



A stroke gene panel for whole-exome sequencing

Andreea Ilinca^{1,2} · Sofie Samuelsson³ · Paul Piccinelli³ · Maria Soller^{3,4} · Ulf Kristoffersson³ · Arne G. Lindgren^{1,5}

Received: 26 April 2018 / Revised: 19 August 2018 / Accepted: 28 August 2018 / Published online: 24 October 2018
© European Society of Human Genetics 2018

Abstract

Extensive analyses of known monogenic causes of stroke by whole-exome/genome sequencing are technically possible today. We here aimed to compile a comprehensive panel of genes associated with monogenic causes of stroke for use in clinical and research situations. We systematically searched the publically available database Online Mendelian Inheritance in Man, and validated the entries against original peer-reviewed publications in PubMed. First, we selected known pathogenic or putatively pathogenic stroke genes reported in at least one person with stroke, and classified the stroke phenotype for each gene into eight subgroups: (1) large artery atherosclerotic, (2) large artery non-atherosclerotic (tortuosity, dolichoectasia, aneurysm, non-atherosclerotic dissection, occlusion), (3) cerebral small-vessel diseases, (4) cardioembolic (arrhythmia, heart defect, cardiomyopathy), (5) coagulation dysfunctions (venous thrombosis, arterial thrombosis, bleeding tendency), (6) intracerebral hemorrhage, (7) vascular malformations (cavernoma, arteriovenous malformations), and (8) metabolism disorders. Second, we selected other genes that may plausibly cause stroke through diseases related to stroke, but without any documented stroke patient description. A third section comprised SNPs associated with stroke in genome-wide association studies (GWAS). We identified in total 214 genes: 120 associated with stroke, 62 associated with diseases that may cause stroke, and 32 stroke-related genes from recent GWAS. We describe these 214 genes and the clinical stroke subtype(s) associated with each of them. The resulting gene panel can be used to interpret exome sequencing results regarding monogenic stroke. Based on the panel's clinical phenotype description, the pathogenicity of novel variants in these genes may be evaluated in specific situations.

Introduction

Stroke is a leading cause of disability worldwide [1]. Previous studies have shown that a genetic background is

significantly associated with stroke risk [2–7]. Family clustering occurs, which supports that a monogenic cause for the underlying disease may sometimes be present [8, 9]. To date, a monogenic cause has been identified only in a minority of families with clustering of stroke. In clinical practice, it is common that routine genetic testing is performed for only a minority of selected well-defined pathologies, for example, CADASIL. It is therefore likely that many other gene variations possibly related to familial aggregation of stroke currently remain undiagnosed, resulting in an under-representation of the contribution of these monogenic forms to overall genetic stroke risk.

The relation between genetic variation and different stroke subtypes is sometimes unclear. In some families, the phenotype related to stroke and a certain genetic variation is well defined. In other families, the genetic change can result in several different phenotypes both regarding stroke subtype and other clinical characteristics. Nevertheless, a systematic characterization of phenotypes and possibly related genotypes may be useful to identify genetic variation expressing different types of stroke.

Supplementary information The online version of this article (<https://doi.org/10.1038/s41431-018-0274-4>) contains supplementary material, which is available to authorized users.

✉ Andreea Ilinca
Andreea.Ilinca@med.lu.se

- ¹ Department of Clinical Sciences Lund, Neurology, Lund University, Lund, Sweden
- ² Department of Neurology and Rehabilitation Medicine, Skåne University Hospital, Malmö, Sweden
- ³ Department of Clinical Genetics and Pathology, Laboratory Medicine, Region Skåne and Lund University, Lund, Sweden
- ⁴ Department of Clinical Genetics, Karolinska University Hospital, Solna, Sweden
- ⁵ Department of Neurology and Rehabilitation Medicine, Skåne University Hospital, Lund, Sweden

Technical advances, including new testing methods like whole-exome sequencing and whole-genome sequencing, have accelerated the identification of new genes for several monogenic diseases, including stroke [10, 11], and today provide a possibility of economical and time-efficient investigation not only for the most well-known diseases but also for the less well characterized.

We aimed to review the present knowledge on reported genetic variations that may be related to Mendelian stroke, and to systematically connect the identified possible stroke genes with specific stroke subtypes and other specific clinical characteristics. We also aimed to summarize these data in a comprehensive panel of genes, providing practical information to be used both for stroke research purposes and for possible genetic testing strategies in clinical practice. The panel can be used in whole-exome/genome sequencing evaluation of monogenic stroke, and easily updated when new genes associated with stroke are detected in the future.

