

## Supplement

### Jaworek et al., Contribution of Common Genetic Variants to Risk of Early Onset Ischemic Stroke

#### 1. eTables

eTable 1. Sites participating in the Early Onset Stroke Genetics Consortium, grouped by analysis strata .....	3
eTable 2. Study Characteristics and Stroke Subtype Distribution .....	5
eTable 3. Genotyping platforms, quality control, and imputation parameters of non-SiGN studies providing Genotype Data or Summary statistics.....	10
eTable 4. TRANS GWAS: Peak associations of early onset stroke from genomewide analyses ( $p < 10^{-6}$ ) .....	13
eTable 5. EUR GWAS: Peak associations of early onset stroke from genomewide analyses ( $p < 10^{-6}$ ) .....	14
eTable 6. Peak associations of early onset stroke with stroke subtypes from genomewide analyses ( $p < 10^{-6}$ ) in TRANS and EUR meta-analyses.....	15
eTable 7. Association of early and late onset stroke (TRANS) with established stroke loci in MEGASTROKE and UKB.....	17

#### 2. eFigures

eFigure 1. Locuszoom plots of 2 SNPs associated with AIS in EUR.....	20
eFigure 2. Association of HABP2 with early onset stroke in the EOSC.....	18
eFigure 3. Association of ABO SNPs rs529565 (blood group O1) and rs635634 (blood group A1) with early and late onset ischemic stroke in European only analysis.....	19

#### 3. eMethods

3.1. Cohort Descriptions .....	21
3.1.1. Case cohorts .....	21
3.1.2. Control-only cohorts .....	40
3.2. Genome-wide association analysis .....	44
3.3 Conditional analysis at the ABO locus.....	45

3.4 Serologic ABO blood groups, ABO blood subgroups-determining SNPs, and stroke .....	45
3.4.1 Distribution of ABO blood group among early onset stroke, late onset stroke, and controls .....	45
3.4.2 Association of ABO haplotype-defining SNPs with stroke .....	45
3.5 Association of ABO SNPs with early and late onset stroke in UKB .....	46
3.6 Association of ABO SNPs with early and late onset VTE in UKB.....	47
3.7 Association of VTE PRS with early and late onset stroke .....	47
3.8 Co-location of associations at the ABO locus between EOS and other thrombotic-related disorders and biomarkers .....	47
4. eAppendix 1: Funding and acknowledgments for contributing studies .....	49
5. eReferences .....	61

eTable 1. Sites participating in the Early Onset Stroke Genetics Consortium, grouped by analysis strata

Group	Site	Site Location	Ancestral background	Cases	Controls	Gtyping platform
<b>SiGN studies</b>						
1	BRAINS	UK	EUR	66		650Q
	ISGS	USA	EUR	115		610
	GASROS	USA	EUR	30		610
	SWISS	USA	EUR	6		610
	HABC	USA	EUR		1,586	1M
2	EDIN	UK	EUR	105		660
	OXVASC	UK	EUR	54		660
	STGEORGE	UK	EUR	49		660
	MUNICH	GER	EUR	270		660
	KORA	GER	EUR		804	550
	WTCCC	UK	EUR		5,150	660
3A, 3E	GEOS	USA	EUR, AFR	843	881	1M
	CIDR <sup>1</sup>	USA,UK	EUR, AFR	1158		5M
4A,4E	HRS	USA	EUR, AFR		10,342	2.5M
	OAI	USA	EUR, AFR		3,882	2.5M
5	KRAKOW	POLAND	EUR	218	716	5M
6	LEUVEN	BELGIUM	EUR	137	453	5M
7	BARCELONA	SPAIN	EUR	413		HumanCore Exome
	BASICMAR	SPAIN	EUR	807		1M, 2.5M, HumanCore Exome
	INMA	SPAIN	EUR		807	1M
	ADHD	SPAIN	EUR		411	1M
8	GRAZ	AUSTRIA	EUR	142	816	5M (cases), 610 (controls)
9	GOTHENBURG (SAHLSIS)	SWEDEN	EUR	769	597	Illumina 5M array + exome, Illumina HumanOmni Express 700K, Illumina 2.5M, Illumina HumanOmni Express Exome BeadChip version 1.0, Illumina HumanOminExpress-24v1-1_A
	LUND	SWEDEN	EUR	103		5M
	MALMO	SWEDEN	EUR	73	1362	610
10	ASGC	AUSTRALIA	EUR	183	1200	610
11	FUTURE	NETH	EUR	283		Illumina GSA
	ODYSSEY	NETH	EUR	469		Illumina GSA
	MINE	NETH	EUR		2839	700K, 2.5M
13A,13E	VISP	USA	EUR, AFR	521		1M
	MELANOMA	USA	EUR		3033	Omni1
	HANDLS	USA	AFR		971	550K,370K,240S,1M
14	CEDIR	ITALY	EUR	385		610, 660W
	PROCARDIS	ITALY	EUR		948	610

eTable 1. Sites participating in the Early Onset Stroke Genetics Consortium, grouped by analysis strata

Group	Site	Site Location	Ancestral background	Cases	Controls	Gtyping platform
15	RACE1	PAKISTAN	ASIAN	902		660W
	PROMIS	PAKISTAN	ASIAN		1,253	???
16	SIFAP	GERMANY	EUR	830		2.5M
	KORA	GERMANY	EUR		1773	2.5M
17 SS	BIOBANK JAPAN	JAPAN	ASIAN	2,806	153,299	OmniExpress
18 SS	CADISP	EUR	EUR	555	9,259	610
19 SS	EPIC CVD	EUR	EUR	428	11,774	HumanCore Exome
20 SS	GEISINGER	USA	EUR	291	1,550	GSA
21 SS	GENESIS (DNA Lacunar)	UK	EUR	703	965	HumanCoreExome
21A SS	INTERSTROKE	Multiple <sup>2</sup>	EUR	493	2,384	Axiom PMRA
21B SS	INTERSTROKE	Multiple <sup>3</sup>	AFR	182	531	Axiom PMRA
21C SS	INTERSTROKE	Multiple <sup>4</sup>	AMR	189	948	Axiom PMRA
21D SS	INTERSTROKE	Multiple <sup>5</sup>	EAS	164	440	Axiom PMRA
22 SS	SLESS	UK	AFR CARIB	230	868	MEGA array
23 SS	UK BIOBANK	UK	EUR	1,491	375,128	UK BiLEVE Axiom / UK Biobank Axiom
24 SS	RACE2	PAKISTAN	ASIAN	339	3,295	OmniExpress
25 SS	HELSINKI	FINLAND	EUR	558	1,148	HumanCore Exome /GSA
TOTAL				17,360*	601,413	

Abbreviations: SS = summary stats; E or EUR = European; A or AFR =African; AFR CARIB= Afro-Caribbean; AMR= Admixed Americans; EAS= East Asians.

<sup>1</sup> CIDR = BRAINS, GASROS, GCNKSS, ISGS, MCISS, MIAMISR, NHS, NOMAS, REGARDS, SPS3, SWISS, WHI, WUSTL

<sup>2</sup> Canada, Australia, Germany, Denmark, Argentina, Brazil, Croatia, Poland, Peru, UAE, Sweden, United Kingdom, Russia, Ireland

<sup>3</sup> Canada, Germany, South Africa, Mozambique, Uganda, Sudan, Nigeria, UAE, United Kingdom

<sup>4</sup> Canada, Argentina, Brazil, Chile, Colombia, Ecuador, Peru, Sweden, United Kingdom

<sup>5</sup> Canada, Australia, Brazil, Philippines, Ecuador, Malaysia, UAE, Ireland

\* 17,360 with phenotypes and genotypes, of whom 16,927 used in analysis.

eTable 2. Study Characteristics and Stroke Subtype Distribution									
Group	Helsinki	Geisinger	GENESIS	INTERSTROKE (ALL)	INTERSTROKE (EUR)	INTERSTROKE (AFR)	INTERSTROKE (AMR)	INTERSTROKE (EAS)	SLESS
Total Cases	558	291	703	1028	493	182	189	164	230
Total Controls	1148	1550	965	4303	2384	531	948	440	868
Source of Controls	Community (1)	Internal	Community (2)	Community (3)	Community (3)	Community (3)	Community (3)	Community (3)	Community (2)
Age (and SD) of 1st stroke in cases	47.9 (9.4)	51.4 (7.2)	50.4 (7.5)	48.8 (8.9)	50.2 (8.2)	46.0 (10.3)	49.8 (9.0)	47.7 (8.7)	49.6 (7.7)
Age (and SD) at last exam in controls	69.8 (6.6)	78.0 (5.8)	59.7 (4.3)	63.0 (14.0)	66.4 (12.4)	56.8 (14.1)	63.7 (14.0)	55.1 (14.0)	58.7 (12.0)
% male in case; % male in controls	67.7; 41.5	49.8; 46.7	71.0; 52.5	61.8; 56.0	66.3; 56.9	53.3; 52.2	56.6; 52.1	59.8; 57.3	40.4; 52.1
Number of Cases who are 1st ever ischemic stroke (%)	521 (93%)	291(100%)	N/A*	1028 (100%)	493 (100%)	182 (100%)	189 (100%)	164 (100%)	N/A*
Number of Cases with imaging (MRI or CT) (%)	558 (100%)	291(100%)	830 (100%)	1028 (100%)	493 (100%)	182 (100%)	189 (100%)	164 (100%)	230 (100%)
TOAST subtypes:									
# of LAA:	62	73	23	218	142	19	20	37	34
# of CE:	118	69	22	179	88	32	34	25	45
# of SAO:	54	43	603	290	125	40	37	88	75
# of Other:	104	19	17	105	29	36	36	4	69
# of Undetermined:	220	87	30	236	109	55	62	10	--

eTable 2. Study Characteristics and Stroke Subtype Distribution										
Group	UK Biobank	Biobank Japan	EPIC-CVD	VISP	CEDIR	RACE1	RACE2	SIFAP	CADISP	FUTURE/ODYSSEY
Total Cases	1491	2806	428	383 (EUR)/98 (AFR)	368	898	283	830	555	698
Total Controls	375,128	153299	11774	2982 (EUR)/966 (AFR)	948	2139	2838	1773	9259	601
Source of Controls	Internal	Internal	Internal (4)	MELANOMA / HANDLS (5)	PROCARDIS (6)	PROMISE (7)	PROMISE (7)	KORA (8)	Various (9)	MinE (10)
Age (and SD) of 1st stroke in cases	52.8(5.71)	52.7 (6.14)	54.0 (5.5)	52.5 (5.1) / 52.3 (5.5)	45.0 (10.4)	46.8 (9.0)	50.2 (9.2)	41.7 (7.35)	43.73(9.9)	40.8 (7.6)
Age (and SD) at last exam in controls	63.45 (8.01)	62.9 (14.3)	68.9 (8.4)	52.1 (14.4) / 48.5 (9.0)	N/A*	51.8 (8.0)	60.9 (13.2)	55.2 (11.6)	NA*	63.0 (9.8)
% male in case; % male in controls	66.3; 45.8		39.4; 45.2	70.5/55.1; 58.8/44.6	56.0; 64.1	51.9; 56.4	80.4 ;55.8	60.7; 49.2	61.1; NA*	48.3; 57.1
Number of Cases who are 1st ever ischemic stroke (%)	1491 (100%)	2806 (100%)	428 (100%)	305 (79.6%)/73 (74.5)	0 (0.0%)	N/A*		715 (86.1%)	100%	686 (98.3%)
Number of Cases with imaging (MRI or CT) (%)	N/A*	MRI or CT: 2464 (88%); MRI or CT or MRA or cerebral angiography: 2517 (90%); MRI or CT or MRA or cerebral angiography or echoencephalography : 2521 (90%)	0	N/A*	N/A*	N/A*		100%	100%	100%
TOAST subtypes:										
# of LAA:	N/A*	259	N/A*	N/A*	50	N/A*	N/A*	157	67	100
# of CE:	N/A*	102	N/A*	N/A*	47	N/A*	N/A*	158	211	55
# of SAO:	N/A*	662	N/A*	N/A*	81	N/A*	N/A*	98	31	94
# of Other:	N/A*	120	N/A*	N/A*	35	N/A*	N/A*	140	18	113
# of Undetermined:	N/A*	1663	N/A*	N/A*	155	N/A*	N/A*	277	228	195

eTable 2. Study Characteristics and Stroke Subtype Distribution			
Group	Barcelona	Gothenburg	SiGN (CIDR subset)**
Total Cases	404	769	748 (EUR)/410 (AFR)
Total Controls	46	327	11820 (EUR)/2022 (AFR)
Source of Controls	GRECOS & ISSYS projects	population controls (11)	HRS, OAI, SOL and others (12)
Age (and SD) of 1st stroke in cases	49.7 (8.3)	48.5 (9.7)	50.1 (7.7)/50.1 (7.9)
Age (and SD) at last exam in controls	53.7 (5.1)	50.0 (9.3)	67.1 (10.5)/64.2 (10.2)
% male in case; % male in controls	71.8; 80.4	63.2; 62.1	65.6/49.8; 42.4/34.5
Number of Cases who are 1st ever ischemic stroke (%)	313 (82.6%)	647 (88.4%)	655 (87.6%)/322 (78.7%)
Number of Cases with imaging (MRI or CT) (%)	100%	100%	720 (96.3%)/409 (99.8%)
TOAST subtypes:			
# of LAA:	74	63	94/36 (EUR/AFR)
# of CE:	87	99	91/32
# of SAO:	51	104	215/142
# of Other:	34	114	65/11
# of Undetermined:	129	209	124/94

## eTable 2: Notes and References

\*N/A, Not Available

\*\* Description of SiGN in: NINDS Stroke Genetics Network, International Stroke Genetics Consortium. Loci associated with ischaemic stroke and its subtypes (SiGN): a genome-wide association study. *Lancet Neurol* 2016;15:174-184

- (1) Salomaa V. Genetic and Environmental Contributions to Cardiovascular Risk: Lessons From North Karelia and FINRISK. *Glob Heart*. 2016;11(2):229-233.
- (2) Traylor M, Bevan S, Baron JC, Hassan A, Lewis CM, Markus HS. Genetic Architecture of Lacunar Stroke. *Stroke*. 2015 Sep;46(9):2407-12.
- (3) O'Donnell M, Xavier D, Diener C, Sacco R, Lisheng L, et al.; INTERSTROKE investigators. Rationale and design of INTERSTROKE: a global case-control study of risk factors for stroke. *Neuroepidemiology*. 2010;35(1):36-44.
- (4) Danesh J, Saracci R, Berglund G, Feskens E, Overvad K, et al.; EPIC-Heart. EPIC-Heart: the cardiovascular component of a prospective study of nutritional, lifestyle and biological factors in 520,000 middle-aged participants from 10 European countries. *Eur J Epidemiol*. 2007;22(2):129-41.
- (5) Cotugna N, Fanelli-Kuczmarski M, Clymer J, Hotchkiss L, Zonderman AB, Evans MK. Sodium intake of special populations in the Healthy Aging in Neighborhoods of Diversity Across the Life Span (HANDLS) study. *Prev Med*. 2013 Oct;57(4):334-8.
- (6) Barlera S, Chiodini BD, Franzosi MG, Tognoni G. Il PROCARDIS: un approccio attuale allo studio della genetica dell'infarto miocardico [PROCARDIS: A current approach to the study of the genetics of myocardial infarct]. *Ital Heart J Suppl*. 2001 Sep;2(9):997-1004.
- (7) Saleheen D, Zaidi M, Rasheed A, Ahmad U, Hakeem A, et al. The Pakistan Risk of Myocardial Infarction Study: a resource for the study of genetic, lifestyle and other determinants of myocardial infarction in South Asia. *Eur J Epidemiol*. 2009;24(6):329-38.
- (8) Holle R, Happich M, Löwel H, Wichmann HE; MONICA/KORA Study Group. KORA--a research platform for population based health research. *Gesundheitswesen*. 2005 Aug;67 Suppl 1:S19-25.
- (9) Debette S, Metso TM, Pezzini A, Engelter ST, Leys D, CADISP-group. CADISP-genetics: an International project searching for genetic risk factors of cervical artery dissections. *Int J Stroke*. 2009 Jun;4(3):224-30.
- (10) Project MinE ALS Sequencing Consortium. Project MinE: study design and pilot analyses of a large-scale whole-genome sequencing study in amyotrophic lateral sclerosis. *Eur J Hum Genet*. 2018 Oct;26(10):1537-1546.



- (11) Jood K, Ladenvall C, Rosengren A, Blomstrand C, Jern C. Family history in ischemic stroke before 70 years of age: the Sahlgrenska Academy Study on Ischemic Stroke. Stroke. 2005 Jul;36(7):1383-7
- (12) NINDS Stroke Genetics Network, International Stroke Genetics Consortium. Loci associated with ischaemic stroke and its subtypes (SiGN): a genome-wide association study. Lancet Neurol 2016;15:174-184

eTable 3. Genotyping platforms, quality control, and imputation parameters of non-SiGN studies providing Genotype Data or Summary statistics

Group	Helsinki	Geisinger	GENESIS	INTERSTROKE (ALL)	INTERSTROKE (EUR)	INTERSTROKE (AFR)	INTERSTROKE (AMR)	INTERSTROKE (EAS)
<b>SNP level quality control</b>								
Array (cases)	ILLumina HumanCore Exome / GSA	ILLumina Infinium OmniExpress Exome Array & ILLumina GSA-24v1-0 Array	HumanCoreExome	Axiom PMRA	Axiom PMRA	Axiom PMRA	Axiom PMRA	Axiom PMRA
Array (controls)	ILLumina HumanCore Exome	ILLumina Infinium OmniExpress Exome Array & ILLumina GSA-24v1-0 Array	HumanCoreExome	Axiom PMRA	Axiom PMRA	Axiom PMRA	Axiom PMRA	Axiom PMRA
<b>Pre-imputation variant exclusion criteria:</b>								
Per SNP call rate	95%	99%	99%	99%	99%	99%	99%	99%
HWE exclusion	1x10-6	1x10-5	1x10-5	1x10-8	1x10-8	1x10-8	1x10-8	1x10-8
MAF Cutoff	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01
<b>Subject level quality control</b>								
Exclude if gender mismatch?	Y	Y	Y	Y	Y	Y	Y	Y
Exclude if highly related? (if so, what PI-HAT?)	Y, > 0.1875	Y, > 0.125	Y, > 0.125	Y, KING 2nd degree relative or closer	Y, KING 2nd degree relative or closer	Y, KING 2nd degree relative or closer	Y, KING 2nd degree relative or closer	Y, KING 2nd degree relative or closer
genotyping rate cutoff	> 95 %	>95%	>95%	>95%	>95%	>95%	>95%	>95%
Imputation								
Pre-phasing software:	Eagle	Eagle	SHAPEIT	Eagle	Eagle	Eagle	Eagle	Eagle
Imputation software	Beagle	minimacv4	minimacv4 (Michigan Imputation Server)	PBWT	PBWT	PBWT	PBWT	PBWT
Genomics NCBI Build used	Hg38	Hg19(GRCh37)	Hg19	Hg19	Hg19	Hg19	Hg19	Hg19
Lifted Over to Build Hg38	N	Y	Y	Y	Y	Y	Y	Y
# of SNPs used as input for imputation:	241,704 / 404,044	?	241,789	697,521	697,521	697,521	697,521	697,521
Imputation reference panel:	Sequencing Initiative Suomi v3	HRC.r1-1 EUR	HRC 1.1	1KGUK10K Combined Reference Panel	1KGUK10K Combined Reference Panel	1KGUK10K Combined Reference Panel	1KGUK10K Combined Reference Panel	1KGUK10K Combined Reference Panel
Type of reported imputation quality (INFO/RSQ)	INFO	INFO >0.7	INFO	INFO	INFO	INFO	INFO	INFO
<b>Genetic association results</b>								
Genetic association software used:	PLINK(GLM)	PLINK (GLM)	SAIGE Logistic Mixed Model	SAIGE Logistic Mixed Model	SAIGE Logistic Mixed Model	SAIGE Logistic Mixed Model	SAIGE Logistic Mixed Model	SAIGE Logistic Mixed Model
# of PCs used	10	5	10	20	20	20	20	20
Number of total SNPs used for analysis:	8,760,341	6,514,283	6,839,658	10,856,933	10,856,933	10,856,933	10,856,933	10,856,933
Genomic control parameter/lambda all IS:	0.857	1.01	1.00	NA	0.97	1.00	0.99	0.96

eTable 3. Genotyping platforms, quality control, and imputation parameters of non-SiGN studies providing Genotype Data or Summary statistics							
					13	14	15
Group	SLESS	UK Biobank	Biobank Japan	EPIC-CVD	VISP	CEDIR	RACE1
SNP level quality control							
Array (cases)	IIlumina Multi-Ethnic Genotyping Array (MEGA)	UK BiLEVE Axiom array / UK Biobank Axiom array	IIlumina HumanOmni Express Exome BeadChip or a combination of the IIlumina HumanOmni Express and Human Exome BeadChips	HumanCore Exome	1M Array	720K, 660K, 610K	IIlumina Human660W-Quad
Array (controls)	IIlumina Multi-Ethnic Genotyping Array (MEGA)	UK BiLEVE Axiom array / UK Biobank Axiom array	IIlumina HumanOmni Express Exome BeadChip or a combination of the IIlumina HumanOmni Express and HumanExome BeadChips	HumanCoreExome	1M Array	610K	IIlumina Human660W-Quad
Pre-imputation variant exclusion criteria:							
Per SNP call rate	99%	99%	99%	97%	95%	95%	95%
HWE exclusion	1x10 <sup>-5</sup>	1x10 <sup>-5</sup>	1x10 <sup>-6</sup>	1x10 <sup>-6</sup>	1x10 <sup>-6</sup>	1x10 <sup>-6</sup>	1x10 <sup>-6</sup>
MAF Cutoff	0.01	0.01	Minor allele count of heterozygous variants ≥ 5	0.01	0.01	0.01	0.01
Subject level quality control							
Exclude if gender mismatch?	Y	Y	Y	Y	Y	Y	Y
Exclude if highly related? (if so, what PI-HAT?)	Y, >0.125	Y, >0.884 (KING)	Y, > 0.25	N	Y, >0.125	Y, >0.126	Y, >0.127
genotyping rate cutoff	>95%	>99%	>=98%	>95%	>95%	>95%	>95%
Imputation							
Pre-phasing software:	EAGLE2	see UKB documentation	Eagle (v2.3)	Eagle	Eagle	Eagle	Eagle
Imputation software	PBWT (Sanger server)		minimac3 (v2.0.1)	Eagle	minimacv4 (Michigan Imputation Server)	minimacv4 (Michigan Imputation Server)	minimacv4 (Michigan Imputation Server)
Genomics NCBI Build used	Hg19	Hg19	Hg19	Hg19	Hg38	Hg38	Hg38
Lifted Over to Build Hg38	Y	Y	Y	Y	N	N	N
# of SNPs used as input for imputation:	791,305	784,256	516,432	366,641	878,486	271,849	519,214
Imputation reference panel:	African Genome Resources haplotype reference panel	Haplotype Reference Consortium and UK10K haplotype resources.	1KGp3v5 ALL	HRC 1.1	TOPMED	TOPMED	TOPMED
Type of reported imputation quality (INFO/RSQ)	Rsq	INFO	Rsq	INFO	Rsq	Rsq	Rsq
Genetic association results							
Genetic association software used:	PLINK2 (GLM)	PLINK (GLM)	PLINK (GLM)	SAIGE Logistic Mixed Model	PLINK (GLM)	PLINK (GLM)	PLINK (GLM)
# of PCs used	10	10	n=10 for total cases GWAS // n=5 for subtype GWAS	10	10	10	10
Number of total SNPs used for analysis:	13,713,025	9,778,013	13,540,512	7,666,454			
Genomic control parameter/lambda all IS:	1.06	1.01	1	1.02	0.92 (EUR), 0.96 (AFR)	1.01	1.02

eTable 3. Genotyping platforms, quality control, and imputation parameters of non-SiGN studies providing Genotype Data or Summary statistics

