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: $\alpha = 1 - (1 - 0.05)^{1/Meff}$ [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 α=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 $5x10^{-8}$. 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 [2]. 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_{eff}) among all SNPs was estimated using the Nyholt method [4] based on observed patterns of linkage disequilibrium in WTCCC2-UK controls. The adjusted significance threshold was derived from M_{eff} using the Sidak adjustment [3].

We identified 40 SNPs showing genome-wide association ($P<5x10^{-8}$) 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 [4], 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 $P=2x10^{-4}$) 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 [5], and retinal vessel calibre in the previous CKDGen cross-trait analysis [6]. It also reached genome-wide significance ($P<5x10^{-8}$) in a recent large GWAS of broadly defined ischaemic stroke [7].

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 $P\sim3x10^{-3}$), 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 $P_{threshold}=1$ was used as the base set. From these, we randomly selected M SNPs, where M was the observed number of SNPs reaching $P_{threshold}<0.5$. This process was continued for decreasing P-value thresholds ($P_{threshold}=0.4$, 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 *increased* CVD risk, 688 simulations (P=0.069) produced a set of ten *positive* effects.

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

Study	Sample size eGFRcrea	Sample size eGFRcys	Women %(n)	Age mean(SD)	eGFRcrea [ml/min/ 1.73 m^2] mean(SD)	eGFRcys [ml/min/ 1.73 m^2] mean(SD)
AGES	3219	NA	58.0(1867)	76.4(5.46)	73.0(20.0)	NA
Amish Studies	1211	783	48.9(592)	49.5(16.9)	93.7(19.7)	114.9(18.0)
ARIC	8982	7145	53.1(4767)	61.8(6.1)	81.4(17.5)	84.2(19.7)
ASPS	848	NA	56.8(482)	65.2(8.0)	96.5(39.9)	NA
Baltimore Longitudinal Study of Aging (BLSA)	723	NA	46.1(333)	70.4(15.2)	80.3(23.1)	NA
Cardiovascular Health Study (CHS)	2820	2475	61.3(1729)	71.9(5.0)	77.3(20.8)	81.0(17.9)
ERF	2079	NA	56.3(1171)	49.2(14.0)	93.5(21.4)	NA
Family Heart Study (FamHS)	883	NA	51.1(451)	55.5(11.1)	88.5(19.4)	NA
Framingham Heart Study (FHS)	7782	2992	54.3(4229)	51.2(14.0)	91.7(21.7)	83.8(17.8)
GENOA	1163	NA	56.3(655)	59(10.2)	87.7(24)	NA
Health ABC	1663	1663	47.1(784)	73.8(2.8)	71.2(14.8)	77.0(19.9)
Health Professionals Follow-Up Study (HPFS)	818	NA	-	64.7(8.3)	85.2(22.7)	NA
KORA F3	1641	1642	50.5(831)	62.5(10.1)	83.9(21)	111.8(26.3)
KORA F4	1814	1811	51.3(930)	60.9(8.9)	85.1(20.2)	109.7(26.2)
Korcula	888	NA	64.0(568)	56.3(13.9)	87.3(20.6)	NA
Microisolates in South Tyrol (MICROS)	1201	1198	56.5(678)	46.2(16.1)	94.6(20.9)	107.4(23.8)
Northern Sweden Population Health Survey (NSPHS)	565	NA	53.1(300)	51.7(18.3)	91.0(22.1)	NA
Nurses' Health Study (NHS)	786	NA	100(786)	59.5(6.5)	86.2(22.1)	NA
Orkney Complex Disease Study (ORCADES)	704	NA	53.6(377)	54.2(15.2)	89.4(20.7)	NA

Popgen	1163	NA	44.4(516)	54.8(13.9)	88.1(18.8)	NA
Rotterdam Study – I	4390	NA	61.4(2696)	70.0(9.0)	77.1(17.2)	NA
Rotterdam Study – II	1863	NA	54.5(1015)	64.8(8.0)	81.3(17.2)	NA
SHIP	3228	3228	51.7(1670)	54.5(15.3)	90.4(23.6)	97.1(25.3)
Sorbs	856	NA	58.5(501)	48.8(15.7)	92.2(19.0)	NA
Vis	768	NA	58.6(450)	56.9(15.2)	88.2(22.1)	NA
Women's Genome Health Study (WGHS)	21940	NA	100(21940)	55(7.1)	90(22.5)	NA
Total	73,998	22,937				

