




# Whole-exome sequencing analyses in a Saudi Ischemic Stroke Cohort reveal association signals, and shows polygenic risk scores are related to Modified Rankin Scale Risk

Fahad A. Alkhamis<sup>1</sup> · Majed M. Alabdali<sup>1</sup> · Abdulla A. Alsulaiman<sup>1</sup> · Abdullah S. Alamri<sup>1</sup> · Rudaynah Alali<sup>2</sup> · Mohammed S. Akhtar<sup>3</sup> · Sadiq A. Alsalman<sup>4</sup> · Cyril Cyrus<sup>3</sup> · Aishah I. Albakr<sup>1</sup> · Anas S. Alduhalan<sup>1</sup> · Divya Gandla<sup>5</sup> · Khaldoun Al-Romaih<sup>6</sup> · Mohamed Abouelhoda<sup>6</sup> · Bao-Li Loza<sup>5</sup> · Brendan Keating<sup>5</sup> · Ameen K. Al-Ali<sup>3</sup> 

Received: 9 February 2023 / Revised: 22 March 2023 / Accepted: 23 March 2023 / Published online: 27 March 2023  
© The Author(s) 2023

## Abstract

Ischemic stroke represents a significant societal burden across the globe. Rare high penetrant monogenic variants and less pathogenic common single nucleotide polymorphisms (SNPs) have been described as being associated with risk of diseases. Genetic studies in Saudi Arabian patients offer a greater opportunity to detect rare high penetrant mutations enriched in these consanguineous populations. We performed whole exome sequencing on 387 ischemic stroke subjects from Saudi Arabian hospital networks with up to 20,230 controls from the Saudi Human Genome Project and performed gene burden analyses of variants in 177 a priori loci derived from knowledge-driven curation of monogenic and genome-wide association studies of stroke. Using gene-burden analyses, we observed significant associations in numerous loci under autosomal dominant and/or recessive modelling. Stroke subjects with modified Rankin Scale (mRSs) above 3 were found to carry greater cumulative polygenic risk score (PRS) from rare variants in stroke genes (standardized PRS mean > 0) compared to the population average (standardized PRS mean = 0). However, patients with mRS of 3 or lower had lower cumulative genetic risk from rare variants in stroke genes (OR (95%CI) = 1.79 (1.29–2.49),  $p = 0.0005$ ), with the means of standardized PRS at or lower than 0. In conclusion, gene burden testing in Saudi stroke populations reveals a number of statistically significant signals under different disease inheritance models. However, interestingly, stroke subjects with mRS of 3 or lower had lower cumulative genetic risk from rare variants in stroke genes and therefore, determining the potential mRS cutoffs to use for clinical significance may allow risk stratification of this population.

**Keywords** Exome sequencing · Stroke · Saudi · Variants · Dominant · Recessive

✉ Ameen K. Al-Ali  
aalali@iau.edu.sa

- <sup>1</sup> Department of Neurology, King Fahd Hospital of The University, Imam Abdulrahman Bin Faisal University, Dammam 31441, Saudi Arabia
- <sup>2</sup> Department of Internal Medicine, King Fahd Hospital of The University, Imam Abdulrahman Bin Faisal University, Dammam 31441, Saudi Arabia
- <sup>3</sup> Department of Clinical Biochemistry, College of Medicine, Imam Abdulrahman Bin Faisal University, Dammam 31441, Saudi Arabia
- <sup>4</sup> Department of Neurology, King Fahd Hospital, Alhafaf, Saudi Arabia
- <sup>5</sup> Department of Surgery, Perelman School of Medicine, University of Pennsylvania, Pennsylvania, PA 19104, USA
- <sup>6</sup> Centre for Genomic Medicine, KFSH&RC, Riyadh 11211, Saudi Arabia

## Introduction

Stroke is a major cause of morbidity and is the second leading cause of death worldwide. Over 13 million people have a stroke each year and around 5.5 million people will die as a result (Virani et al. 2021). Incidence, type of stroke and mortality rates vary markedly between countries and ancestral groups. Ischemic strokes are the most common type of stroke, and typically involve a disruption of blood flow to the brain parenchyma which causes brain cell death from a lack of oxygen. It may result from a number of processes including small vessel occlusions (SVOs) and large-artery atherosclerosis (Adams et al. 1993). More young people are affected with ischemic stroke in low- and middle-income countries than higher income countries (Lehman and Fullerton 2013; Lehman et al. 2018). Initial twin and family

studies, primarily using linkage analysis, contributed to the initial knowledge of heritability studies in stroke (Flossmann et al. 2004). While such approaches found causal variants in various genes for monogenic stroke disorders, they had limited value in finding common variants impacting polygenic risk (Ekkert et al. 2021; Li et al. 2021). Genome-Wide Association Studies (GWAS) utilizing genome-wide genotyping arrays and/or whole exome sequencing (WES) have been successful in elucidating rare and common variants in various stroke subtypes (Li et al. 2021; Dichgans et al. 2021; Kumar et al. 2021).

Polygenic risk scores (PRS) utilizing cumulative, weighted risk scores for multiple genetic variants, with specific diseases/phenotypes, typically integrate known clinical risk covariates. They have been used with varying success in common diseases with multifactorial disease risk, including those with common and rare genetic underpinnings. Malik and colleagues combined stroke PRS data with Framingham risk scores and observed a significant association with ischemic stroke risk, although the prognostic value of the PRS was not substantially different from that of conventional clinical risk factors (Malik et al. 2014). However, more recent studies have shown utility with the use of PRSs in stroke-related phenotypes (Hachiya et al. 2020).

Saudi Arabia has seen an unprecedented adverse rise in modifiable risk factors for vascular disease over the last 30–40 years including poor diet, smoking and sedentary lifestyle, resulting in increases in dyslipidemia, type 2 diabetes, and hypertension which further exacerbate stroke and other vascular disease progression (Alhazzani et al. 2021). There are wide incidence differences across reported Saudi stroke studies ranging from ~ 16 to 58 cases per 100,000 person years (Alhazzani et al. 2018; Al Rajeh & Awada 2002). Such differences may be due to an interplay of study ascertainment biases as well as from significant primary and secondary health care delivery differences between private and public payer systems in Saudi Arabia, leading to a higher likelihood of undiagnosed diseases (Alqahtani et al. 2020). Recent studies indicate that the incidence of stroke is increasing rapidly with ischemic stroke being the dominant subtype affecting the Saudi populations (Alhazzani et al. 2018; Alqahtani et al. 2020; Al Rajeh & Awada 2002; Alqahtani et al. 2020).