		16		11			
Group	RACE2	SIFAP	CADISP	FUTURE/ ODYSSEY	Barcelona	Gothenburg	SIGN *
<b>SNP level quality control</b>							*
Array (cases)	II lumina HumanOmni Express	II lumina HumanOmni 2.5M-Quad	II lumina Human 610K BeadChip	II lumina GSA	II lumina HumanCore ExomeChip	II lumina 5M array + exome, II lumina HumanOmni Express Exome BeadChip version 1.0, II lumina HumanOmniExpress-24v1-1_A	*
Array (controls)	II lumina HumanOmni Express	II lumina HumanOmni 2.5M-Quad	II lumina Human 610K BeadChip	II lumina HumanOmni Express 700K and II lumina 2.5M	II lumina HumanCore ExomeChip, 1M, 2.5M	II lumina HumanOmni Express 700K, II lumina 2.5M, II lumina HumanOmni Express Exome BeadChip version 1.0	*
<b>Pre-imputation variant exclusion criteria:</b>							*
Per SNP call rate	95%	95%	95%	95%	95%	95%	*
HWE exclusion	1x10 <sup>-4</sup>	1x10 <sup>-6</sup>	1x10 <sup>-6</sup>	1x10 <sup>-6</sup>	-	1x10 <sup>-6</sup>	*
MAF Cutoff	NA	0.01	0.01	0.01	0.001	0.01	0.01
<b>Subject level quality control</b>							*
Exclude if gender mismatch?	Y	Y	Y	Y	Y	Y	Y
Exclude if highly related? (if so, what PI-HAT?)	N	Y, >0.128	Y, >0.129	Y, >0.130	Y, >0.130	Y, >0.130	*
genotyping rate cutoff	NA	>95%	>95%	>95%	0.95	>95%	*
Imputation							*
Pre-phasing software:	IMPUTE2	Eagle	SHAPEIT	Eagle	Eagle	Eagle	*
Imputation software	IMPUTE2	minimacv4 (Michigan Imputation Server)	IMPUTE2	minimacv4 (Michigan Imputation Server)	minimacv4 (Michigan Imputation Server)	minimacv4 (Michigan Imputation Server)	minimacv4 (Michigan Imputation Server)
Genomics NCBI Build used	Hg19	Hg38	Hg19	Hg38	Hg38	Hg38	*
Lifted Over to Build Hg38	Y	N	Y	N	Y	Y	
# of SNPs used as input for imputation:	NA	1,486,019	482,891	272,937	206,257	975,118	*
Imputation reference panel:	1KG phase 1 interim set (June 2011)	TOPMED	1KG phase 1 integrated variant set (March 2012)	TOPMED	TOPMED	TOPMED	TOPMED
Type of reported imputation quality (INFO/RSQ)	INFO	Rsq	Rsq	Rsq	Rsq	Rsq	Rsq
<b>Genetic association results</b>							*
Genetic association software used:	PLINK (GLM)	PLINK (GLM)	PLINK (GLM)	PLINK (GLM)	PLINK (GLM)	PLINK (GLM)	PLINK (GLM)
# of PCs used	10	10	10	10	10	10	10
Number of total SNPs used for analysis:	NA		31,008,974		8,402,119	8,375,448	*
Genomic control parameter/lambda all IS:	NA	1.14	NA	1.02	1.01	1.02	0.99-1.03

eTable 4. TRANS GWAS: Peak associations of early onset stroke from genomewide analyses ( $p < 1E-06$ )

rsNum	Chr37	Pos37	Ref	Alt	Type	Function	GENE	OR	pVal_SNP	Reference allele frequency (1000G)						Association with late onset stroke	
										1kG_ALL	1kG_AFR	1kG_AMR	1kG_EAS	1kG_SAS	1kG_EUR	OR	pVal_SNP
rs975381715*	9	136,137,106	G	A	intronic		'ABO	0.883	3.72E-10	NA	0.568	NA	NA	NA	0.632	0.955	2.73E-02
rs8176685**	9	136,138,765	GCGCCCACTA	G	intronic	deletion	'ABO	1.155	2.06E-09		0.069				0.194	1.052	1.03E-01
rs299624	5	34,066,317	C	T	ncRNA_intronic		'C1QTNF3-AMACR	0.927	6.81E-08	0.570	0.612	0.592	0.388	0.633	0.621	0.999	9.71E-01
rs201527389	3	169,611,109	A	AT	intergenic	insertion	'SAMD7	2.193	6.94E-08	0.008	0.002	0.020	0.000	0.001	0.023	NA	NA
rs201684748	9	118,559,652	TACAA	T	intergenic	deletion	'RP11-284G10.1	1.316	1.34E-07	0.030	0.003	0.069	0.001	0.051	0.047	0.948	3.47E-01
rs2830952	21	28,821,893	A	G	intergenic		'LINC00113	1.085	1.87E-07	0.209	0.119	0.223	0.050	0.374	0.318	0.992	6.79E-01
rs17004869	4	81,205,033	A	T	intronic		'FGF5	1.152	1.88E-07	0.080	0.002	0.117	0.185	0.091	0.044	1.065	2.78E-01
rs12445022	16	87,575,332	G	A	intergenic		'LOC101928737	1.082	2.14E-07	0.207	0.123	0.212	0.082	0.332	0.317	1.052	6.87E-03
rs3783840	14	89,871,574	G	A	intronic		'FOXN3	1.078	3.07E-07	0.416	0.464	0.231	0.586	0.522	0.206	1.048	2.78E-02
rs118027213	14	56,039,076	T	C	downstream		'KTN1-AS1	0.811	4.30E-07	0.039	0.002	0.069	0.012	0.075	0.058	0.962	2.35E-01
rs1037320	18	42,703,908	G	A	intergenic		'SETBP1	0.923	5.03E-07	0.279	0.028	0.329	0.687	0.184	0.258	0.973	1.93E-01
rs35597151	12	67,298,347	A	AG	intergenic	insertion	'RP11-123O10.4	0.904	5.05E-07	0.285	0.150	0.195	0.445	0.359	0.292	1.000	9.88E-01
rs144408492	3	129,137,347	C	T	intronic		'EFCAB12	0.790	5.37E-07	0.016	0.001	0.012	0.001	0.053	0.019	1.021	7.38E-01
rs148530193	4	148,567,824	G	A	intronic		'PRMT9	1.376	5.60E-07	0.030	0.076	0.006	0.041	0.003	0.000	NA	NA
rs367716945	4	96,514,044	T	C	intergenic		'UNC5C	1.595	5.86E-07	0.017	0.002	0.029	0.035	0.015	0.012	NA	NA
rs377424471	12	122,588,387	GT	G	intronic	deletion	'MLXIP	0.908	5.93E-07	0.530	0.685	0.445	0.537	0.461	0.444	1.000	9.97E-01
rs1879380	12	123,739,111	G	A	intronic		'C12orf65	0.908	6.41E-07	0.118	0.025	0.193	0.002	0.201	0.222	1.000	9.87E-01
rs9620654	22	27,430,724	G	A	intergenic		'LINC01422	1.098	6.97E-07	0.225	0.324	0.166	0.216	0.269	0.100	1.009	7.49E-01
rs56962280	19	33,543,980	C	T	intronic		'RHPN2	0.602	7.30E-07	0.026	0.093	0.007	0.000	0.000	0.000	NA	NA
rs78411354	9	27,129,468	C	T	intronic		'TEK	0.340	8.28E-07	0.008	0.029	0.001	0.000	0.000	0.000	NA	NA
rs11513729	12	112,273,499	C	T	downstream		'MAPKAPK5-AS1	1.086	8.87E-07	0.137	0.031	0.225	0.009	0.060	0.417	1.103	1.30E-07
peak SNP at each locus shown; correlated SNPs ( $r^2 \geq 0.50$ ) removed																	
* rs975381715 tags ABO subgroup O1. Allele frequency for this SNP from gnomAD V2.1.1																	
** rs8176685 tags ABO subgroup A1. Allele frequency for this SNP from gnomAD V2.1.1																	

eTable 5. EUR GWAS: Peak associations of early onset stroke from genomewide analyses ( $p < 1E-06$ )

rsNum	Chr37	Pos37	Ref	Alt	Type	Function	GENE	OR	pVal_SNP	Reference allele frequency						Association with late onset stroke	
										1kG_ALL	1kG_AFR	1kG_AMR	1kG_EAS	1kG_SAS	1kG_EUR	OR	pVal_SNP
rs529565*	9	136,149,500	T	C	intronic		'ABO	0.876	4.31E-14	0.632	0.579	0.785	0.619	0.607	0.632	0.960	3.63E-02
rs635634**	9	136,155,000	T	C	upstream		'ABO	1.160	6.54E-13	0.140	0.086	0.120	0.187	0.132	0.185	1.050	3.20E-02
rs118091666	19	41,087,999	A	G	intronic		SHKBP1	1.739	2.65E-08	0.004	0.000	0.000	0.000	0.000	0.018	0.938	5.27E-01
rs117470359	16	55,661,683	C	T	intergenic		'SLC6A2	1.342	1.52E-07	0.005	0.000	0.009	0.000	0.002	0.015	0.986	8.31E-01
rs61528440	14	89,179,735	CA	C	intronic	deletion	'EML5	1.119	1.67E-07	0.742	0.873	0.686	0.920	0.692	0.481	0.995	7.88E-01
rs142012236	18	58,109,417	C	T	intergenic		'MC4R	1.264	1.91E-07	0.014	0.002	0.030	0.001	0.011	0.035	0.987	8.06E-01
rs201527389	3	169,611,109	A	AT	intergenic	insertion	'SAM7	2.205	2.02E-07	0.008	0.002	0.020	0.000	0.001	0.023	NA	NA
rs28537480	12	122,605,846	T	C	intronic		'MLXIP	0.899	2.70E-07	0.503	0.634	0.408	0.509	0.457	0.435	0.999	9.77E-01
rs201684748	9	118,559,652	TACAA	T	intergenic	deletion	'RP11-284G10.1	1.319	3.24E-07	0.030	0.003	0.069	0.001	0.051	0.047	0.948	3.47E-01
rs11513729	12	112,273,499	C	T	downstream		'MAPKAPK5-AS1	1.091	4.11E-07	0.137	0.031	0.225	0.009	0.060	0.417	1.103	1.71E-07
rs10766994	11	23,050,783	A	G	intergenic		'CCDC179	1.107	4.98E-07	0.174	0.026	0.369	0.013	0.164	0.405	1.026	1.71E-01
rs1601239	1	43,332,039	C	T	ncRNA_intronic		'LOC339539	0.909	5.59E-07	0.622	0.624	0.496	0.521	0.710	0.723	1.006	7.76E-01
rs79582859	13	106,301,114	A	C	intergenic		'LINC00343	1.212	7.34E-07	0.014	0.004	0.026	0.000	0.003	0.045	0.959	3.31E-01
rs140393706	11	117,269,842	C	T	intronic		'CEP164	1.248	7.40E-07	0.014	0.005	0.026	0.000	0.005	0.039	0.974	6.00E-01
rs115085929	4	31,241,575	A	G	intergenic		'LOC102723778	1.365	7.87E-07	0.005	0.005	0.010	0.000	0.000	0.012	1.000	9.99E-01
rs141554609	5	125,137,437	C	CT	intergenic	insertion	'RP11-756H20.1	1.386	8.45E-07	0.087	0.055	0.144	0.000	0.057	0.205	NA	NA
rs10960888	9	13,037,088	C	G	intergenic		'SNORD137	0.904	9.48E-07	0.119	0.006	0.151	0.038	0.265	0.187	0.970	1.78E-01
peak SNP at each locus shown; correlated SNPs ( $r^2 \geq 0.50$ ) removed																	
* rs529565 tags ABO subgroup O1. rs912805253 was merged into rs529565 on October 12, 2018 (Build 152)																	
** rs635634 tags ABO subgroup A1.																	

**eTable 6. Peak associations of early onset stroke with stroke subtypes from genomewide analyses (p < 10E-6) in TRANS and EUR meta-analyses**

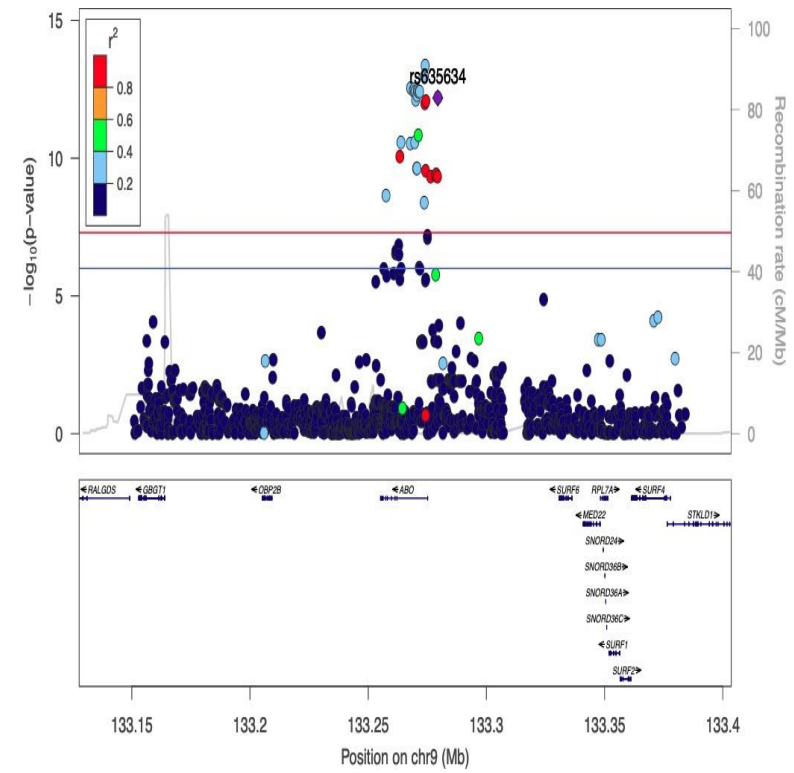
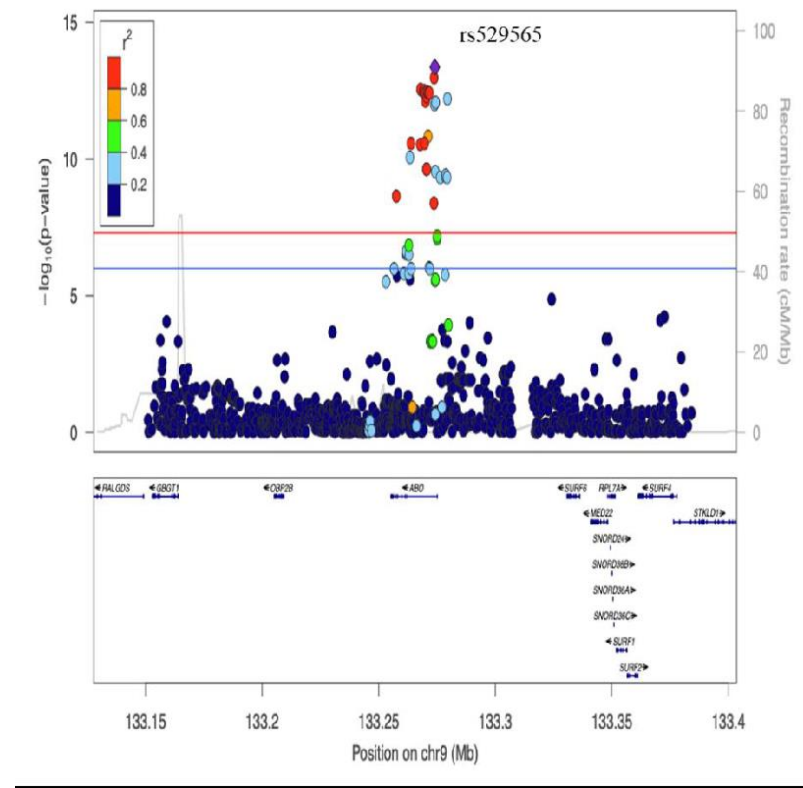
rsNum	Group	Phenotype	Chr37	Pos37	Chr38	Pos38	Ref	Alt	Type	Function	GENE	OR_SNP	pVal_SNP	1kG_ALL	1kG_AFR	1kG_AMR	1kG_EAS	1kG_SAS	1kG_EUR
rs138546363	EUR	toastLAA	1	10975110	1	10915053	C	T	intergenic		'C1orf127	4.211	6.65E-10	0.008	0.001	0.012		0.013	0.016
rs138061501	EUR	toastSAO	2	220234308	2	219369586	C	T	downstream		'DNPEP	4.721	2.49E-09	0.006		0.009	0.001	0.002	0.019
rs114766274	EUR	toastOTHER	4	152230379	4	151309227	A	C	intergenic		'PRSS48	3.532	1.25E-08	0.004		0.006		0.001	0.015
rs73140277	EUR	toastLAA	2	4451222	2	4403632	A	G	intergenic		'LINC01249	1.667	1.49E-08	0.171	0.192	0.228	0.168	0.187	0.089
rs186359940	EUR	toastCE	7	73637477	7	74223147	C	T	intronic		'LAT2	11.365	2.64E-08	0.004		0.009		0.009	0.004
rs718897	EUR	toastUNDETER	7	119953489	7	120313435	T	A	intronic		'KCND2	2.979	3.07E-08	0.004		0.007			0.017
rs117171806	EUR	toastLAA	20	37261272	20	38632629	T	A	intronic		'ARHGAP40	2.954	4.11E-08	0.005		0.006		0.010	0.009
rs17775754	EUR	toastLAA	7	14423673	7	14384048	C	T	intronic		'DGKB	2.617	4.31E-08	0.010	0.002	0.026	0.001	0.001	0.026
rs144758055	EUR	toastUNDETER	20	37692879	20	39064236	G	A	intergenic		'DHX35	2.422	4.32E-08	0.002	0.002	0.001			0.006
rs73555189	EUR	toastCE	19	44850272	19	44346120	C	G	intronic		'ZNF112	1.823	4.54E-08	0.071	0.138	0.022	0.002	0.089	0.071
rs117040291	EUR	toastLAA	10	34184136	10	33895208	T	C	intergenic		'LINC00838	3.680	4.59E-08	0.056	0.039	0.159	0.074	0.025	0.017
rs76553746	EUR	toastCE	19	44684507	19	44180354	G	A	downstream		'ZNF226	1.822	5.32E-08	0.035	0.002	0.016	0.002	0.092	0.068
rs10413396	EUR	toastCE	19	44813702	19	44309549	G	T	upstream		'ZNF235	1.817	5.33E-08	0.086	0.180	0.027	0.002	0.103	0.071
rs117614886	EUR	toastCE	9	79497797	9	76882881	G	T	intronic		'PRUNE2	2.846	7.05E-08	0.004	0.002	0.006	0.002	0.004	0.010
7:146788844	EUR	toastLAA	7	146788844	7	147091752	C	G	intronic		'CNTNAP2	3.670	7.31E-08	0.083	0.278	0.042		0.005	0.016
rs150108900	EUR	toastCE	19	44735425	19	44231272	GC	G	intronic	deletion	'ZNF227	1.800	1.07E-07	0.035	0.002	0.016	0.002	0.093	0.068
rs41348950	EUR	toastCE	19	44657679	19	44153526	G	A	intronic		'ZNF234	1.794	1.12E-07	0.043	0.056	0.020	0.001	0.058	0.069
rs143971940	EUR	toastLAA	7	3894970	7	3855338	G	T	intronic		'SDK1	2.868	1.14E-07	0.004				0.003	0.016
rs192999654	EUR	toastSAO	7	83763608	7	84134292	C	T	intronic		'SEMA3A	4.793	1.15E-07	0.003		0.009		0.001	0.006
rs145988306	EUR	toastCE	19	44766054	19	44261901	C	T	intronic		'ZNF233	1.798	1.15E-07	0.035	0.002	0.016	0.002	0.092	0.068
rs116092652	EUR	toastLAA	1	106812399	1	106269777	G	A	intergenic		'LOC101928476	2.237	1.28E-07	0.008	0.001	0.014		0.003	0.027
rs79205015	EUR	toastCE	2	126277949	2	125520372	G	A	intergenic		'CNTNAP5	2.282	1.41E-07	0.011		0.007		0.009	0.040
rs117735921	EUR	toastOTHER	17	35206634	17	36849369	G	T	intergenic		'LHX1	4.156	1.56E-07	NF					
rs146756136	EUR	toastCE	2	220862104	2	219997383	T	C	intergenic		'MIR4268	2.993	1.58E-07	0.003		0.003			0.012
rs17138265	EUR	toastCE	10	16305792	10	16263793	G	A	intergenic		'PTER	2.103	1.60E-07	0.105	0.142	0.040	0.130	0.142	0.038
rs62031324	EUR	toastSAO	16	9073716	16	8979859	G	A	intergenic		'USP7	4.123	1.62E-07	0.015	0.002	0.004	0.002	0.051	0.017
rs114392362	EUR	toastLAA	3	138695121	3	138976279	A	G	intergenic		'FOXL2NB	2.752	1.64E-07	0.008		0.016			0.027
rs148164535	EUR	toastSAO	2	220186877	2	219322155	C	T	intergenic		'RESP18	3.861	1.64E-07	0.006		0.009	0.001	0.007	0.016
rs55747216	EUR	toastCE	8	76067440	8	75155205	T	C	intergenic		'CAS9	1.842	1.70E-07	0.082	0.151	0.112	0.027	0.050	0.059
rs141967697	EUR	toastCE	7	30496937	7	30457321	C	T	intronic		'NOD1	2.795	1.70E-07	0.005	0.002	0.007			0.018
rs139338593	EUR	toastSAO	1	245210593	1	245047291	G	A	intronic		'EFCAB2	3.114	1.72E-07	0.004		0.007			0.015
rs6492022	TRANS	toastOTHER	13	107270064	13	106617716	A	T	upstream		'LINC00551	2.089	1.92E-07	0.092	0.188	0.055	0.067	0.057	0.052
rs148945476	TRANS	toastUNDETER	18	4372752	18	4372752	C	T	intronic		'DLGAP1	1.878	2.15E-07	0.009	0.004	0.007		0.003	0.032
rs11910117	TRANS	toastLAA	21	44707356	21	43287476	G	A	intergenic		'LINC00322	1.757	2.39E-07	NF					
rs57955269	TRANS	toastSAO	12	114981180	12	114543375	T	C	intergenic		'TBX3	1.720	2.39E-07	0.058	0.092	0.063		0.051	0.077
rs111314264	TRANS	toastUNDETER	12	102075496	12	101681718	C	T	intronic		'MYBPC1	2.022	2.43E-07	0.021	0.059	0.010		0.005	0.017
rs112652149	TRANS	toastSAO	12	114978974	12	114541169	T	C	intergenic		'TBX5-AS1	1.719	2.44E-07	0.059	0.092	0.063		0.052	0.077
rs72807423	TRANS	toastLAA	16	88591000	16	88524592	C	T	intronic		'ZFPM1	1.818	2.50E-07	0.104	0.102	0.062	0.095	0.171	0.081
rs79078388	TRANS	toastCE	2	180755578	2	179890851	T	A	intergenic		'ZNF385B	2.462	2.51E-07	0.006		0.019		0.003	0.014
rs10135707	TRANS	toastUNDETER	14	22497992	14	22029753	A	G	intergenic		'OR4E1	1.730	2.56E-07	0.100	0.159	0.030	0.169	0.066	0.037
rs12709100	TRANS	toastLAA	16	88591385	16	88524977	A	G	intronic		'ZFPM1	1.797	2.63E-07	0.197	0.418	0.092	0.100	0.194	0.081
rs2108605	TRANS	toastSAO	17	10352579	17	10449262	G	C	intronic		'MYH4	2.162	2.74E-07	0.010	0.000	0.007	0.000	0.011	0.034
rs73072592	TRANS	toastCE	7	24235100	7	24195481	G	A	intergenic		'NPY	2.949	2.85E-07	0.013		0.004		0.046	0.016
rs181437257	TRANS	toastSAO	13	54866997	13	54292862	G	A	intergenic		'MIR1297	2.905	2.91E-07	0.003	0.001	0.004		0.001	0.008
rs139154970	TRANS	toastCE	1	105155552	1	104612930	G	A	intergenic		'LOC100129138	4.412	3.10E-07	0.003			0.005	0.002	0.008
rs112859132	TRANS	toastSAO	16	77728363	16	77694466	C	A	intergenic		'NUDT7	2.478	3.40E-07	0.009	0.001	0.009		0.011	0.027
rs58501135	TRANS	toastCE	12	98107826	12	97714048	T	C	ncRNA_intronic		'LOC643711	2.427	3.52E-07	0.015	0.025		0.006	0.008	0.028
rs118049538	TRANS	toastLAA	11	100433497	11	100562766	C	T	intergenic		'LOC100128386	3.469	3.84E-07	0.004		0.012			0.010
rs74637497	TRANS	toastSAO	19	16184288	19	16073478	G	A	intronic		'TPM4	2.147	3.87E-07	0.007	0.002	0.009		0.003	0.023
rs11102917	TRANS	toastSAO	1	115839196	1	115296575	A	G	intronic		'NGF	2.557	4.13E-07	0.107	0.272	0.069	0.083	0.033	0.014
rs114984878	TRANS	toastLAA	3	138744275	3	139025433	C	G	upstream		'PRR23B	2.625	4.29E-07	0.008		0.016			0.028
rs79349941	TRANS	toastLAA	3	138758411	3	139039569	G	A	downstream		'PRR23C	2.625	4.29E-07	0.008		0.016			0.027

