Beyond the Brain

American Heart Association | American Stroke Association

Stroke

Manuscript Submission and Peer Review System

URL: http://stroke-submit.aha-journals.org

Title: Beyond the Brain: Systematic Review of Extra-Cerebral Phenotypes associated with Monogenic Cerebral Small Vessel Disease

Manuscript number: STROKE/2020/029517R2

Author(s): Kristiina Rannikmae, David Henshall, Sophie Thrippleton, Qiu Kong, Michael Chong, Nickrooz Grami, Isaac Kuan, Tim Wilkinson, Blair Wilson, Kirsty Wilson, Guillaume Paré, and Cathie Sudlow

Beyond the Brain: Systematic Review of Extra-Cerebral Phenotypes associated with Monogenic Cerebral Small Vessel Disease

Kristiina Rannikmäe[1*] (MD, PhD), David E Henshall[1*] (MB ChB), Sophie Thrippleton[2], Qiu Ginj Kong[2], Mike Chong[3] (MsC), Nickrooz Grami[3] (BHSc), Isaac Kuan[2] (MBChB), Tim Wilkinson[1] (MBChB), Blair Wilson[2] (MB ChB), Kirsty Wilson[1] (MBChB), Guillaume Paré[3] (MD), Cathie Sudlow[1] (FRSE).

1. Centre for Medical Informatics, Usher Institute, University of Edinburgh, United Kingdom.

2. Edinburgh Medical School, University of Edinburgh, United Kingdom.

3. Genetic and Molecular Epidemiology Laboratory, McMaster University, Canada.

*Authors contributed equally

Corresponding author: Kristiina Rannikmäe; kristiina.rannikmae@ed.ac.uk; NINE Edinburgh BioQuarter, 9 Little France Road, Edinburgh, EH16 4UX; +44(0)1316517881

Cover title: Extra-Cerebral Phenotypes in Small Vessel Disease

Figures: 4; **Tables**: 2; Supplemental Material;

Key words: Stroke, Gene, Systematic Review, SVD, Mendelian, Phenotype.

Subject terms: Genetics, Cerebrovascular Disease/Stroke

Word count: 5500

Twitter: @nickroozgrami; @krannikmae; @tim_neuro;

Abstract

Background and Purpose: An important minority of cerebral small vessel disease (cSVD) is monogenic. Many monogenic cSVD genes are recognized to be associated with extra-cerebral phenotypes. We assessed the frequency of these phenotypes in existing literature.

Methods: We performed a systematic review following the PRISMA guidelines, searching Medline/Embase for publications describing individuals with pathogenic variants in *COL4A1/2*, *TREX1*, *HTRA1*, *ADA2* and *CTSA* genes (PROSPERO 74804). We included any publication reporting on ≥1 individual with a pathogenic variant and their clinically-relevant phenotype. We extracted individuals' characteristics, and information about associated extra-cerebral phenotypes and stroke/TIA. We noted any novel extra-cerebral phenotypes and looked for shared phenotypes between monogenic cSVDs.

Results: After screening 6048 publications, we included 96 *COL4A1* (350 individuals), 32 *TREX1* (115 individuals), 43 *HTRA1* (38 homozygous/61 heterozygous individuals), 16 *COL4A2* (37 individuals), 119 *ADA2* (209 individuals) and 3 *CTSA* (14 individuals) publications. The majority of individuals originated from Europe/North America, except for *HTRA1*, where most were from Asia. Age varied widely, *ADA2* individuals being youngest and heterozygous *HTRA1/CTSA* individuals oldest. Sex distribution appeared equal. Extra-cerebral phenotypes were common: 14% to 100% of individuals with a pathogenic variant manifested at least one extra-cerebral phenotype (14% *COL4A2*, 43% *HTRA1* heterozygotes, 47% *COL4A1*, 57% *TREX1*, 91% *ADA2*, 94% *HTRA1* homozygotes and 100% *CTSA* individuals). Indeed, for 4 of 7 genes an extra-cerebral phenotype was observed more frequently than stroke/TIA. Ocular, renal, hepatic, muscle and hematological systems were each involved in more than one monogenic cSVD.

Conclusions: Extra-cerebral phenotypes are common in monogenic cSVD with extra-cerebral system involvement shared between genes. However, inherent biases in the existing

literature mean that further data from large-scale population-based longitudinal studies collecting health outcomes in a systematic unbiased way is warranted. The emerging knowledge will help to select patients for testing, inform clinical management, and provide further insights into the underlying mechanisms of cSVD.

Non-standard Abbreviations and Acronyms:

cSVD = cerebral small vessel disease

CADASIL = cerebral autosomal dominant arteriopathy with subcortical infarcts and leukoencephalopathy

ICH = intracerebral hemorrhage

HetZ=heterozygous

HomZ=homozygous/compound heterozygous

Introduction

Non-amyloid cerebral small vessel disease (hereafter referred to in this report as cSVD) is a term used to describe a variety of pathological processes that affect the deep small penetrating arteries, arterioles, venules and capillaries of the brain. The main clinical phenotypes of cSVD include small vessel ischemic stroke, deep intracerebral hemorrhage and vascular cognitive impairment. It can also manifest as a primarily radiological syndrome, with lesions in the subcortical regions of the brain[1,2]. The overall burden of cSVD is growing as the world's population continues to age. Other than management of hypertension, we currently lack effective treatments to reduce the risk of cSVD[2]. Hence pathways involved in cSVD pathogenesis must be better understood to develop new effective prevention and treatment strategies.

A minority of cSVD is monogenic – i.e. caused by rare pathogenic variants in several genes thought to cause cSVD as a primary syndrome or a dominant feature alongside other systemic manifestations[3]. The first such gene discovered was *NOTCH3* in 1996, causing cerebral autosomal dominant arteriopathy with subcortical infarcts and leukoencephalopathy (CADASIL)[4]. Subsequently, over the last >20 years, further genes have been reported, including (in order of their discovery) *COL4A1*, *TREX1*, *HTRA1*, *COL4A2*, *ADA2* and *CTSA*. Although these genes are known for causing a primarily cSVD phenotype, all but *NOTCH3* are also recognized to a smaller or larger extent to be associated with symptomatically manifest extra-cerebral phenotypes, and there is evidence of *NOTCH3* associated asymptomatic skin involvement in CADASIL[5,6]. Understanding which genes are associated with which extra-cerebral phenotypes, and knowing the proportion of individuals with a pathogenic variant in these genes who exhibit specific phenotypes during their lifetime (i.e. genetic variant penetrance), can inform the selection of patients to test for monogenic diseases, as well as provide information relevant for counselling and screening of both

affected and non-affected family members. In addition, improved understanding of monogenic cSVDs has broader implications for the direction of future research into common cSVD. Finally, as treatment options for *ADA2*-related disease already exist[7], and are emerging for other monogenic cSVDs (e.g. phenyl butyric acid for *COL4A1*[8], and an anthracycline antibiotic for *TREX1*-associated disease [https://clinicaltrials.gov/ct2/show/study/NCT02723448]), understanding the full phenotypic spectrum of monogenic cSVD manifestations is of utmost importance in enabling relevant studies to monitor treatment effects and side effects. For example, preventative oral treatment of Col4a1 mutant mice with phenylbutyric acid has shown differential effects on different organ system pathologies, reducing adult ICH with no effect on eye and kidney defects[8].