Performing genetic studies in Saudi Arabian populations offers a unique opportunity for the discovery of novel genetic variants impacting disease risk due to a high rate of consanguinity amongst tribal pedigrees that make up the majority of the national population (Kari et al. 2014; Alkuraya 2012). We performed whole exome sequencing on 387 Saudi subjects with clinically diagnosed ischemic stroke. We focused on the analyses of exonic sequence data from a panel of 177 gene regions derived from highly curated stroke studies

and utilized approximately 20,230 controls from the Saudi Human Genome Project. We then assessed the association of rare variants, primarily with ischemic stroke, and also evaluated PRSs within our study population.

## Methods

### Study participants' samples and data

During 2019–2020, samples and data from 387 subjects (inpatients and outpatients) who had been diagnosed with ischemic stroke and attending the following Neurology Clinics were collected for inclusion in this study: King Fahd Hospital of the University (KFHU), Al Khobar; King Fahd Hospital, Al Hafof; and Al Wajh Hospital, Dammam. Participants ranged in age from 19 to 81. The phenotype data of all subjects were reviewed by a neurology consultant to ascertain and verify the diagnoses and the phenotype uniformity among sites as well as eligibility according to the study criteria. Table 1 outlines the demographic and clinical characteristics of the 387 ischemic stroke subjects included in this study. The subtypes of ischemic stroke were determined according to the Trial of Org 10,172 in Acute Stroke

**Table 1** Saudi Ischemic Stroke Study Participants Clinical Characteristics. Clinical characteristics data collated from 387 ischemic stroke subjects diagnosed at neurology clinics at the following three Saudi Hospitals: King Fahd Hospital University (KFHU), Al-Khobar, and King Fahd Hospital, Alhafof and Al Wajh Hospital, Dammam. \*Trial of Org 10,172 in Acute Stroke Treatment (TOAST) classification

Parameter	
Total number of patients	387
Age (Mean ± SD)	56.5 ± 15.8
Age at diagnosis (Mean ± SD)	55.6 ± 15.7
Male sex <i>N</i> (%)	223 (57.6)
Hospital	
AL-WAJAH, Dammam, <i>n</i> (%)	94 (24.3)
King Fahd Hospital, Al Hafof, <i>n</i> (%)	107 (27.6)
KFHU, Alkhobar, <i>n</i> (%)	186 (48.1)
TOAST* classification	
Large artery atherosclerosis, <i>n</i> (%)	78 (20.2)
Cardio aortic embolism, <i>n</i> (%)	56 (14.5)
Small artery occlusion, <i>n</i> (%)	253 (65.4)
Modified Rankin Scale (mRS)	
0 <i>N</i> (%)	51 (13.2)
1 <i>N</i> (%)	44 (11.4)
2 <i>N</i> (%)	219 (56.6)
3 <i>N</i> (%)	38 (9.8)
4 <i>N</i> (%)	17 (4.4)
5 <i>N</i> (%)	12 (3.1)
6 <i>N</i> (%)	6 (1.6)

Treatment (TOAST) classification (Adams et al. 1993). The functional outcome of stroke patients was determined using modified Rankin Scale (mRS) at admission and at one-month post stroke follow-up.

### DNA extraction and sequencing

Peripheral blood samples were collected into EDTA tubes and stored in  $-20^{\circ}\text{C}$  freezers at the research laboratories at the College of Medicine, Imam Abdulrahman bin Faisal University. Genomic DNA extraction from all samples was performed using Genra Puregene Blood kits (Qiagen, USA) according to the manufacturer's protocol. Whole exome sequencing libraries were generated using the SureSelect Human All Exon Kit v5 (Agilent, CA, USA) and sequenced on a HiSeq instrument (Illumina, CA, USA) using a standard paired-end sequencing protocol for SureSelectXT Library Prep and Target Enrichment System Version B.2 (Illumina, CA, USA).

### Read alignment, variant calling, and QC

Reads in the FASTQ files were aligned to the standard human genome reference (GRCh37) using Illumina's Dynamic Read Analysis for GENomics (DRAGEN) Genomic Pipeline. Resultant BAM files were position-sorted and duplicate reads marked. Single-sample genomic variant call files (gVCF) were generated by the DRAGEN Germline Pipeline, and joint calling of all samples in the study cohort was performed by DRAGEN Joint Genotyping (Illumina, CA, USA).

### Principal components analysis (PCA) and Kinship

The *KING* algorithm was used for relatedness inference based on the genotype of exome SNPs (minor allele frequency [MAF]  $> 0.01$ ). Estimated kinship coefficient and number of SNPs with zero shared alleles (identity by state [IBS]<sub>0</sub>) between a pair of individuals were plotted. Parent-offspring, sibling pairs, and unrelated pairs were visualized on the scatterplot to distinguish any separate clusters. Ancestry and Kinship Toolkit (AKT) was used to calculate PCAs and plot the results using 1000 genome project data. The study participant samples demonstrated a genetically matched background consistent with typical Saudi populations including African admixture, which is known to be evident in Saudi tribes primarily from East Africa (Fernandes et al. 2019).

### Variant annotation, filtering and prioritization

Variants were annotated with a program for annotating and predicting the effects of single nucleotide polymorphisms

(SnPEff,v5.0) to predict the effects of the variants. Rare variants were defined as minor allele frequency (MAF)  $< 1\%$  in The Genome Aggregation Database (gnomAD) (v2.1.1). Intronic, synonymous, 3' and 5' UTR, upstream and downstream variants were identified and excluded from the analysis. The remaining rare variants were considered to be potentially deleterious variants. Genetic variants classified in ClinVar as "Likely pathogenic" or "Pathogenic," and in Human Gene Mutation Database (HGMD) as disease-causing mutations (DM) for stroke were collected and curated together with research literature to serve as the knowledge-base for variant prioritization and classification (Stenson et al. 2003).

### Use of a comprehensive stroke gene panel

Numerous reports have identified genes associated with stroke by using data from monogenic and genome-wide association studies (GWAS). We used a comprehensive collection of genes with associations for monogenic causes of stroke which has been used in many clinical and research studies (Ilinca et al. 2019). Malek et al. in a large multi-ancestry GWAS of up to 67,162 stroke cases and 454,450 controls discovered 22 new stroke risk loci and validated 10 known stroke loci, bringing the total to 32 loci which are encompassed in the panel (Malik et al. 2018).

