

# Detailed Analysis of Gene Polymorphisms Associated with Ischemic Stroke in South Asians

Sunaina Yadav<sup>1</sup>, Nazeeha Hasan<sup>1</sup>, Thomas Marjot<sup>1</sup>, Muhammad S. Khan<sup>1</sup>, Kameshwar Prasad<sup>2</sup>, Paul Bentley<sup>1</sup>, Pankaj Sharma<sup>1\*</sup>

**1** Imperial College Cerebrovascular Research Unit (ICCRU), Imperial College London, London, United Kingdom, **2** Department of Neurology, All India Institute of Medical Sciences, New Delhi, India

## Abstract

The burden of stroke is disproportionately high in the South Asian subcontinent with South Asian ethnicity conferring a greater risk of ischemic stroke than European ancestry regardless of country inhabited. While genes associated with stroke in European populations have been investigated, they remain largely unknown in South Asians. We conducted a comprehensive meta-analysis of known genetic polymorphisms associated with South Asian ischemic stroke, and compared effect size of the MTHFR C677T-stroke association with effect sizes predicted from homocysteine-stroke association. Electronic databases were searched up to August 2012 for published case control studies investigating genetic polymorphisms associated with ischemic stroke in South Asians. Pooled odds ratios (OR) for each gene-disease association were calculated using a random-effects model. We identified 26 studies (approximately 2529 stroke cases and 2881 controls) interrogating 33 independent genetic polymorphisms in 22 genes. Ten studies described MTHFR C677T (108 with TT genotype and 2018 with CC genotype) -homocysteine relationship and six studies (735 stroke cases and 713 controls) described homocysteine-ischemic stroke relationship. Risk association ORs were calculated for ACE I/D (OR 5.00; 95% CI, 1.17–21.37;  $p=0.03$ ), PDE4D SNP 83 (OR 2.20; 95% CI 1.21–3.99;  $p=0.01$ ), PDE4D SNP 32 (OR 1.57; 95% CI 1.01–2.45,  $p=0.045$ ) and IL10 G1082A (OR 1.44; 95% CI, 1.09–1.91,  $p=0.01$ ). Significant association was observed between elevated plasma homocysteine levels and MTHFR/677 TT genotypes in healthy South Asians (Mean difference ( $\Delta X$ ) 5.18  $\mu\text{mol/L}$ ; 95% CI 2.03–8.34;  $p=0.001$ ). Our results demonstrate that the genetic etiology of ischemic stroke in South Asians is broadly similar to the risk conferred in Europeans, although the dataset is considerably smaller and warrants the same clinical considerations for risk profiling.

**Citation:** Yadav S, Hasan N, Marjot T, Khan MS, Prasad K, et al. (2013) Detailed Analysis of Gene Polymorphisms Associated with Ischemic Stroke in South Asians. PLoS ONE 8(3): e57305. doi:10.1371/journal.pone.0057305

**Editor:** Jean-Claude Baron, University of Cambridge, United Kingdom

**Received:** October 15, 2012; **Accepted:** January 20, 2013; **Published:** March 7, 2013

**Copyright:** © 2013 Yadav et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Funding:** PS and PB hold senior fellowships from the Dept of Health (UK). SY is funded by the United Kingdom-India Education and Research Initiative award (UKIERI). NH is funded by the Medical Research Council (MRC). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing Interests:** The authors have declared that no competing interests exist.

\* E-mail: pankaj.sharma@imperial.ac.uk

## Introduction

South Asia comprising of India, Pakistan, Sri Lanka and Bangladesh, forms 20% of the world's populous and shoulders much of the global death burden from cardiovascular disease [1] with India reporting 930,985 cases of stroke in 2004 leading to 639,455 deaths and loss of 6.4 million disability adjusted life years [2]. The past decade has seen prevalence rates of stroke rise in South Asia from between ~200 [3–6] to 545 per 100,000 persons [7], while incidence studies demonstrate exponential increases in stroke incidence rates of >800% over the past 30 years (e.g. from 13 to 123 per 100 000 persons per year between 1969 and 1993) [8,9]. Studies of South Asian diasporas in the West have also shown a heightened prevalence and incidence of stroke and coronary artery disease compared to Caucasians [10,11]. This is compounded by the fact that South Asians are developing stroke at a relatively young age, despite lower rates of alcohol and tobacco use [12]. Although 90% of the population attributable risk (PAR) for stroke has been suggested to be attributed to ten modifiable risk factors [13], this does not account for the occurrence of stroke in young unexposed populations and also fails to explain the

development of stroke in only some individuals within a population that is uniformly exposed to environmental risk factors. As with Europeans [14,15], it is likely that the causality of stroke in South Asians involves the complex interaction between genetic and environmental risk factors.

