Global Assessment of Mendelian Stroke Genetic Prevalence in 101 635 Individuals From 7 Ethnic Groups

SUPPLEMENTAL MATERIAL

Supplemental Methods

Identifying Candidate Mendelian Stroke Genes

To identify relevant Mendelian stroke genes, MEDLINE and Embase were searched for relevant articles published in English from January 1, 1990 to June 12, 2018. In consultation with a health sciences librarian, a search strategy was created in four parts: (1) terms specific to the outcome including "stroke", "ictus" and "intracerebral hemorrhage"; (2) terms highlighting the occurrence of the outcome including "prevalence" and "penetrance"; (3) terms related to disease inheritance such as "monogenic", "Mendelian", and "hereditary"; and (4) terms related to the cause of disease such as "mutation" or "variation".

The search strategy encompassed the following:
January 1st, 1990 to June 12th, 2018 [only in humans, only English] in Embase, OVID MEDLINE Epub Ahead of Print, In-Process & Other Non-Indexed Citations, Ovid MEDLINE(R) Daily and Ovid MEDLINE(R) 1946-Present
exp STROKE/ OR stroke* OR ictus OR ischemi* OR hemorrhag* OR intracerebral adj1 hemorrhage OR infarct* AND exp PREVALENCE/ OR prevalen* OR occurrence OR penetran* AND monogenic OR Mendelian OR hereditary OR exp GENES, DOMINANT OR dominant OR exp GENES, RECESSIVE OR recessive OR single adj1 gene AND exp MUTATION OR mutation* OR variation OR non-synonymous

Eligible publications were screened for potentially relevant genetic loci and new studies discussing novel loci published after the initial search were also included. Only studies that denoted genetic variants in certain individuals were included. The maximum age referenced in articles was limited to 70 years old and stroke attributable to antecedent head trauma or sympathomimetic drugs led to article exclusion. Most importantly, only genes that were shown to lead to stroke as a prevailing clinical manifestation were analyzed. Genes were excluded if 1) stroke was not a primary feature, 2) there was an absence of clear evidence of Mendelian inheritance, 3) the gene was encoded in the mitochondrial genome. A total of 39 articles were included from 880 unique publications. Overall, 18 candidate genes (Table VI) were found to be directly implicated in the development of stroke and followed a Mendelian inheritance pattern.

gnomAD Population Data

The Genome Aggregation Database (gnomAD) is the largest publicly-available population dataset to date and makes summary-level genetic data available for the wider scientific community.[1] Research was approved by the Hamilton Integrated Research Ethics Board. gnomAD ethics approval is overseen by the Broad Institute's Office of Research Subject Protection and the Partners Human Research Committee, and informed consent was obtained from all participants. Only samples from individuals who were not ascertained for having a neurological condition in a neurological case/control study were used in this study (gnomAD v2.1.1 non-neuro subset). The dataset comprised genetic information from 101,635 unrelated individuals sequenced by whole-exome sequencing from dozens of international contributing projects. Individuals known to be affected by severe pediatric disease as well as their first-degree relatives were removed.

Principal component analysis (PCA) organized gnomAD participants into seven distinct ethnic groups by ancestry: African/African-American (AFR), Latino/Admixed American (AMR), Ashkenazi Jewish (ASJ), East Asian (EAS), Finnish European (FIN), Non-Finnish European (NFE), and

South Asian (SAS). A number of unassigned individuals classified as Other (OTH) did not visibly cluster with any of the seven major populations and were not analyzed.

Variant Annotation, Filtering and Classification

A web-based version of ANNOVAR was used to annotate variant mutation effects based on ENSEMBL transcripts.[2] Only rare non-synonymous mutations within the 18 genes were analyzed in the present study. Non-synonymous mutations included any variants which are predicted to alter the primary amino acid sequence of the protein product; missense, inframe insertion/deletion, frameshift, start gain/loss, stop gain/loss, and splice site variants were included. In gnomAD these variants were listed as "loss of function (LoF)" and "missense". Variants with dubious annotation and/or quality were removed. Rare variants were defined as those having a minor allele frequency (MAF) <0.001 within the gnomAD v2.1.1 (non-neuro) dataset. This MAF threshold was applied within each ethnic subdivision of gnomAD (AFR, AMR, ASJ, EAS, FIN, NFE, SAS) and for each variant class. If a variant was common (MAF≥0.001) in a single ethnic group, then it was excluded from analysis. The remaining variants were deemed to be globally rare. Several *bona fide* pathogenic mutations were spiked into MAF calculations after the MAF threshold was established. Two well-established pathogenic founder *NOTCH3* variants with MAF>0.001 were included in the analysis (R544C and R1231C). One well-established *RNF213* pathogenic founder variant with MAF>0.001 was included (R4810K). These founder mutations were included irrespective of frequency. Based on cumulative minor allele frequencies, carrier frequencies were calculated and used to estimate the prevalence of putatively disease-causing mutations for each gene.