### Gene burden testing

The open-source software package Test Rare vAriants with Public Data (TRAPD) was used to perform a gene-based burden testing against public control databases. The software allows for adaptable filtering on various quality and frequency fields to ensure a well-controlled burden test. Gene burden test on our ischemic stroke WES cohort datasets was performed against whole exome or whole genome sequencing data available from approximately 20,230 individuals from the Saudi Human Genome Project (<https://shgp.kacst.edu.sa>).

### Polygenetic risk score generation

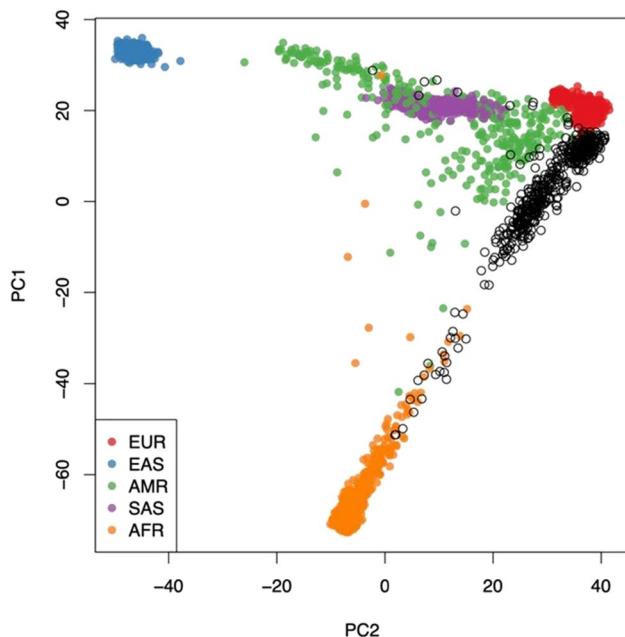
Rare, impactful variants (listed in Supplementary Table S1) identified in stroke genes were included to calculate polygenetic risk scores (PRS) for stroke for each individual study subject to represent the cumulative risk of carrying one or more of these rare variants. (Malik et al. 2018). PRSice-2 software was employed to calculate PRS by setting an equal effect size ( $\beta = 1$ ) for each variant (Choi and O'Reilly 2019). The resulting PRS was standardized for association tests with clinical variables such as age of diagnosis, mRS, and classifications of stroke. It is acknowledged that there may be an under-

over-estimation of the effect size of these rare variants. However, when there is an actual effect, the association, albeit with less accurate estimates, could be detected. The validity of this approach was evident in the initial effort to estimate the combined effect of multiple genetic variants before a more sophisticated statistical approach was developed (Zheng et al. 2008). In addition, as the rare variant PRS was calculated using PRSice-2, it was restricted to the specific set of rare variants that were identified without consideration of other common variants based on LD patterns.

## Results

### Principal component analysis

The common variants of the whole exome sequencing data from the stroke study participants show expected clustering when compared to the world's major populations using a standard principal component analysis pipeline (Fig. 1). This population stratification data was used to mitigate false attributions in the association analyses components.



**Fig. 1** Principal component analysis (PCA) of common SNP genotypes with 1000 genome projects (1KGP) populations reference panel. The x- and y-axes denote the value of two components of PCA (PC1, PC2), with each dot in the figure representing one individual. The color for individuals from 1000 genome projects, Europeans (EUR), East Asians (EAS), Admixed Americans (AMR), South Asians (SAS), and Africans (AFR) are red, blue, green, purple, and orange, respectively. The color for individuals belonging to the Stroke Disease study group is illustrated in black

### Gene burden testing

We were able to utilize exonic data from 177 genes from the initial set of 214 stroke loci panel (48) after removal of mitochondrial genes; non-exonic regions such as 9p.21 and non-coding RNAs, and genes that had poor quality control and quality assurance filtering data in the Saudi Genome Project control dataset including *ATP7A*, *FLNA*, *GLA*, *GPR143*, *PGAM4*, and *F9*. Focusing on the a priori 177 loci derived from the stroke panel, we identified rare putative impactful variants within the cohort (Supplementary Table S1). The equivalent analysis was performed using the TRAPD burden testing pipeline and we observed that our cohort of 387 ischemic stroke subjects was significantly enriched for impactful alleles in several loci on the a priori stroke panel, compared to the control populations of approximately 20,230 Saudi population-based controls derived from the Saudi Genome Project.

The top 20 most significant signals for gene burden analyses under a dominant model ( $p < 1 \times 10^{-5}$ ) and 14 signals with  $p < 0.05$  under a recessive model are shown on Table 2 with FDR correction and a full list of association signals under both dominant and recessive models are shown for all 177 genes (Supplementary Table S2). Table 2 (top) shows the top 20 genes with  $p < 1 \times 10^{-5}$  under a dominant model (*F13A1*, *NF1*, *ACAD9*, *NOTCH3*, *MYLK*, *SH3PXD2A*, *TSC2*, *ADAMTS13*, *COL4A2*, *APP*, *FOXC1*, *COL1A1*, *TGFBR2*, *PDE4D*, *MYH7*, *DPM1*, *PGM1*, *FCGR2C*, *ZFH3*, *PDE3A*). Table 2 (top) shows the 14 most significant signals at  $p < 0.05$  under a recessive model (*PDE4D*, *KCNQ1*, *TREX1*, *CYP11B1*, *F5*, *HTRA1*, *CACNA1A*, *ZCCHC14*, *NBEAL2*, *FGA*, *PCNT*, *DPM1*, *LOC100505841*, and *SH3PXD2A*). Three genes, *SH3PXD2A*, *PDE4D*, and *DPM1*, were significant under both the dominant and recessive models. Twenty-eight of the genic 32 GWAS loci contained among the 177 genes on the stroke panel were available for analysis and 12 showed gene burden associations at  $p < 0.05$  under either dominant or recessive models (*ZCCHC14*, *FGA*, *LOC100505841*, *SH3PXD2A*, *PRPF8*, *PDE3A*, *ZFH3*, *SH2B3*, *TM4SF4*, *HDAC9-TWIST1*, *SMARCA4*, and *FOXF2*).

