Polygenic Overlap Between Kidney Function and Large Artery Atherosclerotic Stroke

Elizabeth G. Holliday, PhD; Matthew Traylor, PhD; Rainer Malik, PhD; Stephen Bevan, PhD; Jane Maguire, PhD; Simon A. Koblar, MBBS, PhD; Jonathan Sturm, PhD; Graeme J. Hankey, MD; Christopher Oldmeadow, PhD; Mark McEvoy, PhD; Cathie Sudlow, Dphil; Peter M. Rothwell, FMedSci; Josef Coresh, MD, PhD; Pavel Hamet, MD, PhD; Johanne Tremblay, PhD; Stephen T. Turner, MD; Mariza de Andrade, PhD; Madhumathi Rao, MD, PhD; Reinhold Schmidt, MD; Peter A. Crick, BAppSci; Antonietta Robino, PhD; Carmen A. Peralta, MD; J. Wouter Jukema, MD, PhD; Paul Mitchell, MD, PhD; Sylvia E. Rosas, MD; Jie Jin Wang, PhD; Rodney J. Scott, PhD; Martin Dichgans, MD; Braxton D. Mitchell, PhD; W.H. Linda Kao, PhD†; Caroline S. Fox, MD; Christopher Levi, MBBS; John Attia, MD, PhD; Hugh S. Markus, DM; on behalf of the CKDGen Consortium and the International Stroke Genetics Consortium

Background and Purpose—Epidemiological studies show strong associations between kidney dysfunction and risk of ischemic stroke (IS), the mechanisms of which are incompletely understood. We investigated whether these associations may reflect shared heritability because of a common polygenic basis and whether this differed for IS subtypes.

Methods—Polygenic models were derived using genome-wide association studies meta-analysis results for 3 kidney traits: estimated glomerular filtration rate using serum creatinine (eGFRcrea: n=73998), eGFR using cystatin C (eGFRCys: n=22937), and urinary albumin to creatinine ratio (n=31580). For each, single nucleotide polymorphisms passing 10 P value thresholds were used to form profile scores in 4561 IS cases and 7094 controls from the United Kingdom, Germany, and Australia. Scores were tested for association with IS and its 3 aetiological subtypes: large artery atherosclerosis, cardioembolism, and small vessel disease.

Results—Polygenic scores correlating with higher eGFRcrea were associated with reduced risk of large artery atherosclerosis, with 5 scores reaching $P<0.05$ (peak $P=0.004$) and all showing the epidemiologically expected direction of effect. A similar pattern was observed for polygenic scores reflecting higher urinary albumin to creatinine ratio, of which 3 associated with large artery atherosclerosis (peak $P=0.01$) and all showed the expected directional association. One
Epidemiological evidence supports an association between kidney dysfunction and risk of cardiovascular diseases, including stroke. In fact, the majority of individuals with chronic kidney disease (CKD) die of a cardiovascular cause before developing end-stage renal disease.1 In relation to stroke, kidney dysfunction is associated with multiple outcomes, including incident stroke,2 recurrent stroke,3 and mortality after stroke.4 These relationships seem related more to ischemic stroke (IS) rather than hemorrhagic stroke.5,6

The mechanisms for these phenotypic correlations are not completely understood. Sharing of established cardiovascular risk factors, such as age, sex, blood pressure, cholesterol, smoking, and diabetes mellitus, can explain some but not all excess stroke risk in patients with CKD.4 Sharing of pathophysiological correlates of vascular disease, including carotid atherosclerosis,7 arterial stiffness,8 and cerebral white matter hyperintensity lesions,9 also seem to explain an additional component, but not all, of the excess risk.

Kidney traits and IS both have substantial heritable components, with ≈30% to 50% of variation in glomerular filtration rate (GFR),10 15% to 45% of variation in albuminuria,11 and 30% to 40% of variation in IS risk,12 attributable to genetic effects. Thus, associations between kidney dysfunction and IS may partly reflect a shared genetic component.

In recent years, genome-wide association studies (GWAS) have identified several single nucleotide polymorphisms (SNPs) associated with kidney traits and aetiological subtypes of IS. However, these variants explain only a minority (1%–2%) of population variation in their respective traits. This missing heritability partly reflects a genetic architecture comprising numerous risk variants with effects too small to show significant association in available sample collections. However, numerous SNPs aggregated into polygenic scores may show stronger association and explain more trait variation.13

The polygenic basis of complex traits also hampers attempts to demonstrate pleiotropy for individual SNPs, because effect sizes are typically small for both traits. Using the largest available data sets, a recent study assessed individual SNPs for joint contribution to CKD and cardiovascular disease.14 Significant cross-trait association was observed for SNPs at only 1 locus (SH2B3), suggesting that if genetic overlap between CKD and cardiovascular disease exists, the effect sizes of pleiotropic variants are likely too small to permit their individual detection.

Building on this earlier work, we hypothesised that phenotypic correlations between kidney dysfunction and IS may result from polygenic overlap, that is, sharing of a genetic component consisting of numerous, small-effect SNPs. This hypothesis was tested by deriving polygenic scores using GWAS results for kidney traits and testing their performance in stroke GWAS data sets. These analyses used GWAS meta-analysis results from the CKDGen Consortium and individual-level genotype data from 3 IS case–control collections.

**Key Words:** genetic epidemiology ■ kidney ■ stroke

**Conclusions**—This study suggests possible polygenic correlation between renal dysfunction and IS. The shared genetic components may be specific to stroke subtypes, particularly large artery atherosclerotic stroke. Further study of the genetic relationships between these disorders seems merited. (Stroke. 2014;45:00-00.)

**Methods**

**Data Sources and Study Samples**

For the derivation stage, we used genome-wide association meta-analysis results for 3 established kidney function traits: (1) estimated glomerular filtration rate (eGFR) based on serum creatinine (eGFRcrea: n=73998);15 (2) eGFR based on Cystatin C (eGFRcys: n=229937);16 and (3) urinary albumin to creatinine ratio (UACR: n=31580).17 Higher GFR describes better kidney function, whereas elevated UACR suggests kidney disease. Details of individual studies are provided in Tables I and II in the online-only Data Supplement.

Testing of polygenic scores was performed using individual-level genotype data for IS cases and controls from the Wellcome Trust Case Control Consortium 2 Ischemic Stroke Study17 and the Australian Stroke Genetics Collaborative.18 Three major aetiological subtypes of IS were determined: (1) large artery atherosclerosis (LAA); (2) cardioembolism (CE); and (3) small vessel disease (SVD). Stroke subtyping was performed using the Trial of Org 10172 in Acute Stroke Treatment (TOAST) system.19 All studies were approved by appropriate ethics committees, and participants provided written informed consent.

**Genotyping, Imputation, and Quality Control**

For IS studies, genotyping was performed using Illumina arrays followed by quality control and imputation to the HapMap Phase II CEU reference. Principal components analysis was performed using Eigenstrat,20 based on 95016 approximately independent, directly genotyped SNPs. Principal component covariates of ancestry were calculated after 3 iterations of principal components analysis with outlier removal.

**SNP Enrichment Assessment, Polygenic Scoring, and Association Analyses**

Using GWAS meta-analysis results for the 3 kidney traits, sets of SNPs passing 10 graded P values (Pthreshold=0.0001, 0.001, 0.01, 0.05, 0.1, 0.2, 0.3, 0.4, 0.5, and 1) were extracted. SNP sets were pruned by removing correlated SNPs (r2>0.2) within 1 Mb, preferentially retaining the most significantly associated SNP.21 Pruned SNP sets were used to form polygenic scores for IS cases and controls as the sum of reference alleles for each SNP weighted by the summary regression coefficient for the relevant kidney trait. Polygenic scores were tested for association with IS and its subtypes using mixed effects logistic regression adjusted for 3 ancestry principal components, incorporating study site as a random effect. The proportion of stroke case–control variation explained by polygenic scores (R2) was estimated as previously described for mixed effects logistic models.22 Polygenic score tests were 1-sided because a priori, we sought to identify effects consistent with established epidemiological evidence. For eGFRcrea and eGFRcys, the prespecified effect direction was negative because higher eGFR correlates with reduced stroke risk.2 For UACR, a positive effect was prespecified.23 The study-wise significance threshold was derived using the method proposed by Galwey24 (online-only...
Data Supplement). A flow-chart describing the polygenic analysis is shown in Figure I in the online-only Data Supplement.