We undertook a systematic review of the literature, aiming to identify all reported individuals with a pathogenic variant in a monogenic cSVD gene, and to summarize their extra-cerebral disease phenotypes. Our aim was to: (i) improve our understanding of the frequency of already recognized associated extra-cerebral phenotypes; (ii) discover novel extra-cerebral phenotypes; and (iii) search for shared clinical presentations between different monogenic cSVDs.

Methods

All supporting data are available within the article and its online supplementary files.

Study protocol

The present study is a systematic review including all known monogenic cSVD genes, whose protocol is available on PROSPERO (https://www.crd.york.ac.uk/PROSPERO/display_record.php?RecordID=74804). In this review, we included only genes that are widely recognized to cause monogenic cSVD as a

dominant phenotype and are associated with at least one clinically-relevant extra-cerebral phenotype. We followed the PRISMA guidelines.

Search strategy

We searched Ovid Medline and Embase (from inception until July 2019) for publications about humans with pathogenic variants in *COL4A1*, *TREX1*, *HTRA1*, *COL4A2*, *ADA2* and *CTSA* genes. Our search strategy included: (a) text words, phrases and Medical Subject Headings (MeSH) for relevant monogenic syndromes/diseases associated with variants in these genes; or (b) text words, phrases and MeSH terms associated with cSVD combined with those for genes of interest and their proteins (https://www.ahajournals.org/journal/str:Appendix I). We also checked the bibliographies of all relevant studies and reviews identified and used Google Scholar for forward citation searching. We included conference abstracts and foreign language publications. At least two authors (DEH, IK, BW, KR) screened all publications, reviewing all titles and abstracts and full texts as required. They worked independently, blinded to each other's decisions, and then reviewed disagreements through discussion and mutual consensus.

Study inclusion and exclusion criteria

We included all publications of any type (including abstracts, letters, clinical notes, comments, case reports, original articles describing case series or other study designs) that reported on ≥1 human individual(s) (hereafter referred to as 'individuals') who were heterozygous (HetZ), homozygous or compound heterozygous (HomZ) for a variant in *COL4A1*$^{\text{HetZ}}$, *TREX1*$^{\text{HetZ}}$, *HTRA1*$^{\text{HetZ or HomZ}}$, *COL4A2*$^{\text{HetZ}}$, *ADA2*$^{\text{HomZ}}$ or *CTSA*$^{\text{HetZ}}$, and described clinically-relevant phenotype(s) of the included individual(s).

From each eligible publication we included individuals carrying a variant which the study authors considered to be probably or definitely pathogenic, and reported on the clinically-

relevant phenotype(s). This reporting could range from a simple comment to say that an individual was 'healthy' (e.g. asymptomatic family members of affected individuals) to a detailed case report. We excluded individuals whose variant was not confirmed by genotyping and those with a presumed pathogenic variant in ≥1 gene. We excluded publications studying only non clinically-relevant or purely radiological phenotypes (e.g. studies performing specialized psychological tests or measurements of specific brain structures). Where the same individual(s) were described in >1 publication, we included each individual only once, summarizing the information from all relevant publications.

Data extraction

From each included **publication**, we extracted data on the first author, publication year, journal and number of eligible individuals and pedigrees.

For each eligible **individual**, we extracted data on their genetic variant and the resulting amino acid change (if relevant), country of origin and ethnicity, sex, age at clinical assessment, clinically-relevant extra-cerebral phenotype(s) and stroke, and age at respective phenotype diagnosis.

One of KR, DEH, ST, QGK or IK extracted data from each publication. To assess agreement in assigning the status of major phenotype groups, at least two members of the team extracted data from 10% of publications, working independently and blinded to each other's decisions.

Data synthesis

We counted the number of relevant publications (overall and by specialty), pedigrees, individuals and variants per gene. We summarized the characteristics of the individuals for each gene, including their geographic region of origin, sex and age at assessment (mean, median and range).

We looked up variant frequencies among the gnomAD database v2.1.1 non-neuro 101,605 participants[9], summarizing the proportion of variants present with any allele frequency, and the proportion of variants with allele frequencies of >0.1% and >0.5% for autosomal dominant and recessive inheritance mode genes respectively. The 0.1% and 0.5% thresholds were chosen as a pragmatic cut-off suggestive of the possibility of the variant being benign and we performed subgroup analyses to see if excluding these variants changes the overall results.

For each gene, we summarized the clinically-relevant phenotype (hereafter referred to in this report as 'phenotype') data with respect to frequency (i.e. the percentage of individuals with a pathogenic variant in whom the respective associated phenotype was reported). We did not make assessments at the variant level because of the high proportion of private variants (e.g. variants described in only one pedigree) with only small numbers of pedigrees and individuals.

Frequency of already recognized associated extra-cerebral phenotypes and of stroke/transient ischemic attack (TIA)

We used the Online Mendelian Inheritance in Man database (OMIM, https://www.omim.org/) to compile a list of extra-cerebral phenotypes for each gene (treating associations for *HTRA1*^{HetZ}/HTRA1^{HomZ} and *COL4A1/COL4A2* as applying to both genes respectively), that are already recognized to be associated with pathogenic variants in that gene (https://www.ahajournals.org/journal/str:Table I, Appendix II). Our primary definition of hemorrhagic stroke for these purposes included porencephaly. We performed additional subgroup analyses excluding individuals presenting only with porencephaly, because while it often occurs as a result of intracerebral hemorrhage, it may also result from other destructive influences.[10] For these selected extra-cerebral phenotypes and for stroke/TIA (including stroke subtype where information was available), we summarized data on the presence versus

absence versus uninvestigated status of the phenotype. A phenotype was considered uninvestigated if no mention was made about its presence or absence. We assessed whether the proportion of uninvestigated individuals for each phenotype changed significantly when excluding individuals reported only in abstract form (Fisher's exact test significance p-value ≤0.01).

Novel potentially associated extra-cerebral phenotypes

We identified extra-cerebral phenotypes not already recognized by OMIM, and considered these potentially associated when reported in >5% of individuals and ≥2 pedigrees per gene. We selected the 5% cut-off as a pragmatic method to identify phenotypes of potential interest, due to the substantial challenges in obtaining robust estimates of background disease prevalence in populations of different ethnicities and age for comparison. (Indeed, even if such robust estimates were available, meaningful comparison with the reported frequency among individuals in this study would be limited by incomplete age/ethnicity reporting). Based on the clinical expert knowledge of the authors, we also noted any particularly rare or unusual disease phenotypes present in ≥2 individuals from ≥2 pedigrees.

Shared extra-cerebral phenotypes between monogenic cSVDs

We looked for shared patterns and differences in extra-cerebral manifestations among the monogenic cSVD genes. We stratified these manifestations into presumed vascular versus other phenotypes, and by anatomical organ system involvement (https://www.ahajournals.org/journal/str:Supplemental methods). We also explored corresponding gene expression data from the Expression Atlas (https://www.ebi.ac.uk/gxa/home).