### Associations of polygenic risk with clinical variables

The overall distribution of polygenic risk score (PRS) representing the overall genetic burden from rare, impactful variants in stroke genes among all 387 ischemic stroke study participants subjected to whole exome sequencing deviated from normality and was right-skewed (Fig. 2) with median [IQR] of 8.61 [6.59,10.53], minimum of 1.53, and maximum of 18.75. PRS was not associated with age of diagnosis, ischemia diagnosis, and classification of stroke such as large artery atherosclerosis, cardio aortic embolism,

**Table 2** (Top) 20 most significant dominant model signals and (bottom)  $p < 0.05$  Recessive model signals from TRAPD gene burden testing in Saudi Ischemic Stroke subjects compared in up to 20,230 individuals from Saudi Human Genome Project

Gene	Heterozygote cases	Homozygote cases	Heterozygote controls	Homozygote controls	P: dominant model	P: recessive model
<i>F13A1</i>	10	0	6	0	3.13E-11	1
<i>NF1</i>	11	0	12	2	4.84E-10	1
<i>ACAD9</i>	9	0	11	0	1.52E-08	1
<i>NOTCH3</i>	42	2	411	18	1.77E-08	0.405288
<i>MYLK</i>	31	0	231	13	3.60E-08	0.170765
<i>SH3PXD2A</i>	21	1	126	5	5.16E-08	0.042198
<i>TSC2</i>	48	0	499	15	5.68E-08	0.791086
<i>ADAMTS13</i>	26	0	182	10	1.54E-07	0.106345
<i>COL4A2</i>	27	1	217	9	2.70E-07	0.435679
<i>APP</i>	12	0	44	2	1.19E-06	1
<i>FOXC1</i>	6	1	4	6	1.36E-06	0.234242
<i>COL1A1</i>	21	0	146	3	1.39E-06	1
<i>TGFBR2</i>	5	0	2	0	1.41E-06	1
<i>PDE4D</i>	23	6	253	11	1.58E-06	2.20E-06
<i>MYH7</i>	7	0	11	0	2.16E-06	1
<i>DPM1</i>	4	1	3	0	3.64E-06	0.037399
<i>PGM1</i>	9	0	25	1	3.88E-06	0.073402
<i>FCGR2C</i>	8	1	26	2	6.40E-06	0.108062
<i>ZFH3</i>	105	4	1791	80	7.23E-06	0.308726
<i>PDE3A</i>	14	0	76	3	8.31E-06	0.17355
Gene	Heterozygote Cases	Homozygote Cases	Heterozygote Controls	Homozygote Controls	P: Dominant model	P: Recessive model
<i>PDE4D</i>	23	6	253	11	1.58E-06	2.20E-06
<i>KCNQ1</i>	16	1	486	15	0.746853	5.99E-05
<i>TREX1</i>	7	0	126	11	0.293476	0.000408
<i>CYP11B1</i>	19	0	209	5	0.001179	0.001053
<i>F5</i>	25	1	314	8	0.000606	0.001374
<i>HTRA1</i>	2	2	42	0	0.092068	0.001395
<i>CACNA1A</i>	20	2	217	14	0.000202	0.009701
<i>ZCCHC14</i>	11	2	204	7	0.077487	0.013894
<i>NBEAL2</i>	72	2	1840	118	0.623318	0.015113
<i>FGA</i>	15	1	253	6	0.053339	0.016894
<i>PCNT</i>	79	6	1374	88	0.000112	0.029154
<i>DPM1</i>	4	1	3	0	3.64E-06	0.037399
<i>LOC100505841</i>	3	1	39	0	0.075819	0.037399
<i>SH3PXD2A</i>	21	1	126	5	5.16E-08	0.042198

and small artery occlusion (Table 3). However, a significant association of PRS and mRS, which measures the degree of disability post-stroke, was identified (regression coefficient (95%CI)=0.15 (0.03–0.27),  $p=0.02$ ). PRS was found to increase the risk of developing greater disability post-stroke event (mRS > 3), when compared to patients with lower mRS (mRS ≤ 3) (OR (95%CI) = 1.79 (1.29–2.49),  $p=0.0005$ ) (Table 3).

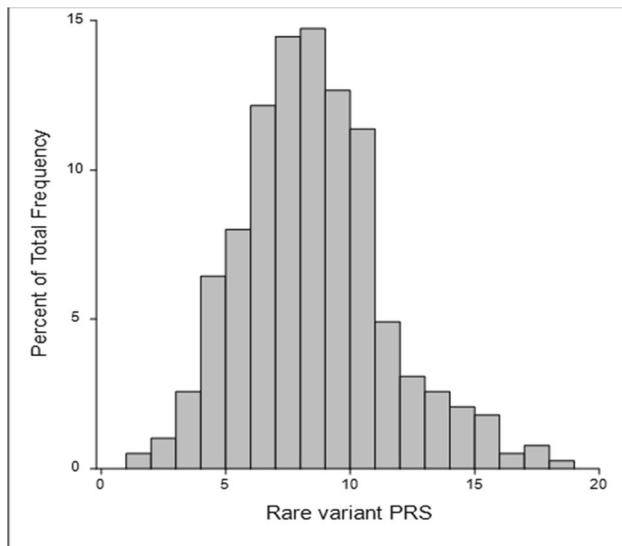
Stroke subjects with mRSs above 3 were found to carry greater cumulative genetic risk from rare variants in stroke genes (standardized PRS mean > 0) compared to

the population average (standardized PRS mean = 0). However, patients with mRS of 3 or lower had lower cumulative genetic risk from rare variants in stroke genes, with the means of standardized PRS at or lower than 0 (Fig. 3).