Note: Further study-specific information is provided in the original publication by Pattaro et al [8]

Table II. Characteristics of studies included in the discovery GWAS meta-analysis of UACR

Study	Sample size UACR	Women %	Age mean	UACR median(IQR)
Amish Studies	744	45	42.8	3.8 (2.2, 6.1)
ARIC	6525	53.1	63.1	3.9 (2.0, 7.7)
Baltimore Longitudinal Study of Aging (BLSA)	354	46.9	69.8	7 (4.4, 11.0)
Cardiovascular Health Study (CHS)	1865	60.8	71.2	9.3 (5.3, 19.9)
Colaus	5311	53.2	53.4	5.1 (3.4, 9.1)
EPIC	2371	53.3	59.2	3.6 (1.5, 8.3)
Fenland	1398	56.2	44.9	4.5 (3.2, 7.1)
Framingham Heart Study	6523	53.2	47.6	4.6 (2.6, 9.9)
KORA F3	1530	50.5	62.5	4.9 (2.1, 11.1)
KORA F4	1803	51.3	60.9	10.7 (7.0, 18.4)
Microisolates in South Tyrol (MICROS)	503	56.3	46	NA
SHIP	2653	48.9	55.3	8.9 (5.0, 20.6)
Total	31,580			

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.

	eGFRcrea			eGFRcys			UACR		
$P_{threshold}$ [†]	LAA	CE	SVD	LAA	CE	SVD	LAA	CE	SVD
0.0001	0.01%	0.01%	0.03%	<0.01%	0.01%	0.02%	0.04%	0.03%	0.14%
0.001	0.12%	0.01%	0.13%	<0.01%	0.01%	0.03%	0.11%	0.05%	<0.01%
0.01	0.13%	0.01%	0.05%	<0.01%	0.02%	<0.01%	0.07%	0.02%	0.03%
0.05	0.26%	0.01%	<0.01%	0.02%	0.01%	0.02%	0.06%	<0.01%	0.02%
0.1	0.12%	<0.01%	<0.01%	0.13%	0.01%	<0.01%	0.03%	0.01%	0.03%
0.2	0.09%	0.01%	0.01%	0.05%	<0.01%	<0.01%	0.03%	<0.01%	0.02%
0.3	0.08%	0.01%	<0.01%	0.03%	<0.01%	0.01%	0.02%	0.01%	0.01%
0.4	0.08%	0.02%	0.01%	0.03%	<0.01%	0.01%	0.02%	<0.01%	0.01%
0.5	0.09%	0.01%	0.01%	0.03%	<0.01%	0.01%	0.02%	<0.01%	0.02%
1	0.10%	0.02%	0.02%	0.03%	<0.01%	<0.01%	0.02%	<0.01%	0.01%

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 [10]. 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

		Ischaemic stroke		LAA*		CE*		SVD*	
$P_{threshold}$ [†]	N_{SNPs} [‡]	*P*	Direction [§]	*P*	Direction	*P*	Direction	*P*	Direction
0.0001	144	0.70	+	0.60	+	0.71	+	0.77	+
0.001	803	0.35	-	0.65	+	0.35	-	0.22	-
0.01	5,968	0.39	-	0.41	-	0.22	-	0.63	+
0.05	22,807	0.38	-	0.27	-	0.28	-	0.84	+
0.1	40,551	0.24	-	0.07	-	0.30	-	0.74	+
0.2	71,653	0.35	-	0.14	-	0.38	-	0.70	+
0.3	98,626	0.38	-	0.16	-	0.42	-	0.72	+
0.4	122,702	0.41	-	0.16	-	0.45	-	0.73	+
0.5	144,116	0.38	-	0.14	-	0.50	+	0.73	+
1	214,626	0.39	-	0.16	-	0.46	+	0.66	+

[*]LAA, large artery atherosclerosis; CE, cardioembolism; 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 V. Association with stroke subtypes for SNPs previously showing genome-wide association with kidney traits.