At a significance threshold of 0.05, we had 98% power to detect polygenic scores explaining ≥0.2% of variance in case/control status for any validation subtype, 81% to 82% power to detect scores explaining ≥0.1% of variance (varying by subtype), and 51% to 53% power for scores explaining ≥0.05% of variance. If at a significance threshold of 0.001, the corresponding power estimates were 76% to 78%, 32% to 34%, and 10% to 11%, respectively.

Lookup for Individual SNPs Associated With Kidney or Stroke Traits

As a secondary analysis, we conducted targeted, cross-trait analyses for individual SNPs previously associated with kidney traits or IS subtypes. These analyses are described in the online-only Data Supplement.

Results

SNP Enrichment

SNP set enrichment across P value thresholds for the 3 kidney traits is shown in Figure. Enrichment was strongest for eGFR based on serum creatinine (eGFRcrea). Modest SNP enrichment was also observed for eGFR based on Cystatin C (eGFRcys). Results for UACR showed less marked evidence for enrichment.

Polygenic Scoring Results

A total of 4561 IS cases and 7094 controls were used for polygenic score association testing (Table 1). Based on observed correlation among traits (online-only Data Supplement), the adjusted study-wise significance threshold was α=0.001.

Two scores derived from eGFRcrea showed nominal association with broad IS, reaching P=0.02 and P=0.05 (for Pthreshold=0.001 and 0.01, respectively, see Table 2). Five eGFRcrea-based scores showed nominal association with LAA, with peak association (P=0.004) observed for Pthreshold=0.05. This score explained an estimated 0.26% of LAA case–control variation (Table III in the online-only Data Supplement). For both IS and LAA, scores for all 10 P value thresholds demonstrated the expected negative direction of effect, with genetic scores indicative of higher eGFRcrea correlating with reduced (negative) stroke risk. No eGFRcrea-based score showed association with either CE or SVD, and effect directions were also inconsistent within these subtypes.

Polygenic scores derived from eGFRcys showed no association with broad IS or any of its subtypes (all P>0.05; see Table IV in the online-only Data Supplement). Further, effect directions were inconsistent across tests within each stroke type.

For UACR, one score (Pthreshold=0.01) showed nominal association with broad IS, (P=0.04; Table 3). Three scores were also nominally associated with LAA (peak P=0.01 at Pthreshold=0.001, R2=0.11%) and 1 score showed nominal association with SVD (P=0.03 at Pthreshold=0.0001, R2=0.14%; see Table III in the online-only Data Supplement). For the 3 traits (IS, LAA, and SVD), all effects were in the expected direction, with scores indicative of higher UACR (microalbuminuria) correlating with positive stroke risk.

None of the polygenic associations passed the study-wise significance threshold of 0.001. However, within each of the 5 trait combinations showing nominal association (eGFRcrea-IS, eGFRcrea-LAA, UACR-IS, UACR-LAA, and UACR-SVD), the direction of effect across all 10 tests was in accordance with prior evidence. We conducted simulations to empirically estimate the probability of this pattern occurring by chance for any given kidney–stroke trait combination (online-only Data Supplement). For eGFRcrea, the probability of a set of 10 tests showing a negative effect was P=0.080, based on 10000 simulations. For eGFRcys and UACR, the corresponding 1-sided probabilities were 0.067 and 0.069, respectively. Thus, the observed pattern was unlikely to have arisen by chance in any given set of 10 tests.

If we consider all kidney–stroke combinations where such consistency was observed, the results seem even less likely to have occurred by chance. Excluding results for broad IS, we conducted 3 sets of approximately independent stroke subtype tests for each of the 3 kidney traits. Among these 9 sets of tests, we observed 3 in which consistent effects in the expected direction were uniformly observed (eGFRcrea-LAA, UACR-LAA, and UACR-SVD). The probability of this occurring by chance is =0.023 (online-only Data Supplement), or ≥1 in every 43 studies such as ours. This suggests our

Table 1. Ischemic Stroke Samples Used for Polygenic Score Testing

<table>
<thead>
<tr>
<th>Study</th>
<th>IS</th>
<th>LAA</th>
<th>CE</th>
<th>SVD</th>
<th>Controls</th>
</tr>
</thead>
<tbody>
<tr>
<td>ASGC</td>
<td>1071</td>
<td>375</td>
<td>226</td>
<td>287</td>
<td>1212</td>
</tr>
<tr>
<td>WTCCC2-Munich</td>
<td>1140</td>
<td>338</td>
<td>322</td>
<td>104</td>
<td>775*</td>
</tr>
<tr>
<td>WTCCC2-UK</td>
<td>2350</td>
<td>494</td>
<td>450</td>
<td>471</td>
<td>5107</td>
</tr>
<tr>
<td>Total</td>
<td>4561</td>
<td>1207</td>
<td>998</td>
<td>862</td>
<td>7094</td>
</tr>
</tbody>
</table>

ASGC indicates Australian Stroke Genetics Collaborative; CE, cardioembolism; LAA, large artery atherosclerosis; SVD, small vessel disease; UK, United Kingdom; and WTCCC2, Wellcome Trust Case Control Consortium 2.

*WTCCC2-Munich controls were selected from the German KORA study, which also contributed to kidney trait meta-analyses.
nominal associations likely to represent true polygenic correlations of small effect.

**Lookup for Individually Associated SNPs**

Of 40 SNPs previously associated with kidney traits, 1 (rs653178) was significantly associated (1-sided \( P = 2 \times 10^{-4} \)) with LAA25 after multiple testing adjustment. Of 7 stroke-associated SNPs, none were significantly associated with kidney traits (Tables V and VI in the online-only Data Supplement).

**Discussion**

This study suggests that reported epidemiological associations between renal disease and stroke may be partly explained by shared genetic factors and that this association may differ for IS subtypes, in this study being most marked for large artery stroke. However, the causal genetic variants are likely of small individual effect, detectable in the current study only when aggregated into highly polygenic scores and, even then, achieving only nominal significance. Independent replication will be necessary to confirm the validity of these results.

An important factor affecting the significance of our results was power for individual stroke subtypes. For nominally significant associations, the kidney trait profile scores explained from 0.1 to 0.26% of case–control variation for different stroke subtypes. Power analyses indicated larger samples would be necessary to identify the observed effects at more stringent significance levels.