Definition of Pathogenic Clinical Variants

Pathogenic clinical variants encompassed three subgroups of mutation types.

1. Firstly, canonical disease-causing (CDC) mutations were extracted from gnomAD. These variants were directly relevant to clinical cases and followed highly stereotypical patterns consistent with known disease pathways. Rare variants were categorized as CDC mutations if they strongly resembled reported disease-causing mutations from genetic databases: The Online Mendelian Inheritance of Man compendium (OMIM) and/or the National Institutes of Health's Genetic Home reference (NIH-GHR).
2. Next, suspected pathogenic variants associated with the gene of interest systematically found in literature or large case series were extracted from gnomAD.
3. Lastly, variants associated with the gene and disorder of interest that were unambiguously classified as pathogenic or likely pathogenic in ClinVar were extracted from gnomAD.[3]

Note: *NOTCH3* cysteine-sparing variants, while searchable in gnomAD non-neuro, were excluded from analysis since their pathogenicity is still debated.[4]

Baseline disease prevalence estimates were found by searching the rare disease registry, Orphanet, and clinical presentation was described in the National Institutes of Health's Genetics Home Reference.

Twelve genes (*ABCC6*, *CECR1*, *COL3A1*, *COL4A1*, *COL4A2*, *COLGALT1*, *GLA*, *HTRA1*, *KRIT1*, *NOTCH3*, *RNF213*, and *TREX1*) were found to exhibit pathogenic clinical mutations. Only *NOTCH3* and *COL4A1*, however, presented CDC mutations. *NOTCH3* mutations that involved a gain

or loss of cysteine in one of the 34 epidermal growth factor-like repeat (EGFR) domains (ENSG00000074181) (residues 40-1373) and *COL4A1* glycine-altering variants in the triple-helical domain (ENSG00000187498) (residues 173-1440) were included.[5,6]

Using CADD to Predict Variant Effect

CADD v1.4 utilizes in-silico algorithms to predict variant deleteriousness of any non-synonymous mutations based on biochemical characteristics and sequence conservation. CADD scores expanded variant inclusion criteria to encompass all rare non-synonymous mutations with PHRED-scaled C-scores >20.00 and MAF<0.001. These mutations rank among the top 1% of deleterious variants and are more likely to be implicated in disease development. This mutation class lacks the specificity of pathogenic clinical variants but may allow for novel variants associated with disease to be determined.[7]

All Non-Synonymous Variants

The most inclusive mutation class encompassed all non-synonymous variants located anywhere in the gene of interest, regardless of deleteriousness (MAF<0.001).

Supplemental Tables

Gene Locus	P-Value
ABCC6	1.9E-06
CECR1	2.5E-04
COL3A1	9.5E-02
COL4A1	7.8E-07
COL4A2	3.6E-01
COLGALT1	7.3E-02
GLA	6.4E-01
HTRA1	2.3E-05
KRIT1	1.3E-02
NOTCH3	< 2.2E-16
RNF213	< 2.2E-16
TREX1	7.9E-02

Table I: Comparing Wildtype and Alternate Allele Counts for Pathogenic Clinical Variants among gnomAD v2.1.1 (non-neuro) Subpopulations

Ethnicity							
Gene	**SAS**	**EAS**	**ASJ**	**AFR**	**NFE**	**FIN**	**AMR**
ABCC6	21.94	27.92	5.16	19.82	15.00	5.58	17.59
APP	8.70	10.32	1.61	11.88	6.34	1.67	7.87
CCM2	7.06	5.08	1.94	5.84	6.01	1.56	2.49
CECR1	6.01	2.38	0.64	3.71	3.49	2.16	6.49
COL3A1	13.07	12.82	1.29	8.59	8.44	1.55	9.92
COL4A1	12.74	8.23	0.98	12.35	11.42	3.62	12.73
COL4A2	17.39	18.90	6.40	18.42	17.71	7.52	21.39
COLGALT1	9.97	10.58	1.94	11.34	9.04	0.91	9.75
CST3	1.09	0.30	0.78	0.77	0.77	0	1.44
CTSA	3.46	5.59	0.64	6.43	3.58	2.75	4.00
GLA	0.42	0.79	0	2.15	1.52	0	0.74
HTRA1	4.67	4.23	0.64	3.24	4.61	3.20	7.01
ITM2B	1.21	2.39	0.64	1.98	1.24	0.24	2.20
KRIT1	6.15	4.77	2.26	6.54	5.24	1.91	4.33
NOTCH3	27.39	34.64	8.28	32.09	20.67	5.26	24.59
PDCD10	0.85	0.45	0.32	0.99	0.76	0	0.66
RNF213	49.96	41.50	8.95	27.71	29.13	8.74	34.61
TREX1	5.10	6.27	0.32	5.24	3.83	0.84	2.36
Total	197.18	197.15	42.79	179.08	148.81	47.51	170.18