Kidney trait [ref]	SNP	Chr	Position	A1	*P* (LAA)	*P* (CE)	*P* (SVD)
eGFRcrea [8]	rs12124078	1	15,869,898	G	0.55	0.40	0.29
eGFRcrea [11]	rs267734	1	150,951,476	NR	0.44	0.06	0.10
eGFRcrea [8]	rs6431731	2	15,863,001	T	0.14	0.87	0.74
eGFRcrea [11]	rs1260326	2	27,730,939	NR	0.29	0.44	0.46
eGFRcrea [11]	rs13538	2	73,868,327	NR	0.70	0.24	0.70
eGFR [12]	rs10206899	2	73,900,899	C	0.96	0.24	0.48
eGFRcrea [11]	rs347685	3	141,807,136	NR	0.44	0.98	0.98
eGFRcrea [13]	rs17319721	4	77,368,846	A	0.11	0.98	0.54
eGFR [12]	rs10032549	4	77,398,014	A	0.10	0.43	0.94
eGFR [12]	rs4859682	4	77,410,317	A	0.09	0.74	0.58
eGFRcrea [14]	rs13146355	4	77,412,139	A	0.16	0.86	0.53
eGFRcrea [11]	rs11959928	5	39,397,131	NR	0.60	0.45	0.67
eGFRcrea [11]	rs6420094	5	176,817,635	NR	0.74	0.97	0.60
eGFRcrea [14]	rs3828890	6	31,440,668	G	NA	NA	NA
eGFRcrea [11]	rs881858	6	43,806,608	NR	0.93	0.56	0.54
eGFRcrea [14]	rs10277115	7	1,285,194	T	0.95	0.64	0.06
CKD [11]	rs7805747	7	151,407,800	NR	0.56	0.42	0.74
eGFRcrea [11]	rs10109414	8	23,751,150	NR	0.85	0.73	0.92
eGFRcys [13]	rs1731274	8	23,766,318	G	0.81	0.76	0.71
eGFRcrea [11]	rs4744712	9	71,434,706	NR	0.68	0.48	0.78
UACR [9]	rs1801239	10	16,919,051	C	0.64	0.30	0.91
eGFRcrea [14]	rs963837	11	30,749,089	T	0.15	0.35	0.64
eGFRcrea [8]	rs3925584	11	30,760,334	T	0.07	0.60	0.83
eGFRcys [11]	rs653178	12	112,007,755	C	**4.81E-04**[*]	0.17	**6.74E-03**[*]
eGFRcrea [11]	rs626277	13	72,347,695	NR	0.99	0.48	0.56
eGFRcrea [8]	rs2928148	15	41,401,549	G	0.35	0.25	0.42
eGFRcrea [11]	rs2453533	15	45,641,224	NR	0.86	0.88	0.61
eGFRcrea [13]	rs2467853	15	45,698,792	G	0.91	0.86	0.63
eGFR [12]	rs17536527	15	45,719,186	C	0.71	0.97	0.36
eGFRcrea [14]	rs17730436	15	53,942,927	T	0.33	0.92	0.31
eGFRcrea [11]	rs1394125	15	76,158,982	NR	0.45	0.60	0.82
CKD [15]	rs4293393	16	20,364,587	T	0.41	0.56	0.28
eGFRcrea [13]	rs12917707	16	20,367,689	G	0.26	0.33	0.28
eGFRcrea [14]	rs11864909	16	20,400,838	C	0.91	0.42	0.34
eGFRcrea [8]	rs2453580	17	19,438,320	C	0.35	0.54	0.78
eGFRcrea [8]	rs11078903	17	37,631,923	A	0.55	0.27	0.92
eGFRcrea [14]	rs9895661	17	59,456,588	C	0.90	0.27	0.24
eGFRcrea [11]	rs12460876	19	33,356,890	NR	0.29	0.76	0.30
eGFRcys [13]	rs13038305	20	23,610,261	C	0.60	0.51	0.76
eGFRcrea/eGFRcys [11]	rs911119	20	23,612,736	NR	0.57	0.48	0.71

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.