We undertook a comprehensive meta-analysis of all known genetic associations with ischemic stroke in South Asians and compare it to published results in different ethnic groups. We further sought to establish whether homocysteine, the putative biochemical intermediary of the MTHFR gene is associated with quantitative levels of risk [16] in South Asians as similarly shown in Europeans [14,16].

## Methods

### (1) SNP analysis

**Data searching.** We identified all published case-control studies evaluating any gene polymorphisms and ischemic stroke in South Asians residing in their native countries and diasporas around the world. Electronic searches were conducted using Medline, EMBASE and Google Scholar. All published manu-

scripts up until and including August 2012 as well as letters, previous meta-analyses and abstracts were included. The retrieved studies were examined thoroughly to assess their appropriateness for inclusion in our study. The references of all identified publications were manually reviewed for additional studies and the PUBMED 'relevant articles' option was utilized. The following index terms along with 'and/or' as a Boolean operator were used: "South Asia" "India" "Pakistan" "Sri Lanka" "Bangladesh" for ancestry and "stroke genetics" "gene polymorphism" "gene mutation" "stroke genes" for genetics and "stroke" "cerebrovascular disease" "ischemic stroke" "brain infarction" "brain ischemia" for clinical phenotype.

**Study selection.** Study inclusion criteria were: (1) studies in populations of South Asian descent defined as Indian, Pakistani, Sri Lankan or Bangladeshi; (2) case-control studies where ischemic stroke was analyzed as a dichotomous trait; (3) stroke was confirmed using brain imaging with sub-acute (within 10 days) CT or MRI, and; (4) genotype frequency for both cases and controls was reported. Studies were excluded if: (1) subjects were <18 years age; (2) the genotype frequency was not reported and could not be obtained from authors and (3) stroke other than ischemic.

**Data extraction.** Data extracted from each study included: first author, journal, year of publication, stroke sub-type, and number of cases and controls for each genotype and SNP. Baseline characteristics for cases and controls were documented including mean age, gender, ethnicity and geographical location.

**Data analysis.** For each genetic variant for which data were available from at least two studies, a meta-analysis was carried out. Data was analyzed using Review Manager v5.0 and Comprehensive Meta-Analysis (CMA) v2.0. To test for strength of association for each gene variant, a pooled odds ratio (OR) and 95% confidence interval (CI) was calculated using a random effects model [17]. The strength of genetic association (risk or protection) for each polymorphism was considered statistically significant if a p-value of <0.05 was obtained. For each analysis an  $I^2$  test for heterogeneity [18] was performed, with significance set at  $p < 0.05$ . Funnel plots and Egger regression intercept p-value (two-tailed) were used to determine probability of publication bias [19]. Comparison of ORs of genes with significant associations was made between ethnic populations by observing whether 95% confidence intervals overlapped.

## (2) MTHFR C677T – homocysteine phenotype comparison

For MTHFR C677T gene variant which had an associated biomarker homocysteine, we performed a separate analysis that produced an estimate of expected risk based upon genotype–biomarker, and biomarker–stroke, association studies, using mendelian randomization [14,16,20].

**Data searching.** Electronic searches were conducted using Medline, EMBASE and Google Scholar and all published manuscripts up until and including August 2012 were considered. The following index terms along with 'and/or' as Boolean operators were used for MTHFR: (MTHFR OR *Methylenetetrahydrofolate reductase*) AND (gene OR genetic OR genotype OR polymorphism OR mutation), in combination with (MTHFR OR *Methylenetetrahydrofolate reductase*) AND (homocysteine OR activity OR level); and 2) (MTHFR OR *Methylenetetrahydrofolate reductase*) AND (activity OR level) in conjunction with (cerebrovascular disease OR brain infarction OR stroke OR cerebral ischemia) and (South Asia or India or Pakistan or Sri Lanka or Bangladesh).

