***MAP3K6* mutations in a neurovascular disease causing stroke, cognitive impairment and tremor**

Authors: Andreea Ilinca, Elisabet Englund, Sofie Samuelsson, Katarina Truvé, Efthymia Kafantari, Nicolas Martinez-Majander, Jukka Putaala, Claes Håkansson, Arne G. Lindgren, Andreas Puschmann

Supplemental data:

table e1: Detailed clinical information on the studied individuals

Please see separate pdf file at: links.lww.com/NXG/A365.

Abbreviations in table e1:

AF, atrial fibrillation; Angio, angiography; ANS, autonomous nervous system; Ap, angina pectoris; BP, blood pressure; COLD, chronic obstructive lung disease; CT, computed tomography; CTA, CT angiography; Cran CTA, CT angiography of arteries from aortic arch to vertex; CVI, cerebrovascular insult; Dm, diabetes mellitus; DVT, dep vein thrombosis; ECG, electrocardiogram; HT, arterial hypertension; ICH, intracerebral hemorrhage; LP, lumbar puncture; LVH, left ventricular hypertrophy; MI, myocardial infarction; MMSE, Mini mental status examination; MoCA, Montreal Cognitive Assessment; MRI, magnetic resonance imaging; mut, mutated; ND, not determined; Obe, obesity; OH, orthostatic hypotension; SDH, subdural hematoma; Smo, smoking; TTE, transthoracic echography of the heart; WMH, white matter hyperintensities; WT, wild type; y, years;

This table does not include individuals from generation IV whose status as clinically affected or unaffected remains undetermined.

Abbreviations in table e1:

AF, atrial fibrillation; Angio, angiography; ANS, autonomous nervous system; Ap, angina pectoris; BP, blood pressure; COLD, chronic obstructive lung disease; CT, computed tomography; CTA, CT angiography; Cran CTA, CT angiography of arteries from aortic arch to vertex; CVI, cerebrovascular insult; Dm, diabetes mellitus; DVT, dep vein thrombosis; ECG, electrocardiogram; HT, arterial hypertension; ICH, intracerebral hemorrhage; LP, lumbar puncture; LVH, left ventricular hypertrophy; MI, myocardial infarction; MMSE, Mini mental status examination; MoCA, Montreal Cognitive Assessment; MRI, magnetic resonance imaging; mut, mutated; ND, not determined; Obe, obesity; OH, orthostatic hypotension; SDH, subdural hematoma; Smo, smoking; TTE, transthoracic echography of the heart; WMH, white matter hyperintensities; WT, wild type; y, years;

This table does not include individuals from generation IV whose status as clinically affected or unaffected remains undetermined.

supplemental data e1: Genetic examinations of the proband and family

Whole exome and subsequently whole genome sequencing data of the **proband III.18** was analyzed for potentially disease-causing variants in the following genes implicated in known monogenic disorders with cerebral vasculopathy:

supplemental data e1 – table 1:

Location	Inheritance	Gene	Associated main neurological phenotype acc. to OMIM	Association with cSVD	Documented stroke§
Genes for cerebral small vessel disease from Stroke Gene Panels 1 and 2 (ref. [10] in main article)					
3p21.31	AD	TREX1	CRV	cSVD (with WMH, lacunar)	yes
20q13.12	AD	CTSA	CARASAL	cSVD (with WMH, ICH)	yes
10q26.13	AD	HTRA1	CARASIL /(CADASIL type2)	cSVD (with WMH, lacunar)	yes
19p13.12	AD	NOTCH3	CADASIL	cSVD (with WMH, lacunar)	yes
13q34	AD	COL4A1	Brain SVD	cSVD (with WMH, lacunar, ICH)	yes
13q34	AD	COL4A2	Brain SVD	cSVD (with WMH, lacunar, ICH)	yes
Xq22.1	X-L	GLA	Fabry disease	cSVD (with WMH, lacunar)	yes
20p11.21	AD	CST3	CAA	cSVD (ICH)	yes
13q14.2	AD	ITM2B	CAA	cSVD (with WMH, stroke-like)	yes
21q21.3	AD	APP	CAA	cSVD (with WMH, ICH)	yes
18q12.1	AD	TTR	CAA	cSVD (with WMH, ICH)	yes
22q11.1	AR	CECR1	Sneddon syndrome	cSVD (with WMH)	yes
16p13.11	AR	ABCC6	Pseudoxanthoma elasticum	cSVD (with WMH, lacunar)	yes
1p34.1	AR	MMACHC	Methylmalonic aciduria/ homocystinuria	cSVD (with WMH)	yes
1p36.22	AD	MTHFR	Homocystinuria	cSVD (with WMH, lacunar)	yes
21q22.3	AR	CBS	Homocystinuria	cSVD (with WMH, lacunar)	yes
6p12.3	AD/AR	MUT	Methylmalonic aciduria	cSVD (with WMH)	yes
17q25.3	AR	GAA	Glycogen storage disease II	cSVD (with ICH)	yes
Xq21.1	X-LR	ATP7A	Menkes disease	cSVD (with WMH)	yes
6p25.3	nd/AD	FOXC1	Axenfeld-Rieger syndrome	cSVD	yes
12p12.2	AD	PDE3A	Hypertension and brachydactyly	cSVD (secondary to HBP*)	yes
16q22.1	AD	HSD11B2	Apparent mineralocorticoid excess	cSVD (secondary to HBP)	yes
10q23.31	AD	ACTA2	Moya Moya 5	cSVD (with WMH)	yes
10q21.1	AD	PRKG1	Aortic aneurysm, familial thoracic 8	cSVD*	no
8q24.3	AD/AR	CYP11B1	Aldosteronism	cSVD*(-HBP)	no
16p12.2	AD	SCNN1B	Pseudohypoaldosteronism	cSVD*(-HBP)	no
16p12.2	AD	SCNN1G	Pseudohypoaldosteronism	cSVD*(-HBP)	no
Genes for monogenic forms of SVD/WMH identified since the publication of Stroke Gene Panels (ref. [10] in main article)					
19p13.11	AR	COLGALT1	Brain SVD	cSVD	yes (2 cases)[1]
17p13.1	AR	CTC1	Cerebroretinal microangiopathy	presumed cSVD	yes (2 cases)[2]
17p13.1	AR	SNORD118	Leukoencephalopathy	presumed cSVD	yes (2 cases)[3]