SNP [ref]	Chr	Position	Stroke subtype	Risk Allele	*P* eGFRcrea	*P* eGFRcys	*P* UACR
rs2200733 [16]	4	111,710,168	CE	T	8.2E-03*	0.20	0.29
rs6843082 [2]	4	111,718,066	CE	G	6.9E-03*	0.19	0.11
rs556621 [17]	6	44,594,158	LAA	A	0.46	0.039	1.00
rs11984041 [18]	7	19,031,934	LAA	A	0.43	0.22	0.81
rs2107595 [2]	7	19,049,387	LAA	A	0.94	0.22	0.38
rs12425791 [19]	12	783,483	IS	A	0.68	0.67	0.33
rs879324 [2]	16	73,068,677	CE	A	0.32	0.98	0.17

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

Supplemental Figure

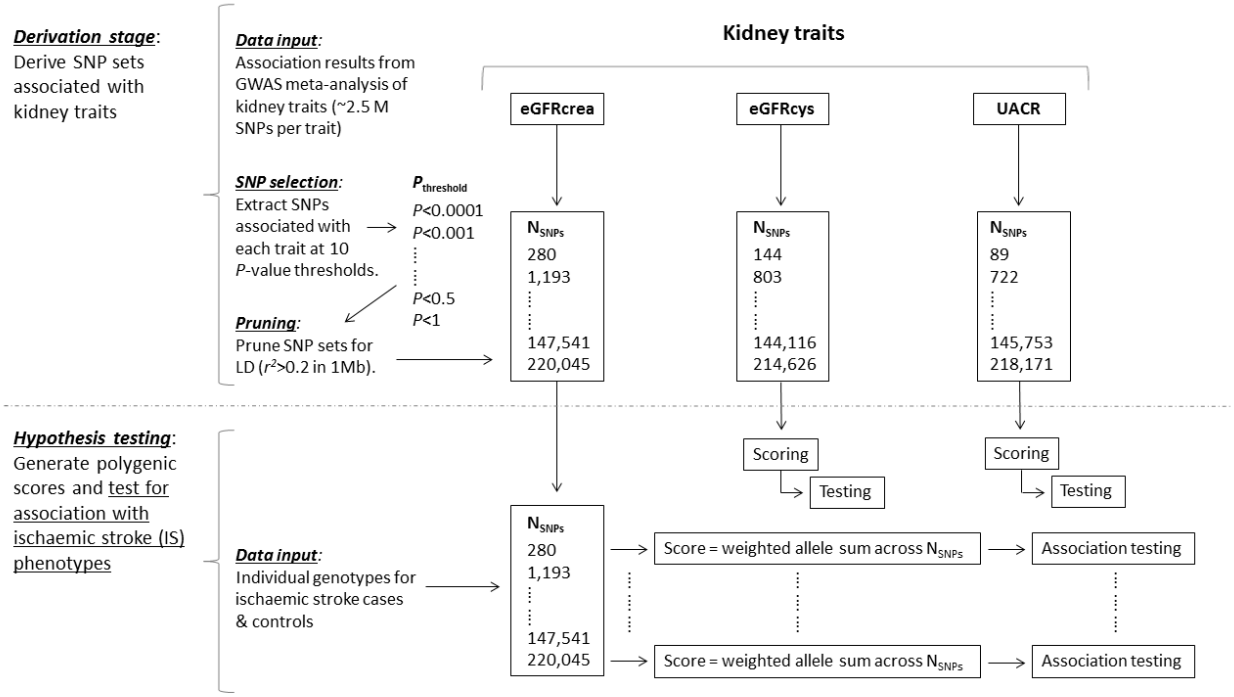

Figure I. Flow-chart representing the polygenic analysis approach.