Results

From 6048 publications identified for screening, we included 305 eligible publications: 96 *COL4A1*, 32 *TREX1*, 25 *HTRA1*^{HomZ}, 16 *COL4A2*, 119 *ADA2*, 18 *HTRA1*^{HetZ}, and three *CTSA* publications (Figure 1). As expected, there was a positive correlation between the number of years since the gene was first discovered and the number of publications to date (https://www.ahajournals.org/journal/str: Figure I). An exception was *ADA2*, which was first reported as recently as 2013 but still has the largest number of publications. A likely explanation is the combination of existing treatment options and the severe early-onset systemic phenotype of *ADA2*, prompting more widespread genetic testing. The percentage of publications in specialist neurology or stroke journals varied by gene from 7-100% (*ADA2* 7%; *COL4A2* 44%; *COL4A1* 60%; *TREX1* 75%; *HTRA1*^{HomZ} 84%; *HTRA1*^{HetZ} 94%; *CTSA* 100%). The number of individuals per gene ranged from 14 (*CTSA*) to 350 (*COL4A1*) and the number of pedigrees from three (*CTSA*) to 191 (*COL4A1*). The proportion of pedigrees carrying a private variant ranged from 0% to 83% (*CTSA* 0%; *TREX1* 44%; *ADA2* 65%; *COL4A1/HTRA1*^{HomZ} 74%; *HTRA1*^{HetZ} 79%; *COL4A2* 83%) (Figure 2).

We identified 0% to 25% variants per gene in the gnomAD database, however only one *COL4A1* variant and three *COL4A2* variants were present at allele frequencies >0.1% (https://www.ahajournals.org/journal/str:Table II).

Summary of individuals' characteristics

The majority (or all) of *COL4A1*, *TREX1*, *COL4A2*, *ADA2* and *CTSA* individuals were European or North American (77%[239/309], 66%[76/115], 65%[22/34], 61%[121/200] and 100%[14/14] respectively), while the majority (76%[29/38]) of *HTRA1*^{HZ} and just over a half (54%[33/61]) of *HTRA1*^{HetZ} individuals were Asian. However, data on individuals' geographic region of origin was not reported for over a quarter of individuals for 5/7 genes; in these cases it was assumed from the first author's country of origin (Table 1). Where the number of individuals was large enough to enable robust comparison (>100 individuals per

gene), sex distribution was generally approximately equal (44-50% female) (Table 1). Data about the age at the time of assessment was missing for over a quarter of *COL4A1*, *TREX1* and *HTRA1*^{HomZ} individuals. Accepting this limitation, mean (median) age ranged from 17 (13) years for the 209 *ADA2* individuals to 60 (61) years for the 61 *HTRA1*^{HetZ} individuals. Age ranges were very broad for *COL4A1*, *COL4A2* and *ADA2*, ranging from less than one to over 70 years. (Table 1). Age at specific phenotype diagnoses was generally not reported, hence it was not possible to comment on typical ages of phenotype presentation (https://www.ahajournals.org/journal/str: Table VI).

Frequency of already recognized associated extra-cerebral phenotypes

Comparison of results of the dual extraction showed 95% agreement.

On an individual-level, extra-cerebral phenotypes were common: 14-100% of individuals per gene were reported to manifest at least one associated extra-cerebral phenotype (*COL4A2* 14%[5/37]; *HTRA1*^{HetZ} 43%[26/61]; *COL4A1* 47%[147/311]; *TREX1* 57%[21/37]; *ADA2* 91%[187/205]; *HTRA1*^{HomZ} 94%[33/35] and *CTSA* 100%[14/14]) (Figure 3). Excluding six *COL4A2* and one *COL4A1* individuals carrying variants with allele frequencies of >0.1% in the gnomAD population changed these proportion to 16% for *COL4A2* (3/31 individuals) while the proportion for *COL4A1* remained the same (47%, 147/310 individuals).

The frequency of specific extra-cerebral phenotypes ranged widely across genes: 3-19% for *COL4A1*; 0-68% for *TREX1*; 37-74% for *HTRA1*^{HomZ}; 0-8% for *COL4A2*; 11-66% for *ADA2*; 7-31% for *HTRA1*^{HetZ}; and 64-93% *CTSA* (Figure 4, https://www.ahajournals.org/journal/str:Figure II). Detailed definitions of included phenotypes are available from https://www.ahajournals.org/journal/str: Table I.

For 4/7 genes (*TREX1*, *HTRA1*^{HomZ}, *ADA2*, *CTSA*) an extra-cerebral phenotype was reported more frequently than stroke/TIA.

Frequency of stroke/TIA

The frequency of stroke/TIA ranged from 40-57% for five genes (*ADA2*, *COL4A1*, *HTRA1*^{HetZ}, *CTSA* and *COL4A2*). In contrast, only 26% (10/38) of *HTRA1*^{HomZ} and 4% (5/115) of *TREX1* individuals were reported to have suffered a stroke/TIA. The latter proportion did not include episodes of progressive neurological deficit characteristic for *TREX1*, often also attributed to an underlying vascular cause. The majority of *COL4A1/2* strokes (73%[115/157] and 100%[21/21] respectively) were hemorrhagic, while ischemic stroke/TIA was more commonly reported for the other genes. Hemorrhagic stroke remained the most common stroke type for COL4A1/2 individuals when excluding those presenting only with porencephaly. (Figure 4, https://www.ahajournals.org/journal/str:Table III, Figure II).

Proportion of individuals with uninvestigated phenotypes

We estimated the proportion of individuals where a phenotype was uninvestigated, and therefore its presence or absence unknown. This proportion ranged from 49-96% for *COL4A1*, 20-98% for *TREX1*, 11-63% for *HTRA1*^{HomZ}, 43-100% for *COL4A2*, 26-86% for *ADA2*, 16-89% for *HTRA1*^{HetZ}, and 0-7% for *CTSA* phenotypes (https://www.ahajournals.org/journal/str: Figure II). There was no statistically significant difference when excluding individuals reported only in an abstract form, except for degenerative spine disease in *HTRA1*^{HomZ} individuals (37% vs 8% uninvestigated individuals in all publications versus in publications excluding abstracts, respectively) (https://www.ahajournals.org/journal/str:Table IV).

Novel potentially associated extra-cerebral phenotypes

For the complete list of novel potentially associated extra-cerebral phenotypes see https://www.ahajournals.org/journal/str:Supplemental Results/Table V. More noteworthy

associations were: *COL4A1* and congenital kidney/urinary tract anomalies; *TREX1* and hypertension, anemia, gastrointestinal bleeding; *ADA2* and arthritis/arthralgia, lymphadenopathy, Diamond Blackfan anemia; and *CTSA* and venous thrombosis. Concerning rare disease phenotypes, two *COL4A1* individuals had spontaneous bleeding into extra-cerebral organs, and two *TREX1* individuals had avascular bone necrosis in the absence of risk factors, suggestive of a systemic vulnerable vasculature[11].

