Differential and shared genetic effects on kidney function between diabetic and non-diabetic individuals

Thomas W. Winkler  
*University of Regensburg*

E Warwick Daw  
*Washington University School of Medicine in St. Louis*

Mary F. Feitosa  
*Washington University School of Medicine in St. Louis*

Mary K. Wojczynski  
*Washington University School of Medicine in St. Louis*

Michael A. Province  
*Washington University School of Medicine in St. Louis*

See next page for additional authors

Follow this and additional works at: https://digitalcommons.wustl.edu/open_access_pubs

Please let us know how this document benefits you.

**Recommended Citation**  
Winkler, Thomas W.; Daw, E Warwick; Feitosa, Mary F.; Wojczynski, Mary K.; Province, Michael A.; and et al, "Differential and shared genetic effects on kidney function between diabetic and non-diabetic individuals." Communications Biology. 5, 1. 580 (2022).  
https://digitalcommons.wustl.edu/open_access_pubs/12056

This Open Access Publication is brought to you for free and open access by Digital Commons@Becker. It has been accepted for inclusion in Open Access Publications by an authorized administrator of Digital Commons@Becker. For more information, please contact vanam@wustl.edu.
Authors
Thomas W. Winkler, E Warwick Daw, Mary F. Feitosa, Mary K. Wojczynski, Michael A. Province, and et al
Differential and shared genetic effects on kidney function between diabetic and non-diabetic individuals

Reduced glomerular filtration rate (GFR) can progress to kidney failure. Risk factors include genetics and diabetes mellitus (DM), but little is known about their interaction. We conducted genome-wide association meta-analyses for estimated GFR based on serum creatinine (eGFR), separately for individuals with or without DM \( (n_{DM} = 178,691, n_{noDM} = 1,296,113) \). Our genome-wide searches identified (i) seven eGFR loci with significant DM/noDM-difference, (ii) four additional novel loci with suggestive difference and (iii) 28 further novel loci (including CUBN) by allowing for potential difference. GWAS on eGFR among DM individuals identified 2 known and 27 potentially responsible loci for diabetic kidney disease. Gene prioritization highlighted 18 genes that may inform reno-protective drug development. We highlight the existence of DM-only and noDM-only effects, which can inform about the target group, if respective genes are advanced as drug targets. Largely shared effects suggest that most drug interventions to alter eGFR should be effective in DM and noDM.
Impaired kidney function can progress to kidney failure requiring dialysis or transplantation, which implicates high early mortality\(^1\) and public health burden\(^2\). The most common kidney function measure is the glomerular filtration rate (GFR). Chronic kidney disease (CKD) can be defined as GFR < 60 ml/min per 1.73 m\(^2\), and has a prevalence of \(\sim\)10% in the population\(^3\). In large-scale epidemiological studies, GFR can be estimated from serum creatinine (eGFR) using the CKD-Epi formula\(^4\).

Diabetes mellitus (DM) is a major risk factor for impaired kidney function\(^5\) and is present in 30% of individuals starting dialysis in Europe\(^6\). Most current pharmacotherapy in DM has substantial side effects\(^7,8\). Recent work showed that SGLT2 inhibitors and GLP1R are renoprotective for individuals with DM\(^9,10\) and early evidence indicates that this might be also true for individuals without DM (noDM)\(^11\). The biological processes related to lower kidney function might differ between individuals with DM and individuals without DM. Understanding the mechanisms of kidney function variability within and between these groups is pivotal for understanding pathogenesis. Genome-wide association studies (GWAS) have pinpointed hundreds of loci for eGFR\(^14-16\). These loci help understand kidney function variability in the overall population. However, commonalities and differences in genetic kidney function effects between DM and noDM individuals are not well understood.

Genes underneath GWAS loci have been shown to double the success rate in drug development pipelines\(^17,18\). This renders each gene in eGFR loci a drug target to alter kidney function, particularly when the gene maps to association-driving variants that are relevant for the protein or for the gene’s expression in kidney tissue\(^19\). This opens up a clinically important question in the context of differential or shared genetic kidney function effects between DM and noDM individuals: when a locus identified for eGFR in the overall population turns out to have no effect in DM individuals, the underlying causal gene might not be informative for reno-protection in DM individuals. This is not an unlikely scenario, since only 10% of individuals in previous population-based GWAS on eGFR overall had DM\(^16,20\).

Previous DM-/noDM-stratified searches for genetic eGFR-effects have been limited\(^21\) and without identification of a significant DM/noDM-difference. It is unclear whether this was due to limited power or lack of such differential effects. GWAS for eGFR-based kidney function outcomes in DM, used as one approach to identify the genetics of Diabetic Kidney Disease (DKD), identified four loci in \(\sim\)40,000 type 1 or type 2 DM patients\(^22\) and four other loci in \(\sim\)19,500 type 1 DM patients\(^23\). When genetic effects on eGFR in DM versus noDM have little overlap, such focused searches will remain the best option. For shared genetics, GWAS on eGFR overall will have much better power.

The quest for understanding commonalities and differences of genetic eGFR-effects between DM and noDM has two aspects: a search for eGFR-associated loci with DM/noDM-difference will help understand whether such differential effects exist and a search for SNP-association on eGFR allowing for DM/noDM-difference may detect novel eGFR loci. Statistically, DM/noDM-difference in eGFR association and genetics-by-DM-status interaction on eGFR association are equivalent. There are various approaches to do so\(^24\). These approaches have been applied successfully for other phenotypes\(^25-29\), but not yet for GWAS interaction analyses for kidney function.

We thus set out to search for eGFR loci with DM/noDM-difference and for novel eGFR loci allowing for difference. For this, we gathered GWAS data on eGFR separately for 178,691 DM and 1,296,113 noDM individuals. We prioritized genes underneath identified loci using in-silico functional evidence. We also evaluated the impact of DM-/noDM-specific weights on the genetic risk score (GRS) for eGFR in data independent from the variant- and weight-identifying step\(^30\).

Results

Overview of the GWAS meta-analyses. We first analyzed GWAS data on 7,046,926 single nucleotide polymorphisms (SNPs) and their association with logarithm-transformed eGFR separately for 109,993 DM (type 2 or type 1) and 1,070,999 noDM individuals (stage 1, including 72 studies from CKDGen Consortium\(^31\) and UK Biobank, UKB\(^32\); mostly European-ancestry; Supplementary Data 1, Methods, Fig. 1a). These DM-/noDM-stratified GWAS summary statistics allowed us to apply the difference test, the joint test and the stratified tests (Methods). Based on these tests, we searched for eGFR-associated loci with DM/noDM-difference and for novel eGFR-associated loci allowing for DM/noDM-difference.

We sought replication of identified loci in independent DM-/noDM-stratified data (stage 2, \(n\)\(_{DM}\) = 68,698, \(n\)\(_{noDM}\) = 225,114; Million Veterans Program, MVP\(^33\); Michigan Genomics Initiative, MGI\(^34\); Trøndelag health study, HUNT\(^35\); all European-ancestry; Supplementary Data 1, Methods). While the discovery+replication design augments confidence in identified loci, this comes at the cost of lower power\(^36\). We also conducted GWAS searches in a combined stage design (stage 1 + 2: \(n\)\(_{DM}\) = 178,691, \(n\)\(_{noDM}\) = 1,296,113), to yield significant loci exploiting the full sample size, yet without independent replication.

The two stages of data were also used to separate the variant identification and weight quantification for the GRS (stage 1 data) from the GRS association analyses (stage 2 data). The analysis workflow is shown in Fig. 1.

Seven eGFR loci identified with DM-/noDM-differential effects. To search for eGFR loci with DM/noDM-difference, we applied two approaches\(^24,37\) (Fig. 1b): (i) a genome-wide difference test (\(P\)\(_{\text{diff}}\) < 5 \times 10^{-8}, difference test approach), and (ii) a search for genome-wide significant association with overall eGFR followed by a difference test in the same data (\(P\)\(_{\text{overall}}\) < 5 \times 10^{-8} and \(P\)\(_{\text{diff}}\) < 0.05, k = number of followed SNPs, overall+difference test approach, Methods).

In the discovery search (stage 1, \(n\)\(_{DM}\) = 109,993; \(n\)\(_{noDM}\) = 1,070,999), we identified four eGFR loci with significant DM/noDM-difference (Supplementary Fig. 1): (i) two by the difference test approach (\(rs77924615\) near UMOD-PDILT, \(rs12233328\) near PDE9A; \(P\)\(_{\text{diff}}\) < 5 \times 10^{-8}, Supplementary Data 2), and (ii) two further loci by the overall+difference test approach (near TPPP and MEDI-NEUROD2; \(P\)\(_{\text{Overall}}\) < 5 \times 10^{-8} and \(P\)\(_{\text{diff}}\) < 0.05/610 = 8.2 \times 10^{-5}; corrected for 610 followed variants\(^16\), Supplementary Data 2). Details of the overall+difference test results are provided in Supplementary Note 1 and Supplementary Data 3, conditional difference analyses on known \(P\)\(_{\text{Overall}}\) < 5 \times 10^{-8} variants in UMOD-PDILT in Supplementary Fig. 2. In stage 2 data (\(n\)\(_{DM}\) = 68,698, \(n\)\(_{noDM}\) = 225,114), three of the four loci replicated (near UMOD-PDILT, MEDI-NEUROD2, and TPPP; one-sided \(P\)\(_{\text{diff}}\) < 0.05/4 = 0.0125; Table 1, Supplementary Data 2), while the PDE9A locus variant did not (\(P\)\(_{\text{diff}}\) = 0.30).

In the combined stage design (stage 1 + 2: \(n\)\(_{DM}\) = 178,691; \(n\)\(_{noDM}\) = 1,296,113), we identified seven eGFR loci with significant DM/noDM-difference (Table 1, Fig. 2a, Supplementary Fig. 3): (i) six loci by the difference test approach (\(P\)\(_{\text{diff}}\) < 5 \times 10^{-8}, Supplementary Data 2) including the three loci already found by the discovery+replication approach now all with \(P\)\(_{\text{diff}}\) < 5 \times 10^{-8} (near UMOD-PDILT, TPPP, and MEDI-NEUROD2), and three additional loci (near CSRNP, DCDC5, and NRIP1), (ii) one locus...
from the overall+difference test approach (near SLC22A2; \( P_{\text{overall}} < 5 \times 10^{-8} \) and \( P_{\text{diff}} < 0.05/610 = 8.2 \times 10^{-5} \); 610 followed variants as described above, Supplementary Data 2).