eTable 6. Peak associations of early onset stroke with stroke subtypes from genomewide analyses (p < 10E-6) in TRANS and EUR meta-analyses																			
rsNum	Group	Phenotype	Chr37	Pos37	Chr38	Pos38	Ref	Alt	Type	Function	GENE	OR_SNP	pVal_SNP	1kG_ALL	1kG_AFR	1kG_AMF	1kG_EAS	1kG_SAS	1kG_EUR
rs117380508	TRANS	toastSAO	17	53550274	17	55472913	C	T	intergenic		'MMD	4.521	4.31E-07	0.004	0.002	0.004			0.014
rs192179817	TRANS	toastLAA	15	81407227	15	81114886	A	G	intergenic		'CFAP161	5.571	4.36E-07	0.004	0.001	0.007		0.005	0.009
rs201952797	TRANS	toastSAO	12	114970491	12	114532686	AC	A	intergenic	deletion	'AC069240.1	1.694	4.62E-07	0.060	0.098	0.063		0.051	0.077
rs117755663	TRANS	toastOTHER	7	50609381	7	50541683	A	G	intronic		'DOC	2.716	4.70E-07	0.005		0.006		0.009	0.011
rs143040676	TRANS	toastUNDETER	16	77501101	16	77467204	T	TCA	intergenic	insertion	'RP11-449J10.1	2.142	4.76E-07	0.018	0.037	0.026		0.002	0.019
rs375259532	TRANS	toastSAO	2	233167625	2	232302915	TATTA	T	intronic	deletion	'DIS3L2	6.162	4.78E-07	0.003	0.002	0.006		0.002	0.006
rs34544149	TRANS	toastSAO	2	105833193	2	105216736	C	T	intergenic		'GPR45	1.520	4.81E-07	0.161	0.004	0.163	0.203	0.325	0.162
rs201924657	TRANS	toastOTHER	12	85384061	12	84990282	C	CT	intergenic	insertion	'TSPAN19	3.865	4.91E-07	0.005	0.001	0.009		0.004	0.013
rs78159697	TRANS	toastCE	2	109692580	2	109076124	G	A	intergenic		'SH3RF3-AS1	1.976	4.93E-07	0.012	0.001	0.019		0.001	0.044
rs72835462	TRANS	toastCE	6	17552500	6	17552269	T	C	intronic		'CAP2	2.482	4.97E-07	0.018	0.006	0.013	0.036	0.021	0.018
rs4788801	TRANS	toastSAO	16	71427960	16	71394057	A	G	upstream		LOC105371332	0.580	5.56E-07	0.170	0.027	0.079	0.435	0.181	0.144
rs114984274	TRANS	toastSAO	3	173272699	3	173554909	C	T	intronic		'NLGN1	2.119	5.70E-07	0.008	0.003	0.017			0.025
rs62479873	TRANS	toastCE	7	125093011	7	125452957	G	A	intergenic		'LOC101928283	2.136	5.75E-07	0.012	0.002	0.027		0.007	0.032
rs74777948	TRANS	toastCE	9	134763590	9	131888203	C	T	intronic		'MED27	1.687	5.87E-07	0.022	0.002	0.022		0.010	0.082
rs831610	TRANS	toastSAO	11	33719536	11	33697990	A	G	downstream		'C11orf91	1.459	5.90E-07	0.259	0.186	0.238	0.265	0.375	0.250
rs144328299	TRANS	toastCE	1	164086318	1	164117081	C	T	intergenic		'PBX1	3.896	5.90E-07	0.062	0.001	0.108	0.190	0.030	0.013
rs200017069	TRANS	toastCE	1	164088989	1	164119752	GA	G	intergenic	deletion	'U3	3.896	5.90E-07	0.062	0.001	0.108	0.190	0.030	0.013
rs73974808	TRANS	toastSAO	17	10560674	17	10657357	C	T	upstream		'MYH3	2.146	6.00E-07	0.039	0.058	0.012	0.009	0.073	0.030
rs116996267	TRANS	toastCE	18	24217756	18	26637792	T	C	intronic		'KCTD1	3.003	6.11E-07	0.006		0.004		0.007	0.019
rs141455408	TRANS	toastCE	20	35393978	20	36765575	C	T	intronic		'DSN1	2.323	6.14E-07	0.008	0.002	0.013		0.006	0.025
rs544695789	TRANS	toastCE	2	126682015	2	125924438	C	T	intergenic		'GYPC	2.328	6.19E-07	0.005	0.001	0.004		0.002	0.021
rs58224000	TRANS	toastCE	13	94673516	13	94021262	A	G	intronic		'GPC6	3.620	6.28E-07	0.020	0.061	0.012	0.001	0.004	0.007
rs76400184	TRANS	toastSAO	3	187025074	3	187307286	C	T	intergenic		'MASP1	3.116	6.32E-07	0.019	0.001	0.006	0.048	0.033	0.009
rs5002319	TRANS	toastOTHER	15	52586967	15	52294770	G	A	intronic		'MYO5C	1.619	6.41E-07	0.2919	0.325	0.264	0.4831	0.214	0.1471
rs113852703	TRANS	toastOTHER	9	110219234	9	107456953	G	A	ncRNA_intronic		'LINC01509	2.579	6.52E-07	0.011		0.009	0.001	0.022	0.028
rs34697222	TRANS	toastUNDETER	8	93360581	8	92348353	T	C	intergenic		'MIR7641-2	1.841	6.55E-07	0.005		0.012		0.008	0.010
rs113381921	TRANS	toastLAA	10	17588136	10	17546137	G	A	intergenic		'HACD1	1.743	6.65E-07	0.034	0.002	0.084		0.047	0.063
rs116954034	TRANS	toastSAO	15	84261443	15	83592691	T	C	intronic		'SH3GL3	2.517	6.79E-07	0.007	0.001	0.016		0.002	0.021
rs111280593	TRANS	toastOTHER	13	21568814	13	20994675	G	T	intronic		'LATS2	2.217	6.89E-07	0.018	0.002	0.043		0.018	0.039
rs77512494	TRANS	toastCE	6	39811813	6	39844037	A	C	intronic		'DAAM2	3.152	6.90E-07	0.019	0.003	0.012	0.057	0.007	0.019
rs73425643	TRANS	toastCE	6	37371860	6	37404084	G	A	intergenic		'RNF8	2.430	6.93E-07	0.063	0.105	0.020	0.106	0.036	0.015
rs769073055	TRANS	toastOTHER	4	142800387	4	141879234	T	G	intergenic		'IL15	2.032	6.96E-07	NF					
rs189094658	TRANS	toastCE	10	21687072	10	21398143	G	C	intergenic		'CASC10	2.124	7.23E-07	0.012		0.012		0.026	0.026
rs77947583	TRANS	toastCE	14	79299269	14	78832926	T	A	intronic		'NRXN3	3.731	7.27E-07	0.005		0.012		0.002	0.013
rs149402160	TRANS	toastUNDETER	5	159674312	5	160247305	C	T	downstream		'CCNJL	2.344	7.38E-07	0.001				0.001	0.006
rs202184545	TRANS	toastUNDETER	19	3977413	19	3977415	G	A	intronic		'EEF2	2.090	7.45E-07	0.004		0.003		0.006	0.010
rs2125962	TRANS	toastLAA	17	77135611	17	79139529	C	T	intronic		'RBFox3	2.534	7.79E-07	0.101	0.013	0.029	0.355	0.090	0.017
rs12213871	TRANS	toastSAO	6	47895747	6	47928011	T	C	intronic		'PTCHD4	2.881	7.89E-07	0.005		0.004		0.006	0.017
rs117918491	TRANS	toastLAA	10	108137531	10	106377773	A	C	intergenic		'SORCS1	2.974	8.29E-07	0.006	0.001	0.006		0.010	0.013
rs903701204	TRANS	toastSAO	7	64098115	7	64637737	CT	C	intergenic	deletion	'ZNF107	2.179	8.37E-07	NF					
rs72958156	TRANS	toastSAO	4	136157411	4	135236256	T	C	intergenic		'LINC00613	2.862	8.39E-07	0.024	0.065	0.006		0.003	0.024
rs149860703	TRANS	toastLAA	2	61398255	2	61171120	G	A	intergenic		'C2orf74	2.094	8.61E-07	0.019	0.001	0.016		0.061	0.022
rs17289631	TRANS	toastCE	7	55212613	7	55144920	C	T	intronic		'EGFR	2.809	8.69E-07	0.004	0.001	0.006			0.016
rs112164772	TRANS	toastLAA	5	49715802	5	50419968	G	C	intronic		'EMB	5.557	8.73E-07	0.004				0.002	0.017
rs148746448	TRANS	toastUNDETER	1	242658751	1	242495449	A	T	intronic		'PLD5	2.962	8.92E-07	0.012	0.001	0.009	0.002	0.036	0.016
rs77161568	TRANS	toastLAA	10	98474317	10	96714560	A	G	intronic		'PIK3AP1	3.808	8.93E-07	0.002	0.001	0.001			0.006
rs72835483	TRANS	toastCE	6	17584057	6	17584826	T	C	intergenic		'FAM8A1	2.399	9.04E-07	0.018	0.006	0.013	0.036	0.018	0.019
rs80076841	TRANS	toastLAA	3	44473438	3	44431946	C	A	intergenic		'ZNF445	2.668	9.06E-07	0.007	0.002	0.012			0.023
rs72946196	TRANS	toastUNDETER	11	36318090	11	36296540	T	A	intronic		'PRRS5	1.422	9.18E-07	0.093	0.026	0.092	0.243	0.051	0.071
rs7275586	TRANS	toastCE	21	19071529	21	17699211	A	T	intergenic		'C21orf91-OT1	3.466	9.24E-07	0.052	0.086	0.017	0.008	0.116	0.012
rs7727177	TRANS	toastOTHER	17	31426564	17	33099546	A	G	intronic		'ASIC2	2.810	9.41E-07	0.007	0.002	0.012		0.004	0.022
rs3104890	TRANS	toastSAO	8	96524590	8	95512362	T	C	ncRNA_intronic		'C8orf37-AS1	1.403	9.51E-07	0.535	0.873	0.411	0.410	0.444	0.392
rs149176481	TRANS	toastSAO	2	48934481	2	48707342	G	A	intronic		'LHCGR	3.204	9.69E-07	0.004	0.002	0.006		0.008	0.006
rs112837347	TRANS	toastCE	6	137559333	6	137238196	A	G	intergenic		'IFNGR1	1.890	9.75E-07	0.022	0.004	0.033		0.030	0.054
rs115384957	TRANS	toastCE	1	117646282	1	117103660	C	T	downstream		'TTF2	6.003	9.82E-07	0.003		0.001			0.012
rs184127974	TRANS	toastCE	3	20793336	3	20751844	C	A	intergenic		'LOC101927829	3.322	9.89E-07	0.007	0.002	0.006		0.018	0.009
rs143170653	TRANS	toastLAA	9	135850076	9	132974689	T	A	intronic		'GFI1B	3.585	9.90E-07	0.005	0.003	0.004		0.004	0.014
rs77227638	TRANS	toastLAA	5	140783285	5	141403718	C	T	exonic	missense	'PCDHGA9	3.463	9.92E-07	0.004	0.001	0.001			0.017



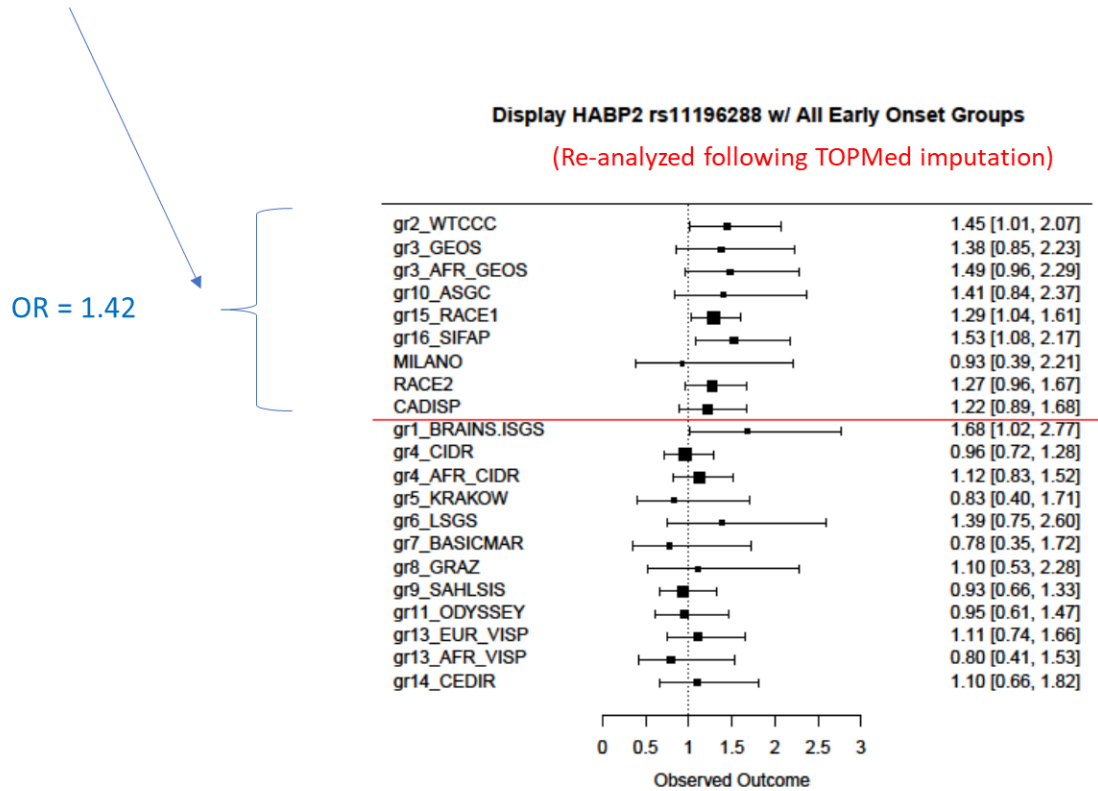
eTable 7. Association of early and late onset stroke (TRANS) with established stroke loci in MEGASTROKE and UKBB														
									MEGASTROKE		Early (age stroke onset < 60)		Late (age stroke onset ≥ 60)	
SNP	CHR	POS (37)	POS (38)	GENE	REF	ALT	1kg_ALL	PlatformFreq	OR	p-value	OR	p-value	OR	p-value
rs 880315		1	10796866	10736809 CASZ1	T	C	0.392	0.319	1.05	3.62E-10	1.04	6.97E-02	1.01	6.64E-01
rs 12037987		1	113042822	112500200 WNT2B	T	C	0.12	0.087	1.07	2.73E-08	1.05	4.69E-02	0.99	8.50E-01
rs 12124533		1	115657799	115115178 TSPAN2	C	T	0.113	0.171	1.17	1.22E-08	1.04	4.58E-01	1.06	2.48E-03
rs 1052053		1	156202173	156232382 PMF1-BGLAP	A	G	0.447	0.471	1.06	2.70E-14	0.94	2.71E-04	0.95	2.61E-02
rs 146390073		1	241306248	241142948 RGS7	C	T	0.007	0.011	1.95	2.20E-08	0.99	9.66E-01	1.33	1.65E-04
rs 12476527		2	26915624	26692756 KCNK3	G	T	0.426	0.477	1.05	6.44E-08	0.95	3.12E-02	1.00	9.39E-01
rs 7610618		3	149157706	149439919 TM4SF4	C	T	0.058	0.042	2.33	1.44E-08	1.12	1.74E-01	1.06	4.81E-01
rs 13143308		4	111714419	110793263 MIR297	T	G	0.602	0.695	1.32	1.86E-47	0.95	1.90E-03	0.92	1.55E-03
rs 34311906		4	113732090	112810934 ANK2	T	C	0.452	NA	1.07	1.07E-08	1.03	2.68E-01	NA	NA
rs 17612742		4	148414651	147493499 EDNRA	T	C	0.186	0.154	1.19	1.46E-11	1.08	4.09E-03	1.04	1.75E-01
rs 6825454		4	155501188	154580036 FGA	T	C	0.281	0.255	1.06	7.43E-10	1.04	1.85E-02	1.07	2.81E-03
rs 11957829		5	121515195	122179500 LOC100505841	A	G	0.18	0.19	1.07	7.51E-09	0.95	4.58E-02	0.96	2.59E-01
rs 6891174		5	172640590	173213587 NKX2-5	A	G	0.785	0.746	1.11	5.82E-09	0.99	7.25E-01	0.99	6.92E-01
rs 4959130		6	1356916	1356681 FOXF2	G	A	0.054	0.088	1.08	1.42E-09	1.06	1.81E-02	1.08	4.61E-03
rs 16896398		6	43262704	43294966 SLC22A7	A	T	0.459	0.439	1.05	1.30E-08	1.05	1.31E-01	1.06	2.34E-03
rs 2107595		7	19049388	19009765 HDAC9	G	A	0.253	0.198	1.21	3.65E-15	1.06	6.47E-04	1.10	8.17E-05
rs 42039		7	92244422	92615108 CDK6	C	T	0.127	0.179	1.07	6.55E-09	0.97	5.71E-02	0.93	4.24E-04
rs 1799983*		7	150696111	150999023 NOS3	T	G	0.824	0.771	1.05	2.20E-08	0.94	2.10E-04	0.97	2.33E-01
rs 7859727		9	22102165	22102166 CDKN2B-AS1	C	T	0.62	0.594	1.05	4.22E-10	1.05	5.04E-03	1.00	9.51E-01
rs 10820405		9	106010237	103247955 LINC01492	G	A	0.158	0.14	1.20	4.51E-08	0.98	1.88E-01	0.95	3.77E-02
rs 635634		9	136155000	133279427 ABO	C	T	0.86	0.836	1.08	9.18E-09	0.95	2.99E-01	0.95	1.15E-01
rs 2295786		10	105616482	103856724 SH3PXD2A	A	T	0.382	0.307	1.05	1.80E-10	0.95	5.75E-04	0.98	4.07E-01
rs 2005108		11	102770353	102899623 MMP12	C	T	0.067	0.095	1.08	3.33E-08	1.01	8.66E-01	1.05	8.16E-02
rs 7304841		12	20577593	20424659 PDE3A	A	C	0.426	0.437	1.05	4.93E-08	0.93	3.96E-03	NA	NA
rs 3184504		12	111884608	111446804 SH2B3	T	C	0.853	0.702	1.08	2.17E-14	0.93	6.07E-05	0.91	1.26E-07
rs 35436		12	115554523	115116718 TBX3	C	T	0.402	0.401	1.05	2.87E-08	0.96	2.66E-02	0.95	1.16E-02
rs 9526212		13	47225745	46651610 LRCH1	A	G	0.787	0.744	1.06	5.03E-10	1.07	1.92E-03	1.07	1.79E-03
rs 9521634*		13	110833899	110181552 COL4A1	T	C	0.321	0.335	1.04	3.80E-08	1.03	9.25E-02	1.02	3.50E-01
rs 4932370		15	91404705	90861475 FURIN	G	A	0.313	0.375	1.05	2.88E-08	1.04	2.20E-02	1.03	1.31E-01
rs 12932445		16	73069888	73035989 ZFHX3	T	C	0.213	0.176	1.20	6.86E-18	1.05	6.53E-03	1.07	2.71E-02
rs 12445022		16	87575332	87541726 LOC101928737	G	A	0.207	0.252	1.06	1.05E-10	1.08	2.14E-07	1.05	6.87E-03
rs 11867415		17	1571818	1668524 PRPF8	A	G	0.214	0.209	1.09	4.81E-08	1.04	4.59E-01	1.06	6.82E-02
rs 2229383		19	10794630	10683954 ILF3	G	T	0.664	0.662	1.05	4.72E-08	1.03	7.30E-02	1.06	1.08E-03
rs 8103309		19	11174935	11064259 SMARCA4	T	C	0.384	0.402	1.05	3.40E-08	0.95	4.42E-03	0.94	9.46E-04
rs 720470		21	38933888	37561586 DYRK1A	T	C	0.336	0.293	1.05	6.10E-09	0.95	9.64E-02	0.99	8.61E-01
From Traylor, Lancet Neurol, 2021. Lacunar stroke, EUR (6030 cases, inc SIGN)														
									Lacunar stroke (Traylor)		Early (age stroke onset < 60)		Late (age stroke onset ≥ 60)	
rs 72934535		2	203968973	203104250 ICA1L-WDR12-CA C	T		0.113	0.106	0.83	3.7E10-09	0.90	1.02E-04	0.90	3.49E-04
rs 4621303		3	41839370	41797878 ULK4	A	T	0.190	NA	1.15	1.7E10-07*	0.96	2.32E-01	NA	NA
rs 2293576		11	47434986	47413435 SPI1-SLC39A13-PS A	G		0.311	0.34	0.88	7.2E10-10	0.98	1.13E-01	0.95	8.80E-03
rs 12445022		16	87575332	87541726 ZCCHC14	G	A	0.317	0.364	1.13	2.5E10-08	1.08	2.14E-07	1.05	5.87E-03
rs 9958650		18	5389832	5389833 ZBTB14-EPB41L3	G	A	0.116	0.1	1.18	9.9E10-07	1.05	1.64E-02	1.08	1.75E-02
Traylor et al., Lancet Neurol 2021;20:351-61														
* genome-wide significant in transethnic analysis														

eFigure 1. Locuszoom plots of 2 SNPs in *ABO* associated with AIS in EUR ( $r^2 = 0.39$ ,  $D' = 1$ )

