_{\text{crea}}$  level was 92.2 (IQR: 81.3, 103.2)  $\text{ml/min/1.73m}^2$   
236 (Table 1; additional clinical characteristics are reported in Supplementary Table S1). The  
237 genomic inflation factor  $\lambda$  was 1.02 (Supplementary Fig. S3), supporting appropriate modeling of  
238 the sample structure. In the CHRIS study, we replicated 11 of the 147 CKDGen loci (1-sided P-  
239 value between  $3.34 \times 10^{-4}$  and  $4.51 \times 10^{-7}$ ; Table 2; Supplementary Table S2). For each replicated  
240 locus we retained all variants in LD ( $r^2 > 0.8$ ) with the lead SNP, totaling 163 variants over the 11  
241 loci (Supplementary Table S3). The replicated loci displayed very similar LD pattern in CHRIS



242 as in CKDGen (Supplementary Fig. S4). Among the replicated loci, variants' effect magnitude at  
243 *CASZ1*, *DDX1*, *PIP5K1B*, *GAB2*, and *IGF1R* was significantly larger in CHRIS than in CKDGen  
244 (Fig. 2; Table 2). Replicated loci explained a larger proportion of  $\ln(\text{eGFR}_{\text{crea}})$  variance (Fig.  
245 3A). For them, the CHRIS-to-CKDGen effect ratio varied between 1 and 5.4 and was not  
246 entirely explained by MAF differences (Fig. 3B; Supplementary Table S3). Three broad  
247 groupings characterized by MAF and effect size differences could be distinguished:

- 248 1. at *PDILT*, *SLC34A1* and *SHROOM3*, we observed similar effect magnitude,  
249 regardless whether MAF was similar or lower in CHRIS, consistent with the fact that  
250 these loci were amongst the ones identified by the earliest GWAS thanks to their  
251 large effects, counterbalancing the still relatively limited sample sizes and the  
252 coarser genomic imputation<sup>34,35</sup>;
- 253 2. at *DAB2*, *TMEM60*, *PIP5K1B*, *GAB2*, and *IGF1R*, MAF was similar or slightly larger  
254 in CHRIS (effects in CHRIS were 1.3-to-5 times larger than in CKDGen);
- 255 3. at *STC1*, *DDX1*, and *CASZ1*, we observed lower MAF in CHRIS and about 1.8-to-5.4  
256 times larger effects in CHRIS.

257

### 258 *Mediation analysis*

259 To identify possible reasons for observed larger effects, we conducted mediation  
260 analysis of the association between  $\text{eGFR}_{\text{crea}}$  and each of the 163 variants at the 11 loci  
261 across 70 quantitative health traits available in CHRIS (Supplementary Table S1). Adjusting the  
262  $\text{eGFR}_{\text{crea}}$ -SNP association for each trait in turn (step 2 of the mediation analysis; see Methods)  
263 resulted in a substantial change of the  $\text{eGFR}_{\text{crea}}$ -SNP association coefficient in between 4 and  
264 12 variant-trait associations per locus (Fig. 4). As expected, adjustment for the positive control  
265 SCr resulted in the nearly complete knockdown of the SNP effect at all loci (effect sizes toward  
266 null). Often, the effect change happened when adjusting for traits closely related to kidney  
267 function such as urate (*GAB2*, *IGF1R*, *PIP5K1B*, *SHROOM3*, and *STC1*), urinary creatinine  
268 (*SHROOM3*), and urinary albumin (*IGF1R*). Serum electrolytes such as magnesium  
269 (*SHROOM3*), corrected calcium (*PIP5K1B*), and sodium (*DDX1*, *GAB2*, *RSBN1L*, *SLC34A1*,  
270 and *STC1*) were also widely involved. Of note was the widespread effect of FT3, FT4, and the  
271 activated partial thromboplastin time (aPTT) across all loci. When adjusting for FT3 and FT4,  
272 the SNP effect on  $\ln(\text{eGFR}_{\text{crea}})$  was often completely altered. This behavior across all loci may  
273 suggest the presence of either a measurement artifact or a general related condition affecting  
274 the underlying population.

275 In the 3<sup>rd</sup> mediation analysis step, we evaluated whether the traits identified as effect  
276 modifiers in step 2 were associated with the SNP of which they modified the association with  
277 eGFRcrea (Table 3; Supplementary Table S4): SNP rs3812036 at *SLC34A1* was associated  
278 with aPTT levels ( $P=3.54\times 10^{-6}$ ) and five SNPs at *SHROOM3* were associated with serum  
279 magnesium levels ( $P$ -values between  $5.33\times 10^{-5}$  and  $6.12\times 10^{-5}$ ). Three of our associations  
280 reproduce previously reported associations found using Phenoscanner: at *SLC34A1*, Tang *et*  
281 *al.*<sup>36</sup> reported association of rs3812036 with aPTT ( $P=2.88\times 10^{-18}$ ), and at *SHROOM3*, Meyer *et*  
282 *al.*<sup>37</sup> reported associations of magnesium with rs4859682 ( $P=2.39\times 10^{-9}$ ) and rs13146355  
283 ( $P=6.27\times 10^{-13}$ ).

284 All SNP-trait associations that successfully passed the 2<sup>nd</sup> mediation analysis step were  
285 also interrogated against the Phenoscanner, GWAS Catalog, and ThyroidOmics consortium  
286 results (Methods) to identify associations that might have been missed by the 3<sup>rd</sup> step of the  
287 mediation analysis in CHRIS due to lack of power (Supplementary Table S4). This analysis  
288 provided additional support for our findings at *SHROOM3* and *SLC34A1*. It also showed  
289 associations between variants at *IGF1R* and serum urate, unraveling how this additional  
290 mediation was likely missed by our study due to lack of power.

291 In summary, aPTT and serum magnesium qualified as partial mediators of eGFRcrea  
292 associations at *SLC34A1* and *SHROOM3*, respectively. At *SLC34A1*, the partial mediation of  
293 aPTT corresponded to a 21% larger effect of the SNP on  $\ln(\text{eGFRcrea})$ ; at *SHROOM3*, the  
294 partial mediation of serum magnesium corresponded to an effect attenuation of about 11%  
295 (Table 3).

296

### 297 *Interaction with thyroid dysfunction*

298 Despite the pervasive effect change caused by FT3 and FT4 in step-2 analysis, these  
299 two traits did not result in significant step-3 associations and so did not qualify as mediators.  
300 However, motivated by the fact that hypothyroidism is amongst the leading causes of  
301 hospitalization in the region<sup>38</sup>, thyroid problems are common in Alpine areas, and the FT3 and  
302 FT4 adjustment affected the SNP-eGFRcrea association at all 11 loci, we considered it  
303 reasonable to evaluate a potential role of thyroid dysfunction as a modifier of the SNP-  
304 eGFRcrea relation.

305 Preliminarily, given that in the CHRIS study FT3 and FT4 were measured only in  
306 individuals with TSH levels above or below specific thresholds, we first conducted sensitivity

307 analyses to exclude the presence of artifacts. We did not observe any genotype stratification by  
308 measured versus unmeasured FT3 and FT4 levels (Supplementary Table S5). Adjustment for  
309 municipality of residence, to exclude the presence of local clusters of thyroid dysfunction, did  
310 not alter the results (Supplementary Fig. 3).

311 We then tested for the presence of linear interaction with TSH levels, which identified a  
312 significant interaction at *STC1* SNPs rs819185 ( $P=0.00177$ ) and rs819196 ( $P=0.00154$ ) both  
313 below the multiple testing threshold of 0.05/11 loci (Fig. 5; Supplementary Table S6). Since for  
314 most loci the adjustment both for FT3 and for FT4 were causing an increase of the effect  
315 magnitude, we also postulated a U-shaped interaction model, with the SNP effects on  
316  $\ln(\text{eGFR}_{\text{crea}})$  being larger both in hyper- and in hypothyroidism. We classified individuals as  
317 healthy (88.5%), with hypothyroidism (9.8%) or with hyperthyroidism (1.7%; Table 1). Our  
318 results did not support the presence of interaction with these two conditions at any locus after  
319 multiple testing control, despite nominally significant  $P$ -values observed for interaction between  
320 hyperthyroidism and SNPs in *SHROOM3* ( $P=0.03$ ) and between hypothyroidism and SNPs in  
321 *PIP5K1B* ( $P=0.03$ ) and *GAB2* ( $P=0.04$ ; Supplementary Table S7).

322

323

## 324 Discussion

325 Out of 147 loci known for their association with kidney function, our analysis identified 11 that  
326 were more strongly associated with  $\text{eGFR}_{\text{crea}}$  in the CHRIS study sample than in average  
327 European-ancestry population samples. Observed differences in effect magnitude could not be  
328 explained by MAF differences. Extensive mediation analysis identified serum magnesium and  
329 aPTT as partial mediators of the association of  $\text{eGFR}_{\text{crea}}$  with variants at *SHROOM3* and  
330 *SLC34A1*, respectively. SNP-by-TSH interaction analysis highlighted that the effect of variants  
331 at *STC1* on  $\text{eGFR}_{\text{crea}}$  would vary by TSH levels.

332 *PDILT*, *SHROOM3*, and *SLC34A1* explained the same fraction of  $\text{eGFR}_{\text{crea}}$  variance in  
333 CHRIS and CKDGen. In fact, for these loci, a relatively small sample size proved to be sufficient  
334 for their detection.<sup>34,35</sup> For all other replicated loci, the replication was due to an interplay  
335 between allele frequency differences and effect estimate differences. Figures 3A and 3B show  
336 that the effect difference was probably more relevant than the allele frequency differences. But  
337 the question remains as to why precisely these and no other loci showed such large effects in  
338 CHRIS. SNPs may show larger effects at lower MAF. This situation was observed at just *DDX1*

339 and *CASZ1*, whose effects got larger as the CHRIS-to-CKDGen MAF ratio got smaller.  
340 However, the MAF difference was unrelated to the larger effects observed at *GAB2*, *TMEM60*,  
341 *PIP5K1B*, and *IGF1R*: in these cases, MAF was similar or even smaller in CKDGen than in  
342 CHRIS (Fig. 2). The presence of un-modeled sample structure causing genomic inflation can be  
343 excluded as relatedness was appropriately modeled. We can also exclude the presence of  
344 genetic admixture as the CHRIS study was conducted in a small area, on a local population of  
345 homogeneous ancestry<sup>15,39-41</sup>. Differences in LD structure between the local and the overall  
346 population can also be excluded as regional association plots showed similar shapes between  
347 CHRIS and CKDGen. Epistatic gene-gene interaction would be a further possibility that cannot  
348 be easily verified. The most plausible remaining reasons would be the presence of local effect  
349 modifiers or gene-by-environment interaction.

350 Mediation analysis based on commonly measured biochemical and anthropometric traits  
351 provided limited answers in this respect but highlighted relevant mechanisms at three loci. The  
352 first finding was the identified presence of partial mediation of serum magnesium levels on the  
353 association of eGFRcrea with the *SHROOM3* locus. Genetic variants at *SHROOM3* have been  
354 associated with CKD<sup>8,9</sup>, reduced eGFRcrea<sup>9</sup>, increased albumin-to-creatinine ratio<sup>10</sup>, and low  
355 serum magnesium levels<sup>11</sup>. *SHROOM3* is necessary to maintain the glomerular filtration barrier  
356 integrity<sup>42</sup>. Variants in this gene have been shown to be associated with increased risk of CKD<sup>43</sup>,  
357 likely through disruption of the transcription factor TCF7L2 in podocyte cells. At the same locus,  
358 *FAM47E* is highly expressed in the human glomeruli and its transcription levels were associated  
359 with eGFRcrea<sup>44</sup>. Additional genes in this locus include *STBD1* and *CCDC158*, for which no  
360 connection with kidney function was demonstrated. The kidney is one of the primary regulators  
361 of serum magnesium concentration. It has been suggested that *SHROOM3* might increase  
362 serum magnesium levels through decreased renal magnesium excretion<sup>45</sup>. At the *SHROOM3*  
363 rs13146355 SNP, a previous GWAS on serum magnesium showed that the allele related with  
364 lower magnesium levels was associated with higher eGFRcrea levels<sup>37</sup>. The authors observed  
365 that the rs13146355 association with magnesium did not change when adjusting for eGFRcrea,  
366 suggesting pleiotropic independent effects. In our analysis, the effect of rs13146355 on  
367 eGFRcrea was reduced by 11% when adjusting for magnesium, suggesting partial mediation  
368 and thus the presence of partially overlapping mechanisms. Deeper analyses at the molecular  
369 level are warranted to identify the relevant mechanism.

370 A similar situation was observed at *SLC34A1* which, in addition to eGFRcrea<sup>35</sup>, was also  
371 previously associated with aPTT in a GWAMA of 9240 European ancestry individuals<sup>36</sup>

372 including participants from the same areas where the CHRIS study was sampled. By quantifying  
373 the clotting time from the activation of factor XII (*F12*) until the formation of a fibrin clot, aPTT is  
374 a measure of the integrity of the intrinsic and common coagulation pathways<sup>46</sup>. The observed  
375 mediation, in which adjustment for aPTT increases the SNP-eGFR<sub>crea</sub> association by 21%, is  
376 likely related to the proximity between *SLC34A1* and the coagulation factor gene *F12*. The LD at  
377 this locus is compatible with the presence of common haplotypes tagging both genes, inducing  
378 pleiotropic independent effects on eGFR<sub>crea</sub> and aPTT.