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
2. Traylor M, Farrall M, Holliday EG, Sudlow C, Hopewell JC, Cheng YC, et al. Genetic risk factors for ischaemic stroke and its subtypes (the metastroke collaboration): A meta-analysis of genome-wide association studies. *Lancet Neurol*. 2012;11:951-962
3. Sidak ZK. Rectangular confidence regions for the means of multivariate normal distributions. *Journal of the American Statistical Association*. 1967;62:626-633
4. Nyholt DR. A simple correction for multiple testing for single-nucleotide polymorphisms in linkage disequilibrium with each other. *Am J Hum Genet*. 2004;74:765-769
5. Wain LV, Verwoert GC, O'Reilly PF, Shi G, Johnson T, Johnson AD, et al. Genome-wide association study identifies six new loci influencing pulse pressure and mean arterial pressure. *Nat Genet*. 2011;43:1005-1011
6. Olden M, Teumer A, Bochud M, Pattaro C, Kottgen A, Turner ST, et al. Overlap between common genetic polymorphisms underpinning kidney traits and cardiovascular disease phenotypes: The ckdgen consortium. *Am J Kidney Dis*. 2013;61:889-898
7. Kilarski LL, Achterberg S, Devan WJ, Traylor M, Malik R, Lindgren A, et al. Meta-analysis in more than 17,900 cases of ischemic stroke reveals a novel association at 12q24.12. *Neurology*. 2014;83:678-685
8. Pattaro C, Kottgen A, Teumer A, Garnaas M, Boger CA, Fuchsberger C, et al. Genome-wide association and functional follow-up reveals new loci for kidney function. *PLoS Genet*. 2012;8:e1002584
9. Boger CA, Chen MH, Tin A, Olden M, Kottgen A, de Boer IH, et al. Cubn is a gene locus for albuminuria. *J Am Soc Nephrol*. 2011;22:555-570
10. Snijders TAB, Bosker RJ. *Multilevel analysis*. Sage Publications Ltd; 2012:305-306.
11. Kottgen A, Pattaro C, Boger CA, Fuchsberger C, Olden M, Glazer NL, et al. New loci associated with kidney function and chronic kidney disease. *Nat Genet*. 2010;42:376-384
12. Tin A, Colantuoni E, Boerwinkle E, Kottgen A, Franceschini N, Astor BC, et al. Using multiple measures for quantitative trait association analyses: Application to estimated glomerular filtration rate. *J Hum Genet*. 2013;58:461-466
13. Kottgen A, Glazer NL, Dehghan A, Hwang SJ, Katz R, Li M, et al. Multiple loci associated with indices of renal function and chronic kidney disease. *Nat Genet*. 2009;41:712-717
14. Okada Y, Sim X, Go MJ, Wu JY, Gu D, Takeuchi F, et al. Meta-analysis identifies multiple loci associated with kidney function-related traits in east asian populations. *Nat Genet*. 2012;44:904-909
15. Gudbjartsson DF, Holm H, Indridason OS, Thorleifsson G, Edvardsson V, Sulem P, et al. Association of variants at umod with chronic kidney disease and kidney stones-role of age and comorbid diseases. *PLoS Genet*. 2010;6:e1001039
16. Gretarsdottir S, Thorleifsson G, Manolescu A, Styrkarsdottir U, Helgadottir A, Gschwendtner A, et al. Risk variants for atrial fibrillation on chromosome 4q25 associate with ischemic stroke. *Ann Neurol*. 2008;64:402-409
17. Holliday EG, Maguire JM, Evans TJ, Koblar SA, Jannes J, Sturm JW, et al. Common variants at 6p21.1 are associated with large artery atherosclerotic stroke. *Nat Genet*. 2012;44:1147-1151

18. Bellenguez C, Bevan S, Gschwendtner A, Spencer CC, Burgess AI, Pirinen M, et al. Genome-wide association study identifies a variant in hdac9 associated with large vessel ischemic stroke. *Nat Genet*. 2012;44:328-333
19. Ikram MA, Seshadri S, Bis JC, Fornage M, DeStefano AL, Aulchenko YS, et al. Genomewide association studies of stroke. *N Engl J Med*. 2009;360:1718-1728