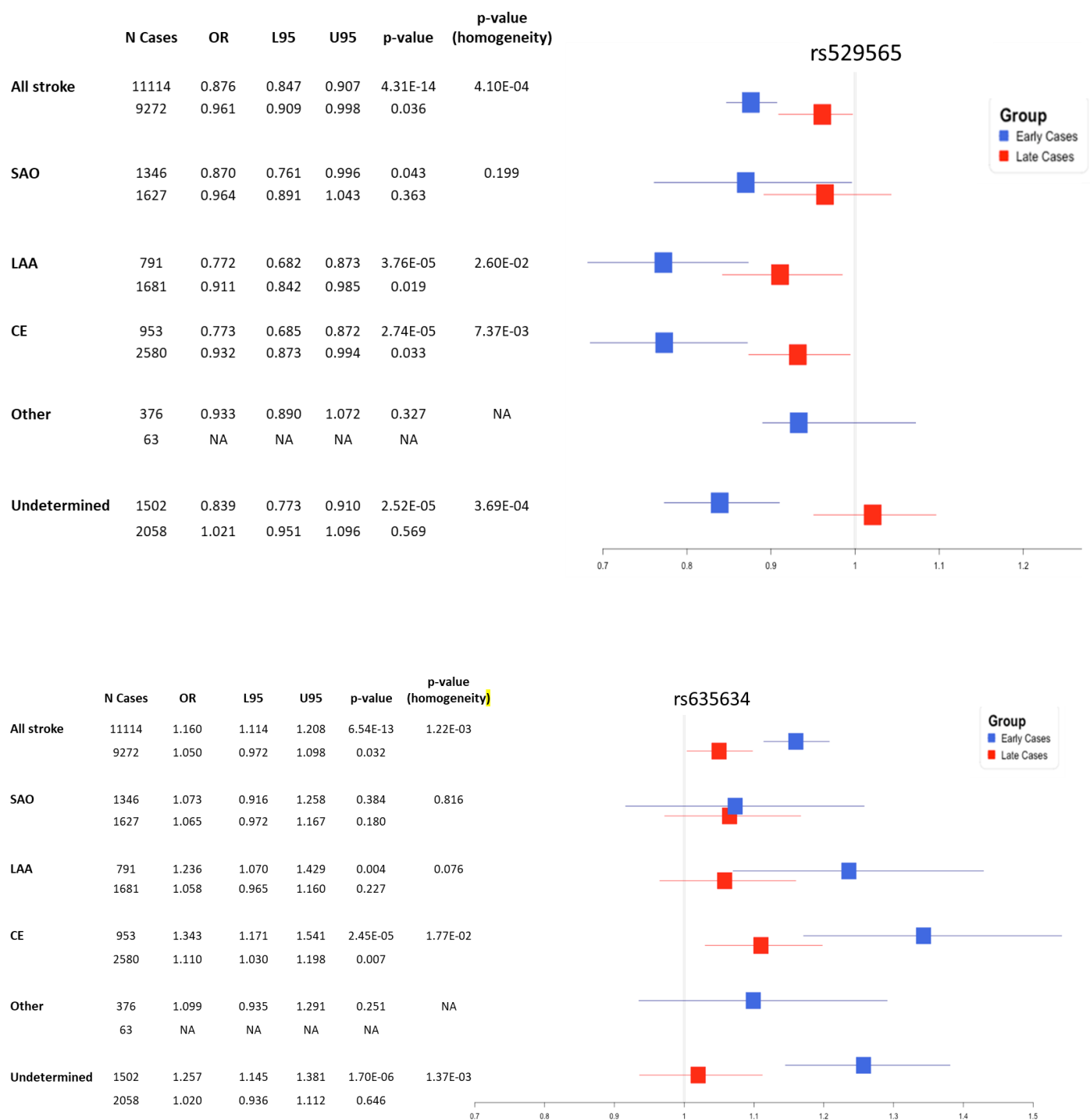


eFigure 2. Association of HABP2 with early onset stroke in the EOSC

**Genome-Wide Association Analysis of Young-Onset Stroke Identifies a Locus on Chromosome 10q25 Near *HABP2*. Cheng et al., *Stroke* 2016**



eFigure 3. Associations of *ABO* rs529565 (blood group O1; top panel) and rs635634 (blood group A1; bottom panel) with early and late onset ischemic stroke; European only



### 3. eMethods

#### **3.1 COHORT DESCRIPTIONS**

##### **3.1.1 Case cohorts**

###### **Australian Stroke Genetics Collaborative (ASGC)**

ASGC stroke cases comprised European-ancestry stroke patients admitted to four clinical centres across Australia (The Neurosciences Department at Gosford Hospital, Gosford, New South Wales (NSW); the Neurology Department at John Hunter Hospital, Newcastle, NSW; The Queen Elizabeth Hospital, Adelaide; and the Royal Perth Hospital, Perth) between 2003 and 2008. Stroke was defined by WHO criteria as a sudden focal neurologic deficit of vascular origin, lasting more than 24 hours and confirmed by imaging such as computerized tomography (CT) and/or magnetic resonance imaging (MRI) brain scan. Other investigative tests such as electrocardiogram, carotid doppler and transoesophageal echocardiogram were conducted to define IS mechanism as clinically appropriate. Cases were excluded from participation if aged <18 years, diagnosed with haemorrhagic stroke or transient ischemic attack rather than IS, or were unable to undergo baseline brain imaging. IS subtypes were assigned using TOAST criteria, based on clinical, imaging and risk factor data. Consenting participants completed five detailed self-report questionnaires and attended the HCS data collection centre, at which time a series of clinical measures were obtained. All study participants gave informed consent for participation in genetic studies. Approval for the individual studies was obtained from relevant institutional ethics committees.

ASGC controls were participants in the Hunter Community Study (HCS), a population-based cohort of individuals aged 55-85 years, predominantly of European Caucasian ancestry and residing in the Hunter Region, NSW, Australia. Briefly, participants were randomly selected from the NSW State electoral roll and contacted by mail between 2004 and 2007.

###### **BASEde datos de ICts del hospital del MAR (BASICMAR)**

BASICMAR is an ongoing prospective study of all acute strokes assessed since 2005 at the IMIMHospital Universitari del Mar (Barcelona, Spain). It includes both first-ever and recurrent strokes. There were no exclusion criteria regarding age or race-ethnicity of the individuals. All patients had an electrocardiogram (ECG), a blood analysis, and neuroimaging at the acute stage. Additional diagnostic procedures were performed when clinically indicated. A follow-up of three months after stroke was completed for all survivors. Ischemic stroke etiologic subtypes were classified according to TOAST criteria. For this study, only individuals of European origin with ischemic stroke were selected from BASICMAR, with eligible events defined as a clinical syndrome of any duration associated with a radiographically proven acute infarct, without radiographic evidence of a demyelinating or neoplastic disease or other structural disease including primary intracerebral hemorrhage.

## **Biobank Japan**

BioBank Japan Project was started in 2003 and collected DNA and clinical information from a total of 200,000 patients suffering from at least one of 47 common diseases, including ischemic stroke over age 40. Ischemic stroke was diagnosed by physicians at the 66 collaborating hospitals and its subtypes were determined by medical record review according to the TOAST criteria. Samples without ischemic stroke or intracranial aneurysm were used as controls.

## **Biorepository of DNA in Stroke (BRAINS)**

BRAINS is a hospital-based study. Cases participating in the current study were of European descent and recruited within the United Kingdom between September 25, 2009 and August 4, 2011. Extensive phenotype information is collected including subtype of stroke, past and family cardiovascular history, blood pressure data, MRI or CT brain imaging, carotid anatomy, and blood tests (including cholesterol). All hospital admitted participants over the age of 18 years with first-ever or recurrent stroke that provided informed consent (or caregivers on their behalf) were recruited. Participants must have image-positive lesions. Exclusion criteria are mainly for those that were brain image-negative, even if the clinical presentation is that of stroke. There are no eligibility criteria based on stroke severity or participation in a treatment trial. Inability to obtain consent results in mandatory exclusion.

Controls were European-Ancestry, stroke-free participants from the shared WTCCC controls, a prospectively collected cohort of individuals born in 1958 (1958 Birth Cohort).

## **Cervical Artery Dissections and Ischemic Stroke Patients (CADISP)**

The CADISP Study aims primarily at identifying genetic and environmental risk factors for cervical artery dissection (CeAD), a major cause of ischemic stroke in young adults. We included 942 CeAD patients in the CADISP study in 2004-2009 (CADISP-1; 170 Finns and 772 non-Finnish Europeans). Additional 451 CeAD patients of European origin were recruited in 2008-2010, exclusively for the CADISP-genetics project, in some CADISP centers and additional European and US centers (CADISP2). To assess the specificity of risk factors for CeAD, we also recruited 583 patients with an IS attributable to other causes (non-CeAD IS, 162 Finns and 421 non-Finnish Europeans from Belgium, France, Germany, Italy, and Switzerland), frequency-matched on age, sex and geographical origin with CADISP-1 CeAD patients. Of these, a total of 555 non-CeAD IS patients aged < 60 years, who were successfully genotyped and met genotyping quality control criteria, were available for the present analysis. The abstracted hospital records of cases were reviewed and adjudicated for IS subtype by a neurologist in each participating center. Each item required for the subtype classification was also recorded in a standardized fashion. Based on this, IS subtypes were then centrally readjudicated by a panel of neurologists, in agreement with the TOAST system, using the more detailed subtype description from Ay et al.<sup>1</sup> DNA samples were genotyped on an Illumina Human610-Quad or Human660W-Quad BeadChip® at the Centre National de Génotypage [CNG], Evry, France.

The majority of controls (N=9,046, of which 74 Finns and 8,972 non-Finnish Europeans) were selected from an anonymized control genotype database at the CNG, in order to match cases for ethnic background, based on principal component analysis. European reference samples from the genotype repository at the CNG were also analyzed simultaneously to provide improved geographical resolution. Additional Finnish controls were recruited within the CADISP study, both from the general population and among spouses and unrelated friends of CADISP patients, within the Helsinki area. A total of 234 individuals were eligible for genotyping at the CNG. Of these, 213 individuals who were genotyped successfully and met quality control criteria were available for the present analysis. Individuals were excluded if they were unexpected duplicates, gender discrepancy and unexpected relatedness. After quality control, we performed genotype imputation to the nonmonomorphic SNPs described in the 1000 genomes phase I v3 reference panel.

### **Edinburgh Stroke Study**

Between 2002 and 2005, consecutive consenting patients with stroke who were admitted to or seen as outpatients at the Western General Hospital, Edinburgh were prospectively recruited from stroke centers in Edinburgh, Scotland, U.K. There were no exclusion criteria for cases based on age, stroke severity, or inclusion in other clinical research studies. Cases in this study were of European origin, with a clinically evident stroke, demonstrated by brain imaging (CT or MRI) to be ischemic. An experienced stroke physician assessed each participant as soon as possible after stroke onset, prospectively recording demographic and clinical details, including vascular risk factors and results of brain imaging and other investigations. Ischemic subtypes were assigned according to the TOAST criteria and, subsequently, using CCS, specifically for the purposes of the SIGN study. ESS cases were collected as part of the WTCCC2 effort. All WTCCC2 cases were genotyped as part of the WTCCC2 Ischemic Stroke study using the Illumina Human660W-Quad array. Quality control procedures in the WTCCC2 excluded SNPs not genotyped on all case and control collections and SNPs with Fisher information measure < 0.98, genotype call rate < 0.95, MAF < 0.0 or Hardy-Weinberg P-value <  $1 \times 10^{-20}$  in either the case or control collections. Samples were excluded if identified as outliers on call rate, heterozygosity, ancestry and average probe intensity based on a Bayesian clustering algorithm. Samples were also removed if they exhibited discrepancies between inferred and recorded sex or were shown to have cryptic relatedness with other WTCCC2 samples (pairwise identity-by-descent > 0.05).

Controls from the UK samples were drawn from shared WTCCC controls obtained from the 1958 Birth Cohort. This is a prospectively collected cohort of individuals born in 1958 (<http://www.b58cgene.sgul.ac.uk/>) and ascertained as part of the national child development study (<http://www.cls.ioe.ac.uk/studies.asp?section=000100020003>).

### **EPIC-CVD**

EPIC is a multi-centre prospective cohort study of 519,978 participants (366,521 women and 153,457 men, mostly aged 35–70 years) recruited between 1992 and 2000 in 23 centres located in 10 European countries. Participants were invited mainly from population-based registers (Denmark, Germany, certain Italian centres, the Netherlands, Norway, Sweden, UK.

Other sampling frameworks included: blood donors (Spain and Turin and Ragusa in Italy); screening clinic attendees (Florence in Italy and Utrecht in the Netherlands); people in health insurance programmes (France); and health-conscious individuals (Oxford, UK). About 97% of the participants were of white European ancestry. Prevalent CVD was ascertained through self-reported history of MI or angina, or registry-ascertained CVD event prior to baseline. Stroke ascertainment in EPIC: Centres were asked to ascertain suspected stroke cases from registries, hospital records or self-report (i.e. follow-up questionnaires). Stroke events were defined by ICD10 codes as follows: Ischemic I63, Haemorrhagic I61, SAH I60, Unclassified I64, Other CRBV I62, I65- I69, F01. Incident stroke cases have been defined as fatal and non-fatal. All centres have recorded cause-specific mortality through mortality registries and/or active follow-up and have ascertained and validated incident fatal and non-fatal stroke through a combination of methods. Ascertained non-fatal stroke events were validated by clinical symptoms and imaging evidence (CT/MRI) or confirmed through hospital/GP records (with assessment of notes) or confirmed through hospital records (without assessment of notes). Individuals were excluded if they had clinical symptoms, but no validation was possible e.g. there was no imaging evidence, nor GP/primary care records (without assessment of notes) or registry information. Fatal stroke events were validated either by autopsy or hospital records and death certificate or by death certificate if they died in hospital. Individuals where validation was not possible were excluded. Participants with a history of stroke or MI at baseline were excluded. No further stroke subtyping was performed.

EPIC-CVD employs a nested case-cohort design, analogous to the EPIC-InterAct study for type-2 diabetes which established a common set of referents through selection of a random sample of the entire cohort ("sub cohort").

### **Follow-Up of Transient ischemic attack and stroke patients and Unelucidated Risk Factor Evaluation study (FUTURE)**

The FUTURE study is a prospective cohort study on risk factors and prognosis of young ischemic stroke and hemorrhagic stroke among 1,006 patients, aged 18-50 years, included in the study database between 1-1-1980 and 1-11-2010. Follow-up visits at the research centre occurred from the end of 2009 until the end of 2011. Control subjects were recruited among the patients' spouses, relatives or social environment. Information on mortality and incident vascular events will be retrieved via structured questionnaires. In addition, participants were invited to the research centre to undergo an extensive sub-study including MRI. Inclusion criteria for this consortium was ischemic stroke of presumed arterial origin. Exclusion criteria include traumatic hemorrhagic stroke, intracerebral hemorrhage in known cerebral metastasis or primary brain tumor, ischemic/hemorrhagic stroke due to cerebral venous sinus thrombosis, intracerebral hemorrhage due to ruptured cerebral aneurysm, any subarachnoid hemorrhage, and retinal infarct. Controls have to be at least 18 years old without a history of any TIA or stroke before the age of 50 at the moment of inclusion.

### **Mass General/GASROS**

MGH-GASROS enrolled ischemic stroke subjects as part of a single-center prospective cohort study of consecutive patients with ischemic stroke aged  $\geq 18$  years admitted to the



Massachusetts General Hospital Stroke Unit (Boston, MA, U.S.A.) between 2003 and 2011 after presenting to the emergency department within 24 hours of symptom onset. Ischemic stroke was defined as a clinical syndrome of any duration associated with a radiographically proven acute infarct consistent with a vascular pattern of involvement and without radiographic evidence of a demyelinating or neoplastic disease or other structural disease, including vasculitis, subacute bacterial endocarditis, vasospasm due to subarachnoid hemorrhage or cocaine abuse, or primary intracerebral hemorrhage. Diagnosis of acute cerebral ischemia was confirmed for all subjects in the present study by admission diffusion weighted imaging (DWI) completed within 48 hours after symptom onset. Vascular and critical care neurologists subtyped ischemic strokes by systematic medical record review using the TOAST and CCS systems. Controls were matched to cases on the basis of age, sex and race/ethnicity and drawn from stroke-free individuals who received care at primary care practices within Massachusetts General Hospital.

### **Greater Cincinnati/Northern Kentucky Stroke Study (GCNKSS)**

The GCNKSS is a population-based epidemiologic study of stroke in blacks and whites that is designed to measure temporal trends and racial differences in incidence of stroke. The catchment area includes two southwestern Ohio, U.S.A., counties (Hamilton, which includes the city of Cincinnati, and Clermont to the east) and three Northern Kentucky, U.S.A., counties (Boone, Kenton, and Campbell) to the south of Cincinnati across the Ohio River. As part of the GCNKSS, for calendar years 1999 and 2005, prospective cohorts of first-ever and recurrent ischemic stroke cases were assembled using “hot pursuit” methodology at all local hospitals in the region (18 in 1999, and 17 in 2005), except for one hospital that is solely devoted to treating pediatric cases. Subjects with all degrees of severity of stroke were eligible, and no particular racial group was intentionally oversampled (about 80% were white participants and 20% black). Study research nurses prospectively screened inpatient admission and emergency department logs to identify acute ischemic stroke patients. After consent was granted from the patient or legally authorized representative, a study nurse performed an extensive interview, and a blood sample was obtained for genetic analysis. In addition, a study nurse abstracted information about the individual, the subject’s medical history, the stroke event, and imaging studies from the hospital chart. A study physician reviewed every abstract, along with the imaging studies, to verify that an acute stroke had occurred, and to classify the event according to TOAST and CCS criteria.

### **Geisinger**

The study cohort was made up of participants of the Geisinger’s MyCode Community Health Initiative consisting of 946 patients with acute ischemic stroke with validated European ancestry. The informed consent was obtained for all MyCode patients. This study was approved by the Geisinger Institutional Review Board. Patient characteristics, clinical Variables, and outcome Measures were based on the neurological examination and corresponding neuroimaging. Controls were non-stroke patients with validated European ancestry.

## **Genetics of Early Onset Stroke (GEOS) Study**

The GEOS study is a population-based case-control study designed to identify the genetic determinants of early-onset ischemic stroke and to characterize interactions of stroke-associated genes with environmental risk factors. Cases with a first-ever ischemic stroke were identified by discharge surveillance from one of 59 hospitals in the U.S. greater Baltimore-Washington area and by direct referral from regional neurologists.

The abstracted hospital records of potential cases were reviewed and adjudicated for ischemic stroke, ischemic stroke subtype, and modified Rankin Scale <sup>2</sup> at discharge by a pair of vascular neurologists according to previously published procedures <sup>3 4</sup> with disagreements resolved by a third vascular neurologist. Stroke was defined according to the criteria of the WHO <sup>5</sup> and ischemic stroke was defined based on the criteria of the National Institute of Neurological Disorders and Stroke Data Bank.<sup>6</sup> Cases had a head CT and/or brain MRI that was consistent with cerebral infarction. Visualization of the infarct was not required, only that no alternative etiology was identified. The ischemic stroke subtype classification system retains information on all probable and possible causes and is reducible to the more widely used TOAST1 system that assigns each case to a single category. Cases were subsequently subtyped using the CCS.

Ischemic strokes with the following characteristics were excluded from participation: stroke occurring as an immediate consequence of trauma, stroke within 48 hours after a hospital procedure, stroke within 60 days after the onset of a non-traumatic subarachnoid hemorrhage, and cerebral venous thrombosis. Additional exclusions for genetic analyses modified from(10) were as follows: known single-gene or mitochondrial disorders recognized by a distinctive phenotype (e.g. cerebral autosomal dominant arteriopathy with subcortical infarcts and leukoencephalopathy (CADASIL), mitochondrial encephalopathy with lactic acidosis and stroke-like episodes (MELAS), homocystinuria, Fabry disease, or sickle cell anemia); mechanical aortic or mitral valve at the time of index stroke; untreated or actively treated bacterial endocarditis at the time of the index stroke; neurosyphilis or other central nervous system infections; neurosarcoidosis; severe sepsis with hypotension at the time of the index stroke; cerebral vasculitis by angiogram and clinical criteria; post-radiation arteriopathy; left atrial myxoma; major congenital heart disease; and cocaine use in the 48 hours prior to stroke onset. There were no exclusions based on race or ethnicity, stroke severity, or participation in clinical trial research.

Control participants without a history of stroke were identified by random-digit dialing. Controls were balanced to cases by age and region of residence in each study and were additionally balanced for ethnicity in SPYW-2 and SPYM.

## **GRAZ**

Between 1994 and 2003, subjects with first-ever and recurrent ischemic strokes admitted to the stroke unit of the Department of Neurology, Medical University of Graz (Graz, Austria) were included. All race-ethnic groups were eligible and there was no intentional oversampling. All age groups were allowed, though only subjects above the age of 18 were admitted to our department. Ischemic stroke was defined as an episode of focal neurological deficits with acute onset and lasting > 24 hours. There were no selection criteria based on stroke severity. Those individuals in treatment trials were excluded. 685 subjects were eligible to participate in this study (278 women, 407 men). All cases were Caucasian. Mean age was  $68.9 \pm 13.8$  years with an age range from 19 – 101 years. In addition to a standardized protocol including a laboratory examination and carotid ultrasound or magnetic resonance angiography and ECG, 304 subjects

underwent neuroimaging by CT and 381 by MRI. More extensive cardiac examination, including transesophageal echocardiography or transthoracic echocardiography and Holter, was performed in subjects with suspected cardiac embolism. Stroke subtypes were assessed according to modified TOAST criteria and were conducted by trained stroke neurologists. Controls were from the Austrian Stroke Prevention Study.