Approximate conditional analyses using the difference test (combined stage 1 + 2, European only, Methods) did not identify any further independent variants within the 7 identified difference loci (regional difference plots in Supplementary Fig. 4). Interestingly, there was also no further independent variant with DM/noDM-difference in the UMOD-PDILT locus besides rs77924715, despite the multiple independent variants known from the overall analysis (Supplementary Note 2, Supplementary Data 2).

Approximate conditional analyses using the difference test (combined stage 1 + 2, European only, Methods) did not identify any further independent variants within the 7 identified difference loci (regional difference plots in Supplementary Fig. 4). Interestingly, there was also no further independent variant with DM/noDM-difference in the UMOD-PDILT locus besides rs77924715, despite the multiple independent variants known from the overall analysis (Supplementary Note 2, Supplementary Data 2).

Genetic risk score (GRS) analyses

634 index SNPs from Stanickz et al. (CKDGen + UKB i.e. stage 1 data); overall or DM/noDM-specific weights from stage 1 data

**Fig. 1 Data and analysis workflow.** a Overview on datasets and meta-analyses. b Approaches to identify DM/noDM-differences. c Approaches to identify novel eGFR loci and suggestive DM/noDM-differences. d Genetic risk score (GRS) analyses in HUNT.

Allowing for difference identifies 32 novel eGFR-associated loci including CUBN. To search for novel eGFR-associated loci allowing for the difference between DM and noDM, we applied two approaches (Fig. 1c, Methods): (i) a genome-wide joint test, which considers the main SNP effect on eGFR and the SNP-by-DM-status interaction effect jointly (joint-test approach) and (ii) GWAS separately in DM or noDM (stratified tests approach).

In the discovery (stage 1; \( n_{\text{DM}} = 109,993, n_{\text{noDM}} = 1,070,999 \)), we identified 25 novel eGFR loci after excluding previously identified GWAS eGFR loci (near CUBN; Table 2). The 29 additional novel loci include the 32 DM-only loci replicated (near SLC22A2, SERTAD2, and ABCC2, Table 2).

In summary, we established seven eGFR loci with DM/noDM-difference (near UMOD-PDILT, TPPP, DCDC5, NRIPI, and SLC22A2) including one with DM-only (NRIPI) and one with noDM-only effect (MED1-NEUROD2). While all seven difference loci were known loci for eGFR, these are the first eGFR loci reported with significant DM/noDM-difference from a genome-wide search for the difference to the best of our knowledge.
Table 1 Search for difference loci identified seven eGFR loci with established difference between DM and noDM.

<table>
<thead>
<tr>
<th>Locus</th>
<th>rsid</th>
<th>Chr:Pos</th>
<th>ea</th>
<th>eaf</th>
<th>beta</th>
<th>DM</th>
<th>noDM</th>
<th>PDiff</th>
<th>PnoDM</th>
<th>Diff</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Overall</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1.5E-04</td>
<td>2.6E-07</td>
<td>2.6E-04</td>
<td>7.8E-04</td>
<td>3.8E-08</td>
<td>5.4E-09</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>4.2E-05</td>
</tr>
</tbody>
</table>

Approximate conditional analyses (combined stage, European only, Methods) identified two additional variants independently associated with eGFR at two joint-test loci (near SERTAD2 and PIK3CG, $P_{\text{Joint} \text{cond}} < 5 \times 10^{-8}$, Supplementary Table 2). This raised the total number of newly identified eGFR-associated variants to 34.

Among the 34 novel eGFR variants, four showed a suggestive DM/noDM-difference for eGFR ($P_{\text{Diff}} < 0.0015 = 0.05/34$, Table 2, Fig. 3b): one was a DM-only effect (near SH3BP4) and three showed more pronounced effects in DM (near LOX1A, ALPL, and PIK3CG). These patterns were consistent also in stage 1 and stage 2 separately (Supplementary Fig. 5). Of note, the difference test is mathematically dependent on the joint test and the stratified tests, but not on the overall test. Therefore, the observed difference in the overall + difference test approach applied to the same data is an established significant difference, while the observed difference in loci established by the joint or stratified tests is termed here as suggestive difference.

In summary, the joint test and stratified tests identified 34 genome-wide significant independent eGFR variants across 32 novel eGFR loci. These included four loci with suggestive DM/noDM-difference, which yielded a total of 11 eGFR loci with established or suggestive DM/noDM-difference.

Interaction with Hba1C and overlap with associations for DM risk, glycemic and other traits. We explored the lead variants of the 11 eGFR loci identified with DM/noDM-difference (i.e., established/suggestive SNP-by-DM interaction) for SNP-by-Hba1C interaction in UK Biobank ($n = 368,005$), making full use of the continuous variable Hba1C instead of binary DM status. For all 11 variants, the SNP main effects and SNP-by-Hba1C interaction effect sizes were directionally consistent with main and interaction effects sizes in the SNP-by-DM interaction analysis (Supplementary Data 7, Methods). This underscored again the negative interaction effect for the 10 of the 11 variants with larger effects in DM or DM-only effects, while the one variant (near MED1/NEUROD2) showed a negative interaction effect for the SNP-by-DM as well as the SNP-by-Hba1C interaction.

We were also interested in whether any of the 11 loci with DM/noDM-difference overlapped with known genome-wide significant loci for type 1 DM,$^{42}$ type 2 DM,$^{49}$ or glycemic traits.$^{43}$ None of the 11 loci overlapped with type 1 DM, type 2 DM, or glycemic traits loci (all 11 lead variants $P > 5 \times 10^{-8}$ in published GWAS for DM risk, glucose, or insulin levels, Supplementary Data 8). None of the 11 variants was associated with type 1 DM judged at Bonferroni-corrected significance ($P > 0.05/11 = 4.5 \times 10^{-3}$). Three of the 11 variants were associated with type 2 DM ($P < 0.05/11 = 4.5 \times 10^{-3}$, rs77924615 near UMOD-PDILT, rs55722796 near MED1/NEUROD2) or fasting glucose (rs963387 near DCDC5). For all three variants, the eGFR-lowering allele was associated with decreased type 2 DM risk or fasting glucose. Thus, the SNP effect on eGFR cannot be fully explained by DM status, which would have yielded a DM risk increasing or glucose-increasing effect by the eGFR-lowering allele. The observation is in line with a pleiotropic effect on DM and eGFR, but with adverse effects for one of the two (eGFR or DM/glucose) and a beneficial effect for the other, which should be considered in drug design when applicable. When taking into account the DM/noDM-stratum where the eGFR effect was more pronounced, the three variants consisted of (i) two variants with stronger eGFR-effects in DM (still significant eGFR effect in noDM, near UMOD/PDILT and DCDC5) and (ii) one variant with a noDM-only effect on eGFR (near MED1/NEUROD2).

We also queried the Open Targets Genetics database$^{44}$ for associations of the 11 variants with other traits. We found 126
GWAS for eGFR in individuals with DM identified 29 genome-wide significant loci. Previous searches for SNP-effects on eGFR-based kidney function outcome in DM patients had identified eight genome-wide significant loci (i.e., eGFR-based DKD loci; four identified in type 1 DM patients, four other in type 1 or type 2 DM patients). We wanted to understand the overlap of these with the genome-wide significant loci in our DM-only GWAS on eGFR ($n_{DM} = 178,691$, combined stage, Methods): we observed 29 genome-wide significant eGFR loci in DM ($P_{DM} < 5 \times 10^{-8}$, Fig. 5, Supplementary Data 10). These included: (i) 27 novel eGFR loci in DM, and (ii) two of the four previously identified DKD loci among type 1 or type 2 DM patients (near UMOD/PDILT and PRKAG2; $r^2 = 0.60$ and 0.94, respectively, between eGFR and DKD lead variants). None of the previously identified DKD loci among type 1 DM patients were identified. Two of the 27 novel eGFR loci in DM were detected here as completely novel eGFR loci (near SH3BP4 and LOXL4; also detected by the joint test). The 29 identified eGFR loci in DM included 6 of our 11 difference loci (near UMOD/PDILT, TPPP, DCDC5, SH3BP4, NRIP1, and LOXL4).

In silico evidence prioritized 18 genes by protein- or expression-altering variants or as human kidney disease monogene. Genes at GWAS loci for eGFR might pinpoint relevant drug targets for kidney function. Dissecting eGFR loci as DM-only, noDM-only, or shared will help define the target population for potential therapy. Gene PrioritiSation (GPS) was conducted previously for all genes under the 424 eGFR-associated loci in a meta-analysis of CKDGen and UKB (i.e., our stage 1 studies) focused on European-ancestry. Since all 7 difference loci were among these 424 loci, we used this established GPS tool to extract in-silico functional evidence for the 159 genes.

GWAS for eGFR in individuals with DM identified 29 genome-wide significant loci. Previous searches for SNP-effects on eGFR-based kidney function outcome in DM patients had identified eight genome-wide significant loci (i.e., eGFR-based DKD loci; four identified in type 1 DM patients, four other in type 1 or type 2 DM patients). We wanted to understand the overlap of these with the genome-wide significant loci in our DM-only GWAS on eGFR ($n_{DM} = 178,691$, combined stage, Methods): we observed 29 genome-wide significant eGFR loci in DM ($P_{DM} < 5 \times 10^{-8}$, Fig. 5, Supplementary Data 10). These included: (i) 27 novel eGFR loci in DM, and (ii) two of the four previously identified DKD loci among type 1 or type 2 DM patients (near UMOD/PDILT and PRKAG2; $r^2 = 0.60$ and 0.94, respectively, between eGFR and DKD lead variants). None of the previously identified DKD loci among type 1 DM patients were identified. Two of the 27 novel eGFR loci in DM were detected here as completely novel eGFR loci (near SH3BP4 and LOXL4; also detected by the joint test). The 29 identified eGFR loci in DM included 6 of our 11 difference loci (near UMOD/PDILT, TPPP, DCDC5, SH3BP4, NRIP1, and LOXL4).