Methods

Systematic search for stroke genes

A systematic search into the publically available database Online Mendelian Inheritance in Man (OMIM) was conducted until 1 August 2017. OMIM is a comprehensive database of human genes and genetic phenotypes that is freely available and contains information on Mendelian disorders and over 15,000 genes. OMIM focuses on the relationship between phenotype and genotype, which makes the information provided ideal for the purpose of this objective. We systematically identified genes related to stroke in OMIM, by using six different combinations of search terms in the following sequential order: (1) (stroke), (2) (cerebrovascular), (3) (cerebral OR intracerebral OR intracranial OR brain OR encephalic) AND (infarct OR infarction OR ischemia OR ischemic), (4) (ischemic OR ischemic) AND (event OR stroke), (5) (transitory OR transient) AND (event OR ischemic OR ischemic), or (6) (intracranial OR cerebral OR intracerebral OR encephalic OR brain) AND (hemorrhage OR hemorrhage OR bleeding OR hematoma). The OMIM filtering function “Phenotype-only entries” was used. Only disorders with known molecular basis or contiguous gene duplication or deletion syndromes involving multiple genes were considered.

Detection and grouping of genes for the stroke gene panel

In the first section of a stroke gene panel, SGP1, genes were included that contain at least one variant for which a causative role has been shown or postulated and reported from

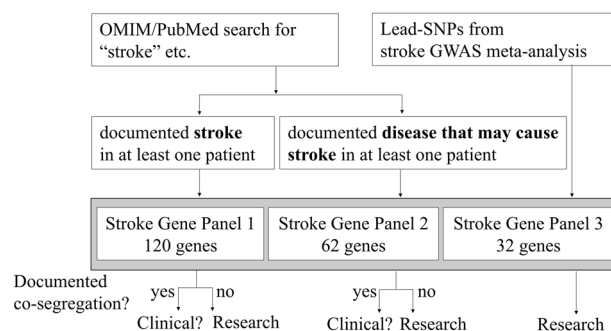


Fig. 1 Systematic identification of genes related to monogenic stroke, and inclusion of genes in panels. *OMIM* = Online Mendelian Inheritance in Man; *GWAS* = genome-wide association studies, *SNP* = single-nucleotide polymorphism

at least one well-documented human patient in the literature (Fig. 1). For this purpose, the literature cited in the OMIM entries was evaluated.

In a second section of the stroke gene panel, SGP2, we listed genes that were retrieved in our OMIM search for genes related to stroke, as outlined above, but where no documented human patients with stroke was identified; these genes were associated with a disease suggested to be related to stroke. The mechanism through which stroke could be expected to be caused was entered, as documented in the literature listed in OMIM. The disease caused by variants in this gene had to be documented in at least one patient in publications from PubMed (Fig. 1). Genes already included in the first SGP1 section were excluded from this second SGP2 section.

SGP1 and SGP2 include all the genes that are presently known to contain pathogenic variants for stroke. We define as pathogenic those variants that affect gene or protein function and that cause, or cause under certain circumstances (for example when penetrance is incomplete) a specific subtype of stroke.

A third section, SGP3, included genes located in the proximity of loci related to stroke risk identified in a recent meta-analysis of stroke genome-wide association studies (GWAS) [12].

Clinical evaluation

Clinical stroke subtypes

Relevant articles cited in OMIM database were assessed to identify the different possible stroke subcategories (Table 1), which were subsequently used to describe the known phenotype(s) for each gene in the panels. The most recent clinical reviews on each of these genetic disorders were also examined, as well as the reports of clinical cases of stroke related to the disease, as found in the references of

Table 1 Clinical stroke subgroups used in SPG1 and SPG2

LAA	Unspecified
	Hypercholesterolemia
	Hypertension
LAN	Tortuosity/dolichoectasia
	Dissection
	Occlusion: MoyaMoya-like/ fibromuscular dysplasia
VM	Aneurysm
	Arteriovenous Cavernoma
SVD	Unspecified: isolated lacunary infarct
	Deep white matter intensities
	Intracerebral hemorrhage/microbleeds
	Hypertension
CE	Arrhythmia: atrial fibrillation/flutter
	Morphological defect, such as patent foramen ovale
	Myopathy
Coag	Venous thrombosis
	Arterial thrombosis
	Bleeding
	Hyperviscosity
MB	Mitochondrial
	Defect of intermediary metabolism
ICH or subarachnoidal bleeding	Unspecified
	Specified: LAN/SVD/VM/Coag

LAA large artery atherosclerotic, LAN large artery non-atherosclerotic, VM vascular malformation, SVD small-vessel disease, CE cardioembolic, Coag coagulation, MB metabolic, ICH intracerebral hemorrhage

these articles. No structured critical analysis of the consulted articles is presented, as our aim was to compile relevant stroke genes and relevant clinical correlations between these genes and defined stroke subtypes, for analyzing the occurrence of familial clustering of stroke.

