Association between genetic predisposition and disease burden of stroke in China: a genetic epidemiological study

Qiya Huang,[a,h] Xianmei Lan,[b,f,h] Hebing Chen,[c] Hao Li,[c] Yu Sun,[c] Chao Ren,[c] Chao Xing,[d] Xiaochen Bo,[c,****] Jizheng Wang,[a,***] Xin Jin,[e,f,**] and Lei Song[a,g,*]

[a]State Key Laboratory of Cardiovascular Disease, Fuwai Hospital, National Center for Cardiovascular Diseases, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, China
[b]College of Life Sciences, University of Chinese Academy of Sciences, Beijing 100049, China
[c]Institute of Health Service and Transfusion Medicine, Beijing, China
[d]Eugene McDermott Center for Human Growth and Development, Department of Bioinformatics, Department of Population and Data Sciences, University of Texas Southwestern Medical Center, Dallas, TX, USA
[e]School of Medicine, South China University of Technology, Guangzhou, Guangdong, China
[f]BGI-Shenzhen, Shenzhen, China
[g]National Clinical Research Center of Cardiovascular Diseases, Fuwai Hospital, National Center for Cardiovascular Diseases, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, China

The Lancet Regional Health - Western Pacific 2023;36: 100779

Published Online 22 May 2023
https://doi.org/10.1016/j.lanwpc.2023.100779

Summary

Background Stroke ranks second worldwide and first in China as a leading cause of death and disability. It has a polygenic architecture and is influenced by environmental and lifestyle factors. However, it remains unknown as to whether and how much the genetic predisposition of stroke is associated with disease burden.

Methods Allele frequency from the whole genome sequencing data in the Chinese Millionome Database of 141,418 individuals and trait-specific polygenic risk score models were applied to estimate the provincial genetic predisposition to stroke, stroke-related risk factors and stroke-related drug response. Disease burden including mortality, disability-adjusted life years (DALYs), years of life lost(YLLs), years lived with disability (YLDs) and prevalence in China was collected from the Global Burden Disease study. The association between stroke genetic predisposition and the epidemiological burden was assessed and then quantified in both regression-based models and machine learning-based models at a provincial resolution.

Findings Among the 30 administrative divisions in China, the genetic predisposition of stroke was characterized by a north-higher-than-south gradient ($p < 0.0001$). Genetic predisposition to stroke, blood pressure, body mass index, and alcohol use were strongly intercorrelated (rho >0.6; $p < 0.05$ after Bonferroni correction for each comparison). Genetic risk imposed an independent effect of approximately 1–6% on mortality, DALYs and YLLs.

Interpretation The distribution pattern of stroke genetic predisposition is different at a macroscopic level, and it subtly but significantly impacts the epidemiological burden. Further research is warranted to identify the detailed aetiology and potential translation into public health measures.

Funding Beijing Municipal Science and Technology Commission (Z191100006619106), CAMS Innovation Fund for Medical Sciences (CAMS-I2M, 2023-I2M-1-001), the National High Level Hospital Clinical Research Funding (2022-GSP-GG-17), National Natural Science Foundation of China (32000398, 32171441 to X.J.), Natural Science

Abbreviations: PRS, Polygenic risk score; WGS, Whole genome sequencing; GWAS, Genome-wide association studies; SNP, Single nucleotide polymorphism; MAF, Minor allele frequency; CMDB, Chinese millionome database; DALY, Disability-adjusted life years; YLL, Years of life lost; YLD, Years lived with disability; GBD, Global burden of disease; BP, Blood pressure; BMI, Body mass index; T2D, Type 2 diabetes; TC, Total cholesterol; LDL-C, Low-density lipoprotein cholesterol

*Corresponding author. Department of Cardiomyopathy Center Fuwai Hospital, National Center for Cardiovascular Disease 167, Beilishilu, Xicheng District, 100037, Beijing, China.
**Corresponding author. BGI-Shenzhen, School of Medicine, South China University of Technology, Beishan Industrial Zone Complex, No. 146 Beishan Road, Yantian District, Shenzhen, 518083, China.
***Corresponding author. State Key Laboratory of Cardiovascular Disease Fuwai Hospital, National Center for Cardiovascular Disease 167, Beilishilu, Xicheng District, 100037, Beijing, China.
****Corresponding author. Institute of Health Service and Transfusion Medicine, No. 27, Taiping Road, Haidian District, 100850, Beijing, China.
E-mail addresses: songlqd@126.com (L. Song), jinxin@genomics.cn (X. Jin), jzwang@hotmail.com (J. Wang), boxiaoc@163.com (X. Bo).
[h]These authors contributed equally to this study.

Foundation of Guangdong Province, China (2017A030306026 to X.J.), and National Key R&D Program of China (2022YFC2502402).

Keywords: Stroke; Genetic risk; Disease burden; Geographical variation; China

Research in context

Evidence before this study

We searched PubMed, Embase, Cochrane, China National Knowledge Internet, China Science and Technology Journal Database, and Wanfang Data for original research on the association between stroke genetic predisposition and its disease burden published up to August 31st, 2022 in English and Chinese. We also focused on stroke-related pharmacogenetics. We used the keywords "stroke", "genetic risk", "disease burden", and "pharmacogenetics". Previous studies have yielded achievements on genetic variations associated with stroke, stroke-related cardiometabolic and behavioural traits, and stroke-related drug response. And different indicators of stroke disease burden in every country and subregion are available from the Global Burden Disease study. However, the distribution landscape of stroke genetic predisposition and pharmacogenetics, and the association between genetic risk and disease burden remain unknown.

Added value of this study

By applying both a Chinese-only polygenic risk score (PRS) model with 500 modelling selected genetic variants and an East Asian-specific PRS model with 4,856,268 genome-wide variants to the whole genome sequencing data of 141,418 individuals, we sketched the stroke genetic predisposition landscape at a provincial geographic scale in China, from which we discovered a north-higher-than-south gradient. We also depicted the genetically predicted metabolism of clopidogrel, statins and warfarin, indicating lower dosage for the Chinese population compared to Europeans. And we discovered that the genetic predisposition to stroke has an association with mortality, disability-adjusted life years (DALYs), years of life lost(YLLs), and years lived with disability (YLDs) (but not prevalence). We quantified that genetics accounts for 1–6%, a minor but nonnegligible part of the mortality, DALYs, and YLLs of stroke.

Implications of all the available evidence

Altogether, acquired factors play a more influential role in disease burden. However, differences in genetic risk for stroke are observed at a provincial resolution, and it has a small effect on the death-related disease burden. Early screening for genetic risk of stroke would help identify high-risk individuals and obtain more clinical benefits, as well as reduce the epidemiological burden. Further investigation is warranted to determine the precise aetiology and potential translation into public health measures.

Introduction

Among the leading causes of death and disability throughout the world, stroke ranks second worldwide and first in China,[1–3] reaching a total of 101 million prevalent cases, 6.55 million deaths, and 143 million disability-adjusted life years (DALYs) globally,[4] as well as 28.8 million prevalent cases, 2.19 million deaths, and 45.9 million DALYs in China,[2] in 2019. The global cost of stroke is estimated to exceed US$891 billion (1.12% of the global GDP).[5] Stroke burden continues to increase, particularly in lower-income and lower-middle-income countries, accounting for 86.0% of deaths and 89.0% of DALYs.[6] The considerable disease burden poses threats to public health and calls for urgent attention to its prevention, as well as an early warning.

Stroke is a common complex disease caused by various environmental, lifestyle and genetic factors.[4,7,8] Genome-wide association studies (GWAS) have identified ~100 genetic loci underlying stroke,[7–9] and the polygenic risk score (PRS), a weighted sum of the effects of genetic variants, has become a promising and powerful tool to evaluate the lifetime risk of stroke.[10,11] The incorporation of genetic predisposition into the conventional early risk stratification framework is appealing for individual clinical encounters.[7,12] However, it is unclear whether the genetic predisposition of stroke is associated with disease burden at a population level.