5q32	AD	CSF1R	Leukoencephalopathy	not considered vascular	yes (3 cases)[4]
Genes for cerebral amyloid angiopathy not known to cause stroke					
20p13	AD	PRNP	CAA	cSVD	no
9q33.2	AD	GSN	CAA	cSVD	no
Monogenic causes of Familial basal ganglia calcification					
1q25.3	AD	XPR1	IBGC	presumed vascular	no
5q32	AD	PDGFRB	IBGC	presumed vascular	no
8p11.21	AD	SLC20A2	IBGC	presumed vascular	no
22q13.1	AD	PDGFB	IBGC	presumed vascular	no
9p13.3	AR	MYORG	IBGC	presumed non-vascular	no

[§]Documented stroke in at least one carrier of a pathogenic variant in this gene, as defined in reference 6 of the main article. *, no clinical case with this phenotype; AD, autosomal dominant; AR, autosomal recessive; CAA, cerebral amyloid angiopathy; cSVD, cerebral small vessel disease; HBP, high blood pressure; ICH, intracerebral hemorrhage; LAA, large artery atherosclerosis; LAN, large artery non-atherosclerotic; nd, not determined; PMC, PubMed Central PMID, PubMed identification number; WMH, white matter hyperintensities; X-LR, X-chromosome linked recessive.

Genes associated with small vessel disease from our previously published stroke gene panel SGP1 and 2 (reference 6 in the main article), and additional 4 possible stroke-genes that more recently were associated with stroke episodes in humans are analyzed. Genes associated only with intracerebral cavernoma were not analyzed. Basal ganglia calcifications-associated mutations are also known to cause cerebrovascular disease, but to date they are not associated with clinically manifest stroke. The brain pathology phenotype reported for *CSF1R*, *SNORD118* and *CTC1* is different, showing cysts and spheroids. These were not identified at the post mortem examination of our patient, when specifically investigated for this purpose.

In **whole exome sequencing** data, no likely pathogenic/pathogenic variant (according to ACGM/ACP criteria) was found with an allele frequency of below 0.01.

Whole genome sequencing revealed the following variants in the (genomic regions of the) genes implicated in known monogenic disorders with cerebral vasculopathy (listed above). Here, we used a cutoff minor allele frequency of 0.05 to also assess low-frequency variants.

supplemental data e1 – table 2:

Gene name	Variant	Impact	Frequency in gnomAD genomes from Non-Finnish Europeans	Frequency in Swefreq	CADD_phred	Human Splicing Finder
PDE3A	ENST00000359062.3: c.961-4041C>T [intron 2-3], rs117966728	MODIFIER	0,033191	0,034	1,382	Creation of an intronic ESE site. Probably no impact on splicing.
COL4A2	ENST00000360467.5: c.649-294A>G [intron 10-11], rs184039734	MODIFIER	0,012395	0,0085	0,312	Activation of an intronic cryptic acceptor site. Potential alteration of splicing. Creation of an intronic ESE site. Probably no impact on splicing.
COL4A2	ENST00000360467.5: c.685-177T>C [intron 11-12], rs74124328	MODIFIER	0,046090	0,034	1,403	Creation of an intronic ESE site. Probably no impact on splicing.
COL4A2	ENST00000360467.5: c.727-390A>T [intron 12-13], rs192501203	MODIFIER	0,010725	0,007	1,224	No significant splicing motif alteration detected. This mutation has probably no impact on splicing.
COL4A2	ENST00000360467.5: c.1432+494A>G [intron 21-22], s182203175	MODIFIER	0,007420	0,008	6,969	Creation of an intronic ESE site. Probably no impact on splicing.
COL4A2	ENST00000360467.5: c.1433-549C>T [intron 21-22], rs186477441	MODIFIER	0,007287	0,008	3,271	No significant splicing motif alteration detected. This mutation has probably no impact on splicing.
COL4A2	ENST00000360467.5: c.1776+449G>A [intron 24-25], rs184033102	MODIFIER	0,007403	0,008	0,571	No significant splicing motif alteration detected. This mutation has probably no impact on splicing.
COL4A2	ENST00000360467.5: c.1776+497C>T [intron 24-25], rs189013305	MODIFIER	0,007397	0,008	0,256	Creation of an intronic ESE site. Probably no impact on splicing.
COL4A2	ENST00000360467.5: c.1776+1145T>A [intron 24-25], rs145381623	MODIFIER	0,007397	0,008	1,604	Creation of an intronic ESE site. Probably no impact on splicing.
COL4A2	ENST00000360467.5: c.1777-204G>A [intron 24-25], rs185724740	MODIFIER	0,007335	0,008	1,396	Creation of an intronic ESE site. Probably no impact on splicing.
COL4A2	ENST00000360467.5: c.1979-6C>T [intron 25-26], rs190632602	LOW	0,005424	0,008	4,465	No significant splicing motif alteration detected. This mutation has probably no impact on splicing.
COL4A2	ENST00000360467.5: c.2038+75A>G [intron 26-27], rs185332559	MODIFIER	0,007461	0,008	0,465	No significant splicing motif alteration detected. This mutation has probably no impact on splicing.
COL4A2	ENST00000360467.5: c.2759-245C>G [intron 31-32], rs74468721	MODIFIER	0,012653	0,027	2,982	Activation of an intronic cryptic donor site. Potential alteration of splicing. Creation of an intronic ESE site. Probably no impact on splicing.
COL4A2	ENST00000360467.5: c.2759-74C>G [intron 31-32], rs78589207	MODIFIER	0,012593	0,028	0,034	Alteration of WT Branch Point. Potential alteration of splicing
ITM2B	ENST00000378565.5: c.715+167G>A [intron 5-6], rs9332295	MODIFIER	0,013800	0,01	0,145	Alteration of an intronic ESS site. Probably no impact on splicing.

HSD11B2	NC_000016.10: g.67431114G>A, rs56265397	MODIFIER	0,010681	0,018	7,307	
GAA	ENST00000302262.3: c.271G>A p.(Asp91Asn) [**exon 2**], rs1800299	MODERATE	0,034974	0,03	25,6	Alteration of an exonic ESE site. Potential alteration of splicing.
CTC1	NC_000017.10: g.8130048C>T rs3027250	MODIFIER	0,039026	0,033	2,078	
NOTCH3	ENST00000263388.2: c.*837G>A [**exon 33**], rs12082	MODIFIER	0,026137	0,024	3,778	Alteration of an exonic ESE site. Potential alteration of splicing.

All variants were heterozygous. One exonic variant in the *GAA* gene received a CADD score above 20. Bi-allelic variants in *GAA* cause Glycogen storage disease II (Pompe disease), but this variant was relatively common, has been reported in ClinVar as benign (https://www.ncbi.nlm.nih.gov/clinvar/variation/4020/). Furthermore, heterozygous *GAA* variants are not known to cause Pompe disease, and we consider the clinical phenotype in the present family not compatible with Pompe disease, although both may have white matter hyperintensities on MRI. The table above shows a summary of the variants and only some of the prediction tools used.

Detailed description of the bioinformatics pipeline used for WGS analyses:

The obtained VCF file was annotated with VEP v98.2.[5] The GRCh37/hg19 reference sequence was used. Variant frequencies for non-Finnish Europeans and Swedish were retrieved from gnomAD genomes and exomes database[6] and SweGen Variant Frequency Dataset of SweFreq[7] respectively. Only variants in the genes of the list above were selected and analyzed further.

Variants with low genotype quality and read depth were filtered out. Only rare variants (MAF<1% or absent from databases with known variants) and variants of low frequency (MAF<5%) in the Swedish population were selected.