Although the proportion of stroke subtype variance explained by kidney-based scores was low, this does not mean the true genetic overlap is small. Profile scoring combines errors in effect estimates across all SNPs in the score, which usually produces estimates of explained variance markedly lower than true values.26 We observed the strongest polygenic correlations between eGFR defined based on serum creatinine and LAA. Using eGFRcrea as the derivation trait, nominal significance of polygenic scores with LAA was sustained across nearly the full

### Table 2. Association of eGFRcrea-Based Polygenic Scores With Ischemic Stroke

<table>
<thead>
<tr>
<th>( P_{\text{threshold}} )</th>
<th>( N_{\text{SNPs}} )</th>
<th>Ischemic Stroke</th>
<th>LAA</th>
<th>CE</th>
<th>SVD</th>
</tr>
</thead>
<tbody>
<tr>
<td>( P_{\text{value}} )</td>
<td>( P_{\text{value}} )</td>
<td>Direction</td>
<td>( P_{\text{value}} )</td>
<td>Direction</td>
<td>( P_{\text{value}} )</td>
</tr>
<tr>
<td>0.0001</td>
<td>280</td>
<td>0.13</td>
<td>–</td>
<td>0.29</td>
<td>–</td>
</tr>
<tr>
<td>0.001</td>
<td>1193</td>
<td>0.02</td>
<td>–</td>
<td>0.03</td>
<td>–</td>
</tr>
<tr>
<td>0.01</td>
<td>6769</td>
<td>0.05</td>
<td>–</td>
<td>0.07</td>
<td>–</td>
</tr>
<tr>
<td>0.05</td>
<td>24621</td>
<td>0.07</td>
<td>–</td>
<td>0.004</td>
<td>–</td>
</tr>
<tr>
<td>0.1</td>
<td>42961</td>
<td>0.18</td>
<td>–</td>
<td>0.04</td>
<td>–</td>
</tr>
<tr>
<td>0.2</td>
<td>74252</td>
<td>0.16</td>
<td>–</td>
<td>0.06</td>
<td>–</td>
</tr>
<tr>
<td>0.3</td>
<td>101719</td>
<td>0.21</td>
<td>–</td>
<td>0.07</td>
<td>–</td>
</tr>
<tr>
<td>0.4</td>
<td>125887</td>
<td>0.16</td>
<td>–</td>
<td>0.06</td>
<td>–</td>
</tr>
<tr>
<td>0.5</td>
<td>147541</td>
<td>0.14</td>
<td>–</td>
<td>0.04</td>
<td>–</td>
</tr>
<tr>
<td>1</td>
<td>220045</td>
<td>0.10</td>
<td>–</td>
<td>0.04</td>
<td>–</td>
</tr>
</tbody>
</table>

CE indicates cardioembolism; LAA, large artery atherosclerosis; SNP, single nucleotide polymorphism; and SVD, small vessel disease.

*Threshold for selecting SNPs into the score.
†Number of SNPs forming the score.
‡Indicates whether a score predicting higher values of the kidney trait predicts increased (+) or decreased (–) stroke risk.

### Table 3. Association of UACR-Based Polygenic Scores With Ischemic Stroke

<table>
<thead>
<tr>
<th>( P_{\text{threshold}} )</th>
<th>( N_{\text{SNPs}} )</th>
<th>Ischemic Stroke</th>
<th>LAA</th>
<th>CE</th>
<th>SVD</th>
</tr>
</thead>
<tbody>
<tr>
<td>( P_{\text{value}} )</td>
<td>( P_{\text{value}} )</td>
<td>Direction</td>
<td>( P_{\text{value}} )</td>
<td>Direction</td>
<td>( P_{\text{value}} )</td>
</tr>
<tr>
<td>0.0001</td>
<td>89</td>
<td>0.29</td>
<td>+</td>
<td>0.08</td>
<td>+</td>
</tr>
<tr>
<td>0.001</td>
<td>722</td>
<td>0.27</td>
<td>+</td>
<td>0.01</td>
<td>+</td>
</tr>
<tr>
<td>0.01</td>
<td>5541</td>
<td>0.04</td>
<td>+</td>
<td>0.03</td>
<td>+</td>
</tr>
<tr>
<td>0.05</td>
<td>22476</td>
<td>0.15</td>
<td>+</td>
<td>0.04</td>
<td>+</td>
</tr>
<tr>
<td>0.1</td>
<td>40818</td>
<td>0.38</td>
<td>+</td>
<td>0.11</td>
<td>+</td>
</tr>
<tr>
<td>0.2</td>
<td>72518</td>
<td>0.35</td>
<td>+</td>
<td>0.07</td>
<td>+</td>
</tr>
<tr>
<td>0.3</td>
<td>99798</td>
<td>0.28</td>
<td>+</td>
<td>0.09</td>
<td>+</td>
</tr>
<tr>
<td>0.4</td>
<td>124037</td>
<td>0.26</td>
<td>+</td>
<td>0.09</td>
<td>+</td>
</tr>
<tr>
<td>0.5</td>
<td>145753</td>
<td>0.27</td>
<td>+</td>
<td>0.07</td>
<td>+</td>
</tr>
<tr>
<td>1</td>
<td>218171</td>
<td>0.33</td>
<td>+</td>
<td>0.08</td>
<td>+</td>
</tr>
</tbody>
</table>

CE indicates cardioembolism; LAA, large artery atherosclerosis; SNP, single nucleotide polymorphism; and SVD, small vessel disease.
range of $P$ value derivation thresholds. If these results reflect a true genetic correlation, the pattern of results is consistent with a complex genetic model involving numerous small-effect variants influencing diverse biological processes.

The primary pathophysiological mechanism for LAA is atherosclerosis of the large cerebral arteries, a surrogate marker of which is carotid intima media thickness. Various epidemiological studies have reported inverse associations between eGFR and carotid intima media thickness. The majority show that this association can be explained by traditional cardiovascular risk factors, including age, smoking, hypertension, obesity, diabetes mellitus, and dyslipidaemia. Thus, polygenic correlations between eGFR and LAA — if confirmed — may reflect genetic variants influencing atherosclerosis or its heritable risk factors.

We observed no cointrait association for eGFR scores derived from cystatin C at $P < 0.05$. The lack of similar results between eGFRcrea and eGFRcreys may reflect the smaller sample size for the latter; the larger discovery set for eGFRcrea will increase polygenic score precision. Greater discriminatory power of eGFRcrea-based scores was also supported by stronger SNP enrichment across $P$ value thresholds (Figure). Previous GWAS meta-analyses have also identified considerably more SNPs associated with eGFRcrea than eGFRcreys. Given that both are measures of GFR and have similar heritability, the different results may largely reflect differences in sample size and power to identify variants of modest effect.

Polygenic scores derived from microalbuminuria (UACR) showed nominal associations with LAA across various $P$ value thresholds. This is consistent with epidemiological associations between UACR, carotid intima media thickness, and cardiovascular disease, although the pathophysiological basis of these relationships is less clear. In contrast to GFR, microalbuminuria seems not to reflect generalized atherosclerosis, but may represent another common pathophysiological process such as endothelial dysfunction or low-grade inflammation.

We observed no evidence for polygenic overlap between renal function and cardioembolic stroke (CE), in spite of epidemiological associations between CKD and atrial fibrillation, the major CE risk factor. However, increased atrial fibrillation prevalence has mainly been shown in patients with advanced kidney disease. Furthermore, factors associated with atrial fibrillation in the general population seem not to be associated with atrial fibrillation in CKD, suggesting pathophysiological differences.

Our results suggested possible polygenic overlap between microalbuminuria and SVD. These results are consistent with epidemiological associations between renal function and cerebral SVD, which have been interpreted as suggesting a systemic generalized microvascular disease that underlies both pathologies. Given the epidemiological evidence, more significant genetic associations might have been expected. Our modest results may reflect 2 factors. First, accuracy of diagnosis of small vessel stroke is greatly improved by the routine use of MRI, which was only used for $\approx 50\%$ of stroke patients in the current study. Second, pathological and imaging data suggest that SVD is phenotypically heterogeneous, incorporating 2 distinct subtypes. One is characterized by single larger lacunar infarcts and thought to primarily relate to atherosclerosis; the other is characterized by multiple small lacunar infarcts and leukoaraiosis and is related to a diffuse small vessel arteriopathy. The latter subtype has been particularly associated with microalbuminuria. If renal disease is genetically correlated with only 1 SVD subtype, the heterogeneity of broadly defined SVD will have reduced our ability to detect any genetic overlap. The study of SVD samples with finer-scale phenotyping will likely provide better insights into genetic pleiotropy for SVD.