379 An additional relevant finding was the presence of an interaction between two SNPs at  
380 *STC1* and TSH levels. *STC1* is chiefly expressed in the thyroid gland  
381 (<https://gtexportal.org/home/gene/STC1>). It is also highly expressed in the kidney, where it is  
382 exclusively expressed in the collecting duct<sup>47</sup>. *STC1* is also associated with urate and urea  
383 levels and is involved in the physiological response to dehydration and protein-rich diet<sup>47</sup>. The  
384 *STC1*-encoded protein, stanniocalcin-1, is involved in phosphate reabsorption in the kidney  
385 proximal tubules and in multiple pathophysiological mechanisms including ischemic kidney  
386 injury<sup>48</sup>. Stanniocalcin-1 behaves like an endocrine hormone, with both autocrine and paracrine  
387 functions. Our observation of a SNP effect on eGFR<sub>crea</sub> depending on TSH levels would  
388 suggest that altered thyroid function levels might affect Stanniocalcin-1 levels with consequent  
389 effect on kidney function phenotypes. While intriguing, this hypothesis should be verified by  
390 appropriate studies.

391 Our study had both strengths and limitations. The exploitation of a local homogeneous  
392 population sample was at the same time the most relevant strength and limitation of our study.  
393 While homogeneity may enhance the identification of specific genetic associations, it may also  
394 hinder the possibility to explain the underlying reasons of the observed larger effects compared  
395 to the average European ancestry population. The availability of a very large number of clinical  
396 parameters has allowed us to conduct an exhaustive and unbiased screening on potential  
397 mediation mechanisms. However, the analyzed traits lacked molecular specificity to identify  
398 underlying mechanisms. Given eGFR<sub>crea</sub> might also reflect creatinine metabolism in addition to  
399 kidney function, another limitation of our study was the lack of additional kidney function  
400 markers such as cystatin C or BUN. For this reason, we limited our analyses to loci whose  
401 kidney function relevance was already demonstrated by previous studies<sup>17</sup>. Assessment of  
402 environmental exposures may also be a tool to explain the observed larger effects and would be  
403 ground for further research. Finally, our choice of adjustment for multiple testing is also worth a  
404 brief reflection. We have considered 147 loci that were already validated by at least one

405 previous study through discovery and independent replication<sup>17</sup>, with several of the same loci  
406 being already reported previously<sup>35,49,50</sup>. As such, they wouldn't need further validation and a  
407 nominal significance level would be commonly accepted for statistical testing. However, the aim  
408 of our analysis was to identify loci with substantially large effect in the CHRIS study to further  
409 conduct meaningful mediation analyses. For this reason, we choose to penalize the significance  
410 level by the number of independent loci. To us, this seems to be a sensible choice also by the  
411 fact that 147 independent tests covered 98% of the total variability of the 6337 SNPs in strong  
412 LD observed overall.

413 In conclusion, our study has identified loci that have particularly large effect on kidney  
414 function in the specific Alpine population sample considered in this study, namely *CASZ1*,  
415 *IGF1R*, *GAB2*, and *DDX1*. Allele frequency and LD analyses did not explain the observed larger  
416 effects nor did the extensive mediation analyses conducted. On the other hand, we showed that  
417 the effects of *SHROOM3* and *SLC34A1* on eGFR<sub>crea</sub> are partially mediated by serum  
418 magnesium and aPTT, respectively. Finally, the observed *STC1*-TSH interaction implicates  
419 thyroid function involvement, which might have been favored by the high burden of thyroid-  
420 related diseases in the region. Further investigations are warranted to link the identified loci to  
421 environmental and molecular characteristics of specific population samples that may elucidate  
422 mechanisms not otherwise identifiable in large genetic meta-analyses.

423

424

## 425 **Acknowledgements**

426 The CHRIS study is a collaborative effort between the Eurac Research Institute for Biomedicine  
427 and the Healthcare System of the Autonomous Province of Bozen/Bolzano. We thank all study  
428 participants, the general practitioners, the personnel of the Hospital of Schlanders/Silandro, the  
429 field study team and the personnel of the CHRIS Biobank (BRIF code BRIF6107) for their  
430 support and collaboration. Extensive acknowledgement is reported at  
431 <https://translationalmedicine.biomedcentral.com/articles/10.1186/s12967-015-0704-9>.

432

## 433 **Funding**

434 The CHRIS study was funded by the Department of Innovation, Research and University of the  
435 Autonomous Province of Bolzano-South Tyrol. This work was carried out within the TrainCKDis

436 project, funded by the European Union's Horizon 2020 research and innovation programme  
437 under the Marie Skłodowska-Curie grant agreement H2020-MSCA-ITN-2019 ID:860977  
438 (TrainCKDis).

439

#### 440 **Author contributions**

441 Conceptualization of the research project: C.P., D.G.

442 Recruitment and study management: M.G., P.P.P., C.P.

443 Bioinformatics: D.G., E.K., D.E., C.F.

444 Data quality control and harmonization: M.G., L.B., L.F.

445 Statistical Analysis: D.G.

446 Results interpretation: D.G., C.P., C.F., D.E., M.G., L.F., R.F.

447 Manuscript drafting: D.G., C.P.

448 Manuscript critical revision: D.E., E.K., R.F., L.F., L.B., M.G., D.J.M.P., P.P.P., C.F., D.G., C.P.

449

450

#### 451 **Competing interests**

452 The authors declare no competing financial interests.

453

#### 454 **Data availability**

455 The data used in the current analyses can be requested with an application to

456 [access.request.biomedicine@eurac.edu](mailto:access.request.biomedicine@eurac.edu) at the Eurac Research Institute for Biomedicine.

457 **References**

458

- 459 1 Eckardt, K. U. *et al.* Evolving importance of kidney disease: from subspecialty to  
460 global health burden. *Lancet* **382**, 158-169, doi:10.1016/s0140-6736(13)60439-0  
461 (2013).
- 462 2 Foreman, K. J. *et al.* Forecasting life expectancy, years of life lost, and all-cause  
463 and cause-specific mortality for 250 causes of death: reference and alternative  
464 scenarios for 2016-40 for 195 countries and territories. *Lancet* **392**, 2052-2090,  
465 doi:10.1016/s0140-6736(18)31694-5 (2018).
- 466 3 Global, regional, and national burden of chronic kidney disease, 1990-2017: a  
467 systematic analysis for the Global Burden of Disease Study 2017. *Lancet* **395**,  
468 709-733, doi:10.1016/s0140-6736(20)30045-3 (2020).
- 469 4 Naik, R. P. *et al.* Association of sickle cell trait with chronic kidney disease and  
470 albuminuria in African Americans. *Jama* **312**, 2115-2125,  
471 doi:10.1001/jama.2014.15063 (2014).
- 472 5 Brück, K. *et al.* CKD prevalence varies across the European general population.  
473 *Journal of the American Society of Nephrology* **27**, 2135,  
474 doi:10.1681/ASN.2015050542 (2016).
- 475 6 Köttgen, A. & Pattaro, C. The CKDGen Consortium: ten years of insights into the  
476 genetic basis of kidney function. *Kidney Int* **97**, 236-242,  
477 doi:10.1016/j.kint.2019.10.027 (2020).
- 478 7 Stanzick, K. J. *et al.* Discovery and prioritization of variants and genes for kidney  
479 function in >1.2 million individuals. *Nature Communications* **12**, 4350,  
480 doi:10.1038/s41467-021-24491-0 (2021).
- 481 8 Chong, Jessica X. *et al.* The Genetic Basis of Mendelian Phenotypes:  
482 Discoveries, Challenges, and Opportunities. *The American Journal of Human*  
483 *Genetics* **97**, 199-215, doi:<https://doi.org/10.1016/j.ajhg.2015.06.009> (2015).
- 484 9 Montasser, M. E. *et al.* Genetic and functional evidence links a missense variant  
485 in B4GALT1 to lower LDL and fibrinogen. *Science* **374**, 1221-1227,  
486 doi:10.1126/science.abe0348 (2021).
- 487 10 Hamet, P. *et al.* PROX1 gene CC genotype as a major determinant of early onset  
488 of type 2 diabetes in slavic study participants from Action in Diabetes and  
489 Vascular Disease: Preterax and Diamicron MR Controlled Evaluation study. *J*  
490 *Hypertens* **35 Suppl 1**, S24-s32, doi:10.1097/hjh.0000000000001241 (2017).
- 491 11 Humphries, E. M. *et al.* Genome-wide significant risk loci for mood disorders in  
492 the Old Order Amish founder population. *Mol Psychiatry*, doi:10.1038/s41380-  
493 023-02014-1 (2023).
- 494 12 Pattaro, C. *et al.* The Cooperative Health Research in South Tyrol (CHRIS)  
495 study: rationale, objectives, and preliminary results. *Journal of Translational*  
496 *Medicine* **13**, 348, doi:10.1186/s12967-015-0704-9 (2015).
- 497 13 Pattaro, C. *et al.* Genome-wide linkage analysis of serum creatinine in three  
498 isolated European populations. *Kidney International* **76**, 297-306,  
499 doi:<https://doi.org/10.1038/ki.2009.135> (2009).
- 500 14 Pattaro, C. *et al.* A meta-analysis of genome-wide data from five European  
501 isolates reveals an association of COL22A1, SYT1, and GABRR2 with serum



- 502 creatinine level. *BMC Medical Genetics* **11**, 41, doi:10.1186/1471-2350-11-41  
503 (2010).
- 504 15 Gögele, M., Pattaro, C., Fuchsberger, C. & Pramstaller, P. P. Fertility pattern and  
505 family structure in three Alpine settlements in South Tyrol (Italy): marriage  
506 cohorts from 1750 to 1949. *J Biosoc Sci* **41**, 697-701,  
507 doi:10.1017/s0021932009003423 (2009).
- 508 16 Bolzano, A. o. t. P. o. *Prevalence of the Chronic Diseases*,  
509 <[https://www.provincia.bz.it/salute-benessere/osservatorio-salute/diffusione-](https://www.provincia.bz.it/salute-benessere/osservatorio-salute/diffusione-delle-malattie-croniche.asp)  
510 [delle-malattie-croniche.asp](https://www.provincia.bz.it/salute-benessere/osservatorio-salute/diffusione-delle-malattie-croniche.asp)> (2019).
- 511 17 Wuttke, M. *et al.* A catalog of genetic loci associated with kidney function from  
512 analyses of a million individuals. *Nat Genet* **51**, 957-972, doi:10.1038/s41588-  
513 019-0407-x (2019).
- 514 18 Noce, D. *et al.* Sequential recruitment of study participants may inflate genetic  
515 heritability estimates. *Hum Genet* **136**, 743-757, doi:10.1007/s00439-017-1785-8  
516 (2017).
- 517 19 Bates, D., Mächler, M., Bolker, B. & Walker, S. Fitting Linear Mixed-Effects  
518 Models Using lme4. *Journal of Statistical Software* **67**, 1 - 48,  
519 doi:10.18637/jss.v067.i01 (2015).
- 520 20 Levey, A. S. *et al.* A new equation to estimate glomerular filtration rate. *Ann*  
521 *Intern Med* **150**, 604-612, doi:10.7326/0003-4819-150-9-200905050-00006  
522 (2009).
- 523 21 Pattaro, C. *et al.* Estimating the glomerular filtration rate in the general population  
524 using different equations: effects on classification and association. *Nephron Clin*  
525 *Pract* **123**, 102-111, doi:10.1159/000351043 (2013).
- 526 22 O'Connell, J. *et al.* A general approach for haplotype phasing across the full  
527 spectrum of relatedness. *PLoS Genet* **10**, e1004234,  
528 doi:10.1371/journal.pgen.1004234 (2014).
- 529 23 Das, S. *et al.* Next-generation genotype imputation service and methods. *Nat*  
530 *Genet* **48**, 1284-1287, doi:10.1038/ng.3656 (2016).
- 531 24 Yang, J., Lee, S. H., Goddard, M. E. & Visscher, P. M. GCTA: a tool for genome-  
532 wide complex trait analysis. *Am J Hum Genet* **88**, 76-82,  
533 doi:10.1016/j.ajhg.2010.11.011 (2011).
- 534 25 Kang, H. M. *et al.* Variance component model to account for sample structure in  
535 genome-wide association studies. *Nat Genet* **42**, 348-354, doi:10.1038/ng.548  
536 (2010).
- 537 26 Devlin, B. & Roeder, K. Genomic control for association studies. *Biometrics* **55**,  
538 997-1004 (1999).
- 539 27 Nolte, I. M. Metasubtract: an R-package to analytically produce leave-one-out  
540 meta-analysis GWAS summary statistics. *Bioinformatics* **36**, 4521-4522,  
541 doi:10.1093/bioinformatics/btaa570 (2020).
- 542 28 Zhao, H. *et al.* CrossMap: a versatile tool for coordinate conversion between  
543 genome assemblies. *Bioinformatics* **30**, 1006-1007,  
544 doi:10.1093/bioinformatics/btt730 (2014).
- 545 29 Gögele, M. *et al.* Methods for Meta-Analyses of Genome-wide Association  
546 Studies: Critical Assessment of Empirical Evidence. *American Journal of*  
547 *Epidemiology* **175**, 739-749, doi:10.1093/aje/kwr385 (2012).