Table II: CADD-Predicted Deleterious Variant Carrier Frequency (per 1000) Stratified by gnomAD v2.1.1 (non-neuro) Subpopulation

Ethnicity							
Gene	**SAS**	**EAS**	**ASJ**	**AFR**	**NFE**	**FIN**	**AMR**
ABCC6	38.98	47.61	7.19	40.86	31.30	12.80	32.73
APP	14.93	13.13	1.93	15.62	9.05	2.65	10.95
CCM2	9.42	8.16	2.38	9.00	7.94	1.80	6.78
CECR1	13.73	6.41	1.61	9.26	10.40	3.24	14.23
COL3A1	16.68	20.95	2.03	13.22	11.68	1.55	13.52
COL4A1	30.00	23.30	2.99	22.88	20.61	4.93	24.12
COL4A2	37.27	34.28	9.68	28.55	27.40	11.09	36.57
COLGALT1	14.10	14.31	3.22	18.90	14.15	3.12	15.00
CST3	5.88	2.25	1.54	2.90	2.09	0.57	3.15
CTSA	6.09	13.43	2.48	14.51	9.15	3.11	9.64
GLA	4.30	1.38	0.43	6.11	4.61	0.49	4.11
HTRA1	6.42	4.90	1.50	5.58	5.73	3.44	8.23
ITM2B	2.78	5.82	1.29	5.01	2.76	0.60	3.68
KRIT1	8.96	10.44	2.26	10.49	7.32	4.19	6.56
NOTCH3	42.58	47.10	10.52	46.99	32.69	8.37	39.14
PDCD10	1.77	1.34	0.32	2.10	2.75	0	1.38
RNF213	107.47	107.69	22.92	91.86	78.52	22.00	96.94
TREX1	9.22	12.39	3.22	10.12	6.95	3.08	6.69
Total	370.56	374.90	77.51	353.98	285.11	87.03	333.39

Table III: Non-Synonymous Variant Carrier Frequency (per 1000) Stratified by gnomAD v2.1.1 (non-neuro) Subpopulation