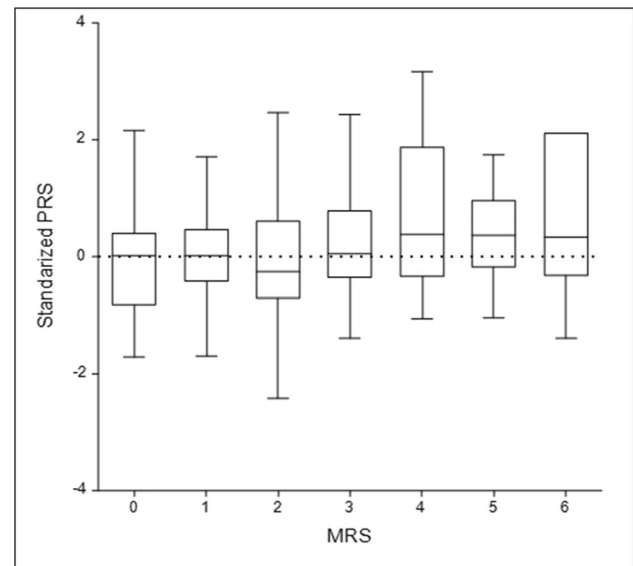
### Discussion

In this study, WES was performed on 387 individuals diagnosed with ischemic stroke. We initially focused on known, or putative, pathogenic exonic variants evident in 177 genes





**Fig. 2** Distribution of rare variant Polygenic Risk Score (PRS) among 387 Saudi Ischemic Stroke subjects. PRS statistics: Mean  $\pm$  SD =  $8.7 \pm 3.0$ ; Median[IQR] = 8.6 [6.6,10.5]; Min = 1.53; Max = 18.6



**Fig. 3** Box-whisker plot of standardized Stroke Polygenic risk score in a Saudi Stroke Cohort by Modified Rankin Scale (mRS) groups. The standardized Polygenic Risk Score (PRS) is illustrated on the y-axis with the Modified Rankin Scale (mRS) groups shown on the x-axis

**Table 3** Association of clinical phenotypes with polygenic risk score (PRS) constructed with rare impactful variants in stroke genes. The PRS was standardized to mean at 0 and SD of 1

Continuous phenotype	Regression coefficient (95%CI)	<i>p</i> -value
Age of Diagnosis	1.15 (−0.43–2.72)	0.15
Modified Rankin Scale (mRS)	0.15 (0.03–0.27)	0.02
Dichotomized phenotype	OR (95%CI)	<i>p</i> -value
High mRS (4/5/6, <i>N</i> = 35) vs low mRS (0/1, <i>N</i> = 95)	1.56 (1.09–2.23)	0.015
High mRS (4/5/6, <i>N</i> = 35) vs lower mRS (0/1/2/3, <i>N</i> = 351)	1.79 (1.29–2.49)	0.0005
Diagnosis (= ischemia)	1.14 (0.89–1.47)	0.31
Large artery atherosclerosis	1.00 (0.75–1.34)	0.98
Cardio aortic embolism	0.87 (0.71–1.07)	0.2
Small artery occlusion	1.15 (−0.43–2.72)	0.15

prioritized from a comprehensive panel of loci associated with monogenic causes of stroke as well as recent GWAS statistically significant signals (Ilinca et al. 2019). This panel has been used in exome sequencing interpretation of individuals from multi-incident stroke families to generate and assess putative pathogenic variants amongst the probands and wider family members (Ilinca et al. 2020). It has also been utilized in cerebral small vessel disease (CSVD) which revealed putative pathogenic variants in multiple loci (Monkare et al. 2022). Using the TRAPD gene burden testing pipeline, filtered on our list of 177 genes, we found significant enrichment for pathogenic or likely pathogenic

stroke variants in numerous genes in our cohort using datasets from the Saudi Genome Project as controls.

A number of the top signals in our study are outlined below. The coagulation factor XIII A chain (*F13A1*) gene encodes the coagulation factor that is the last component activated in the blood coagulation cascade. Factor XIII deficiency is typically classified into two categories: type I deficiency, characterized by the lack of both the A and B subunits; and type II deficiency, characterized by the lack of the A subunit alone. These defects can result in defective wound healing and a tendency for lifelong bleeding. A review of common variants association analyses of *F13A1* with stroke phenotypes has shown mixed findings. However, a large Women's Health Initiative (WHI) study of 2,045 of post-menopausal women, significant risk was observed for both ischemic stroke and for combined ischemic and hemorrhagic between two common *F13A1* variants and with hormone replacement therapy (Huang et al. 2012). A Dutch study also showed significant associations between common variants in *F13A1* with ischemic stroke in young women with the effect more pronounced with oral contraception use (Pruissen et al. 2008).

Neurofibromin 1 (*NFI*) gene encodes a negative regulator of the Ras signal transduction pathway and a number of *NFI* mutations have been linked to neurofibromatosis type 1 which is commonly associated with malignant tumors and cardiovascular or cerebrovascular complications (Napolitano et al. 2022). Neurofibromatosis type 1 also increases the risk of vasculopathies and arterial wall weakness and can

lead to complications such as hemorrhagic stroke, ischemic stroke, and multi-domain cognitive impairment (Napolitano et al. 2022).

Acyl-CoA dehydrogenase family member 9 (ACAD9) encodes a member of the acyl-CoA dehydrogenase family, a family of proteins that are localized in the mitochondria and are involved in beta-oxidation of fatty acyl-CoA. Mutations in this gene cause acyl-CoA dehydrogenase family member type 9 deficiency which often results in intellectual disability and neurologic dysfunction. A homozygous variant in ACAD9 was identified in the proband of a Swedish family where family members reported stroke with intracerebral bleeding and progressive muscle and heart failure (Ilinca et al. 2020). A case report of a death of a teenager with ACAD9 deficiency with Reye-like episode and cerebellar stroke has also been reported (Huang et al. 2012).

Phosphodiesterase 4D (*PDE4D*) encodes up to 9 different isoforms whose functional proteins degrade the second messenger Cyclic adenosine monophosphate (cAMP), a key signal transduction molecule in multiple cell types, including vascular cells (Munshi & Kaul 2008). Numerous case-control studies have been performed to assess the association between *PDE4D* variants and ischemic stroke risk among different ancestral populations. In particular, the so-called “SNP83” (rs966221) association with has been robust in many of these studies (Munshi & Kaul 2008; Xu et al. 2010). *PDE4D* is associated with inflammation and reduced *PDE4D* is thought to increase the risk of atrial fibrillation, which in turn increases stroke risk (Jørgensen et al. 2015). Several studies have shown a robust association of *PDE4D* variants with ischemic stroke in young individuals (Yue et al. 2019).