**Study selection.** The literature was searched for two types of studies: (Study 1) case-control studies reporting dichotomous and

continuous data for plasma homocysteine levels in South Asians with ischemic stroke, and; (Study 2) studies linking plasma homocysteine levels with the MTHFR 677 wild type (CC) and homozygous mutant (TT) genotype in healthy South Asians. Control populations with genotype-homocysteine data from case-control studies on psoriasis, coronary artery disease and glaucoma were also included.

**Data extraction.** Data extracted from each study included: first author, journal, year of publication, stroke sub-type, and total number of participants. Baseline characteristics for cases and controls were documented including mean age, gender, ethnicity and geographical location. In addition for Study 1, plasma homocysteine levels as means with standard deviations and median with ranges in ischemic stroke cases and controls, and; for Study 2, plasma homocysteine levels as means with standard deviations and median with ranges in MTHFR 677 wild type (CC) and homozygous mutant (TT) genotype in healthy South Asians.

**Data analysis.** The methodology for a literature based mendelian randomization for MTHFR and homocysteine is well described [14,16]. Where values of plasma homocysteine were reported as medians and ranges, the mean and standard deviation were estimated using established models dependent upon sample size [21]. For both types of studies, mean homocysteine difference between ischemic stroke cases vs. controls and MTHFR TT vs. CC genotypes were calculated using the continuous data type inverse variance method in Review Manager v5.0. A pooled mean difference ( $\Delta X$ ) with 95% confidence interval was calculated for MTHFR TT vs CC genotypes using a random effects model. Mean difference in homocysteine levels between ischemic stroke cases vs. controls were converted to ORs and 95% confidence interval using Comprehensive Meta-Analysis (CMA) v2.0. Assuming a log-linear relationship, this OR was scaled with the  $\Delta X$  change in homocysteine levels conferred by the MTHFR TT genotype to give the expected OR [16].

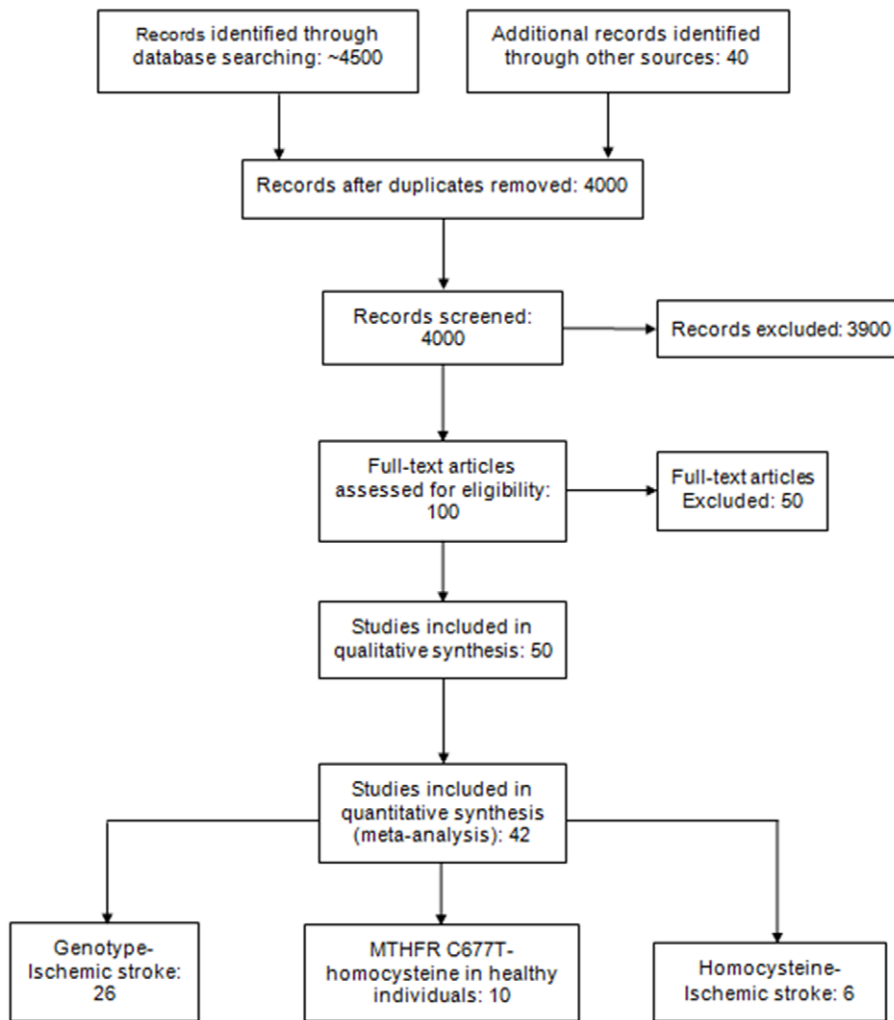
For each meta-analysis an  $I^2$  test for heterogeneity was performed, with significance set at  $p < 0.05$ . Funnel plots and Egger regression intercept p-value (two-tailed) were used to determine probability of publication bias.

