

# Neglected Mendelian causes of stroke in adult Chinese patients who had an ischaemic stroke or transient ischaemic attack

Wei Li,<sup>1,2</sup> Hao Li ,<sup>1</sup> Chaoxia Lu,<sup>1</sup> Jialu Zhao,<sup>1</sup> Huichun Xu ,<sup>3</sup> Zhe Xu,<sup>1</sup> Braxton Mitchell,<sup>3</sup> Yong Jiang ,<sup>1</sup> Hong-Qiu Gu ,<sup>1</sup> Qin Xu ,<sup>1</sup> Anxin Wang,<sup>1</sup> Xia Meng ,<sup>1</sup> Jinxi Lin,<sup>1</sup> Jing Jing,<sup>1,2</sup> Zixiao Li,<sup>1,2</sup> Wanlin Zhu,<sup>1</sup> Zhigang Liang ,<sup>4</sup> Mengxing Wang,<sup>1</sup> Yongjun Wang <sup>2,5</sup>

**To cite:** Li W, Li H, Lu C, *et al.* Neglected Mendelian causes of stroke in adult Chinese patients who had an ischaemic stroke or transient ischaemic attack. *Stroke & Vascular Neurology* 2024;**9**: e002158. doi:10.1136/svn-2022-002158

► Additional supplemental material is published online only. To view, please visit the journal online (<http://dx.doi.org/10.1136/svn-2022-002158>).

Received 11 November 2022  
Accepted 27 June 2023  
Published Online First  
26 July 2023



© Author(s) (or their employer(s)) 2024. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ.

For numbered affiliations see end of article.

## Correspondence to

Dr Yongjun Wang;  
yongjunwang@ncrcnd.org.cn

## ABSTRACT

**Background and purpose** Multiple factors play important roles in the occurrence and prognosis of stroke. However, the roles of monogenic variants in all-cause ischaemic stroke have not been systematically investigated. We aim to identify underdiagnosed monogenic stroke in an adult ischaemic stroke/transient ischaemic attack (TIA) cohort (the Third China National Stroke Registry, CNSR-III).

**Methods** Targeted next-generation sequencing for 181 genes associated with stroke was conducted on DNA samples from 10 428 patients recruited through CNSR-III. The genetic and clinical data from electronic health records (EHRs) were reviewed for completion of the diagnostic process. We assessed the percentages of individuals with pathogenic or likely pathogenic (P/LP) variants, and the diagnostic yield of pathogenic variants in known monogenic disease genes with associated phenotypes.

**Results** In total, 1953 individuals harboured at least one P/LP variant out of 10 428 patients. Then, 792 (7.6%) individuals (comprising 759 individuals harbouring one P/LP variant in one gene, 29 individuals harbouring two or more P/LP variants in different genes and 4 individuals with two P/LP variants in *ABCC8*) were predicted to be at risk for one or more monogenic diseases based on the inheritance pattern. Finally, 230 of 792 individuals manifested a clinical phenotype in the EHR data to support the diagnosis of stroke with a monogenic cause. The most diagnosed Mendelian cause of stroke in the cohort was cerebral autosomal dominant arteriopathy with subcortical infarcts and leukoencephalopathy. There were no relationships between age or family history and the incidence of first symptomatic monogenic stroke in patients.

**Conclusion** The rate of monogenic cause of stroke was 2.2% after reviewing the clinical phenotype. Possible reasons that Mendelian causes of stroke may be missed in adult patients who had an ischaemic stroke/TIA include a late onset of stroke symptoms, combination with common vascular risks and the absence of a prominent family history.

## INTRODUCTION

Deleterious mutations in a single gene can cause a Mendelian form of ischaemic stroke

## WHAT IS ALREADY KNOWN ON THIS TOPIC

⇒ Strokes caused by single-gene mutations are an important type of stroke aetiology. However, the prevalence of strokes with Mendelian causes in all-cause ischaemic stroke is unknown.

## WHAT THIS STUDY ADDS

⇒ We identified that 7.6% individuals harboured at least one pathogenic or likely pathogenic variant associated with one or multiple monogenic diseases in a Chinese all-cause ischaemic stroke cohort of 10 428 individuals.  
⇒ After reviewing electronic health record data, the rate of a monogenic cause of stroke was 2.2%.

## HOW THIS STUDY MIGHT AFFECT RESEARCH, PRACTICE, OR POLICY

⇒ Care should be taken to diagnose monogenic causes in all-cause ischaemic strokes, and an effective Mendelian stroke gene panel should be used to aid diagnosis.

(IS), either as a primary or a secondary manifestation.<sup>1</sup> Despite numerous genetic studies of IS,<sup>2–4</sup> very few monogenic causes of strokes have been identified to date, and these have been found only in patients with a very early age of onset or with some types of small-vessel disease (SVD).<sup>5,6</sup> Few studies have attempted to systemically quantify the prevalence of monogenic stroke or to identify corresponding pathogenic variants in patients who had a stroke in the general population, especially in older individuals. The recent progress in genomic technology and the reduction of sequencing costs now make such an investigation feasible.<sup>7</sup> Understanding the rare genetic variants associated with stroke is an essential step to identify the causes of stroke, decipher the underlying mechanisms, facilitate the identification of novel therapeutic targets and optimise prevention strategies.<sup>6</sup> The current