We matched each of the detected stroke genes with one or more specific subtype(s) of stroke to enhance the panel's usage for more specific genotype–phenotype evaluation. For this purpose, we used the main groups as in the Trial of Org 10172 in Acute Stroke Treatment (TOAST) or Causative Classification System for Ischemic Stroke (CSS) classifications [13, 14]: large artery atherosclerosis (LAA), small-vessel disease (SVD), and cardioembolic (CE), as well as specific forms of stroke that are summarized as “Other causes” in TOAST/CSS and that reflect a certain underlying pathomechanisms: coagulation (Coag) disturbances, vascular malformations (VMs), metabolic (MB) disorders, and large artery non-atherosclerotic (LAN, e.g., caliber changings of the large and medium arteries). Several monogenic forms of ischemic stroke (IS), particularly those causing structural blood vessel abnormalities, also may

cause cerebral hemorrhage, and therefore cerebral bleeding is included as a possible stroke subtype, although hemorrhagic stroke is not included in the TOAST/CSS classification (Table 1):

LAA also included possible underlying mechanisms for LAA, for example, defects in lipid metabolism, pronounced high blood pressure (HBP), and vessel dissection caused by atherosclerosis.

LAN comprised large or middle large wall vessel caliber changes not related to the systemic process of atherosclerosis or trauma, and included caliber-diminishing phenomena such as fibromuscular dysplasia and MoyaMoya (-like) disease, caliber dilatation as in dolichoectasia, aneurysms, and kinking or tortuosity, as well as large/middle LAN dissection.

Cerebral SVDs included lacunar infarcts with or without bleeding or microbleeds, white matter hyperintensities (WMH) or conditions related to HBP. The underlying mechanism was specified when possible.

CE strokes included arrhythmias (CE-A) as atrial fibrillation or flutter, cardiac morphological defect (CE-D) such as patent foramen ovale, or cardiomyopathy of any type (CE-M).

Coag anomalies were further subclassified as venous thrombosis (Coag-VT), arterial thrombosis (Coag-AT), or bleeding tendency (Coag-B).

MB phenotype included disorders caused by dysfunction of the mitochondrial respiratory chains or other specific intermediary metabolism diseases associated with stroke. These syndromes may involve other organs, causing hearing or vision impairment, myopathy, exercise intolerance or muscle weakness or cramps, cardiac defects, seizures, lactic acidosis, or hypoglycemia.

VM comprised disease of different types of vessels such as cavernoma (VM-C) or arteriovenous malformation (VM-AV) including pulmonary VM as a possible embolic source for stroke.

For intracerebral hemorrhage (ICH) the underlying mechanism is specified when known as ICH (LAN-aneurysm/dissection/MoyaMoya), ICH (SVD), ICH (Coag), or ICH (VM).

In SGP3 the clinical stroke subtype as described in the original article [12] was used and the results were classified into: any stroke, cardioembolic stroke, large artery stroke, any IS.

Relation of the genes with the clinical subtypes

For each of the detected genes in the OMIM search, the articles cited in OMIM were retrieved, and information on the stroke subtype(s) associated with variants in this gene

was listed in the panels SGP1 and SGP2. Further, PubMed database (until August 2017) was searched for each of the genes to identify any additional clinical description of carriers of pathogenic variants in this gene who may have displayed additional clinical stroke subtypes. When relevant information could be extracted from abstracts, these were used, otherwise the entire articles were reviewed. If there were any review articles on the gene, or if there were relevant entries in the GeneReviews database (www.genereviews.org), relevant additional information on the clinical subtypes were extracted from these as well.

If a gene had already been included in SGP1 based on association with a specific stroke subtype, and the same gene was subsequently—when using the SGP2 criteria—found to also be associated with another subtype, this gene was only included in SGP1. The additional SGP2 stroke subtype was then marked with an “*” in the SGP1 panel (Supplementary Table 1 and Supplemental material).

For each of the included genes, we also constructed a detailed list with organs and systems affected by the underlying disease as reported in the literature. This may enable a second matching procedure between patient’s or family’s phenotype and the gene panel SGP1 and SGP2. Other characteristics such as biological markers or brain image characteristics of a specific disease were specified as described in the literature, and are mentioned in the same column, in the panels (see online versions of SGP1 and SGP2).