We observed largely negative evidence for cross-trait association for individual SNPs strongly associated with either kidney function or stroke. Only 1 SNP previously associated with eGFR was associated with both large artery and small vessel stroke; this SNP has been recently associated with broad IS and was also the only variant showing cross-trait association in the previous CKDGen analysis.

An important limitation of this study was modest sample sizes for stroke subtypes. Sample size is a challenge for genetic studies of stroke, reflecting the technical nature of case ascertainment and the presence of multiple aetiological types. These are among the largest current IS samples with GWAS, but our results should be validated in larger, well-phenotyped stroke GWAS data sets as they become available.

Summary
This study suggests a potential polygenic basis for epidemiological associations between renal dysfunction and IS. The effects of the putative shared genetic components seem small and potentially specific to distinct stroke types.

Sources of Funding
E.G. Holliday is supported by the Australian Heart Foundation and National Stroke Foundation (100071). M. Traylor is funded by a UK Stroke Association project grant. H.S. Markus is supported by an Institute for Health Research (UK) Senior Investigator award. H.S. Markus and S. Bevan’s work in this area is supported by the Cambridge University Hospital Trust NIHR Biomedical Research Centre. P.M. Rothwell holds NIHR and Wellcome Trust Senior Investigator Awards and is funded by the NIHR Biomedical Research Centre, Oxford. C. Sudlow, H.S. Markus and S. Bevan have received research funding from the UK Binks Trust. J. Coresh, M. de Andrade, C.S. Fox, W.H.L. Kao, B.D. Mitchell, and S.T. Turner have received research funding from the US National Institutes of Health, P. Hamet and J. Tremblay have received research funding from Genome Quebec, Canada. J. Tremblay has received funding from the Canadian Institutes for Health Research. C. Levi and R.J. Scott have received research funding from the Australian National Health and Medical Research Council.

Disclosures
None.

References


Polygenic Overlap Between Kidney Function and Large Artery Atherosclerotic Stroke
on behalf of the CKDGen Consortium and the International Stroke Genetics Consortium

Stroke. published online October 28, 2014;
Stroke is published by the American Heart Association, 7272 Greenville Avenue, Dallas, TX 75231
Copyright © 2014 American Heart Association, Inc. All rights reserved.
Print ISSN: 0039-2499. Online ISSN: 1524-4628

The online version of this article, along with updated information and services, is located on the World Wide Web at:
http://stroke.ahajournals.org/content/early/2014/10/28/STROKEAHA.114.006609

Data Supplement (unedited) at:
http://stroke.ahajournals.org/content/suppl/2014/10/28/STROKEAHA.114.006609.DC1

Permissions: Requests for permissions to reproduce figures, tables, or portions of articles originally published in Stroke can be obtained via RightsLink, a service of the Copyright Clearance Center, not the Editorial Office. Once the online version of the published article for which permission is being requested is located, click Request Permissions in the middle column of the Web page under Services. Further information about this process is available in the Permissions and Rights Question and Answer document.

Reprints: Information about reprints can be found online at:
http://www.lww.com/reprints

Subscriptions: Information about subscribing to Stroke is online at:
http://stroke.ahajournals.org//subscriptions/
SUPPLEMENTAL MATERIAL

Polygenic overlap between kidney function and large artery atherosclerotic stroke

Elizabeth G. Holliday PhD, Matthew Traylor PhD, Rainer Malik PhD, Stephen Bevan PhD, Jane Maguire PhD, Simon A. Koblar MBBS PhD, Jonathan Sturm PhD, Graeme J. Hankey MD, Christopher Oldmeadow PhD, Mark McEvoy PhD, Cathie Sudlow Dphil, Peter M. Rothwell FMedSci, Josef Coresh MD PhD, Pavel Hamet MD PhD, Johanne Tremblay PhD, Stephen T. Turner MD, Mariza de Andrade PhD, Madhumathi Rao MD PhD, Reinhold Schmidt MD, Peter A. Crick BAppSci, Antonietta Robino PhD, Carmen A. Peralta MD, J. Wouter Jukema MD PhD, Paul Mitchell MD PhD, Sylvia E. Rosas MD, Jie Jin Wang PhD, Rodney J. Scott PhD, Martin Dichgans MD, Braxton D. Mitchell PhD, W. H. Linda Kao PhD, Caroline S. Fox MD, Christopher Levi MBBS, John Attia MD PhD, Hugh S Markus DM, on behalf of the CKDGen Consortium and the International Stroke Genetics Consortium

Supplemental Methods

Assessing SNP enrichment at nominal P-value thresholds in kidney trait GWAS results

We assessed the extent of SNP enrichment at each threshold, as an initial assessment of polygenicity for each kidney trait. Under the null hypothesis of no SNPs being associated, the distribution of genome-wide P-values is uniform on [0,1]. Alternatively, under polygenic inheritance, the distribution of P-values will be more weighted towards 0 and the proportion of SNPs passing nominally significant P-value thresholds will exceed expected values. For each trait, the expected and observed numbers of SNPs passing each threshold were recorded with the quotient (observed/expected) reflecting enrichment as a fold-change compared to the null.

Multiple testing adjustment

For each of the three kidney traits, 10 overlapping SNP sets were used for polygenic scoring. Across the three kidney traits, the effective number of independent tests ($M_{eff\_kidney}$) represented among the 30 correlated scores was estimated using the method proposed by Galwey \(^1\), based on eigenvalues of the 30x30 correlation matrix of polygenic scores calculated for the IS sample.

Within the four stroke traits, the effective number of independent tests ($M_{eff\_stroke}$) was estimated using the Galwey method based on eigenvalues of the 4x4 correlation matrix of SNP effect (beta coefficient) estimates from a previous GWAS meta-analysis of ischaemic stroke and its subtypes \(^2\).

The total number of independent tests was estimated as $M_{eff} = M_{eff\_stroke} \times M_{eff\_kidney}$. Polygenic score P-values were compared with a study-wise significance threshold derived from $M_{eff}$ using the Sidak adjustment: $\alpha = 1 - (1 - 0.05)^{1/M_{eff}}$ \(^3\).

The number of effectively independent polygenic tests among the 30 kidney-based polygenic scores was 12.53 and among the 4 stroke traits was 3.67, yielding 46 independent tests and a study-wise significance threshold of $\alpha=0.001$. 
Identifying individual SNPs associated with kidney or stroke traits

As a secondary analysis, we conducted targeted, cross-trait analyses for individual SNPs previously associated with kidney traits or ischaemic stroke subtypes. Associated SNPs were identified in the NIH “Catalog of Published Genome-Wide Association Studies” using the disease/trait definitions “Glomerular filtration rate, Renal function and chronic kidney disease, Renal function-related traits, Urinary albumin excretion (eGFRcrea), Stroke and Stroke (ischemic)”, and a significance threshold of $5 \times 10^{-8}$. Original publications were also searched to corroborate results. Kidney trait-associated SNPs were assessed for association with ischaemic stroke subtypes using results from the Metastroke Study. Stroke-associated SNPs were assessed for kidney trait association using meta-analysis results for eGFRcrea, eGFRcys and UACR. The number of effectively independent tests ($M_{\text{eff}}$) among all SNPs was estimated using the Nyholt method based on observed patterns of linkage disequilibrium in WTCCC2-UK controls. The adjusted significance threshold was derived from $M_{\text{eff}}$ using the Sidak adjustment.