## Results

### Meta-Analysis of Gene Variants Associated with Ischemic Stroke

Our search strategy identified ~4500 potentially relevant studies of which 26 met the inclusion criteria allowing interrogation of 33 independent genetic polymorphisms in 22 different genes across 2529 stroke cases and 2881 controls (Figure 1 PRISMA flowchart). The majority of studies were from India (North and south) followed by Pakistan, Bangladesh, Malaysia and the United Kingdom (Table S1)

Relevant studies identified were; five for MTHFR C677T [22–26], three for PDE4D SNP 83 [27–29] and 2 studies each for eNOS 4a/4b [30,31], ACE [32,33], ApoE E4/E4 [34,35], FVL G1691A [36,37], PDE4D SNP 87 and 32 [38,39] and IL10 [40,41] genes. Two studies each reporting different stroke sub types (Cerebral Venous Thrombosis (CVT) and arterial paediatric ischemic stroke) were found for genes MTRR G66A [42,43], MTR A2756G [42,43], MTHFR A1298C [42,43] and FVL A4070G [36,42,43], and therefore ORs for risk were not estimable. Prothrombin G20210A polymorphism was investigated in two studies [37,44]. Homozygous or heterozygous variants were completely absent in all individuals studied and with only the wild type genotype being expressed an OR for risk was not estimable. One relevant study was identified for genes variants IL-1 $\alpha$



**Figure 1. PRISMA statement flow diagram illustrating search strategy and studies included in the meta-analysis.**\*Articles excluded because they did not meet inclusion criteria and were missing genotype data. doi:10.1371/journal.pone.0057305.g001

(C889T) [45], CYP11B2 (C344T) [46], ESR1 (PVUII and XbaI) [47],  $\alpha$  ADD1 (WG) [33], TNF  $\alpha$  (G488A and G308A) [48], CYP4F2 (G1347A) [49], MDR-1 [50], t-PA (C7351T and I/D) [51], PAI-1 (4G/5G) [51], CBS (T833C) [52], Klotho (KL-VS and C1818T) [53], Factor XIIIIB (V34L) [54],  $\alpha$ 1 antichymotrypsin (Ala15Thr) [55] and MMP3 (5A/6A) [48]. Of these IL-1 $\alpha$ , CYP11B2, ESR1 (PVUII),  $\alpha$  ADD1, TNF  $\alpha$  (G488A), CYP4F2, MDR-1 and t-PA (I/D) were found to be significantly associated with ischemic stroke (Table S1).

#### PDE4D SNP 83

The PDE4D SNP83 polymorphism was investigated in three studies covering Pakistan and North India [27–29] in a total of 1338 subjects (626 Ischemic stroke cases; 712 controls). A pooled OR of 2.20 (95% CI, 1.21–3.99,  $p=0.01$ ) was generated with a recessive random-effects model. There was no evidence of inter-study heterogeneity ( $[P_{HET}] = 0.06$ ,  $I^2 = 65\%$ ).

#### ACE I/D

The homozygous ACE D/D polymorphism was investigated in a total of 2 studies covering India [32,33] in 693 subjects (355 ischemic stroke cases; 338 controls) providing a pooled OR of 5.00