Chromosome	Position	rsID	Reference	Alternate	Protein Consequence	Annotation	SAS AF	EAS AF	ASJ AF	AFR AF	NFE AF	FIN AF	AMR AF
19	15289890	rs199638166	A	C	p.Cys1222Gly	Missense	0	0	0	6.22E-05	2.25E-04	0	0
19	15289986	rs377099118	G	A	p.Arg1190Cys	Missense	1.63E-04	1.50E-04	0	0	3.37E-05	0	0
19	15295828	rs532100840	G	A	p.Arg767Cys	Missense	0	4.60E-04	0	0	0	4.02E-04	0
19	15296215	rs144163298	G	A	p.Arg717Cys	Missense	3.29E-05	1.51E-04	0	1.32E-04	4.65E-05	0	0
19	15289953	rs772172068	G	A	p.Arg1201Cys	Missense	3.27E-05	7.49E-05	0	0	6.74E-05	0	6.56E-05
19	15297997	rs754554486	G	A	p.Arg587Cys	Missense	0	2.24E-04	0	0	0	0	1.97E-04
19	15298024	rs769773673	G	A	p.Arg578Cys	Missense	0	7.47E-05	0	6.20E-05	2.25E-05	0	1.64E-04
19	15297722	rs760768552	G	A	p.Arg640Cys	Missense	0	2.24E-04	0	6.17E-05	1.12E-05	0	6.55E-05
19	15298084	rs75068032	G	A	p.Arg558Cys	Missense	0	0	0	0	5.60E-05	0	0
19	15290208	rs60373464	G	A	p.Arg1143Cys	Missense	0	7.45E-05	0	0	2.23E-05	0	0
19	15295249	rs776115188	TGGCATCGTG	T	p.Pro805_Cys807del	Inframe Deletion	0	7.46E-05	0	0	0	0	0
19	15291915	rs775964142	G	A	p.Arg951Cys	Missense	0	7.84E-05	1.68E-04	0	0	0	0
19	15297737	rs753801611	G	A	p.Arg635Cys	Missense	0	7.45E-05	0	0	0	0	3.28E-05
19	15289926	rs758961316	G	A	p.Arg1210Cys	Missense	3.27E-05	0	0	0	1.12E-05	0	0
19	15291029	rs1320508682	A	G	p.Cys1061Arg	Missense	3.27E-05	0	0	0	1.12E-05	0	0
19	15291810	rs763321998	A	G	p.Cys986Arg	Missense	3.64E-05	0	0	0	1.33E-05	0	0
19	15291942	rs777577687	C	A	p.Gly942Cys	Missense	0	0	0	0	2.46E-05	0	0
19	15296182	rs1057519101	G	A	p.Arg728Cys	Missense	0	0	0	0	2.32E-05	0	0
19	15295135	rs558392935	C	T	p.Cys846Tyr	Missense	0	7.46E-05	0	0	0	0	0
19	15295804	rs1383763025	A	G	p.Cys775Arg	Missense	0	1.15E-04	0	0	0	0	0
19	15297937	rs777751303	G	A	p.Arg607Cys	Missense	0	7.46E-05	0	0	0	0	0
19	15288794	rs1396345163	G	C	p.Cys1315Trp	Missense	0	0	0	0	5.00E-05	0	0
19	15289641	rs1339695535	C	T	p.Cys1277Tyr	Missense	3.64E-05	0	0	0	0	0	0
19	15289747	rs769660847	G	A	p.Arg1242Cys	Missense	0	0	0	0	1.13E-05	0	0
19	15289949	rs754523402	C	G	p.Cys1202Ser	Missense	0	0	0	0	0	0	3.28E-05
19	15289982	rs1192888680	C	G	p.Cys1191Ser	Missense	0	0	0	0	0	0	3.28E-05
19	15290279	rs1266914122	C	T	p.Cys1119Tyr	Missense	0	0	0	0	1.12E-05	0	0
19	15291004	rs1438064001	T	C	p.Tyr1069Cys	Missense	0	0	0	0	0	5.98E-05	0
19	15291918	rs1378535955	A	C	p.Cys950Gly	Missense	0	0	0	0	0	6.19E-05	0
19	15291969	rs749778923	A	C	p.Cys933Gly	Missense	0	0	0	0	0	0	3.34E-05
19	15292508	rs1447534769	A	C	p.Cys891Gly	Missense	0	0	0	0	0	0	0
19	15292598	rs757098265	C	A	p.Gly861Cys	Missense	0	0	0	0	0	0	4.64E-05
19	15296177	rs760081167	G	T	p.Cys729Ter	Stop Gained	0	0	1.67E-04	0	0	0	0
19	15296404	rs1250956327	G	A	p.Arg680Cys	Missense	3.27E-05	0	0	0	0	0	0
19	15296452	rs1350049644	T	A	p.Ser664Cys	Missense	0	0	0	0	1.12E-05	0	0
19	15297933	rs778350156	C	T	p.Cys608Tyr	Missense	0	0	0	0	1.13E-05	0	0
19	15297982	rs764148985	G	A	p.Arg592Cys	Missense	0	0	0	0	1.12E-05	0	0
19	15298087	rs1317994194	C	A	p.Gly557Cys	Missense	3.27E-05	0	0	0	0	0	0
19	15298704	rs1202763005	G	A	p.Arg532Cys	Missense	0	0	0	1.18E-04	0	0	0
19	15299808	rs1325111374	C	CA	p.Cys457LeufsTer15	Frameshift	0	0	0	0	0	0	3.34E-05
19	15302831	rs775267348	G	A	p.Arg207Cys	Missense	0	0	0	0	1.13E-05	0	0
19	15303013	rs1236699193	C	T	p.Cys146Tyr	Missense	0	0	0	0	0	0	3.33E-05
19	15303053	rs137852642	G	A	p.Arg133Cys	Missense	0	0	0	0	0	6.33E-05	0
19	15296443	rs376046941	C	A	p.Gly667Cys	Missense	0	0	0	0	0	0	0
19	15298126	rs201118034	G	A	p.Arg544Cys	Missense	9.81E-05	3.66E-03	0	0	0	0	3.28E-05
19	15289863	rs201680145	G	A	p.Arg1231Cys	Missense	5.36E-03	0	0	0	3.15E-04	0	6.56E-04

Table IV: Pathogenic Clinical Variants Associated with CADASIL in gnomAD v2.1.1 (non-neuro)
Legend: SAS = South Asian, EAS = East Asian, ASJ = Ashkenazi Jewish, AFR = African/African-American, NFE = Non-Finnish European, FIN = Finnish European, AMR = Latino/Admixed American, AF = Allele Frequency