We identified 40 SNPs showing genome-wide association ($P < 5 \times 10^{-8}$) with eGFRcrea, eGFRcys or UACR in previous GWAS (Table I) and seven SNPs previously associated with ischaemic stroke subtypes (Table II). The number of effectively independent SNPs among these 47 total SNPs was 40, yielding an adjusted significance threshold of 0.0013. One kidney-associated SNP (rs653178 near the $SH2B3$ gene on chromosome 12) showed significant evidence for association (one-sided $P = 2 \times 10^{-4}$) with large artery atherosclerotic stroke (LAA). The same SNP also demonstrated nominal evidence for association with SVD ($P = 0.003$). In both cases, the allele (C) indicative of decreased GFR correlated with increased stroke risk, consistent with epidemiological evidence. This SNP – located near the immune-related $SH2B3$ gene on chromosome 12 – is a pleiotropic variant also associated with blood pressure and retinal vessel calibre in the previous CKDGen cross-trait analysis. It also reached genome-wide significance ($P < 5 \times 10^{-8}$) in a recent large GWAS of broadly defined ischaemic stroke. Of the seven stroke-associated SNPs, none were significantly associated with kidney traits after multiple testing adjustment. Two SNPs (rs2200733 and rs6843082 near $PITX2$ on chromosome 4) associated with cardioembolism were nominally associated with eGFRcrea (one-sided $P \sim 3 \times 10^{-3}$), but this result may be due to chance.

Description of simulation study to estimate the probability of observing ten polygenic score effects in the expected direction by chance for each kidney trait

For each kidney trait we performed 10,000 simulations. Each simulation produced a set of ten sequential effect directions from logistic regression tests of permuted case-control status against randomly generated, overlapping polygenic scores.

For each kidney trait, the set of ~220,000 approximately independent SNPs passing $P_{\text{threshold}}=1$ was used as the base set. From these, we randomly selected $M$ SNPs, where $M$ was the observed number of SNPs reaching $P_{\text{threshold}}<0.5$. This process was continued for decreasing $P$-value thresholds ($P_{\text{threshold}}=0.4, 0.3, 0.2$ etc) to generate ten sets of SNPs of the same size and with the same number of overlapping SNPs as observed in our actual study. Using these ten SNP sets, polygenic scores were calculated for 500 randomly selected WTCCC2-UK control samples, to which case or control status was randomly assigned in a 1:6 ratio to reflect the mean, observed case:control ratio for ischaemic stroke subtypes. Logistic regression of case-control status on the polygenic score was performed for each of the ten scores, and the observed sign of the estimated log-odds ratio (+/-) recorded across the ten tests.
Of the 10,000 simulations performed for eGFRcrea, 804 \((P=0.080)\) produced a set of ten negative effects (one-sided test to accord with prior epidemiological evidence). For eGFRcys, 674 simulations \((P=0.067)\) produced ten negative effects. For UACR, for which increasing values correlate with \textit{increased} CVD risk, 688 simulations \((P=0.069)\) produced a set of ten \textit{positive} effects.

\textit{Estimating the probability of observing consistent effect directions for three of nine hypothesis sets}

Excluding results for broad IS, which is non-negligibly correlated with stroke subtypes, we conducted three sets of approximately independent stroke subtype tests for each of the three kidney traits. Among these nine sets of tests, we observed three in which consistent effects in the expected direction were uniformly observed (eGFRcrea-LAA, UACR-LAA, and UACR-SVD). For a binomial distribution with sample size nine and probability of ‘success’ 0.072 (mean of empirical probabilities above) the probability of three or more ‘successes’ is 0.023, or about one in every 45 studies such as this (see below). This simplifying model assumes accuracy of the ‘success’ probability and independence among stroke subtypes and kidney traits; thus the result is approximate. Nevertheless, it suggests the consistent effects underlying the nominal associations are unlikely to have arisen simply by chance.

If a random variable \(X\) follows a binomial distribution with parameters \(n\) and \(p\), the probability of observing exactly \(k\) successes is:

\[
\Pr(X = k) = \binom{n}{k} (p)^k (1 - p)^{n-k}
\]

Thus, for parameter values \(n=9\) and \(p=0.072\), the probability of exactly three successes \((k=3)\) is:

\[
\Pr(X = 3) = \binom{9}{3} (0.072)^{3} (1 - 0.072)^{9-3} = 0.020
\]

Summing probabilities for discrete values of \(X\), the probability of observing three or more successes \((k \geq 3)\) is:

\[
\Pr(X \geq 3) = \sum_{k=3}^{9} \binom{9}{k} (0.072)^{k} (1 - 0.072)^{9-k} = 0.023
\]
# Supplemental Tables

## Table I. Characteristics of studies included in discovery GWAS meta-analyses of eGFRcrea and eGFRcys

<table>
<thead>
<tr>
<th>Study</th>
<th>Sample size eGFRcrea</th>
<th>Sample size eGFRcys</th>
<th>Women % (n)</th>
<th>Age mean(SD)</th>
<th>eGFRcrea [ml/min/1.73 m²] mean(SD)</th>
<th>eGFRcys [ml/min/1.73 m²] mean(SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>AGES</td>
<td>3219</td>
<td>NA</td>
<td>58.0(1867)</td>
<td>76.4(5.46)</td>
<td>73.0(20.0)</td>
<td>NA</td>
</tr>
<tr>
<td>Amish Studies</td>
<td>1211</td>
<td>783</td>
<td>48.9(592)</td>
<td>49.5(16.9)</td>
<td>93.7(19.7)</td>
<td>114.9(18.0)</td>
</tr>
<tr>
<td>ARIC</td>
<td>8982</td>
<td>7145</td>
<td>53.1(4767)</td>
<td>61.8(6.1)</td>
<td>81.4(17.5)</td>
<td>84.2(19.7)</td>
</tr>
<tr>
<td>ASPS</td>
<td>848</td>
<td>NA</td>
<td>56.8(482)</td>
<td>65.2(8.0)</td>
<td>96.5(39.9)</td>
<td>NA</td>
</tr>
<tr>
<td>Baltimore Longitudinal Study of Aging (BLSA)</td>
<td>723</td>
<td>NA</td>
<td>46.1(333)</td>
<td>70.4(15.2)</td>
<td>80.3(23.1)</td>
<td>NA</td>
</tr>
<tr>
<td>Cardiovascular Health Study (CHS)</td>
<td>2820</td>
<td>2475</td>
<td>61.3(1729)</td>
<td>71.9(5.0)</td>
<td>77.3(20.8)</td>
<td>81.0(17.9)</td>
</tr>
<tr>
<td>ERF</td>
<td>2079</td>
<td>NA</td>
<td>56.3(1171)</td>
<td>49.2(14.0)</td>
<td>93.5(21.4)</td>
<td>NA</td>
</tr>
<tr>
<td>Family Heart Study (FamHS)</td>
<td>883</td>
<td>NA</td>
<td>51.1(451)</td>
<td>55.5(11.1)</td>
<td>88.5(19.4)</td>
<td>NA</td>
</tr>
<tr>
<td>Framingham Heart Study (FHS)</td>
<td>7782</td>
<td>2992</td>
<td>54.3(4229)</td>
<td>51.2(14.0)</td>
<td>91.7(21.7)</td>
<td>83.8(17.8)</td>
</tr>
<tr>
<td>GENOA</td>
<td>1163</td>
<td>NA</td>
<td>56.3(655)</td>
<td>59(10.2)</td>
<td>87.7(24)</td>
<td>NA</td>
</tr>
<tr>
<td>Health ABC</td>
<td>1663</td>
<td>1663</td>
<td>47.1(784)</td>
<td>73.8(2.8)</td>
<td>71.2(14.8)</td>
<td>77.0(19.9)</td>
</tr>
<tr>
<td>Health Professionals Follow-Up Study (HPFS)</td>
<td>818</td>
<td>NA</td>
<td>-</td>
<td>64.7(8.3)</td>
<td>85.2(22.7)</td>
<td>NA</td>
</tr>
<tr>
<td>KORA F3</td>
<td>1641</td>
<td>1642</td>
<td>50.5(831)</td>
<td>62.5(10.1)</td>
<td>83.9(21)</td>
<td>111.8(26.3)</td>
</tr>
<tr>
<td>KORA F4</td>
<td>1814</td>
<td>1811</td>
<td>51.3(930)</td>
<td>60.9(8.9)</td>
<td>85.1(20.2)</td>
<td>109.7(26.2)</td>
</tr>
<tr>
<td>Korcula</td>
<td>888</td>
<td>NA</td>
<td>64.0(568)</td>
<td>56.3(13.9)</td>
<td>87.3(20.6)</td>
<td>NA</td>
</tr>
<tr>
<td>Microisolates in South Tyrol (MICROS)</td>
<td>1201</td>
<td>1198</td>
<td>56.5(678)</td>
<td>46.2(16.1)</td>
<td>94.6(20.9)</td>
<td>107.4(23.8)</td>
</tr>
<tr>
<td>Northern Sweden Population Health Survey (NSPHS)</td>
<td>565</td>
<td>NA</td>
<td>53.1(300)</td>
<td>51.7(18.3)</td>
<td>91.0(22.1)</td>
<td>NA</td>
</tr>
<tr>
<td>Nurses’ Health Study (NHS)</td>
<td>786</td>
<td>NA</td>
<td>100(786)</td>
<td>59.5(6.5)</td>
<td>86.2(22.1)</td>
<td>NA</td>
</tr>
<tr>
<td>Orkney Complex Disease Study (ORCADES)</td>
<td>704</td>
<td>NA</td>
<td>53.6(377)</td>
<td>54.2(15.2)</td>
<td>89.4(20.7)</td>
<td>NA</td>
</tr>
</tbody>
</table>
Table II. Characteristics of studies included in the discovery GWAS meta-analysis of UACR