In this study, we characterized the stroke genetic predisposition landscape at a provincial resolution in China by applying two stroke PRS models on allele frequency estimates from the nationwide whole genome sequencing (WGS) dataset. Subsequently, we showed that genetic predisposition was correlated with mortality and DALY, but not prevalence. We inferred that genetics accounts for a nonnegligible part of the disease burden of stroke, thus providing novel insights into the genetic epidemiology of stroke.

Methods

Data collection

Allele frequency from the whole genome sequencing data

The genetic variants allele frequency at a provincial resolution were obtained from the Chinese Millionome

Database (CMDB)[13,14] (http://cmdb.bgi.com/), the most representative and comprehensive Chinese population genome database to date, consisting of WGS sequencing of 141,431 unrelated Chinese recruited individuals from 31 out of the 34 administrative divisions in China including Han and 36 other ethnic minorities. We included allele frequency data from 30 administrative divisions of 141,418 individuals after quality control (Supplementary Methods). Genetic data that supports the findings of this study are available from the corresponding author upon reasonable request. The study was approved by the Institutional Review Board of BGI (BGI-202212281301).

Polygenic risk score models and clinical drug response SNPs

We calculated the polygenic risk score (PRS) based on two models. PRS Model I is a Chinese-only PRS (https://www.pgscatalog.org/publication/PGP000285/) of 500 modelling selected genetic variants[15] (Supplementary File 1). PRS Model II is an East Asian-specific PRS model (https://www.pgscatalog.org/score/PGS002725/) of 4,856,268 genome-wide variants[7] (Supplementary File 2) (Supplementary Methods).

Based on the American Heart Association (AHA) guidelines, we focused on 4 stroke-related cardiometabolic traits (including body mass index (BMI), blood pressure (BP), type 2 diabetes (T2D), and low-density lipoprotein cholesterol (LDL-C) and 2 behavioural traits (alcohol use[16] and smoking susceptibility[17]) that have available trait-specific PRS models.[18] Moreover, we focused on clinical drug response SNPs associated with the metabolism of three categories of stroke-related drugs including warfarin,[19] clopidogrel,[20] and statins.[21] By comparison, we used the European (non-Finnish) population from the Genome Aggregation Database (gnomAD)[22] v3.1, consisting of 76,156 genomes of unrelated individuals. The SNPs from the latest weighted genetic risk score (wGRS) model of caffeine metabolism[23,24] were also included (Supplementary Methods and Supplementary Table S2).

Epidemiological burden data

We use five dimensions of standard epidemiological measures: mortality, prevalence, years of life lost (YLLs), years lived with disability (YLDs), and disability-adjusted life-years (DALYs). We collected the number and age-standardized rates of the aforementioned metrics available of each province in 1990, 2016, and 2019 from the GBD study.[2,25] We used age-standardized rates to correct for different age distributions and improve comparability for each province (Supplementary Methods and Supplementary Table S3 for details).

Based on the American Heart Association (AHA) Guidelines, we focused on smoking rate, physical inactivity rate, salt intake level (g/d), insufficient intake rate of vegetables and fruits, obesity rate, the morbidity rate of hypertension, the morbidity rate of hyperlipoidemia as covariates when quantifying the impact of genetic components on disease burden. We split 30 provinces evenly into the northern and southern geographical groups. Additionally, we included the concentration of PM2.5, the use of medication interventions and the composite socioeconomic index (RCDI). We collected covariates data for each province in 2019 (Supplementary Methods and Supplementary Table S4 for details).

Estimating genetic predisposition to stroke and other traits

Our study proposed to characterize the stroke genetic predisposition landscape at a provincial resolution in China, which is calculated as the following Equation (1) (see Supplementary Methods for detailed formula derivation):

$$\widehat{PRS}_k = \sum_{j=1}^{M} \left(\beta_j^{'} \cdot MAF_j\right) \qquad \text{(Equation 1)}$$

For province k, $\widehat{PRS}_k$ is the average polygenic risk score for administrative division k, M is the number of SNPs in the PRS model, and β_j is the effect size of SNP_j from the PRS model.

We calculated the weighted sums of provincial-level MAF for SNPs in the trait-specific models as genetic predispositions and mapped them by percentile, with darker colours representing higher genetic predispositions. Particularly, for the two stroke PRS models, our results were discriminated by colour, with orange representing PRS Model I and blue representing PRS Model II.

Study design

The overall design is shown in Fig. 1.

All data were derived from 30 of 34 provincial-level administrative division units in China, including 22 provinces, 4 municipalities, and 4 autonomous regions. Both municipalities and autonomous regions are termed provincial administrative units in China, although they are not named provinces. Hong Kong, Macao, Taiwan and Tibet were not included due to the limited number of sequencing participants.

Our study used each province as a proxy unit, and we aimed to evaluate the association of genetic predisposition to stroke with disease burden (mortality, prevalence, DALYs, YLLs, YLDs) and estimate the predictive value of the PRS at the provincial geographic scale. First, we estimated the genetic predisposition for stroke and its risk factors (including BP, BMI, LDL-C, T2D, alcohol use, and smoking) at a provincial resolution in China. Specifically, for stroke, we utilized two different validated PRS models—the Chinese-only PRS with 500 modelling selected variants (PRS Model I) and the East Asian-specific PRS with 4,856,268 genome-wide variants (PRS Model II). Afterwards, we performed

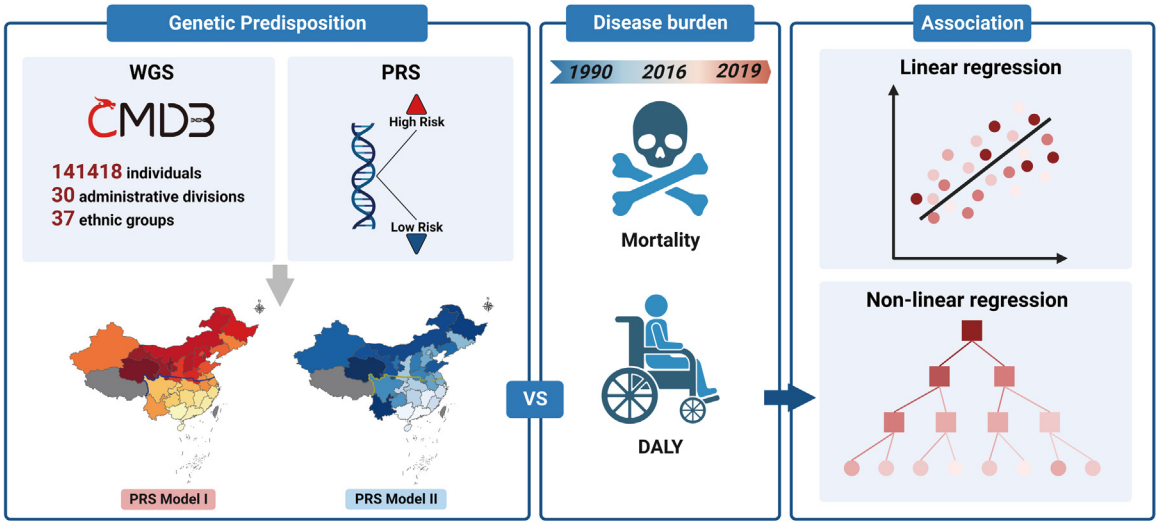

***Fig. 1:* The overall design of the study**. Our study used each province as a proxy unit, and we aimed to evaluate the association of genetic predisposition to stroke with disease burden (mortality, prevalence, DALYs, YLLs, YLDs) and estimate the predictive value of the PRS at the provincial geographic scale. Genetic predisposition distributions for stroke and its risk factors (including BP, BMI, LDL-C, T2D, alcohol use, and smoking) of each administrative division in China were first evaluated from the provincial allele frequencies from the Chinese Millionome Database (CMDB) and trait-specific polygenic risk score (PRS) models. Specifically, for stroke, we utilized two types of validated models, i.e., the Chinese-only PRS (PRS Model I) with 500 modelling-selected variants and the East Asian-specific PRS (PRS Model II) with 4,856,268 genome-wide variants. Afterwards, we performed association analyses between the stroke genetic predisposition of both models and different epidemiological burden metrics of each province. To ensure comparability among different provinces, we used the age-standardized rates of the above epidemiological measurements, from which significant correlations were revealed. Finally, we conducted regression analyses of mortality, DALYs, and YLLs in both linear and nonlinear models for PRS Model I and II to correct for covariates and obtain independent estimates of the genetic predisposition of stroke on the epidemiological burden, thus discovering that genetic risk accounts for a minor (but significant) role. DALYs: disability-adjusted life-years. YLLs: years of life lost. YLDs: years lived with disability. PRS: Polygenic risk score.

association analyses between the stroke genetic predisposition of both PRS models and different epidemiological burden metrics of each province, including mortality, DALYs, YLLs, YLDs, and prevalence. Subsequently, we conducted regression analyses of mortality, DALY, and YLL in both linear models and nonlinear (random forest) models to correct for covariates and obtain independent estimates of the genetic predisposition of stroke on the epidemiological burden.