**Table 1** Characteristics of the included patients in CNSR-III

	NGS analysis cohort (n=10428)	CNSR-III cohort (n=15166)	Remaining cohort (n=4738)
Age at time of study entry, mean±SD	62.25±11.32	62.23±11.30	62.19±11.26
≤45 years	735 (7.05)	1074 (7.08)	339 (7.15)
>45 years	9693 (92.95)	14092 (92.92)	4399 (92.85)
Male, n (%)	7137 (68.44)	10364 (68.34)	3277 (68.11)
Ethnicity, Han, n (%)	10118 (97.03)	14726 (97.10)	4608 (97.26)
Current smoker, n (%)	3320 (31.84)	4752 (31.33)	1432 (30.22)
Heavy drinker, n (%)	1482 (14.21)	2126 (14.02)	644 (13.59)
BMI (≥25 kg/m <sup>2</sup> )	4464 (42.81)	6468 (42.65)	2004 (42.30)
Medical history, n (%)			
Ischaemic stroke	2349 (22.53)	3355 (22.12)	1006 (21.23)
Coronary heart diseases	1152 (11.05)	1608 (10.60)	456 (9.62)
Atrial fibrillation	737 (7.07)	1019 (6.72)	282 (5.95)
Hypertension	6540 (62.72)	9494 (62.60)	2954 (62.35)
Diabetes mellitus	2490 (23.88)	3510 (23.14)	1020 (21.53)
Dyslipidaemia	882 (8.46)	1191 (7.85)	309 (6.52)
Stroke type			
IS	9728(93.3)	14146 (93.27)	4418 (93.25)
TIA	700 (6.7)	1020 (6.73)	320 (6.75)
Family history of Stroke	1395 (13.38)	2013 (13.27)	618 (13.04)
CCS			
Large artery atherosclerosis	2999 (28.76)	4364 (28.77)	1365 (28.81)
Cardioaortic embolism	718 (6.89)	971 (6.40)	253 (5.34)
Small arterial occlusion	2614 (25.07)	3747 (24.71)	1133 (23.91)
Other aetiologies	86 (0.82)	152 (1.00)	66 (1.39)
Undetermined aetiology	4011 (38.46)	5932 (39.11)	1921 (40.54)

Note: Heavy drinker: alcohol ≥20 g/day.  
 BMI, body mass index; CCS, Causative Classification System for Ischaemic Stroke; CNSR-III, the Third China National Stroke Registry; IS, ischaemic stroke; NGS, next-generation sequencing; TIA, transient ischaemic attack.

study was based on a cohort from the Third China National Stroke Registry (CNSR-III), which enrolled more than 10000 consecutive patients who had an IS or transient ischaemic attack (TIA). We sought to determine the prevalence of pathogenic variants associated with Mendelian causes of stroke and to estimate the extent of potentially missed genetic diagnoses in adult patients who had an IS.

## METHODS

### Study population and classification of Mendelian causes of stroke

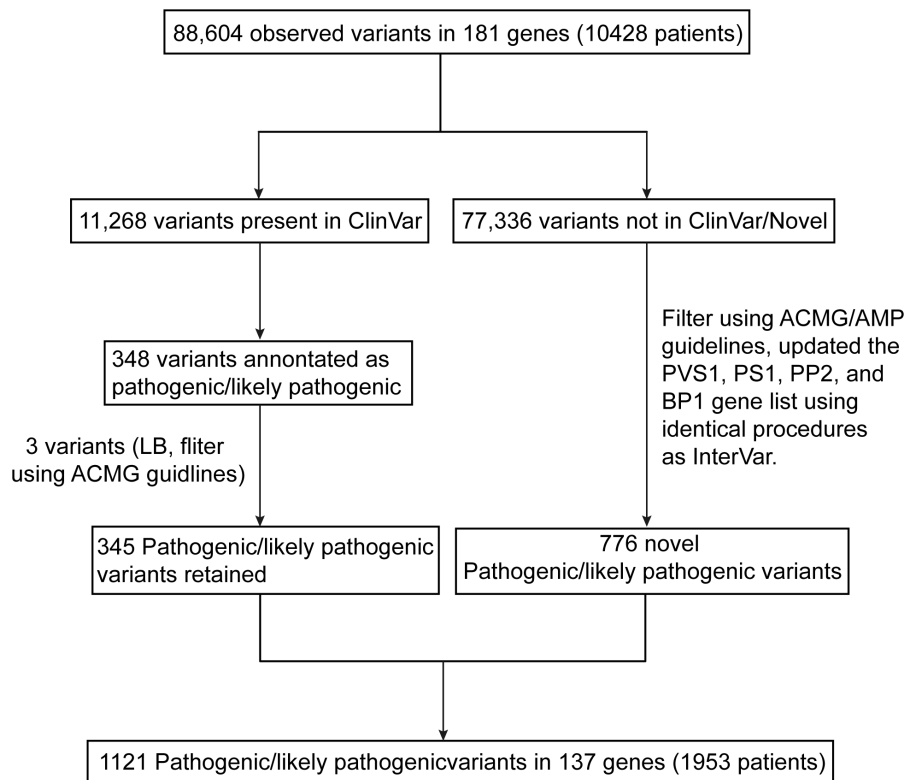
The CNSR-III is a nationwide prospective registry for hospitalised patients who had IS/TIA between August 2015 and March 2018 in China. A total of 15166 stroke patients were enrolled. The detailed CNSR-III protocol has been published.<sup>8</sup> Most CNSR-III participants were also included in a genetic sub-study (n=12603), for which targeted next-generation sequencing (NGS) was successfully conducted for 10613 patients (online supplemental eFigure 1).

Monogenic disorders with a stroke phenotype were classified into the following subgroups: large-artery disease, SVD, embolic stroke, a prothrombotic state and other diseases (including neurofibromatosis 1, polycystic kidney disease, Fabry disease and cerebral cavernous malformations), based on the references<sup>2,9</sup> (online supplemental eTable1).