### **Helsinki Ischemic Stroke Genetics Study**

Helsinki Ischemic Stroke Genetics Study was designed for investigating genetic factors underlying ischemic stroke in the Finnish population and in the long-term to be incorporated to multicenter multinational similar datasets. Ischemic stroke cases were recruited from 2012 to 2017 from the Helsinki University Hospital, Department of Neurology, Helsinki, Finland, which is the only neurological emergency unit for a population of 1.7 million inhabitants. 1848 patients with positive neuroimaging findings for a new-onset brain infarction were included. Furthermore, we enrolled young patients aged 18 to 55 years with MRI-positive acute ischemic stroke from 2008 to 2010, and of these 72 patients have GWAS data available. Stroke subtyping was performed according to the Trial of Org 10172 in Acute Stroke Treatment (TOAST) classification. Detailed phenotypic patient data were recorded in the study database. All subjects are of white, European origin. All the cases have provided a blood sample for the genetic analyses and have been genotyped on the Illumina HumanCoreExome or Illumina Global Screening array. Altogether 558 cases aged 18 to 59 years were available for the present analysis. The study has been approved by the Ethics Committee of Medicine, Helsinki University Hospital. All the participants or their legal representatives have signed an informed consent. Control samples were obtained from the national FINRISK study 1992, 1997, 2002, 2007 and 2012 cohorts.<sup>7</sup> All the stroke free individuals aged 18 to 59 years at examination, residing in the same geographic area (Greater Helsinki region) and who were followed up until at least 60 years of age were considered as eligible controls.

### **INTERSTROKE**

INTERSTROKE (African Americans) is an international, multi-centered, case-control study of stroke investigating the global burden of risk factors across 32 countries and 18 different ethnic groups around the world. Briefly, cases were patients with acute first stroke (within 5 days of symptoms onset and 72 hours of hospital admission) in whom neuroimaging (CT or MRI) was performed. The TOAST classification system was used to define ischemic stroke subtypes. Cases were excluded if 1) they were unable to communicate due to severe stroke without a valid surrogate respondent (e.g. first-degree relative or spouse), 2) they were hospitalized for acute coronary syndrome/myocardial infarction, or 3) stroke was attributed to non-vascular causes (e.g. tumor).

Controls were selected from the community and had no history of stroke. The study was approved by the ethics committees in all participating centres. All participants, or their proxy, provided written informed consent before taking part in the study.

### **Ischemic Stroke Genetics Study (ISGS)**

The ISGS is a multicenter prospective, hospital-based inception cohort study of first-ever ischemic stroke. Enrollment for ISGS began in December 2002 and was completed in July 2007. All cases required meeting the WHO definition for stroke and head imaging, by either head MRI or CT, confirmed no alternative cause for the stroke symptoms other than focal cerebral ischemia. All participants had to be over age 18 years. There were no eligibility criteria based on stroke severity or enrollment status in a treatment trial. Cases were excluded if they had CADASIL, MELAS, homocystinuria, or sickle cell anemia or if their stroke was due to vasculitis, vasospasm due to subarachnoid hemorrhage, mechanical aortic valve or mechanical mitral valve, or occurred within 30 days of a vascular surgical procedure. Baseline assessment of patients included standardized assessment of demographics; medical history; vital signs; results of baseline blood tests, pre-stroke functional status per modified Rankin Scale; and National Institutes of Health Stroke Score (NIHSS) by certified examiner. Functional outcomes at 90 days post stroke onset were assessed using telephonic structured interview to obtain Oxford Handicap Scale, Glasgow Outcome Scale and Barthel Index. To minimize center-to-center variability, a single vascular neurologist (Robert D. Brown, Jr., MD) reviewed all available records of every ischemic stroke case for purposes of classification by etiology and syndrome using the TOAST criteria, along with the Baltimore-Washington criteria and the Oxford Community Stroke Project criteria. Medical records were received from the five centers and were stripped of personal identifiers, coded with study ID, and compiled in standard fashion. A separate neurologist independently reclassified all cases using the CCS system.

656 cases of first-ever ischemic stroke and 648 stroke-free controls were enrolled across the 5 centers (Mayo Clinic, Jacksonville, FL, U.S.A.; Mayo Clinic, Rochester, MN, U.S.A.; University of Virginia, Charlottesville, VA, U.S.A.; Shands Hospital, Jacksonville, FL, U.S.A.; Grady Hospital, Atlanta, GA, U.S.A.).

## **Krakow**

All consecutive subjects with ischemic stroke (fulfilling WHO criteria) who were admitted to the Stroke Unit at the Jagiellonian University (Krakow, Poland) and who provided informed consent were included in the study. The Stroke Unit serves as a stroke emergency center for one district of Krakow, Poland (200,000 inhabitants) and as a referral center for South East Poland (up to 15% of all admissions). For this on-going, prospective single-center, hospital-based study participants with first ever or recurrent strokes were recruited from January 22, 2002 to September 9, 2010. The Jagiellonian University Bioethics Committee approved the study. Participants in treatment trials were excluded. All subjects were of European origin. Stroke severity was not a criterion for inclusion or exclusion. All cases had performed clinically relevant diagnostic workup, including brain imaging with CT (100%) and/or MRI (up to 20%) as well as ancillary diagnostic investigations including duplex ultrasonography of the carotid and vertebral arteries (approximately 90%), and transthoracic echocardiography (approximately 70%). Magnetic resonance angiography (MRA), computed tomographic angiography (CTA), and ambulatory ECG monitoring, transesophageal echocardiography and blood tests for hypercoagulability were performed. Stroke cases were classified into etiologic subtypes according to TOAST. All cases were phenotyped independently by two experienced stroke neurologists with review of original imaging. Cases were subsequently classified additionally using the CCS system.

The control group included unrelated subjects taken from the population of southern Poland. Control subjects had no apparent neurological disease based on the findings in a structured questionnaire and a neurological examination. The Jagiellonian University Bioethics Committee approved the study and informed consent was obtained from all participants.

### **Leuven Stroke Genetics Study (LSGS)**

Cases of European descent with cerebral ischemia, defined as a clinical stroke with imaging confirmation or a TIA with a new ischemic lesion on diffusion-weighted imaging, who were admitted to the Stroke Unit of the University Hospitals (Leuven, Belgium) were enrolled in the LSGS between 2005 and 2009. All participants from the LSGS study underwent brain imaging (MRI in 91% of patients, CT in the remainder) and a standardized protocol including lab examination, carotid ultrasound or CTA and cardiac examination (echocardiography and ambulatory ECG monitoring) in all patients. Based on clinical presentation and results from the diagnostic work-up, cases were classified into ischemic stroke etiologic subtypes according to modified TOAST criteria<sup>10</sup> by a single reviewer. Large-vessel disease was defined as either occlusive or significant stenosis (corresponding to > 50% luminal diameter reduction according to North American Symptomatic Carotid Endarterectomy Trial (NASCET) criteria of a clinically relevant pre-cerebral or cerebral artery, presumably due to atherosclerosis. In case CTA was used as the primary imaging modality, stenosis was confirmed by carotid ultrasound. In case of posterior circulation infarcts on imaging, CTA or MRA was used as the primary imaging modality to determine the degree of stenosis. Probable causes of cardiac embolism were excluded.

Cardioembolic stroke was defined as ischemic stroke in the presence of atrial fibrillation, sick sinus syndrome, myocardial infarction in the past four weeks, cardiac thrombus, infective endocarditis, atrial myxoma, prosthetic mitral or aortic valve, valvular vegetations, left ventricular akinetic segment, dilated cardiomyopathy, or patent foramen ovale or atrial septal aneurysm. Significant stenosis/occlusion due to atherosclerosis of an appropriate pre-cerebral or cerebral artery should be excluded. Other determined cause of stroke included those with arterial dissection, vasculitis, hematologic disorders, monogenic syndromes, and complications of cardiovascular procedures. Dissection was diagnosed by typical findings on contrast-enhanced MRA and T1-fat suppressed MRI. Cryptogenic stroke was defined when no cause was identified despite an extensive evaluation. Strokes associated with significant aortic arch atheroma with plaques of  $\geq 4$  mm were also considered cryptogenic strokes. In addition to this primary classification, cases were reclassified using CCS.

Control individuals were recruited in the same population amongst healthy individuals, spouses of patients suffering from neurological diseases (amyotrophic lateral sclerosis, ischemic stroke, or multiple sclerosis), and from the Leuven University Gerontology Database as previously described.

### **Lund Stroke Register (LSR)**

The LSR is an ongoing study including consecutive subjects with first-ever stroke since March 1, 2001 from the local uptake area of Skåne University Hospital, Lund (Sweden). Stroke was defined using the WHO criteria. Subjects aged 18 years or older with stroke caused by cerebral

infarct, intracerebral hemorrhage or subarachnoid hemorrhage are included. Cases are included regardless of stroke severity, race-ethnic group belonging, or participation in any treatment trial. Those with iatrogenic or traumatic stroke are excluded. In the discovery phase of the SiGN study, subjects from LSR with first-ever ischemic stroke between March 1, 2001 and February 28, 2010 were included if they or their next of kin provided informed consent. Age over 90 years was set to 90 years to maintain anonymity. Every participant underwent CT, MRI, or autopsy of the brain, and ECG. Echocardiography, ultrasound, CTA or MRA of cerebral arteries was performed when judged clinically relevant. The subtype of ischemic stroke was determined using CCS.

For the secondary phase of SiGN, LSR individuals not included in the SiGN discovery phase participated after genotyping in the South Swedish genome-wide association study as follows: first ever ischemic stroke cases recruited in 2006 and 2010 to 2012, and age- and sex-matched LSR control subjects without stroke recruited in 2001 to 2002 and 2006 to 2007 from the same geographical area with use of the official Swedish population register.

### **Middlesex County Ischemic Stroke Study (MCISS)**

The MCISS was initiated as a prospective hospital-based stroke registry at the New Jersey Neuroscience Institute (Edison, NJ, U.S.A.). All cases over age 18 years were included, and no specific ethnic/racial group was targeted or excluded. From 2000 to 2009, 1,139 subjects with ischemic strokes were enrolled in this registry. There was no selection criterion based upon stroke severity, and both first-ever and recurrent strokes were included. Cases that were participants in treatment trials were not excluded. The major race/ethnic groups are Whites (67.2%), African Americans (14.3%), Asian Indians (8.2%), Hispanic (5.5%) and others (4.8%, Chinese and other Asians). All subjects with clinical suspicion of a stroke were admitted through the emergency room to a dedicated stroke unit supervised by a vascular neurologist. After a history and neurological examination, a standardized series of investigations were performed: complete blood count and differential, comprehensive metabolic panel, electrolytes, blood urea nitrogen, creatinine, lipid panel (total cholesterol, low-density lipoprotein, high-density lipoprotein, triglyceride levels, homocysteine levels, a cerebral MRI/MRA (if the MRI could not be performed, a head CT scan was done), carotid duplex ultrasound, ECG and an echocardiogram. The diagnosis of cerebral infarct was confirmed by the imaging studies. The epidemiological and clinical data on these participants was collected prospectively. Two independent investigators (one of which was a board-certified neurologist with expertise in vascular neurology) reviewed the data, and all strokes were classified into etiological subtypes using TOAST criteria. In addition, the Oxfordshire stroke classification was applied, and the vascular distribution of stroke was tabulated. All procedures, including the generation of the databases and recruitment of the stroke subjects, were conducted following Institutional Review Board policies and procedures at the New Jersey Neuroscience Institute/JFK Hospital.

Inclusion criteria was all subjects over the age of 18 years with clinical suspicion of a stroke. Exclusion criteria was subjects under the age of 18 years.

A control cohort of stroke free patients was also established. These volunteers have been recruited from the offices of local primary care physicians in Middlesex County, New Jersey and from the neurology clinic based at the New Jersey Neuroscience Institute. In addition to the

clinical information, DNA samples have been obtained from blood samples for the control groups.

### **Miami Stroke Registry and Biorepository (MIAMISR)**

The MIAMISR at the University of Miami/Jackson Memorial Hospital (Miami, FL, U.S.A.) is an ongoing prospective hospital registry of consecutive patients subjects with prevalent stroke (ischemic and hemorrhagic) and TIA with available neuroimaging (CT or MRI) who provide informed consent. There are no specific exclusion criteria with the respect to age, stroke severity, disability or participation in treatment trials. It was established in November of 2008 in order to investigate stroke type, ischemic stroke subtypes, stroke genetics and stroke outcomes in diverse ethnic population of Miami. The stroke population is predominately Hispanic (63%), with Cuba (32%), Nicaragua (4.8%), Colombia (4.8%), and Puerto Rico (4.1%) contributing the most subjects. Jackson Memorial Hospital is a 1,550-bed county hospital affiliated with the University of Miami with approximately 900 stroke and TIA admissions per year. Demographic and clinical data along with blood samples for genetic and other research have been collected prospectively during the hospitalizations. Follow-up information was obtained at 90 days by telephone interview or in person. Trained research staff obtained written informed consent from the stroke patients or the health care proxy when available for participation in MIAMISR.

### **Milano**

This study includes consecutive Italian patients referred to Besta Institute from 2000 to 2009 with stroke and included in the Besta Cerebrovascular Diseases Registry (CEDIR). Ischemic stroke cases, first ever or recurrent, confirmed on brain imaging, were selected for this study. All cases were of self-reported Caucasian ancestry and had clinically relevant diagnostic workup performed. All cases were phenotyped by an experienced stroke neurologist according to TOAST criteria, based on relevant clinical imaging and available information on cardiovascular risk factors. Controls are Italian individuals enrolled within the PROCARDIS Study, with no personal or sibling history of coronary heart disease before age 66 years.

### **Munich**

Subjects with first-ever or recurrent ischemic stroke were recruited consecutively from a single dedicated stroke unit (University Hospital, LMU Munich, Munich, Germany) from 2002 onward. All participants were over the age of 18 years and of European descent. Brain imaging was performed in all cases, with most patients (> 80%) undergoing MRI, including DWI. Diagnosis of ischemic stroke was based on neurological symptoms in combination with a documented acute infarct on neuroimaging. Subjects were not excluded based on stroke severity or whether they were enrolled in a treatment trial. Diagnostic workup included ECG and duplex ultrasonography of the extracranial carotid arteries in all cases. Transcranial ultrasonography, CTA and/or MRA, transthoracic and transesophageal echocardiography, and ambulatory ECG were performed if clinically indicated.

For the German MUNICH discovery samples and the Stroke in Young Fabry Patients (SIFAP) samples, independent control groups were selected from Caucasians of German origin participating into the population KORAgen study. This survey represents a sex- and age stratified random sample of all German residents of the Augsburg area and consists of individuals 25 – 74 years of age, with about 300 subjects for each 10-year increment. All controls were free of a history of stroke or transient ischemic attack.

### **Nurses' Health Study**

The NHS cohort consists of 121,700 female registered nurses aged 30 – 55 years who were residing in 11 U.S. states and who were enrolled in 1976 through responding to a mailed questionnaire on their medical history and lifestyle practices. They have been followed with biennial mailed questionnaires collecting information on disease risk factors and health status. From 1989 – 1990, blood samples were collected from 32,826 participants. Among these participants, we prospectively identified incident strokes and confirmed ischemic stroke cases by medical record review. Clinical symptoms consistent with stroke and exclusion of alternate etiologies were required for classification of stroke. Virtually all cases had imaging, but confirmation on CT or MRI was not required. No participants were excluded based on race/ethnicity. Neither stroke severity nor enrollment in a treatment trial was part of the eligibility criteria.

Incident ischemic strokes that were part of R01 biomarkers for ischemic stroke were included if sufficient clinical data was available.

### **Northern Manhattan Study (NOMAS)**

NOMAS is an ongoing population-based study designed to determine stroke incidence, risk factors and outcome in an urban multiethnic population. NOMAS started in 1993 as a case-control study of index ischemic stroke cases admitted to the Columbia University Presbyterian Medical Center (New York, NY, U.S.A.) and affiliated hospitals and matching community controls (Northern Manhattan Stroke Study, NOMASS) and continued as a prospective stroke incidence study by following up controls in 1997 (NOMAS). Demographic and clinical data were collected prospectively during the hospitalizations and annually by phone or in person. Genetic samples were derived from two sources: (a) the population-based case-control study conducted from 1993-98 (NOMASS) and (b) the ongoing prospective cohort study (NOMAS). First-ever ischemic stroke cases were identified for the case control study by screening of patient admissions, discharge codes, and referrals for neuroimaging at 15 acute care hospitals in the defined study area and multiple approaches to monitor for non-hospitalized cases. Incident ischemic stroke cases were identified from the prospective cohort study through follow-up visits and scheduled telephone contacts. Ischemic stroke cases from both sources were followed at 6 months by telephone and then annually afterwards in order to assess functional status and other outcomes. The administrative coordinating center of NOMAS moved from New York to Miami in 2007. The Institutional Review Boards of both institutions, Columbia University, and the University of Miami (Miami, FL, U.S.A.), approved the study.



NOMAS started in 1993 as a case-control study of index ischemic stroke cases admitted to the Columbia University Presbyterian Medical Center (New York, NY, U.S.A.) and affiliated hospitals and matching community controls (Northern Manhattan Stroke Study, NOMASS) and continued as a prospective stroke incidence study by following up controls in 1997 (NOMAS).

### **Odyssey**

ODYSSEY is a multicenter prospective cohort study on the prognosis and risk factors of patients with a first-ever TIA, ischemic stroke or intracerebral hemorrhage aged 18 to 50 years among 1490 patients. Primary outcome includes all-cause mortality and risk of recurrent vascular events. Secondary outcome will be the risk of post-stroke epilepsy and cognitive impairment. Patients have completed structured questionnaires on outcome measures and risk factors. Both well-documented and less well-documented risk factors and potentially acute trigger factors will be investigated. Patients will be followed for at least 10 years.

### **OXVASC**

OXVASC is an on-going population-based study of the incidence and outcome of cerebrovascular, cardiovascular, and peripheral vascular events since April 1, 2002. The OXVASC study population comprises all 91,105 individuals, irrespective of age, registered with 101 general practitioners in 9 general practices in Oxfordshire, UK. Multiple overlapping methods of “hot” and “cold” pursuit are used to achieve near complete ascertainment of as many cases as possible. All subjects are consented and seen by study physicians as soon as possible after their initial presentation. In the SiGN study, cases of all ethnic groups from OXVASC with any ischemic stroke between April 1, 2002 and August 31, 2010 were included if they consented to have research DNA samples extracted. Ischemic stroke was defined as an episode of focal neurological deficits with acute onset lasting > 24 hours or until death, with no apparent non-vascular cause, and no signs of primary hemorrhage on brain imaging. An infarct did not need to be seen on CT or MRI to be included in this study. Cases were not excluded if they were of a treatment trial or for their stroke severity. Demographic data, major vascular risk factors (hypertension, diabetes, smoking, hyperlipidemia, prior TIA and history of coronary disease or peripheral vascular disease), and symptomatology were recorded in all patients. Cases routinely had brain imaging (CT or MRI), vascular imaging (carotid Doppler or CTA /MRA or digital subtraction angiography), and 12-lead ECG. Echocardiography and 24-hour ambulatory ECG monitoring were done in selected patients. A senior neurologist subsequently reviewed all cases, and stroke etiology was classified according to modified TOAST criteria. Risk factors such as hypertension and diabetes were not included in the criteria. The subjects were classified as undetermined stroke only if the diagnostic workup was complete (any form of brain imaging plus ECG and any form of vascular imaging), but no clear etiology was found. Those with incomplete investigation were classified as unknown stroke while stroke of multiple causes was classified separately. Controls are collected through spouse or friend volunteers identified by TIA/stroke patients.

### **The Risk Assessment of Cerebrovascular Events Study (RACE 1 & RACE 2)**

The Risk Assessment of Cerebrovascular Events (RACE) Study is a retrospective case-control study designed to identify and evaluate genetic, lifestyle and biomarker determinants of stroke and its subtype in Pakistan. Samples were recruited from six hospital centres in Pakistan. Cases were eligible for inclusion in the study if they: (1) are aged at least 18 years; (2) presented with a sudden onset of neurological deficit affecting a vascular territory with sustained deficit at 24 hours verified by medical attention within 72 hours after onset (onset is defined by when the patient was last seen normal and not when found with deficit); the diagnosis was supported by CT/MRI; and (4) presented with a Modified Rankin Score of < 2 prior to the stroke. TOAST and Oxfordshire classification systems were used to sub-phenotype all stroke cases.

Control participants were individuals enrolled in the Pakistan Risk of Myocardial Infarction Study (PROMIS), a case/control study of acute MI based in Pakistan.<sup>8</sup> Controls in PROMIS were recruited following procedures and inclusion criteria as adopted for RACE cases. In order to minimize any potential selection biases, PROMIS controls selected for this stroke study were frequency matched to RACE cases based on age and gender and were recruited in the following order of priority: (1) non-blood related or blood related visitors of patients of the out-patient department; (2) non-blood related visitors of stroke patients; (3) patients of the out-patient department presenting with minor complaints.

### **Reasons for Geographic and Racial Differences in Stroke (REGARDS)**

The REGARDS study is a U.S. national, population-based, longitudinal cohort of 30,239 African American and white adults aged ≥ 45 years, recruited January 2003 to October 2007 with ongoing follow-up. Suspected stroke is queried every six months and triggered by participant self-report of stroke, stroke symptom(s), hospitalization, or proxy report of death. Stroke severity and participation in a treatment trial did not limit inclusion in this study. Medical records for these reported events were retrieved and reviewed by at least two members of a committee of stroke experts with disagreements resolved by a third adjudicator. A symptom-based approach, independent of neuroimaging outcome, was used to confirm events using the WHO definition of stroke. An infarct did not need to be seen on brain imaging to be included in this study. Ischemic stroke subtype classification was conducted using the TOAST system.

### **Sahlgrenska Academy Study on Ischemic Stroke (SAHLSIS)**

SAHLSIS is a case-control study of ischemic stroke based in Gothenburg, Sweden. Adult subjects who presented with first-ever or recurrent acute ischemic stroke before 70 years of age were recruited consecutively at stroke units in western Sweden from 1998 to 2012. All participants were of European origin. Patients were not excluded based on stroke severity or whether they were enrolled in a treatment trial. All participants underwent ECG and neuroimaging at the acute stage (all by CT and 58% also by MRI). Additional diagnostic work-up was performed as clinically indicated. Inclusion criteria was ischemic stroke which was defined as an episode of focal neurological deficits with acute onset and lasting > 24 hours or until death, with no apparent non-vascular cause, and no signs of primary hemorrhage on brain imaging. Subjects were excluded if they had a diagnosis of cancer at advanced stage, infectious hepatitis, or human immunodeficiency virus. Ischemic stroke was assigned according to modified TOAST criteria.

Healthy Caucasian community controls (n=600) matched for age and sex from the same geographic area were randomly selected from participants in a population-based survey or the Swedish Population Register.

### **Stroke in Young Fabry Patients (SIFAP)**

The SIFAP study is a multicenter study carried out to determine the frequency of Fabry disease in an unselected group of young adult patients with acute cerebrovascular events defined as having had an acute ischemic stroke or transient ischemic attack less than three months before enrollment into the study. First-ever (80.5%) and recurrent ischemic strokes were included. MRI was a mandatory procedure but, in the case of negative or missing MRI, a qualified stroke neurologist could confirm the clinical diagnosis. For this project, ischemic stroke cases recruited from 15 sites throughout Germany and determined not to have Fabry Disease were included in the analysis. All were of European ancestry and had age of first stroke of 18 – 55 years. The diagnosis of Fabry disease was based in males as well as in females in the first level on the sequencing data of the entire exon structure including promoter of the  $\alpha$ -galactosidase gene. In cases where a mutation was detected, biochemical analysis was done. Stroke cases from SIFAP were genotyped at CIDR (Baltimore, MD) using the Illumina Human Omni 2.5MQuad array. Only those cases without Fabry disease were selected for genotyping.