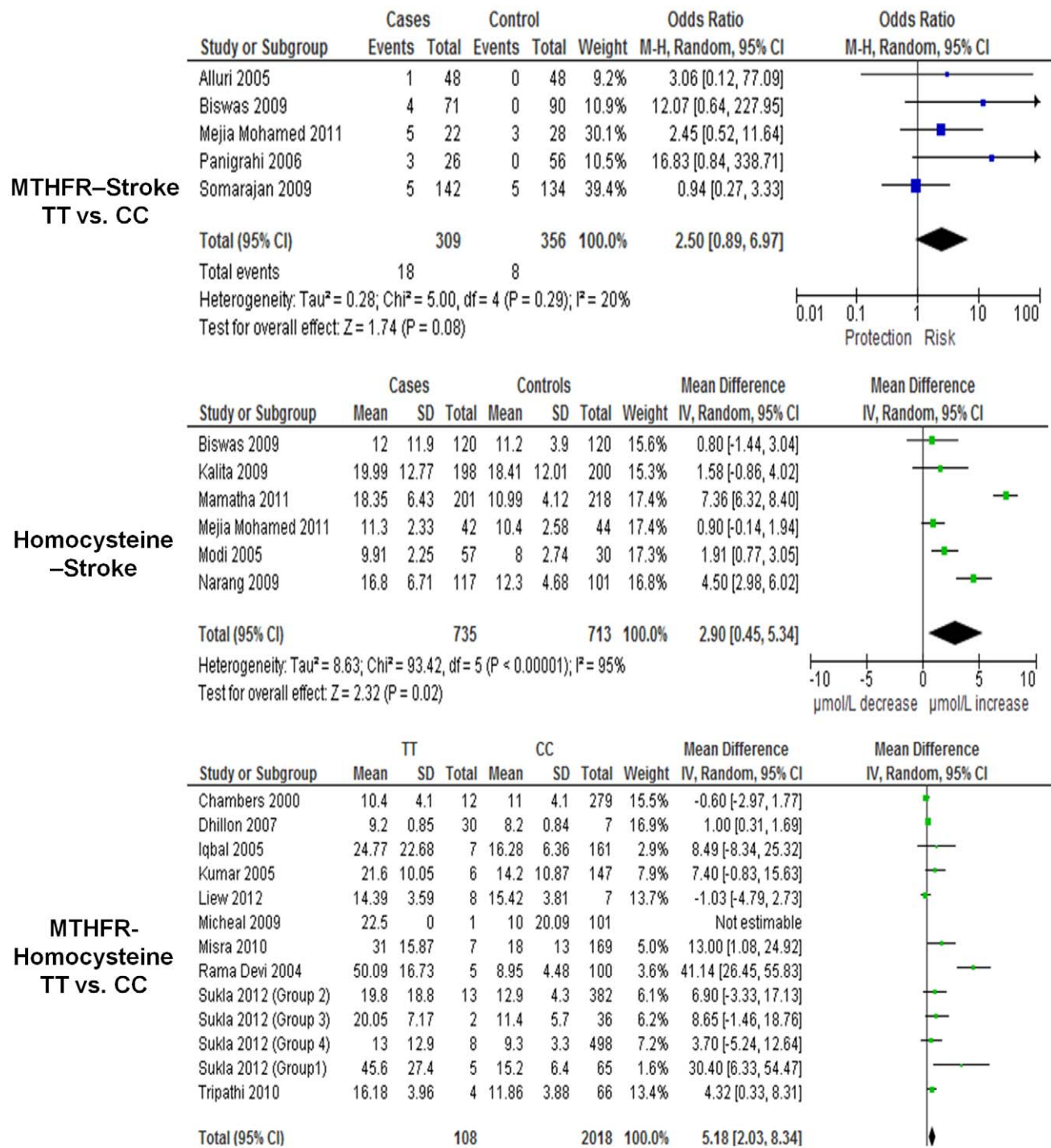
(95% CI, 1.17–21.37,  $p=0.03$ ) with a recessive random-effects model. There was evidence of inter-study heterogeneity ( $[P_{HET}] = 0.002$ ,  $I^2 = 90\%$ ).

#### IL10 G1082A

The homozygous G1082A polymorphism was investigated in a total of 2 studies covering India in 1414 subjects (718 ischemic stroke cases; 696 controls) providing a pooled OR of 1.44 (95% CI, 1.09–1.91,  $p=0.01$ ) with a recessive random-effects model. There was no evidence of inter-study heterogeneity ( $[P_{HET}] = 0.12$ ,  $I^2 = 59\%$ ).

#### MTHFR/C677T

The homozygous MTHFR C677T polymorphism was investigated in seven studies [22–25,36,43,56] out of which five studies covering North India and Malaysian Indians [22–25,36] were included in the final analysis. These totalled 615 subjects (309 Ischemic stroke cases; 356 controls) providing a pooled OR of 2.50 (95% CI, 0.89–6.97,  $p=0.08$ ) with a recessive random-effects model (Fig. 2). There was no evidence of inter-study heterogeneity ( $[P_{HET}] = 0.33$ ,  $I^2 = 14\%$ ).