Clinical classification of IS was performed according to the Causative Classification System for Ischaemic Stroke (5-item CCS).<sup>10</sup>

### NGS and data analysis

Briefly, DNA was isolated from peripheral leukocytes using a DNA Isolation Kit (Bioteke, AU1802, Beijing, China). DNA libraries were prepared using a KAPA Library Preparation Kit (Kapa Biosystems, KR0453, Wilmington, Massachusetts, USA) following the manufacturer's instructions. Genomic DNA capture, library construction and targeted NGS using a panel for Mendelian strokes were conducted as previously described.<sup>11</sup> Paired-end sequencing (150bp) was performed on HiSeq X Ten or NovaSeq (Illumina, San Diego, California, USA). The



**Figure 1** Identification and selection of pathogenic/likely pathogenic variants in the CNSR-III cohort. In total, 88604 observed variants in 181 genes were enrolled in this study. We used 345 pathogenic/likely pathogenic variants after filtering using the ClinVar database (left), and 776 novel pathogenic/likely pathogenic variants after annotation using the ACMG/AMP guidelines (right). ACMG, American College of Medical Genetics and Genomics; AMP, Association for Molecular Pathology; CNSR-III, the Third China National Stroke Registry; LB, likely benign.

sensitivity and specificity of the targeted sequencing were evaluated by comparing the results with the results of Sanger sequencing from a previous study by our group.<sup>11</sup> Variant calling and quality control are described in online supplemental file 1. For the current analysis, we focused only on 181 candidate genes associated with Mendelian stroke or stroke-related risk factors (online supplemental eTable 2). The pathogenicity was evaluated using InterVar software and customised scripts (V.2.0.1) according to the guidelines of the American College of Medical Genetics and Genomics (ACMG) and the Association for Molecular Pathology (AMP).<sup>12</sup> The ClinVar database (ClinVar 20200622 version) was used to aid the evaluation.

#### Evaluation of the concordance between clinical phenotypes and the genetic classification of monogenic stroke

Through electronic health record (EHR) review and based on the availability of diagnostic criteria and the manifestation of relevant disease phenotypes, we classified the extent of the confidence of a diagnosis of monogenic stroke into several categories: undetermined (ie, without phenotypic expression of the relevant monogenic disease), possible (ie, with some features of the monogenic disease), definite (ie, met diagnostic criteria for the monogenic disease) or insufficient information. Based on the literature, some but not all monogenic diseases have well-established diagnostic criteria. For those without existing diagnostic criteria, we used the

disease-related phenotypes listed on OMIM and published data to classify the diagnosis. Full details of the classification scheme for each phenotype can be found in the Phenotyping section of online supplemental file 1.

#### Statistical analysis

We used R software (V.3.6.1) to perform analysis. Multivariable logistic regression analysis was used to predict the relationships between age or family history and the incidence of monogenic stroke in patients without a history of stroke, controlling for hypertension, hyperlipidaemia, diabetes, coronary heart disease, atrial fibrillation, smoking history, drinking history and body mass index (BMI)  $\geq 25$  kg/m<sup>2</sup>.

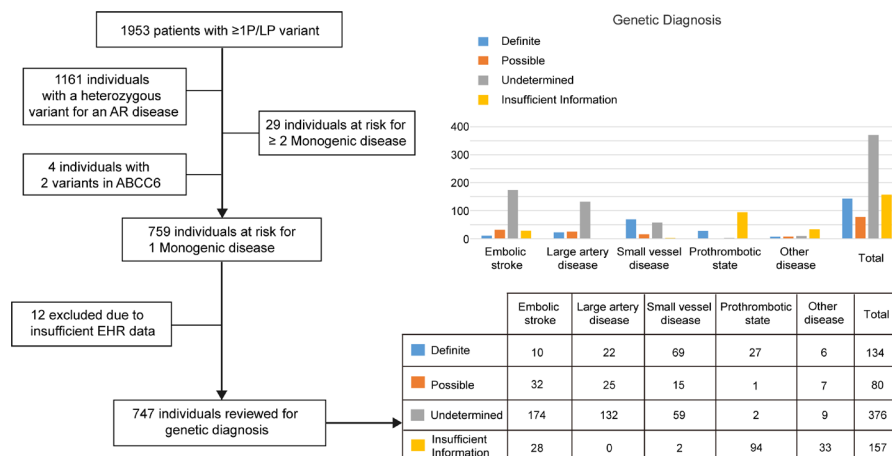
#### Data availability

The data that support the findings of this study are available from the corresponding authors on reasonable request.

## RESULTS

### Description of NGS analysis cohort

After filtering out 147 contaminated samples and 38 duplicated samples from 10613 individuals, 10428 patients remained for NGS data analysis (online supplemental eFigure 1). The final set of 10428 samples had



**Figure 2** Individuals diagnosed and potentially missed diagnoses. Flow chart (left) illustrating the number of individuals harbouring one or more P/LP variant, and the number of individuals predicted to develop one or more monogenic disease. Bar plot (right) illustrating the proportions of the groups at risk for one monogenic disease, showing their likelihood of a missed diagnosis. EHR, electronic health record; P/LP, pathogenic or likely pathogenic.

an average mean depth of coverage of 192, and 96.2% of targeted bases had a coverage depth of at least 20.

In this study cohort, patients who had an IS accounted for 93.3% (9728/10 428) and patients who had a TIA accounted for 6.7% (700/10 428). The ages ranged from 19 to 95 with a mean (SD) of 62.3 (11.3) years old. Of the patients, 93.0% were 45 years old or older, 7137 (68.4%) were men, 2349 (22.53%) had a history of IS and 1395 (13.38%) had a family history of stroke (table 1).