Shared extra-cerebral phenotypes between monogenic cSVDs

Of the vascular phenotypes, Raynaud's phenomenon and ocular, renal and hepatic involvement were reported in association with ≥2 monogenic cSVD genes. Among non-vascular phenotypes, muscle and hematological pathologies were reported in association with ≥2 monogenic cSVD genes. (Table 2).

Protein expression data from the Human Protein Atlas and Human Proteome Map projects did not clearly correspond to these phenotype associations. However, interpretation is limited by incomplete coverage of genes of interest as well as of tissue and cell types in the included experiments (https://www.ahajournals.org/journal/str:Figure III).

Discussion

To our knowledge, this is the first systematic review to summarize extra-cerebral manifestations across a range of monogenic cSVD genes. We found that extra-cerebral manifestations were frequently reported, in many cases more often than stroke/TIA. We also found shared extra-cerebral phenotypes between monogenic cSVDs caused by variants in different genes. Our results suggest that genes traditionally viewed as monogenic stroke genes may be more appropriately viewed as genes underlying a systemic monogenic vasculopathy, sometimes with additional extra-vascular manifestations.

Extracerebral phenotypes appeared common despite: (i) >40% of publications coming from neurology/stroke journals, which might favor the reporting of cerebral over extra-cerebral phenotypes; (ii) a large proportion of individuals with uninvestigated phenotypes, meaning the true prevalence of these phenotypes may be even higher than estimated.

The observation of widespread pleiotropy of monogenic cSVD genes has implications for clinical practice when diagnosing and managing patients with stroke. It should also inform research into the pathogenesis of monogenic cSVD genes more broadly. Although these genes have distinct functions, several of them may ultimately converge on a common pathologic pathway affecting the cerebrovascular (and perhaps more general vascular) matrisome to produce the cSVD phenotype.[12] Future research may also identify shared mechanisms for the array of extra-cerebral manifestations. Finally, observations from large-scale genetic associations studies have shown common variation in monogenic cSVD genes to be associated with sporadic cSVD.[13-16] Whether the same applies to their pleiotropic associations is yet to be seen, but examples are beginning to emerge in the common disease literature of common genetics associations for diseases previously considered to be entirely separate entities.[17]

The high proportion of private variants, together with the observation that this proportion has not fallen over time, is also noteworthy, particularly as regards the design of gene panels for clinical testing. The continuing emergence of novel pathogenic variants suggests that sequencing-based gene panels may be preferable over known variant microarray genotyping.

The predominance of individuals originating from Europe/North America could suggest a higher prevalence among European origin populations. Alternatively, it could arise from the widespread issue of research bias in genetics research, where these populations are the most widely studied. In contrast to other genes, *HTRA1* variants appeared more common in Asia, and in this case, the first report of *HTRA1* pathogenic variants from Asia could have biased

subsequent further testing towards individuals of this ethnicity. This potential for testing bias is supported by the observation that pathogenic *HTRA1* variants have a higher frequency in non-Finnish Europeans than in East Asians in the gnomAD database.[18]

HTRA1^{HomZ} individuals appeared to have a lower frequency of stroke/TIA than *HTRA1*^{HetZ} individuals (26%[10/38] versus 56%[34/61], respectively). This difference may be caused due to selective testing bias, since the first publication of *HTRA1* pathogenic heterozygous variants did not report an association with extra-cerebral features[19], as well as by the higher mean age of *HTRA1*^{HetZ} individuals.

All genes apart from *ADA2* are expressed more broadly than in the brain. *ADA2* is primarily expressed by blood cells. Therefore, it is not surprising that these genes have prevalent manifestations extra-cerebrally. It is interesting however that *NOTCH3*, which has a vascular expression broader than the brain, leads to a primarily brain-restricted phenotype. It has been suggested that vessel-specific expression of the matrisome may explain some of the variable clinical presentations and research into these expression patterns might provide further insights.[12]

The strengths of our study are: (1) a comprehensive search strategy including abstracts and foreign language papers; (2) a systematic approach to data extraction, including consideration the proportion on individuals for whom the phenotype is uninvestigated; (3) a consistent approach not limited to just one gene, but across several cSVD genes to allow similarities and differences to be assessed. As such, our review complements the knowledge from other comprehensive recent review papers for *ADA2*[20] and *COL4A1/2*[21].

Our study also has some limitations: (1) the focus of this work on extra-cerebral phenotypes meant that we included only stroke and TIA as the cerebral manifestations of cSVD (excluding cognitive and radiological phenotypes, which would require the development of

thresholds for normal versus pathological); (2) we looked at gene-phenotype level associations rather than variant-phenotype associations because of small numbers of reported individuals with non-private variants; (3) our stringent efforts to avoid double counting of any individual means that we may have underestimated the overall number of individuals; (4) poor case characterization (including geographic origin and ethnicity) in many included studies; (5) the inherent biases of case reports and series including case investigation, publication and reporting biases; (6) we considered any variant pathogenic if reported as such by the author, hence may have included variants with less supportive evidence of pathogenicity. However, subgroup analyses informed by allele frequencies in the gnomAD population did not change the overall conclusions.

Our study highlights some general challenges of interpreting research into gene-disease associations for rare monogenic disease. First, case reports are often difficult to publish, limited by word count, and lack a reporting structure. Guidelines (e.g. CARE guidelines)[22,23] for publishing case reports are not always enforced and are generally aimed at case reports of treatment effects[24] rather than reports of rare genetic diseases (e.g. the guidelines lack advice about reporting mutation details, case ethnicity, and age of phenotype onset). Second, such reports are inherently biased, because case selection for investigation is not random, and diagnosed cases are more likely to be reported in publications if they are part of a large case series or present a novel phenotype. Third, due to small numbers of rare disease cases, true disease associations are difficult to establish, and alleged associations may arise through chance. Existing specialist-curated databases such as OMIM have established various criteria trying to overcome this, but even these involve subjectivity and are difficult to reproduce[25].

A study design that could overcome some of these issues is a large-scale population-based longitudinal study, where genetic information and health outcomes are collected in a systematic and unbiased way. The emergence of prospective population-based studies with

bio-samples yielding genetic data at scale, such as the UK Biobank (www.ukbiobank.ac.uk), will allow further assessment of the true clinical consequences of rare variants in monogenic cSVD genes, complementing the knowledge acquired from more traditional case reports and series.

In conclusion, we found that extracerebral phenotypes are common in monogenic cSVD. This knowledge can help select patients for testing for the genetic variants as well as inform clinical management of already diagnosed cases (e.g., encourage screening for pre-symptomatic extra-cerebral involvement). Notwithstanding the difficulties in establishing gene-disease associations robustly, we found new relationships of potential interest between monogenic cSVD genes and extra-cerebral phenotypes. Finally, we highlight common phenotype patterns across different monogenic cSVDs, where further research may provide additional mechanistic insights. We recommend that future investigations should focus on emerging population-based longitudinal resources to complement the existing insights gained from more traditional study designs for rare disease.

Acknowledgements: Dr Chung-Fen Tsai and Natalia Monteiro Calanzani translated foreign language articles.