Non coding variants were prioritized using CADD[8], GWAVA[9] and FATHMM-MKL.[10,11]

Genomic Evolutionary Rate Profiling (GERP) scores were also obtained.[12]

In order to determine if a variant may disrupt splicing motifs and if it is predicted to affect splicing, Human Splicing Finder (HSF version 3.1)[13] and SpliceAI (version 1.3)[14] were used.

Disruption of transcription regulatory motifs and DNase I hypersensitive sites that represent open chromatin regions accessible to transcription factors were also investigated. wgEncodeRegTfbsClusteredV3 table was downloaded from TFBS Clusters track which derived from a large collection of ChIP-seq experiments.[15-17] wgEncodeRegDnaseClusteredV3 table was downloaded from DNAse Clusters track which contains DNaseI Hypersensitive Sites .[18] Both tables are stored in the UCSC genome browser.[19-21]

Coding variants were further annotated by using a number of in silico tools like SIFT,[22] PolyPhen2,[23] LRT,[24] Mutation Taster,[25] Mutation Assessor,[26] FATHMM,[11] PROVEAN,[27] VEST3,[28,29] REVEL[30] (data not shown).

Co-segregation analyses of WES data from eight family members: Identification of *MAP3K6* c.322G>A p.(Asp108Asn) variant.

All heterozygous variants with a frequency of below 0.05 in databases were selected from the WES datasets. They were combined in the following order:

supplemental data e1 – table 3:

Step 1: Variants shared by III.17, III. 18, and III.20	661 variants
Step 2: Variants from Step 1 shared by III.19	487 variants
Step 3: Variants from Step 2 NOT in III.24	219 variants
Step 4: Variants from Step 3 shared by III.22	37 variants
Step 5: Variants from Step 4 shared by III.11	12 variants
Step 6: Variants from Step 5 shared by III.4	11 variants (table below)

At Step 6, the following 11 variants remained:

supplemental data e1 – table 4:

Chromosome-Position hg19	Gene / transcript	Nucleotide change, location	Protein change	Frequency in Exac	Frequency in Swegen	Sift	Poly-phen
1-24394835	MYOM3 NM_152372.3	c.3173G>C **exonic**	**p.(Arg1058Pro)**	0.00053	0.00100	0.02	0.08
1-27692767	**MAP3K6 NM_004672.4**	**c.322G>A exonic**	**p.(Asp108Asn)**	**ND**	**ND**	**0.00**	**1.00**
1-144916676*	NBPF20 NM_001278267.1| NBPF9 NM_001277444.1| PDE4DIP NM_001198834.3	c.5357+91920C>T intronic | c.2583+88748C>T intronic | c.1679G>A **exonic**	- | - | **p.(Trp560Ter)**	ND	ND	**0.00**	0.00
1-144922523*	NBPF20 NM_001278267.1| NBPF9 NM_001277444.1| PDE4DIP NM_001198834.3	c.5357+97767C>T intronic | c.2583+94595C>T intronic | c.884G>A **exonic**	- | - | **p.(Arg295His)**	ND	ND	**0.00**	0.00

6-2890836*	LOC101927730 NR_110841.1| SERPINB9 NM_004155.5	intronic_nc| c.724-2C>T intronic	- | -	0.05000	0.09000	0.00	0.00
10-8106135	GATA3 NM_001002295.1	c.924+34A>G intronic	-	0.02800	0.02100	0.00	0.00
19-5245977	PTPRS NM_002850.3	c.798C>T **exonic**	p.(=)	0.02000	0.03300	0.00	0.00
19-5604875	SAFB2 NM_014649.2	c.1369A>**G exonic**	**p.(Thr457Ala)**	0.00724	0.01400	**0.01**	0.12
19-10081951	COL5A3 NM_015719.3	c.3775-17A>C intronic	-	0.01600	0.04500	0.00	0.00
19-11031712	CARM1 NM_199141.1	c.1538-14C>T intronic	-	0.02500	0.03900	0.00	0.00
19-11170680	SMARCA4 NM_001128849.1	c.4865-41C>G intronic	-	0.00003	-1.00000	0.00	0.00

*Since the time of original co-segregation analyses, genomic databases have increased and with renewed annotation using gnomAD v2.0 exomes and genomes, the three variants at positions 1-144916676, 1-144922523, and 6-2890836 were excluded as they were too frequent in gnomAD.

Of these, *MAP3K6* c.322G>A p.(Asp108Asn) was the only one that was located in an exon and was predicted to have a highly deleterious effect in both the Sift (low values indicate deleteriousness, lowest possible: 0.0) and PolyPhen (high values indicate deleteriousness, highest possible: 1.0) prediction tools.