### Pathogenic/likely pathogenic (P/LP) variants

In total, 88604 variants were found in the 181 candidate genes among the 10428 individuals. We implemented two pipelines to annotate the variants: one for variants annotated by ClinVar (11268 variants) and the other for 77336 variants that were not present in the ClinVar database. The first pipeline focused on P/LP variants classified in ClinVar (348 variants in 1031 individuals) followed by verification through manual review according to the ACMG/AMP guidelines, which filtered out three variants re-annotated as likely benign. The second pipeline used our own customised scripts based on ACMG/AMP principles to classify the remaining 77336 not present in the ClinVar database. The second pipeline included updating the PVS1, PS1, PP2 and BP1 gene lists based on identical procedures used in InterVar. This identified a total of 1121 variants, in 137 genes, presented in 1953 individuals, that were classified as P/LP and were further analysed for evaluation of their inheritance patterns and genotype–phenotype concordance (figure 1).

We further considered the inheritance pattern of the disease, excluding individuals with a heterozygous variant of an autosomal recessive disease. A total of 759 (online supplemental eTable 3) individuals harboured one P/LP variant in 80 genes and were predicted to be at risk for one monogenic disease (figure 2), while 29 individuals harboured more than two P/LP variants and were predicted to be at risk for multiple monogenic diseases (online supplemental eTable 4). In addition, four individuals harboured two P/LP

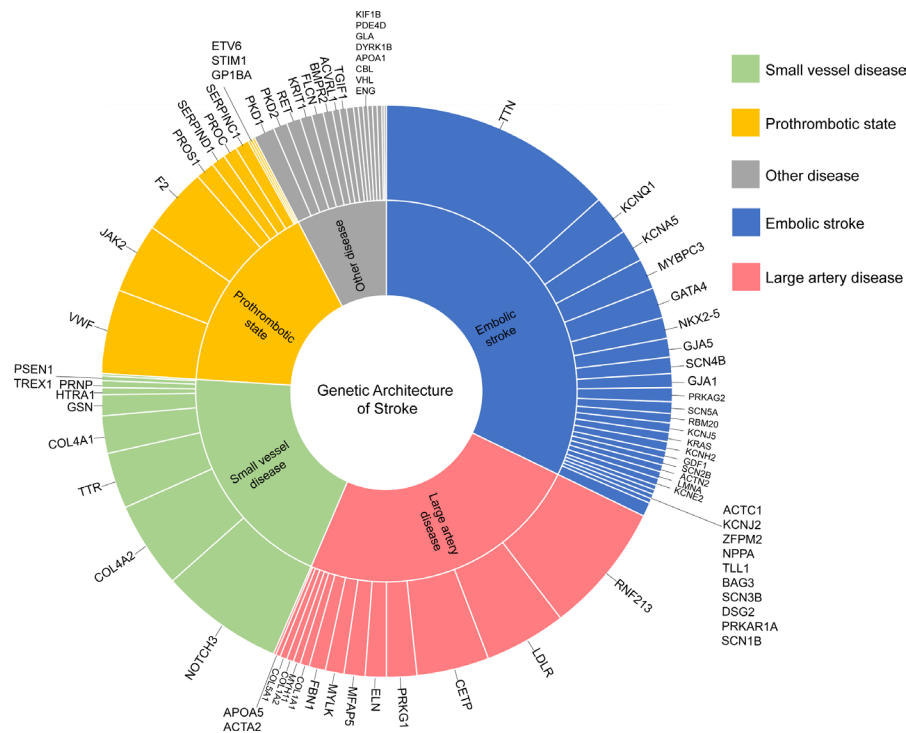
variants (without confirmation of paternity and maternity) in *ABCC6* (online supplemental eTable 5). The Mendelian causes of stroke identified in our cohort included 245 embolic stroke cases (32.3%), 184 large-artery disease cases (24.2%), 148 SVD cases (19.4%), 124 cases of a prothrombotic state (16.3%) and 58 other disease cases (7.6%) (total, 759 individuals; online supplemental eTable 6). Detailed aetiological classifications are shown in online supplemental eTable 3.

### Diagnostic rate of individuals predicted to develop one monogenic disease

EHR data registered in the CNSR-III cohort were available for 747 of these 759 individuals with one P/LP variant at risk for one monogenic disease, to verify the genetic diagnosis (figure 2). Classification of the monogenic stroke and the corresponding genes involved are shown in figure 3. Among the 747 individuals, 157 individuals were classified as having insufficient information, as although EHR data were present in the registry, we anticipated that the phenotypes of their monogenic diseases would not be evaluated through EHR review. After reviewing clinical information for the remaining 590 individuals, we classified them into three groups according to the level of support from clinical evidence: definite genetic diagnosis (134 individuals), possible genetic diagnosis (80 individuals) and inconclusive/undetermined genetic diagnosis because of the absence of clinical phenotypes (376 individuals, figure 2). The positive diagnosis rates (definite+possible diagnosis) were 19.4% (42/216) for embolic stroke, 26.3% (47/179) for large-artery disease, 58.7% (84/143) for SVD, 93.3% (28/30) for a prothrombotic state and 59.1% (13/22) for other diseases. Overall, the positive diagnosis yield among patients with genetically diagnosed monogenic stroke showed the highest yield for monogenic prothrombotic state.

We also found four individuals with two P/LP variants in the *ABCC6* gene (online supplemental eTable





**Figure 3** Genetic architecture of stroke. Each gene related to monogenic stroke identified in 759 individuals was classified into five subgroups: large-artery disease, small-vessel disease, embolic stroke, a prothrombotic state and other diseases (shown in the middle text circle). The proportions of affected genes are shown in the outermost circle.

5), predicted to have pseudoxanthoma elasticum in an autosomal recessive inheritance pattern. We reviewed the EHR data from these four patients and found no evidence to support a clinical diagnosis of pseudoxanthoma elasticum.<sup>13</sup>

#### Diagnostic rate of individuals predicted to develop two or more monogenic diseases

Surprisingly, we identified 29 individuals who harboured two P/LP variants in multiple genes and were predicted to develop two or more relevant monogenic diseases based on the inheritance pattern (online supplemental eTable 4). Two of them (patients #CNSR302050 and #CNSR303839) harboured three variants, and one (patient #CNSR306857) harboured four variants. Of these 29 individuals, three showed definite or possible clinical evidence to support the presence of two monogenic diseases. Thirteen of them had definite or possible clinical evidence to support the presence of only one monogenic disease. The remaining 12 patients did not have sufficient clinical phenotypes to support a genetic diagnosis. This group showed a clinical concordance rate (55.2%, 16/29) (online supplemental eTable 4).