**Figure 2. Meta-analysis, forest plot and pooled ORs of risk from studies investigating MTHFR C677T polymorphism (TT vs CC) - stroke, homocysteine-stroke and homocysteine-MTHFR C677T.**  
doi:10.1371/journal.pone.0057305.g002

### Biochemical Marker of Risk: MTHFR C677T and Homocysteine

Seven studies [22,26,57–61] that presented plasma homocysteine levels in South Asian ischemic stroke patients were subjected to a meta-analysis. One study [60] did not have normally distributed data with a threefold difference in mean homocysteine levels between cases and controls, and the study was excluded from the final analysis. The remaining six studies presented data as means

and standard deviations (SD) as well as median with their lower and higher range. A mean difference in homocysteine between ischemic stroke cases and controls was found to be 2.90  $\mu\text{mol/L}$  ( $p = 0.02$ , 95% CI 0.45–5.34) (Fig. 2) and this corresponded to an OR of 1.68 (95% CI 1.10–2.58), calculated using the CMA  $\nu 2.0$  software. We found 10 studies [62–71] linking homocysteine levels to the MTHFR 677 wildtype (CC) and homozygous mutant (TT) genotypes in healthy South Asians and calculated a mean

difference ( $\Delta X$ ) of 5.18  $\mu\text{mol/L}$  ( $p = 0.001$ , 95% CI 2.03–8.34) between the TT vs. CC genotypes in healthy individuals (Fig. 2). The expected odds ratio was calculated using the following formula:

$$\text{Expected OR} = 1.68^{5.18/2.90}$$

where 1.68 = OR associated with 2.90  $\mu\text{mol/L}$  difference in homocysteine levels between stroke cases and controls

5.18 = 5.18  $\mu\text{mol/L}$  mean difference in homocysteine levels between the MTHFR TT vs. CC genotypes in healthy individuals

2.90 = Mean difference of 2.90  $\mu\text{mol/L}$  associated with an OR of 1.68

The above calculation revealed an expected OR of 2.52. A 95% confidence interval for the logged odds ratio of 0.92 (LN 2.52) was obtained as 1.96 standard errors on either side of the point estimate as previously described by Bland et al [72]. Standard error (SE) was calculated as the square root of the sum of reciprocals of the frequencies i.e. number of cases and controls. The 95% confidence interval was calculated using the formula LN (OR)  $\pm 1.96 \times \text{SE}$ , which generated a range of  $-1.17$  to  $3.02$ . The antilog of these limits generated a 95% confidence interval for OR 2.52 as  $\text{Exp}(-1.17) = 0.31$  to  $\text{Exp}(3.02) = 20.65$ . A meta-analysis relating MTHFR 677/TT vs. CC risk genotype to disease in our current study generated a pooled observed OR of 2.50 (95% CI, 0.89–6.97). The observed OR was close to the expected OR and its 95% confidence interval fell entirely within the confidence interval for the expected OR (Fig. 3).

## Discussion

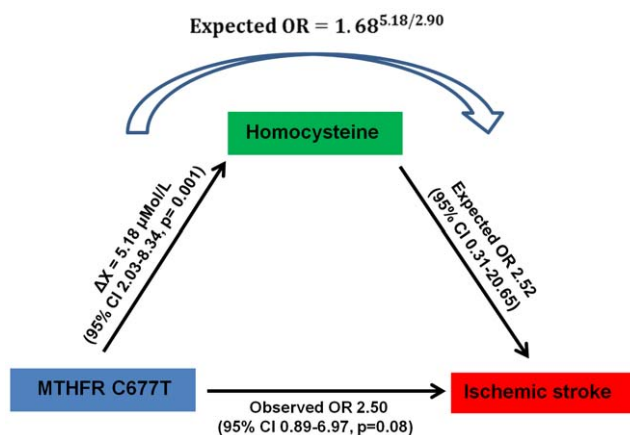
We present, to the best of our knowledge, the most comprehensive genetic meta-analysis of ischemic stroke in South Asians from India, Pakistan, Sri Lanka and Bangladesh. Our findings suggest three genes (PDE4D, ACE I/D and IL10) to have statistically significant ORs for risk of ischemic stroke in a South Asian adult population and the totality of data supported MTHFR C677T as a further likely risk factor. Results from 12 other gene

variants, although significant, were of insufficient power to allow robust conclusions. The remaining 17 gene polymorphisms (Table S1) failed to support any significant association either due to lack of sufficient studies, paucity of subjects studied and/or heterogeneity in the stroke sub-types investigated.