- 548 30 Quick, C. *et al.* emeraLD: rapid linkage disequilibrium estimation with massive  
549 datasets. *Bioinformatics* **35**, 164-166, doi:10.1093/bioinformatics/bty547 (2019).
- 550 31 Kamat, M. A. *et al.* PhenoScanner V2: an expanded tool for searching human  
551 genotype-phenotype associations. *Bioinformatics* **35**, 4851-4853,  
552 doi:10.1093/bioinformatics/btz469 (2019).
- 553 32 Sollis, E. *et al.* The NHGRI-EBI GWAS Catalog: knowledgebase and deposition  
554 resource. *Nucleic Acids Res* **51**, D977-d985, doi:10.1093/nar/gkac1010 (2023).
- 555 33 Teumer, A. *et al.* Genome-wide analyses identify a role for SLC17A4 and AADAT  
556 in thyroid hormone regulation. *Nature Communications* **9**, 4455,  
557 doi:10.1038/s41467-018-06356-1 (2018).
- 558 34 Köttgen, A. *et al.* Multiple loci associated with indices of renal function and  
559 chronic kidney disease. *Nat Genet* **41**, 712-717, doi:10.1038/ng.377 (2009).
- 560 35 Köttgen, A. *et al.* New loci associated with kidney function and chronic kidney  
561 disease. *Nat Genet* **42**, 376-384, doi:10.1038/ng.568 (2010).
- 562 36 Tang, W. *et al.* Genetic associations for activated partial thromboplastin time and  
563 prothrombin time, their gene expression profiles, and risk of coronary artery  
564 disease. *Am J Hum Genet* **91**, 152-162, doi:10.1016/j.ajhg.2012.05.009 (2012).
- 565 37 Meyer, T. E. *et al.* Genome-wide association studies of serum magnesium,  
566 potassium, and sodium concentrations identify six Loci influencing serum  
567 magnesium levels. *PLoS Genet* **6**, doi:10.1371/journal.pgen.1001045 (2010).
- 568 38 Volpato, C. B. *et al.* Linkage and association analysis of hyperthyrotropinaemia in  
569 an Alpine population reveal two novel loci on chromosomes 3q28-29 and 6q26-  
570 27. *J Med Genet* **48**, 549-556, doi:10.1136/jmg.2010.088583 (2011).
- 571 39 Pattaro, C. *et al.* The genetic study of three population microisolates in South  
572 Tyrol (MICROS): study design and epidemiological perspectives. *BMC Med*  
573 *Genet* **8**, 29, doi:10.1186/1471-2350-8-29 (2007).
- 574 40 Gögele, M. *et al.* Heritability analysis of life span in a semi-isolated population  
575 followed across four centuries reveals the presence of pleiotropy between life  
576 span and reproduction. *J Gerontol A Biol Sci Med Sci* **66**, 26-37,  
577 doi:10.1093/gerona/glq163 (2011).
- 578 41 Riegler, A., Marroni, F., Pattaro, C., Gueresi, P. & Pramstaller, P. P. Isolation and  
579 marriage patterns in four South Tyrolean villages (Italy) during the nineteenth  
580 century. *J Biosoc Sci* **40**, 787-791, doi:10.1017/s0021932007002568 (2008).
- 581 42 Yeo, N. C. *et al.* Shroom3 contributes to the maintenance of the glomerular  
582 filtration barrier integrity. *Genome Res* **25**, 57-65, doi:10.1101/gr.182881.114  
583 (2015).
- 584 43 Prokop, J. W. *et al.* Characterization of Coding/Noncoding Variants for  
585 SHROOM3 in Patients with CKD. *J Am Soc Nephrol* **29**, 1525-1535,  
586 doi:10.1681/asn.2017080856 (2018).
- 587 44 Ledo, N. *et al.* Functional genomic annotation of genetic risk loci highlights  
588 inflammation and epithelial biology networks in CKD. *J Am Soc Nephrol* **26**, 692-  
589 714, doi:10.1681/asn.2014010028 (2015).
- 590 45 Larsson, S. C., Drca, N. & Michaëlsson, K. Serum Magnesium and Calcium  
591 Levels and Risk of Atrial Fibrillation. *Circulation: Genomic and Precision*  
592 *Medicine* **12**, e002349, doi:doi:10.1161/CIRCGEN.118.002349 (2019).

- 593 46 Kitchen, S., McCraw, A. & Echenagucia, M. Diagnosis of hemophilia and other  
594 bleeding disorders. *World Federation of Hemophilia* (2010).
- 595 47 Park, J. *et al.* Single-cell transcriptomics of the mouse kidney reveals potential  
596 cellular targets of kidney disease. *Science* **360**, 758-763,  
597 doi:10.1126/science.aar2131 (2018).
- 598 48 Zhao, F. *et al.* Expression, function and clinical application of stanniocalcin-1 in  
599 cancer. *J Cell Mol Med* **24**, 7686-7696, doi:10.1111/jcmm.15348 (2020).
- 600 49 Pattaro, C. *et al.* Genome-Wide Association and Functional Follow-Up Reveals  
601 New Loci for Kidney Function. *PLOS Genetics* **8**, e1002584,  
602 doi:10.1371/journal.pgen.1002584 (2012).
- 603 50 Pattaro, C. *et al.* Genetic associations at 53 loci highlight cell types and biological  
604 pathways relevant for kidney function. *Nature Communications* **7**, 10023,  
605 doi:10.1038/ncomms10023 (2016).
- 606

607 **Tables**

608 **Table 1.** Main characteristics of the 10,146 study individuals. Additional clinical characteristics  
609 are described in Supplementary Table 1.

Participants' characteristics	Median (IQR) or N (%)
Age, years	46.5 (32.8 – 57.6)
Females	5,585 (55.1%)
eGFR <sub>crea</sub> , ml/min/1.73 m <sup>2</sup> (n=10,141)	92.2 (81.3 – 103.2)
Mg, mg/dl (n=10,143)	2.0(1.9 – 2.1)
aPTT, seconds (n=6,369)	29.9 (28.6 – 31.4)
TSH, $\mu$ UI/mL (n=10,142)	1.38 (0.98 – 1.93)
FT3, pg/ml (n=466)	3.00 (2.70 – 3.30)
FT4, ng/dl (n=469)	0.98 (0.87 – 1.08)
Hyperthyroidism (TSH < 0.4 or drug treatment)	172 (1.7%)
Normal (0.4 < TSH < 3.8 and no drug treatment)	8,972 (88.5%)
Hypothyroidism (TSH > 3.8 or drug treatment)	998 (9.8%)

610 Abbreviations: IQR, Interquartile range; TSH, Thyroid-stimulating hormone; FT3, Free triiodothyronine;  
611 FT4, Free thyroxine; Mg, serum magnesium; aPTT, activated partial thromboplastin time.

612 **Table 2.** The 11 loci identified from the CKDGen Consortium GWAS that showed significant effects in CHRIS. Loci were considered  
613 replicated if any variant in strong LD with the CKDGen lead variant was significantly associated with ln(eGFRcrea) in CHRIS.  
614 Consequently, for each locus, reported are either the results on the CKDGen lead variant or the results of the replicated variant in the  
615 locus when the lead variant did not replicate itself (see Methods).

Locus (gene name)	CKDGen lead SNP	Chr:Pos (build 38)	EA/OA	CKDGen Trans-Ethnic <sup>1</sup>			CKDGen EUR <sup>1</sup>			CHRIS			
				EAF	b(SE)*	P-value	EAF	b(SE)*	P-value	EAF	b(SE)	1-sided P- value	2-sided P- value
<b>CASZ1</b>	rs74748843	1:10,670,853	T/C	0.071	-0.00480 (0.00078)	7.84×10 <sup>-10</sup>	0.021	-0.00598 (0.00123)	1.09×10 <sup>-6</sup>	0.016	-0.02776 (0.00806)	2.85×10 <sup>-4</sup>	5.70×10 <sup>-4</sup>
<b>DDX1</b>	rs807624	2:15,642,347	G/T	0.580	-0.00320 (0.00029)	8.28×10 <sup>-28</sup>	0.660	-0.00338 (0.00035)	1.70×10 <sup>-21</sup>	0.701	-0.00886 (0.00221)	3.00×10 <sup>-5</sup>	5.99×10 <sup>-5</sup>
<b>SHROOM3</b>	rs28817415	4:76,480,299	T/C	0.400	-0.00730 (0.00029)	3.17×10 <sup>-137</sup>	0.440	-0.00744 (0.00034)	7.32×10 <sup>-109</sup>	0.428	-0.00840 (0.00214)	4.21×10 <sup>-5</sup>	8.42×10 <sup>-5</sup>
<b>DAB2</b>	rs10062079*	5:39,393,631	A/G	0.039	-0.00480 (0.00029)	2.07×10 <sup>-60</sup>	0.430	-0.00549 (0.00035)	6.19×10 <sup>-56</sup>	0.453	-0.00721 (0.00211)	3.24×10 <sup>-4</sup>	6.49×10 <sup>-4</sup>
<b>SLC34A1</b>	rs3812036	5:177,386,403	T/C	0.260	-0.00650 (0.00039)	2.96×10 <sup>-62</sup>	0.260	-0.00687 (0.00040)	2.31×10 <sup>-67</sup>	0.240	-0.00836 (0.00245)	3.27×10 <sup>-4</sup>	6.55×10 <sup>-4</sup>
<b>TMEM60</b>	rs57514204*	7:77,714,744	T/C	0.380	-0.00340 (0.00029)	3.53×10 <sup>-31</sup>	0.410	-0.00332 (0.00034)	1.48×10 <sup>-22</sup>	0.439	-0.00723 (0.00212)	3.34×10 <sup>-4</sup>	6.67×10 <sup>-4</sup>
<b>STC1</b>	rs819196*	8:23,885,208	T/A	0.410	-0.00410 (0.00029)	1.47×10 <sup>-44</sup>	0.450	-0.00406 (0.00034)	1.43×10 <sup>-33</sup>	0.437	-0.00718 (0.00210)	3.21×10 <sup>-4</sup>	6.41×10 <sup>-4</sup>
<b>PIP5K1B</b>	rs2039424	9:68,817,258	G/A	0.360	-0.00440 (0.00029)	4.75×10 <sup>-51</sup>	0.380	-0.00483 (0.00035)	1.12×10 <sup>-42</sup>	0.421	-0.01042 (0.00212)	4.51×10 <sup>-7</sup>	9.01×10 <sup>-7</sup>
<b>GAB2</b>	rs7113042*	11:78,324,786	A/G	0.740	-0.00270 (0.00039)	4.62×10 <sup>-12</sup>	0.830	-0.00271 (0.00045)	1.86×10 <sup>-9</sup>	0.823	-0.01096 (0.00286)	6.41×10 <sup>-5</sup>	1.28×10 <sup>-4</sup>
<b>IGF1R</b>	rs59646751	15:98,733,292	T/G	0.300	-0.00230 (0.00029)	3.97×10 <sup>-15</sup>	0.310	-0.00203 (0.00036)	2.47×10 <sup>-8</sup>	0.321	-0.00909 (0.00223)	2.32×10 <sup>-5</sup>	4.63×10 <sup>-5</sup>
<b>PDILT</b>	rs77924615	16:20,381,010	G/A	0.800	-0.00980 (0.00039)	4.38×10 <sup>-139</sup>	0.800	-0.00958 (0.00044)	1.52×10 <sup>-104</sup>	0.826	-0.01030 (0.00281)	1.27×10 <sup>-4</sup>	2.53×10 <sup>-4</sup>

616 <sup>1</sup>Trans-ethnic and European ancestry CKDGen results after subtracting the effect of 4661 CHRIS participants included in the  
617 CKDGen meta-analysis.

618 Abbreviations: Chr, chromosome; Pos, position; EA, effect allele; OA, other allele; EAF, effect allele frequency; b, coefficient of  
619 association; SE, standard error of b.