Sources of Funding: KR: Health Data Research UK (Medical Research Council-funded) fellowship MR/S004130/1; GP, MC, NG: Canadian Institutes of Health Research, Natural Sciences and Engineering Research Council of Canada.

Conflicts-of-Interest/Disclosures: Michael Chong has received personal fees from Bayer.

Ethics: This is a systematic review based on data from published studies and does not require an approval from an ethical standards committee.

References

1. Pantoni L. Cerebral small vessel disease: From pathogenesis and clinical characteristics to therapeutic challenges. *Lancet Neurol*. 2010;9:689-701.
2. Shi Y, Wardlaw JM. Update on cerebral small vessel disease: A dynamic whole-brain disease. *Stroke Vasc Neurol*. 2016;1:83-92.
3. Haffner C, Malik R, Dichgans M. Genetic factors in cerebral small vessel disease and their impact on stroke and dementia. *J Cereb Blood Flow Metab*. 2016;36:158-171.
4. Joutel A, Corpechot C, Ducros A, Vahedi K, Chabriat H, Mouton P, Alamowitch S, Domenga V, Cécillion M, Marechal E, et al. Notch3 mutations in Cadasil, a hereditary adult-onset condition causing stroke and dementia. *Nature*. 1996;383:707-710.
5. Søndergaard CB, Nielsen JE, Hansen CK, Christensen H. Hereditary cerebral small vessel disease and stroke. *Clin Neurol Neurosurg*. 2017;155:45-57.

6. Tan RY, Markus HS. Monogenic causes of stroke: Now and the future. *J Neurol*. 2015;262:2601-2616.

7. Ombrello AK, Qin J, Hoffmann PM, Kumar P, Stone D, Jones A, Romeo T, Barham B, Pinto-Patarroyo G, Toro C, et al. Treatment strategies for deficiency of adenosine deaminase 2. *N Engl J Med*. 2019;380:1582-1584.

8. Jones FE, Murray LS, McNeilly S, Dean A, Aman A, Lu Y, Nikolova N, Malomgré R, Horsburgh K, Holmes WM, et al. 4-Sodium phenyl butyric acid has both efficacy and counter-indicative effects in the treatment of Col4a1 disease. *Hum Mol Genet*. 2019;28:628-638.

9. Karczewski KJ, Francioli LC, Tiao G, Cummings BB, Alföldi J, Wang Q, Collins RL, Laricchia KM, Ganna A, Birnbaum DP,et al. Variation across 141,456 human exomes and genomes reveals the spectrum of loss-of-function intolerance across human protein-coding genes. *BioRxiv*. 2019:531210.

10. Pasternak JF, Mantovani JF, Volpe JJ. Porencephaly From Periventricular Intracerebral Hemorrhage in a Premature Infant. *Am J Dis Child*. 1980;134:673-675.

11. Moya-Angeler J, Gianakos AL, Villa JC, Ni A, Lane JM. Current concepts on osteonecrosis of the femoral head. *World J Orthop*. 2015;6:590-601.

12. Joutel A, Haddad I, Ratelade J, Nelson MT. Perturbations of the cerebrovascular matrisome: A convergent mechanism in small vessel disease of the brain? *J Cereb Blood Flow Metab*. 2016;36:143-157.

13. Rannikmäe K, Davies G, Thomson PA, Bevan S, Devan WJ, Falcone GJ, Traylor M, Anderson CD, Battey TWK, Radmanesh F, et al. Common variation in COL4A1/COL4A2 is associated with sporadic cerebral small vessel disease. *Neurology*. 2015;84:918-926.

14. Rannikmäe K, Sivakumaran V, Millar H, Malik R, Anderson CD, Chong M, Dave T, Falcone GJ, Fernandez-Cadenas I, Jimenez-Conde J, et al. COL4A2 is associated with lacunar ischemic stroke and deep ICH: Meta-analyses among 21,500 cases and 40,600 controls. *Neurology*. 2017;89:1829-1839.

15. Malik R, Rannikmäe K, Traylor M, Georgakis MK, Sargurupremraj M, Markus HS, Hopewell JC, Debette S, Sudlow CLM, Dichgans M, et al. Genome-wide meta-analysis identifies 3 novel loci associated with stroke. *Ann Neurol*. 2018;84:934-939.

16. French CR, Seshadri S, Destefano AL, Fornage M, Arnold CR, Gage PJ, Skarie JM, Dobyns WB, Millen KJ, Liu T, et al. Mutation of FOXC1 and PITX2 induces cerebral small-vessel disease. *J Clin Invest*. 2014;124:4877-4881.

17. Paré G, Bhatt DL. Linking spontaneous coronary artery dissection, cervical artery dissection, and fibromuscular dysplasia: Heart, brain, and kidneys. *J Am Coll Cardiol*. 2019;73:67-69

18. Grami N, Chong M, Lali R, Mohammadi-Shemirani P, Henshall DE, Rannikmäe K, Paré G. Global Assessment of Mendelian Stroke Genetic Prevalence in 101,635 Individuals from 7 Ethnic Groups. *Stroke*. 2020;51:1290-1293.

19. Verdura E, Hervé D, Scharrer E, Amador MDM, Guyant-Maréchal L, Philippi A, Corlobé A, Bergametti F, Gazal S, Prieto-Morin C, et al. Heterozygous HTRA1 mutations are associated with autosomal dominant cerebral small vessel disease. *Brain*. 2015;138:2347-2358.

20. Hashem H, Kelly SJ, Ganson NJ, Hershfield MS. Deficiency of Adenosine Deaminase 2 (DADA2), an Inherited Cause of Polyarteritis Nodosa and a Mimic of Other Systemic Rheumatologic Disorders. *Curr Rheumatol Rep*. 2017;19:70.

21. Jeanne M, Gould DB. Genotype-phenotype correlations in pathology caused by collagen type IV alpha 1 and 2 mutations. *Matrix Biol*. 2017;57:29-44.

22. Riley DS, Barber MS, Kienle GS, Aronson JK, von Schoen-Angerer T, Tugwell P, Kiene H, Helfand M, Altman DG, Sox H, et al. Care guidelines for case reports: Explanation and elaboration document. *J Clin Epidemiol*. 2017;89:218-235.

23. Gagnier JJ, Kienle G, Altman DG, Moher D, Sox H, Riley D, CARE Group. The care guidelines: Consensus-based clinical case reporting guideline development. *Glob Adv Health Med*. 2013;2:38-43.

24. Murad MH, Sultan S, Haffar S, Bazerbachi F. Methodological quality and synthesis of case series and case reports. *BMJ Evid Based Med*. 2018;23:60-63.

25. Amberger JS, Bocchini CA, Scott AF, Hamosh A. Omim.Org: Leveraging knowledge across phenotype-gene relationships. *Nucleic Acids Res*. 2019;47:D1038-D1043.

Figure legends:

Figure 1. Selection of included studies.