We did not include age at stroke onset for each gene because we anticipated that the phenotypic variability regarding age at first stroke onset can be substantial.

Selections of genes for clinical versus research evaluations

The pathogenicity of genes included in SGP1 and SGP2 is not equally well documented in the literature. Co-segregation of genotype and phenotype within pedigrees is an important criterion for pathogenicity. Information regarding co-segregation of the putative pathogenic variants were collected from the reviewed literature in order to identify genes recommended for clinical testing.

For autosomal dominant inheritance, we considered suitable for clinical screening those genes where co-segregation of rare or very rare variants (minor allele frequency below 1% in the target population) related to disease have been described in either: (a) two or more unrelated pedigrees, with at least one of the pedigrees containing 10 or more affected individuals, of whom at least 2 had to be third degree or more remote relatives of the proband, or (b) three or more unrelated smaller pedigrees with at least two affected individuals each.

For autosomal recessive inheritance, we considered suitable for clinical screening those genes where

co-segregation of variants (with a minor allele frequency below 2% in the target population) related to disease has been described in (c) at least three unrelated pedigrees, with at least two of them containing two or more individuals with the disease.

Other considerations

The cytogenetic location, genetic coordinates and the name of the associated disease as specified in OMIM are provided in Table 3 and Supplementary Table 1. A detailed analysis of biological and pathological mechanisms possibly leading to stroke for each detected gene was considered to be out of the scope for this review and is therefore not addressed.

Results and discussion

Based on the systematic review of the OMIM database and identification of relevant stroke-causing genes as described in Table 2, 214 genes with possible relation to stroke were detected. Of these, 120 genes were associated with stroke (at least one patient with stroke carrying a pathogenic variant) and included in the SGP1 (Supplementary Table 1), 62 associated with pathologies that may cause stroke were included in SGP2 (Supplementary Table 1) and 32 additional genes previously reported in recent GWAS meta-analyses were included in SGP3 (Table 3).

These genes were further classified according to their relevance and scientific background. We mainly relied on the evidence of at least one clinically documented case of stroke in a variant carrier, and evidence of co-segregation as described in the subsection “Selections of genes for clinical versus research evaluations.” For the first time, a comprehensive list of stroke genes were correlated with defined stroke subtypes and the complex relationship between the

Table 2 Number of genes identified for each clinical subcategory in the three panels

	Total	LAA	LAN	SVD	CE	Coag	MB	ICH	VM
SGP1	120	17	47	24	43	38	24	58	9
SGP2	62	9	4	5	25	29	4	20	0
	Total	AIS	AS	CE	LAS				
SGP3	32	11	13	4	4				

This table shows the number of genes identified in total in each of the three gene panels, and specifies the number of genes per panel associated with each of the stroke subtypes. Variants in many genes were associated with more than one subtype

SGP stroke gene panel, *LAA* large artery atherosclerosis, *LAN* large artery non-atherosclerosis, *LAS* large artery stroke *SVD* small-vessel disease, *CE* cardioembolic, *Coag* coagulation, *MB* metabolic, *ICH* hemorrhage, *VM* vascular malformation, *IS* ischemic stroke, *LVD* large vessel disease, *AIS* any ischemic stroke, *AS* any stroke.

Table 3 SGP3: Genes near the lead SNPs reported in recent stroke GWAS meta-analyses