Statistical analysis

Comparisons of genetic predisposition between two different regions (the northern group and southern group) were performed by using the Mann–Whitney U test. Correlations between genetic predisposition and disease burden metrics, as well as pairwise correlations among genetic risks of different traits, were compared by using the Spearman rank correlation via the cor.test() function from the psych (2.2.3) package of R (4.0.5) software. Correlation coefficients and significance test p values are reported. The correlation coefficient rho $\geq$0.8 was considered a very strong correlation, rho = 0.60 to 0.79 was considered a strong correlation, rho = 0.4 to 0.59 was considered a moderate correlation, rho = 0.20 to 0.39 was considered a weak correlation, and rho $\leq$0.19 was considered a very weak correlation. A p value < 0.05 was considered statistically significant. For pairwise comparisons among the genetic risks of stroke and stroke-related traits, the significant P value for Bonferroni correction was 0.05/28 = 0.0018. For the correlation between the genetic risk of stroke and epidemiological burden metrics, the significant P value for Bonferroni correction was 0.05/13 = 0.0038 for the Chinese-specific model and East Asian-specific model.

Scatter plots and heatmaps were visualized by using the ggplot2 (3.3.5) package and corrplot (0.92) package in R (4.0.5) software. Bonferroni corrections were made for all multiple significance testing.

Regression modelling

Linear model. A simple linear regression between genetic predisposition and disease burden of stroke is as follows:

$$\text{Burden}_i \sim \beta * \text{PRS}_i + \mu$$

where Burden_i is the stroke disease burden of province i; PRS_i is a polygenic risk score for province i; β is the fitted slope; and μ is the intercept. R square and significance test p values were reported. Fitting was performed by using the lm() function from the stats (4.0.5) package in R (4.0.5) software. The multivariate linear

regression of disease burden is as follows:

$$\text{Burden}_i \sim \beta_1 * \text{PRS}_i + \beta_2 * x_{2i} + \beta_3 * x_{3i} + \ldots + \beta_n * x_{ni} + \mu$$

where x_{2i}, x_{3i}, ..., x_{ni} are the covariates of each province, and β_{2i}, β_{3i}, ..., β_{ni} are the fitted coefficients of the corresponding covariates.

Linear Model 1 is the full model, and covariates include RCDI, the use of medication interventions for risk factors, physical inactivity rate, salt intake level (g/d), insufficient intake rate of vegetables and fruits, obesity rate, the morbidity rate of hypertension, the morbidity rate of hyperlipoidemia, the concentration of PM2.5, smoking rate, and geographical regions (dichotomized into the north and the south). Afterwards, we performed a backward variable selection of the full model until we obtained a minimized AIC (Akaike information criterion) as Linear Model 2. Then, we excluded the PRS component of the second model as Linear Model 3. Fitting was performed by using the lm() function, and stepwise variable selection was performed by using the step() function, both included in the stats (4.0.5) package in R (4.0.5) software. Residual plots were used to verify that normal distributions of residuals and constant variance assumptions were satisfied for linear regressions. Adjusted R-square values and p values were reported.

Nonlinear (Machine learning) model. The random forest as a machine learning algorithm confronting multicollinearity[26] was implemented by using the a3() function from the A3 (1.0.0) package in R (4.0.5) software. The significance p value and R-square values were implemented based on 1000 permutation tests.

Role of the funding source

The funders of the study had no role in the study design, data analysis, data interpretation, or writing of the manuscript.

Results

Genetic predisposition distribution landscape of stroke, risk factors, and metabolism

The overall design is shown in Fig. 1 (see Methods). First, we depicted the provincial-level genetic predisposition distribution landscape of stroke (Fig. 2a–b), stroke-related risk factors (Supplementary Fig. S1), stroke-related clinical drug response SNPs (Supplementary Fig. S2a–i), and caffeine metabolism-related SNPs (Supplementary Fig. S2j–l) in China, according to the provincial allele frequencies derived from whole genome sequencing (WGS) data from the Chinese Millionome Database (CMDB),[13] trait-specific polygenic risk score (PRS) models and the Clinical Pharmacogenetics Implementation Consortium (CPIC) guidelines (Methods, Supplementary Methods and Supplementary Tables S1 and S2 in the Appendix).

We estimated the provincial genetic risks of stroke from a Chinese-only model (PRS Model I) consisting of 500 modelling selected SNPs and an East Asian-specific model (PRS Model II) consisting of 4,856,268 genome-

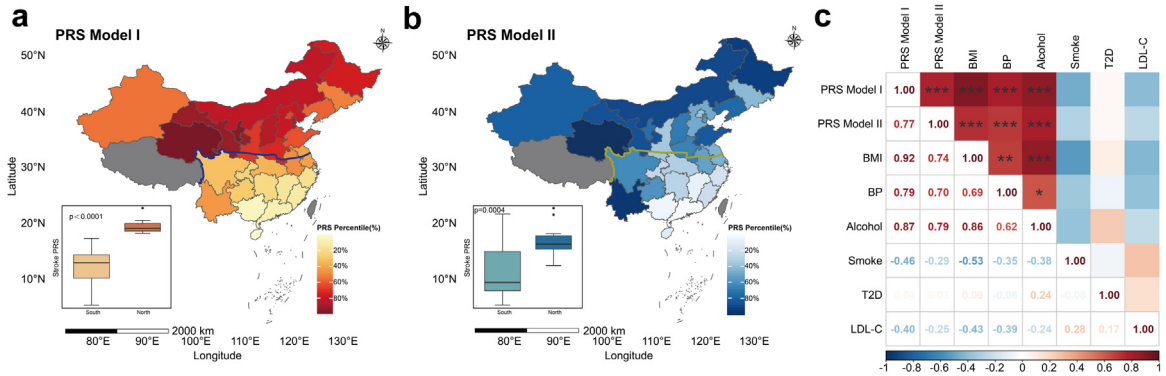

***Fig. 2:* Genetic predisposition for stroke and risk factors in China**. a. Genetic predisposition for stroke in China derived from PRS Model I (Methods). Colour shades represent percentile rankings. The North has a higher score than the South ($p < 0.0001$, two-sided Mann–Whitney U test). b. Genetic predisposition for stroke in China derived from PRS Model II (Methods). Colour shades represent percentile rankings. The North has a higher score than the South ($p < 0.0004$, two-sided Mann–Whitney U test). c. Pairwise correlation of genetic predisposition for stroke and risk factors, including BMI, BP, LDL-C, T2D, alcohol use, and smoking. The bottom-right triangular matrix shows the Spearman correlation coefficient. The correlation coefficient rho ≥0.8 was considered a very strong correlation, rho = 0.60 to 0.79 was considered a strong correlation, rho = 0.4 to 0.59 was considered a moderate correlation, rho = 0.20 to 0.39 was considered a weak correlation, and rho ≤0.19 was considered a very weak correlation. The upper-left triangular matrix highlights significance via asterisks, and blank spaces represent non-significant effects (Spearman correlation, α threshold = 0.05/28 [total number of comparisons]; specifically, $p < 0.0018$. Asterisks indicate significance, with *** denoting $p < 0.001$ and ** denoting $p < 0.01$). PRS: Polygenic risk score. BMI: body mass index. BP: blood pressure. LDL-C: low-density lipoprotein cholesterol. T2D: Type 2 Diabetes.

wide variants, and the risks were generally consistent (Fig. 2c, rho = 0.77, p < 0.0001, Spearman correlation, after Bonferroni correction). A coherent north-to-south gradient of stroke genetic predisposition was observed, with the north being higher than the south (Fig. 2a–b, p < 0.0001 for PRS Model I, p = 0.0004 for PRS Model II, two-sided Mann–Whitney U test). Overall, there was a precipitous change observed in the provinces along the Huai River-Qinling Mountain line (blue line in Fig. 2a and yellow line in Fig. 2b), whereas the trend was more moderately gradated within the provinces on either side of the dividing line.