Potassium voltage-gated channel subfamily Q member 1 (*KCNQ1*) encodes a voltage-gated potassium channel protein which is required for the repolarization phase of the cardiac action potential and forms multimers with KCNE1, KCNE3 and potassium channel proteins. Mutations in *KCNQ1* are associated with familial atrial fibrillation and hereditary long QT syndrome 1 which can impact stroke risk (Jørgensen et al. 2015; Lavy et al. 1974). An increased genetic burden of rare deleterious *KCNQ1* variants in Polish subjects with large-vessel ischemic stroke were identified using second-generation sequencing (Janicki et al. 2019).

Three prime repair exonuclease 1 (*TREX1*) encodes a nuclear protein with 3' exonuclease activity and play a role in DNA repair and serve as a proofreading function for DNA polymerase. Mutations in this gene result in diseases of the immune system including Aicardi-Goutieres syndrome and other autoimmune-type diseases which can put subjects at a higher risk of ischemic stroke. Uemura 2023 investigating the prevalence of Mendelian stroke genes mutations in monogenic cerebral small vessel stroke patients aged 55 years or younger from a Japanese stroke

registry identified a *TREX1* pathogenic genetic variants in one stroke subject (Uemura et al. 2023). A large Mendelian Stroke Consortium also identified pathogenic clinical variants in *TREX1* (Grami et al. 2020).

The distribution of the polygenic risk score analyses showed that the stroke subjects in this Saudi cohort each, on average, carry over 8 rare, impactful variants. PRS analyses performed on a weighted cumulative risk from rare, impactful variants among ischemic stroke participants with high Modified Rankin Scale (mRS), i.e., 4, 5, 6 versus mRS 0, 1, 2, 3 showed a significant association (OR (95%CI) = 1.79 (1.29–2.49),  $p = 0.0005$ ). Determining the potential mRS cutoffs to use for clinical significance within a highly consanguineous population like that in Saudi Arabia may yield translational value, such as risk stratification, especially with the additional of common and rare variants to the PRS from ongoing stroke genome-wide studies.

In Saudi Arabia, undetected or untreated vascular disease is a significant health and financial burden (Walli-Attai et al. 2020a, b; Bindawas & Vennu 2016). There is a compelling need for implementation of primary and secondary stroke prevention strategies in Saudi Arabia due to an increasing incidence rate with the mortality rate projected to almost double by 2030 (Robert et al. 2018; Bindawas & Vennu 2016). In this study using gene-burden analyses in 387 Saudi Arabian ischemic stroke subjects and 20,230 controls from the Saudi Human Genome Project, we observed significant associations in dozens of loci under autosomal dominant and/or recessive modelling. Stroke subjects with Modified Rankin Scale (mRSs) above 3 were observed to have a greater cumulative PRS from rare variants in stroke genes when compared to the population average. Interestingly, stroke subjects with mRS of 3 or lower had lower cumulative genetic risk from rare variants in stroke genes (Alokley and Albakr 2022). Determining the potential mRS cutoffs to use for clinical significance may allow risk stratification of this population.

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1007/s10142-023-01039-7>.

**Acknowledgements** The authors would like to acknowledge the financial support provided by King Abdulaziz City for Science & Technology. We are also grateful to the nurses, technical staff, and students for their efforts, commitment, and dedication. We would like to thank the Saudi Human Genome Project (SHGP), King Abdulaziz City for Science and Technology.

**Author contribution** All authors contributed to the study conception and design. Material preparation, patient recruitment, and data collection were performed by Fahad A. Alkhamis, Majed M. Alabdali, Abdulla A. Alsulaiman, Abdullah S. Alamri, Rudaynah Alali, Sadiq A. Alsalman, Aishah I. Albakr, and Anas S. Alduhalan. Laboratory work, analysis, and interpretation of data were performed by Mohammed S.

Akhtar, Cyril Cyrus, Divya Gandla, Khaldoun Al-Romaih, Mohamed Abouelhoda, Bao-Li Loza, Brendan Keating, and Ameen K. Al-Ali and the manuscript was written by Brendan Keating and Ameen K. Al-Ali. All authors read and approved the final manuscript.

**Funding** This work is supported by King Abdulaziz City for Science and Technology, Riyadh, Saudi Arabia, grant number 14-MED1918-46. This is a government agency which plays no role in the design, collection, analysis and interpretation of the data and writing of the manuscript.

**Data availability** The datasets generated during the current study are available in the European Variation Archive (EVA) repository (<https://www.ebi.ac.uk/ebisearch/search?query=PRJEB59227&submit=Search&db=allebi&requestFrom=global-masthead>), under the title *Whole-Exome Sequencing Analyses in a Saudi Ischemic Stroke Cohort Reveal Association Signals and shows Polygenic Risk Scores are related to Modified Rankin Scale Risk with accession number PRJEB59227*. All requests for data can be sent to the corresponding author (AKA) and verified academic investigators will be granted full access.

## Declarations

**Ethics approval** Ethical approval for the study was obtained from the Imam Abdulrahman bin Faisal University Institutional Review Board (IRB) committees (IRB# 20–1401349) and the study was conducted according to the ethical principles of the Declaration of Helsinki and Good Clinical Practice guidelines. All subjects included in the study signed a written informed consent form in English, with a verified translation provided in Arabic. For those participants who were unable to sign due to the effects of the stroke, the next of kin was requested to sign the informed written consent.