	Location	Reference sequence	Gene(s)/chromosome location	Rs number	Clinical
1	1p13, intronic	hg19 chr1:g.113042822T < C	<i>WNT2B</i>	Rs12037987	AS
2	1p13, intergenic	hg19 chr1:g.115657799C < T	<i>TSPAN2</i>	Rs12124533	AS
3	1p22, exonic	hg19 chr1:g.156202173A < G	<i>PMF1-SEMA4A</i>	Rs1052053	AS
4	1p36, intronic	hg19 chr1:g.10796866T < C	<i>CASZ1</i>	Rs880315	AS
5	1q43, intronic	hg19 chr1:g.241306248C < T	<i>RGS7</i>	Rs146390073	CES
6	2p23, 5'-UTR	hg19 chr2:g.26915624T < G	<i>KCNK3</i>	Rs12476527	AS
7	3q25, intergenic	hg19 chr3:g.149157706C < T	<i>TM4SF4-TM4SF1</i>	Rs7610618	LAS
8	4q25, intergenic	hg19 chr4:g.111714419G < T	<i>PITX2</i>	Rs13143308	CES
9	4q25, intergenic	hg19 chr4:g.113732090T < C	<i>ANK2</i>	Rs34311906	AIS
10	4q31, intronic	hg19 chr4:g.148414651T < C	<i>EDNRA</i>	Rs17612742	LAS
11	4q31, intergenic	hg19 chr4:g.155501188T < C	<i>FGA</i>	Rs6825454	AIS
12	5q23, intronic	hg19 chr5:g.121515195G < A	<i>LOC100505841</i>	Rs11957829	AIS
13	5q35, intergenic	hg19 chr5:g.172640590G < A	<i>NKX2-5</i>	Rs6891174	CES
14	6p21, intergenic	hg19 chr6:g.43262704A < T	<i>SLC22A7-ZNF318</i>	Rs16896398	AS
15	6p25, intergenic	hg19 chr6:g.1356916G < A	<i>FOXF2</i>	Rs4959130	AS
16	7p21, intergenic	hg19 chr7:g.19049388G < A	<i>HDAC9-TWIST1</i>	Rs2107595	LAS
17	7q21, 3'-UTR	hg19 chr7:g.92244422T < C	<i>CDK6</i>	Rs42039	AIS
18	9p21, ncRNA intronic	hg19 chr9:g.22102165C < T	<i>Chr9p21</i>	Rs7859727	AS
19	9q31, ncRNA intronic	hg19 chr9:g.106010237A < G	<i>LINC01492</i>	Rs10820405	LAS
20	9q34, intergenic	hg19 chr9:g.136155000C < T	<i>ABO</i>	Rs635634	AIS
21	10q24, intergenic	hg19 chr10:g.105616482T < A	<i>SH3PXD2A</i>	Rs2295786	AS
22	11q22, intergenic	hg19 chr11:g.102770353C < T	<i>MMP12</i>	Rs2005108	AIS
23	12p12, intronic	hg19 chr12:g.20577593C < A	<i>PDE3A</i>	Rs7304841	AIS
24	12q24, exonic nonsynonymous	hg19 chr12:g.111884608C < T	<i>SH2B3</i>	Rs3184504	AIS
25	12q24, intergenic	hg19 chr12:g.115554523T < C	<i>TBX3</i>	Rs35436	AS
26	13q14, intronic	hg19 chr13:g.47225745A < G	<i>LRCH1</i>	Rs9526212	AS
27	15q26, intergenic	hg19 chr15:g.91404705G < A	<i>FURIN-FES</i>	Rs4932370	AIS
28	16q22, intronic	hg19 chr16:g.73069888T < C	<i>ZFHX3</i>	Rs12932445	CES
29	16q24, intergenic	hg19 chr16:g.87575332G < A	<i>ZCCHC14</i>	Rs12445022	AS
30	17p13, intronic	hg19 chr17:g.1571818A < G	<i>PRPF8</i>	Rs11867415	AIS
31	19p13, exonic synonymous	hg19 chr19:g.10794630G < T	<i>ILF3-SLC44A2</i>	Rs2229383	AIS
32	19p13 intergenic	hg19 chr19:g.11174935C < T	<i>SMARCA4-LDLR</i>	Rs8103309	AS

SGP3 stroke gene panel 3, AS any stroke, CES cardioembolic stroke, LAS large artery stroke, AIS any ischemic stroke [12]

stroke genotype and phenotypes was summarized in a complex but systematic and evidence-based panel.

A systematic review of OMIM database and identification of relevant stroke-causing genes was done. Using six different combinations of search words, we systematically identified stroke-causing genes (Table 4). The genes were grouped into:

SGP1: Of 120 genes included in SGP1, 17 genes were associated with LAA phenotype and 47 with LAN phenotype (Supplementary Table 1). The SVD phenotype was associated with 24 genes, most of which had a clinical phenotype of WMH. SVD-related stroke with ICH was associated with eight genes, whereas two genes associated with SVD were related to HBP. A total of 36 genes were classified as CE, of which 19 genes were associated with

CE-A, 28 with CE-D, and 30 with CE-M. Coag dysfunctions were associated with stroke in 38 genes: Coag-VT with 16 genes, Coag-AT with 15, and Coag-B with 22 genes. MB phenotype was related to stroke in 24 genes and 58 genes were classified as ICH: ICH related to Coag-B in 22 genes, LAN in 14 genes, SVD in 9 genes, and VM in 6 genes. A probable mechanism for ICH could not be specified for 8 of the genes. VM was associated with eight genes, five associated with VM-AV phenotype and three with cavernomas.