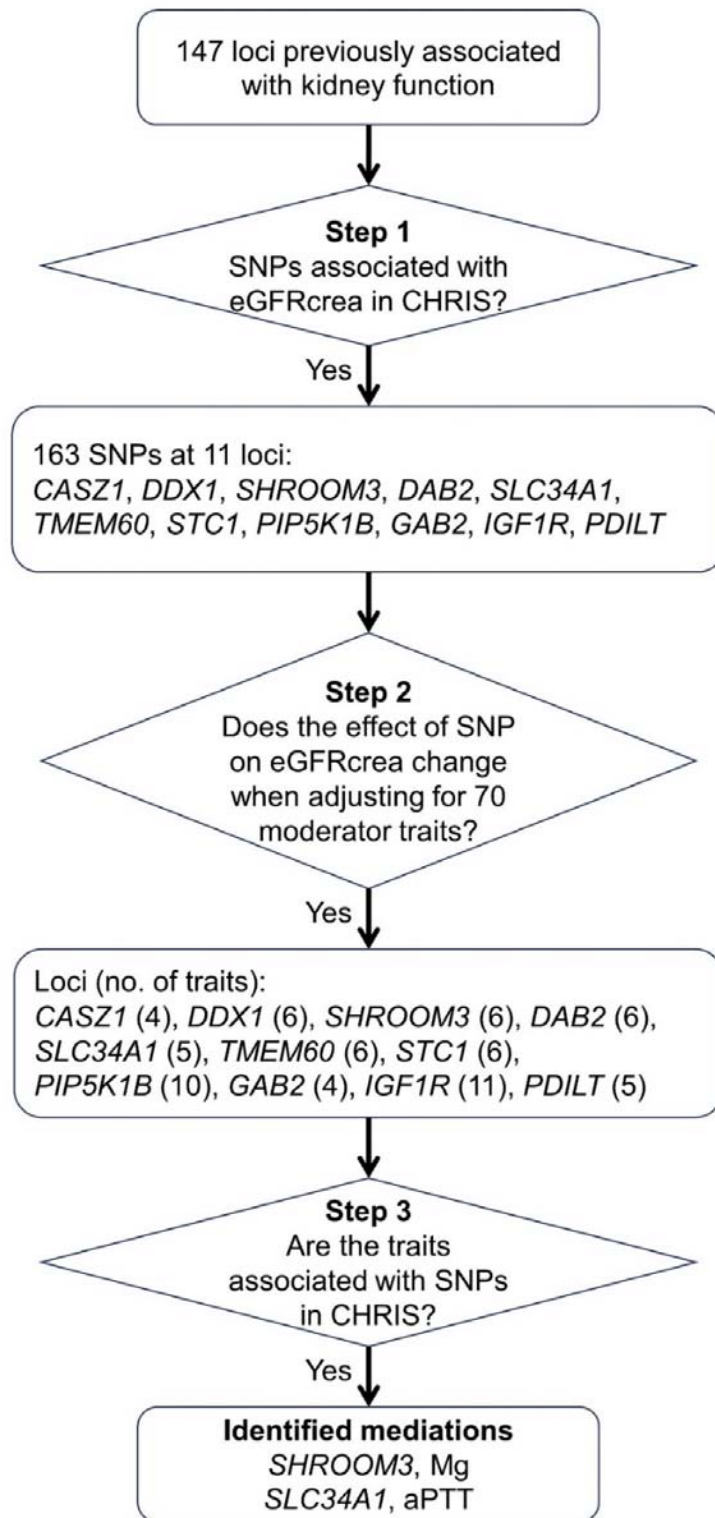
620 \*Replication of a variant in strong LD with the CKDGen lead variant: in these cases, the replicated and not the CKDGen lead variant  
621 is reported. See Supplementary Table S2 for lead variant comparisons.

622 **Table 3** Results of the mediation analysis at *SHROOM3* and *SLC34A1*. Extensive mediation analysis results for all loci and all traits  
 623 are reported in Supplementary Table S4.

Locus (gene name)	GWAS Step 1: association with ln(eGFRcrea)					Step 2: association with ln(eGFRcrea) adjusted for the mediator				Step 3: association with mediator in this study		Step 4: association with mediator adjusted for ln(eGFRcrea)	
	rsID	Chr:Pos (build 38)	EA/OA	b(SE)	P	Trait	b(SE)	P	Change %	b(SE)	P	b(SE)	P
<i>SHROOM3</i>	rs28394165	4:76472865	C/T	-0.00841 (0.00214)	8.35×10 <sup>-5</sup>	Mg	-0.00744 (0.00196)	1.51×10 <sup>-4</sup>	-12	0.00826 (0.00204)	5.33×10 <sup>-5</sup>	0.00773 (0.00204)	1.54×10 <sup>-4</sup>
<i>SHROOM3</i>	rs10025351	4:76472942	T/C	-0.00840 (0.00214)	8.51×10 <sup>-5</sup>	Mg	-0.00743 (0.00196)	1.53×10 <sup>-4</sup>	-12	0.00823 (0.00204)	5.73×10 <sup>-5</sup>	0.00769 (0.00204)	1.65×10 <sup>-4</sup>
<i>SHROOM3</i>	rs28817415	4:76480299	T/C	-0.00840 (0.00214)	8.42×10 <sup>-5</sup>	Mg	-0.00744 (0.00196)	1.51×10 <sup>-4</sup>	-12	0.00823 (0.00204)	5.68×10 <sup>-5</sup>	0.00770 (0.00204)	1.64×10 <sup>-4</sup>
<i>SHROOM3</i>	rs4859682	4:76489165	A/C	-0.00840 (0.00210)	6.59×10 <sup>-5</sup>	Mg	-0.00750 (0.00193)	1.02×10 <sup>-4</sup>	-11	0.00812 (0.00201)	5.49×10 <sup>-5</sup>	0.00758 (0.00201)	1.63×10 <sup>-4</sup>
<i>SHROOM3</i>	rs13146355	4:76490987	A/G	-0.00840 (0.00214)	8.38×10 <sup>-5</sup>	Mg	-0.00749 (0.00196)	1.33×10 <sup>-4</sup>	-11	0.00819 (0.00204)	6.12×10 <sup>-5</sup>	0.00765 (0.00204)	1.77×10 <sup>-4</sup>
<i>SLC34A1</i>	rs3812036	5:177386403	T/C	-0.00836 (0.00245)	6.55×10 <sup>-4</sup>	aPTT	-0.01013 (0.00269)	1.64×10 <sup>-4</sup>	21	-0.23110 (0.04980)	3.54×10 <sup>-6</sup>	-0.23046 (0.04986)	3.88×10 <sup>-6</sup>

624 Abbreviations: Chr, chromosome; Pos, position; EA, effect allele; OA, other allele; b, coefficient of association; SE, standard error of  
 625 b; Mg, Magnesium; aPTT, activated Partial Prothrombin Time.

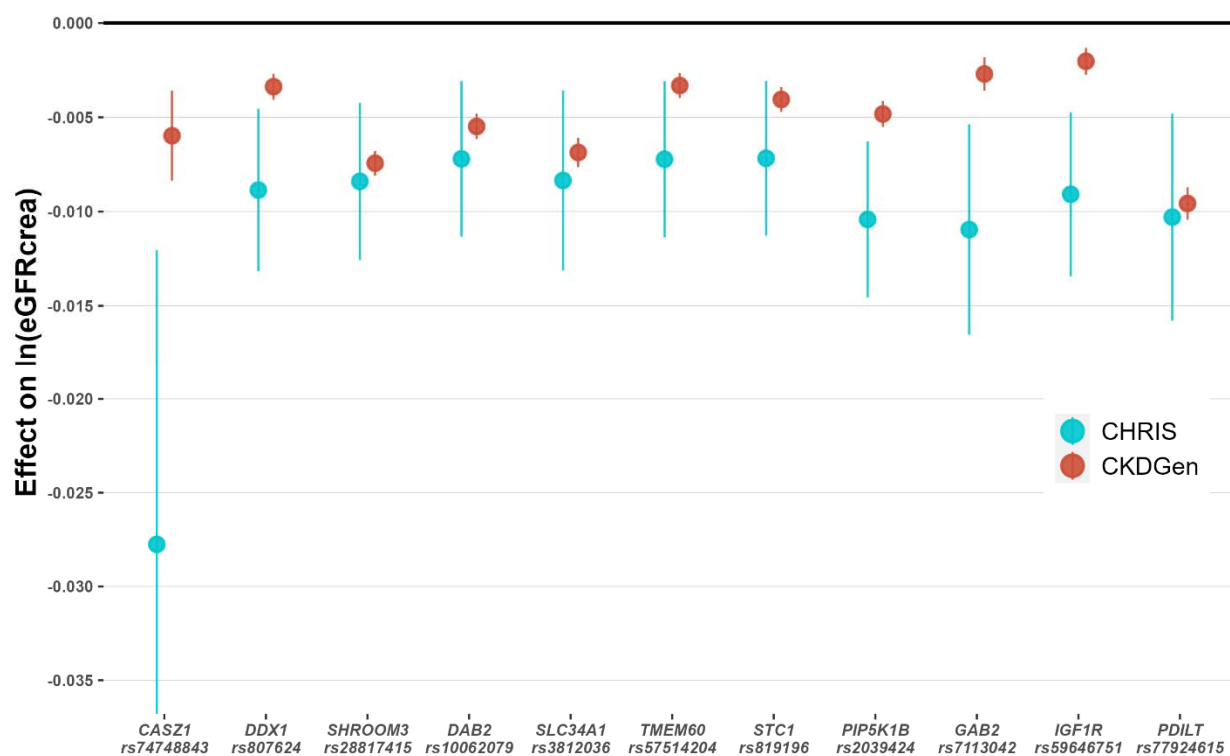
626 **Figures**



627

628 **Figure 1. Flowchart of the mediation analysis.** Abbreviations: Mg, serum magnesium; aPTT,

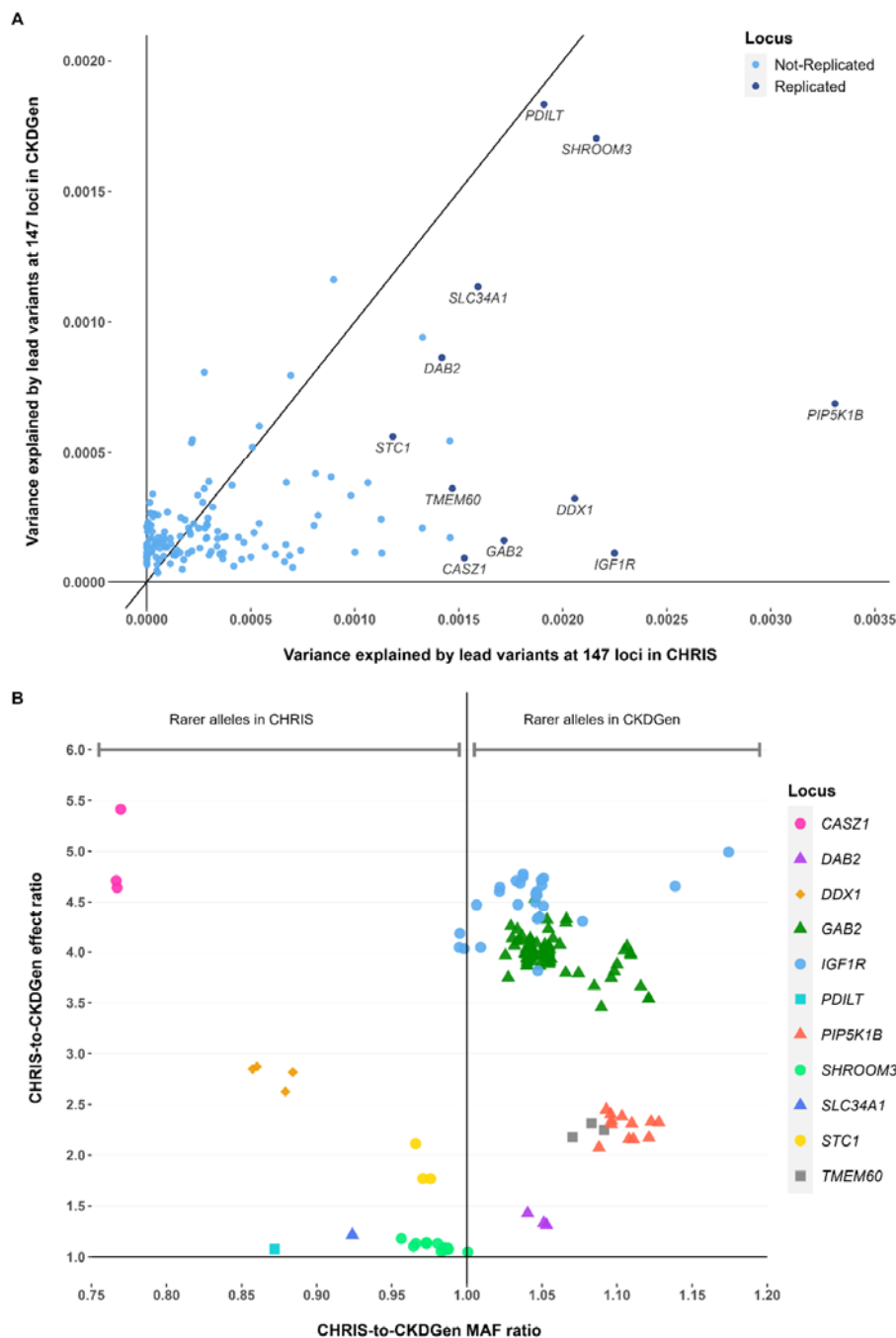
629 activated partial thromboplastin time.