Moreover, according to American Heart Association (AHA) guidelines, we further evaluated the genetic predisposition distribution landscape of several cardiovascular health (CVH) metrics, including body mass index (BMI), blood pressure (BP), type 2 diabetes (T2D), and low-density lipoprotein cholesterol (LDL-C), as well as two lifestyle behaviours (alcohol consumption and smoking) (Supplementary Fig. S1). We observed north-south differences for all of the traits except T2D and LDL-C. A trend of higher-in-north and lower-in-south was observed among genetic risk for BMI (p < 0.0001, two-sided Mann–Whitney U test), BP (p = 0.0001, two-sided Mann–Whitney U test), and alcohol consumption (p < 0.0001, two-sided Mann–Whitney U test). In contrast, the genetic risk of smoking showed a higher-in-south and lower-in-north trend (p = 0.0027, two-sided Mann–Whitney U test). No significant differences were observed for T2D (p = 0.44, two-sided Mann–Whitney U test) or LDL-C (p = 0.098, two-sided Mann–Whitney U test).

We then conducted pairwise comparisons among genetic susceptibility to stroke and stroke-related traits (Fig. 2c). In addition to the consistency in stroke genetic risk derived from PRS Model I and II, strong correlations also exist between the genetic risk of stroke and that of BMI (PRS Model I: rho = 0.92, p < 0.0001, Spearman correlation after Bonferroni correction; PRS Model II: rho = 0.74, p < 0.0001, Spearman correlation after Bonferroni correction), between the genetic risk of stroke and that of BP (PRS Model I: rho = 0.79, p < 0.0001, Spearman correlation after Bonferroni correction; PRS Model II: rho = 0.70, p < 0.0001, Spearman correlation after Bonferroni correction), as well as between genetic risk of stroke and that of alcohol consumption (PRS Model I: rho = 0.87, p < 0.0001, Spearman correlation after Bonferroni correction; PRS Model II: rho = 0.79, p < 0.0001, Spearman correlation after Bonferroni correction).

In addition, we illustrated the population frequency of stroke-related drug response SNPs (clopidogrel for antiplatelet therapy, statins for lipid-lowering therapy, and warfarin for anticoagulation therapy) among different provinces (Supplementary Fig. S2a–i). Notably, for 8 of 9 SNPs that indicate decreased metabolism of medications, including rs12769205 (*CYP2C19*), rs4244285 (*CYP2C19*), and rs4986893 (*CYP2C19*) as no function alleles for clopidogrel; rs2306283 (*SLCO1B1*) and rs4149056 (*SLCO1B1*), rs2231142 (*ABCG2*), and rs1057910 (*CYP2C9*), as poor function alleles for statins; rs1057910 (*CYP2C9*), rs9923231 (*VKORC1*), and rs9934438 (*VKORC1*) as the decreased alleles for warfarin, most provinces demonstrated a higher-frequency pattern than that of European (the MAF of the aforementioned SNPs in non-Finnish European are 0.15, 0.15, 0.00018, 0.40, 0.16, 0.11, 0.066, 0.38, and 0.38, accordingly), implying lower dosages for most Chinese patients than European patients.

We also displayed the provincial allele frequency of SNPs that genetically predict the increased metabolism of caffeine, indicating increased caffeine intake. The weighted average distribution pattern (Supplementary Fig. S2j–l, and Supplementary Table S3) appeared to be generally equal across the regions with the exception of Qinghai, tagged with an asterisk, with all three SNPs displaying high frequency.

Correlation between the genetic predisposition and disease burden of stroke

We hypothesized that at the population level, the genotype–phenotype correlation of stroke exists. To test this hypothesis, we compared stroke genetic predisposition with different epidemiological burden metrics at the provincial geographic scale, including the age-standardized rate of mortality, DALYs, YLLs, YLDs, and prevalence, in 1990, 2016, and 2019 of each province.

In both PRS Model I and II (Supplementary Figs. S3 and S4), we observed that across different years, mortality and DALYs were associated with genetic predisposition of a moderate to a strong degree rather than prevalence. Further stratified analysis of YLL and YLD (Supplementary Figs. S5 and S6), which are the two sources of DALYs, demonstrated a stronger association of YLLs than that of YLDs with the genetic predisposition. We also observed that more recently (2016 and 2019), the most developed administrative divisions (Beijing and Shanghai) have a lower rate of mortality and DALY than the genetically predicted ones. While less developed provinces, such as Qinghai, have higher rates of mortality and DALY than genetically predicted provinces.

Estimating the importance of genetic predisposition to the disease burden of stroke

To further determine the significance of genetic predictors for stroke burden, first, we performed simple linear regression analyses and then conducted multivariable regression analyses in both linear regression and nonlinear regression in both PRS models, which were corrected for the potential confounders including smoking, physical inactivity, salt intake, intake of vegetables and fruits, obesity, hypertension, hyperlipidemia,