GWAS for eGFR in individuals with DM identified 29 genome-wide significant loci. Previous searches for SNP-effects on eGFR-based kidney function outcome in DM patients had identified eight genome-wide significant loci (i.e., eGFR-based DKD loci; four identified in type 1 DM patients, four other in type 1 or type 2 DM patients). We wanted to understand the overlap of these with the genome-wide significant loci in our DM-only GWAS on eGFR ($n_{DM} = 178,691$, combined stage, Methods): we observed 29 genome-wide significant eGFR loci in DM ($P_{DM} < 5 \times 10^{-8}$, Fig. 5, Supplementary Data 10). These included: (i) 27 novel eGFR loci in DM, and (ii) two of the four previously identified DKD loci among type 1 or type 2 DM patients (near UMOD/PDILT and PRKAG2; $r^2 = 0.60$ and 0.94, respectively, between eGFR and DKD lead variants). None of the previously identified DKD loci among type 1 DM patients were identified. Two of the 27 novel eGFR loci in DM were detected here as completely novel eGFR loci (near SH3BP4 and LOXL4; also detected by the joint test). The 29 identified eGFR loci in DM included 6 of our 11 difference loci (near UMOD/PDILT, TPPP, DCDC5, SH3BP4, NRIP1, and LOXL4).
Table 2 Allowing for DM/noDM difference identified 32 novel eGFR loci.

<table>
<thead>
<tr>
<th>Locus</th>
<th>rsid</th>
<th>Source</th>
<th>Stage 1</th>
<th>Stage 2</th>
<th>Combined stage</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>dp</td>
<td>dp</td>
<td></td>
</tr>
<tr>
<td>Discovery-1 replications design</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[ATPI2A]</td>
<td>rs7324815</td>
<td>joint</td>
<td>C</td>
<td>0.15</td>
<td>-0.0039</td>
</tr>
<tr>
<td>[SERTAD2]</td>
<td>rs12614829</td>
<td>joint/noDM</td>
<td>T</td>
<td>0.76</td>
<td>-0.0007</td>
</tr>
<tr>
<td>[ABC2]</td>
<td>rs56080571</td>
<td>joint/noDM</td>
<td>G</td>
<td>0.06</td>
<td>-0.0037</td>
</tr>
<tr>
<td>Combined stage design</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>with suggestive difference</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[SHBP4]</td>
<td>rs4663771</td>
<td>joint/dm</td>
<td>A</td>
<td>0.16</td>
<td>-0.0075</td>
</tr>
<tr>
<td>[ALPL]</td>
<td>rs36503309</td>
<td>joint</td>
<td>C</td>
<td>0.79</td>
<td>-0.0025</td>
</tr>
<tr>
<td>[LOXL4]</td>
<td>rs11589562</td>
<td>joint/dm</td>
<td>G</td>
<td>0.32</td>
<td>-0.0028</td>
</tr>
<tr>
<td>[PIK3CG]</td>
<td>rs2392929</td>
<td>joint/dm</td>
<td>G</td>
<td>0.19</td>
<td>-0.0037</td>
</tr>
<tr>
<td>without suggestive difference</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[DOC7]</td>
<td>rs10789120</td>
<td>joint</td>
<td>G</td>
<td>0.71</td>
<td>-0.0039</td>
</tr>
<tr>
<td>[EBF2]</td>
<td>rs13499370</td>
<td>joint</td>
<td>C</td>
<td>0.63</td>
<td>-0.0064</td>
</tr>
<tr>
<td>[KIF18A]</td>
<td>rs1730049</td>
<td>joint</td>
<td>T</td>
<td>0.82</td>
<td>-0.0058</td>
</tr>
<tr>
<td>[BACH1]</td>
<td>rs379592</td>
<td>joint</td>
<td>G</td>
<td>0.48</td>
<td>-0.0029</td>
</tr>
<tr>
<td>[CUBN]</td>
<td>rs1258238</td>
<td>joint</td>
<td>C</td>
<td>0.30</td>
<td>-0.0042</td>
</tr>
<tr>
<td>[RASSF6]</td>
<td>rs17804499</td>
<td>joint/dm</td>
<td>G</td>
<td>0.05</td>
<td>-0.0064</td>
</tr>
<tr>
<td>[TET2]</td>
<td>rs6749069</td>
<td>joint/noDM</td>
<td>C</td>
<td>0.83</td>
<td>-0.0036</td>
</tr>
<tr>
<td>[FXL17]</td>
<td>rs58585707</td>
<td>joint/noDM</td>
<td>A</td>
<td>0.15</td>
<td>-0.0037</td>
</tr>
<tr>
<td>[NRBF2]</td>
<td>rs130995</td>
<td>joint</td>
<td>G</td>
<td>0.52</td>
<td>-0.0026</td>
</tr>
<tr>
<td>[TTCL2]</td>
<td>rs4623646</td>
<td>joint/noDM</td>
<td>C</td>
<td>0.64</td>
<td>-0.0029</td>
</tr>
<tr>
<td>[PSEN2]</td>
<td>rs1295641</td>
<td>joint/noDM</td>
<td>T</td>
<td>0.50</td>
<td>-0.0014</td>
</tr>
<tr>
<td>[REM2]</td>
<td>rs2299504</td>
<td>joint</td>
<td>C</td>
<td>0.17</td>
<td>-0.0015</td>
</tr>
<tr>
<td>[DBL6]</td>
<td>rs7815867</td>
<td>joint/noDM</td>
<td>C</td>
<td>0.80</td>
<td>-0.0022</td>
</tr>
<tr>
<td>[AUTS2]</td>
<td>rs3750170</td>
<td>joint</td>
<td>A</td>
<td>0.33</td>
<td>-0.0022</td>
</tr>
<tr>
<td>[SLC2A2]</td>
<td>rs7630490</td>
<td>joint/noDM</td>
<td>C</td>
<td>0.84</td>
<td>-0.0033</td>
</tr>
<tr>
<td>[ZFP361]</td>
<td>rs7375164</td>
<td>joint/noDM</td>
<td>T</td>
<td>0.19</td>
<td>-0.0014</td>
</tr>
<tr>
<td>[FBX2]</td>
<td>rs1944933</td>
<td>joint/noDM</td>
<td>T</td>
<td>0.53</td>
<td>-0.0008</td>
</tr>
<tr>
<td>[EEF1D3]</td>
<td>rs703214</td>
<td>joint</td>
<td>G</td>
<td>0.49</td>
<td>-0.0022</td>
</tr>
<tr>
<td>[DUSP6]</td>
<td>rs4722122</td>
<td>joint</td>
<td>C</td>
<td>0.83</td>
<td>-0.0023</td>
</tr>
<tr>
<td>[ARM4]</td>
<td>rs7896951</td>
<td>joint/noDM</td>
<td>A</td>
<td>0.47</td>
<td>-0.0011</td>
</tr>
<tr>
<td>[SLC2A5A]</td>
<td>rs2280206</td>
<td>joint/noDM</td>
<td>G</td>
<td>0.46</td>
<td>-0.0001</td>
</tr>
<tr>
<td>[MSC]</td>
<td>rs3228719</td>
<td>joint/noDM</td>
<td>C</td>
<td>0.91</td>
<td>-0.0018</td>
</tr>
<tr>
<td>[LIMCH1]</td>
<td>rs75228450</td>
<td>joint</td>
<td>A</td>
<td>0.87</td>
<td>-0.0024</td>
</tr>
<tr>
<td>[SLC2A4]</td>
<td>rs117643180</td>
<td>noDM</td>
<td>G</td>
<td>0.02</td>
<td>0.0051</td>
</tr>
<tr>
<td>[FA14]</td>
<td>rs1506363</td>
<td>joint</td>
<td>G</td>
<td>0.83</td>
<td>0.0035</td>
</tr>
</tbody>
</table>

Shown are 32 novel variants identified with genome-wide significant eGFR association allowing for DM/noDM-difference by two designs and two approaches (joint or strafified test approaches) using stage 1 (n_{noDM} = 109,993; n_{DM} = 1,070,993), stage 2 (n_{noDM} = 68,698; n_{DM} = 22,5114), or combined data (n_{noDM} = 178,691; n_{DM} = 129,6113). Discovery-1 replications design: 3 variants with P_{noDM}/P_{DM} > 5 × 10^{-10} in stage 1 were replicated in stage 2 (P_{noDM}/P_{DM} < 0.05/25 = 0.002). (ii) Combined stage design: 29 additional variants were identified with P_{noDM}/P_{DM} > 5 × 10^{-10} in stage 1 + 2 (with suggestive difference, P_{noDM}/P_{DM} < 0.05/20). Details are shown in Supplementary Data 6. Source is the identifying test.
Fig. 3 Accounting for potential DM-/noDM-differences identified 32 novel eGFR loci. We searched for novel loci associated with eGFR allowing for DM-/noDM-difference using the joint test or DM-/noDM-stratified tests approaches in combined stage 1 + 2 (n_{DM} = 178,691; total n_{noDM} = 1,296,113). We found 32 novel genome-wide significant eGFR loci (P < 5 × 10^{-8}, >500 kbp distant of known eGFR loci compared to previous work [16,20]: 30 by joint, 17 by noDM-only and 2 by DM-only test. **a** Shown are P-values for eGFR based on joint, noDM-only, and DM-only test over chromosomal position. Highlighted in red are loci with suggestive DM/noDM-difference (P_{Diff} < 0.05/34; corrected for 34 independent variants across 32 loci), blue for loci identified by joint and noDM-only test (15 loci), and purple for loci that were only identified by joint test (upper panel) or noDM-only test (middle panel). Loci were annotated by nearest genes if P_{Diff} < 0.10 or if they were also identified by the discovery+replication design (the latter also indicated by asterix, Table 2). **b** Shown is a comparison of DM-/noDM-specific eGFR-effect sizes for the 32 novel eGFR locus lead variants. Highlighted in red are the locus names of loci with suggestive DM/noDM-difference (P_{Diff} < 0.05/34; corrected for 34 independent variants across 32 loci). Effect sizes are aligned to the eGFR-decreasing alleles in noDM. Error bars reflect 95% confidence intervals of the estimated genetic effect.