SGP2: In the SGP2 section of the panel with genes identified in the database search for stroke but without a documented description of a patient with stroke but nonetheless representing diseases that may cause cerebrovascular disease, 62 genes were compiled. Nine genes were included based on LAA stroke risk, four genes

Table 4 The number of genes identified in OMIM database by using different searchterms

Genes detected by these search-word into OMIM and their phenotypic series were analyzed						
Variants	1	2	3	4	5	6
OMIM search with the following 6 search-words	(stroke)	(cereb rovas cular)	(ischemic OR ischaemic) AND (event OR stroke)	(cerebral OR intracerebral OR intracranial OR brain OR encephalic) AND (infarct OR infarc on OR ischemi OR ischaemi)	(intracranial OR cerebral OR intracerebral OR encephalic OR brain) AND (hemorrhage OR bleeding OR hematoma)	(transitory OR transient) AND (event OR ischemic OR ischaemic)
Results						
Total entries	313	88	213	209	351	357
Gene Map Table	288	84	200	179	306	402
Pheno map key 3/4	278	79	186	167	289	396
SGP1	81	30	36	40	58	21
SGP2	25	2	10	12	19	10
New included in SGP1	0	2	6	10	20	0
New included in SGP2	0	2	7	9	15	4

OMIM Online Mendelian Inheritance in Man, *SGP1* stroke gene panel 1, *SGP2* stroke gene panel 2

related to LAN stroke, and five genes to SVD, of which three with SVD-HBP (Supplementary Table 1 and Supplemental material). CE risk was suggested for 35 genes: 12 genes related to CE-A, 5 to CE-D, and 9 to CE-M. Coag disorders were associated with 29 genes: 2 genes with Coag-AV, 16 genes with Coag-VT, and 14 genes with Coag-B. MB phenotype was associated with four genes. ICH was associated 21 genes, 16 of them with ICH-Coag-B, 2 with ICH-LAN, and 3 with ICH-HBP.

SGP3: SGP3 contains 32 stroke genes located in the proximity of a lead SNP with evidence for stroke pathogenicity [12]. Of these, 11 were associated with any ischemic stroke, 13 with any stroke, 4 genes with cardioembolic stroke and 4 with large artery stroke (Table 3).

Selection of genes for clinical evaluations

Sixty-one stroke-related genes in SGP1 and 27 in SGP2 may be more relevant to consider in the evaluation of monogenic stroke in clinical practice, as these genes contained variants more solidly documented as described in the method, in the subsection “Selections of genes for clinical versus research evaluations.”

An increasing number of patients and families with stroke are examined by WES in clinical or research context. Not infrequently, WES detects variants of unknown significance, for example, not previously reported rare variants in a gene where other variants have been associated with stroke. Our panels can aid in the interpretation of the consequence of such variants by providing information on stroke subtypes where variants in the specific gene have been reported. If the stroke subtype of the new patient tested is the same as

already reported, and if the novel variant otherwise appears a likely cause of disease, for example, based on in silico prediction tools, phylogenetic conservation, variant frequency, and so on, this increases the likelihood that the novel variant is pathogenic. The interpretation of detected novel variants in the genes included in the SGPs needs to be done carefully, taking into consideration the possibility that the true genetic cause of stroke in the examined patient or family may be a different one. Further, our gene panels summarize the current status of genes associated with the different clinical stroke subtypes.

Previous panels of genes related to monogenic stroke or subtypes of monogenic stroke have been reported, for example, for all types of stroke, such as for targeted sequencing [15] or for massively parallel sequencing [16], as well as compilations of genes intended to be used for the evaluation of a certain stroke subgroup [17]. Although these previous valuable studies presented many genes associated with monogenic stroke, we could not identify in the literature any comprehensive panel that contains a systematic compilation of currently relevant genes for possible monogenic stroke [18–21]. Citations of some articles containing relevant information on the connection between stroke genes and different identified stroke subtypes are provided in our panels so that the original clinical description of a disease or syndrome can be easily identified. As the different possible pathologies that cause stroke may manifest in a diversity of clinical stroke subtypes, the clinical information provided in the panels allow for matching between the specific phenotype of the patient and the relevant genes.

The obtained panels SGP1, SGP2, and SGP3 can be seen as easy to adapt, evidence-based tools for genetic evaluation

of suspected monogenic stroke. Each genetic laboratory can individually choose the best method, for example, exome sequencing to analyze various possible pathogenic variants (substitutions, missense, nonsense, short deletions, insertions, or duplications) followed by Sanger sequencing confirmation of relevant findings, and/or complementary tests such as to detect copy-number variation.