Controls free of cardiovascular diseases were selected from the KORA Study previously genotyped at CIDR in the same platform. The Cooperative Health Research in the Region of Augsburg (KORA) study is a population-based study of cardiovascular and metabolic traits carried out in the region of Augsburg, Southern Germany. A subset of control subjects (N = 28) was re-genotyped together with cases to provide cross-set duplicates. This joint clustering was used to minimize possible artifactual differences in allelic frequency between cases and controls due to genotyping at different times, and the cross-set duplicates were used to detect such artifacts that may have occurred.

### **The South London Ethnicity and Stroke Study (SLESS)**

The South London Ethnicity and Stroke Study (SLESS) is a prospective study begun in 1999 that has recruited consecutive black patients with stroke from a contiguous catchment area covered by 3 hospitals in South London (Guy's and St Thomas' Hospitals, King's College Hospital, and St George's Hospital). Ethnicity was defined according to the UK Census 2001 definition and classified as Black African or Black Caribbean. The study was reviewed and approved by the Wandsworth Local Research Ethics Committee, and informed consent was obtained from all participants. One consultant neurologist performed stroke subtyping using data collected on a standard proforma with additional review of all original brain imaging in all patients, as well as review of original notes when necessary. The pathophysiological Trial of Org 10172 in Acute Stroke Treatment (TOAST) subtyping classification was used for subtyping of ischemic stroke. Stroke ascertainment in SLESS: One consultant neurologist performed stroke subtyping using data collected on a standard proforma with additional review of all original brain imaging in all patients, as well as review of original notes when necessary. The pathophysiological Trial of Org 10172 in Acute Stroke Treatment (TOAST) subtyping classification was used for subtyping of ischemic stroke.

Recruitment of black controls was done by random selection from General Practice lists in the catchment areas of St George's, Guys and St Thomas, and King's College Hospital between 1999 and 2012. Potential controls were selected from age and gender strata matched to stroke cases. Furthermore, controls were identified within St George's University of London and St George's Hospital staff and contacted via email. Additionally, posters inviting healthy Black African and Black Caribbean individuals were displayed in local leisure centres, General Practice surgeries, churches, and community centres within the same catchment area as the that of the cases.

### **Secondary Prevention of Small Subcortical Strokes (SPS3)**

The SPS3 trial (NCT00059306) is a randomized, multicenter, Phase 3 trial of antiplatelet therapy and antihypertensive therapy. Participants are randomized to aspirin alone or the combination of aspirin and clopidogrel. Participants are also randomized to two groups of blood pressure control: either to a target systolic blood pressure of 130 – 149 mm Hg or < 130 mm Hg. Principal eligibility criteria include men or women at least 30 years of age with clinical evidence of small subcortical stroke and brain MRI evidence of small subcortical infarct. Subjects were required to not have evidence of ipsilateral symptomatic cervical carotid stenosis or high-risk cardioembolic sources for embolism. Primary outcomes included ischemic and hemorrhagic stroke. DNA samples were collected from 38% (1,139/3,020) of participants in the trial. These samples were obtained from 46% (37/81) participating centers across the U.S., Canada, Spain, Mexico, Chile, Ecuador, and Peru. No additional eligibility criteria were necessary beyond informed consent for participating in the DNA sub-study.

### **St. George**

First-ever and recurrent ischemic stroke cases of European descent attending a cerebrovascular service were recruited from 1995 to 2008. All cases were phenotyped by one experienced stroke neurologist with review of original imaging. All participants had clinically relevant diagnostic workup performed, including brain imaging with CT and/or MRI as well as ancillary diagnostic investigations including duplex ultrasonography of the carotid and vertebral arteries or MRA/CTA, blood tests, and ECG, and where clinically indicated echocardiography and ambulatory ECG monitoring was performed. Cases were enrolled only if a symptomatic acute infarct was detected on head imaging. Participants had to be over the age of 18 years and have provided informed consent. No case was excluded for participation in a treatment trial or because of stroke severity. An algorithm was established to use the clinical trials database to automatically populate the web-based CCS tool to generate CCS stroke subtype diagnoses. St. George cases were combined with DNA Lacunar Plus (removing any duplicates) and utilized the same set of controls.

### **Siblings with Ischemic Stroke Study (SWISS)**

SWISS is a prospective, hospital-based affected sibling pair study of ischemic stroke. Enrollment for SWISS began in December 2000 and was completed in February 2011. DNA samples were collected from 312 ischemic stroke-affected sibling pairs. During this time, 1,026

cases with first-ever or recurrent ischemic stroke were enrolled across 70 centers in North America (66 in the U.S. and 4 in Canada). All probands required at least one living sibling with a history of stroke and required meeting the WHO definition for stroke with head imaging, by either head MRI or CT, confirming no alternative cause for the stroke symptoms other than focal cerebral ischemia. Probands were excluded if they had CADASIL, MELAS, homocystinuria, or sickle cell anemia or if their stroke was due to vasculitis, vasospasm due to subarachnoid hemorrhage, mechanical aortic valve or mechanical mitral valve, or occurred within 30 days of a vascular surgical procedure. Baseline assessment of cases included standardized assessment of demographic and medical history. Siblings were recruited primarily using proband-initiated contact. Stroke-affected siblings were screened using the Questionnaire for Verifying Stroke-free Status (QVSS). Eligibility criteria for affected siblings were the same as for probands. The Stroke Verification Committee, composed of two vascular neurologists, confirmed ischemic stroke status in affected siblings by medical record review. For probands, the center principal investigator classified ischemic stroke using the original TOAST classification system. Center principal investigators were neurologists certified in TOAST classification using stroke vignette training and certification process. The Stroke Verification Committee classified all affected siblings using TOAST based on medical record review. The Committee received medical records stripped of personal identifiers, coded with study identification number, and compiled in standard fashion. All participants gave written informed consent for participation in the study, and the local institutional review boards of each individual clinical center and the Mayo Clinic institutional review board approved the study. Discordant siblings of the proband were confirmed to be stroke-free using the Questionnaire for Verifying Stroke-free Status.

### **UK Biobank Stroke Study**

The UK Biobank (UKB) <https://www.ukbiobank.ac.uk/> was established to improve understanding of common diseases including stroke. Participants were recruited from the general adult population and, in addition to having provided self-reported medical history at recruitment, are followed prospectively, chiefly through linkage to their National Health Service records.

For definition of stroke cases, all ischemic stroke cases from the algorithmically defined stroke outcomes (UKB fields 42008 and 42990) were used. If there was no date of stroke defined, this case was removed from the analysis. Furthermore, cases with an incident hemorrhagic stroke before the index ischemic stroke were also removed from the analysis. Only individuals of White British ancestry, as defined by UK Biobank, were selected.

Controls were all individuals in the UKB of White British Ancestry that had no history of any type of stroke (including both ischemic and hemorrhagic stroke).

### **VHIR-FMT- Barcelona**

The Barcelona cohort is a subset of Caucasian ischemic stroke subjects that were enrolled as a part of the Genetic contribution to functional Outcome and Disability after Stroke (GODs) project. Cases were selected through demonstration of acute ischemic stroke in a neuroimaging study during the first 7 days after stroke. We included cases with a first-ever and with a

recurrent stroke. We did not include lacunar strokes due to the study was focused only on disability after stroke of non-lacunar cases. We did not use age or stroke severity as exclusion criteria. Participants were not part of a treatment trial. Etiologic subgroups were classified following TOAST criteria. All the samples (cases and controls) were genotyped using the Infinium Human Core Exome Chip (Illumina). Written informed consent was obtained from all subjects with approval from the ethics committee of all participating institutions.

The control cohort was collected in primary care centers from Barcelona city and some hospital in the Spanish Network as a part of the Investigating Silent Stroke in hYpertensives: A magnetic resonance imaging Study (ISSYS) and Genotyping REcurrence Risk of Stroke (GRECOS) study. Controls were healthy subjects older than 40 years old without history of ischemic stroke.

### **Vitamin Intervention for Stroke Prevention (VISP)**

The VISP trial was a multicenter, randomized double-blind controlled clinical trial that enrolled subjects aged 35 years or older with homocysteine levels above the 25th percentile at screening and a non-disabling cerebral infarction within 120 days of randomization. Non-disabling cerebral infarction was defined as an ischemic brain infarction not due to embolism from a cardiac source, characterized by the sudden onset of a neurological deficit. The deficit must have persisted for at least 24 hours, or, if not, an infarction in the part of the brain corresponding to the symptoms must have been demonstrated by CT or MRI. The trial was designed to determine if daily intake of a multivitamin tablet with high-dose folic acid, vitamin B6 and vitamin B12 reduced recurrent cerebral infarction (primary endpoint), and nonfatal myocardial infarction or mortality (secondary endpoints). Subjects were randomly assigned to receive daily doses of the high-dose formulation (N = 1,827), containing 25mg pyridoxine (B6), 0.4mg cobalamin (B12), and 2.5mg folic acid; or the low-dose formulation (N = 1,853), containing 200µg pyridoxine, 6µg cobalamin and 20µg folic acid. Enrollment began in August 1997 and ended in December 2001, with 3,680 participants enrolled, from 55 clinic sites across the US and Canada and one site in Scotland. All participants provided written informed consent, and all local governing institutional review boards approved the trial. A subset of VISP participants provided separate consent for genetic analyses.

Control data for comparison with European ancestry VISP stroke cases were obtained through the database of genotypes and phenotypes (dbGAP) High Density SNP Association Analysis of Melanoma: Case-Control and Outcomes Investigation (phs000187.v1.p1; R01CA100264, 3P50CA093459, 5P50CA097007, 5R01ES011740, 5R01CA133996, HHSN268200782096C; PIs Christopher Amos, Qingyi Wei, Jeffrey E. Lee). For VISP stroke cases of African ancestry, a subset of the Healthy Aging in Neighborhoods of Diversity across the Life Span study (HANDLS) were used as stroke free controls.

### **Women's Health Initiative Observational Study (WHI-OS)**

The Women's Health Initiative Observational Study (WHI-OS) is a long-term follow-up study of postmenopausal women to identify and assess the effects of biological, genetic and lifestyle risk factors for cancer, cardiovascular disease, osteoporosis and other diseases of older women. The cases submitted here came from a case-control ancillary study nested within the WHI-OS

of the first 972 strokes occurring after WHI-OS baseline. This case-control study was the Hormones and Biomarkers Predicting Stroke Study (HaBPS), conducted to examine blood biomarkers in relation to stroke. Forty clinical centers throughout the United States enrolled 93,676 women ages 50 to 79 years at baseline into the parent study, the WHI-OS, between September 1993 and February 28, 1997. Follow-up for clinical events and exposures is ongoing. Recruitment into WHI-OS was mostly through mass mailings to age-eligible women from large mailing lists such as voter registration, driver's license, Health Care Financing Administration, or other insurance lists. Recruitment of minorities and older women was a particular study objective. Women were either specifically recruited for the Observational Study or entered it because they were ineligible or unwilling to be randomized into the Women's Health Initiative Clinical Trials of hormone therapy and/or dietary modification. Exclusions from WHI-OS were participation in other randomized trials, predicted survival of < 3 years, alcoholism, drug dependency, mental illness, dementia, or other conditions making them unable to participate in the study. Exclusions for the HaBPS case-control study of biomarkers of stroke were women with prior history of myocardial infarction or stroke or those who did not have adequate blood sample for biomarker assays. Strokes were first identified through annual mail and/or telephone follow-up, and participant or third-party reports of overnight hospitalizations which were further investigated by obtaining laboratory results, medical records, and available imaging study reports. Trained local physician adjudicators assigned a diagnosis according to standard criteria. Locally adjudicated strokes were sent for central adjudication by three neurologists. Two neurologists adjudicated each potential case, and disagreements were resolved by conference call consensus of the three neurologists. Only centrally confirmed ischemic strokes that required hospitalization were used in this study. TIAs and hemorrhagic strokes (determined on review of reports of brain imaging studies) were excluded. Ischemic stroke was defined as the rapid onset of a persistent neurologic deficit attributed to a vessel occlusion lasting more than 24 hours and without evidence for other causes. The deficit must have lasted > 24 hours unless death supervened or there was a lesion compatible with acute stroke demonstrated on CT or MRI scan. Ischemic strokes were also centrally classified by TOAST and CCS criteria.

### **Washington University St. Louis (WUSTL) Study**

The WUSTL patient collection included ischemic stroke cases admitted to Barnes-Jewish Hospital/Washington University Medical Center (St. Louis, MO, U.S.A.) for genetic studies starting from August 1, 2008. Participants were identified for the genetic studies by screening admissions at our tertiary care hospital (both in the Emergency Department and on the Inpatient Stroke Service) without regard to age, race or ethnicity, including both first-ever and recurrent strokes. Subjects were retained in the study if their discharge diagnosis was ischemic stroke (without requirement for the stroke to be visualized on CT or MRI). Demographic and clinical data were collected prospectively during the hospitalization and at 90 days, by phone or in person. Genetic samples were derived from subjects enrolled in 3 different studies: (a) Acute tPA pharmacogenomics study (Ischemic stroke cases who received tPA and were admitted to BJH/Washington University; serial NIHSS scores, and data on hemorrhagic transformation was collected) (b) Recovery Genomics after Ischemic Stroke Study (ReGenesIS, Ischemic stroke cases with NIHSS > 3 points without underlying chronic neurological disease, and expected survival up to 3 months after stroke), and (c) the Cognitive Recovery and Rehabilitation Group

(CRRG) Registry (all ischemic stroke cases admitted to BJH/Washington University who consent to entering their clinical data into a stroke registry, and the collection of blood for genetic analysis). Cases that were part of a treatment trial were excluded from the tissue plasminogen activator pharmacogenomics and ReGenesIS study, but not the CRRG registry.

Stroke subtyping using TOAST criteria/etiologies was done for all subjects participating in the studies. A subset of subjects was further subtyped according to the CCS (Causative Classification System of Ischemic Stroke) as part of the SiGN study.

### **Young Lacunar Stroke DNA Resource Plus**

A total of 1,029 Caucasian patients with lacunar stroke, aged  $\leq 70$  years, were recruited from 72 specialist's stroke centres throughout the UK between 2002 and 2012, as part of the Young Lacunar Stroke DNA Resource. DNA samples were available in 930 patients. An additional 82 Caucasian patients of all ages with lacunar stroke were recruited from St. George's Hospital, London as part of the GENESIS study. Lacunar stroke was defined as a clinical lacunar syndrome, with an anatomically compatible lesion on MRI (subcortical infarct  $\leq 15$  mm in diameter). All patients underwent full stroke investigation including brain MRI, imaging of the carotid arteries and ECG. Echocardiography was performed when appropriate. All MRIs and clinical histories were reviewed centrally by one physician. Exclusion criteria were: stenosis  $> 50\%$  in the extra- or intracranial cerebral vessels, or previous carotid endarterectomy; cardioembolic source of stroke, defined according to the TOAST (Trial of Org 10172 in Acute Stroke Treatment) criteria as high or moderate probability; cortical infarct on MRI; subcortical infarct  $> 15$  mm in diameter, as these can be caused by embolic mechanisms (striatocapsular infarcts); any other specific cause of stroke (e.g. lupus anticoagulant, cerebral vasculitis, dissection, monogenic cause of stroke). All cases were screened for *NOTCH3* CADASIL and Fabry disease mutations and positive cases excluded. All patients and controls underwent a standardized clinical assessment and completed a standardized study questionnaire. MRI was not performed in controls. The study was approved by the Multi-Centre Research Ethics Committee (04/MRE00/36) and informed consent was obtained from all participants.

Unrelated Caucasian controls, free of clinical cerebrovascular disease, were obtained by random sampling, stratified for age and sex, from general practice lists from the same geographical location as the patients.

#### **3.1.2 Control-Only Cohorts:**

##### **Attention-deficit Hyperactivity Disorder (ADHD)**

The Vall d'Hebron Research Institute (VHIR) cohort included 435 blood donors of Caucasian origin recruited from 2004 to 2008 at the Hospital Universitari Vall d'Hebron (Barcelona, Spain) to identify loci conferring susceptibility to Attention-Deficit Hyperactivity Disorder. Seventy-six percent of participants were male ( $N = 330$ ) and the average age at assessment was 43.8 years (s.d. = 14.3). Genome-wide genotyping was performed with the Illumina HumanOmni1-Quad BeadChip platform. The study was approved by the ethics committee of the institution and



informed consent was obtained from all participants in accordance with the Declaration of Helsinki.

## **FINRISK**

FINRISK surveys are cross-sectional, population-based studies conducted every 5 years since 1972 to monitor the risk of chronic diseases. For each survey, a representative random sample was selected from 25- to 74-year-old inhabitants of different regions in Finland. The survey included a questionnaire and a clinical examination, at which a blood sample was drawn, with linkage to national registers of cardiovascular and other health outcomes. The current study included eligible individuals from FINRISK surveys conducted in 1992, 1997, 2002, and 2007. The GWAS genotyping has been done earlier in several phases for different substudies: PredictCVD, Corogene and CoreExome. Participants with a history of stroke were excluded. Stroke ascertainment in FINRISK: During follow-up, participants were monitored for stroke through linkage of the study database with the National Hospital Discharge Register and the National Causes-of-Death Register. The clinical outcomes were linked to study subjects using their unique national social security ID, which is assigned to every permanent resident of Finland. The registers are countrywide covering all cardiovascular events that have led either to hospitalization or death in Finland. With both registers the diagnostic classification was done using the Finnish adaptation of ICD-codes: I63; not I63.6, I64 (ICD-10) / 4330A, 4331A, 4339A, 4340A, 4341A, 4349A, 436 (ICD-9) / 433, 434, 436 (ICD-8) for Ischemic stroke excluding any hemorrhagic strokes, and I60-I61, I63-I64 (not I63.6) (ICD-10) / 430, 431, 4330A, 4331A, 4339A, 4340A, 4341A, 4349A, 436 (ICD-9) / 430, 431 (except 431.01, 431.91), 433, 434, 436 (ICD-8) for allstroke including SAH. ICD-8 codes 430, 431 (excluding codes 431.01, 431.91 of the Finnish adaptation of ICD-8\*), 432, 433, 434 or with ICD-9 codes 430, 431, 433 (excluding codes 4330X, 4331X, 4339X of the Finnish adaptation of ICD-9\*), 434 (excluding code 4349X of the Finnish adaptation of ICD-9\*), 436, 437, 438 or with ICD-10 codes I60, I61, I63 (excluding I63.6), I64 or I69.34. An event found in either register was sufficient for diagnosis.

## **The Hispanic Community Health Study/Study of Latinos (HCHS/SOL)**

The Hispanic Community Health Study/Study of Latinos (HCHS/SOL) was initiated in 2006 to investigate the prevalence and risk factors affecting several health conditions, including heart, lung and blood disorders, kidney and liver function, diabetes, cognitive function, dental conditions and hearing disorders.<sup>9</sup> Participants aged 18 – 74 self-identified as Hispanic or Latino, with substantial representation of Mexican, Puerto Rican, Dominican, Cuban, Central and South American groups. They were recruited from four field centers in the United States: San Diego, CA; the Bronx, NY; Chicago, IL; and Miami, FL. 12,803 study participants consented to genetic studies and will be included in the HCHS/SOL dbGaP posting. Genotyping of the HCHS/SOL participants was performed at Illumina Microarray Services using the SOL HCHS Custom 15041502 array (annotation version “B3”, genome build 37), which includes 2,575,443 variants (of which 2,427,090 are in common with the Illumina HumanOmni2.5 and 148,353 are custom content).

### **Health and Retirement Study (HRS)**

The University of Michigan Health and Retirement Study (HRS) is a longitudinal panel study that surveys a representative sample of more than 20,000 Americans over the age of 50 every two years. Supported by the National Institute on Aging (NIA U01AG009740) and the Social Security Administration, the HRS explores the changes in labor force participation and the health transitions that individuals undergo toward the end of their work lives and in the years that follow. Since its launch in 1992, the study has collected information about income, work, assets, pension plans, health insurance, disability, physical health and functioning, cognitive functioning, and health care expenditures. HRS is intended to be a nationally representative sample with 2:1 oversampling of minority groups including African American and Hispanic/Latino populations.<sup>10</sup> In Phases I – II, 12,507 study participants with DNA samples taken in 2006 or 2008 were included in the dbGaP posting. Genotyping of the HRS Phase I – II participants was performed at CIDR using the Illumina HumanOmni2.5-4v1 array (annotation version “D”, genome build 37) and released a total of 2,443,179 variants. Only participants in the European ancestry and African ancestry analytic samples (defined by self-reported race/ethnicity and genetic principal components as described in the dbGaP documentation) were included in this analysis. Participants with self-reported stroke or self-reported heart disease were excluded.

### **Infancia y Medio Ambiente (INMA)**

The Infancia y Medio Ambiente (Environment and Childhood) project is a research project comprising a Spanish population-based birth cohort created to study the role of the environmental pollutants during pregnancy and first stages of life and their effects on childhood growth and development. The cohort was established between 2003 and 2008 from mothers enrolled in four regions within Spain and included their infants.

### **KORA**

For the German MUNICH discovery samples and the Stroke in Young Fabry Patients (SIFAP) samples, independent control groups were selected from Caucasians of German origin participating into the population-based KORA study. This survey represents a sex- and age stratified random sample of all German residents of the Augsburg area and consists of individuals 25 – 74 years of age, with about 300 subjects for each 10-year increment. All controls were free of a history of stroke or transient ischemic attack. KORA samples were genotyped on the Illumina Human 550k platform. QC was identical for all WTCCC cohorts.

### **Malmo Diet and Cancer (MDC) Study**

The MDC study is a population-based prospective cohort study. A total of 30,447 individuals, 45 to 73 years old, 60% women, attended a baseline examination between February 1991 and September 1996. Between 1992 and 1994, a total of 6,103 randomly selected subjects attended an extended baseline examination with the purpose of studying the epidemiology of cardiovascular diseases (the MDCCardiovascular cohort, MDC-CC). At the baseline

examination, 23% of the participants were smokers, 16% used anti-hypertensive medication, 14% were obese (body mass index > 30 kg/m<sup>2</sup>), 88% were born in Sweden and > 99% were born in Europe. Genotyping was performed using the Illumina Infinium Omni5 platform with exome content. Incidence of stroke was monitored prospectively from the baseline examination in 1992 to 1994 until December 31, 2008. The case-finding procedures included a broad search among patients with neurological symptoms that could indicate stroke. Stroke was defined according to the WHO criteria. By definition, patients with transient ischemic attacks are excluded. The stroke subtypes are coded according to International Classification of Diseases revision 9. Cerebral infarction (International Classification of Diseases code 434) is diagnosed when CT, MRI, or autopsy verifies the infarction in location corresponding to the focal neurology or excludes hemorrhage and nonvascular disease. The ischemic strokes were retrospectively classified into etiological subtypes by review of hospital records. A board-certified neurologist with expertise in cerebrovascular diseases and a specialized research nurse reviewed the records. The TOAST<sup>11</sup> and CCS<sup>12</sup> criteria were applied.

### **Osteoarthritis Initiative (OAI)**

The OAI is a publicly and privately funded prospective longitudinal cohort with a primary objective of identifying risk factors for incidence and progression of tibiofemoral knee osteoarthritis (OA). The OAI utilized a focused population-based recruitment to enroll 4,674 men and women between the ages of 45 – 79 years who either had radiographic symptomatic knee OA or who were without radiographic symptomatic OA in both knees but were considered high risk for OA because they had two or more known risk factors for knee OA. Subjects were recruited into the baseline phase of the OAI at multiple sites throughout the US between 2004 and 2006. All subjects were invited back for follow-up examinations to assess incidence or progression of OA annually, for up to 5 years. Phenotype data from the baseline and follow-up examinations are available for public access from the Osteoarthritis Initiative database. The Genetic Components of Knee Osteoarthritis (GeCKO) Study was initiated in 2009 as a genetic ancillary study to perform a genome-wide association study to identify genetic variants associated with radiographic osteoarthritis. This study included 4,482 individuals participating in the parent OAI study genotyped on the Illumina HumanOmni2.5M.