Fig. 4 The variants associated with eGFR in the CUBN locus differ from those associated with urinary albumin-to-creatinine ratio. Shown are P-values for associations at the wider (top) and more narrow (bottom) CUBN locus region for **a** eGFR (joint test P-values, n_{DM} = 178,691 and n_{noDM} = 1,296,113) and **b** urinary albumin-to-creatinine ratio (UACR; P-values from ref. 39, n = 564,257). Lead variant for eGFR is rs11254238; color codes variants’ correlation r^2 to rs11254238 in all panels.
at the 7 difference loci: here, we prioritized a gene when it mapped to an association-driving variant (i.e., 99% credible set variant with posterior probability of association ≥ 5%, Methods) that was relevant (i) for the protein (with high predicted deleteriousness\(^45\), i.e., CADD score ≥ 15) or (ii) for expression in kidney tissue (eQTL, false-discovery rate ≥ 5%)\(^46,47\) or (iii) when the gene was a known kidney disease monogene in human (Methods). We found 6 prioritized genes (Fig. 6, Supplementary Fig. 8): (i) \(\text{PGAP3}\) mapping to the locus with noDM-only effect on eGFR (near \(\text{MED1/NEUROD2}\)) and (ii) five genes in loci with more pronounced eGFR-effects in DM but direction-consistent effects in noDM (\(\text{TPPP}, \text{UMOD}, \text{SLC6A19}, \text{NRIP1},\) and \(\text{SLC22A2}\)).

We also used the same approach as in the previous GPS tool\(^16\) to prioritize the genes underneath the 32 novel eGFR loci (34 independent signals, Supplementary Data 11–15, Supplementary Table 3, Supplementary Fig. 9): (i) among the 40 genes in the four novel loci with suggestive DM/noDM-difference, 2 genes were human kidney disease monogenes (\(\text{HPS1} \) and \(\text{HPSE2}\), Fig. 7a); (ii) for the 341 genes in the other 28 novel eGFR loci (i.e., noDM/noDM-difference), 10 genes were prioritized (Fig. 7b): 6 genes contained an association-driving variant that was protein-deleterious (\(\text{AUTS2}, \text{CUBN}, \text{DVL2}, \text{RASSF6}, \text{SLC2A4},\) and \(\text{ZFP36L1}\)), one gene mapped to an eQTL in glomerular tissue (\(\text{TNIK}\)), and 3 additional genes were human kidney disease monogenes (\(\text{SLC2A2}, \text{SLC30A9},\) and \(\text{SLC7A7}\)). Particularly interesting was an association-driving variant, rs1801232, in \(\text{CUBN}\) that was a missense variant (\(r^2 = 0.73\) to eGFR lead variant rs11254238). The missense variant rs17804499 in \(\text{RASSF6}\) had a particularly high probability of being the association-driving variant (posterior probability of association = 86%), which rendered this variant and gene a compelling candidate for

**Fig. 5** DM-only eGFR GWAS identified 29 loci, including 27 novel for eGFR in DM. Shown are eGFR association P-values in individuals with DM over chromosomal position in combined stage (\(n_{\text{DM}} = 178,691\)). This DM-specific analysis identified 29 independent eGFR-associated loci in DM. Compared to known DKD loci\(^22,23\) (i.e., association with eGFR or CKD in type 1 and/or type 2 DM individuals) and known overall eGFR loci\(^16,20\), 2 loci are novel for eGFR overall and novel for DKD (red), 24 are novel for DKD but known for eGFR (orange), and 3 are known DKD and known eGFR loci (purple).

**Fig. 6** Gene prioritization highlights six genes at loci with established DM/noDM-difference. Shown are gene prioritization results for the seven loci with established difference (Table 1, Supplementary Fig. 8). We highlighted six genes based on association-driving variants (PPA > 5%) that were deleteriously protein-relevant or expression-modulating, genes that were known as human kidney monogenes (OMIM or ref. \(^78\) with subsequent manual curation) and in addition SLC22A2 due to its known link to metformin response.
Two genes, SLC2A2 and SLC2A4, encode the glucose transporters GLUT2 and GLUT4 and reside in two novel eGFR loci. In summary, across the identified seven loci identified with established or suggestive DM-/noDM-difference in eGFR-effects and the 32 novel eGFR loci, we prioritized 18 genes with evidence of protein-deleterious or expression-modulating variants or being human kidney disease monogenes. These genes are compelling candidates for functional follow-up and drug development pipelines.

DM/noDM-stratified weights for GRS show similar explained variance as overall weights. We assessed whether the GRS for eGFR, computed as an effect size weighted sum of eGFR-lowering alleles, could be improved by using DM/noDM-specific eGFR-effects as weights rather than overall eGFR-effects. We built the GRS based on the previously identified 634 eGFR variants (i.e., identified in stage 1 studies) or based on the seven variants with established DM/noDM-difference using different weighting schemes (overall effects, DM-/noDM-specific effects, or a combination; all weights from stage 1 data). We computed the GRS for each individual in the stage 2 study HUNT (unrelated, European-ancestry; n_DM = 3023, n_noDM = 25,484, Fig. 1d, Methods).

Across all weighting schemes, we found no significant DM/noDM-difference in the 634-variant GRS association with eGFR (P_Diff > 0.05, Table 3). The 634-variant GRS explained more of the eGFR variance in noDM compared to DM (e.g., R² = 6.0% vs. 4.0% using overall-effect weights). Since the absolute 634-variant GRS effect was similar in DM and noDM (e.g., beta per sdGRS = −2.54 versus −2.84 ml/min/1.73 m² using overall-effect weights), this larger relative GRS effect in noDM can be attributed...
—at least in part—to the smaller eGFR variance in noDM versus DM (HUNT standard deviation = 11.8 versus 13.4 ml/min/1.73 m², respectively). Yet, the 7-variant GRS explained more of the eGFR variance in DM compared to noDM (e.g., \( R^2 = 0.98 \% \) vs. 0.62\% using overall-effect weights). The reason for this is that six of the seven variants had larger eGFR-effects in DM than in noDM, which accumulated to a larger absolute GRS effect in DM (e.g., beta per sd GRS = −1.02 in DM compared to −0.68 in noDM using overall-effect weights).

When comparing different weighting schemes, we found no notable improvement in \( R^2 \) values by using DM-/noDM-specific weights compared to overall weights (Table 3). In the 634-variant GRS, using DM-specific weights even reduced \( R^2 \) in DM individuals (\( R^2 = 3.1\% \) versus 4.0\%). This may be attributable to the larger uncertainty in DM-specific weights estimated from the meta-analysis restricted to DM individuals. For the 7-variant GRS, DM-specific weights slightly increased \( R^2 \) in DM individuals, but not markedly (\( R^2 = 1.0\% \) versus 0.98\%). These findings underscore a similar performance of overall weights and DM-/noDM-specific weights when building a GRS for eGFR. The comparison of the 634-variant with the 7-variant GRS suggests that more differential effect loci and higher precision in DM-specific weights might substantially improve the genetically explained variance also in DM individuals.

### Discussion

In this GWAS on eGFR in ~180,000 individuals with DM and ~1.3 million individuals without DM, we established seven loci with significant DM/noDM-differential effects on eGFR near UMOD/PDLT, TPPI, MEDI/NEUROD2, CSRNP1, DCCD5, NRIP1, and SLC22A2. We also identified 32 novel eGFR-associated loci when allowing for potential difference, including four loci with a suggestive difference near SH3BP4, LOXL4, ALPL, and PIK3CG. The 11 loci with established or suggestive difference included two loci with DM-only effects on eGFR (near CSRNP1, SH3BP4), one with noDM-only effect (near MEDI/NEUROD2), and all others showed more pronounced effects in DM compared to noDM. Our GWAS focused on DM individuals confirmed two known DKD loci, but also identified 27 novel loci for eGFR in DM, which make these potential new DKD loci. The DM-/noDM-stratified GRS analyses showed no improvement in explained trait variance when using DM-/noDM-stratified weights, but the seven identified DM/noDM-differential eGFR-effects explained more of the eGFR variance in DM individuals than in noDM individuals.

Our gene prioritization at the 11 loci with DM/noDM-difference and at the 28 other novel eGFR loci highlighted 18 genes: 10 genes mapped to eGFR-association-driving variants that were deleteriously protein-relevant or eQTLs in kidney tissue, which suggests them as potential drug targets to alter eGFR. Further eight genes were prioritized because they were known as human kidney disease monogenes, which made them plausible causal genes for these common variant findings. Particularly interesting was \( CUBN \) in a novel eGFR locus. The \( CUBN \) locus is the GWAS locus with the strongest effect on UACR and microalbuminuria and identified here for eGFR at genome-wide significance, but respective lead variants for eGFR and UACR are uncorrelated. The lead variant for eGFR, rs11254238, is highly correlated with a variant that alters the encoded protein, cubulin (rs1801232, \( r^2 = 0.91 \)). We observed a twice as large effect on eGFR in DM compared to noDM in both discovery and replication data, but not statistically significantly different. This is in line with previously observed larger \( CUBN \) variants’ effects on microalbuminuria and UACR in DM, which provides further evidence for an interaction of \( CUBN \) with DM status on kidney function. In fact, our previous sequencing of the \( CUBN \) gene found rare variants in \( CUBN \) that were associated with higher UACR and with better eGFR. Our study now reports such a parallel association between UACR and eGFR now also for a common variant at the genome-wide significance levels. Together, our results provide further support for the importance to assess the physiological role of \( CUBN \) by functional studies not only for microalbuminuria, but also for finding pathophysiological explanations related to impaired filtration rate.