630

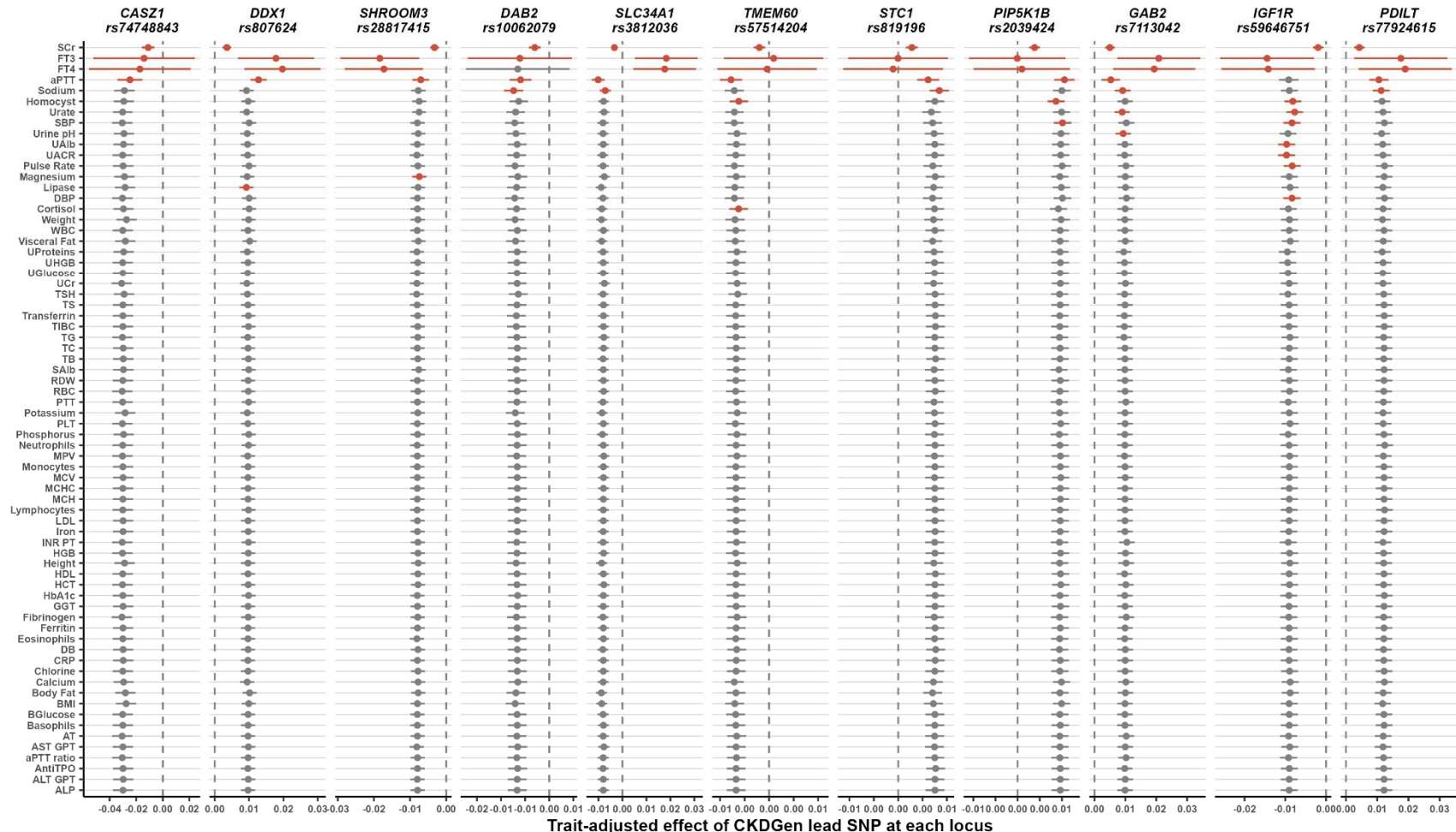
631 **Figure 2. Effect sizes with 95% confidence intervals for the most associated variants at**  
632 **each locus in CKDGen and CHRIS following harmonization to the eGFRcrea-lowering**  
633 **allele.** The effect is meant per copy of the effect allele on the natural logarithm of eGFRcrea in  
634 ml/min/1.73m<sup>2</sup> (y-axis). For *CASZ1*, *DDX1*, *PIP5K1B*, *GAB2*, and *IGF1R* the magnitude of the  
635 effect in CHRIS was larger than in CKDGen.





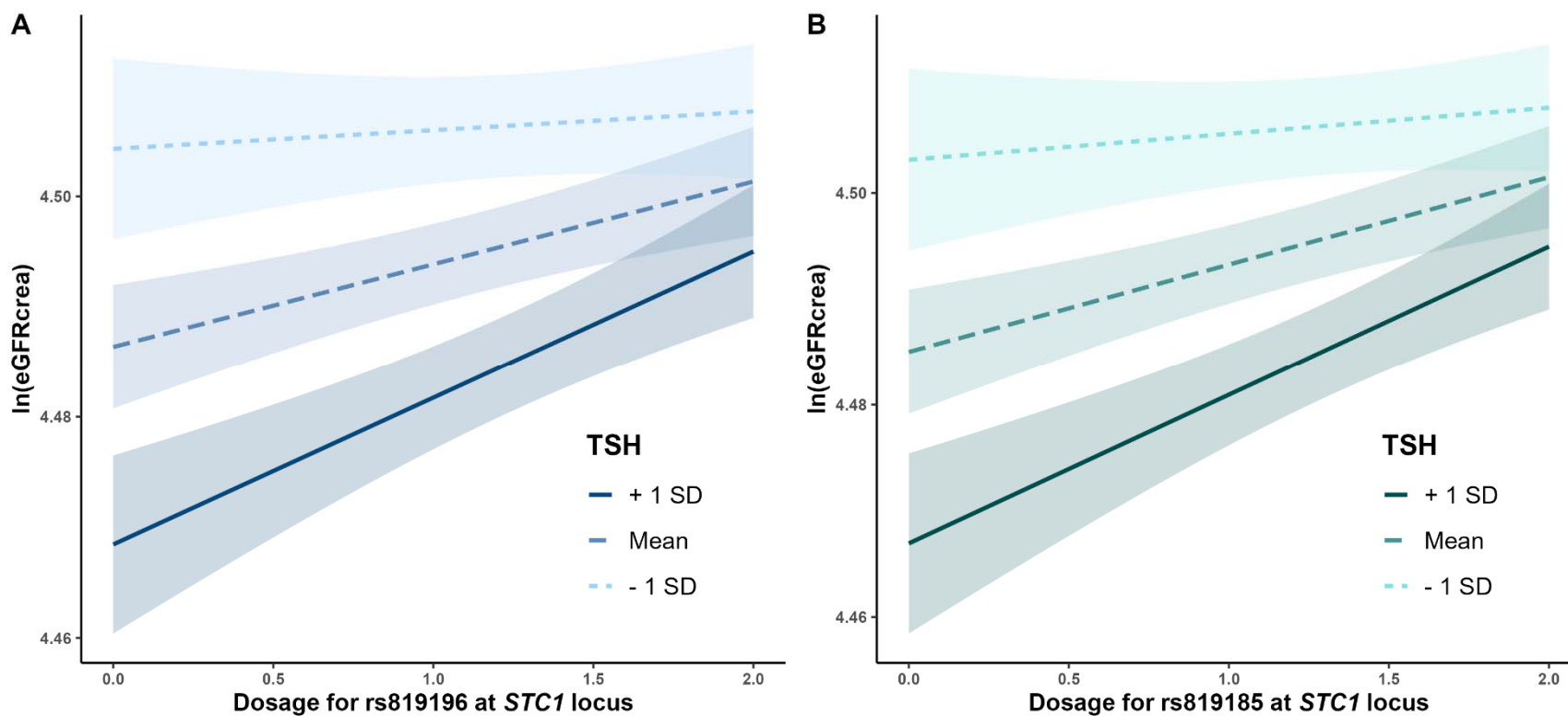
636

637 **Figure 3. Comparison of the characteristics of all 163 variants in the 11 replicated loci in**  
 638 **CKDGen versus CHRIS after alignment of the effect alleles. Panel A:** Variance of age- and  
 639 sex-adjusted  $\ln(eGFR_{crea})$  in CHRIS versus CKDGen EUR explained by the lead variants at  
 640 each of the 147 loci. Replicated loci are labeled with the locus name. **Panel B:** The ratio  
 641 between CHRIS and CKDGen minor allele frequencies (MAF, x-axis) is plotted against the ratio  
 642 between CHRIS and CKDGen effect estimates (y-axis). Effects are expressed in terms of  
 643 change of the natural logarithm of  $eGFR_{crea}$  in  $ml/min/1.73m^2$  per copy of the effect allele.



644

645 **Figure 4. Results of the 2<sup>nd</sup> step of the mediation analysis for the most associated variants at each locus.** SNP effect size and  
 646 their 95% confidence intervals are reported. Effects are expressed in terms of change of the natural logarithm of eGFRcrea in  
 647 ml/min/1.73m<sup>2</sup> per copy of the effect allele. Large departures from the expected effect size ('outlier', see Methods) are indicated in  
 648 red. The vertical dashed lines identify the null. Large effect size changes were observed when adjusting for SCr (positive control; all  
 649 loci), FT3 (all loci) and FT4 (10 loci), aPTT (10 loci), sodium (5 loci), homocysteine (3 loci), SBP (2 loci), urate (2 loci), and cortisol,  
 650 DBP, lipase, magnesium, pulse rate, UACR, urinary albumin and urine pH (1 locus each).



651

652 **Figure 5. Interaction of ln(eGFRcrea)-associated variants with TSH levels.** Effects of *STC1* SNPs rs819185 (P=0.00177; **Panel**  
653 **A**) and rs819196 (P=0.00154; **Panel B**) are larger at lower TSH levels.

## 654 **Supplementary Material**

### 655 **Table of Contents**

656	Supplementary method. Quantile normalization of the 70 health traits used in the mediation	
657	analysis.....	29
658	Supplementary Figure S1. Cumulative variance of 6337 variants over 147 loci explained by	
659	principal components in the CHRIS study. Adjusting for 147 independent tests covers 98% of	
660	the cumulative variance (red dot).....	30
661	Supplementary Figure S2. Fitting PC-adjusted simple linear models came for practical reasons	
662	after observing substantial equivalence with kinship-adjusted linear mixed models obtained with	
663	EMMAX.....	31
664	Supplementary Figure S3. Quantile-quantile (QQ) plot of the P-values from the GWAS of	
665	ln(eGFRcrea) in the CHRIS study.....	32
666	Supplementary Figure S4. Regional association plots for the 11 replicated loci. Highlighted in	
667	purple is the most associated SNP in the CKDGen GWAS meta-analysis. All SNP positions are	
668	referred to the NCBI Build 38. Plots were generated with LocusZoom version 1.4 (.....	33
669	Supplementary Figure S5. Comparison of effect size and standard error ratio for the variants in	
670	the model regressing ln(eGFRcrea) on age, sex, FT3 (or FT4), genetic variants with and without	
671	adjusting for municipality.....	39
672		
673		

674 **Supplementary method.** Quantile normalization of the 70 health traits used in the mediation  
675 analysis.

676 We used *normalize2Reference* function in the *caret* R package version 6.0-34 (the function is  
677 available in the later versions here: <https://github.com/topepo/caret/tree/master/deprecated>).

678 Each trait was measured partially with an older method and partially with a newer, most recent  
679 method. The observations from the newer method (method = 1) were set as the reference to  
680 derive the quantiles of the trait distribution on which to standardize the observations from the  
681 older method (method = 0). We then replaced the observation from the older method with the  
682 standardized ones and included together with the values from the newer method.

### 683 **Notation:**

- 684 - **trait:** the quantitative variable to which quantile normalization was applied;
- 685 - **method:** the newer method for measuring the trait (1); the older method (0);
- 686 - **df:** dataframe; the data in which the trait of interest is available

687

```
688 library(caret)
689 quantile_norm <- normalize2Reference(
690   data = df[df$method == 0, "trait"],
691   refData = quantile(
692     df[df$method == 1, "trait"],
693     probs = seq(0, 1, length.out = length(df[df$method == 0,
694 "trait"])),
695     na.rm = TRUE,
696     names = TRUE,
697     type = 7,
698     digits = 7),
699   ties = TRUE)
```