Analyses were performed independently by three bioinformaticians (E.K., S.S. and K.T.) at different centers, starting with variant annotation step. All three identified this variant as the topmost candidate. We also analyzed the data using III.23 instead of III.24 as unaffected, and found no additional variant that we considered potentially pathogenic (data not shown).

In silico prediction of the effect of the *MAP3K6* c.322G>A p.(Asp108Asn) variant was also evaluated by additional tools:

supplemental data e1 – table 5:

Prediction tool	**Score**	**Prediction**
SIFT	0.01	deleterious
Polyphen2_HDIV	1.0	deleterious
Polyphen2_HVAR	0.996	deleterious
MutationTaster	1.000	deleterious
MutationAssessor	2.195	moderate
FATHMM	-0.9	tolerated
RadialSVM	0.227	deleterious
LR	0.573	deleterious
VEST3	0.252	--
CADD_raw	5.193	--
CADD_phred	32	--
GERP++_RS	4.76	--
phyloP46way_placental	2.457	--
phyloP100way_vertebrate	4.704	--
SiPhy_29way_logOdds	17.535	--

Determining the shared haplotype around *MAP3K6* c.322G>A in affected members

Data were phased and missing genotypes were imputed by using Beagle 5.1.[31,32] Then identity-by-descent (IBD) segments were detected by using hap-id program.[33] The smallest segment containing the variant of interest found to be shared across any two affected members was 1.39 Mb long (chr1:26499281-27889773), spanning 36 genes.

Analysis of sequence alterations, short tandem repeats and copy number variants within the shared haplotype around *MAP3K6* c.322G>A

Within this 1.39Mb shared segment, we analyzed WGS data from III.18, looking for any other variant that may not have been detected by WES but that may be potentially disease-causing. WGS data was analyzed for sequence alterations, copy number variants and short tandem repeats with an algorithm that is also used for clinical genetic testing at the Dept. of Clinical Genetics, Lund.

For sequence alterations, we used the same annotation and filtering steps as for WES data. For the detection of copy number variants, variants were called using a combination of CNVnator[34], TIDDIT[35] och Manta[36]. Locally developed software SVBD was used to combine all variants that showed a 70% or more overlap between these tools. Detected variants were annotated and ranked in a pipeline with the following tools: VEP[37], AnnotSV[38], Pre-score (local Pearlscript), Genmod and Compound finder (local Pearlscript). Short tandem repeats were analysed using ExpansionHunter (Illumina, https://github.com/Illumina/ExpansionHunter) and annotated using Stranger (https://github.com/moonso/stranger).

We found the following single nucleotide variants and small indels (with MAF<0.05):

supplemental data e1 – table 6:

Position (hg38)	Change	Gene	Variant	Function	dbSNP	PopFreq	CADD	Also detected by WES	Co-segregating in family?
1-26473999	C→T	HMGN2 NM_005517	c.91-86C>T	ncRNA_exonic	rs114289390	0.0139	0.35	YES	NO
1-26890065	C→CTTCG	GPN2 NM_018066	c.31_32insCGAA	p.Gly11AlafsTer22	rs770679034	~	28.0	YES	NO
1-27355361	T→C	MAP3K6 NM_004672	c.*30A>G	3UTR	rs201154544	0.0004	2.0	YES	YES*
1-27366276	C→T	MAP3K6 NM_004672	c.322G>A	p.Asp108Asn	rs947285063	~	29.0	**YES**	**YES**
1-27534293	A→C	AHDC1 XM_024446461	c.*667T>G	3UTR	NA	0.0005	19.0	NO	NA
1-28690203	G→GTT	GMEB1 NM_006582	c.241+37_241+38dup	ncRNA_exonic	rs878890649	~	3.0	NO	NA

~ not reported. * MAP3K6 c.*30A>G in the 3' untranslated region was not detected in our co-segregation analyses where only exonic and splice site variants were considered. It fully co-segregates with MAP3K6 c.322G>A in all individuals analyzed by NGS, marking a common haplotype. We consider MAP3K6 c.322G>A the more likely disease cause.

We found 6 larger deletions and 1 larger insertion in this region of chromosome 1:

supplemental data e1 – table 7:

Type	Start loc (hg38)	Stop loc (hg38)	Length	Region	Function	Frequency	Gene(s)	Comment:
DEL	25264901	25335500	70600	RSRP1:5UTR	RSRP1:5_prime_UTR_variant		RSRP1	Homozygous (copy number = 0). Spans entire RHD gene.
				RHD:exonic	RHD:transcript_ablation	0.3814	RHD	
				SDHDP6:exonic	SDHDP6:transcript_ablation		SDHDP6	
DEL	26282320	26282398	78	exonic	inframe_deletion	~	UBXN11	Seen in several *in house* samples.
DEL	26803744	26803842	98			~		Low coverage. No gene.
DEL	26826601	26827000	400	exonic	coding_sequence_variant	~	ZDHHC18	Overlaps with exon1 of ZDHHC18.
DEL	28294766	28295236	470			~		Technical artefact.
INS	29100326	29100657	331	ncRNA_exonic	non_coding_transcript_variant	~	EPB41	Uncertain*
DEL	30908401	30908900	500	exonic	coding_sequence_variant	~	SDC3	Exon 1 of SDC3.