#### Summary of the diagnostic rate of all individuals with one or more P/LP variant

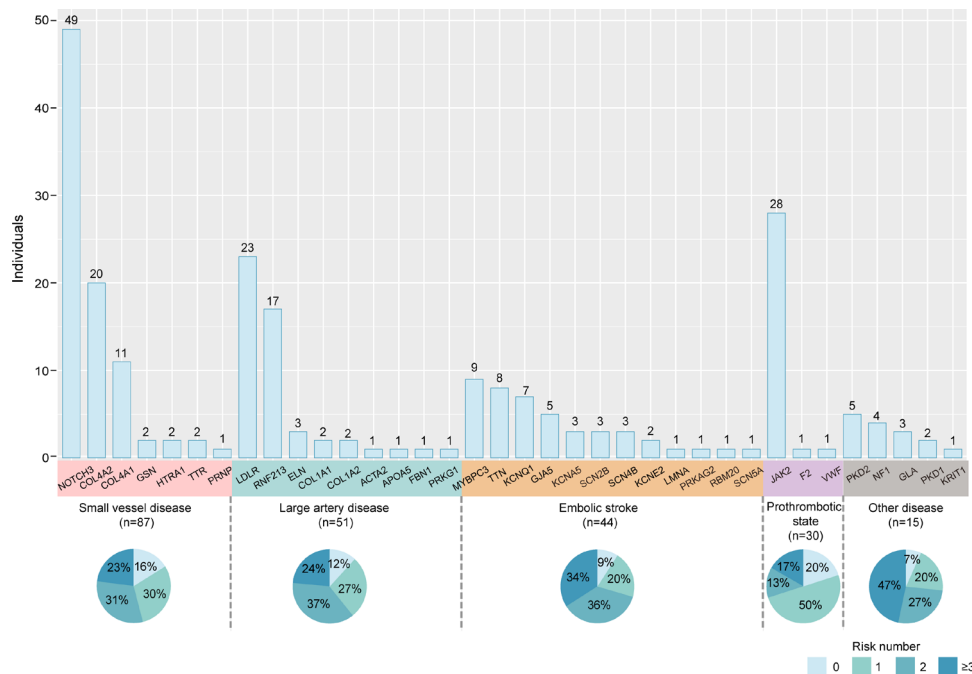
In total, 792 of 10428 individuals (7.6% of all patients) were identified as carrying at least one P/LP variant for monogenic disease, according to the ACMG/AMP guidelines or the ClinVar database. EHR data were available for 780 individuals, and 624 individuals had relevant

phenotypic information for evaluation in the EHR data that corresponded to their genetic diagnoses of a monogenic disease. A total of 230 individuals (36.9%, 230/624) exhibited definite or possible clinical evidence to support their genetic diagnoses, including 227 individuals with one monogenic disease and three individuals with two monogenic diseases. In other words, 2.2% (230/10428) of individuals from our cohort not only carried at least one P/LP variant related to monogenic stroke but also demonstrated definite or possible clinical phenotypic evidence to support a genetic diagnosis.

At the gene level, individuals with *NOTCH3* P/LP variants had the highest rate of positive genetic diagnosis (89.3%, 50/56). Mutations in exon 11 of *NOTCH3* accounted for 44.0% (22/50), with R544C and R587C as the most common (28.0% and 14.0%, respectively). Variants in exon 6–24 accounted for 88.0% (44/50). Surprisingly, we identified a *JAK2* variant (p.V617F) in 33 individuals, 29 of whom had corresponding phenotypes (ie, thrombocytopenia or erythrocytosis). The third and fourth monogenic diseases with relatively high genetic diagnosis were familial hypercholesterolemia caused by heterozygous *LDLR* mutations (60%, 24/40) and *COL4A2* microangiopathy caused by heterozygous *COL4A2* mutations (52.5%, 21/40).

#### The characteristics of 230 individuals with Mendelian causes of stroke

Patients in our cohort with Mendelian causes of stroke had a mean age of 61.8 years old, and 65.4% were men.



**Figure 4** Characteristics of 227 individuals diagnosed with Mendelian causes of stroke. The bars represent the number of individuals diagnosed with Mendelian causes of stroke identified for each gene, which are coloured according to the classification of stroke aetiology. Three individuals with two monogenic diseases were ruled out. The percentages in the boxes indicate the distribution of risk factors, including hypertension, hyperlipidaemia, diabetes, coronary heart disease, atrial fibrillation, smoking history, drinking history and body mass index  $\geq 25$ .

Only 17 individuals of the 230 (7.4%, 17/230) had been diagnosed with an identified aetiology in EHR prior to genetic testing (online supplemental eTable 7), including eight cases of cerebral autosomal dominant arteriopathy with subcortical infarcts and leukoencephalopathy (CADASIL) with *NOTCH3* mutation, six cases of idiopathic thrombocytopenia with V617F mutation in the *JAK2* gene and three cases of Moyamoya disease with a *RNF213* R4810K mutation. The common risk factors for IS, such as hypertension, hyperlipidaemia, diabetes, coronary heart disease, atrial fibrillation, smoking history, drinking history and BMI  $\geq 25$  kg/m<sup>2</sup>, were carried out by 86.5% (199/230) of individuals. Of them, 30% (69/230) of all patients with monogenic causes carried one risk factor, 30.4% (70/230) carried two risk factors and 26.1% (60/230) carried three or more risk factors (figure 4). Only 10.87% of individuals among these patients with Mendelian causes of stroke had a family history of stroke. According to the multivariable logistic regression model, after eliminating the confounding effect of common risk factors, there were no relationships between age (OR=0.99, 95% CI: 0.98 to 1.01, p=0.23) or family history (OR=0.66, 95% CI: 0.36 to 1.11, p=0.14) and the incidence of first symptomatic monogenic stroke in patients.