Chromosome	Position	rsID	Reference	Alternate	Protein Consequence	Annotation	SAS AF	EAS AF	ASJ AF	AFR AF	NFE AF	FIN AF	AMR AF
19	15308327	rs200595885	G	A	p.Arg61Trp	Missense	6.57E-05	7.56E-05	3.32E-04	1.99E-04	2.55E-04	0	3.30E-05
19	15302857	rs140914494	G	A	p.Ala198Val	Missense	0	2.25E-04	0	0	5.65E-05	0	0
19	15302858	rs375682932	C	T	p.Ala198Thr	Missense	0	0	0	0	3.39E-05	5.99E-05	3.28E-05
19	15299803	rs370186772	C	T	p.Ala459Thr	Missense	0	0	0	0	0	0	0
19	15272105	rs372833545	C	T	p.Gly2112Ser	Missense	0	3.89E-04	0	0	4.37E-05	0	0
19	15298793	rs778571943	G	A	p.Ser502Phe	Missense	0	0	0	1.17E-04	1.41E-04	7.40E-05	4.84E-05
19	15296089	rs1442324683	T	A	p.Thr759Ser	Missense	0	0	0	0	1.14E-05	0	0
19	15289998	rs1400946198	C	G	p.Val1186Leu	Missense	0	0	0	0	1.12E-05	0	0
19	15303235	rs1470690834	A	G	p.Val98Ala	Missense	0	0	0	0	1.15E-05	0	0
19	15289686	rs143684274	C	A	p.Arg1262Leu	Missense	0	0	0	6.61E-05	1.17E-04	6.13E-05	0
19	15302288	rs769567750	G	A	p.Thr328Ile	Missense	0	0	0	0	1.13E-05	0	0
19	15299072	rs376728138	T	C	p.Asn489Ser	Missense	7.14E-05	0	0	0	7.39E-05	1.89E-04	4.54E-04
19	15291825	rs143695196	G	A	p.His981Tyr	Missense	0	0	0	0	9.32E-05	0	1.11E-04

Table V: Cysteine-Sparing Variants Associated with CADASIL in gnomAD v2.1.1 (non-neuro) (MAF<0.001)

Gene	Genetic Disorder	Mode of Inheritance	Clinical Manifestation	Disease Prevalence
NOTCH3	CADASIL	AD	Migraine with aura, mood disturbance, depression, dementia, epileptic seizures	1 in 25,000 - 50,000
HTRA1	CARASIL	AR	Spasticity, dementia, alopecia, gait disturbance, lumbago, spondylosis deformans, disc herniation	Unknown
KRIT1, CCM2, PDCD10	Familial cerebral cavernous malformations	AD	Epileptic seizures, headaches, paralysis, auditory and visual impairment	1 in 5000 -10,000
APP, CST3, ITM2B	Hereditary cerebral amyloid angiopathy	AD	Drowsiness, headaches, seizures, vomiting, dementia	Unknown
RNF213	Moyamoya disease	Unknown	New small fragile cerebral vessel networks, headaches, chorea	1 in 9500 - 30,000
COL3A1	Vascular Ehlers-Danlos Syndrome (Type IV)	AD	Hypermobility, hypotonia, highly elastic skin, abnormal scar formation	1 in 10,000 - 100,000
ABCC6	Pseudoxanthoma elasticum	AR	Cervical and axillary papules, Ocular peau d'orange lesions and angioid streaks, hypertension, peripheral artery disease, coronary artery disease, gastrointestinal bleeding	1 in 25,000 -100,000
GLA	Fabry disease	X-Linked	Acroparesthesia, hypohidrosis, angiokeratoma, corneal opacity, tinnitus, chronic kidney disease, cardiomyopathy	1 in 3000
COL4A1	*COL4A1*-related small vessel diseases	AD	Porencephaly, intracranial aneurysms, migraine with aura, retinal arteriolar tortuosities and hemorrhage, cataracts, Axenfeld-Rieger anomaly, nephropathy, muscle cramps	Unknown
COL4A2	*COL4A2*-related small vessel diseases	AD	Porencephaly, cataracts	Unknown
CECR1	Deficiency of *ADA2*	AR	Livedo reticularis, livedo racemosa, hepatosplenomegaly, systemic vasculitis	< 1 in 1,000,000
TREX1	Retinal vasculopathy with cerebral leukodystrophy	AD	Visual impairment, dementia, headache	Unknown
CTSA	CARASAL	AD	Lower cranial nerve dysfunction (vertigo, dysphagia, dry mouth, dry eyes, central facial paresis, dysarthria), refractory hypertension	Unknown
COLGALT1	*COL4A1* and *COL4A2*-related small vessel diseases	AR	Porencephaly, frequent skin eruption, microscopic hematuria	Unknown