**Competing interests** The authors declare no competing interests.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Adams HP, Bendixen BH, Kappelle LJ, Biller J, Love BB, Gordon DL, Marsh EE (1993) 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* 24(1):35–41. <https://doi.org/10.1161/01.str.24.1.35>
- Al Rajeh S, Awada A (2002) Stroke in Saudi Arabia. *Cerebrovasc Dis* 13(1):3–8. <https://doi.org/10.1159/000047738>
- Alhazzani AA, Mahfouz AA, Abolyazid AY, Awadalla NJ (2021) Risk factors of the first-time stroke in the Southwest of Saudi Arabia: a case-control study. *Brain Sci* 11(2). <https://doi.org/10.3390/brainsci11020222>
- Alhazzani AA, Mahfouz AA, Abolyazid AY, Awadalla NJ, Aftab R, Faraheen A, Khalil SN (2018) Study of stroke incidence in the Aseer Region, Southwestern Saudi Arabia. *Int J Environ Res Public Health* 15(2). <https://doi.org/10.3390/ijerph15020215>
- Alkuraya, FS (2012) Discovery of rare homozygous mutations from studies of consanguineous pedigrees. *Curr Protoc Hum Genet*, Chapter 6, pp. Unit6.12. <https://doi.org/10.1002/0471142905.hg0612s75>
- Alokley AA, Albakr A (2022) Intracerebral hemorrhages related to clinical conditions and associated outcomes among Saudi and Non-Saudi Patients in a University Hospital: a retrospective study. *Prim Care Companion CNS Disord* 24(5). <https://doi.org/10.4088/PCC.21m03192>
- Alqahtani BA, Alenazi AM, Hoover JC, Alshehri MM, Alghamdi MS, Osaïlan AM, Khunti K (2020) Incidence of stroke among Saudi population: a systematic review and meta-analysis. *Neurol Sci* 41(11):3099–3104. <https://doi.org/10.1007/s10072-020-04520-4>
- Bindawas SM, Vennu VS (2016) Stroke rehabilitation A call to action in Saudi Arabia. *Neurosciences (Riyadh)* 21(4):297–305. <https://doi.org/10.17712/nsj.2016.4.20160075>
- Choi SW, O'Reilly PF (2019) PRSice-2: Polygenic Risk Score software for biobank-scale data. *Gigascience* 8(7). <https://doi.org/10.1093/gigascience/giz082>
- Dichgans M, Beaufort N, Debette S, Anderson CD (2021) Stroke genetics: turning discoveries into clinical applications. *Stroke* 52(9):2974–2982. <https://doi.org/10.1161/STROKEAHA.121.032616>
- Ekkert A, Šliachtenko A, Grigaitė J, Burnytė B, Utkus A, Jatužis D (2021) Ischemic stroke genetics: what is new and how to apply it in clinical practice?. *Genes (Basel)* 13(1). <https://doi.org/10.3390/genes13010048>
- Fernandes V, Brucato N, Ferreira JC, Pedro N, Cavadas B, Ricaut FX, Alshamali F, Pereira L (2019) Genome-wide characterization of arabian peninsula populations: shedding light on the history of a fundamental bridge between continents. *Mol Biol Evol* 36(3):575–586. <https://doi.org/10.1093/molbev/msz005>
- Flossmann E, Schulz UG, Rothwell PM (2004) Systematic review of methods and results of studies of the genetic epidemiology of ischemic stroke. *Stroke* 35(1):212–227. <https://doi.org/10.1161/01.STR.0000107187.84390.AA>
- Grami N, Chong M, Lali R, Mohammadi-Shemirani P, Henshall DE, Rannikmäe K, Paré G (2020) Global assessment of Mendelian stroke genetic prevalence in 101 635 individuals from 7 ethnic groups. *Stroke* 51(4):1290–1293. <https://doi.org/10.1161/STROKEAHA.119.028840>
- Hachiya T, Hata J, Hirakawa Y, Yoshida D, Furuta Y, Kitazono T, Shimizu A, Ninomiya T (2020) Genome-wide polygenic score and the risk of ischemic stroke in a prospective cohort: the Hisayama study. *Stroke* 51(3):759–765. <https://doi.org/10.1161/STROKEAHA.119.027520>
- Huang Y, Ballinger DG, Stokowski R, Beilharz E, Robinson JG, Liu S, Robinson RD, Henderson VW, Rossouw JE, Prentice RL (2012) Exploring the interaction between SNP genotype and postmenopausal hormone therapy effects on stroke risk. *Genome Med* 4(7):57. <https://doi.org/10.1186/gm358>
- Ilinca A, Samuelsson S, Piccinelli P, Soller M, Kristofferson U, Lindgren AG (2019) A stroke gene panel for whole-exome sequencing. *Eur J Hum Genet* 27(2):317–324. <https://doi.org/10.1038/s41431-018-0274-4>
- Ilinca A, Martinez-Majander N, Samuelsson S, Piccinelli P, Truvé K, Cole J, Kittner S, Soller M, Kristofferson U, Tatlisumak T, Puschmann A, Putaala J, Lindgren A (2020) Whole-exome sequencing in 22 Young ischemic stroke patients with familial clustering of stroke. *Stroke* 51(4):1056–1063. <https://doi.org/10.1161/STROKEAHA.119.027474>