## DISCUSSION

In this study, at least 2.2% of our cohort had definite or possible clinical evidence to support genetically diagnosed Mendelian causes of stroke/TIA, which is similar to other studies on complex diseases. For example, it was

found that the diagnosis rate of monogenic disease was 1.7% in a cardiovascular disease cohort.<sup>14</sup>

Several features among the individuals identified as having a Mendelian cause of stroke in our cohort presented complexity and obstacles for a correct diagnosis of monogenic stroke, including late-onset symptoms of stroke, coexisting common risk factors and a low prevalence of a positive family history. Most of the monogenic stroke individuals with a first symptomatic stroke in our cohort were relatively old with a mean age of 61, and most of the patients carried common risk factors similar to other stroke patients with non-Mendelian causes in our adult IS cohort. However, similar exceptions were already known for some monogenic diseases. For example, patients with CADASIL can have stroke events that occur after the age of 60 and can carry common cerebrovascular risk factors.<sup>15–18</sup> Hypertension is present in 20% of patients with CADASIL, and hyperlipemia and smoking are present in 50% of patients with CADASIL.<sup>19</sup> More than 90% of the patients with *COL4A1/COL4A2* mutations in our cohort did not present with haemorrhage, either now or previously, or had only microbleeds with other characteristics of cerebral SVD, a result that somewhat contradicts prior literature indicating that *COL4A1/COL4A2* mutations are a cause of haemorrhagic stroke.<sup>11 20</sup> Similarly, among 16 cases of Moyamoya disease caused by *RNF213* mutation, only three cases were clinically diagnosed as Moyamoya disease in EHR before genetic testing while the remaining cases were diagnosed as vascular stenosis,



either owing to coexisting common risk factors or only unilateral internal carotid artery involvement. However, a similar complex presentation has been reported in *COL4A1/COL4A2* microangiopathy,<sup>21–24</sup> Moyamoya disease<sup>25</sup> and CADASIL,<sup>17 26–28</sup> whose patients can present with mild signs or symptoms, or even have a negative history of stroke and family history.

The *NOTCH3* gene contains 33 exons encoding the Notch3 protein, which includes an extracellular domain that consists of 34 epidermal growth factor-like repeats (EGFr).<sup>29</sup> Most P/LP variants (89.29%, 50/56) of the *NOTCH3* gene in our cohort were located in exon 6 to exon 22, encoding EGFr 7–34, which results in milder phenotypes than mutations located in the region encoding EGFr 1–6.<sup>29–32</sup> Another example can be found in individuals with a V617F mutation in the *JAK2* gene, leading to essential thrombocythemia or polycythemia vera. These patients present with only an increased platelet count, which is easily confused with the increased platelet count secondary to stroke complications such as infection or anaemia. Additionally, aspirin is effective for the vascular symptoms caused by the V617F mutation in *JAK2*, which would also mask the clinical signs.<sup>33 34</sup> Diagnosis will be missed if the mutations lead to the occurrence of risk factors that then cause IS. For example, heterozygous mutations in *LDLR* result in familial hypercholesterolemia that can then cause IS.<sup>35–37</sup> Clinicians often ignore the differential diagnosis of hypercholesterolemia and do not differentiate between monogenetic and complex aetiologies.

The genetic screening for Mendelian cause of stroke is critical for correct aetiological diagnosis in adult stroke patients. Almost all causes of stroke are included, such as large-artery atherosclerotic, cerebral SVD, cardioembolic, as well as coagulation disturbances, vascular malformations, metabolic disorders and large-artery non-atherosclerotic, so the panel is suitable for molecular diagnosis of all-cause IS. However, due to the significantly higher proportion of Mendelian stroke detected in patients with undetermined aetiology compared with other CCS types of stroke, and the highest genotype–phenotype matching among Mendelian stroke patients with coagulation abnormalities and cerebral SVD types, these patients are the most beneficial population in the clinical setting.

This study had some limitations. We determined whether an individual with P/LP variants predicted to be at risk for monogenetic disease had corresponding phenotypes by reviewing EHR data; however, not every phenotype would have been available in the EHR system from our registry, so most cases with systemic monogenetic diseases, such as congenital heart diseases and pseudoxanthoma elasticum, were classified as having insufficient clinical information. We also used an automated interpretation tool (InterVar), based on the ACMG/AMP guidelines, to evaluate the pathogenicity of the variants by updating the gene list. Our pipeline only used 18 categories of ACMG/AMP criteria to classify the variants, while additional

information such as familial segregation, family history and de novo status could not be obtained in this cohort for further analysis. Some variants of unknown significance (VUS) may therefore be pathogenic with inclusion of those additional criteria and may have been missed. In addition, some of the variants currently classified in ClinVar as VUS may actually be P/LP in future acquired data. Thus, the prevalence of monogenetic stroke in this cohort may have been underestimated. Furthermore, copy number variants were not analysed. In addition, our current study used targeted NGS and would have missed genes associated with other Mendelian causes. For this reason, we performed further whole genome sequencing on these samples and the data analysis is currently ongoing.<sup>38</sup> We will explore the feasibility of following up with those monogenetic stroke patients with insufficient or inconclusive clinical evidence, to either confirm or deny the genetic diagnosis of Mendelian causes through long-term medical observation.