*Algorithm insufficiently evaluated for insertions. ~ not reported.

There was one whole-gene deletion of RHD, which we consider not relevant as there is normal genetic variation of blood group antigens, including deletions.

None of these variants were considered to be a more likely cause for the disease in this family than *MAP3K6* c.322G>A p.(Asp108Asn).

We did not identify any short tandem repeat.

table e2: Variants in *MAP3K6* in WES data from 22 probands with young-onset familial stroke

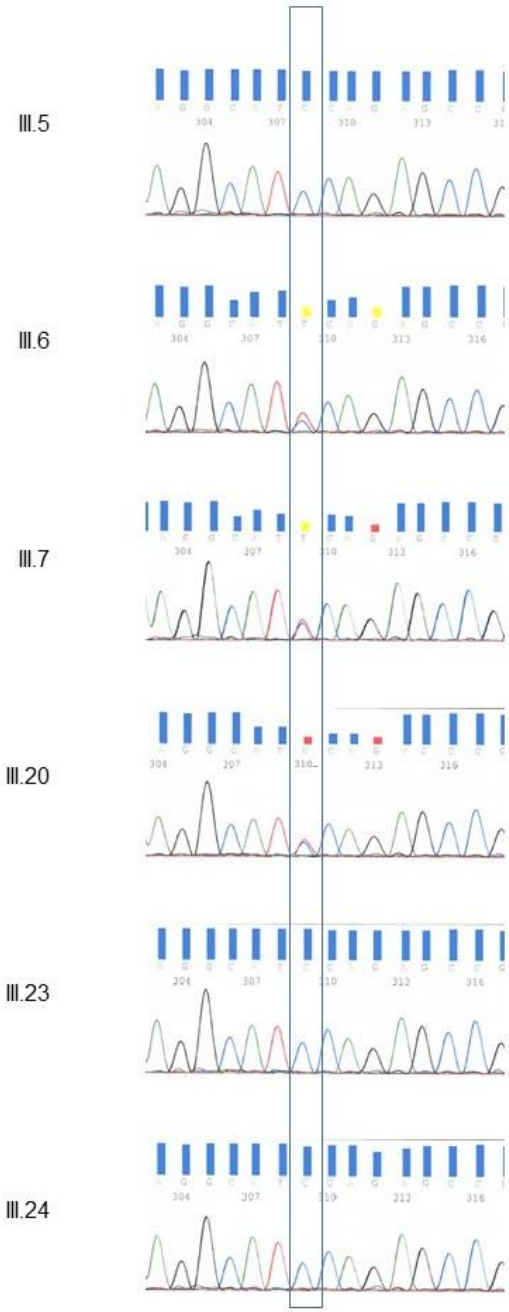
Position (hg17)	1-27692767-C-T	1-27689154-A-G	1-27688743-T-C	1-27685067-G-A
MAP3K6 transcript and variant	NM_001297609 and NM_004672 c.G322A p.(Asp108Asn) [exon 1]	NM_001297609 c.T1209C p.(Asp403Asp) [exon 7], NM_004672 c.T1233C p.(Asp411Asp) [exon 8]	NM_004672 c.1256-2A>G [exon9], NM_001297609 c.1232-2A>G [exon8]	NM_001297609 c.C2595T p.(Ser865Ser) [exon 19], NM_004672 c.C2619T p.(Ser873Ser) [exon 20]
Genotype	heterozygous	heterozygous	heterozygous	heterozygous
Location	**exonic**	**exonic**	**intronic splice site**	**exonic**
Variant type	**nonsynonymous SNV**	**synonymous SNV**	**splicing**	**synonymous SNV**
GnomAD	0	0.0000922	0.00372	0.000007128
1000G_ALL	.	.	0.002	.
1000G_EUR	.	.	0.007	.
ExAC_Freq	.	0.0001	0.0038	.
ExAC_NFE	.	0	0.0046	.
dbSNP	.	.	rs55841735	.
CADD_phred	32	7.727	19.25	9.021
Individual	*Family reported here*	Proband of Family 9 in Ilinca et al. 2020 (reference)	Proband of Family 3 and of Family 12 in Ilinca et al. 2020 (reference)	Proband of Family 20 in Ilinca et al. 2020 (reference)
Subtype of first-ever stroke according to classification in ref. [14]	SVD, ICH	SVD	SVD (both)	Coagulation defect
Presence of WMH	Y	Y	Y (in both)	N

The 22 probands with young-onset familial stroke from Sweden and Finland (from reference 12 in the main manuscript) carried a total of 104 variants in the *MAP3K6* gene; 31 different variants, each detected in 1-19 individuals. Three variants had an allele frequency of below 0.01 and were located in exons or splice site; these are shown above.