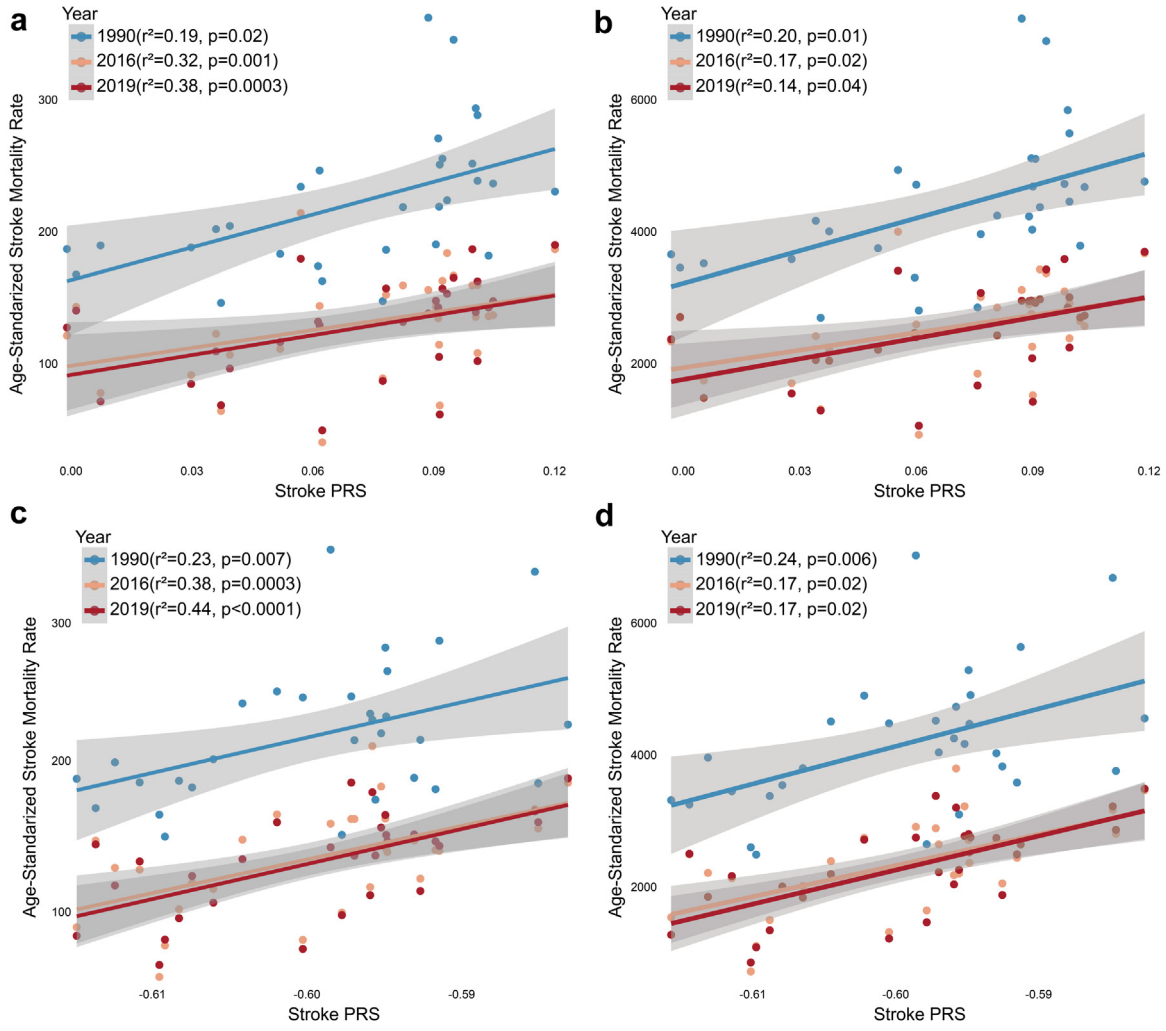

***Fig. 3*: Correlation between the disease burden and genetic predisposition for stroke in 1990, 2016, and 2019**. a. Simple linear Regression of age-standardized mortality rate (per 100,000) and genetic predisposition in PRS Model I. b. Simple linear Regression of age-standardized DALY rate (per 100,000) and genetic predisposition in PRS Model I. c. Simple linear Regression of age-standardized mortality rate (per 100,000) and genetic predisposition in PRS Model II. d. Simple linear Regression of age-standardized DALY rate (per 100,000) and genetic predisposition in PRS Model II. Shadowed areas indicate 95% confidence intervals. DALY: age-standardized disability-adjusted life year.

the concentration of PM2.5, medication use, RCDI, and geographical regions (Methods).

In simple linear regression, genetic risk is a significant predictive variable for mortality (Fig. 3a–b; PRS Model I: $r^2 = 0.19$, $p = 0.02$ in 1990; $r^2 = 0.32$, $p = 0.001$ in 2016; $r^2 = 0.38$, $p = 0.0003$ in 2019; PRS Model II: $r^2 = 0.20$, $p = 0.01$ in 1990; $r^2 = 0.17$, $p = 0.02$ in 2016; $r^2 = 0.14$, $p = 0.04$ in 2019) and DALYs (PRS Model I: $r^2 = 0.23$, $p = 0.007$ in 1990; $r^2 = 0.38$, $p = 0.0003$ in 2016; $r^2 = 0.44$, $p < 0.0001$ in 2019; PRS Model II: $r^2 = 0.24$, $p = 0.006$ in 1990; $r^2 = 0.17$, $p = 0.02$ in 2016; $r^2 = 0.17$, $p = 0.02$ in 2019, Fig. 3c–d) other than prevalence (Supplementary Fig. S7) in both PRS models across different years. Regression validity was confirmed with the normal distribution of residuals and constant variance in most regressions (Supplementary Figs. S8–S13).

Second, we performed multivariable analyses. From linear models (Table 1), although the full model (Linear Model 1) with lower adjusted R-square values did not identify genetic risk as a significant variable, after stepwise regression, Linear Model 2 with the highest adjusted R-square values determined genetic risk as an independent predictor of mortality and DALYs in both PRS models in 3 of 4 tests ($p = 0.0003$ for mortality in PRS Model I; $p = 0.2$ for DALYs in PRS Model I; $p = 0.05$ for mortality in the PRS Model II; $p = 0.03$ for DALYs in PRS Model II). After excluding the genetic factor, the adjusted R-square of Linear Model 3 decreased. Overall, we observed that Linear Model 2 was

Estimated coefficient (SD)	PRS Model I						PRS Model II					
	Age-standardized Mortality Rate (per 100,000)			Age-standardized DALYs Rate (per 100,000)			Age-standardized Mortality Rate (per 100,000)			Age-standardized DALYs Rate (per 100,000)		
	(1)	(2)	(3)	(1)	(2)	(3)	(1)	(2)	(3)	(1)	(2)	(3)
RCDI[b]	−1.83**	−2.17***	−1.68***	−32.38*	−35.58***	−33.17***	−1.64*	−1.96***	−2.00***	(28.08)	−32.23***	−33.17***
	(0.84)	(0.47)	(0.58)	(16.59)	(9.65)	(9.64)	(0.86)	(0.45)	(0.48)	(16.74)	(8.98)	(9.64)
Drug Intervention	−0.51*	−0.39**	−0.62***	−11.55**	−9.44***	−10.51***	−0.51*	−0.37**	−0.46***	−11.53**	−8.55**	−10.51***
	(0.25)	(0.16)	(0.20)	(4.95)	(3.34)	(3.30)	(0.25)	(0.16)	(0.16)	(4.82)	(3.19)	(3.30)
Stroke PRS	301.84	410.84***		5083.80	4412.72		952.00	917.91*		18332.48	20,145.42**	
	(210.98)	(97.13)		(4188.09)	(3229.10)		(601.41)	(449.76)		(11782.89)	(8952.82)	
Obesity	0.16	−0.25*	−0.31*	1.84	−6.93**	−8.33**	0.27			4.24		
	(0.37)	(0.13)	(0.17)	(7.36)	(2.91)	(3.83)	(0.38)			(7.49)		
Hypertension	0.12	0.20	0.10	5.89	5.72	3.44	0.05			0.89		
	(0.23)	(0.16)	(0.20)	(5.39)	(3.41)	(4.46)	(0.22)			(4.27)		
Geographical Regions	16.16			405.17	355.60*	562.30***	20.24	14.59**	23.22***	469.57	372.94**	562.30***
	(16.03)			(318.22)	(192.50)	(121.04)	(15.36)	(7.06)	(5.99)	(300.97)	(140.54)	(121.04)
Smoking	−0.17			−3.63			−0.14			−3.08		
	(0.19)			(3.71)			(0.18)			(3.54)		
Physical inactivity	−0.01			1.17			−0.00			0.87		
	(0.31)			(6.10)			(0.30)			(5.87)		
Salt Intake	0.04			1.69			0.03			1.73		
	(0.31)			(6.15)			(0.30)			(5.93)		
Insufficient Intake of Vegetables and Fruits	0.14			3.38			0.13			3.03		
	(0.17)			(3.38)			(0.17)			(3.31)		
Hyperlipidemia	−0.15			−3.85			−0.20			−4.80		
	(0.23)			(4.61)			(0.23)			(4.54)		
PM2.5	−0.06			0.18			−0.17			−1.88		
	(0.19)			(3.72)			(0.20)			(3.87)		
Constant	305.70***	338.19***	329.12***	5570.00***	6000.96***	6013.98***	880.57**	887.57***	341.52***	16,511.80**	17,998.07***	6013.98***
	(96.16)	(40.40)	(51.44)	(1908.84)	(832.12)	(845.83)	(341.07)	(270.46)	(41.85)	(6682.34)	(5383.62)	(845.83)
Adjusted R^2	0.74	0.80	0.67	0.75	0.79	0.78	0.75	0.80	0.78	0.76	0.81	0.78

Note: *p < 0.1; **p < 0.05; ***p < 0.01. (1) Full mode. (2) Backwards stepwise regression of the full model. (3) Remove the PRS component of the second model. [a]DALY: Disability-adjusted life years. [b]RCDI: Renmin University China Development Index, a composite socioeconomic index.