700

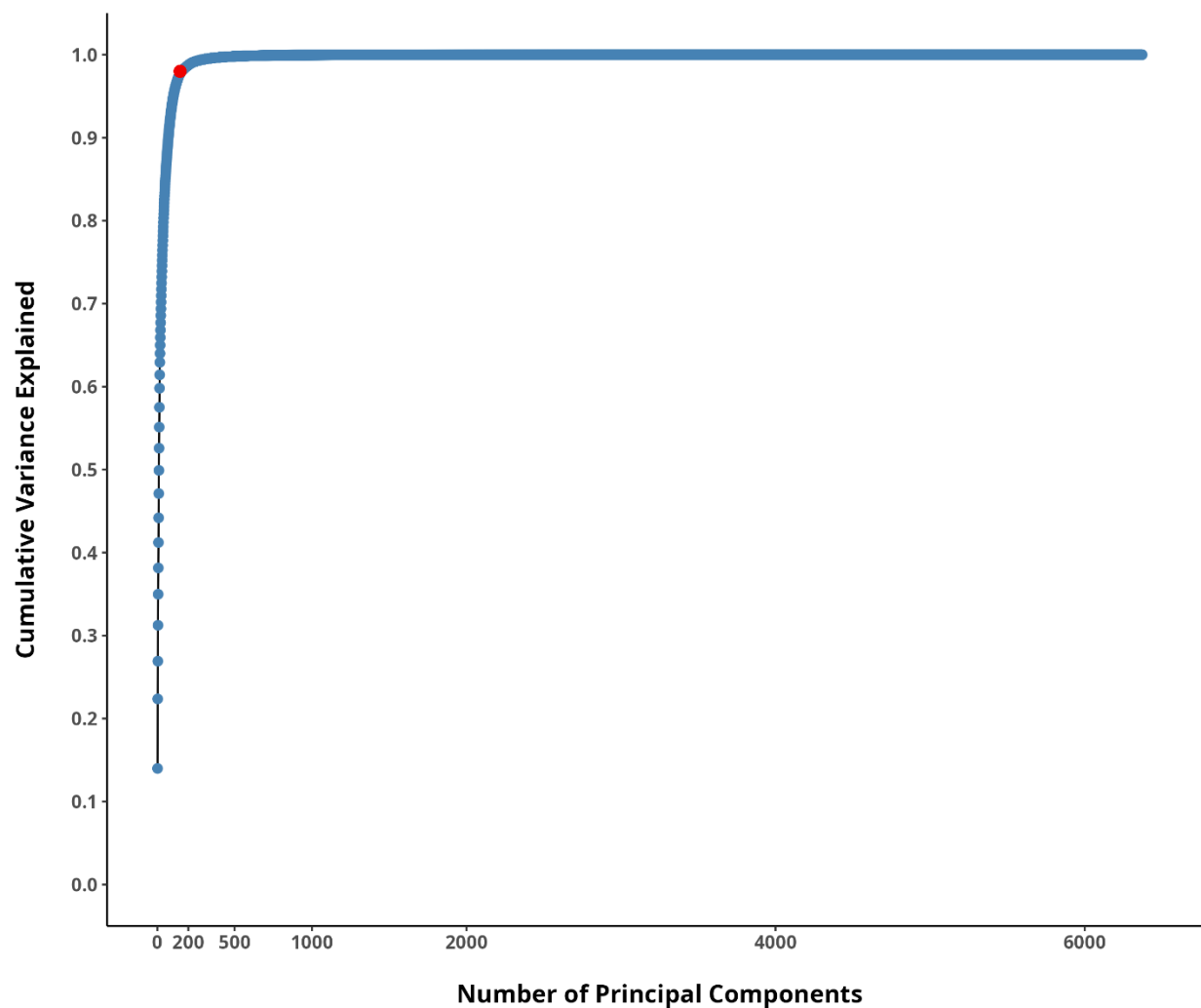
701 # Creating a new variable for the quantile normalized trait

```
702 df$trait_std <- df$trait
```

703 # Replacing non-transformed values with the quantile-normalized values

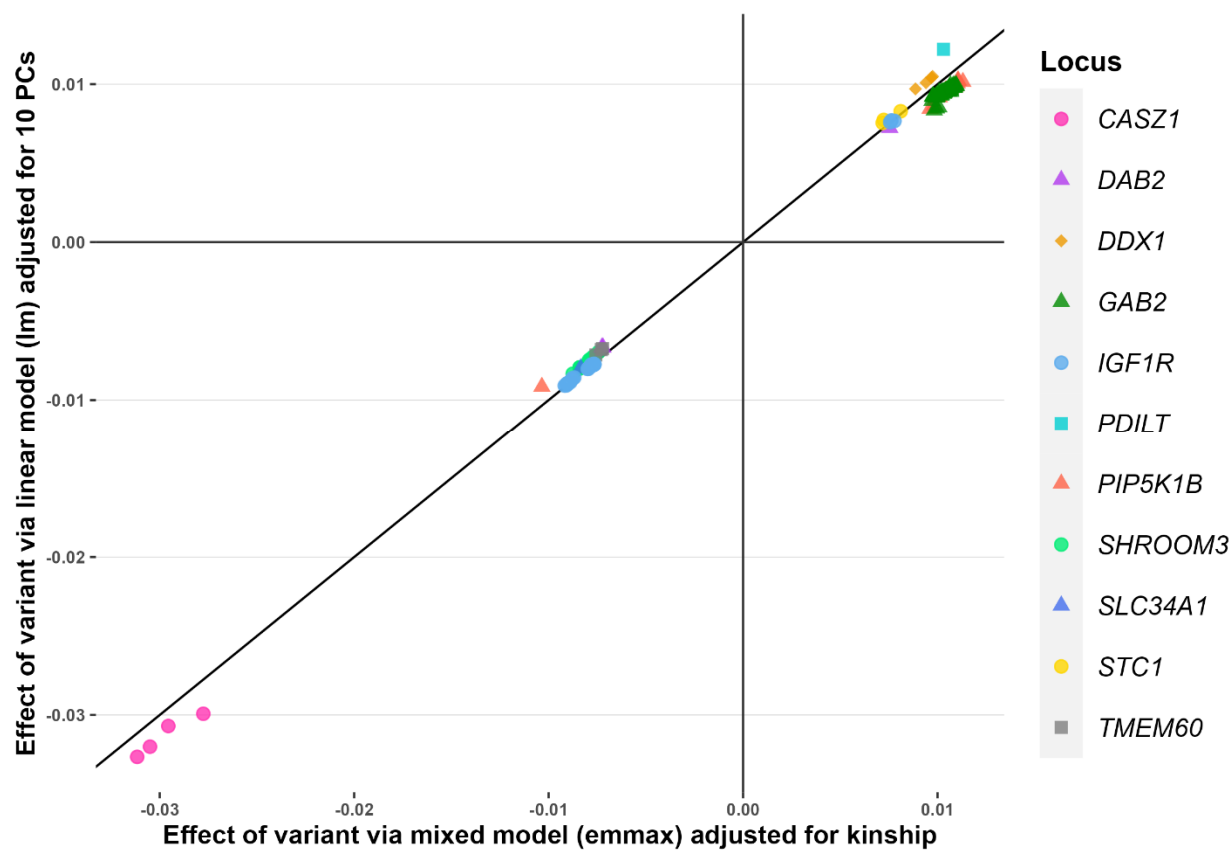
```
704 df[df$method == 0, "trait_std"] <- quantile_norm
```

705 **Supplementary Figure S1.** Cumulative variance of 6337 variants over 147 loci explained by  
706 principal components in the CHRIS study. Adjusting for 147 independent tests covers 98% of  
707 the cumulative variance (red dot).



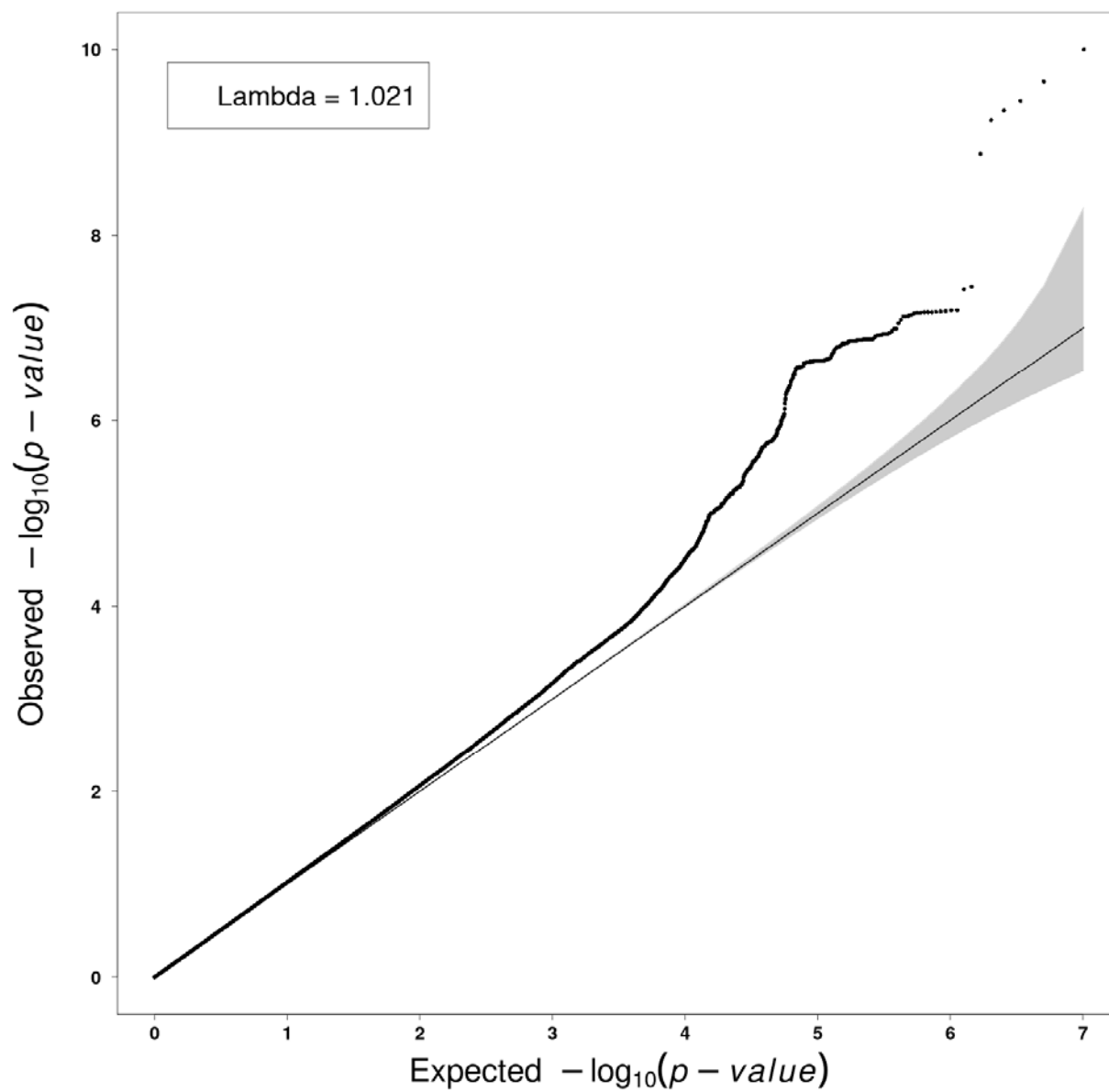
708

709 **Supplementary Figure S2.** Fitting PC-adjusted simple linear models came for practical reasons  
710 after observing substantial equivalence with kinship-adjusted linear mixed models obtained with  
711 EMMAX.



712

713 **Supplementary Figure S3.** Quantile-quantile (QQ) plot of the P-values from the GWAS of  
714  $\ln(\text{eGFR}_{\text{crea}})$  in the CHRIS study.



715

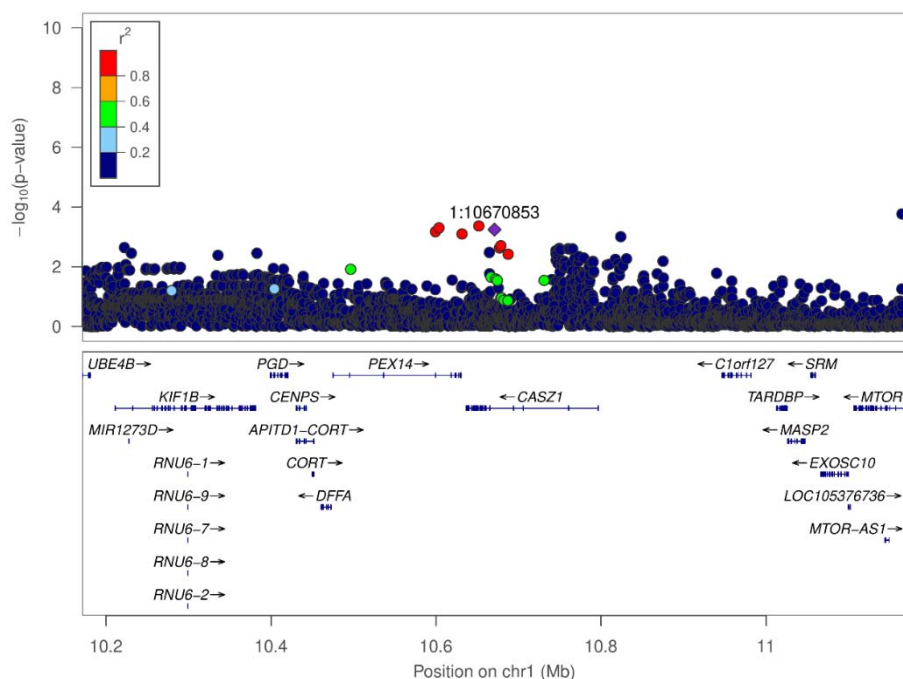


716 **Supplementary Figure S4.** Regional association plots for the 11 replicated loci. Highlighted in  
717 purple is the most associated SNP in the CKDGen GWAS meta-analysis. All SNP positions are  
718 referred to the NCBI Build 38. Plots were generated with LocusZoom version 1.4 (Pruim RJ, et  
719 al., Bioinformatics, 2010. 26(18): 2336-7).

720 \* indicates cases where not the CKDGen lead SNP but a proxy of it was replicated in the CHRIS  
721 study.