<table>
<thead>
<tr>
<th>Study</th>
<th>Sample size</th>
<th>Women %</th>
<th>Age mean</th>
<th>UACR median(IQR)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Amish Studies</td>
<td>744</td>
<td>45.0</td>
<td>42.8</td>
<td>3.8 (2.2, 6.1)</td>
</tr>
<tr>
<td>ARIC</td>
<td>6525</td>
<td>53.1</td>
<td>63.1</td>
<td>3.9 (2.0, 7.7)</td>
</tr>
<tr>
<td>Baltimore Longitudinal Study of Aging (BLSA)</td>
<td>354</td>
<td>46.9</td>
<td>69.8</td>
<td>7 (4.4, 11.0)</td>
</tr>
<tr>
<td>Cardiovascular Health Study (CHS)</td>
<td>1865</td>
<td>60.8</td>
<td>71.2</td>
<td>9.3 (5.3, 19.9)</td>
</tr>
<tr>
<td>Colaus</td>
<td>5311</td>
<td>53.2</td>
<td>53.4</td>
<td>5.1 (3.4, 9.1)</td>
</tr>
<tr>
<td>EPIC</td>
<td>2371</td>
<td>53.3</td>
<td>59.2</td>
<td>3.6 (1.5, 8.3)</td>
</tr>
<tr>
<td>Fenland</td>
<td>1398</td>
<td>56.2</td>
<td>44.9</td>
<td>4.5 (3.2, 7.1)</td>
</tr>
<tr>
<td>Framingham Heart Study</td>
<td>6523</td>
<td>53.2</td>
<td>47.6</td>
<td>4.6 (2.6, 9.9)</td>
</tr>
<tr>
<td>KORA F3</td>
<td>1530</td>
<td>50.5</td>
<td>62.5</td>
<td>4.9 (2.1, 11.1)</td>
</tr>
<tr>
<td>KORA F4</td>
<td>1803</td>
<td>51.3</td>
<td>60.9</td>
<td>10.7 (7.0, 18.4)</td>
</tr>
<tr>
<td>Microisolates in South Tyrol (MICROS)</td>
<td>503</td>
<td>56.3</td>
<td>46.0</td>
<td>NA</td>
</tr>
<tr>
<td>SHIP</td>
<td>2653</td>
<td>48.9</td>
<td>55.3</td>
<td>8.9 (5.0, 20.6)</td>
</tr>
<tr>
<td>Total</td>
<td>31,580</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note: Further study-specific information is provided in the original publication by Boger et al. 9
Table III. Estimates of case/control variation ($R^2$) explained by profile scores for each discovery kidney trait and validation stroke subtype combination.

<table>
<thead>
<tr>
<th>$P_{\text{threshold}}$</th>
<th>LAA</th>
<th>CE</th>
<th>SVD</th>
<th>LAA</th>
<th>CE</th>
<th>SVD</th>
<th>LAA</th>
<th>CE</th>
<th>SVD</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.0001</td>
<td>0.01%</td>
<td>0.01%</td>
<td>0.03%</td>
<td>&lt;0.01%</td>
<td>0.01%</td>
<td>0.02%</td>
<td>0.04%</td>
<td>0.03%</td>
<td>0.14%</td>
</tr>
<tr>
<td>0.001</td>
<td>0.12%</td>
<td>0.01%</td>
<td>0.13%</td>
<td>&lt;0.01%</td>
<td>0.01%</td>
<td>0.03%</td>
<td>0.11%</td>
<td>0.05%</td>
<td>&lt;0.01%</td>
</tr>
<tr>
<td>0.01</td>
<td>0.13%</td>
<td>0.01%</td>
<td>0.05%</td>
<td>&lt;0.01%</td>
<td>0.02%</td>
<td>&lt;0.01%</td>
<td>0.07%</td>
<td>0.02%</td>
<td>0.03%</td>
</tr>
<tr>
<td>0.05</td>
<td>0.26%</td>
<td>0.01%</td>
<td>&lt;0.01%</td>
<td>0.02%</td>
<td>0.01%</td>
<td>0.02%</td>
<td>0.06%</td>
<td>&lt;0.01%</td>
<td>0.02%</td>
</tr>
<tr>
<td>0.1</td>
<td>0.12%</td>
<td>&lt;0.01%</td>
<td>&lt;0.01%</td>
<td>0.13%</td>
<td>0.01%</td>
<td>&lt;0.01%</td>
<td>0.03%</td>
<td>0.01%</td>
<td>0.03%</td>
</tr>
<tr>
<td>0.2</td>
<td>0.09%</td>
<td>0.01%</td>
<td>0.01%</td>
<td>0.05%</td>
<td>&lt;0.01%</td>
<td>&lt;0.01%</td>
<td>0.03%</td>
<td>&lt;0.01%</td>
<td>0.02%</td>
</tr>
<tr>
<td>0.3</td>
<td>0.08%</td>
<td>0.01%</td>
<td>&lt;0.01%</td>
<td>0.03%</td>
<td>&lt;0.01%</td>
<td>0.01%</td>
<td>0.02%</td>
<td>0.01%</td>
<td>0.01%</td>
</tr>
<tr>
<td>0.4</td>
<td>0.08%</td>
<td>0.02%</td>
<td>0.01%</td>
<td>0.03%</td>
<td>&lt;0.01%</td>
<td>0.01%</td>
<td>0.02%</td>
<td>&lt;0.01%</td>
<td>0.01%</td>
</tr>
<tr>
<td>0.5</td>
<td>0.09%</td>
<td>0.01%</td>
<td>0.01%</td>
<td>0.03%</td>
<td>&lt;0.01%</td>
<td>0.01%</td>
<td>0.02%</td>
<td>&lt;0.01%</td>
<td>0.02%</td>
</tr>
<tr>
<td>1</td>
<td>0.10%</td>
<td>0.02%</td>
<td>0.02%</td>
<td>0.03%</td>
<td>&lt;0.01%</td>
<td>&lt;0.01%</td>
<td>0.02%</td>
<td>&lt;0.01%</td>
<td>0.01%</td>
</tr>
</tbody>
</table>

Notes: The proportion of stroke case-control variation explained by each polygenic score ($R^2$) was estimated as previously described for fixed explanatory variables in mixed effects logistic models. The percent variance attributable to the score was calculated by estimating the $R^2$ associated with all fixed effects (polygenic score and three principal components), and subtracting the $R^2$ for principal components calculated in a model that included only principal components as fixed effects.
Table IV. Association of eGFRcys-based polygenic scores with ischaemic stroke