The \( RASSF6 \) gene has been studied in relation to the kidney as well: \( RASSF6 \) mediates apoptosis in various cells. It is shown to be expressed in the slit diaphragm in glomeruli and the apical membranes in proximal renal tubular epithelial cells of rat rodents.
kidney. The same animal study suggests the involvement of the RASSF6 pathway in contrast-induced nephropathy. Thus, the novel detection of the RASSF6 locus for eGFR here and the lead variant, rs17804499, being protein-altering and with 86% probability causal render RASSF6 a compelling candidate for the functional follow-up to help understand its role in kidney function that is elusive so far.

Another interesting gene was PGAP3 in the MEDI-NEUROD2 locus, which mapped to an expression-modulating variant that was associated with eGFR in noDM, but not in DM individuals. PGAP3 is a known gene for CKD and eGFR: eQTL analyses showed colocalization of the association signal with PGAP3 expression.\(^1\) PGAP3 knockout mice developed larger glomeruli with deposition of immunoglobulins, although we acknowledge that this does not provide a mechanistic explanation for its association with eGFR. Eventually, if causal at eGFR loci with significant differences by DM status will be advanced as drug targets for reno-protective therapies, the knowledge about DM- or noDM specificity of the eGFR association can help define the target group for potential subsequent therapy, as DM-only, noDM-only, or both.

The difference in SNP effects between DM and noDM can be interpreted as SNP-by-DM-status interaction effect. To our knowledge, our study is the first genome-wide search for SNP-by-DM-status interaction effects for eGFR. Sensitivity analyses indicated that the observed differences were neither biased from the stratified modeling, nor from log-transformed eGFR, and the SNP-by-DM-status interaction was not explained by SNP-interaction with age, sex, hypertension, or body-mass-index.\(^2\) The seven loci with an established difference are known loci for eGFR.\(^3\) To our knowledge, the significant DM/noDM-differential effects for eGFR or eGFR-based CKD were identified here for the first time, except for the UMOD-PDILT locus.\(^9\) At UMOD-PDILT, a well-known kidney function locus\(^9\), the two-fold higher eGFR effect in DM compared to noDM confirmed previous observations for eGFR\(^4,5\) and CKD.\(^5\) Very interesting was the TPPP locus, where the three-fold higher eGFR effect in DM now draws attention to a locus that was just one of the hundreds of small effects eGFR loci. The lead variant for the difference, rs434215, modulates TPPP expression in tubulo-interstitial tissue, but the role of TPPP in diabetes and kidney disease is yet unknown. Previous studies on TPPP focused on the nervous system: TPPP is highly expressed in the brain and shown to affect neural microtubules.\(^5\)

Our GWAS on eGFR in DM provides a link to the genetics of DKD. Previous GWAS for DKD analyzed eGFR or eGFR-based CKD in type 1 and/or type 2 DM individuals.\(^2,22\) Our GWAS in >178,000 DM individuals was 4- to 7-fold larger. These DM individuals were mostly from population-based studies of adult individuals. As such, the proportion of type 1 DM among DM individuals analyzed here reflects the proportion of type 1 DM of 5–10% among adult DM individuals.\(^7\) Due to the substantially larger number of type 2 DM individuals, the identified DM/noDM-differential effects on eGFR may mostly reflect differences between type 2 DM versus noDM. A better distinction by type of DM would require more granular data focusing on the distinction between these two major diabetes groups and substantially larger data on type 1 DM. While the use of self-reports or one-time measurements of glucose or HbA1c to define DM here is typical also for GWAS on DM,\(^9,58\) this implies heterogeneity in the DM definition and may include some individuals without clinically manifest DM.

Yet, our 29 loci with genome-wide significant association with eGFR in DM confirmed two of eight previously identified DKD loci\(^22,23\) and the 27 other are compelling new candidate loci for further analyses in studies specialized on DKD patients. A large sample size eGFR GWAS with a more heterogeneous spectrum of DM can be a powerful complementary approach to focused searches in DKD patients.\(^22,23\) Several prioritized genes mapped to these novel eGFR loci in DM: TPPP already mentioned above and SLC6A19, NRP1, HPS1 as well as HPS2E2 with strong monogenic impact on kidney. All these genes reside in loci with more pronounced eGFR-effects in DM compared to noDM, but none of these effects was DM-only.

Our GWAS allowing for DM/noDM-differential effects identified 32 novel eGFR-associated loci compared to previous GWAS.\(^6,20\) Some of our novel loci might be identified due to increased power by ~20% increased sample size or due to chance by using alternative statistical tests. However, four of the novel loci showed suggestive DM/noDM-difference, suggesting that their identification was facilitated by using tests allowing for differential effects.\(^9\) The four suggestive difference loci included one locus, near SH3BP4, with eGFR effect only in DM.

In contrast to hundreds of loci found in GWAS for eGFR, the eGFR loci with significant DM/noDM-difference were few. This suggests a largely shared genetics of kidney function between DM and noDM individuals, which has an important implication for drug development: most drug interventions aimed at altering eGFR should thus be effective among persons with and without DM. This is mirroring what is observed for SGLT2 inhibitors—a medication originally developed for individuals with DM that is now also being tested for reno-protection among individuals without DM.\(^13\) While our GWAS here is, to date, the largest for eGFR in DM individuals, we still might have missed loci with effects in DM-only. Power for interaction effects is generally smaller than for overall effect,\(^60\), and particularly reduced when one subgroup, like DM, is substantially smaller (~10%) than the other.\(^28\) Future work with an increased sample size might detect more eGFR loci with differences by DM status, especially those with effects in DM-only. This GWAS included mostly European individuals and our findings require replication in non-European ancestries, particularly because of the large differences in DM prevalence across ancestries.\(^62\)

In summary, our results highlight the existence of DM- and noDM-specific genetic effects on kidney function, but emphasize that the majority of eGFR locus associations do not differ between individuals with and without DM. Larger DM-/noDM-stratified data on eGFR in the future will improve the detectability of differential effect loci and the precision of DM-specific weights. This might also improve DM-/noDM-stratified GRS prediction of eGFR. The identified eGFR loci with difference between DM- and noDM individuals include loci with effects only in DM as well as loci with effects in noDM. This has highly relevant implications, if the respective genes are advanced as a drug target: the specificity of the association might help sharpen the target group for such potentially arising drug therapies.

**Methods**

**Definition of the outcome eGFR and study-specific participant information.** GFR was estimated in all study participants in all studies using serum creatinine measurements via the Chronic Kidney Disease Epidemiology Collaboration (CKD-EPI) formula.\(^20\) Study-specific information on the utilized assay and year of measurement is given in Supplementary Data 1. We used the R package nephro\(^63\), winsorized at 15 and 200 ml min\(^{-1}\)per 1.73 m\(^2\), and logarithmized using a natural logarithm. For a better interpretation of effect size, we there used eGFR on the original scale in GRS analyses. Study-specific information on study design, sample size, sex and age, utilized serum creatinine assay, year of measurement, and average eGFR is given in Supplementary Data 1.

Each study is conducted according to the declaration of Helsinki; local ethics committees approved research protocols and participants provided written informed consent.

**Definition of DM status.** DM of each study participant at the time point of the serum creatinine measurement was defined either (i) as fasting plasma glucose ≥126 mg/dl (7.0 mmol/L) or diabetes therapy, or (ii) (fasting glucose unavailable)
as non-fasting plasma glucose ≥200 mg/dL (11.0 mmol/L) or diabetes therapy, or (iii) glucose unavailable) as self-reported diabetes. For UKB, DM was defined as HbA1c ≥48 mmol/mol (6.6%) or diabetes therapy (i.e., A10 ATC codes obtained from 84).

**Study-specific GWAS analyses stratified by DM status.** We distributed an analysis plan to the 72 participating studies of the CKDGen consortium. Each study conducted analyses separately for individuals with DM and individuals without DM to detect novel loci and to replicate known loci. For Com- sortium v1.1 (HRC) or 1000 Genomes Project phase 3 v5 (1000Gp3v5) ALL or phase 1 v3 (1000Gp1v3) ALL panel. Each study conducted linear regression GWAS on log(eGFR) using natural logarithm, an additive genotype model as well as adjusted for sex, age, and other study-specific covariates. Details on study-specific genotyping, imputation, and GWAS in UKB as well as for stage 2 studies are shown in Supplementary Data 1. For quality control, we excluded variants with low imputation quality, Info < 0.6, or rare variants with minor allele frequency, MAF < 0.1%. We utilized the software packages GWAtoolbox and EasyQC for the quality control of study-specific GWAS results. We conducted a correction for genomic control lambda of the results stratified by diabetes status.

**DM/noDM-stratified GWAS meta-analyses.** In stage 1 of our analysis, separately in DM (nDM = 88,829) and noDM (nnoDM = 620,665) strata, we conducted fixed-effect inverse-variance weighted meta-analyses of 72 GWAS of log(eGFR) using meta85, and then meta-analyzed these results with DM/noDM-stratified GWAS in UKB (nUKB = 11,040; noDM-only). We adjusted for population stratification within studies, we applied genomic control (GC) correction88 to each study prior to the meta-analysis. We applied a second GC correction to the DM- and noDM stage 1 meta-analysis results (GC lambda = 1.02 and 1.20, respectively). We excluded variants that were present only in stage 1 studies (≤50%) and variants with a cumulative minor allele count of ≤400 in the stage 1 meta-analyses. In summary, 109,869 individuals with DM and 1,035,190 with noDM were included in stage 1. We followed variants identified at stage 1 in independent stage 2 meta-analyses. For stage 2, we included DM/noDM-stratified GWAS on log(eGFR) from MVP (nMVP = 37,430, nDM = 122,966, hospital-based), MG1 (nMG1 = 7469, nDM = 36,558, hospital-based) and HUNT (nHUNT = 3799, nDM = 65,590, population-based), totalling 68,698 individuals with DM and 225,114 with noDM, all of European-ancestry. Again, we applied study-specific GC correction prior to the meta-analysis and a second GC correction89 to the stage 2 meta-analysis results (GC lambda = 1.00 and 1.02 for DM- and noDM-stratified, respectively). To maximize power for locus identification, we combined the double GC-corrected stage 1 and 2 meta-analysis separately by DM status via fixed-effect inverse-variance weighted meta-analyses of the two sources using meta88. The GC lambda in this final meta-analysis was comparable to previous GWAS16,76 (GC lambda = 1.03 and 1.15, respectively). The DM-/noDM-specific summary statistics (stratified GWAS) for each variant genome-wide served to investigate potential DM-/noDM-differential genetic effects on log(eGFR) without making any assumptions on the DM association with any other covariate71. These DM-/noDM-specific summary statistics on genetic variants associated with log(eGFR) enabled the implementation of all the following statistical tests to search for DM/noDM-difference loci or novel loci allowing for difference as described below.