The two most significant risk associations, PDE4D SNP 83 and ACE I/D identified in our meta-analysis, were previously described as genetic risk factors for stroke in other ethnic groups [73]. We observed a doubling in the odds (OR 2.20; 95% CI, 1.21–3.99) or a  $\sim 120\%$  increase in the risk of developing stroke per copy of the risk allele of PDE4D SNP 83 for overall ischemic stroke. PDE4D gene encodes a phosphodiesterase enzyme that regulates cAMP levels in the body [74] and was first identified as a candidate gene for stroke by the DeCODE study [75]. However several attempts to replicate these findings failed [76], while some studies reported conflicting results [14,77,78]. These discrepancies have been attributed to possible problems in study design and analytical methods [79]. A recent meta-analysis by Yoon et al [80] identified SNP 83 as having a protective association with stroke in Asians (OR 0.79, 95% CI 0.69–0.90;  $p = 0.0005$ ). The meta-analysis combined 4 studies of South East Asian origin and 2 studies of South Asian origin. While combining different ethnic populations for the analysis maybe a reasonable approach for an exploratory analysis of genetic risk factors, such assumptions may lead to fallacious results as studies have already proved that India has a unique genetic population sub-structure which cannot be imputed from other ethnic groups [81]. Other studies have also identified race-ethnic disparities for stroke risk factors such as hypertension and diabetes [82] as well as genes affecting stroke in different ethnic groups [83]. Our study included individuals of only South Asian ancestry from the Indian sub-continent and though the population size is small, it takes into account the need for independent analysis of South Asians.

ACE gene plays an important role in vascular physiology and structural integrity. Our study finds a high risk of ACE/DD variant in South Asians (OR 5.00; 95% CI, 1.17–21.37) which accounts for a five-fold increase in the risk of overall ischemic stroke per copy of the risk allele of the ACE DD variant relative to Europeans (OR 1.15; 95% CI, 1.06–1.25) [14]. Closer examination of this comparison revealed a very large confidence interval for our study largely due to low sample size (355 cases/338 controls vs. 4897 cases/13949 controls in Europeans) suggesting that the measured effect size is probably inflated. Partially overlapping confidence intervals further suggested that there may not be a statistical difference between the two groups. An alternate explanation could be the difference in the prevalent stroke sub type in South Asians as compared to other ethnic groups. The homozygous *D* allele is associated with hypertension in Asian Indians [84] and is also associated with preferential risk of small vessel disease (SVD) [85] which is the most common subtype of stroke found in South Asians [12,86]. The polymorphism accounts for 47% of the total phenotypic variance of serum ACE [87] which is linked with quantitative levels of risk of disease [14] [47].

Interleukin 10 is an anti-inflammatory cytokine produced primarily by monocytes and type 2 T helper cells. We found IL10 to be associated with a 44% increase (OR 1.44; 95% CI 1.09–1.91) in the risk of overall ischemic stroke per copy of the risk allele of the IL 10 G1082A variant. IL10 is involved in various cellular processes such as inhibition of pro-inflammatory cytokines, suppression of antigen-presenting capacity of antigen presenting cells (APC) and stimulation of B cell maturation. IL10 forms part of an inflammatory genetic profile and elevated levels of IL10 post stroke have been implicated in severe neurological impairment



**Figure 3. Mendelian Randomization to compare estimated risk with observed risk for gene polymorphisms associated with ischemic stroke.** The mean difference ( $\Delta X$ ) was calculated from a meta-analysis relating MTHFR C677T genotype with homocysteine variation in healthy South Asians. The expected OR was calculated using the following formula:  $\text{Expected OR} = 1.68^{5.18/2.90}$  [16] where 1.68 was the OR associated with 2.90  $\mu\text{mol/L}$  mean difference in homocysteine levels between stroke cases and controls. doi:10.1371/journal.pone.0057305.g003

and major adverse clinical outcomes [88,89]. Since acute ischemic brain insult is known to trigger anti-inflammatories, mutations in IL10 may result in uninhibited effects of pro-inflammatory cytokines.