- Janicki PK, Eyileten C, Ruiz-Velasco V, Pordzik J, Czlonkowska A, Kurkowska-Jastrzebska I, Sugino S, Imamura Kawasawa Y, Mirowska-Guzel D, Postula M (2019) Increased burden of rare deleterious variants of the KCNQ1 gene in patients with large-vessel ischemic stroke. *Mol Med Rep* 19(4):3263–3272. <https://doi.org/10.3892/mmr.2019.9987>
- Jørgensen C, Yasmeen S, Iversen HK, Kruuse C (2015) Phosphodiesterase4D (PDE4D)—a risk factor for atrial fibrillation and stroke? *J Neurol Sci* 359(1–2):266–274. <https://doi.org/10.1016/j.jns.2015.11.010>
- Kari JA, Bockenbauer D, Stanescu H, Gari M, Kleta R, Singh AK (2014) Consanguinity in Saudi Arabia: a unique opportunity for pediatric kidney research. *Am J Kidney Dis* 63(2):304–310. <https://doi.org/10.1053/j.ajkd.2013.08.033>
- Kumar A, Chauhan G, Sharma S, Dabla S, Sylaja PN, Chaudhary N, Gupta S, Agrawal CS, Anand KS, Srivastava AK, Vibha D, Sagar R, Raj R, Maheshwari A, Vivekanandhan S, Kaul B, Raghavan S, Gorthi SP, Mohania D, Kaushik S, Yadav RK, Hazarika A, Sharma P, Prasad K (2021) Association of SUMOylation pathway genes with stroke in a genome-wide association study in India. *Neurology* 97(4):e345–e356. <https://doi.org/10.1212/WNL.00000000000012258>
- Lavy S, Yaar I, Melamed E, Stern S (1974) The effect of acute stroke on cardiac functions as observed in an intensive stroke care unit. *Stroke* 5(6):775–780. <https://doi.org/10.1161/01.str.5.6.775>
- Lehman LL, Fullerton HJ (2013) Changing ethnic disparity in ischemic stroke mortality in US children after the STOP trial. *JAMA Pediatr* 167(8):754–758. <https://doi.org/10.1001/jamapediatrics.2013.89>
- Lehman LL, Khoury JC, Taylor JM, Yeramaneni S, Sucharew H, Alwell K, Moomaw CJ, Peariso K, Flaherty M, Khatri P, Broderick JP, Kissela BM, Kleindorfer DO (2018) Pediatric stroke rates over 17 years: report from a population-based study. *J Child Neurol* 33(7):463–467. <https://doi.org/10.1177/0883073818767039>
- Li J, Chaudhary DP, Khan A, Griessenauer C, Carey DJ, Zand R, Abedi V (2021) Polygenic risk scores augment stroke subtyping. *NeuroGenet* 7(2):e560. <https://doi.org/10.1212/NXG.0000000000000560>
- Malik R, Chauhan G, Traylor M, Sargurupremraj M, Okada Y, Mishra A et al (2018) Multiancestry genome-wide association study of 520,000 subjects identifies 32 loci associated with stroke and stroke subtypes. *Nat Genet* 50(4):524–537. <https://doi.org/10.1038/s41588-018-0058-3>
- Malik R, Bevan S, Nalls MA, Holliday EG, Devan WJ, Cheng YC, Ibrahim-Verbaas CA, Verhaaren BF, Bis JC, Joon AY, de Stefano AL, Fornage M, Psaty BM, Ikram MA, Launer LJ, van Duijn CM, Sharma P, Mitchell BD, Rosand J, Meschia JF, Levi C, Rothwell PM, Sudlow C, Markus HS, Seshadri S, Dichgans M, Wellcome Trust Case Control Consortium 2 (2014) Multilocus genetic risk score associates with ischemic stroke in case-control and prospective cohort studies. *Stroke* 45(2):394–402
- Monkare S, Kuuluvainen L, Schleutker J, Bras J, Roine S, Poyhonen M et al (2022) Genetic analysis reveals novel variants for vascular cognitive impairment. *Acta Neurol Scand* 146(1):42–50. <https://doi.org/10.1111/ane.13613>
- Munshi A, Kaul S (2008) Stroke genetics—focus on PDE4D gene. *Int J Stroke* 3(3):188–192. <https://doi.org/10.1111/j.1747-4949.2008.00199.x>
- Napolitano F, Dell'Aquila M, Terracciano C, Franzese G, Gentile MT, Piluso G, Santoro C, Colavito D, Patanè A, De Blasiis P, Sampaolo S, Paladino S, Melone MAB (2022) Genotype-phenotype correlations in neurofibromatosis type 1: identification of novel and recurrent. *Genes (Basel)* 13(7). <https://doi.org/10.3390/genes13071130>
- Pruissen DM, Slooter AJ, Rosendaal FR, van der Graaf Y, Algra A (2008) Coagulation factor XIII gene variation, oral contraceptives, and risk of ischemic stroke. *Blood* 111(3):1282–1286. <https://doi.org/10.1182/blood-2007-08-110254>
- Robert AA, Al-Dawish A, Mujammami M, Dawish MAA (2018) Type 1 diabetes mellitus in Saudi Arabia: a soaring epidemic. *Int J Pediatr* 2018:9408370. <https://doi.org/10.1155/2018/9408370>
- Stenson PD, Ball EV, Mort M, Phillips AD, Shiel JA, Thomas NS, Abeyasinghe S, Krawczak M, Cooper DN (2003) Human gene mutation database (HGMD): 2003 update. *Hum Mutat* 21(6):577–581. <https://doi.org/10.1002/humu.10212>
- Uemura M, Hatano Y, Nozaki H, Ando S, Kondo H, Hanazono A, Iwanaga A, Murota H, Osakada Y, Osaki M, Kanazawa M, Kanai M, Shibata Y, Saika R, Miyatake T, Aizawa H, Ikeuchi T, Tomimoto H, Mizuta I, Mizuno T, Ishihara T, Onodera O (2023) High frequency of. *J Neurol Neurosurg Psychiatry* 94(1):74–81. <https://doi.org/10.1136/jnnp-2022-329917>
- Virani SS, Alonso A, Aparicio HJ, Benjamin EJ, Bittencourt MS, Callaway CW (2021) Heart Disease and Stroke Statistics—2021 update: a report from the American Heart Association. *Circulation* 143(8):e254–e743. <https://doi.org/10.1161/CIR.0000000000000950>
- Walli-Attai M, Joseph P, Rosengren A, Chow CK, Rangarajan S, Lear SA et al (2020a) Variations between women and men in risk factors, treatments, cardiovascular disease incidence, and death in 27 high-income, middle-income, and low-income countries (PURE): a prospective cohort study. *Lancet* 396(10244):97–109. [https://doi.org/10.1016/S0140-6736\(20\)30543-2](https://doi.org/10.1016/S0140-6736(20)30543-2)
- Walli-Attai M, Joseph P, Rosengren A, Chow CK, Rangarajan S, Lear SA, AlHabib KF, Davletov K, Dans A, Lanas F, Yeates K, Poirier P, Teo KK, Bahonar A, Camilo F, Chifamba J, Diaz R, Didkowska JA, Irazola V, Ismail R, Kaur M, Khatib R, Liu X, Mańczuk M, Miranda JJ, Oguz A, Perez-Mayorga M, Szuba A, Tsolekile LP, Prasad Varma R, Yusufali A, Yusuf R, Wei L, Anand SS, Yusuf S (2020b) Variations between women and men in risk factors, treatments, cardiovascular disease incidence, and death in 27 high-income, middle-income, and low-income countries (PURE): a prospective cohort study. *Lancet* 396(10244):97–109. [https://doi.org/10.1016/S0140-6736\(20\)30543-2](https://doi.org/10.1016/S0140-6736(20)30543-2)
- Xu X, Li X, Li J, Ou R, Sheng W (2010) Meta-analysis of association between variation in the PDE4D gene and ischemic cerebral infarction risk in Asian populations. *Neurogenetics* 11(3):327–333. <https://doi.org/10.1007/s10048-010-0235-8>
- Yue X, Lixia L, Yan H, Zhang P, Gui Y, Song J (2019) Association between PDE4D polymorphism and ischemic stroke in young population. *Saudi J Biol Sci* 26(5):1023–1026. <https://doi.org/10.1016/j.sjbs.2019.04.007>
- Zheng SL, Sun J, Wiklund F, Smith S, Stattin P, Li G, Grönberg H (2008) Cumulative association of five genetic variants with prostate cancer. *N Engl J Med* 358(9):910–919. <https://doi.org/10.1056/NEJMoa075819>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.