**Approaches to search for DM/noDM-difference in genetic effects on eGFR.** We used the meta-analyzed SNP-specific summary statistics, to test for difference in eGFR-effects between DM and noDM. For this, we applied a difference test for each variant using

$$Z_{\text{diff}} = \frac{\hat{\beta}_{\text{DM}} - \hat{\beta}_{\text{noDM}}}{\sqrt{\hat{\sigma}^2_{\text{DM}} + \hat{\sigma}^2_{\text{noDM}} - 2 \hat{r}_{\text{DM,noDM}} \hat{\sigma}_{\text{DM,noDM}}}}$$

(1)

where $\hat{\beta}_{\text{DM}}$ and $\hat{\beta}_{\text{noDM}}$ are the genetic effect estimates for eGFR in DM or noDM from the stratified GWAS meta-analysis, respectively, and corresponding standard errors $\hat{\sigma}_{\text{DM}}$ and $\hat{\sigma}_{\text{noDM}}$. The term $\hat{r}_{\text{DM,noDM}}$ reflects the correlation between $\hat{\beta}_{\text{DM}}$ and $\hat{\beta}_{\text{noDM}}$ across all variants (Spearman correlation coefficient, $\rho_{\text{DM,noDM}} = 0.14$).

To search genome-wide for eGFR loci with DM-/noDM-differential effects, we applied two approaches85,86,87,88: (i) a genome-wide difference test ($P_{\text{DM-noDM,diff}} < 5 \times 10^{-8}$, difference test approach), and (ii) a search for genome-wide significant association with overall eGFR followed by a difference test in the same data ($P_{\text{diff noDM, DM}} < 5 \times 10^{-8}$ and $P_{\text{diff noDM, DM}} < 0.05/k$, k = number of followed SNPs, overall-difference test approach). The two approaches complement each other in terms of power to detect difference loci that depend on the magnitude and DM/noDM specificity87. We implemented the two designs in two different pipelines: a discovery + replication design, where we searched for significant differences in the stage 1 meta-analysis and moved the selected SNPs to a replication stage using the stage 2 meta-analysis (applying a Bonferroni-corrected alpha-level accounting for the SNPs tested). Second, to make full use of stage 1 and stage 2 data, we searched for differences in stage 1 + 2 meta-analysis combined (combined stage design). We utilized the R package EasyStrata27 to apply the difference test approaches to the DM/noDM-stratified meta-analysis results.

**Approaches to search for eGFR loci allowing for DM/noDM-differences.** By allowing for differences between DM and noDM in SNP-effects on eGFR, one can investigate gene-disease interactions and identify SNPs that are common to the different stages (with or without DM status stratification). Again, we used the meta-analyzed SNP-specific summary statistics to apply two approaches to search for novel eGFR loci allowing for difference: (i) a screen using a two degrees of freedom joint test ($P_{\text{joint, DM-noDM,diff}} < 5 \times 10^{-8}$, joint-test approach)88 derived from a $\chi^2$-test using

$$C_{\text{joint}} = \frac{\hat{\beta}_{\text{DM}}^2}{\hat{\sigma}^2_{\text{DM}}} + \frac{\hat{\beta}_{\text{noDM}}^2}{\hat{\sigma}^2_{\text{noDM}}}$$

(2)

where $\hat{\beta}_{\text{DM}}$ and $\hat{\beta}_{\text{noDM}}$ are the genetic effect estimates for eGFR in DM or noDM from the stratified GWAS meta-analysis, respectively, and corresponding standard errors $\hat{\sigma}_{\text{DM}}$ and $\hat{\sigma}_{\text{noDM}}$; and (ii) a on eGFR association in individuals with or without DM separately (stratified tests approach, $P_{\text{DM, noDM}} < 5 \times 10^{-8}$) or ($P_{\text{DM, noDM}} < 5 \times 10^{-8}$ as done previously16; the most significant variant was selected genome-wide (first locus lead variant) and the corresponding locus was defined as the smallest physical interval on the corresponding chromosome containing this variant such that there were no genome-wide significant variants within 500 kb outside the two borders. Omitting the identified locus, we repeated the procedure until no further genome-wide significant variants were detected. By this, a locus region is defined by adding ±250 kb to the first and last genome-wide significant variant of an identified locus. This procedure also ensured non-overlapping loci. A locus was considered to be known for eGFR from previous GWAS, if it overlapped with one of the 427 known eGFR loci (424 from ref. 16, 3 additional from ref. 20). If a locus is not known, we call it a novel eGFR locus. For the identified variants, we assessed between-study heterogeneity based on the CKDGen meta-analysis using a Chi-Squared test and an I^2 statistic72 and verified the association statistics with regards to abnormal or unusually large effect sizes.

**GCTA analyses to identify independent secondary signals within loci.** To evaluate whether there were multiple independent signals within loci, we conducted approximate conditional analyses with GCTA84 for each identified locus. These analyses were based on European-ancestry individuals, since appropriate trans-ethnic linkage disequilibrium (LD) reference panels were limited. As LD is expected to be sparse, we used a random subset of 20,000 unrelated individuals of European-ancestry from UKB, as done previously16. To obtain independent signals for associations derived by the difference test, the joint test, or DM- or noDM-specific association tests, we applied a stepwise approach: we condition GCTA analysis on the, respectively, observed lead variant in DM and noDM separately, and, derived DM/noDM-specific conditioned results for the locus. Separately for DM and noDM, we ensured whether any of the conditioned DM- or noDM-specific associations in the locus showed genome-wide significant association at $P_{\text{DM, noDM, cond}} < 5 \times 10^{-8}$ or $P_{\text{DM, noDM, cond}} < 5 \times 10^{-8}$ for loci derived by the difference test or the joint test, we applied the difference test or joint test, respectively, to the conditioned DM/noDM-estimates to infer whether additional signals showed significant differences or joint effects ($P_{\text{DM, noDM, cond}} < 5 \times 10^{-8}$ or $P_{\text{DM, noDM, cond}} < 5 \times 10^{-8}$).

**Prioritization of variants and genes.** For each variant within each identified signal, we derived the variants that were the most likely to drive the association. For this, we calculated approximate Bayes factor and posterior probabilities of association (PPA) based on Z-scores using the Kichaev method. We calculated PPA-based on unconditioned or conditioned summary statistics depending on whether the locus showed only one or multiple independent signals, respectively. Then, we obtained 99% credible sets of variants by sorting the variants within each signal by descending PPA and then summing up PPA until a cumulative PPA of 99% was reached. As a result, we generated a list of 10 known idiopathic (sQTL) and for splice quantitative trait loci in kidney tissue (sQTL)76 (FDR < 5% for the sQTL). We also prioritized genes that
were known as human kidney disease monogene: for this, we queried each gene at identified loci for a documented kidney phenotype in human as done previously. (Online Mendelian Inheritance in Man® database, OMIM®; or Groopman et al.®) with additional manual curation by expert review to focus on kidney disease monogenes.

GRS analyses. GRS analyses based on the previously identified 634 eGFR variants® and based on the variants identified for DM-/noDM-difference on eGFR were conducted in unrelated European-ancestry individuals from HUNT (n_HUNT = 3023, n_NOHM = 25,484, stage 2 study). For each individual, the GRS was computed as the weighted sum of eGFR-lowering alleles across the 634 variants applying three different weighting schemes: (i) weighted by the respective overall per-variant effect as derived previously (i.e., estimated in CKDGen and UK Biobank, our stage 1 data®) (ii) weighted by the variant’s DM-/noDM-specific effects as derived here in stage 1 depending on whether the individual in HUNT had DM or noDM, respectively (and vice versa), (iii) weighted by the variant’s DM-/noDM-specific effect from stage 1 for HUNT individuals with DM/noDM, respectively, when the variant was among the seven with an identified significant difference, and weighted by the overall-effect size otherwise. By this, the GRS association analyses conducted in a stage 2 study were independent of the variant identification and weight estimation, which were based on stage 1 studies. Separately for individuals with and without DM, we estimated the association of the GRS on eGFR by linear regression (original scale; adjusted by sex, age, and genetic principal components). We judged the GRS effect per standard deviation of the GRS and the eGFR variance explained by the GRS separately in DM and noDM separately.

Reporting summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability Summary genetic association results for the DM-status-specific meta-analyses for logic(eGFRcra) can be downloaded from https://ckdgen.imb.uni-freiburg.de/. All other data are available from the corresponding author on reasonable request.

Code availability The analysis plan can be downloaded from https://ckdgen.eurac.edu/mediawiki/index.php/CKDGen_Round_4_EPACTS_analysis_plan and the phenotype command line script from https://github.com/genepi-freiburg/ckdgen-phenom. All other code is available from the corresponding author on reasonable request.