Table VI: Genetic Loci Associated with Mendelian Stroke

Ethnicity	Pathogenic Clinical Variant Carrier Frequency	CADD-Predicted Deleterious Variant Carrier Frequency	Total Non-Synonymous Variant Carrier Frequency
SAS	20.09	197.18	370.56
EAS	28.35	197.15	374.90
ASJ	3.25	42.79	77.51
AFR	7.14	179.08	353.98
NFE	12.89	148.81	285.11
FIN	7.67	47.51	87.03
AMR	11.84	170.18	333.39

Table VII: Mendelian Stroke Carrier Frequency (per 1000) by gnomAD v2.1.1 (non-neuro) Subpopulation and Variant Class

Genotype Call Rate (%) [SD %]							
	Ethnicity						
Gene	AFR	AMR	ASJ	EAS	FIN	NFE	SAS
ABCC6	93.7 [15.1]	95.9 [12.2]	95.6 [13.3]	96.5 [9.7]	95.6 [15.1]	94.1 [15.1]	96.3 [11.8]
APP	96.6 [13.9]	98.1 [7.6]	98.3 [6.4]	97.6 [9.3]	95.8 [16.8]	96.8 [11.8]	98.4 [6.2]
CCM2	96.0 [14.4]	97.9 [7.9]	98.0 [6.4]	97.5 [9.7]	94.8 [18.9]	96.4 [12.3]	98.4 [6.4]
CECR1	99.9 [0.5]	99.9 [0.1]	99.9 [0.2]	100.0 [0.1]	99.7 [1.2]	99.9 [0.4]	100.0 [0.1]
COL3A1	97.0 [9.7]	98.0 [6.9]	98.4 [5.2]	98.1 [7.0]	98.8 [4.2]	97.3 [9.3]	98.3 [5.7]
COL4A1	97.4 [11.4]	98.8 [6.3]	98.8 [4.8]	98.5 [10.3]	98.4 [10.0]	98.2 [8.3]	98.9 [5.4]
COL4A2	90.2 [11.3]	95.2 [10.2]	96.1 [8.5]	93.4 [8.7]	97.3 [5.0]	94.4 [10.9]	96.2 [8.3]
COLGALT1	95.6 [15.1]	96.5 [12.1]	97.3 [9.4]	96.6 [13.1]	95.4 [12.6]	95.9 [13.7]	97.4 [9.7]
CST3	64.6 [35.7]	79.7 [21.7]	83.9 [18.7]	71.3 [29.8]	88.4 [13.0]	72.6 [27.9]	82.4 [19.5]
CTSA	97.9 [7.3]	99.1 [4.1]	99.0 [3.6]	98.9 [4.3]	98.4 [6.0]	98.5 [5.6]	99.3 [3.3]
GLA	80.8 [0.5]	79.8 [0.2]	74.7 [0.4]	75.5 [0.1]	72.7 [0.2]	72.0 [0.4]	62.1 [0.6]
HTRA1	83.3 [33.1]	90.5 [19.8]	92.0 [16.7]	86.1 [29.5]	82.5 [33.2]	86.9 [26.0]	91.7 [17.5]
ITM2B	89.8 [23.0]	95.2 [11.0]	95.9 [9.3]	92.5 [17.2]	97.3 [5.9]	93.7 [14.4]	95.6 [10.1]
KRIT1	99.8 [0.4]	99.9 [0.1]	99.9 [0.1]	100.0 [0.0]	99.7 [0.6]	99.8 [0.5]	99.9 [0.1]
NOTCH3	85.7 [21.5]	91.6 [14.7]	91.7 [13.3]	91.2 [16.2]	88.3 [20.4]	88.0 [18.8]	92.9 [12.5]
PDCD10	99.8 [0.4]	99.8 [0.2]	99.9 [0.1]	99.9 [0.1]	99.6 [0.6]	99.9 [0.1]	99.9 [0.2]
RNF213	92.5 [17.8]	95.5 [11.1]	96.3 [9.0]	94.7 [13.2]	89.5 [26.2]	93.4 [15.7]	96.4 [8.9]
TREX1	99.0 [1.6]	99.9 [0.1]	99.9 [0.3]	99.6 [0.8]	98.7 [3.1]	99.4 [0.9]	100.0 [0.1]

Table VIII: Mean gnomAD v2.1.1 (non-neuro) Genotype Call Rate Stratified by Ethnicity and Gene