Research versus clinical diagnostic usage of the stroke gene panel

We propose the panels to be used for the interpretation of data generated by whole-exome sequencing in future research studies of IS probands where family history or phenotypic information is suggestive of monogenic inheritance. In addition, the panels can also be considered for genetic evaluation of stroke in clinical practice although not all the genes in the table may have sufficiently strong, clinically evidence-based relevance as causative for stroke. We suggest a stratification of the compiled stroke genes in SGP1 and SGP2 into clinical or research-only categories. We considered a gene's pathogenicity to be clinically meaningful when at least one variant has been reported that co-segregates with stroke (SGP1) or with a specific disease that may cause stroke in a defined way (SGP2), in two or more pedigrees, as we described in detail in the Methods, subsection "Selections of genes for clinical versus research evaluations." However, individual analyses of each result using up-to date published data remains essential, as a variant considered disease-causing can be later shown to be benign or a variant considered benign today can turn out to be pathogenic when more data becomes available. The complex interrelations between different genes and environmental factors can result in incomplete penetrance for many diseases, which means further challenges for clinicians in interpreting the meaning of an identified variant for a specific patient. Nevertheless, an accurate genetic diagnosis remains essential for the understanding of the disease, and for choosing the best prevention and treatment strategies.

We expect that other genes and pathologies responsible for monogenic stroke will be identified and defined in the near future. As all the genes related to monogenic stroke in our panels were systematically included based on clearly defined criteria, the panels can be quickly and easily updated.

Although GWASs are mainly designed to identify polygenic contributors to stroke, and rare causal variants are most likely not the same as GWAS lead SNPs associated with stroke, we considered that it might be of interest to evaluate the GWAS-related genes listed in our SGP3 within research studies. A weak association in GWAS studies may suggest a large stroke risk in single individuals, as illustrated by the situation in Parkinson disease where the genes *SNCA* and *MAPT*, both containing well-established variants for monogenic forms of Parkinsonism, also contained top

hits in GWAS studies [22]. Mutations in *COL4A2* are well established as causes of monogenic stroke, and a SNP in the same gene, *COL4A2*, was suggestively associated with stroke in a large GWAS meta-analysis [12].

Limitations

Though systematically compiled, our panels are based on a review of currently available data on genes and related pathologies. The panels are meant to be regularly updated using clinical and genetic data on stroke that the scientific community continuously provides. This means that new genes could be included in the near future, and that some of the genes and related pathologies may be eliminated or adjusted in the panels. We only included genes related to intermediary stroke risk factors (e.g., hypertension, atrial fibrillation) when our search detected such a gene in combination with mentioning of cerebrovascular disease. It is therefore possible that additional genes may also be of interest for evaluation of individuals with suspected monogenic stroke.

Our method to identify stroke genes proved to cover a majority but not all the stroke gene the authors are aware of. This shows how difficult it still is today to gather the actual knowledge on known genetic causes to Mendelian stroke, and also shows that our tables probably are not complete. One stroke gene, *CTSA*, known to be associated with stroke [23] but not identified by our OMIM search was included in the SGP1.

Even though it is not clear to date whether genes or loci associated with stroke risk in large GWAS will be found to harbor monogenic causes of Mendelian stroke, we included genes and loci identified in a very large meta-analysis of GWAS data [14] in SGP3 (Table 3). We suggest that this panel can be used for research purposes to evaluate stroke families for possible rare high-impact variants in those genes/loci, and hypothesize that these genes might sometimes contain rare genetic causes for Mendelian stroke.

The stroke gene panels have not yet been tested in a cohort of patients. We hope that upcoming studies by others and us, using the stroke gene panels, will determine the proportion of monogenic stroke in various proportions and establish the usefulness of the stroke gene panels.

Conclusions

We present a comprehensive, systematically compiled stroke gene panel for evaluation of patients with possible monogenic stroke. The clinical utility of the panels should be further prospectively evaluated. Accordingly, we plan to join on-going clinical efforts from different research groups and further test the stroke gene panels, by analyzing the

WES data of patients with possible monogenic stroke from different stroke cohorts.

We also suggest the development of a publically available website-based stroke gene panel to be permanently updated, as the rate of discovery in the area is high. Such an effort would prospectively evaluate the evidence for or against pathogenicity of variants in the genes in the three panels. For example, identification of the same, rare variant in unrelated patients with clinically identical type of stroke, identified in different international research projects or in clinical diagnostic efforts, would greatly increase the likelihood that the variant truly is pathogenic.