Of these, the splice site variant NM_004672 c.1256-2A>G [exon9], and NM_001297609 c.1232-2A>G [exon8] had the highest CADD-phred score, but DNA was unavailable from additional affected members from the two families where it was detected and further analyses have not been possible.

SNV, single nucleotide variant.

figure e1: Electropherograms of Sanger sequencing for MAP3K6 c.322G>A



For individuals III.20 and III.23, Sanger sequencing confirmed WES results. Four additional members were only examined for this variant after the *MAP3K6* variant was identified by co-segregation analyses.

III.6, III.7 and III.20 are heterozygous carriers of *MAP3K6* c.322G>A p.(Asp108Asn), whereas the other individuals are homozygous for wild type c.322G.

supplemental data e2: VEGF measurements in serum

Results of measurements:

Cases: Affected family members carrying MAP3K6 c.322G>A	VEGF levels (pg/ml)	**Controls:** Age- and sex-matched controls	VEGF levels (pg/ml)
Case 1	312	Control 1	387
Case 2	207	Control 2	632
Case 3	306	Control 3	415
Case 4	367	Control 4	472
Case 5	680	Control 5	247
Case 6	456	Control 6	324
Case 7	715	Control 7	237
Case 8	604	Control 8	172
Mean (±SD)	**456 (±178)**	Mean (±SD)	**361 (±139)**

Statistical evaluation: Mann-Whitney Test

Ranks

	VAR2	N	Mean Rank	Sum of Ranks
VAR1	Controls	8	7,5	60
	Cases	8	9,5	76
	Total	16		

Test Statistics[a]

	VAR1
Mann-Whitney U	24
Wilcoxon W	60
Z	-0,840
Asymp. Sig. (2-tailed)	0,401
Exact Sig. [2*(1-tailed Sig.)]	0,442[b]

a. Grouping Variable: VAR2

b. Not corrected for ties.

references to supplemental material e1:

1. Miyatake S, Schneeberger S, Koyama N. Biallelic COLGALT1 Variants Are Associated With Cerebral Small Vessel Disease. Ann Neurol 2018;84:843-853.

2. Xu W, Zhao J, Zhu Y, Zhang W. Cerebroretinal microangiopathy with calcifications and cysts. A case report. Medicine (Baltimore) 2017; 96: e5545.

3. Osman O ,Labrune P, Reiner P. Leukoencephalopathy With Calcifications and Cysts (LCC): 5 Cases and Literature Review. Rev Neurol (Paris) 2020;176:170-179.

4. Battisti C, Di Donato I, Bianchi S, et al. Hereditary Diffuse Leukoencephalopathy With Axonal Spheroids: Three Patients With Stroke-Like Presentation Carrying New Mutations in the CSF1R Gene. J Neurol. 2014;261:768-72.

5. McLaren W, Gil L, Hunt SE, et al. The Ensembl Variant Effect Predictor. Genome Biology 2016;17:122.

6. Karczewski KJ, Francioli LC, Tiao G, et al. Variation across 141,456 human exomes and genomes reveals the spectrum of loss-of-function intolerance across human protein-coding genes. BioRxiv. Aug 13 2019. doi: https://doi.org/10.1101/531210

7. Ameur A, Dahlberg J, Olason P, et al. SweGen: a whole-genome data resource of genetic variability in a cross-section of the Swedish population. Eur J Hum Genet 2017;25:1253-1260.

8. Kircher M, Witten DM, Jain P, O'Roak BJ, Cooper GM, Shendure J. A general framework for estimating the relative pathogenicity of human genetic variants. Nat Genet 2014;46:310-5.

9. Ritchie GR, Dunham I, Zeggini E, Flicek P. Functional annotation of noncoding sequence variants. Nat Methods. 2014;11:294-6.

10. Shihab HA, Rogers MF, Gough J, et al. An Integrative Approach to Predicting the Functional Consequences of Non-coding and Coding Sequence Variation. Bioinformatics 2015;31:1536-43.

11. Shihab HA, Gough J, Cooper DN, et al. Predicting the Functional, Molecular and Phenotypic Consequences of Amino Acid Substitutions using Hidden Markov Models. Hum Mutat 2013;34:57-65.

12. Cooper GM, Stone EA, Asimenos G, Green ED, Batzoglou S, Sidow A. Distribution and intensity of constraint in mammalian genomic sequence. Genome Res 2005;15:901-13.