In summary, 7.6% individuals carried at least one P/LP variant associated with monogenetic disease with stroke. Moreover, 2.2% patients in the CNSR-III cohort had clinical evidence from EHR data to support their diagnosis of monogenetic causes. The Mendelian causes of stroke are neglected in adult IS cohorts, mainly because of the late onset of symptomatic stroke, combined common vascular risks and no prominent family history.

#### Author affiliations

<sup>1</sup>China National Clinical Research Center for Neurological Diseases, Beijing Tiantan Hospital, Capital Medical University, Beijing, China

<sup>2</sup>Department of Neurology, Beijing Tiantan Hospital, Capital Medical University, Beijing, China

<sup>3</sup>Division of Endocrinology, Diabetes and Nutrition, Department of Medicine, University of Maryland School of Medicine, Baltimore, Maryland, USA

<sup>4</sup>Department of Neurology, Qindao University Medical College Affiliated Yantai Yuhuangding Hospital, Yantai, Shandong, China

<sup>5</sup>Clinical Center for Precision Medicine in Stroke, Capital Medical University, Beijing, China

X Hong-Qiu Gu @hqqg

**Acknowledgements** The authors thank clinicians, stroke patients who were enrolled in CNSR-III and Beijing Genomics institution.

**Contributors** YW take responsibility for the overall content as the guarantor. YW accepts full responsibility for the work and the conduct of the study, had access to the data, and controlled the decision to publish. YW and WL had full access to all of the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis. Concept and design: YW, WL, HL Acquisition, analysis, or interpretation of data: WL, HL, CL, JZ, ZX, HG, QX, AW, ZL, MW Drafting of the manuscript: YW, WL, HL, CL, JZ, ZX, HX, BM Critical revision of the manuscript for important intellectual content: YW, WL, HL, CL, JZ, HX, BM Statistical analysis: JZ, HX, ZX, QX, MW Administrative, technical, or material support: YJ, HG, AW, XM, JL, JJ, ZL, WZ, Beijing Genomics institution Supervision: WL, HL, CL, JZ.

**Funding** This study is supported by grants from the Capital's Funds for Health Improvement and Research (2020-1-2041), Chinese Academy of Medical Sciences Innovation Fund for Medical Sciences (2019-I2M-5-029), National Natural Science Foundation of China (81870905, U20A20358).

**Competing interests** None declared.

**Patient consent for publication** Not applicable.

**Ethics approval** This study involves human participants and was approved by IRB of Beijing Tiantan Hospital Affiliated to Capital Medical University (KY2015-001-01), and



all other research centres in accordance with the Declaration of Helsinki. Participants gave informed consent to participate in the study before taking part.

**Provenance and peer review** Not commissioned; externally peer reviewed.

**Data availability statement** The data that support the findings of this study are available from the corresponding author, YW, upon reasonable request.

**Supplemental material** This content has been supplied by the author(s). It has not been vetted by BMJ Publishing Group Limited (BMJ) and may not have been peer-reviewed. Any opinions or recommendations discussed are solely those of the author(s) and are not endorsed by BMJ. BMJ disclaims all liability and responsibility arising from any reliance placed on the content. Where the content includes any translated material, BMJ does not warrant the accuracy and reliability of the translations (including but not limited to local regulations, clinical guidelines, terminology, drug names and drug dosages), and is not responsible for any error and/or omissions arising from translation and adaptation or otherwise.

**Open access** This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>.

#### ORCID iDs

Hao Li <http://orcid.org/0000-0002-8591-4105>  
 Huichun Xu <http://orcid.org/0000-0001-8118-9607>  
 Yong Jiang <http://orcid.org/0000-0001-7700-6054>  
 Hong-Qiu Gu <http://orcid.org/0000-0003-1608-1856>  
 Qin Xu <http://orcid.org/0000-0002-8039-709X>  
 Xia Meng <http://orcid.org/0000-0002-4918-0845>  
 Zhigang Liang <http://orcid.org/0000-0002-6358-415X>  
 Yongjun Wang <http://orcid.org/0000-0002-9976-2341>