<table>
<thead>
<tr>
<th>$P_{\text{threshold}}$</th>
<th>$N_{\text{SNPs}}$</th>
<th>$P$</th>
<th>Direction$^\ddagger$</th>
<th>$P$</th>
<th>Direction</th>
<th>$P$</th>
<th>Direction</th>
<th>$P$</th>
<th>Direction</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.0001</td>
<td>144</td>
<td>0.70</td>
<td>+</td>
<td>0.60</td>
<td>+</td>
<td>0.71</td>
<td>+</td>
<td>0.77</td>
<td>+</td>
</tr>
<tr>
<td>0.001</td>
<td>803</td>
<td>0.35</td>
<td>-</td>
<td>0.65</td>
<td>+</td>
<td>0.35</td>
<td>-</td>
<td>0.22</td>
<td>-</td>
</tr>
<tr>
<td>0.01</td>
<td>5,968</td>
<td>0.39</td>
<td>-</td>
<td>0.41</td>
<td>-</td>
<td>0.22</td>
<td>-</td>
<td>0.63</td>
<td>+</td>
</tr>
<tr>
<td>0.05</td>
<td>22,807</td>
<td>0.38</td>
<td>-</td>
<td>0.27</td>
<td>-</td>
<td>0.28</td>
<td>-</td>
<td>0.84</td>
<td>+</td>
</tr>
<tr>
<td>0.1</td>
<td>40,551</td>
<td>0.24</td>
<td>-</td>
<td>0.07</td>
<td>-</td>
<td>0.30</td>
<td>-</td>
<td>0.74</td>
<td>+</td>
</tr>
<tr>
<td>0.2</td>
<td>71,653</td>
<td>0.35</td>
<td>-</td>
<td>0.14</td>
<td>-</td>
<td>0.38</td>
<td>-</td>
<td>0.70</td>
<td>+</td>
</tr>
<tr>
<td>0.3</td>
<td>98,626</td>
<td>0.38</td>
<td>-</td>
<td>0.16</td>
<td>-</td>
<td>0.42</td>
<td>-</td>
<td>0.72</td>
<td>+</td>
</tr>
<tr>
<td>0.4</td>
<td>122,702</td>
<td>0.41</td>
<td>-</td>
<td>0.16</td>
<td>-</td>
<td>0.45</td>
<td>-</td>
<td>0.73</td>
<td>+</td>
</tr>
<tr>
<td>0.5</td>
<td>144,116</td>
<td>0.38</td>
<td>-</td>
<td>0.14</td>
<td>-</td>
<td>0.50</td>
<td>+</td>
<td>0.73</td>
<td>+</td>
</tr>
<tr>
<td>1</td>
<td>214,626</td>
<td>0.39</td>
<td>-</td>
<td>0.16</td>
<td>-</td>
<td>0.46</td>
<td>+</td>
<td>0.66</td>
<td>+</td>
</tr>
</tbody>
</table>

*LAA, large artery atherosclerosis; CE, cardioembolism; SVD, small vessel disease.*

$^\dagger$Threshold for selecting SNPs into the score.

$^\ddagger$Number of SNPs forming the score.

$^\ddagger$Indicates whether a score predicting higher values of the kidney trait predicts increased (+) or decreased (-) stroke risk.
Table V. Association with stroke subtypes for SNPs previously showing genome-wide association with kidney traits.