### **Project MinE/Population based ALS registry, The Netherlands**

Dutch controls from cohorts NL4 (IlluminaOmniExpress) and NL5 (Illumina 2.5M) were included, originally described in Rheenen, W. van et al. Genome-wide association analyses identify new risk variants and the genetic architecture of amyotrophic lateral sclerosis. *Nature Genetics* 2016; 48, 1043–1048. The controls were identified through ongoing population-based ALS registry, The Netherlands. (Huisman, M. H. B. et al. Population based epidemiology of amyotrophic lateral sclerosis using capture-recapture methodology. *Journal of Neurology, Neurosurgery & Psychiatry* 2011; 82, 1165–1170).

In brief, cases with ALS are diagnosed with probable or definite ALS according to the revised El-Escorial Criteria by neurologists specialized in motor neuron diseases. Both cases with and without a family history for ALS are included. Tertiary referral centers for ALS were University Medical Center Utrecht, Academic Medical Centre Amsterdam and Radboud University Medical

Center Nijmegen. Both patients with and without a family history for ALS are included. Control individuals are free of any neuromuscular disease and matched for age, gender and ethnicity to cases.

### **3.2 GENOME-WIDE ASSOCIATION ANALYSIS**

Since many contributing study sites had small numbers of cases or may have lacked control samples, we consolidated datasets in analysis strata (Table S1) and matched cohorts based on genotype array, population structure, and whether or not they needed controls.<sup>13</sup> We performed our association testing at the strata level and then used inverse variance meta-analysis (GWAMA) to combine results across strata using both fixed and random effects models.

Genotype data was filtered by individual missingness and SNP call rate. In addition, we removed snps that were out of Hardy-Weinberg equilibrium ( $p < 1 \times 10^{-5}$ ) in controls or if they had a minor allele frequency of less than 1%. Samples were removed if they had a sex-mismatch or had a high degree of relatedness to other samples.

Genotypes were lifted to build Hg38 if necessary and were imputed using the TOPMED imputation panel on the University of Michigan Imputation Server.

#### **Two Step Meta-Analysis Overview**

We first performed population-based meta-analysis for strata of recent African ancestry, European ancestry, and South Asian Ancestry. We required that each SNP be present in at least two analysis strata to be included in the ancestry-specific meta-analysis.

The second step was to perform a trans-ancestry meta-analysis by combining the ancestry-specific results into a single analysis. Following meta-analysis, we excluded SNPs with heterogenous effects across strata (post-meta  $i^2 > 0.5$ ) and those present in only a single stratum. For subtype-specific analysis, we also required a minor allele count (MAC)  $\geq 30$  within each stratum. Association analyses were performed using PLINK for European ancestry and SAIGE for Transethnic ancestry.

In order to compare our effect estimates in our early onset stroke GWAS, we performed a late age onset stroke analysis in cases over the age of 60 using the SiGN stroke dataset as described by Pulit et al.<sup>13</sup> Logistic regression was performed using PLINK. Covariates in the logistic regression model included sex and the first 10 principal components to account for ancestry. Strata level results were combined using GWAMA package. Results were filtered by  $MAF \geq 1\%$  and  $r^2$  imputation quality score  $\geq 0.5$ .

***Proportion of early onset stroke attributable to rs635634 in Europeans:*** We calculated the proportion of early onset cases in the population attributable to rs635634 using the formula  $(Pe \times (OR-1)) / (Pe \times (OR-1) + 1)$ , where  $Pe$  = proportion of subjects with the rs635634 risk variant and an  $OR = 1.18$ .  $Pe$  was estimated based on an allele frequency of 0.20 and proportions of the population heterozygous and homozygous for the variant calculated using Hardy-Weinberg proportions.

### 3.3 CONDITIONAL ANALYSIS AT THE ABO LOCUS

We performed a conditional analysis of stroke with other SNPs at the *ABO* locus after conditioning on associations of stroke with rs635634 and rs529565. In addition to retaining these two SNPs as independent variables in a logistic regression analysis, we also included sex and up to 8 PCs as covariates. We tested the effects of 50 additional SNPs within 50 kb of the boundaries of the *ABO* gene. We flagged SNPs as showing evidence for residual association if the p-value for association was  $< 0.01$ . We restricted this analysis to samples of European ancestry because it is the largest population group within our cohort and has a more defined haplotype structure. As in the main GWAS, we performed the conditional analysis in PLINKv1.9 for each stratum and then combined results using meta-analysis.

### 3.4 SEROLOGIC ABO BLOOD GROUPS, ABO BLOOD SUBGROUP-DETERMINING SNPS, AND STROKE

#### 3.4.1 Distribution of ABO blood group among early onset stroke, late onset stroke, and controls

We defined serologic blood groups based on two SNPs rs8176746 and rs8176719, as described in Groot et al.<sup>15</sup> and then compared the distribution of ABO blood types among early and late onset stroke cases and controls. We initially tested for differences in the distribution of ABO blood groups between early vs late onset cases, early cases vs controls, and late cases vs controls using a 3 df chi-square omnibus (4 blood groups by 2 comparison groups). If an overall difference was observed between groups, we then performed pairwise tests to identify which blood groups differed significantly between the two comparison groups (e.g., distribution of blood group A vs non-A between early and late).

#### Genetically Determined Serologic Blood Group Coding Scheme (from Groot, 2020)

rsID	Alleles							
rs8176719	TC_TC(/T)	0	0	1	2	2	1	2
rs8176746	G:T_T(/G)	0	1	0	0	2	1	1
	<b>Serologic Blood Group Assignment</b>	<b>O</b>	<b>O</b>	<b>A</b>	<b>A</b>	<b>B</b>	<b>B</b>	<b>AB</b>

#### 3.4.2 Association of ABO haplotype-defining SNPs with stroke

Five SNPs tag define the 5 common ABO haplotypes as defined in Table below. We performed a genetic association analysis to determine associations of each blood group haplotype with early and late onset stroke. Analyses were performed within each analysis stratum, adjusting for age, sex and x PCs and results combined in meta-analysis.

Table: 5 haplotypes that tag the major ABO blood groups\*

ABO Blood group	rs2519093	rs8176719	rs1053878	rs8176743	rs41302905	Hapotype freq	Effect on VTE
O1	C	delG	G	C	C	0.63	ref
O2	C	G	G	C	T	0.02	↓
A1	T	G	G	C	C	0.2	↑↑↑
A2	C	G	A	C	C	0.07	↑
B	C	G	G	T	C	0.08	↑↑↑

\* adapted from Table 2 in Goumidi et al. (Blood 137:2394-2400, 2021)<sup>14</sup>

### **3.5 ASSOCIATION OF ABO SNPS WITH EARLY AND LATE ONSET STROKE IN UKB**

Ischemic stroke cases were extracted by ICD code algorithm previously published in the “Definitions of Stroke for UK Biobank Phase 1 Outcomes Adjudication” (see <https://biobank.ndph.ox.ac.uk/showcase/label.cgi?id=43>) and validated by stroke physician review of electronic patient records.<sup>16</sup> Controls were matched to cases on age, sex, and ancestry, with up to 20 controls chosen per case. Any individual related to an already selected individual was removed. Age at Ischemic Stroke diagnosis was determined at either the main or secondary diagnosis, whichever was younger. We defined early onset Ischemic Stroke cases as subjects with age at diagnosis before age 60 and late onset Ischemic Stroke cases as those with age at diagnosis at age 60 or older. Our analysis included 1,803 early onset cases and 4,071 late onset cases.

PLINK was used to perform logistic regression using age, sex and 5 principal components for ancestry.

#### ICD10 Codes used to define Ischemic Stroke

ICD 10 Code I63 Cerebral infarction
ICD 10 Code I63.0 Cerebral infarction due to thrombosis of precerebral arteries
ICD 10 Code I63.1 Cerebral infarction due to embolism of precerebral arteries
ICD 10 Code I63.2 Cerebral infarction due to unspecified occlusion or stenosis of precerebral arteries
ICD 10 Code I63.3 Cerebral infarction due to thrombosis of cerebral arteries
ICD 10 Code I63.4 Cerebral infarction due to embolism of cerebral arteries
ICD 10 Code I63.5 Cerebral infarction due to unspecified occlusion or stenosis of cerebral arteries
ICD 10 Code I63.6 Cerebral infarction due to cerebral venous thrombosis, nonpyogenic
ICD 10 Code I63.8 Other cerebral infarction
ICD 10 Code I63.9 Cerebral infarction, unspecified
ICD 10 Code I64 Stroke, not specified as haemorrhage or infarction

### **3.6 ASSOCIATION OF ABO SNPS WITH EARLY AND LATE ONSET VTE IN UKB**

Individuals were defined as a VTE case based on at least one of the following ICD-10 codes being recorded in their medical record as either a main or secondary diagnosis:

1. Hospitalization for ICD-10 Code I80.1 or I80.2 - phlebitis and thrombophlebitis of the femoral vein or other deep vessels of lower extremities;
2. Hospitalization for ICD-10 Code I82.2 - embolism and thrombosis of vena cava;
3. Hospitalization for ICD-10 Code I26.0 or I26.9 - pulmonary embolism with or without acute cor pulmonale;

Controls were selected after exclusion of related codes. Controls were matched to cases on age, sex, and ancestry, with up to 5 controls chosen per case. Any individual related to an already selected individual was removed. Age at VTE diagnosis was determined at either the main or secondary diagnosis, whichever was younger. We defined early onset VTE cases as subjects with age at diagnosis before age 60 and late onset VTE cases as those with age at diagnosis at age 60 or older. Our analysis included 3,519 early onset cases and 5,046 late onset cases.

### **3.7 ASSOCIATION OF VTE PRS WITH EARLY AND LATE ONSET STROKE**

Additive, effect weighted polygenic risk scores were calculated for VTE based on a published list of 297 VTE risk alleles identified through a large primary GWAS (Klarin et al, 2019).<sup>17</sup> This prior VTE PRS was constructed using pruning and thresholding criteria of  $R^2 < 0.2$  and  $P < 1 \times 10^{-5}$ ) and obtained from European MVP v.2.1 and UK Biobank European VTE meta-analyzed summary statistics. Of these 297 SNPs, 255 were available across all strata of our early and late onset GWAS as used for analysis. We tested the association between the VTE PRS score with stroke in the European ancestry sample using logistic regression with 10 PCs and sex included as covariates. Scores were generated using PRSice<sup>18</sup> and the -score avg option. Association analysis with stroke was performed in RStudio. We computed odds ratios corresponding to the associations of VTE PRS on early and late onset stroke and tested for homogeneity of the odds ratios using the Wald test.

### **3.8 CO-LOCATION OF ASSOCIATIONS AT THE ABO LOCUS BETWEEN EOS AND OTHER THROMBOTIC-RELATED DISORDERS AND BIOMARKERS**

We performed colocalization analyses to evaluate evidence for a shared common variant at the ABO locus between early onset stroke and late onset stroke and between early onset stroke and 3 prothrombotic traits (VTE and plasma levels of von Willebrand factor (VWF) and Factor VIII (FVIII)). For this analysis, we considered all SNPs with 50 kb of the gene boundaries. Analyses were performed using summary genetic association results with the coloc software program.<sup>19</sup> Summary statistics for GWAS results were obtained from the sources indicated in

the table below. Briefly, coloc utilizes a Bayesian approach and summary level association level results for two traits to calculate the posterior probabilities of five competing models that assess whether the associations are due to the same or a different causal variant. The competing models tested are:

$H_0$ : no causal SNP for either trait

$H_1$ : causal SNP for Trait 1, not Trait 2

$H_2$ : causal SNP for Trait 2, not Trait 1

$H_3$ : different causal SNPs for Traits 1 and 2

$H_4$ : shared causal SNP for Traits 1 and 2

For each set of traits, we considered a posterior probability >80% for  $H_4$  as indicating strong evidence for a shared causal variant for both traits.

Disorder/Trait	Source of summary statistics	Sample size	Source/Publication
Late onset stroke	GWAS of late onset stroke from SiGN	9,272 cases 25,124 controls	Described in this manuscript
VTE	UKB	4,620 cases 356,574 controls	Internal analysis; ICD codes
Plasma VWF levels	CHARGE Consortium	46,354 subjects	Sabater-Lleal M, Huffman JE, de Vries PS, et al. Genome-Wide Association Transehtnic Meta-Analyses Identifies Novel Associations Regulating Coagulation Factor VIII and von Willebrand Factor Plasma Levels. Circulation. 2019;139(5):620-635.
Plasma Factor VIII levels	CHARGE Consortium	46,354 subjects	Sabater-Lleal M, et al., Circulation. 2019;139(5):620-635.



## **eAppendix 1: Study Funding and Acknowledgements for Contributing Studies**

### **Case-only & Case and Control Cohort:**

**ASGC:** Australian population control data were derived from the Hunter Community Study. We also thank the University of Newcastle for funding and the men and women of the Hunter region who participated in this study. This research was funded by grants from the Australian National and Medical Health Research Council (NHMRC Project Grant ID: 569257), the Australian National Heart Foundation (NHF Project Grant ID: G 04S 1623), the University of Newcastle, the Gladys M Brawn Fellowship scheme, and the Vincent Fairfax Family Foundation in Australia. Elizabeth G Holliday was supported by a Fellowship from the National Heart Foundation and National Stroke Foundation of Australia (ID: 100071).

**BASICMAR:** The Base de Datos de Ictus del Hospital del Mar (BASICMAR) Genetic Study was supported by the Ministerio de Sanidad y Consumo de España, Instituto de Salud Carlos III (ISC III) with the grants: Registro BASICMAR Funding for Research in Health (PI051737); GWA Study of Leukoaraiosis (GWALA) project from Fondos de Investigación Sanitaria ISC III (PI10/02064) and (PI12/01238); Agència de Gestió Ajuts Universitaris de Recerca (2014 SGR 1213) and Fondos European Regional Development Funding (FEDER/EDRF) Red INVICTUS-PLUS (RD16/0019/0002). Additional support was provided by the Fundació la Marató TV3 with the grant GODS project. Genestroke Consortium (76/C/2011) Recercaixa'13 (JJ086116). Assistance with data cleaning was provided by the Research in Cardiovascular and Inflammatory Diseases Program of Institute Hospital del Mar of Medical Investigations, Hospital del Mar, and the Barcelona Biomedical Research Park.

**Biobank Japan:** We extend our appreciation to the staff of BBJ for their outstanding assistance. This study was funded by the BioBank Japan project, which is supported by the Ministry of Education, Culture, Sports, Sciences and Technology (MEXT) of Japanese government and the Japan Agency for Medical Research and Development (AMED).

**BRAINS:** Biorepository of DNA in Stroke (BRAINS) is partly funded by a Senior Fellowship from the Department of Health (UK) to P Sharma, the Henry Smith Charity and the UK-India Education Research Institute (UKIERI) from the British Council.

**CADISP:** The Cervical Artery Dissections and Ischemic Stroke Patients (CADISP) study has been supported by Inserm, Lille 2 University, Institut Pasteur de Lille and Lille University Hospital and received funding from the ERDF (FEDER funds) and Région Nord-Pas de Calais in the frame of Contrat de Projets Etat-Region 2007-2013 Région Nord-Pas-de-Calais - Grant N°09120030, Centre National de Genotypage, Emil Aaltonen Foundation, Paavo Ilmari Ahvenainen Foundation, Helsinki University Central Hospital Research Fund, Helsinki University Medical Foundation, Päivikki and Sakari Sohlberg Foundation, Aarne Koskelo Foundation, Maire Taponen Foundation, Aarne and Aili Turunen Foundation, Lilly Foundation, Alfred Kordelin Foundation, Finnish Medical Foundation, Orion Farnos Research Foundation, Maud Kuistila Foundation, the Finnish Brain Foundation, Biomedicum Helsinki Foundation, Projet Hospitalier de Recherche Clinique Régional, Fondation de France, Génopôle de Lille, Adrinord, Basel Stroke-Funds, Käthe-Zingg-Schwichtenberg-Fonds of the Swiss Academy of Medical Sciences, Swiss Heart Foundation.

**China Kadoorie Biobank:** CKB acknowledges the contribution of participants, project staff, and the China National Centre for Disease Control and Prevention (CDC) and its regional offices. China Kadoorie Biobank was supported as follows: Baseline survey and first re-survey: Hong Kong Kadoorie Charitable Foundation; long-term follow-up and second re-survey: UK Wellcome Trust (212946/Z/18/Z, 202922/Z/16/Z, 104085/Z/14/Z, 088158/Z/09/Z), National Natural Science Foundation of China (91843302), and National Key Research and Development Program of China (2016YFC 0900500, 0900501, 0900504, 1303904). DNA extraction and genotyping: GlaxoSmithKline, UK Medical Research Council (MC\_PC\_13049, MC-PC-14135). The UK Medical Research Council (MC\_UU\_00017/1, MC\_UU\_12026/2 MC\_U137686851), Cancer Research UK (C16077/A29186; C500/A16896), and the British Heart Foundation (CH/1996001/9454) provide core funding to the Clinical Trial Service Unit and Epidemiological Studies Unit at Oxford University for the project.

**Edinburgh:** The Edinburgh Stroke Study was supported by the Wellcome Trust and the Binks Trust. Sample processing occurred in the Genetics Core Laboratory of the Wellcome Trust Clinical Research Facility, Western General Hospital, Edinburgh, UK. Dr. Rannikmäe was funded from HDR UK fellowship MR/S004130/1

**EPIC-CVD:** This work was supported by core funding from the: UK Medical Research Council (MR/L003120/1), British Heart Foundation (RG/13/13/30194; RG/18/13/33946), BHF Centre for Research Excellence (RE/18/1/34212) and NIHR Cambridge Biomedical Research Centre (BRC-1215-20014) [\*]. EPIC-CVD was also funded by the European Research Council (268834) and the European Commission Framework Programme 7 (HEALTH-F2-2012-279233).

\*The views expressed are those of the author(s) and not necessarily those of the NIHR or the Department of Health and Social Care.

The coordination of EPIC is financially supported by International Agency for Research on Cancer (IARC) and also by the Department of Epidemiology and Biostatistics, School of Public Health, Imperial College London which has additional infrastructure support provided by the NIHR Imperial Biomedical Research Centre (BRC). The national cohorts are supported by: Danish Cancer Society (Denmark); Ligue Contre le Cancer, Institut Gustave Roussy, Mutuelle Générale de l'Education Nationale, Institut National de la Santé et de la Recherche Médicale (INSERM) (France); German Cancer Aid, German Cancer Research Center (DKFZ), German Institute of Human Nutrition PotsdamRehbruecke (DIfE), Federal Ministry of Education and Research (BMBF) (Germany); Associazione Italiana per la Ricerca sul Cancro-AIRC-Italy, Compagnia di SanPaolo and National Research Council (Italy); Dutch Ministry of Public Health, Welfare and Sports (VWS), Netherlands Cancer Registry (NKR), LK Research Funds, Dutch Prevention Funds, Dutch ZON (Zorg Onderzoek Nederland), World Cancer Research Fund (WCRF), Statistics Netherlands (The Netherlands); Health Research Fund (FIS) - Instituto de Salud Carlos III (ISCIII), Regional Governments of Andalucía, Asturias, Basque Country, Murcia and Navarra, and the Catalan Institute of Oncology - ICO (Spain); Swedish Cancer Society, Swedish Research Council and County Councils of Skåne and Västerbotten (Sweden); Cancer Research UK (14136 to EPIC-Norfolk; C8221/A29017 to EPIC-Oxford), Medical Research Council (1000143 to EPIC-Norfolk; MR/M012190/1 to EPIC-Oxford). (United Kingdom).

We thank all EPIC participants and staff for their contribution to the study, the laboratory teams at the Medical Research Council Epidemiology Unit for sample management and Cambridge

Genomic Services for genotyping, Sarah Spackman for data management, and the team at the EPIC-CVD Coordinating Centre for study coordination and administration.

**FUTURE/Odyssey:** F-E d L received a grant from the Dutch Heart Foundation (grant 2014 T060 to FE.d.L) and Bike4Brains. AMT received a grant from the Dutch Heart Foundation (grant 2016 T044 to A.M.T) and from the Netherlands CardioVascular Research Initiative: the Dutch Heart Foundation (CVON 2018-28 & 2012-06 Heart Brain Connection to A.M.T)

**MGH-GASROS:** MGH Genes Affecting Stroke Risk and Outcome Study (MGH-GASROS) was supported by NINDS (U01 NS069208), the American Heart Association/Bugher Foundation Centers for Stroke Prevention Research 0775010N, the NIH and NHLBI's STAMPEED genomics research program (R01 HL087676), and a grant from the National Center for Research Resources. The Broad Institute Center for Genotyping and Analysis is supported by grant U54 RR020278 from the National Center for Research resources.

**GCNKSS:** Greater Cincinnati/Northern Kentucky Stroke Study (GCNKSS) was supported by the NIH (NS030678).

**Geisinger:** Dr Abedi has research support from the Defense Threat Reduction Agency (DTRA) grant No. HDTRA1-18-1-0008 and from National Institutes of Health grant No. R56HL116832. Dr Zand has research support from Bucknell University Initiative Program, Roche—Genentech Biotechnology Company, and the Geisinger Health Plan Quality fund.

**GEOS:** Genetics of Early Onset Stroke (GEOS) Study, Baltimore, USA was supported by the NIH Genes, Environment and Health Initiative (GEI) Grant U01 HG004436, as part of the GENEVA consortium under GEI, with additional support provided by the Mid-Atlantic Nutrition and Obesity Research Center (P30 DK072488), and the Office of Research and Development, Medical Research Service, and the Baltimore Geriatrics Research, Education, and Clinical Center of the Department of Veterans Affairs. Genotyping services were provided by the Johns Hopkins University Center for Inherited Disease Research (CIDR), which is fully funded through a federal contract from the NIH to the Johns Hopkins University (contract number HHSN268200782096C). Assistance with data cleaning was provided by the GENEVA Coordinating Center (U01 HG 004446; PI Bruce S Weir). Study recruitment and assembly of datasets were supported by a Cooperative Agreement with the Division of Adult and Community Health, Centers for Disease Control and Prevention and by grants from NINDS and the NIH Office of Research on Women's Health (R01 NS45012, U01 NS069208- 01).

**GRAZ:** The Austrian Stroke Prevention Study was supported by the Austrian Science Fund (FWF) grant Nos. P20545-P05 and P13180 and I904-B13 (Era-Net). The Medical University of Graz supports the databases of the Graz Stroke Study and the Austrian Stroke Prevention Study.

**Helsinki:** The study was supported by the Finnish Medical Foundation, Sigrid Juselius Foundation, and the Helsinki University Central Hospital governmental subsidiary funds for clinical research. The investigators thank Marja Metso, RN for her support.