722

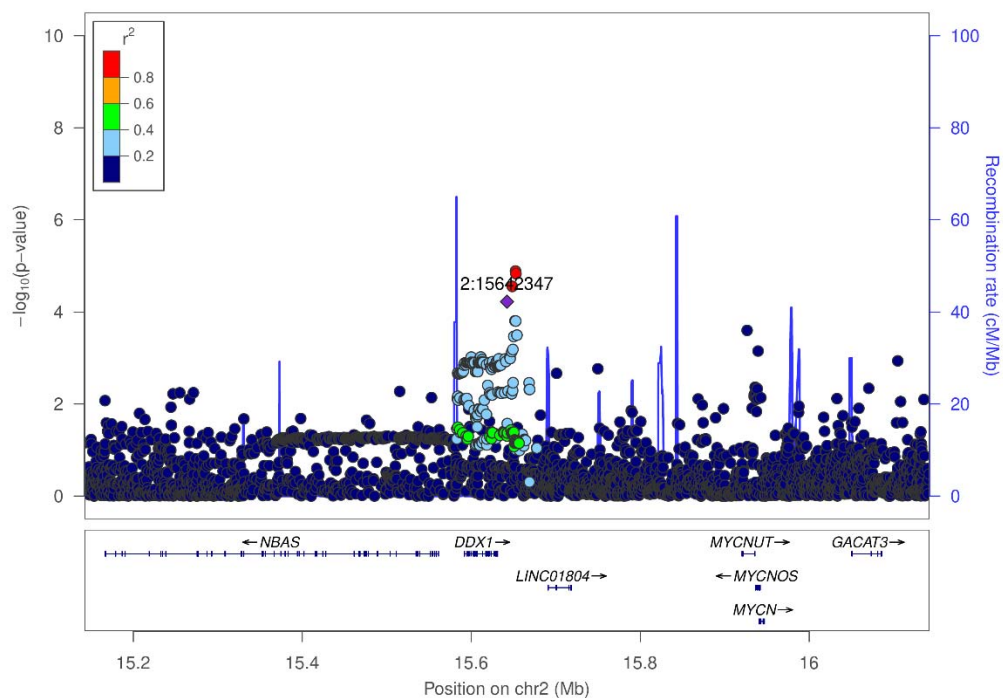
723 Top CHRIS SNP: rs74748843 (1:10,670,853) at **CASZ1**



724

725

Top CHRIS SNP: rs807624 (2:15,642,347) at **DDX1**

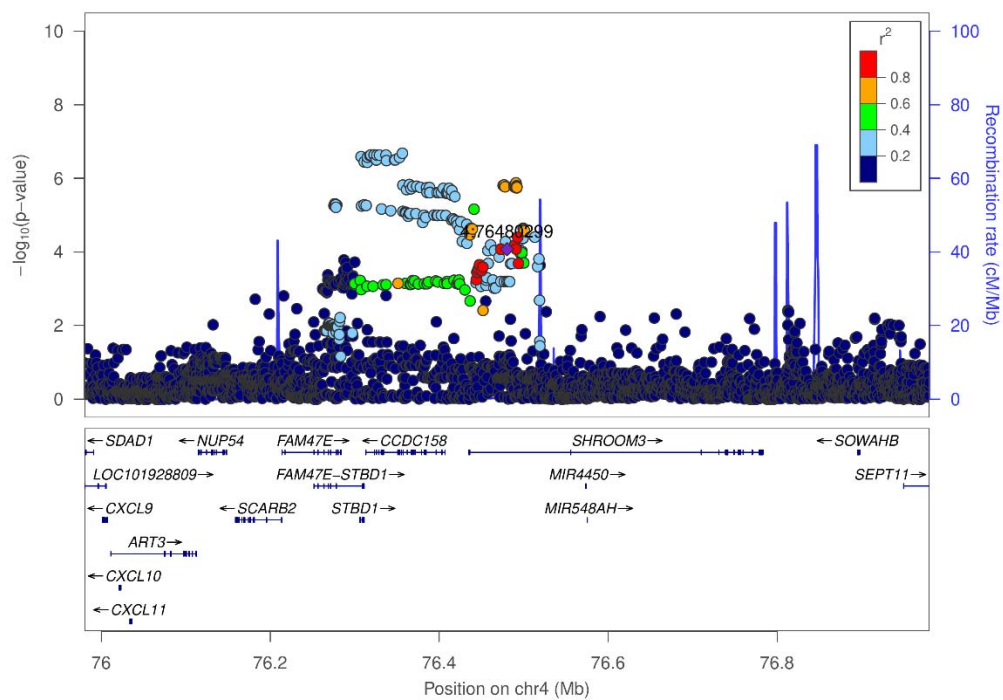


726

727

728

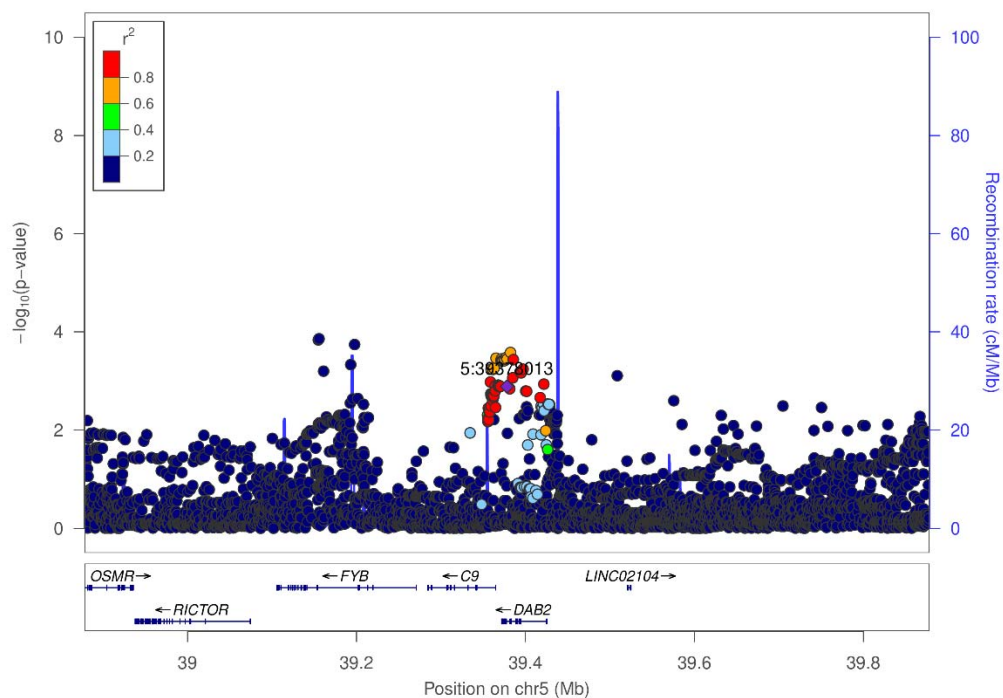
Top CHRIS SNP: rs28817415 (4:76,480,299) at **SHROOM3**



729

730

Top CHRIS SNP: rs1362800\* (5:39,378,013) at **DAB2**\*

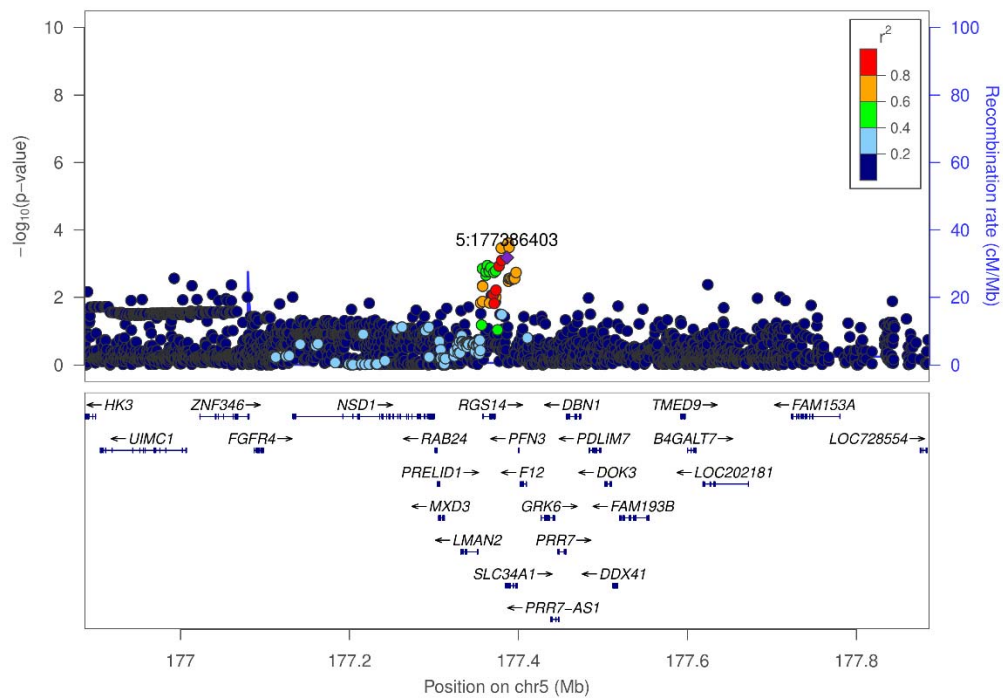


731

732

733

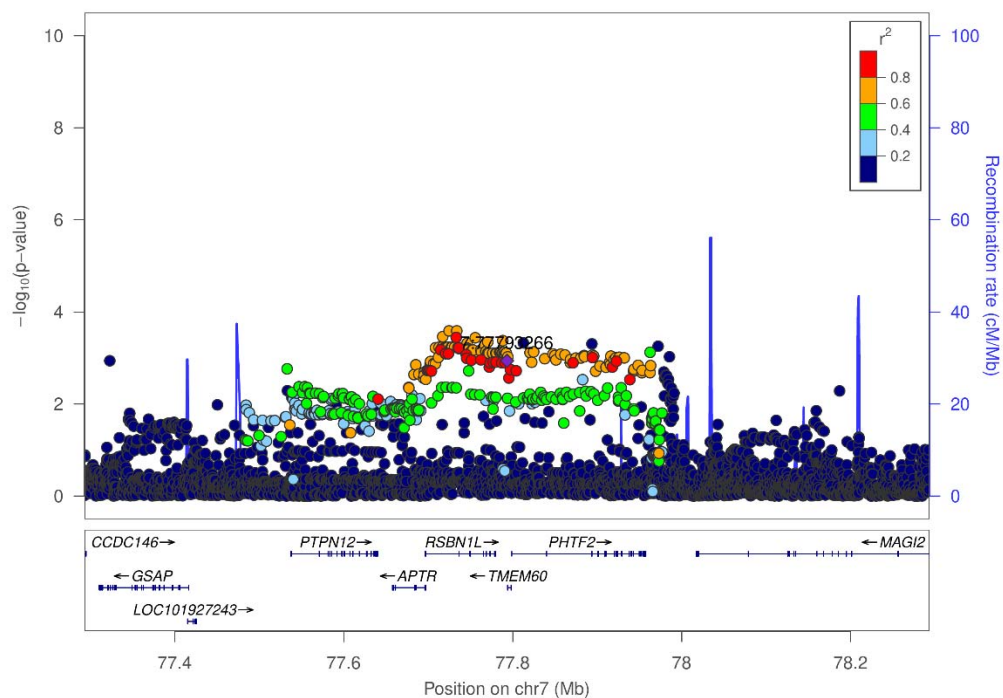
Top CHRIS SNP: rs3812036 (5:177,386,403) at **SLC34A1**