Acknowledgements This study was supported by the Swedish Heart and Lung Foundation; Region Skåne; Skåne University Hospital; the Freemasons Lodge of Instruction EOS, Lund; Lund University; and the Swedish Stroke Association.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

References

- Hay SI, Abajobir AA, Abate KH, et al. Global, regional, and national disability-adjusted life-years (DALYs) for 333 diseases and injuries and healthy life expectancy (HALE) for 195 countries and territories, 1990–2016: a systematic analysis for the Global Burden of Disease Study 2016. *Lancet* 2017;390:1260–344.
- Flossmann E, Schulz UG, Rothwell PM. Systematic review of methods and results of studies of the genetic epidemiology of ischemic stroke. *Stroke*. 2004;35:212–27.
- Meschia JF, Brown RD Jr, Brott TG, et al. Feasibility of an affected sibling pair study in ischemic stroke: results of a 2-center family history registry. *Stroke*. 2001;32:2939–41.
- Jood K, Ladvall C, Rosengren A, et al. Family history in ischemic stroke before 70 years of age: the Sahlgrenska Academy Study on Ischemic Stroke. *Stroke*. 2005;36:1383–7.
- Starby H, Delavaran H, Andberg G, et al. Multiplicity of risk factors in ischemic stroke patients: relations to age, sex, and subtype—a study of 2,505 patients from the Lund Stroke Register. *Neuroepidemiology*. 2014;42:161–8.
- Sharma P, Meschia JF. *Stroke genetics*. London: Springer; 2012.
- Lindgren A, Lovkvist H, Hallstrom B, et al. Prevalence of stroke and vascular risk factors among first-degree relatives of stroke patients and control subjects. A prospective consecutive study. *Cerebrovasc Dis*. 2005;20:381–7.
- Ilinca A, Kristoffersson U, Soller M, Lindgren AG. Familial aggregation of stroke amongst young patients in Lund Stroke Register. *Eur J Neurol*. 2016;23:401–7.
- Silventoinen K, Hjelmborg J, Möller S, et al. Family aggregation of cardiovascular disease mortality: a register-based prospective study of pooled Nordic twin cohorts. *Int J Epidemiol*. 2017;46:1223–9.
- Holliday EG, Traylor M, Malik R, et al. Genetic overlap between diagnostic subtypes of ischemic stroke. *Stroke*. 2015;46:615–9.
- Jiang T, Tan MS, Tan L, Yu JT. Application of next-generation sequencing technologies in neurology. *Ann Transl Med*. 2014;2:125.
- Malik R, Chauhan G, Traylor M, et al. Multiancestry genome-wide association study of 520,000 subjects identifies 32 loci associated with stroke and stroke subtypes. *Nat Genet*. 2018;50:524–37.
- Adams HP Jr, 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;24:35–41.
- 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;38:2979–84.
- Bersano A, Markus HS, Quaglini S, et al. Clinical pregenetic screening for stroke monogenic diseases: results from Lombardia GENS Registry. *Stroke*. 2016;47:1702–9.
- Bersano A, Baron P, Lanfranconi S, et al. Lombardia GENS: a collaborative registry for monogenic diseases associated with stroke. *Funct Neurol*. 2012;27:107–17.
- Bersano A, Zuffardi O, Pantoni L, et al. Next generation sequencing for systematic assessment of genetics of small-vessel disease and lacunar stroke. *J Stroke Cerebrovasc Dis*. 2015;24:759–65.
- Søndergaard CB, Nielsen JE, Hansen CK, Christensen H. Hereditary cerebral small vessel disease and stroke. *Clin Neurol Neurosurg*. 2017;155:45–57.
- Chauhan G, Dabette S. Genetic risk factors for ischemic and hemorrhagic stroke. *Cardiol Rep*. 2016;18:124.
- Tan R, Traylor M, Rutten-Jacobs L, Markus H. New insights into mechanisms of small vessel disease stroke from genetics. *Clin Sci (Lond)*. 2017;131:515–31.
- Lindgren A. Stroke genetics: a review and update. *J Stroke*. 2014;16:114–23.
- Nalls MA, Pankratz N, Lill CM, et al. Large-scale meta-analysis of genome-wide association data identifies six new risk loci for Parkinson's Disease. *Nat Genet*. 2014;46:989–93.
- Bugiani M, Kevelam SH, Bakels HS, et al. Cathepsin A-related arteriopathy with strokes and leukoencephalopathy (CARASAL). *Neurology*. 2016;25:1777–86.