Received: 16 December 2021; Accepted: 4 May 2022; Published online: 13 June 2022

References
Thomas W. Winkler, Bamidele Tayo, Chris H. L. Thio, Daniele Cusi, Jin-Fang Chai, Karsten B. Sieber, and Bristol-Myers Squibb, and grants and non-financial support from Amgen and AstraZeneca. L.S.P. has served on Scientific Advisory Boards for Janssen, and has or had research support from Merck, Pfizer, Eli Lilly, Novo Nordisk, Sanofi, PhaseBio, Roche, Abbvie, Vascular Pharmaceuticals, Janssen, Glaxo SmithKline, and the Cystic Fibrosis Foundation. He is also a co-founder, Officer and Board member and stockholder for a company, Dassart, Inc., which markets software aimed to help improve diabetes management. A.I.P. and D.F.R. are employees of Merck Sharp Dohme Corp. Bruce.M.P. serves on the steering committee of the Yale Open Data Access Project funded by Johnson & Johnson. P.R. received fees to his institution for research support from AstraZeneca and Novo Nordisk; for steering group participation from AstraZeneca, Gilead, Novo Nordisk, and Bayer; for lectures from Bayer, Eli Lilly and Novo Nordisk; and for Advisory boards from Sanofi and Boehringer Ingelheim outside of this work. V.S. has received a modest honorarium from Sanofi for consulting. He also has ongoing research collaboration with Bayer Ltd. (all outside of the present study). L.W. received institutional grants from GlaxoSmithKline, AstraZeneca, BMS, Boehringer Ingelheim, Pfizer, MSD and Roche/Genentech. H.W. has received grant support paid to the institution and fees for serving on Steering Committees of the ODYSSEY trial from Sanofi and Regeneron Pharmaceuticals, the ISCHEMIA and the MINT studies from the National Institutes of Health, the STRENGTH trial from Omthera Pharmaceuticals, the HEART-FID study from American Regent, the DAL-GENE study from DaCor Pharma UK Inc., the AEGIS-II study from CSL Behring, the SCORED and SOLOIST-WHF from Sanofi Aventis Australia Pty. Ltd., and the CLEAR OUTCOMES study from Esperion Therapeutics. M.P. is partly funded by the study FinerenGen (www.finergen6.de), which is jointly funded by a Finnish Governmental agency Business Finland and thirteen international pharmaceutical companies. The HEART-FID study is supported by Amgen, AstraZeneca, Biogen, Boehringer Ingelheim, Bristol-Myers Squibb, Genentech, a member of the Roche Group, GlaxoSmithKline (GSK), Janssen, Maze Therapeutics, MSD (the tradename of Merck & Co., Inc, Kenilworth, NJ USA), Novartis, Pfizer and Sanofi. C.C.K. is an Editorial Board Member for Communications Biology, but was not involved in the editorial review, nor the decision to publish this article. The remaining authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s42003-022-03448-z.

Correspondence and requests for materials should be addressed to Thomas W. Winkler or Iris M. Heid.

Peer review information Communications Biology thanks Despoina Manousaki and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Primary Handling Editor: George Inglis. Peer reviewer reports are available.

Reprints and permission information is available at http://www.nature.com/reprints

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s) 2022

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/.