Gene	Overall Proportion of Sequences with ≥ 20X Coverage
ABCC6	88.3
APP	97.2
CCM2	96.7
CECR1	99.2
COL3A1	94.1
COL4A1	95.7
COL4A2	86.8
COLGALT1	79.0
CST3	65.1
CTSA	97.1
GLA	98.6
HTRA1	74.5
ITM2B	90.9
KRIT1	99.1
NOTCH3	77.0
PDCD10	98.7
RNF213	90.7
TREX1	98.5

Table IX: Overall Proportion of Sequences in gnomAD v2.1.1 (non-neuro) with ≥ 20X Coverage across 18 Mendelian Stroke Genes

Gene	Variant Class	Total Number of Unique Mutations	Frameshift	Inframe Deletion	Inframe Insertion	Missense	Splice Acceptor	Splice Donor	Start Gained	Start Lost	Stop Gained	Stop Lost
ABCC6	Pathogenic Clinical Variants	95	7	0	0	76	3	1	0	0	8	0
	CADD-Predicted Deleterious Variants	478	0	0	0	453	0	0	0	0	25	0
	Non-Synonymous Variants	949	43	6	3	844	10	13	0	0	30	0
APP	Pathogenic Clinical Variants	0	0	0	0	0	0	0	0	0	0	0
	CADD-Predicted Deleterious Variants	267	0	0	0	260	0	0	0	0	7	0
	Non-Synonymous Variants	372	4	9	5	337	4	4	0	0	9	0
CCM2	Pathogenic Clinical Variants	0	0	0	0	0	0	0	0	0	0	0
	CADD-Predicted Deleterious Variants	142	0	0	0	140	0	0	0	0	2	0
	Non-Synonymous Variants	246	3	0	0	235	0	5	0	0	2	1
CECR1	Pathogenic Clinical Variants	19	2	0	0	15	1	0	0	0	1	0
	CADD-Predicted Deleterious Variants	124	0	0	0	115	0	0	0	0	9	0
	Non-Synonymous Variants	296	11	2	0	257	7	4	0	3	12	0
COL3A1	Pathogenic Clinical Variants	2	0	0	0	2	0	0	0	0	0	0
	CADD-Predicted Deleterious Variants	332	0	0	0	330	0	0	0	0	2	0
	Non-Synonymous Variants	475	2	0	2	466	2	1	0	0	2	0
COL4A1	Pathogenic Clinical Variants	49	0	0	0	49	0	0	0	0	0	0
	CADD-Predicted Deleterious Variants	325	0	0	0	325	0	0	0	0	0	0
	Non-Synonymous Variants	649	1	2	0	640	3	3	0	0	0	0
COL4A2	Pathogenic Clinical Variants	2	0	0	0	1	0	0	0	0	1	0
	CADD-Predicted Deleterious Variants	453	0	0	0	434	0	0	0	0	19	0
	Non-Synonymous Variants	860	43	6	3	770	12	5	0	2	19	0
COLGALT1	Pathogenic Clinical Variants	1	0	0	0	1	0	0	0	0	0	0
	CADD-Predicted Deleterious Variants	215	0	0	0	206	0	0	0	0	9	0
	Non-Synonymous Variants	346	17	3	1	308	3	4	0	0	10	0
CST3	Pathogenic Clinical Variants	0	0	0	0	0	0	0	0	0	0	0
	CADD-Predicted Deleterious Variants	36	0	0	0	35	0	0	0	0	1	0
	Non-Synonymous Variants	83	4	1	2	70	0	3	0	0	3	0
CTSA	Pathogenic Clinical Variants	0	0	0	0	0	0	0	0	0	0	0
	CADD-Predicted Deleterious Variants	151	0	0	0	140	0	0	0	1	10	0
	Non-Synonymous Variants	285	12	6	1	247	4	3	0	1	11	0
GLA	Pathogenic Clinical Variants	3	0	0	0	3	0	0	0	0	0	0
	CADD-Predicted Deleterious Variants	29	0	0	0	29	0	0	0	0	0	0
	Non-Synonymous Variants	81	0	0	0	81	0	0	0	0	0	0
HTRA1	Pathogenic Clinical Variants	11	0	0	0	8	0	0	0	0	3	0
	CADD-Predicted Deleterious Variants	140	0	0	0	135	0	0	0	0	5	0
	Non-Synonymous Variants	191	4	0	2	177	1	2	0	0	5	0
ITM2B	Pathogenic Clinical Variants	0	0	0	0	0	0	0	0	0	0	0
	CADD-Predicted Deleterious Variants	69	0	0	0	66	0	0	0	1	2	0
	Non-Synonymous Variants	110	3	4	2	98	0	0	0	1	2	0
KRIT1	Pathogenic Clinical Variants	6	1	0	0	0	0	1	0	0	4	0
	CADD-Predicted Deleterious Variants	201	0	0	0	190	0	0	0	0	10	1
	Non-Synonymous Variants	316	8	4	0	288	1	3	0	0	11	1
NOTCH3	Pathogenic Clinical Variants	46	1	1	0	43	0	0	0	0	1	0
	CADD-Predicted Deleterious Variants	642	0	0	0	631	0	0	0	0	11	0
	Non-Synonymous Variants	1019	16	1	4	972	7	6	0	0	13	0
PDCD10	Pathogenic Clinical Variants	0	0	0	0	0	0	0	0	0	0	0
	CADD-Predicted Deleterious Variants	39	0	0	0	38	0	0	0	1	0	0
	Non-Synonymous Variants	74	0	1	0	72	0	0	0	1	0	0
RNF213	Pathogenic Clinical Variants	1	0	0	0	1	0	0	0	0	0	0
	CADD-Predicted Deleterious Variants	1005	0	0	0	935	0	0	0	0	70	0
	Non-Synonymous Variants	2614	105	23	5	2346	25	24	0	1	83	2
TREX1	Pathogenic Clinical Variants	1	0	0	0	1	0	0	0	0	0	0
	CADD-Predicted Deleterious Variants	106	0	0	0	101	2	0	0	2	1	0
	Non-Synonymous Variants	256	29	2	6	209	3	2	0	4	1	0