13. Desmet FO, Hamroun D, Lalande M, Collod-Beroud G, Claustres M, Beroud C. Human Splicing Finder: an online bioinformatics tool to predict splicing signals. Nucleic Acids Res 2009;37:e67.

14 Jaganathan K, Kyriazopoulou Panagiotopoulou S, McRae JF, et al. Predicting Splicing from Primary Sequence with Deep Learning. Cell 2019;176:535-548.

15. Gerstein MB, Kundaje A, Hariharan M, et al. Architecture of the human regulatory network derived from ENCODE data. Nature 2012;489:91-100.

16. Wang J, Zhuang J, Iyer S, et al. Sequence features and chromatin structure around the genomic regions bound by 119 human transcription factors. Genome Res 2012;22:1798-812.

17. Wang J, Zhuang J, Iyer S, et al. Factorbook.org: a Wiki-based database for transcription factor-binding data generated by the ENCODE consortium. Nucleic Acids Res 2013;41(Database issue):D171-6.

18. Thurman RE, Rynes E, Humbert R, et al. The accessible chromatin landscape of the human genome. Nature. 2012;489:75-82.

19. Kent WJ, Sugnet CW, Furey TS, et al. The human genome browser at UCSC. Genome Res. 2002;12:996-1006.

20. Rosenbloom KR, Armstrong J, Barber GP, et al. The UCSC Genome Browser database: 2015 update. Nucleic Acids Res. 2015;43(Database issue):D670-81.

21. Rosenbloom KR, Sloan CA, Malladi VS, et al. ENCODE data in the UCSC Genome Browser: year 5 update. Nucleic Acids Res. 2013;41(Database issue):D56-63.

22. Kumar P, Henikoff S, Ng PC. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. Nat Protoc 2009;4:1073-81.

23. Adzhubei IA, Schmidt S, Peshkin L, et al. A method and server for predicting damaging missense mutations. Nat Methods 2010;7:248-9.

24. Chun S, Fay JC. Identification of deleterious mutations within three human genomes. Genome Res 2009;19:1553-61.

25. Schwarz JM, Rodelsperger C, Schuelke M, Seelow D. MutationTaster evaluates disease-causing potential of sequence alterations. Nat Methods 2010;7:575-6.

26. Reva, BA, Antipin, YA. Sander, C. Determinants of protein function revealed by combinatorial entropy optimization. Genome Biol 2007;8:R232.

27. Choi Y, Chan AP. PROVEAN web server: a tool to predict the functional effect of amino acid substitutions and indels. Bioinformatics 2015;31:2745-7.

28. Carter H, Douville C, Yeo G, Stenson PD, Cooper DN, Karchin R. Identifying Mendelian disease genes with the Variant Effect Scoring Tool. BMC Genomics 2013;14:1-16.

29. Douville C, Masica DL, Stenson PD, et al. Assessing the pathogenicity of insertion and deletion variants with the Variant Effect Scoring Tool (VEST-indel). Hum Mutat 2016;37:28-35.

30. Ioannidis NM, Rothstein JH, Pejaver V et al. REVEL: An Ensemble Method for Predicting the Pathogenicity of Rare Missense Variants. Am J Hum Genet. 2016;99:877-885

31. Browning S.R. and Browning B.L. Rapid and accurate haplotype phasing and missing data inference for whole genome association studies by use of localized haplotype clustering. Am J Hum Genet. Nov 2007;81(5):1084-1097.

32. Browning B.L, Zhou Y, Browning S.R. A one-penny imputed genome from next generation reference panels. Am J Hum Genet. Sep 6 2018;103(3):338-348.

33. Kopanos C, Tsiolkas V, Kouris A, Chapple CE, Aguilera MA, Meyer R, Massouras A. VarSome: The Human Genomic Variant Search Engine. Oxford Bioinformatics. Oct 30 2018;35(11):1978-1980.

34. Abyzov A. et al., CNVnator: an approach to discover, genotype, and characterize typical and atypical CNVs from family and population genome sequencing. Genome Res. 2011;21:974-84.

35. Eisfeldt J. et al., TIDDIT, an efficient and comprehensive structural variant caller for massive parallel sequencing data. F1000Res. 2017;6:664.

36. Chen X. et al., Manta: rapid detection of structural variants and indels for germline and cancer sequencing applications. Bioinformatics. 2016;32:1220-2.

37. Ensembl Variant Effect Predictor (VEP): https://www.ensembl.org/info/docs/tools/vep/index.html

38. Geoffroy V et al., AnnotSV: an integrated tool for structural variations annotation. Bioinformatics. 2018;34:3572-3574. https://lbgi.fr/AnnotSV