HetZ=heterozygous; HomZ=homozygous/compound heterozygous; *Reference lists and Google Scholar search. †We identified CADASIL (*NOTCH3*), *FOXC1* and *PITX2* individuals as part of another systematic review but excluded them for this review (see Methods, Study protocol). ‡Four publications reported both *COL4A1/2* individuals, hence the total number of unique publications (305) is not the sum total of publications per gene (309). 112/305 publications were conference abstracts: *COL4A1/2* (28), *TREX1* (10), *HTRA1*^{HomZ} (4), *CECR1* (65), *HTRA1*^{HetZ} (5);

Figure 2. Number of included individuals and pedigrees.

Year=year gene first reported to be associated with cSVD; HetZ=heterozygous; HomZ=homozygous/compound heterozygous; N=number of publications; n/r=not reported;

This figure is reporting on the DNA change, not the resulting amino acid change. Where the DNA change was not reported, the individual was included but variant considered 'n/r'. For compound heterozygotes, if either of the variants was private, the pedigree was considered to carry a private variant. Where a range is given for the number of variants/pedigrees, publications had not clearly reported these data (e.g. a publication reporting 5 cases with "pathogenic *COL4A1* variants", but not specifying the variants, could refer to 5 cases all carrying the same variant or each carrying a private variant, resulting in a range of 1-5 private variants).

Figure 3. Proportion of individuals with an extra-cerebral phenotype

HetZ=heterozygous; HomZ=homozygous/compound heterozygous; $N_{phenotype}$=number of individuals with ≥1 extra-cerebral phenotype; N_{all}=total number of individuals with relevant phenotype data.

Figure 4. Frequency of Stroke/TIA and Extra-Cerebral Phenotypes.

HetZ=heterozygous; HomZ=homozygous/compound heterozygous;

Table 1. Summary of case characteristics.

	COL4A1	*TREX1*	*HTRA1*^{HomZ}	*COL4A2*	*ADA2*	*HTRA1*^{HetZ}	*CTSA*
	N=350	N=115	N=38	N=37	N=209	N=61	N=14
REGION OF ORIGIN[*]							
European	69% (212/309)	40% (46/115)	11% (4/38)	53% (18/34)	49% (98/200)	46% (28/61)	100% (14/14)
Asian	15% (46/309)	13% (15/115)	76% (29/38)	24% (8/34)	22% (44/200)	54% (33/61)	0% (0/14)
Turkish	7% (23/309)	<1% (1/115)	5% (2/38)	0% (0/34)	13% (27/200)	0% (0/61)	0% (0/14)
North American	9% (27/309)	26% (30/115)	3% (1/38)	12% (4/34)[8]	12% (23/200)	0% (0/61)	0% (0/14)
South American	<1% (1/309)	16% (18/115)	0% (0/38)	0% (0/34)	1% (3/200)	0% (0/61)	0% (0/14)
African	0% (0/309)	0% (0/115)	0% (0/38)	0% (0/34)	3% (5/200)	0% (0/61)	0% (0/14)
Australian	0% (0/309)	4% (5/115)	5 % (2/38)	12% (4/34)	0% (0/200)	0% (0/61)	0% (0/14)
SEX							
Female / Male	51%/49% (139/135)	44%/56% (50/63)	56%/44% (19/15)	39%/61% (14/22)	50%/50% (91/90)	31%/69% (19/42)	86%/14% (12/2)
% cases sex not reported	22% (76/350)	2% (2/115)	11% (4/38)	3% (1/37)	13% (28/209)	-	-
AGE AT TIME OF ASSESSMENT[†]							
Mean age (years)	22	43	37	24	17	60	57
Median age (years)	14	-	33[‡]	15	13	61	55

Age range (years)	<1 - 77	-	24 – 52‡	<1 - 72	<1 - 76	31 - 86	39 - 74
% cases age not reported	31%	13%	13%	24%	11%	5%	0%

N=number of cases; HetZ: heterozygous; HomZ: homozygous/compound heterozygous;

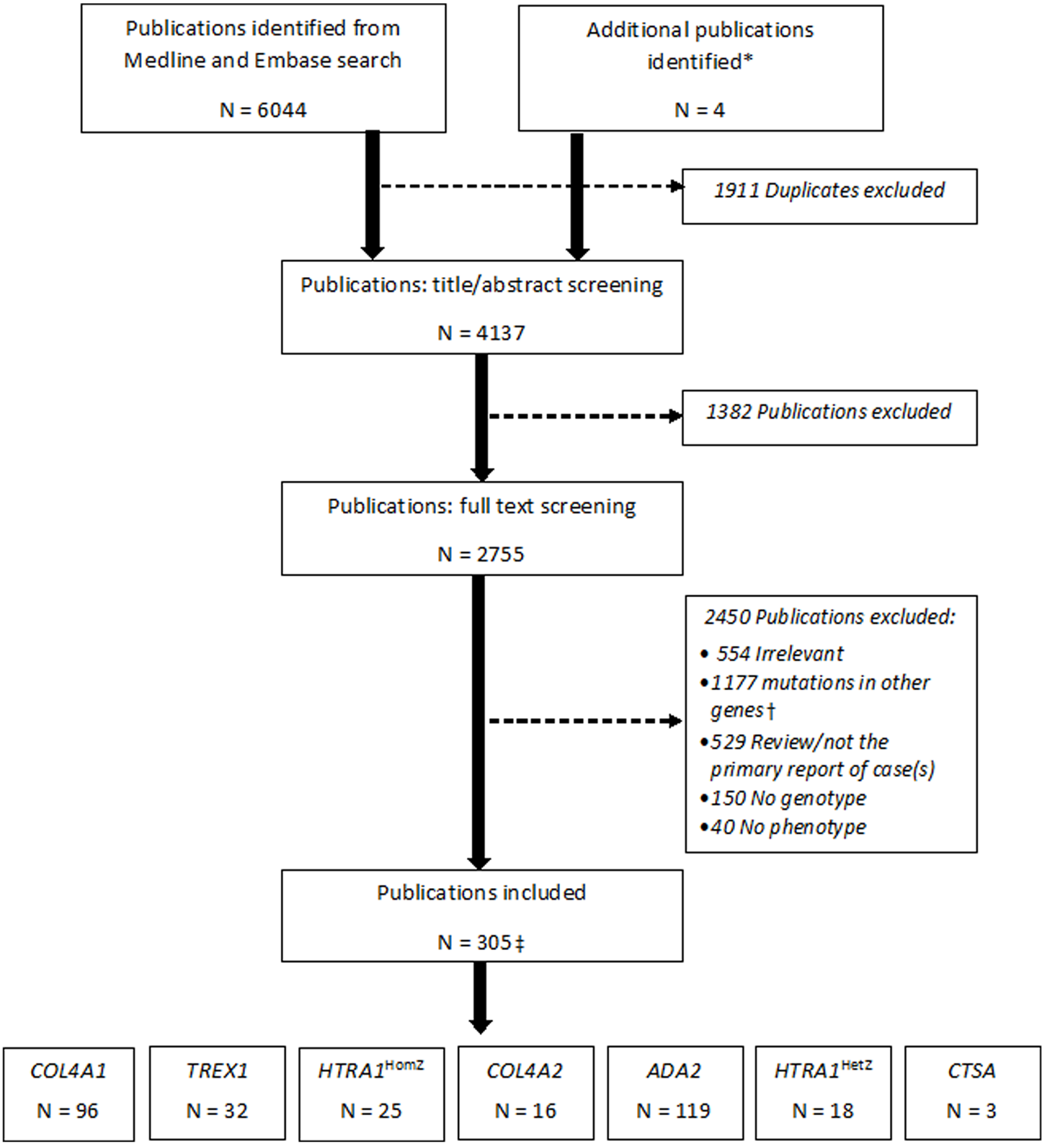
*Region of origin was not reported and therefore assumed from the first author's institution country for 199/350 *COL4A1*, 12/115 *TREX1*, 10/38 *HTRA1*^{HomZ}, 17/37 *COL4A2*, 62/209 *ADA2* and 21/61 *HTRA1*^{HetZ} individuals. It was not reported or possible to derive this for 41 *COL4A1*, 3 *COL4A2* and 9 *ADA2* individuals. If a case was reported to have a different region of origin/ancestry to that of the country where they lived, they were considered to be from their region of origin (e.g. a Chinese origin person living in USA was considered Asian). †If a mean age was available for a group of individuals, the overall summary estimate was weighted by the group size. ‡8 individuals had only a mean age reported, so they were included in the calculations for mean age, but not for median and age range. Turkey as a region of origin was reported on specifically because of the high proportion of ADA2 individuals reported from there.