Table 1: **Multivariate linear regression of mortality and DALY[a].**

best fitted for both mortality and DALY. In addition, as stratified analyses for DALYs, we obtained similar results for YLLs (Supplementary Table S5). Most regressions obtained normal distribution of residuals and constant variance (Supplementary Figs. S14–S19).

On the other hand, we performed nonlinear regression as complementary to the linear regressions combating potential collinearity among different predictive variables by using the random forest algorithm.[26] For mortality and DALYs (Table 2), the genetic risk was still significant as an independent variable, with a partial R-square of approximately 1–6% in both PRS models ($p = 0.03$ for mortality in PRS Model I; $p = 0.01$ for DALYs in PRS Model I; $p = 0.002$ for mortality in PRS Model II, $p = 0.06$ for DALYs in PRS Model II). Meanwhile, we obtained a partial R-square of approximately 4–6% for YLLs as stratified analyses for DALYs (Supplementary Table S6).

We concluded that genetic predisposition plays a small but nonnegligible role in stroke epidemiological burden, especially for mortality, DALYs and YLLs.

Discussion

By applying two different validated PRS models respectively derived from modelling selected SNPs and whole genome-wide variants to the allele frequency data from WGS from 141,418 unrelated Chinese people in 30 provinces (Methods and Fig. 1), we depicted coherent genetic predispositions to stroke in China at a provincial resolution for the first time (Fig. 2a–b). We found the genetic risk was associated with mortality and DALYs, but not prevalence and dissected the small but nonnegligible impact of genetic risk on disease burden (Tables 1 and 2). We also illustrated the provincial genetic risk of stroke-related cardiometabolic traits and behavioural traits (Supplementary Fig. S1). Moreover, we displayed the allele frequency of the stroke-related drug response SNPs and caffeine metabolism-related SNPs (Supplementary Fig. S2). Overall, we revealed stroke's genetic predisposition at a provincial geographic scale and found its correlation with the disease burden.

The interest in personalized early risk stratification for complex diseases has increased.[12] However, its limited trans-ancestry applicability has long been criticized. The latest GWAS, led by the GIGASTROKE consortium,[7] now provides an integrated East Asian-specific PRS model derived from a Japanese-majority population. Although both Japanese and Chinese are East Asians, they do have differences in genetic backgrounds[27] and distinguishable lifestyles. To ensure the validity of our findings, we employed a Chinese-only PRS model with 500 SNPs[15] and an integrated East Asian-specific PRS model with nearly 5 million SNPs to evaluate the provincial genetic predisposition to stroke. The results from PRS Model I and II showed agreement, demonstrating the reliability of the quantitative inferences on the depiction and impact of stroke genetic risk.

Identifying genetic predispositions can help with early stratification. We illustrated genetic predispositions to stroke as well as its risk factors and detected a noticeable north-south gradient for most traits (Fig. 2a and b and Supplementary Fig. S1). The genetic differences are in conjunction with the well-

Estimated coefficient (SD)	PRS Model I				PRS Model II			
	Age-standardized Mortality Rate (per 100,000)		Age-standardized DALYs[a] Rate (per 100,000)		Age-standardized Mortality Rate (per 100,000)		Age-standardized DALYs[a] Rate (per 100,000)	
	r^2	p-value	r^2	p-value	r^2	p-value	r^2	p-value
Full Model	0.52	<0.001	0.55	<0.001	0.60	<0.001	0.61	<0.001
RCDI[b]	0.14	<0.001	0.07	<0.001	0.13	<0.001	0.06	0.001
Drug Intervention	0.14	0.001	0.14	<0.001	0.11	<0.001	0.12	<0.001
Stroke PRS	0.02	0.03	0.01	0.01	0.04	0.002	0.06	<0.001
Geographical Regions	0.00	0.2	0.00	0.1	0.00	0.004	0.01	0.03
Obesity	−0.02	0.6	−0.02	0.4	0.00	0.2	−0.01	0.3
Hypertension	−0.01	0.4	−0.03	0.4	0.01	0.03	−0.02	0.2
Smoking	−0.02	0.4	−0.01	0.6	0.00	0.1	−0.01	0.4
Physical inactivity	−0.01	0.7	−0.02	0.4	0.00	0.04	-0.01	0.1
Salt Intake	−0.02	0.4	−0.03	0.8	0.01	0.1	−0.02	0.6
Insufficient Intake of Vegetables and Fruits	−0.02	0.7	−0.02	0.8	−0.01	0.1	−0.03	1
Hyperlipidemia	−0.03	0.8	−0.02	0.6	−0.01	0.2	−0.01	0.4
PM2.5	−0.02	0.5	−0.01	0.6	0.00	0.1	−0.02	0.4
(Intercept)	0.00	0.2	0.01	0.3	0.00	0.03	0.00	0.2

[a]DALY: Disability-adjusted life years. [b]RCDI: Renmin University China Development Index, a composite socioeconomic index.

Table 2: **Random forest regression of Stroke disease burden.**

known genomic dissection of the population substructure of the Chinese population, which is partly attributed to Huai River-Qinling Mountain as a natural barrier of restricted gene flow. And the north-south genetic gradient is consistent with the long-existing north-south gradient of epidemiological burden[28–30] or the geographical "stroke belt",[31] which is defined as a higher stroke burden of north and west China. Additionally, the strong correlation among the estimated provincial genetic risk of stroke, BP, and BMI (Fig. 2c) is consistent with the deduced cause of "stroke belt": specifically, the higher prevalence of hypertension and excess body weight in stroke belt regions leads to a higher burden of stroke.

Despite the stable correlation between mortality and YLLs and genetic predisposition, however, the disassociation between prevalence and genetic predisposition is counterintuitive. This is partially due to the multiple etiologies of stroke. For instance, GWAS may aim to discover the loci associated with the most prevalent etiologies,[32] such as large-artery atherosclerotic stroke (LAS) and cardioembolic stroke (CES), which tend to have a more severe prognosis. Additionally, during the construction of PRS, only the loci that are more likely to be associated with more than one subtype of stroke[33] are not pruned, thus leading to the possibilities of their greater effects on poor prognosis. In addition, the data quality of mortality may be closer to the actual situation than prevalence because the overall framework for the national surveillance of cardiovascular disease (CVD) incidence and prevalence has just begun,[34] whereas the establishment of the death registration reporting system dates back to 1978 and is more experienced.[35]

Pharmacogenetics guides tailored therapeutic interventions. Different genetic dosages of clinical drug response SNPs can influence individual pharmacokinetics, dose needs, and safety. Awareness of the genetic variations among populations is essential for identifying patients who may experience an adverse drug response or no drug response, thus optimizing patients' prognosis. We found stroke-related drug response SNPs showed varying frequencies (Supplementary Fig. S2). Clopidogrel, as the most widely prescribed antiplatelet medication, has higher resistance among East Asians.[36] This can be partly explained by a higher frequency of CYP2C19 nonfunctional allele carriers than Europeans as we observed (Supplementary Fig. S2a–c). Statins, the cornerstone for primary and secondary prevention, are more likely to induce adverse side effects, such as myalgia, in the Chinese population.[36] It can be attributed to the higher frequency of SNPs for statin metabolic decline as we demonstrated (Supplementary Fig. S2d–g). In addition, clinical prescription of warfarin needs special caution in China because Asians exhibit greater thromboembolic protection at lower INR, lower initiation and maintenance doses of warfarin, and a higher risk of ICH caused by warfarin.[36] The phenomenon is in line with the allele frequencies of the SNPs that decrease warfarin metabolism are more frequent in the Chinese population (Supplementary Fig. S2g–i) when compared to Europeans. Our findings agree with the conclusions of previous clinical trials,[37–39] which demonstrated that Asian ancestry patients require a genotype-guided lower dosage than European guidelines have recommended for clinical benefits. By contrast, Europeans made up more than 80% of recruited patients in pivotal efficacy random control trials implementing new approval of cardiometabolic drugs,[36] raising disparity among underrepresented populations. And due to financial constraints, drug–drug and drug–diet interactions, and a lack of enough high-quality studies in non-white populations, a genotype-guided strategy is not routine in the clinic. However, we do not suggest that the 'one-size-fits-all' therapy strategy would endure, and with the increasing ease and declining cost of sequencing tests, as well as more well-designed and high-quality clinical studies being carried out in more diverse populations, the genotype-guided strategy will play a more important role in the individualized cardiovascular care.