734

735

Top CHRIS SNP: rs6973656\* (7:77,793,266) at **TMEM60\***

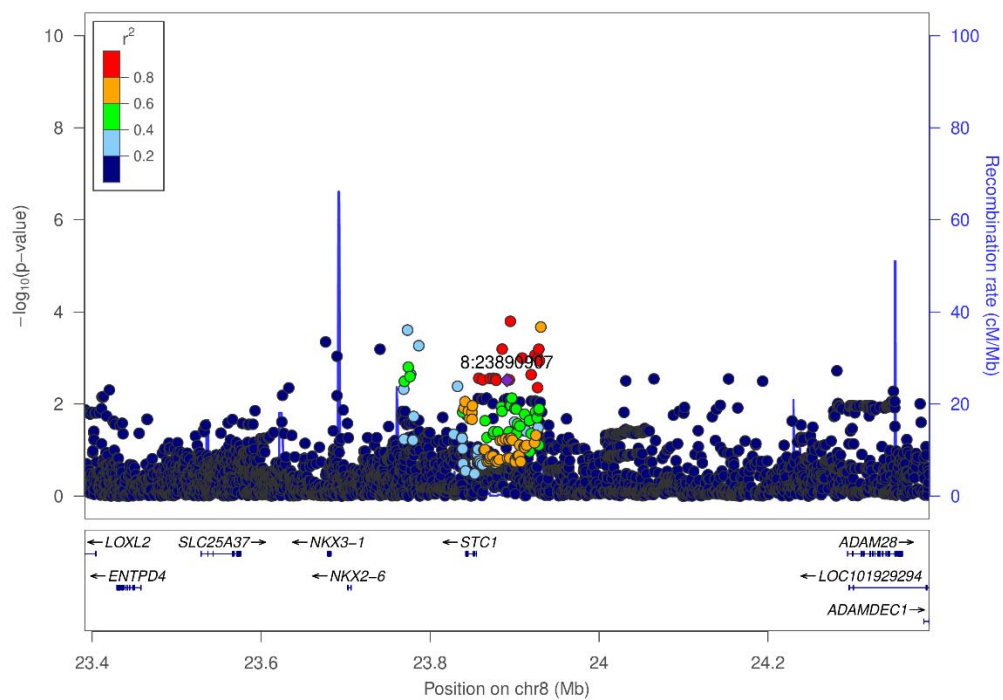


736

737

738

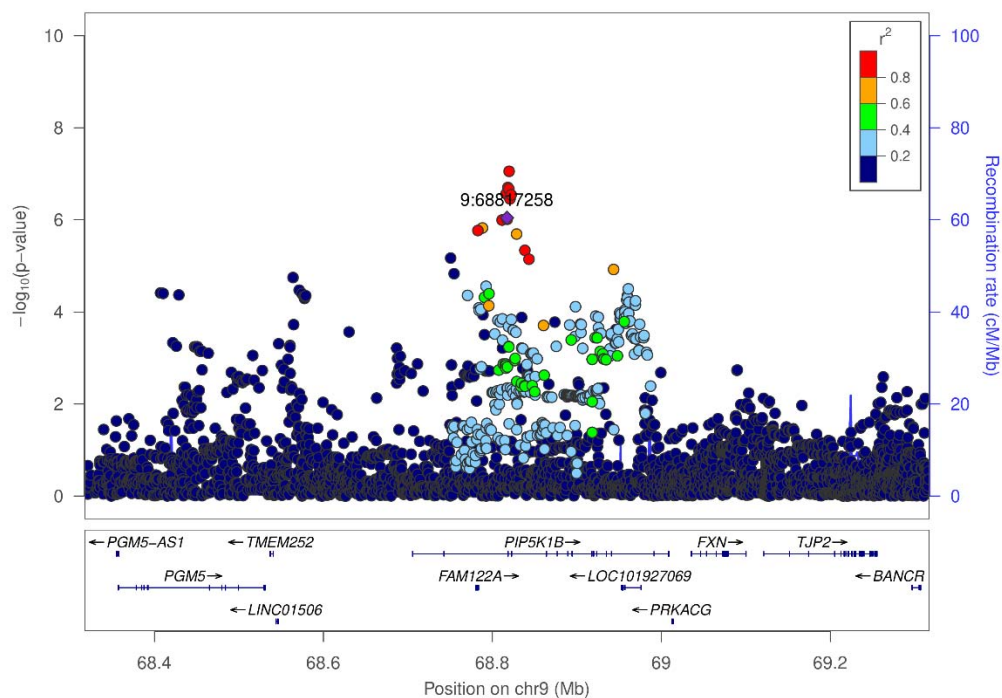
Top CHRIS SNP: rs34861762\* (8:23,890,907) at **STC1\***



739

740

Top CHRIS SNP: rs2039424 (9:68,817,258) at **PIP5K1B**

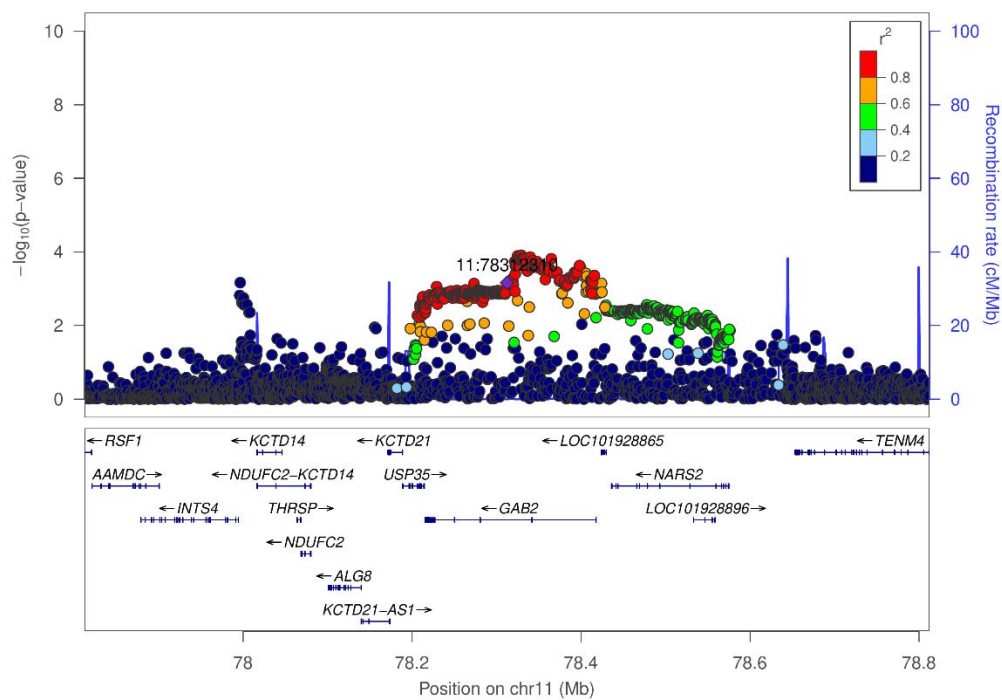


741

742

743

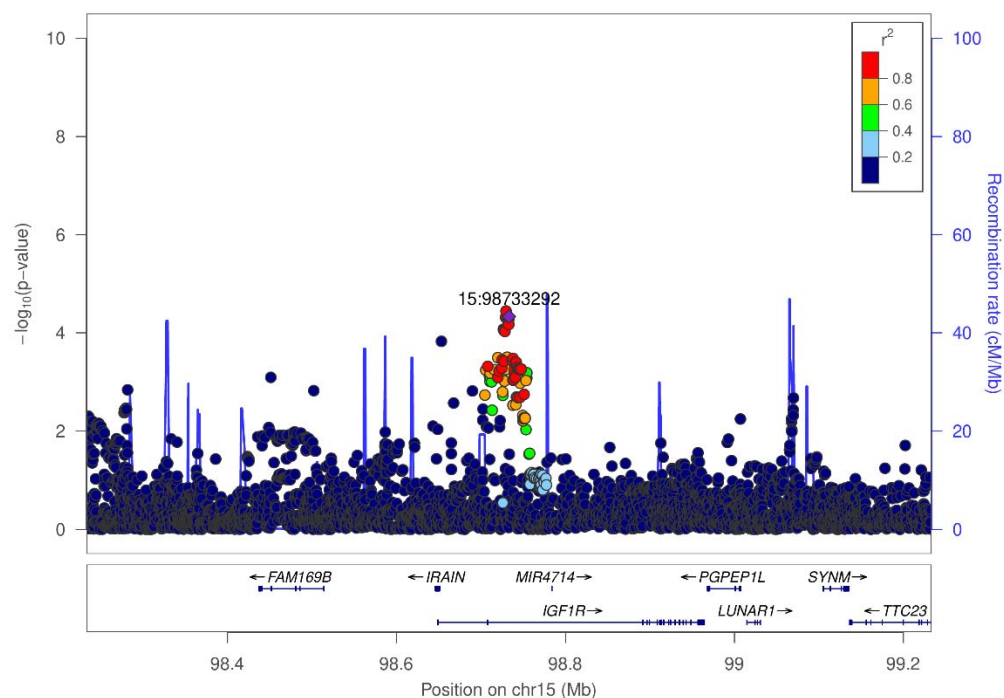
Top CHRIS SNP: rs11237450\* (11:78,312,310) at **GAB2\***



744

745

Top CHRIS SNP: rs59646751 (15:98,733,292) at **IGF1R**

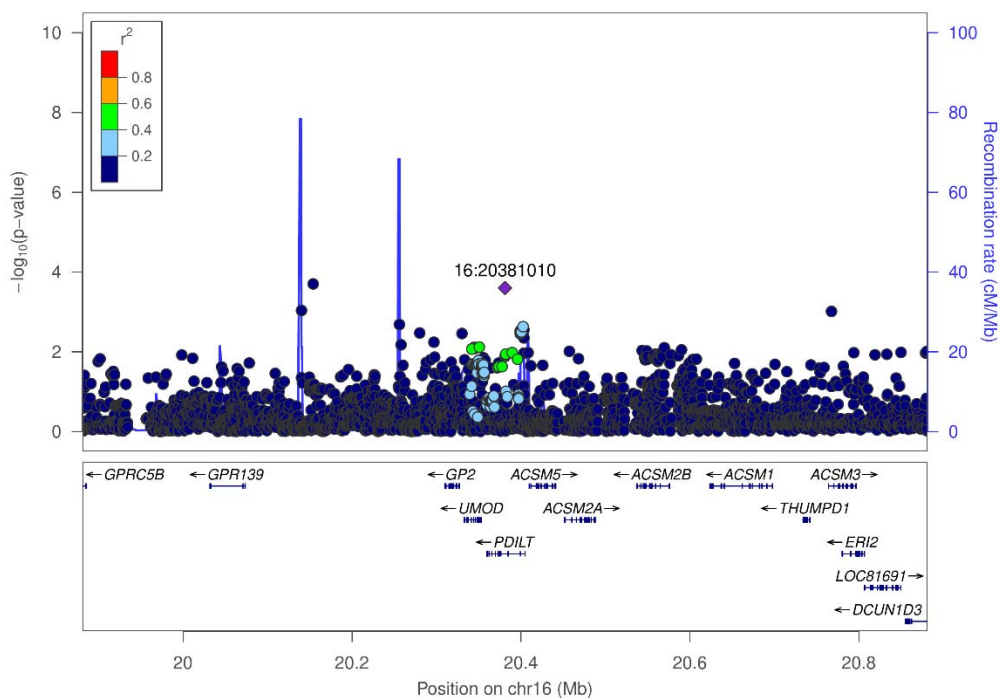


746

747

748

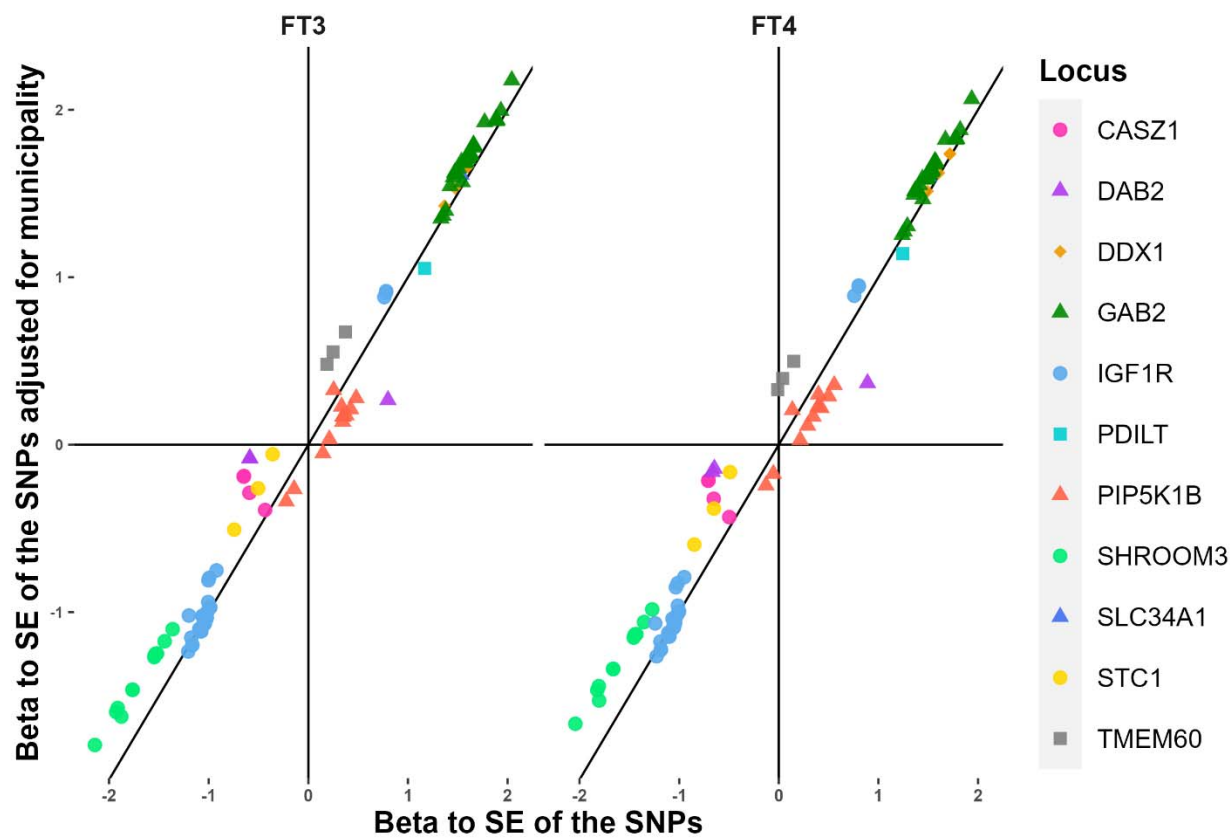
Top CHRIS SNP: rs77924615 (16:20381010) at **PDILT**



749

750 **Supplementary Figure S5.** Comparison of effect size and standard error ratio for the variants  
751 in the model regressing  $\ln(\text{eGFR}_{\text{crea}})$  on age, sex, FT3 (or FT4), genetic variants with and  
752 without adjusting for municipality.

753



754