Table X: Variant Annotations by Gene and Mutation Class in gnomAD v2.1.1 (non-neuro)

Note: Because pathogenic variants were established through literature rather than with bioinformatic criteria, certain variants were included as pathogenic but not deleterious according to CADD's PHRED score.

	Ethnicity						
Gene	**SAS**	**EAS**	**ASJ**	**AFR**	**NFE**	**FIN**	**AMR**
ABCC6	4.52	8.40	2.19	2.48	5.95	3.46	5.68
CECR1	1.18	0.43	0.20	0.49	1.48	0.19	1.74
COL3A1	0	0	0	0.23	0.03	0	0
COL4A1	0.72	0.65	0.20	2.54	2.18	0.65	1.45
COL4A2	0.13	0	0	0	0.02	0	0
COLGALT1	0	0	0	0.12	0	0	0
GLA	0	0	0	0.15	0.05	0	0
HTRA1	0.78	0.33	0	0.25	0.65	1.78	0.22
KRIT1	0.13	0	0	0	0	0	0.23
NOTCH3	11.78	11.26	0.41	0.87	1.90	1.11	2.69
RNF213	0.85	5.65	0	0	0	0	0.06
TREX1	0	0	0	0	0	0	0.12
Total	20.09	26.73	2.99	7.13	12.26	7.19	12.17

Table XI: Pathogenic Clinical Variant Carrier Frequency (per 1000) Stratified by gnomAD v2.1.1 Subpopulation

Supplemental References

1. Karczewski KJ, Francioli LC, Tiao G, Cummings BB, Alföldi J, Wang Q, et al. Variation across 141,456 human exomes and genomes reveals the spectrum of loss-of-function intolerance across human protein-coding genes. *BioRxiv*. 2019:531210.
2. Yang H, Wang K. Genomic variant annotation and prioritization with ANNOVAR and wANNOVAR. *Nat Protoc*. 2015;10:1556-1566.
3. Landrum MJ, Lee JM, Benson M, Brown G, Chao C, Chitipiralla S, et al. ClinVar: public archive of interpretations of clinically relevant variants. *Nucleic Acids Res*. 2016;44:D862-868.
4. Mishra A, Chauhan G, Violleau MH, Vojinovic D, Jian X, Bis JC, Li S, et al. Association of variants in HTRA1 and NOTCH3 with MRI-defined extremes of cerebral small vessel disease in older subjects. *Brain*. 2019;142:1009-1023.
5. Joutel A, Corpechot C, Ducros A, Vahedi K, Chabriat H, Mouton P, et al. Notch3 mutations in CADASIL, a hereditary adult-onset condition causing stroke and dementia. *Nature*. 1996;383:707-710.
6. Volonghi I, Pezzini A, Del Zotto E, Giossi A, Costa P, Ferrari D, et al. Role of COL4A1 in basement-membrane integrity and cerebral small-vessel disease. The COL4A1 stroke syndrome. *Curr Med Chem*. 2010;17:1317-1324.
7. Rentzsch P, Witten D, Cooper GM, Shendure J, Kircher M. CADD: predicting the deleteriousness of variants throughout the human genome. *Nucleic Acids Res*. 2018;47:D886-894.