<table>
<thead>
<tr>
<th>Kidney trait ref</th>
<th>SNP</th>
<th>Chr</th>
<th>Position</th>
<th>A1</th>
<th>P (LAA)</th>
<th>P (CE)</th>
<th>P (SVD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>eGFRreca</td>
<td>rs12124078</td>
<td>1</td>
<td>15,869,898</td>
<td>G</td>
<td>0.55</td>
<td>0.40</td>
<td>0.29</td>
</tr>
<tr>
<td>eGFRreca</td>
<td>rs267734</td>
<td>1</td>
<td>150,951,476</td>
<td>NR</td>
<td>0.44</td>
<td>0.06</td>
<td>0.10</td>
</tr>
<tr>
<td>eGFRreca</td>
<td>rs6431731</td>
<td>2</td>
<td>15,863,001</td>
<td>T</td>
<td>0.14</td>
<td>0.87</td>
<td>0.74</td>
</tr>
<tr>
<td>eGFRreca</td>
<td>rs1260326</td>
<td>2</td>
<td>27,730,939</td>
<td>NR</td>
<td>0.29</td>
<td>0.44</td>
<td>0.46</td>
</tr>
<tr>
<td>eGFRreca</td>
<td>rs13538</td>
<td>2</td>
<td>73,868,327</td>
<td>NR</td>
<td>0.70</td>
<td>0.24</td>
<td>0.70</td>
</tr>
<tr>
<td>eGFR</td>
<td>rs10206899</td>
<td>2</td>
<td>73,900,899</td>
<td>C</td>
<td>0.96</td>
<td>0.24</td>
<td>0.48</td>
</tr>
<tr>
<td>eGFRreca</td>
<td>rs347685</td>
<td>3</td>
<td>141,807,136</td>
<td>NR</td>
<td>0.44</td>
<td>0.98</td>
<td>0.98</td>
</tr>
<tr>
<td>eGFRreca</td>
<td>rs17319721</td>
<td>4</td>
<td>77,368,846</td>
<td>A</td>
<td>0.11</td>
<td>0.98</td>
<td>0.54</td>
</tr>
<tr>
<td>eGFR</td>
<td>rs10032549</td>
<td>4</td>
<td>77,398,014</td>
<td>A</td>
<td>0.10</td>
<td>0.43</td>
<td>0.94</td>
</tr>
<tr>
<td>eGFR</td>
<td>rs4859682</td>
<td>4</td>
<td>77,410,317</td>
<td>A</td>
<td>0.09</td>
<td>0.74</td>
<td>0.58</td>
</tr>
<tr>
<td>eGFRreca</td>
<td>rs13146355</td>
<td>4</td>
<td>77,412,139</td>
<td>A</td>
<td>0.16</td>
<td>0.86</td>
<td>0.53</td>
</tr>
<tr>
<td>eGFR</td>
<td>rs11959928</td>
<td>5</td>
<td>39,397,131</td>
<td>NR</td>
<td>0.60</td>
<td>0.45</td>
<td>0.67</td>
</tr>
<tr>
<td>eGFRreca</td>
<td>rs6420094</td>
<td>5</td>
<td>176,817,635</td>
<td>NR</td>
<td>0.74</td>
<td>0.97</td>
<td>0.60</td>
</tr>
<tr>
<td>eGFRreca</td>
<td>rs3828890</td>
<td>6</td>
<td>31,440,668</td>
<td>G</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>eGFRreca</td>
<td>rs881858</td>
<td>6</td>
<td>43,806,608</td>
<td>NR</td>
<td>0.93</td>
<td>0.56</td>
<td>0.54</td>
</tr>
<tr>
<td>eGFRreca</td>
<td>rs10277115</td>
<td>7</td>
<td>1,285,194</td>
<td>T</td>
<td>0.95</td>
<td>0.64</td>
<td>0.06</td>
</tr>
<tr>
<td>CKD</td>
<td>rs7805747</td>
<td>7</td>
<td>151,407,800</td>
<td>NR</td>
<td>0.56</td>
<td>0.42</td>
<td>0.74</td>
</tr>
<tr>
<td>eGFRreca</td>
<td>rs10109414</td>
<td>8</td>
<td>23,751,150</td>
<td>NR</td>
<td>0.85</td>
<td>0.73</td>
<td>0.92</td>
</tr>
<tr>
<td>eGFRreys</td>
<td>rs1731274</td>
<td>8</td>
<td>23,766,318</td>
<td>G</td>
<td>0.81</td>
<td>0.76</td>
<td>0.71</td>
</tr>
<tr>
<td>eGFRreca</td>
<td>rs4744712</td>
<td>9</td>
<td>71,434,706</td>
<td>NR</td>
<td>0.68</td>
<td>0.48</td>
<td>0.78</td>
</tr>
<tr>
<td>UACR</td>
<td>rs1801239</td>
<td>10</td>
<td>16,919,051</td>
<td>C</td>
<td>0.64</td>
<td>0.30</td>
<td>0.91</td>
</tr>
<tr>
<td>eGFRreca</td>
<td>rs963837</td>
<td>11</td>
<td>30,749,089</td>
<td>T</td>
<td>0.15</td>
<td>0.35</td>
<td>0.64</td>
</tr>
<tr>
<td>eGFRreca</td>
<td>rs3925584</td>
<td>11</td>
<td>30,760,334</td>
<td>T</td>
<td>0.07</td>
<td>0.60</td>
<td>0.83</td>
</tr>
<tr>
<td>eGFRreys</td>
<td>rs653178</td>
<td>12</td>
<td>112,007,755</td>
<td>C</td>
<td>4.81E-04</td>
<td>0.17</td>
<td>6.74E-03</td>
</tr>
<tr>
<td>eGFRreca</td>
<td>rs626277</td>
<td>13</td>
<td>72,347,695</td>
<td>NR</td>
<td>0.99</td>
<td>0.48</td>
<td>0.56</td>
</tr>
<tr>
<td>eGFRreca</td>
<td>rs2928148</td>
<td>15</td>
<td>41,401,549</td>
<td>G</td>
<td>0.35</td>
<td>0.25</td>
<td>0.42</td>
</tr>
<tr>
<td>eGFRreca</td>
<td>rs2453533</td>
<td>15</td>
<td>45,641,224</td>
<td>NR</td>
<td>0.86</td>
<td>0.88</td>
<td>0.61</td>
</tr>
<tr>
<td>eGFRreca</td>
<td>rs2467853</td>
<td>15</td>
<td>45,698,792</td>
<td>G</td>
<td>0.91</td>
<td>0.86</td>
<td>0.63</td>
</tr>
<tr>
<td>eGFR</td>
<td>rs17536527</td>
<td>15</td>
<td>45,719,186</td>
<td>C</td>
<td>0.71</td>
<td>0.97</td>
<td>0.36</td>
</tr>
<tr>
<td>eGFRreca</td>
<td>rs17730436</td>
<td>15</td>
<td>53,942,927</td>
<td>T</td>
<td>0.33</td>
<td>0.92</td>
<td>0.31</td>
</tr>
<tr>
<td>eGFRreca</td>
<td>rs1394125</td>
<td>15</td>
<td>76,158,982</td>
<td>NR</td>
<td>0.45</td>
<td>0.60</td>
<td>0.82</td>
</tr>
<tr>
<td>CKD</td>
<td>rs4293393</td>
<td>16</td>
<td>20,364,587</td>
<td>T</td>
<td>0.41</td>
<td>0.56</td>
<td>0.28</td>
</tr>
<tr>
<td>eGFRreca</td>
<td>rs12917707</td>
<td>16</td>
<td>20,367,689</td>
<td>G</td>
<td>0.26</td>
<td>0.33</td>
<td>0.28</td>
</tr>
<tr>
<td>eGFRreca</td>
<td>rs11864909</td>
<td>16</td>
<td>20,400,838</td>
<td>C</td>
<td>0.91</td>
<td>0.42</td>
<td>0.34</td>
</tr>
<tr>
<td>eGFRreca</td>
<td>rs2453580</td>
<td>17</td>
<td>19,438,320</td>
<td>C</td>
<td>0.35</td>
<td>0.54</td>
<td>0.78</td>
</tr>
<tr>
<td>eGFRreca</td>
<td>rs11078903</td>
<td>17</td>
<td>37,631,923</td>
<td>A</td>
<td>0.55</td>
<td>0.27</td>
<td>0.92</td>
</tr>
<tr>
<td>eGFRreca</td>
<td>rs9895661</td>
<td>17</td>
<td>59,456,588</td>
<td>C</td>
<td>0.90</td>
<td>0.27</td>
<td>0.24</td>
</tr>
<tr>
<td>eGFRreca</td>
<td>rs12460876</td>
<td>19</td>
<td>33,356,890</td>
<td>NR</td>
<td>0.29</td>
<td>0.76</td>
<td>0.30</td>
</tr>
<tr>
<td>eGFRreys</td>
<td>rs13038305</td>
<td>20</td>
<td>23,612,236</td>
<td>C</td>
<td>0.60</td>
<td>0.51</td>
<td>0.76</td>
</tr>
<tr>
<td>eGFRreca/eGFRreca</td>
<td>rs911119</td>
<td>20</td>
<td>23,612,736</td>
<td>NR</td>
<td>0.57</td>
<td>0.48</td>
<td>0.71</td>
</tr>
</tbody>
</table>

A1, effect allele predicting increased values of the kidney trait, NR=not reported; LAA, large artery atherosclerosis; CE, cardioembolism; SVD, small vessel disease. All P-values are two-sided since one-sided P-values could not be derived for SNPs without a reported effect allele. * Corresponding one-sided P-values were 2x10^{-4} and 3x10^{-3}, since the effect allele had the epidemiologically predicted direction of effect.
**Table VI.** Association with kidney traits for SNPs previously showing genome-wide association with stroke subtypes.

<table>
<thead>
<tr>
<th>SNP ref</th>
<th>Chr</th>
<th>Position</th>
<th>Stroke subtype</th>
<th>Risk Allele</th>
<th>P eGFRcrea</th>
<th>P eGFRcys</th>
<th>P UACR</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs2200733</td>
<td>16</td>
<td>4</td>
<td>111,710,168</td>
<td>CE</td>
<td>T</td>
<td>8.2E-03</td>
<td>0.20</td>
</tr>
<tr>
<td>rs6843082</td>
<td>4</td>
<td>111,718,066</td>
<td>CE</td>
<td>G</td>
<td>6.9E-03*</td>
<td>0.19</td>
<td>0.11</td>
</tr>
<tr>
<td>rs556621</td>
<td>17</td>
<td>4</td>
<td>44,594,158</td>
<td>LAA</td>
<td>A</td>
<td>0.46</td>
<td>0.039</td>
</tr>
<tr>
<td>rs11984041</td>
<td>7</td>
<td>19,031,934</td>
<td>LAA</td>
<td>A</td>
<td>0.43</td>
<td>0.22</td>
<td>0.81</td>
</tr>
<tr>
<td>rs2107595</td>
<td>7</td>
<td>19,049,387</td>
<td>LAA</td>
<td>A</td>
<td>0.94</td>
<td>0.22</td>
<td>0.38</td>
</tr>
<tr>
<td>rs12425791</td>
<td>12</td>
<td>7</td>
<td>783,483</td>
<td>IS</td>
<td>A</td>
<td>0.68</td>
<td>0.67</td>
</tr>
<tr>
<td>rs879324</td>
<td>16</td>
<td>16</td>
<td>73,068,677</td>
<td>CE</td>
<td>A</td>
<td>0.32</td>
<td>0.98</td>
</tr>
</tbody>
</table>

*Corresponding one-sided P-values are 4x10^-3 and 3x10^-3, since the effect alleles had the epidemiologically predicted direction of effect.

**Supplemental Figure**

**Figure I.** Flow-chart representing the polygenic analysis approach.
Supplemental References

1. Galwey NW. A new measure of the effective number of tests, a practical tool for comparing families of non-independent significance tests. *Genet Epidemiol.* 2009;33:559-568