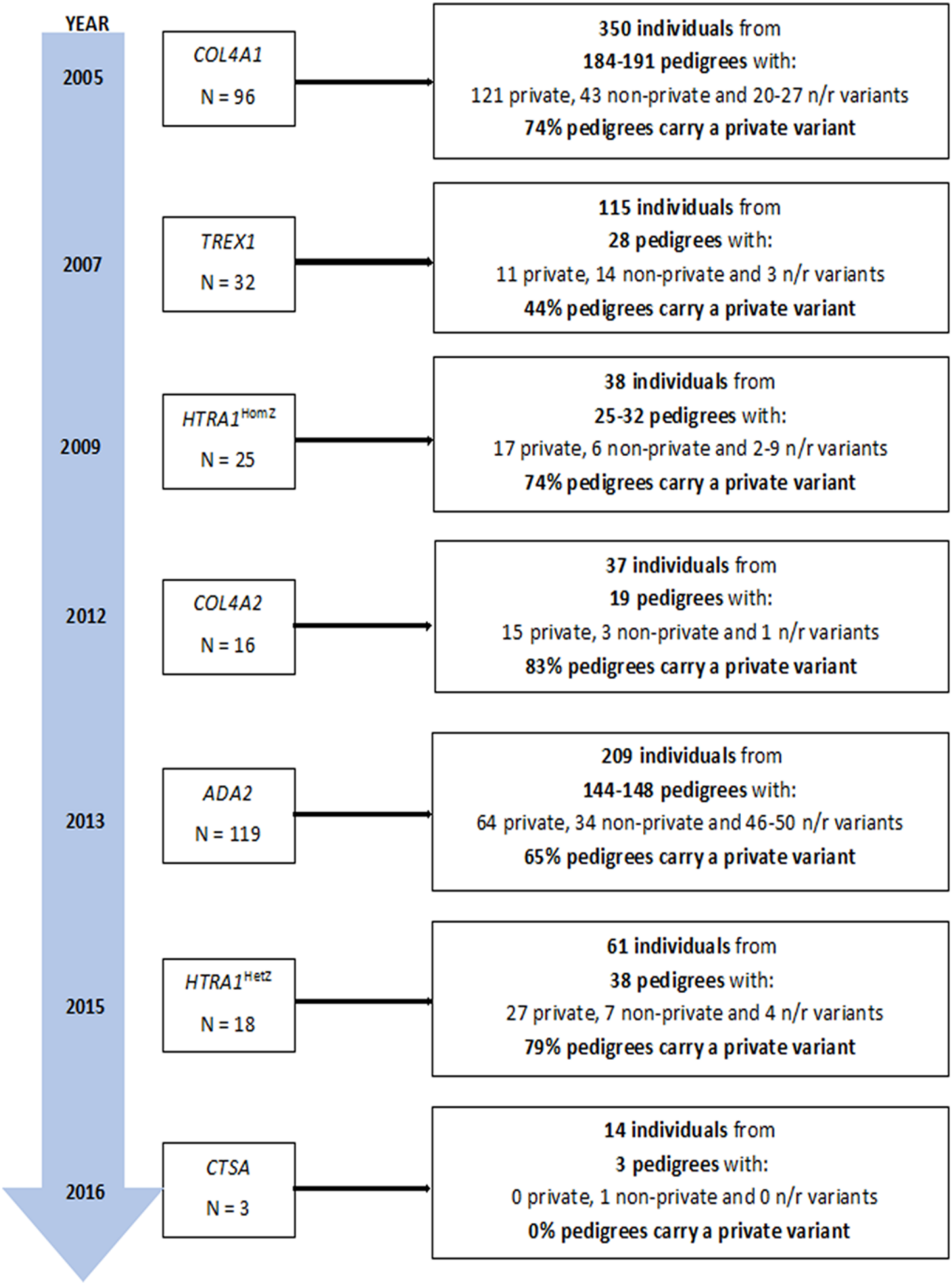
Table 2. Shared extra-cerebral phenotypes across monogenic cSVDs