**INTERSTROKE:** The INTERSTROKE study was supported by the Canadian Institutes of Health Research, Heart and Stroke Foundation of Canada, Canadian Stroke Network, Health Research Board Ireland, Swedish Research Council, Swedish Heart and Lung Foundation, The Health & Medical Care Committee of the Regional Executive Board, Region Västra Götaland (Sweden),

AstraZeneca, Boehringer Ingelheim (Canada), Pfizer (Canada), MSD, Chest, Heart and Stroke Scotland, and The Stroke Association, with support from The UK Stroke Research Network. Microarray genotyping for a subset of INTERSTROKE participants was funded by the Heart and Stroke Foundation of Canada Grant NA-6872 (PI: Drs Pare, Anand, O'Donnell, Xie, Yusuf).

**INVENT:** David-Alexandre Trégouët is supported by the «EPIDEMIOM-VTE» Senior Chair from the Initiative of Excellence of the University of Bordeaux.

**ISGS:** The Ischemic Stroke Genetics Study (ISGS) was supported by the NINDS (R01 NS42733; PI Dr Meschia). Both SWISS and ISGS received additional support, in part, from the Intramural Research Program of the National Institute on Aging (Z01 AG000954-06; PI Andrew Singleton). SWISS and ISGS used samples and clinical data from the NIH-NINDS Human Genetics Resource Center DNA and Cell Line Repository (<http://ccr.coriell.org/ninds>), human subject protocol Nos. 2003-081 and 2004-147. SWISS and ISGS used stroke-free participants from the Baltimore Longitudinal Study of Aging (BLSA) as controls with the permission of Dr Luigi Ferrucci. The inclusion of BLSA samples was supported, in part, by the Intramural Research Program of the National Institute on Aging (Z01 AG000015-50), human subject protocol No. 2003-078. This study used the high-performance computational capabilities of the Biowulf Linux cluster at the NIH (<http://biowulf.nih.gov>). For SWISS and ISGS cases of African ancestry, a subset of the Healthy Aging in Neighborhoods of Diversity across the Life Span study (HANDLS) were used as stroke-free controls. HANDLS is funded by the National Institute of Aging (1Z01AG000513; PI Michele K. Evans).

**Krakow:** Phenotypic data and genetic specimens collection were funded by the grant from the Polish Ministry of Science and Higher Education for Leading National Research Centers (KNOW) and by the grants from the Jagiellonian University Medical College in Krakow, Poland: K/ZDS/002848, K/ZDS/003844.

**Leuven:** The Leuven Stroke genetics study was supported by personal research funds from the Department of Neurology of the University Hospitals Leuven. Dr. Lemmens is a Senior Clinical Investigator of FWO Flanders (FWO 1841918N).

**Lund:** Supported by the Swedish Research Council (2019-01757), The Swedish Heart-Lung Foundation, Region Skåne, Skåne University Hospital, the Freemasons Lodge of Instruction Eos in Lund, King Gustaf V's and Queen Victoria's Foundation, Lund University, CaNVAS NIH (1R01NS114045-01), and The Swedish Government (under the "Avtal om Läkarutbildning och Medicinsk Forskning, ALF"). Biobank services were provided by Region Skåne Competence Centre (RSKC Malmö), Skåne University Hospital, Malmö, Sweden, and Biobank, Labmedicin Skåne, University and Regional Laboratories Region Skåne, Sweden.

**MCISS:** The Middlesex County Ischemic Stroke Study (MCISS) was supported by intramural funding from the New Jersey Neuroscience Institute/JFK Medical Center, Edison, NJ, and The Neurogenetics Foundation, Cranbury, NJ. We acknowledge Dr Souvik Sen for his advice and encouragement in the initiation and design of this study.

**MIAMISR:** The Cerebrovascular Biorepository at University of Miami/Jackson Memorial Hospital (The Miami Stroke Registry, Institutional Review Board No. 20070386) was supported by the Department of Neurology at University of Miami Miller School of Medicine and Evelyn McKnight Brain Institute. Biorepository and DNA extraction services were provided by the Hussmann Institute for Human Genomics at the Miller School of Medicine.

**Milano:** Milano - Besta Stroke Register Collection and genotyping of the Milan cases within CEDIR were supported by the Italian Ministry of Health (Grant Numbers: RC 2007/LR6, RC 2008/LR6; RC 2009/LR8; RC 2010/LR8; GR-2011-02347041). FP6 LSHM-CT-2007-037273 for the PROCARDIS control samples.

**Munich:** This project received funding from the European Union's Horizon 2020 research and innovation programme (666881), SVDs@target (to MD; 667375)); the DFG as part of the Munich Cluster for Systems Neurology (EXC 2145 SyNergy – ID 390857198), the CRC 1123 (B3; to MD), DI 722/16-1 (project ID: 428668490), DI 722/13-1; and the Fondation Leducq (Transatlantic Network of Excellence on the Pathogenesis of Small Vessel Disease of the Brain; to MD).

Institute for Stroke and Dementia Research (ISD), University Hospital, LMU Munich, Munich, Germany. Munich Cluster for Systems Neurology (SyNergy), Munich, Germany. German Center for Neurodegenerative Diseases (DZNE), Munich, Germany.

**NHS:** The Nurses' Health Study work on stroke is supported by grants from the NIH, including HL088521 and HL34594 from the National Heart, Lung, and Blood Institute, as well as grants from the National Cancer Institute funding the questionnaire follow-up and blood collection: CA87969 and CA49449.

**NOMAS:** The Northern Manhattan Study (NOMAS) was supported by grants from the NINDS (R01 NS029993, R01 NS27517). The Cerebrovascular Biorepository at University of Miami/Jackson Memorial Hospital (The Miami Stroke Registry, Institutional Review Board No. 20070386) was supported by the Department of Neurology at University of Miami Miller School of Medicine and Evelyn McKnight Brain Institute. Biorepository and DNA extraction services were provided by the Hussmann Institute for Human Genomics at the Miller School of Medicine.

**OXVASC:** The Oxford Vascular Study was supported by the Wellcome Trust, Wolfson Foundation, Stroke Association, Medical Research Council, Dunhill Medical Trust, NIH Research (NIHR), and NIHR Oxford Biomedical Research Centre based at Oxford University Hospitals NHS Trust and University of Oxford. Dr Rothwell is in receipt of Senior Investigator Awards from the Wellcome Trust and the NIHR.

**RACE:** We are thankful to the RACE study participants. Fieldwork in RACE was funded by the R-21 grant provided by the NINDS and the Fogarty International Center (1R21NS064908-01) and educational grants available to Dr. Saleheen at the Center for Non-Communicable Diseases, Pakistan. We would also like to acknowledge the contributions made by Professor John Danesh, Dr. Ayesha Kamal and Professor Panos Deloukas.

**REGARDS:** The Reasons for Geographic and Racial Differences in Stroke (REGARDS) Study was supported by a cooperative agreement U01 NS041588 from the NINDS, NIH, and Department of Health and Human Service. A full list of participating REGARDS investigators and institutions can be found at <http://www.regardsstudy.org>.

**SAHLSIS:** The Sahlgrenska Academy Study on Ischemic Stroke was supported by the Swedish Research Council (2018-02543), the Swedish Heart and Lung Foundation (20190203), the Swedish state under the agreement between the Swedish government and the county councils, the ALF-agreement (ALFGBG-720081).

**SIFAP:** The SIFAP study (Stroke in Young Fabry Patients, <http://www.sifap.eu>; ClinicalTrials.gov: NCT00414583) has been supported partially by an unrestricted scientific grant from Shire Human Genetic Therapies. Funding for genotyping and analysis of samples were supported by the National Institutes of Health Genes, Environment and Health Initiative (GEI) Grant U01 HG004436, as part of the GENEVA consortium.

**SLESS:** This work was supported by a Stroke Association (UK) Programme Grant (PROG 3), the National Institute for Health Research Biomedical Research Centre (NIHR BRC) at South London and Maudsley NHS Foundation Trust and the National Institute for Health Research (NIHR) Biomedical Research Centre based at Guy's and St Thomas' NHS Foundation Trust and King's College London.

S.Bell and H.S.M. are funded by the British Heart Foundation (RG/16/4/32218). H.S.M. is also supported by an NIHR Senior Investigator award. This research was supported by the NIHR Cambridge Biomedical Research Centre (BRC-1215-20014). The views expressed are those of the author(s) and not necessarily those of the NIHR or the Department of Health and Social Care

**SPS3:** The Secondary Prevention of Small Subcortical Strokes trial was funded by the US National Institute of Health and Neurological Disorders and Stroke grant No. U01NS38529-04A1 (principal investigator, Oscar R. Benavente; coprincipal investigator, Robert G. Hart). The SPS3 Genetic Substudy (SPS3-GENES) was funded by R01 NS073346 (coprincipal investigators, Julie A. Johnson, Oscar R. Benavente, and Alan R. Shuldiner) and U01 GM074492-05S109 (principal investigator, Julie A. Johnson).

**St. George:** The principal funding for this study was provided by the Wellcome Trust, as part of the Wellcome Trust Case Control Consortium 2 project (085475/B/08/Z and 085475/Z/08/Z and WT084724MA). Collection of some of the St George's stroke cohort was supported by project grant support from the Stroke Association.

**SWISS:** The Sibling with Ischemic Stroke Study (SWISS) was supported by the NINDS (R01 NS39987; PI Dr Meschia). Both SWISS and ISGS received additional support, in part, from the Intramural Research Program of the National Institute on Aging (Z01 AG000954-06; PI Andrew Singleton). SWISS and ISGS used samples and clinical data from the NIH-NINDS Human Genetics Resource Center DNA and Cell Line Repository (<http://ccr.coriell.org/ninds>), human subject protocol Nos. 2003-081 and 2004-147. SWISS and ISGS used stroke-free participants from the Baltimore Longitudinal Study of Aging (BLSA) as controls with the permission of Dr Luigi Ferrucci. The inclusion of BLSA samples was supported, in part, by the Intramural Research Program of the National Institute on Aging (Z01 AG000015-50), human subject protocol No. 2003-078. This study used the high-performance computational capabilities of the Biowulf Linux cluster at the NIH (<http://biowulf.nih.gov>). For SWISS and ISGS cases of African ancestry, a subset of the Healthy Aging in Neighborhoods of Diversity across the Life Span study (HANDLS) were used as stroke-free controls. HANDLS is funded by the National Institute of Aging (1Z01AG000513; PI Michele K. Evans).

**UK Biobank:** UK Biobank has received funding from the UK Medical Research Council, Wellcome Trust, Department of Health, British Heart Foundation, Diabetes UK, Northwest Regional Development Agency, Scottish Government, and Welsh Assembly Government.

**VHIR-FMT-Barcelona:** The Barcelona GWAs Study was supported by the Genetic contribution to functional Outcome and Disability after Stroke (GODS) project and EPIGENESIS project Fundació la Marató de TV3, GENERACION Project (Instituto de Salud Carlos III: PI15/01978), Maestro Project (Instituto de Salud Carlos III, Fondo Europeo de Desarrollo Regional (FEDER)).) and by the Miguel Servet grant (Pharmastroke project: CP12/03298). I. F-C. is supported by the Miguel Servet programme (CP12/03298), Instituto de Salud Carlos III.

**VISP:** The GWAS component of the VISP study was supported by the United States National Human Genome Research Institute (NHGRI), Grant U01 HG005160 (PI Michèle Sale & Bradford Worrall), as part of the Genomics and Randomized Trials Network (GARNET). Genotyping services were provided by the Johns Hopkins University Center for Inherited Disease Research (CIDR), which is fully funded through a federal contract from the NIH to the Johns Hopkins University. Assistance with data cleaning was provided by the GARNET Coordinating Center (U01 HG005157; PI Bruce S Weir). Study recruitment and collection of datasets for the VISP clinical trial were supported by an investigator-initiated research grant (R01 NS34447; PI James Toole) from the United States Public Health Service, NINDS, Bethesda, Maryland. Control data for comparison with European ancestry VISP stroke cases were obtained through the database of genotypes and phenotypes (dbGAP) High Density SNP Association Analysis of Melanoma: Case-Control and Outcomes Investigation (phs000187.v1.p1; R01CA100264, 3P50CA093459, 5P50CA097007, 5R01ES011740, 5R01CA133996, HHSN268200782096C; PIs Christopher Amos, Qingyi Wei, Jeffrey E. Lee). For VISP stroke cases of African ancestry, a subset of stroke-free subjects from the Healthy Aging in Neighborhoods of Diversity across the Life Span study (HANDLS) were used as stroke free controls. HANDLS is funded by the National Institute of Aging (1Z01AG000513; PI Michele K. Evans).

Vitamin Intervention for Stroke Prevention (VISP) was funded by the National Institute of Neurological Disorders and Stroke (R01-NS34447). Genome-wide association study data for a subset of VISP participants supported by the National Human Genome Research Institute (U01-HG005160), as part of the Genomics and Randomized Trials Network (PI: Drs Sale and Worrall).

**WHI-OS:** The Women's Health Initiatives (WHI) program was funded by the National Heart, Lung, and Blood Institute, NIH, US Department of Health and Human Services through contracts N01WH22110, 24152, 32100-2, 32105-6, 32108- 9, 32111-13, 32115, 32118 to 32119, 32122, 42107-26, 42129-32, and 44221. The Hormones and Biomarkers Predicting Stroke (HaBPS) was supported by a grant from the National Institutes of Neurological Disorders and Stroke (R01NS042618).

**WUSTL:** Washington University St. Louis Stroke Study (WUSTL): The collection, extraction of DNA from blood, and storage of specimens were supported by 2 NINDS NIH grants (P50 NS055977 and R01 NS085419. Basic demographic and clinical characterization of stroke phenotype was prospectively collected in the Cognitive Rehabilitation and Recovery Group (CRRG) registry. The Recovery Genomics after Ischemic Stroke (ReGenesIS) study was supported by a grant from the Barnes-Jewish Hospital Foundation.

**Young Lacunar Stroke DNA Resource Plus:** Collection of the UK Young Lacunar Stroke DNA Study (DNA Lacunar) was primarily supported by the Wellcome Trust (WT072952) with

additional support from the Stroke Association (TSA 2010/01). Genotyping of the samples, and Dr Traylor, were supported by a Stroke Association Grant (TSA 2013/01).

### **Control-only Cohorts:**

**ADHD:** Financial support was received from the “Instituto de Salud Carlos III-FIS”, grants PI18/01788, PI19/00721, P19/01224 and PI20/00041, cofinanced by the European Regional Development Fund (ERDF), “Agència de Gestió d’Ajuts Universitaris i de Recerca-AGAUR, Generalitat de Catalunya” (2017SGR1461) and “Departament de Salut”, Government of Catalonia, Spain. Authors wish to thank all participants who kindly participated in this research.

**FINRISK:** Veikko Salomaa was supported by the Finnish Foundation for Cardiovascular Research. Disclosure: Veikko Salomaa has received an honorarium from Sanofi for consulting. He also has ongoing research collaboration with Bayer Ltd. (All unrelated to the present study).

**HCHS/SOL:** The Hispanic Community Health Study/Study of Latinos was carried out as a collaborative study supported by contracts from the National Heart, Lung, and Blood Institute (NHLBI) to the University of North Carolina (N01-HC65233), University of Miami (N01-HC65234), Albert Einstein College of Medicine (N01-HC65235), Northwestern University (N01-HC65236), and San Diego State University (N01-HC65237). The following Institutes/Centers/Offices contribute to the HCHS/SOL through a transfer of funds to the NHLBI: National Center on Minority Health and Health Disparities, the National Institute of Deafness and Other Communications Disorders, the National Institute of Dental and Craniofacial Research, the National Institute of Diabetes and Digestive and Kidney Diseases, the National Institute of Neurological Disorders and Stroke, and the Office of Dietary Supplements

**HRS:** HRS is supported by the National Institute on Aging (NIA U01AG009740). The genotyping was funded as a separate award from the National Institute on Aging (RC2 AG036495). Genotyping was conducted by the NIH Center for Inherited Disease Research (CIDR) at Johns Hopkins University. Genotyping quality control and final preparation of the data were performed by the Genetics Coordinating Center at the University of Washington. HRS genotype data have been deposited in the NIH GWAS repository (dbGaP; accession number: phs000428.v2.p2).

**INMA:** This study was funded by grants from Instituto de Salud Carlos III (CB06/02/0041, G03/176, FIS PI041436, PI081151, PI041705, PI061756, PI091958, and PS09/00432, FIS-FEDER 03/1615, 04/1509, 04/1112, 04/1931, 05/1079, 05/1052, 06/1213, 07/0314, 09/02647, 11/01007, 11/02591, 11/02038, 13/1944, 13/2032 and CP11/0178), Spanish Ministry of Science and Innovation (SAF2008- 00357), European Commission (ENGAGE project and grant agreement HEALTH-F4- 2007-201413, HEALTH.2010.2.4.5-1, FP7-ENV-2011 cod 282957), Fundació La Marató de TV3, Generalitat de Catalunya-CIRIT 1999SGR 00241 and Conselleria de Sanitat Generalitat Valenciana. Part of the DNA extractions and genotyping was performed at the Spanish National Genotyping Centre (CEGEN-Barcelona). The authors are grateful to Silvia Fochs, Anna Sánchez, Maribel López, Nuria Pey, Muriel Ferrer, Amparo Quiles, Sandra Pérez, Gemma León, Elena Romero, Maria Andreu, Nati Galiana, Maria Dolores Climent, Amparo Cases and Cristina Capo for their 165 assistance in contacting the families and administering the questionnaires. The authors would particularly like to thank all the participants for their generous collaboration. A full roster of the INMA Project Investigators can be found at <https://www.proyectoinma.org/en/inma-project/inma-project-researchers/>



**KORA:** The KORA research platform (KORA, Cooperative Research in the Region of Augsburg) was initiated and financed by the Helmholtz Zentrum München - German Research Center for Environmental Health, which is funded by the German Federal Ministry of Education and Research and by the State of Bavaria. Furthermore, KORA research was supported within the Munich Center of Health Sciences (MC Health), Ludwig-Maximilians-Universität, as part of LMUinnovativ. Funded by the Bavarian State Ministry of Health and Care through the research project DigiMed Bayern ([www.digimed-bayern.de](http://www.digimed-bayern.de)).

**MDC:** The Malmö Diet and Cancer Study was supported by the Swedish Research Council (Vetenskapsrådet), Heart and Lung Foundation (Hjärt och Lungfonden), and Swedish Stroke Foundation (Strokeförbundet).

**OAI:** The OAI is a public–private partnership comprised of five contracts (N01-AR-2-2258; N01-AR2-2259; N01-AR-2-2260; N01-AR-2-2261; N01-AR-2-2262) funded by the National Institutes of Health, a branch of the Department of Health and Human Services and conducted by the OAI Study Investigators. Genotyping support was provided by grant RC2-AR-058950 from NIAMS/NIH. Private funding partners include Merck Research Laboratories; Novartis Pharmaceuticals Corporation, GlaxoSmithKline; and Pfizer, Inc. Private sector funding for the OAI is managed by the Foundation for the National Institutes of Health.

**Project MinE/Population based ALS registry, The Netherlands:** The collaboration project is co-funded by the PPP Allowance made available by Health~Holland, Top Sector Life Sciences & Health, to stimulate public-private partnerships. This study was supported by the ALS Foundation Netherlands.

## eReferences

- e1. Ay H, Benner T, Arsava EM, et al. A computerized algorithm for etiologic classification of ischemic stroke: the Causative Classification of Stroke System. *Stroke*. 2007;38(11):2979-2984. doi:10.1161/STROKEAHA.107.490896
- e2. Kasner SE. Clinical interpretation and use of stroke scales. *Lancet Neurol*. 2006;5(7):603-612. doi:10.1016/S1474-4422(06)70495-1
- e3. Johnson CJ, Kittner SJ, McCarter RJ, et al. Interrater reliability of an etiologic classification of ischemic stroke. *Stroke*. 1995;26(1):46-51. doi:10.1161/01.str.26.1.46
- e4. Kittner SJ, Stern BJ, Wozniak M, et al. Cerebral infarction in young adults: the Baltimore-Washington Cooperative Young Stroke Study. *Neurology*. 1998;50(4):890-894. doi:10.1212/wnl.50.4.890
- e5. Hatano S. Experience from a multicentre stroke register: a preliminary report. *Bull World Health Organ*. 1976;54(5):541-553.
- e6. Foulkes MA, Wolf PA, Price TR, Mohr JP, Hier DB. The Stroke Data Bank: design, methods, and baseline characteristics. *Stroke*. 1988;19(5):547-554. doi:10.1161/01.str.19.5.547
- e7. Salomaa V. Genetic and Environmental Contributions to Cardiovascular Risk: Lessons From North Karelia and FINRISK. *Glob Heart*. 2016;11(2):229-233. doi:10.1016/j.gheart.2016.04.008
- e8. Saleheen D, Zaidi M, Rasheed A, et al. The Pakistan Risk of Myocardial Infarction Study: a resource for the study of genetic, lifestyle and other determinants of myocardial infarction in South Asia. *Eur J Epidemiol*. 2009;24(6):329-338. doi:10.1007/s10654-009-9334-y
- e9. Lavange LM, Kalsbeek WD, Sorlie PD, et al. Sample design and cohort selection in the Hispanic Community Health Study/Study of Latinos. *Ann Epidemiol*. 2010;20(8):642-649. doi:10.1016/j.annepidem.2010.05.006
- e10. Juster FT, Suzman RM. An Overview of the Health and Retirement Study. *J Hum Resour*. 1995;30:S7-56.
- e11. Adams HP, Bendixen BH, Kappelle LJ, et al. Classification of subtype of acute ischemic stroke. Definitions for use in a multicenter clinical trial. TOAST. Trial of Org 10172 in Acute Stroke Treatment. *Stroke*. 1993;24(1):35-41. doi:10.1161/01.str.24.1.35
- e12. Ay H, Furie KL, Singhal A, Smith WS, Sorensen AG, Koroshetz WJ. An evidence-based causative classification system for acute ischemic stroke. *Ann Neurol*. 2005;58(5):688-697. doi:10.1002/ana.20617
- e13. NINDS Stroke Genetics Network (SiGN), International Stroke Genetics Consortium (ISGC). Loci associated with ischaemic stroke and its subtypes (SiGN): a genome-wide association study. *Lancet Neurol*. 2016;15(2):174-184. doi:10.1016/S1474-4422(15)00338-5

- e14. Goumidi L, Thibord F, Wiggins KL, et al. Association between ABO haplotypes and the risk of venous thrombosis: impact on disease risk estimation. *Blood*. 2021;137(17):2394-2402. doi:10.1182/blood.2020008997
- e15. Groot HE, Villegas Sierra LE, Said MA, Lipsic E, Karper JC, van der Harst P. Genetically Determined ABO Blood Group and its Associations With Health and Disease. *Arterioscler Thromb Vasc Biol*. 2020;40(3):830-838. doi:10.1161/ATVBAHA.119.313658
- e16. Rannikmäe K, Ngoh K, Bush K, et al. Accuracy of identifying incident stroke cases from linked health care data in UK Biobank. *Neurology*. 2020;95(6):e697-e707. doi:10.1212/WNL.00000000000009924
- e17. Klarin D, Busenkell E, Judy R, et al. Genome-wide association analysis of venous thromboembolism identifies new risk loci and genetic overlap with arterial vascular disease. *Nat Genet*. 2019;51(11):1574-1579. doi:10.1038/s41588-019-0519-3
- e18. Choi SW, O'Reilly PF. PRSice-2: Polygenic Risk Score software for biobank-scale data. *GigaScience*. 2019;8(7):giz082. doi:10.1093/gigascience/giz082