© The Author(s) 2022

1 Department of Genetic Epidemiology, University of Regensburg, Regensburg, Germany. 2K. G. Jebsen Center for Genetic Epidemiology, Department of Public Health and Nursing, Faculty of Medicine and Health Sciences, NTNU, Norwegian University of Science and Technology, Trondheim, Norway. 3MRC Integrative Epidemiology Unit, Population Health Sciences, Bristol Medical School, University of Bristol, Bristol, UK. 4Division of Medicine and Laboratory Sciences, University of Oslo, Oslo, Norway. 5Institute for Community Medicine, University Medicine Greifswald, Greifswald, Germany. 6DZHK (German Center for Cardiovascular Research), partner site Greifswald, Greifswald, Germany. 7Department of Population Medicine and Lifestyle Diseases Prevention, Medical University of Bialystok, Bialystok, Poland. 8Department of Nephrology, University Hospital Regensburg, Regensburg, Germany. 9Department of Biostatistics, Vanderbilt University Medical Center, Nashville, TN, USA. 10Department of Veteran’s Affairs, Tennessee Valley Healthcare System (626)/Vanderbilt University, Nashville, TN, USA. 11Department of Clinical and Molecular Medicine, NTNU, Norwegian University of Science and Technology, Trondheim, Norway. 12BioCare—Bioinformatics Core Facility, Norwegian University of Science and Technology, Trondheim, Norway. 13Department of Medicine, University of Mississippi Medical Center, Jackson, MS, USA. 14Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, USA. 15Institute of Genetic Epidemiology, Department of Data Driven Medicine, Faculty of Medicine and Medical Center–University of Freiburg, Freiburg, Germany. 16Genetics, Merck & Co, Inc, Kenilworth, NJ, USA. 17Department of Public Health Sciences, Loyola University Chicago, Maywood, IL, USA. 18Department of Epidemiology, University of Groningen, University Medical Center Groningen, Groningen, The Netherlands. 19Institute of Biomedical Technologies, National Research Council of Italy, Milan, Italy. 20Bio4Dreams—Business Nursery for Life Sciences, Milan, Italy. 21Saw Swee Hock School of Public Health, National University of Singapore and National University Health System, Singapore, Singapore. 22Target Sciences—Genetics, GlaxoSmithKline, Collegeville, PA, USA. 23Institute for Medical Informatics, Statistics and Epidemiology, University of Leipzig, Leipzig, Germany. 24LIFE Research Center for Civilization Diseases, University of Leipzig, Leipzig, Germany. 25Division of Nephrology and Hypertension, Department of Medicine, University of Utah, Salt Lake City, UT, USA. 26Institute for Maternal and Child Health, IRCCS ‘Burlo Garofolo’, Trieste, Italy. 27Renal Division, Department of Medicine IV, Faculty of Medicine and Medical Center—University of Freiburg, Freiburg, Germany. 28Department of Biostatistics, Boston University School of Public Health, Boston, MA, USA. 29Department of Psychiatry and Psychotherapy, University Medicine Greifswald, Greifswald, Germany. 30Institute of Genetics and Bioinformatics ‘Adriano Buzzati-Traverso’—CNR, Naples, Italy. 31EuroC Research Institute, Institute for Biomedicine (affiliated with the University of Lübeck), Bolzano, Italy. 32University of Trento, Department of Cellular, Computational and Integrative Biology—CIBIO, Trento, Italy. 33Statistical Consulting Unit StaBLab, Department of Statistics, LMU Munich, Munich, Germany. 34MRC Centre for Environment and Health, Department of Epidemiology and Biostatistics, School of Public Health, Faculty of Medicine, Imperial College London, London, UK. 35Dementia Research Institute, Imperial College London, London, UK. 36Division of Kidney, Urologic and Hematologic Diseases, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, MD, USA. 37University of Maryland School of Medicine, Baltimore, MD, USA. 38Department of Medicine, Surgery and Health Sciences, University of Trieste, Trieste, Italy. 39Section of Nephrology, Department of Internal Medicine, Leiden University Medical Center, Leiden, The Netherlands. 40Laboratory of Epidemiology and Population Sciences, National Institute on Aging, Intramural Research Program, US National Institutes of Health, Baltimore, MD, USA. 41Faculty of Medicine, School of Health Sciences, University of Iceland, Reykjavik, Iceland. 42CNRS UMR 8199, European Genomic Institute for...
135Department of Pediatrics, Erasmus MC, University Medical Center Rotterdam, Rotterdam, The Netherlands. 136Department of Cardiology, Ealing Hospital, London North West University Healthcare NHS Trust, Middlesex, UK. 137Imperial College Healthcare NHS Trust, Imperial College London, London, UK. 138MRC-PHE Center for Environment and Health, School of Public Health, Imperial College London, London, UK. 139National Heart and Lung Institute, Imperial College London, London, UK. 140Department of Pediatrics, Karolinska Institutet, Stockholm, Sweden. 141Institute of Clinical Chemistry and Molecular Diagnostics, University of Leipzig, Leipzig, Germany. 142Division of Family Medicine and Primary Care, Department of Neurobiology, Care Sciences and Society, Karolinska Institutet, Stockholm, Sweden. 143School of Health and Social Studies, Dalarna University, Stockholm, Sweden. 144University of Eastern Finland, Kuopio, Finland. 145Kuopio University Hospital, Kuopio, Finland. 146Icelandic Heart Association, Kopavogur, Iceland. 147The Center of Public Health Sciences, University of Iceland, Reykjavik, Iceland. 148Montreal University Hospital Research Center, CHUM, Montreal, QC, Canada. 149QCR, CHUM, Montreal, QC, Canada. 150Lee Kong Chian School of Medicine, Nanyang Technological University, Singapore, Singapore. 151QIMR Berghofer Medical Research Institute, Brisbane, QLD, Australia. 152Department of Internal Medicine, Harvard Medical School, Boston, MA, USA. 153VA Cooperative Studies Program, VA Boston Healthcare System, Boston, MA, USA. 154Department of Ophthalmology, Medical Faculty Mannheim, University Heidelberg, Mannheim, Germany. 155Department of Medical Sciences, Uppsala University, Uppsala, Sweden. 156Laboratory of Clinical Genome Sequencing, Graduate School of Frontier Sciences, The University of Tokyo, Tokyo, Japan. 157Institute of Genetic Epidemiology, Helmholtz Zentrum München—German Research Center for Environmental Health, Neuhberg, Germany. 158Chair of Genetic Epidemiology, IBE, Faculty of Medicine, Ludwig-Maximilians-Universität München, Munich, Germany. 159Institute of Medical Biostatistics, Epidemiology and Informatics (IMBEI), University Medical Center, Johannes Gutenberg University Mainz, Mainz, Germany. 160Department of Medical Sciences, Uppsala University, Uppsala, Sweden. 161Department of Neurobiology, Care Sciences and Society, Karolinska Institutet, Stockholm, Sweden. 162Division of Endocrinology, Diabetes and Nutrition, University of Maryland School of Medicine, Baltimore, MD, USA. 163Department of Biostatistics, University of Washington, Seattle, WA, USA. 164Geisinger Research, Biomedical and Translational Informatics Institute, Rockville, MD, USA. 165Department of Nephrology, Geisinger, Danville, PA, USA. 166Department of Cardiology, Heart Center, Tampere University Hospital, Tampere, Finland. 167Department of Nephrology and Hypertension, Friedrich Alexander University Erlangen-Nürnberg (FAU), Erlangen, Germany. 168Department of Anatomy and Cell Biology, University Medicine Greifswald, Greifswald, Germany. 169Laboratory of Clinical Genome Sequencing, Graduate School of Frontier Sciences, The University of Tokyo, Tokyo, Japan. 170Laboratory of Clinical Genome Sequencing, Graduate School of Frontier Sciences, The University of Tokyo, Tokyo, Japan. 171Institute of Genetic Epidemiology, Helmholtz Zentrum München—German Research Center for Environmental Health, Neuhberg, Germany. 172Institute of Genetic Epidemiology, IBE, Faculty of Medicine, Ludwig-Maximilians-Universität München, Munich, Germany. 173Department of Medical Biostatistics, Epidemiology and Informatics (IMBEI), University Medical Center, Johannes Gutenberg University Mainz, Mainz, Germany. 174Department of Medical Sciences, Uppsala University, Uppsala, Sweden. 175Laboratory of Clinical Genome Sequencing, Graduate School of Frontier Sciences, The University of Tokyo, Tokyo, Japan. 176Division of Biomedical Informatics and Personalized Medicine, School of Medicine, University of Colorado Denver-Anschutz Medical Campus, Aurora, CO, USA. 177Department of Epidemiology, Erasmus MC, University Medical Center Rotterdam, Rotterdam, The Netherlands. 178Max Planck Institute of Psychiatry, Munich, Germany. 179SYNLAB MVZ Humangenetik Mannheim, Mannheim, Germany. 180Institute of Genetic and Biomedical Research, National Research Council of Italy, Cagliari, Italy. 181Finnish Institute for Health and Welfare, Helsinki, Finland. 182Research Program for Clinical and Molecular Metabolism, Faculty of Medicine, University of Helsinki, Helsinki, Finland. 183Institute of Medical Biostatistics, Epidemiology and Informatics (IMBEI), University Medical Center, Johannes Gutenberg University Mainz, Mainz, Germany. 184Laboratory of Neurogenetics, National Institute on Aging, National Institutes of Health, Bethesda, MD, USA. 185Data Tecnica International, Glen Echo, MD, USA. 186Department of Public Health and Welfare, Finnish Institute for Health and Welfare, Helsinki, Finland. 187Data Tecnica International, Glen Echo, MD, USA. 188Division of Neurobiology, Department of Internal Medicine, University of Groningen, University Medical Center Groningen, Groningen, The Netherlands. 189Target Sciences—Genetics, GlaxoSmithKline, Albuquerquee, NM, USA. 190Laboratory for Statistical Analysis, RIKEN Center for Integrative Medical Sciences (IMS), Yokohama, Japan. 191Institute of Biomedical Informatics, Harvard Medical School, Boston, MA, USA. 192Department of Ophthalmology, Graduate School of Medical Sciences, Kyushu University, Fukuoka, Japan. 193Department of Ophthalmology, Tohoku University Graduate School of Medicine, Miyagi, Japan. 194Institute of Clinical Chemistry and Laboratory Medicine, University Medicine Greifswald, Greifswald, Germany. 195DZHK (German Center for Cardiovascular Research), Partner Site Munich Heart Alliance, Munich, Germany. 196Cardiovascular Division, Brigham and Women’s Hospital, Boston, MA, USA. 197TIMI Study Group, Boston, MA, USA. 198RIKEN Center for Integrative Medical Sciences (IMS), Yokohama (Kanagawa), Japan. 199Department of Clinical Physiology, Tampere University Hospital, Tampere, Finland. 200Department of Clinical Physiology, Tampere University Hospital, Tampere, Finland. 201Laboratory of Neurogenetics, National Institute on Aging, National Institutes of Health, Bethesda, MD, USA. 202Data Tecnica International, Glen Echo, MD, USA. 203The Department of Public Health and Welfare, Finnish Institute for Health and Welfare, Helsinki, Finland. 204South Texas Diabetes and Obesity Institute and Department of Human Genetics, University of Texas Rio Grande Valley School of Medicine, Brownsville, TX, USA. 205Department of Genetics, School of Medicine, Marshall Medical University, Texas, USA. 206Department of Medical Sciences, Uppsala University, Uppsala, Sweden. 207Division of Nephrology, Department of Medicine, University of Groningen, University Medical Center Groningen, Groningen, The Netherlands. 208Kidney Research Institute, University of Washington, Seattle, WA, USA. 209Institute of Physiology, University of Zurich, Zurich, Switzerland. 210Division of Nephrology, Department of Medicine, University of Washington, Seattle, WA, USA. 211Tuberculosis Research Unit, University of Cape Town, South Africa. 212Laboratory of Clinical Genome Sequencing, Graduate School of Frontier Sciences, The University of Tokyo, Tokyo, Japan. 213MedPharmGene, Montreal, QC, Canada. 214Department of Clinical Science and Education, Karolinska Institutet, Södersjukhuset, Stockholm, Sweden. 215Section of Gerontology and Genomics, Department of Internal Medicine, Leiden University Medical Center, Leiden, The Netherlands. 216Institute of Computational Biology, University of Lausanne, Lausanne, Switzerland. 217Department of Clinical Medicine, Section of Internal Medicine, Umeå University Hospital, Umeå University, Umeå, Sweden. 218Institute of Clinical Chemistry and Laboratory Medicine, University Hospital Regensburg, Regensburg, Germany. 219Vanderbilt Genetics Institute, Vanderbilt University Medical Center, Nashville, TN, USA. 220Section of Gerontology and Genomics, Department of Internal Medicine, Leiden University Medical Center, Leiden, The Netherlands. 221Department of Preventive Medicine, New York University School of Medicine, New York, NY, USA. 222Institute of Preventive Medicine, New York University School of Medicine, New York, NY, USA. 223MedPharmGene, Montreal, QC, Canada. 224Department of Clinical Science and Education, Karolinska Institutet, Södersjukhuset, Stockholm, Sweden. 225Integrated Research and Treatment Center Adiposity Diseases, University of Leipzig, Leipzig, Germany. 226Steno Diabetes Center Copenhagen, Herlev, Denmark. 227Department of Clinical Medicine, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark. 228Department of Genetics, University of Groningen, University Medical Center Groningen, Groningen, The Netherlands. 229Institute of Clinical Chemistry and Laboratory Medicine, University Hospital Regensburg, Regensburg, Germany. 230Department of Neurology, School of Medicine, University of Pennsylvania, Philadelphia, PA, USA. 231Section of Gerontology and Genomics, Department of Internal Medicine, Leiden University Medical Center, Leiden, The Netherlands.
Bioinformatics, Lausanne, Switzerland. 235Department of Exercise and Nutrition Sciences, Milken Institute School of Public Health, The George Washington University, Washington, DC, USA. 236The Mindich Child Health and Development Institute, Icahn School of Medicine at Mount Sinai, New York, NY, USA. 237Novo Nordisk Foundation Center for Basic Metabolic Research, Department of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark. 238Institute of Cardiovascular and Medical Sciences, University of Glasgow, Glasgow, UK. 239Institute of Medical Informatics and Statistics, Kiel University Hospital Schleswig-Holstein, Kiel, Germany. 240Geisinger Research, Biomedical and Translational Informatics Institute, Danville, PA, USA. 241NHBLI’s Framingham Heart Study, Framingham, MA, USA. 242The Center for Population Studies, NHLBI, Framingham, MA, USA. 243Department of Nephrology, St. Olavs Hospital, Trondheim University Hospital, Trondheim, Norway. 244Diabetes Center, Khoo Teck Puat Hospital, Singapore, Singapore. 245The Bioinformatics Center, Department of Biology, Faculty of Science, University of Copenhagen, Copenhagen, Denmark. 246Institute of Human Genetics, Helmholtz Zentrum München, Neuherberg, Germany. 247Institute of Human Genetics, Technische Universität München, Munich, Germany. 248Einthoven Laboratory of Experimental Vascular Research, Leiden University Medical Center, Leiden, The Netherlands. 249Program in Medical and Population Genetics, Broad Institute, Cambridge, MA, USA. 250Interfaculty Institute for Genetics and Functional Genomics, University Medicine Greifswald, Greifswald, Germany. 251Molecular Geriatrics, Department of Public Health and Caring Sciences, Uppsala University, Uppsala, Sweden. 252Department of Genetics, Shanghai—MOST Key Laboratory of Health and Disease Genomics, Chinese National Human Genome Center, Shanghai, China. 253Shanghai Industrial Technology Institute, Shanghai, China. 254Beijing Tongren Eye Center, Beijing Tongren Hospital, Capital Medical University, Beijing, China. 255Synlab Academy, Synlab Holding Deutschland GmbH, Mannheim, Germany. 256Clinical Institute of Medical and Chemical Laboratory Diagnostics, Medical University of Graz, Graz, Austria. 257Deutsches Herzzentrum München, Technische Universität München, Munich, Germany. 258Institute of Epidemiology and Medical Biometry, University of Ulm, Ulm, Germany. 259Institute of Epidemiology and Biobank Popgen, Kiel University, Kiel, Germany. 260School of Public Health and Community Medicine, Hebrew University of Jerusalem, Jerusalem, Israel. 261Laboratory of Complex Trait Genomics, Department of Computational Biology and Medical Sciences, Graduate School of Frontier Sciences, The University of Tokyo, Tokyo, Japan. 262Laboratory for Systems Genetics, RIKEN Center for Integrative Medical Sciences (IMS), Osaka, Japan. 263Department of Statistical Genetics, Osaka University Graduate School of Medicine, Osaka, Japan. 264Massachusetts General Hospital, Boston, MA, USA. 265Department of Genetics and Pharmacology, Institute of Genetic Epidemiology, Medical University of Innsbruck, Innsbruck, Austria. These authors contributed equally: Thomas W. Winkler, Humaira Rasheed, Alexander Teumer, Mathias Gorski, Bryce X. Rowan. 266These authors jointly supervised this work: Adriana M. Hung, Florian Kronenberg, Anna Kötting, Cristian Pattaro, Iris M. Heid. *Lists of authors and their affiliations appear at the end of the paper.

✉email: thomas.winkler@ukr.de; iris.heid@ukr.de

Lifelines cohort study
Chris H. L. Thio19, Peter J. van der Most19 & Martin H. de Borst189

DiscovEHR/MyCode study
Kevin Ho166,167, Navya Shilpa Josyula208 & Sarah A. Pendergrass240

VA Million Veteran Program
Bryce X. Rowan9,10, Cassianne Robinson-Cohen10,80, John M. Gaziano154,155, Lawrence S. Phillips177,178, Ran Tao9,232 & Adriana M. Hung10,80

A full List of members and their affiliations appears in the Supplementary Information.