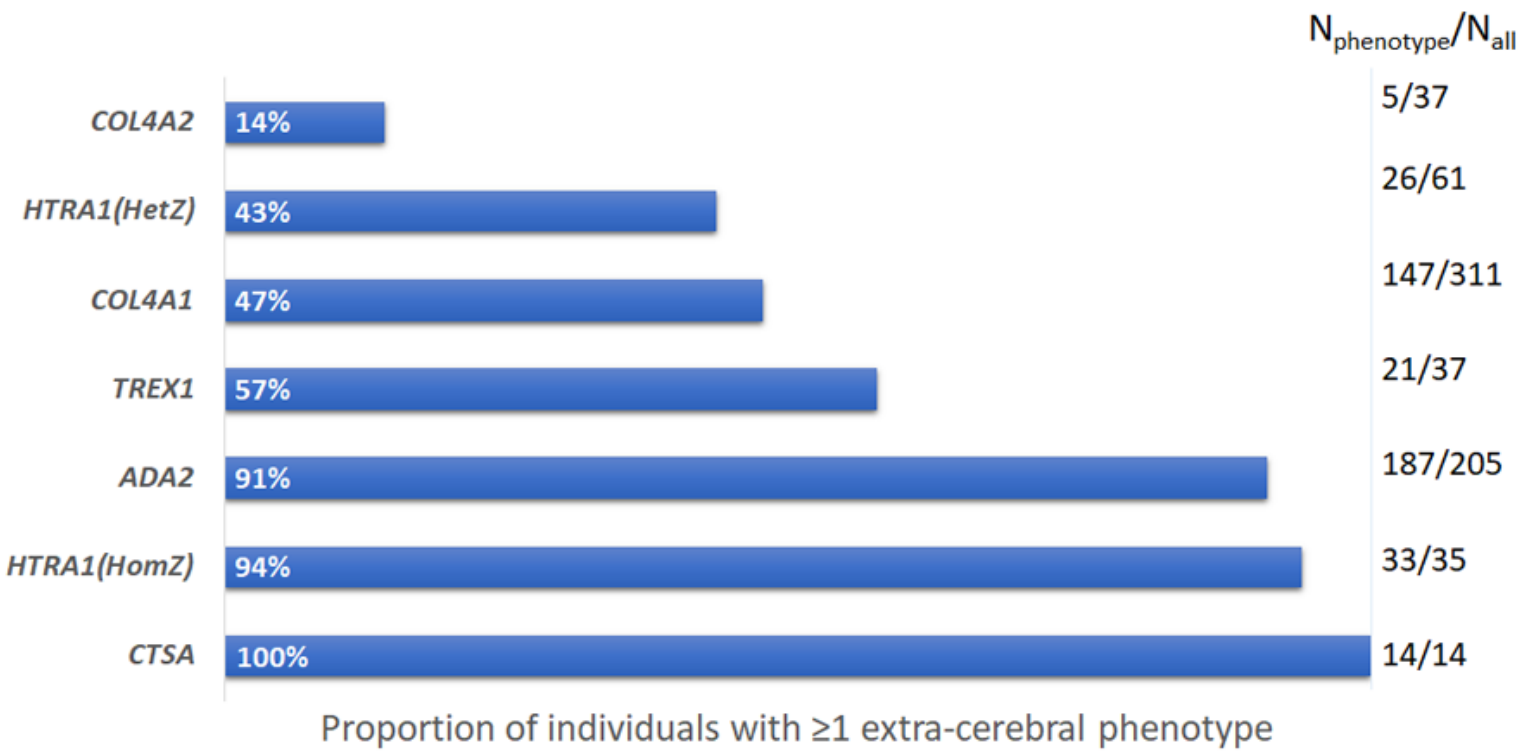
	Presumed vascular phenotypes								Other phenotypes							
	Ocular	Renal	Hepatic	Venous pathology	Skin	Raynaud's	Hypertension	Vasculitis / inflammation	Ocular	Renal	Muscle	Hematological	Hair loss	Joints	Cardiac	Dry mouth / eyes	Spleen
COL4A1	√	√				√	?		√	√	√	√			√		
TREX1	√	√	√			√	?					?					
HTRA1^{HomZ}							?						√	√			
COL4A2	√	√				√	?		√	√	√	√					
ADA2		√	√		√	√	?	√			√	√		?			√
HTRA1^{HetZ}							?						√	√			
CTSA		?		?			√				√					√	
Phenotype associated with >1 gene	√	√	√			√	?		?		√	√		?			

√=extra-cerebral phenotype is already recognized to be associated with the gene in OMIM database (associations for *HTRA1*^{HetZ}/HTRA1^{HomZ} and *COL4A1/COL4A2* were treated as applying to both genes respectively); ?=extra-cerebral phenotype is a novel potentially associated phenotype

not already recognized to be associated with the gene in OMIM database; HetZ=heterozygous; HomZ=homozygous/compound heterozygous; Phenotypes are defined in https://www.ahajournals.org/journal/str:Supplemental Methods;


Publications identified from Medline and Embase search
N = 6044
Additional publications identified*
N = 4
1911 Duplicates excluded
Publications: title/abstract screening
N = 4137
1382 Publications excluded
Publications: full text screening
N = 2755
2450 Publications excluded:
• 554 Irrelevant
• 1177 mutations in other genes†
• 529 Review/not the primary report of case(s)
• 150 No genotype
• 40 No phenotype
Publications included
N = 305‡
COL4A1
N = 96
TREX1
N = 32
HTRA1HomZ
N = 25
COL4A2
N = 16
ADA2
N = 119
HTRA1HetZ
N = 18
CTSA
N = 3

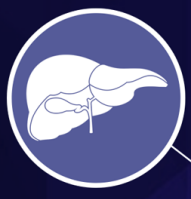

YEAR
2005
COL4A1
N = 96
350 individuals from
184-191 pedigrees with:
121 private, 43 non-private and 20-27 n/r variants
74% pedigrees carry a private variant
2007
TREX1
N = 32
115 individuals from
28 pedigrees with:
11 private, 14 non-private and 3 n/r variants
44% pedigrees carry a private variant
2009
HTRA1HomZ
N = 25
38 individuals from
25-32 pedigrees with:
17 private, 6 non-private and 2-9 n/r variants
74% pedigrees carry a private variant
2012
COL4A2
N = 16
37 individuals from
19 pedigrees with:
15 private, 3 non-private and 1 n/r variants
83% pedigrees carry a private variant
2013
ADA2
N = 119
209 individuals from
144-148 pedigrees with:
64 private, 34 non-private and 46-50 n/r variants
65% pedigrees carry a private variant
2015
HTRA1HetZ
N = 18
61 individuals from
38 pedigrees with:
27 private, 7 non-private and 4 n/r variants
79% pedigrees carry a private variant
2016
CTSA
N = 3
14 individuals from
3 pedigrees with:
0 private, 1 non-private and 0 n/r variants
0% pedigrees carry a private variant


$N_{phenotype}/N_{all}$
COL4A2
14%
5/37
HTRA1(HetZ)
43%
26/61
COL4A1
47%
147/311
TREX1
57%
21/37
ADA2
91%
187/205
HTRA1(HomZ)
94%
33/35
CTSA
100%
14/14
Proportion of individuals with ≥1 extra-cerebral phenotype

Frequency Of Stroke/TIA and Extra-Cerebral Phenotypes

Dry eyes/mouth: *CTSA* 64%
Retinal vasculopathy: *TREX1* 68%
Cataract: *COL4A1* 19%, *COL4A2* 8%
Retinal artery tortuosity: *COL4A1* 11%
Anterior segment defects: *COL4A1* 9%

Stroke/ TIA:
COL4A1 45%, *COL4A2* 57%,
ADA2 40%, *CTSA* 50%, *TREX1* 4%
HTRA1^{HetZ} 56%, *HTRA1*^{HomZ} 26%



Liver disease:
TREX1 34%, *ADA2* 21%

Hair loss:
HTRA1^{HetZ} 20%, *HTRA1*^{HomZ} 74%

Kidney cyst:
COL4A1 8%, *COL4A2* 3%
Hematuria:
COL4A1 12%, *COL4A2* 3%
Nephropathy:
TREX1 37%, *ADA2* 12%

Arrhythmia: *COL4A1* 3%
Hypertension: *CTSA* 93%

Hematological features:
ADA2 56%
Inflammation: *ADA2* 66%
Hemolytic anemia:
COL4A1 3%

Degenerative spine disease:
HTRA1^{HetZ} 31%, *HTRA1*^{HomZ} 63%
Lower back pain:
HTRA1^{HetZ} 7%, *HTRA1*^{HomZ} 37%

Raynaud's:
COL4A1 3%, *TREX1* 28%
Skin lesions: *ADA2* 56%

Muscle:
COL4A1 17%, *COL4A2* 5%,
ADA2 11%, *CTSA* 64%

Spleen: *ADA2* 26%