Caffeine is a widely consumed psychoactive agent in daily beverages such as coffee, tea and soft drinks, attracting increasing attention for cardiovascular health, and it has been recognized as a healthy lifestyle with adequate intake in recent years.[40] The coffee intake inferred from the three accessible SNPs appeared to be relatively even across provinces (except for Qinghai) (Supplementary Fig. S2j–l), and more verification is needed.

Finally, we inferred that genetic risk plays a small but independent role in disease burden (Fig. 3, Tables 1 and 2, Supplementary Figs. S3–S6, Supplementary Tables S5 and S6). Although simple linear regression results showed that hereditary factors can explain approximately 30% of the disease burden for a single variable (Fig. 3), which is comparable with some family lineage studies that have obtained heritability for stroke up to 32%[41]; we believe this result is over-estimated owing to confounding effects by covariates within the same family, such as environment and lifestyle. Therefore, we included covariates of more dimensions, thus yielding a finding that genetic risk affects stroke burden by approximately 1–6% (Tables 1 and 2). In terms of methodology, we use both linear regressions and nonlinear regressions to ensure validity. On the other hand, from quantitative calculations, our estimation is consistent with a multiancestry genome-wide association meta-analysis in 521,612 individuals, which indicated that the phenotypic variance explained by SNPs ranged between 0.6% and 1.8%.[32] Therefore, our findings on the correlation between genetic risk and disease burden reveal the predictive value of PRS from a macroscopic perspective.

Previous studies have revealed that attributable environmental and lifestyle factors account for more

than 90%[2] of stroke disease burden and that genetics plays a nonnegligible role in the remaining 10% of stroke disease burden. Additionally, we observed a much larger effect of RCDI and medication intervention than genetic components, which emphasizes the necessity of intervention of modifiable acquired factors, including environment, lifestyle, healthy cardiometabolic metrics, and timely drug interventions. It is promising that at an individual level, previous studies revealed that people at higher genetic risk can benefit more from maintaining a healthy lifestyle and adhering to preventive treatment, compared to those at lower risk. Thereby, they offset increased genetic risk.[15] We value it as an opportunity to promote public awareness and patient compliance if high-risk individuals can be identified beforehand. On the other hand, at a population level, as previous simulation analysis demonstrated that if CVD risk factors (including smoking, physical inactivity, BMI, fasting glucose levels, total cholesterol levels, and systolic blood pressure) were adequately controlled, more than 700,000 deaths among people aged 30–70 years could be avoided.[42] Hence, early screening for genetic risk of stroke would help identify high-risk individuals and obtain more clinical benefits, as well as reduce the epidemiological burden. While nurture is more influential than nature in the case of stroke, a clearer understanding of nature would make a difference in the outcome.

Our study had several strengths. First, we utilized two PRS models of different magnitudes of SNPs derived from the Chinese-only population and East Asian-specific population, thus avoiding poor trans-ancestry portability as much as possible. Importantly, consistent results were achieved from both models, strengthening the reliability of our conclusions. Second, the genetic data and epidemiological data that we applied are obtained from the most comprehensive and representative genetic database and the most authoritative published statistics in China, respectively, allowing us to fill the gap in the landscape of genetic predisposition to stroke and stroke-related traits as well as genetically predicted drug response at the population level. Third, we made exhaustive efforts to ensure the robustness of the genotype–phenotype correlation. We not only compared epidemiological data from different years but also analyzed multidimensional covariates and managed to fit both linear and nonlinear models. The relatively stable association we identified between genetic risk and death-related disease burden (mortality, DALYs and YLLs) provides a new perspective for future research.

The study results should also be interpreted in light of several limitations. First, we analyzed summary-statistical level data at a provincial resolution, and our findings warrant further individual-level granular validation. Second, the PRS calculation results may not be perfectly accurate to elucidate the complete heritability of stroke due to the loss of a small fraction of SNPs (6.4% for the Chinese-specific model and 19.2% for the East Asian-specific model). Fortunately, the proportion is not enough to twist the overall results. More detailed and diversified data collection in population genetics is needed to expand insights into genetic assessment, etiological explorations, and precision medicine of stroke. Third, epidemiological data are limited in some remote areas wherein poor data availability intensifies regional disparity; thus, this effect could be underestimated.[2] This result also suffers from the shortcomings inherent in the GBD methodological framework, including challenges in fully quantifying all sources of uncertainty, variation in coding practices, and other biases.[25] A better disease surveillance system is currently under construction and is expected to deliver new information to help in reducing the disease burden.

Conclusion

We delineated the distribution of genetic predisposition to stroke and common cardiovascular risk factors from a population-level perspective in China and detected a minor but nonnegligible independent effect of genetic predisposition on stroke burden. Collectively, our findings provide novel insights into the genetic epidemiology of stroke. Further granular research of genetic risk prediction and genotype-guided pharmacotherapy is warranted to mitigate stroke burden clinical outcomes.

Contributors

Q.H. and X.L. conducted the analyses and drafted the manuscript. X.B., J.W., X.J. and L.S. supervised study design, data collection, data analysis and data interpretation. C.X. contributed to data interpretation and manuscript preparation. H.C., H.L., C.R. and Y.S. helped prepare the epidemiological data. All authors participated in the preparation and revision of the manuscript, and all authors approved the final version. Q.H. and X.L. contributed equally as the co-first authors. X.B., J.W, X.J. and L.S. contributed equally as the corresponding authors.

Data sharing statement

Anonymised genetic data from the Chinese Millionome Database (CMDB) is available upon reasonable request with a designated proposal submitted to the corresponding author for consideration. Epidemiological disease burden data, study protocol and the statistical analysis plan are available in the paper.

Editor note

The Lancet Group takes a neutral position with respect to territorial claims in published maps and institutional affiliations.

Declaration of interests

We declare no competing interests.

Acknowledgements

We would like to express our gratitude to the patients and their families for their participation, to all staff members for data collection, sample handling, and genotyping. This work was supported by China National GeneBank (CNGB). We thank Xiangfeng Lu (State Key Laboratory of Cardiovascular Disease, Fuwai Hospital, National Center for Cardiovascular Diseases, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, China), Chaolong Wang (Department of Epidemiology and Biostatistics, Ministry of Education Key Laboratory of Environment and Health and State Key Laboratory of Environmental

Health (Incubating), School of Public Health, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, China), Yi Liu (Department of Biostatistics, School of Public Health, Cheeloo College of Medicine, Shandong University, Jinan, China) for valued scientific guidance. We would also like to acknowledge the support from *BioRender* in figure generation (https://biorender.com/), and Data-V.GeoAtlas from Aliyun (http://datav.aliyun.com/portal) for the map drawing.

Appendix A. Supplementary data

Supplementary data related to this article can be found at https://doi.org/10.1016/j.lanwpc.2023.100779.