#### REFERENCES

- Baird AE. Genetics and genomics of stroke: novel approaches. *J Am Coll Cardiol* 2010;56:245–53.
- Ilinca A, Samuelsson S, Piccinelli P, et al. A stroke gene panel for whole-exome sequencing. *Eur J Hum Genet* 2019;27:317–24.
- Bordes C, Sargurupremraj M, Mishra A, et al. Genetics of common cerebral small vessel disease. *Nat Rev Neurol* 2022;18:84–101.
- Sargurupremraj M, Suzuki H, Jian X, et al. Cerebral small vessel disease genomics and its implications across the lifespan. *Nat Commun* 2020;11:6285.
- Guey S, Lesnik Oberstein SAJ, Tournier-Lasserre E, et al. Hereditary cerebral small vessel diseases and stroke: a guide for diagnosis and management. *Stroke* 2021;52:3025–32.
- Debette S, Markus HS. Stroke genetics: discovery, insight into mechanisms, and clinical perspectives. *Circ Res* 2022;130:1095–111.
- Meng L, Pammi M, Saronwala A, et al. Use of exome sequencing for infants in intensive care units: ascertainment of severe single-gene disorders and effect on medical management. *JAMA Pediatr* 2017;171:e173438.
- Wang Y, Jing J, Meng X, et al. The third China national stroke registry (CNSR-III) for patients with acute ischaemic stroke or transient ischaemic attack: design, rationale and baseline patient characteristics. *Stroke Vasc Neurol* 2019;4:158–64.
- Dichgans M, Pulit SL, Rosand J. Stroke Genetics: discovery, biology, and clinical applications. *Lancet Neurol* 2019;18:587–99.
- 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.
- Fang F, Xu Z, Suo Y, et al. Gene panel for mendelian strokes. *Stroke Vasc Neurol* 2020;5:416–21.
- Richards S, Aziz N, Bale S, et al. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American college of medical genetics and genomics and the association for molecular pathology. *Genet Med* 2015;17:405–24.
- Roach ES, Islam MP. Pseudoxanthoma Elasticum. *Handb Clin Neurol* 2015;132:215–21.
- Abdulrahim JW, Kwee LC, Alenezi F, et al. Identification of undetected monogenic cardiovascular disorders. *J Am Coll Cardiol* 2020;76:797–808.
- Chen S, Ni W, Yin X-Z, et al. Clinical features and mutation spectrum in Chinese patients with CADASIL: a multicenter retrospective study. *CNS Neurosci Ther* 2017;23:707–16.
- Hack RJ, Rutten JW, Person TN, et al. Cysteine-altering notch3 variants are a risk factor for stroke in the elderly population. *Stroke* 2020;51:3562–9.
- Mancuso M, Arnold M, Bersano A, et al. Monogenic cerebral small-vessel diseases: diagnosis and therapy. consensus recommendations of the European academy of neurology. *Eur J Neurol* 2020;27:909–27.
- Liao Y-C, Hu Y-C, Chung C-P, et al. Intracerebral hemorrhage in cerebral autosomal dominant arteriopathy with subcortical infarcts and leukoencephalopathy: prevalence, clinical and neuroimaging features and risk factors. *Stroke* 2021;52:985–93.
- Singhal S, Bevan S, Barrick T, et al. The influence of genetic and cardiovascular risk factors on the CADASIL phenotype. *Brain* 2004;127:2031–8.
- Groopman EE, Povysil G, Goldstein DB, et al. Rare genetic causes of complex kidney and urological diseases. *Nat Rev Nephrol* 2020;16:641–56.
- Rutten JW, Hack RJ, Duering M, et al. Broad phenotype of cysteine-altering notch3 variants in UK biobank: CADASIL to nonpenetrance. *Neurology* 2020;95:e1835–43.
- Zagaglia S, Selch C, Nisevic JR, et al. Neurologic phenotypes associated with COL4A1/2 mutations: expanding the spectrum of disease. *Neurology* 2018;91:e2078–88.
- Rannikmäe K, Sivakumaran V, Millar H, et al. COL4A2 is associated with lacunar ischemic stroke and deep ICH: meta-analyses among 21,500 cases and 40,600 controls. *Neurology* 2017;89:1829–39.
- Maurice P, Guilbaud L, Garel J, et al. Prevalence of COL4A1 and COL4A2 mutations in severe fetal multifocal hemorrhagic and/or ischemic cerebral lesions. *Ultrasound Obstet Gynecol* 2021;57:783–9.
- Bang OY, Chung J-W, Kim DH, et al. Moyamoya disease and Spectrums of RNF213 vasculopathy. *Transl Stroke Res* 2020;11:580–9.
- Chabriat H, Joutel A, Tournier-Lasserre E, et al. CADASIL: yesterday, today, tomorrow. *Eur J Neurol* 2020;27:1588–95.
- Mizuno T, Mizuta I, Watanabe-Hosomi A, et al. Clinical and genetic aspects of CADASIL. *Front Aging Neurosci* 2020;12:91.
- Hack RJ, Gravesteyn G, Cerfontaine MN, et al. Cerebral autosomal dominant arteriopathy with subcortical infarcts and leukoencephalopathy family members with a pathogenic notch3 variant can have a normal brain magnetic resonance imaging and skin biopsy beyond age 50 years. *Stroke* 2022;53:1964–74.
- Hu Y, Sun Q, Zhou Y, et al. Notch3 variants and genotype-phenotype features in Chinese CADASIL patients. *Front Genet* 2021;12:705284.
- Rutten JW, Van Eijsden BJ, Duering M, et al. The effect of notch3 pathogenic variant position on CADASIL disease severity: notch3 EGFr 1-6 pathogenic variant are associated with a more severe phenotype and lower survival compared with EGFr 7-34 pathogenic variant. *Genet Med* 2019;21:676–82.
- Paraskevas GP, Stefanou MI, Constantinides VC, et al. CADASIL in greece: mutational spectrum and clinical characteristics based on a systematic review and pooled analysis of published cases. *Eur J Neurol* 2022;29:810–9.
- Gravesteyn G, Hack RJ, Mulder AA, et al. Notch3 variant position is associated with notch3 aggregation load in CADASIL vasculature. *Neuropathol Appl Neurobiol* 2022;48:e12751.
- Muendlein A, Kinz E, Gasser K, et al. Occurrence of the JAK2 V617F mutation in patients with peripheral arterial disease. *Am J Hematol* 2015;90:E17–21.
- Gotlib J. Treatment and clinical endpoints in polycythemia vera: seeking the best obtainable version of the truth. *Blood* 2022;139:2871–81.
- Leren TP, Bogsrud MP. Molecular genetic testing for autosomal dominant hypercholesterolemia in 29,449 norwegian index patients and 14,230 relatives during the years 1993-2020. *Atherosclerosis* 2021;322:61–6.
- Hori M, Ohta N, Takahashi A, et al. Impact of LDLR and PCSK9 pathogenic variants in Japanese heterozygous familial hypercholesterolemia patients. *Atherosclerosis* 2019;289:101–8.
- Do R, Stitzel NO, Won H-H, et al. Exome sequencing identifies rare LDLR and apoA5 alleles conferring risk for myocardial infarction. *Nature* 2015;518:102–6.
- Cheng S, Xu Z, Liu Y, et al. Whole genome sequencing of 10K patients with acute ischaemic stroke or transient ischaemic attack: design, methods and baseline patient characteristics. *Stroke Vasc Neurol* 2021;6:291–7.