References

1 Roth GA, Mensah GA, Johnson CO, et al. Global burden of cardiovascular diseases and risk factors, 1990-2019: update from the GBD 2019 study. *J Am Coll Cardiol*. 2020;76(25):2982–3021.
2 Ma Q, Li R, Wang L, et al. Temporal trend and attributable risk factors of stroke burden in China, 1990-2019: an analysis for the Global Burden of Disease Study 2019. *Lancet Public Health*. 2021;6(12):e897–e906.
3 Thayabaranathan T, Kim J, Cadilhac DA, et al. Global stroke statistics 2022. *Int J Stroke*. 2022;17(9):946–956.
4 Collaborators GBDS. Global, regional, and national burden of stroke and its risk factors, 1990-2019: a systematic analysis for the Global Burden of Disease Study 2019. *Lancet Neurol*. 2021;20(10):795–820.
5 Owolabi MO, Thrift AG, Mahal A, et al. Primary stroke prevention worldwide: translating evidence into action. *Lancet Public Health*. 2022;7(1):e74–e85.
6 Feigin VL, Brainin M, Norrving B, et al. World stroke organization (WSO): global stroke fact sheet 2022. *Int J Stroke*. 2022;17(1):18–29.
7 Mishra A, Malik R, Hachiya T, et al. Stroke genetics informs drug discovery and risk prediction across ancestries. *Nature*. 2022;611(7934):115–123.
8 Debette S, Markus HS. Stroke genetics: discovery, insight into mechanisms, and clinical perspectives. *Circ Res*. 2022;130(8):1095–1111.
9 Network NSG, International Stroke Genetics C. Loci associated with ischaemic stroke and its subtypes (SiGN): a genome-wide association study. *Lancet Neurol*. 2016;15(2):174–184.
10 Marston NA, Patel PN, Kamanu FK, et al. Clinical application of a novel genetic risk score for ischemic stroke in patients with cardiometabolic disease. *Circulation*. 2021;143(5):470–478.
11 O'Sullivan JW, Shcherbina A, Justesen JM, et al. Combining clinical and polygenic risk improves stroke prediction among individuals with atrial fibrillation. *Circ Genom Precis Med*. 2021;14(3):e003168.
12 O'Sullivan JW, Raghavan S, Marquez-Luna C, et al. Polygenic risk scores for cardiovascular disease: a scientific statement from the American Heart association. *Circulation*. 2022;146(8):e93–e118.
13 Li Z, Jiang X, Fang M, et al. CMDB: the comprehensive population genome variation database of China. *Nucleic Acids Res*. 2022;51(D1):D890–D895.
14 Liu S, Huang S, Chen F, et al. Genomic analyses from non-invasive prenatal testing reveal genetic associations, patterns of viral infections, and Chinese population history. *Cell*. 2018;175(2):347–359.e14.
15 Lu X, Niu X, Shen C, et al. Development and validation of a polygenic risk score for stroke in the Chinese population. *Neurology*. 2021;97(6):e619–e628.
16 Hu C, Huang C, Li J, et al. Causal associations of alcohol consumption with cardiovascular diseases and all-cause mortality among Chinese males. *Am J Clin Nutr*. 2022;116(3):771–779.
17 Geng T, Chang X, Wang L, et al. The association of genetic susceptibility to smoking with cardiovascular disease mortality and the benefits of adhering to a DASH diet: the Singapore Chinese Health Study. *Am J Clin Nutr*. 2022;116(2):386–393.
18 Lu X, Liu Z, Cui Q, et al. A polygenic risk score improves risk stratification of coronary artery disease: a large-scale prospective Chinese cohort study. *Eur Heart J*. 2022;43(18):1702–1711.
19 Johnson JA, Caudle KE, Gong L, et al. Clinical pharmacogenetics implementation consortium (CPIC) guideline for pharmacogenetics-guided warfarin dosing: 2017 update. *Clin Pharmacol Ther*. 2017;102(3):397–404.
20 Lee CR, Luzum JA, Sangkuhl K, et al. Clinical pharmacogenetics implementation consortium guideline for CYP2C19 genotype and clopidogrel therapy: 2022 update. *Clin Pharmacol Ther*. 2022;112(5):959–967.
21 Cooper-DeHoff RM, Niemi M, Ramsey LB, et al. The clinical pharmacogenetics implementation consortium guideline for SLCO1B1, ABCG2, and CYP2C9 genotypes and statin-associated musculoskeletal symptoms. *Clin Pharmacol Ther*. 2022;111(5):1007–1021.
22 Karczewski KJ, Francioli LC, Tiao G, et al. The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature*. 2020;581(7809):434–443.
23 Cornelis MC, Kacprowski T, Menni C, et al. Genome-wide association study of caffeine metabolites provides new insights to caffeine metabolism and dietary caffeine-consumption behavior. *Hum Mol Genet*. 2016;25(24):5472–5482.
24 Loftfield E, Cornelis MC, Caporaso N, Yu K, Sinha R, Freedman N. Association of coffee drinking with mortality by genetic variation in caffeine metabolism: findings from the UK biobank. *JAMA Intern Med*. 2018;178(8):1086–1097.
25 Liu S, Li Y, Zeng X, et al. Burden of cardiovascular diseases in China, 1990-2016: findings from the 2016 global burden of disease study. *JAMA Cardiol*. 2019;4(4):342–352.
26 Lindner T, Puck J, Verbeke A. Beyond addressing multicollinearity: robust quantitative analysis and machine learning in international business research. *J Int Bus Stud*. 2022;53(7):1307–1314.
27 Pan Z, Xu S. Population genomics of East Asian ethnic groups. *Hereditas*. 2020;157(1):49.
28 Ru X, Wang W, Sun H, et al. Geographical difference, rural-urban transition and trend in stroke prevalence in China: findings from a national epidemiological survey of stroke in China. *Sci Rep*. 2019;9(1):17330.
29 Wang W, Jiang B, Sun H, et al. Prevalence, incidence, and mortality of stroke in China: results from a nationwide population-based survey of 480 687 adults. *Circulation*. 2017;135(8):759–771.
30 He J, Klag MJ, Wu Z, Whelton PK. Stroke in the People's Republic of China. I. Geographic variations in incidence and risk factors. *Stroke*. 1995;26(12):2222–2227.
31 Xu G, Ma M, Liu X, Hankey GJ. Is there a stroke belt in China and why? *Stroke*. 2013;44(7):1775–1783.
32 Malik R, Chauhan G, Traylor M, et al. Multiancestry genome-wide association study of 520,000 subjects identifies 32 loci associated with stroke and stroke subtypes. *Nat Genet*. 2018;50(4):524–537.
33 Neurology Working Group of the Cohorts for H. Aging Research in Genomic Epidemiology Consortium tSGN, the International Stroke Genetics C. Identification of additional risk loci for stroke and small vessel disease: a meta-analysis of genome-wide association studies. *Lancet Neurol*. 2016;15(7):695–707.
34 Xue C, Jia-yin C, Zeng-wu W. Experiences and implications of cardiovascular health and disease surveillance. *Chin J Cardiovasc Res*. 2022;20(2):188–192.
35 Mei Y, Yan L, Hai-feng L, Xiang-ying C. Current status of domestic death cause monitoring in China. *Occup Health*. 2020;36(2):280–283.
36 Tamargo J, Kaski JC, Kimura T, et al. Racial and ethnic differences in pharmacotherapy to prevent coronary artery disease and thrombotic events. *Eur Heart J Cardiovasc Pharmacother*. 2022;8(7):738–751.
37 Jia DM, Chen ZB, Zhang MJ, et al. CYP2C19 polymorphisms and antiplatelet effects of clopidogrel in acute ischemic stroke in China. *Stroke*. 2013;44(6):1717–1719.
38 Zhao S, Wang Y, Mu Y, et al. Prevalence of dyslipidaemia in patients treated with lipid-lowering agents in China: results of the DYSlipidemia International Study (DYSIS). *Atherosclerosis*. 2014;235(2):463–469.
39 Syn NL, Wong AL, Lee SC, et al. Genotype-guided versus traditional clinical dosing of warfarin in patients of Asian ancestry: a randomized controlled trial. *BMC Med*. 2018;16(1):104.
40 van Dam RM, Hu FB, Willett WC. Coffee, caffeine, and health. *N Engl J Med*. 2020;383(4):369–378.
41 Bak S, Gaist D, Sindrup SH, Skytthe A, Christensen K. Genetic liability in stroke: a long-term follow-up study of Danish twins. *Stroke*. 2002;33(3):769–774.
42 Li Y, Zeng X, Liu J, et al. Can China achieve a one-third reduction in premature mortality from non-communicable diseases by 2030? *BMC Med*. 2017;15(1):132.