MTHFR C677T genotype is associated with hyperhomocystinemia among South Asians and Europeans [90] and has been shown to have a larger effect on homocysteine concentration in geographical regions of low folate consumption than in regions with high dietary folate intake [91]. Knowing whether there is an ethnic predisposition to hyperhomocystinemia is necessary as homocysteine levels can be effectively lowered by supplementing diet with folic acid, vitamin B12 and vitamin B6. This is important as the majority Hindu South Asian population has low plasma folate and Vitamin B12 levels [92,93] which can be partly accounted for by their predominantly vegetarian diets or high temperature cooking methods which destroy folate. We sought to establish whether the putative biochemical intermediary, homocysteine, of the MTHFR C677T gene variant was associated with equivalent quantitative levels of risk [16] in South Asians. We compared effect sizes of gene-stroke associations with those predicted from independent biochemical data using a mendelian randomization strategy [14].

Although the genetic association between disease and gene for MTHFR C677T was weak ( $p=0.08$ ), the totality of data supported such an association and with its previously documented strong association in Europeans [94] led us to undertake this causal analysis. A mendelian randomization strategy for a causal association, was concordant with literature on Europeans [14,16] with similar observed and expected odds ratios. Our study also showed a significant association ( $p=0.02$ ) between homocysteine levels and ischemic stroke where a 2.9  $\mu\text{Mol/L}$  elevation in plasma homocysteine levels increased the odds of stroke risk by 68%. This is in contrast to reports from European studies which have observed much lower odds of risk (24%) for a similar increment in plasma homocysteine levels [20]. Interestingly, the biochemical change C  $\rightarrow$  T also showed a high difference in homocysteine levels,  $\Delta X$  value of 5.18  $\mu\text{Mol/L}$ , for South Asians, compared to Europeans  $\Delta X$  of  $\sim 2.3$   $\mu\text{Mol/L}$  [14]. While it is possible this is an artifactual difference, another explanation is that healthy South Asians with CC genotype have overall higher levels of homocysteine than healthy CC genotype Caucasians resulting in a greater  $\Delta X$ . Alternately, our study population maybe largely vegetarian and resultantly have high homocysteine levels [95]. Epidemiological studies [93] have already established ethnic differences in plasma homocysteine levels and these need to be followed up with comparative analysis of homocysteine levels in healthy South Asians and Europeans with MTHFR C677C genotypes.

As with all meta-analyses, a number of limitations to interpreting our data need to be reported. Possible publication bias [96] could be introduced by reporting only positive results published in the English language and unconscious exclusion of studies which may have reported contradictory results. Studies were smaller and much fewer in numbers leading to less robust conclusions compared to similar work in Europeans [97]. As an example, for ACE I/D polymorphism the availability of only 2 studies for the meta-analysis prevented us from conducting an iterative analysis to identify and remove the source of inter study heterogeneity ( $[P_{\text{HET}}]=0.002$ ,  $I^2=90\%$ ). The small sample size meant large a confidence interval and less statistical reliability of data. Lack of stroke TOAST classification amongst the selected studies also hampered sub-group analysis which is essential due to the heterogeneous etiology of ischemic stroke sub-types. Fewer studies also limited our ability to take the genetic stratification of

the South Asian population [81] into account and we conducted a pooled analysis of both South and North Indian stroke cases.

Another major limitation was the inability to calculate the F statistic to determine the strength of the genetic instrumental variable (IV). A mendelian randomization strategy on individual patient level data involves the assessment of a causal effect of a phenotype on an outcome by using genetic IV [98]. The magnitude of bias is determined by the F statistic for the strength of association between the IV and phenotypes and is essential to assess the appropriateness of the IV. In a literature based meta-analysis pooled F statistics can be calculated if each study included in the meta-analysis reports the study specific values. However none of the studies included in our meta-analysis reported the study specific F statistics.

In the current climate of GWAS studies and whole genome sequencing, the literature based meta-analysis has its own unique utility. Where large GWAS studies have failed to identify risk associations with gene variants such as *MTHFR* and *ACE*, literature based meta-analysis of comparable (or greater) power and sample size have been able to identify these genetic risk variants to be associated with stroke [14,15]. Although the stand alone validity of individual candidate gene based studies remains inconclusive, a meta-analysis may reveal a true association. Our meta-analysis included studies that examined candidate genes previously identified in other vascular disorders such as MI in Europeans and South East Asians. Though this is a reasonable exploratory approach to understanding the genetics of stroke in South Asians, there is a clear need for larger GWAS strategies to truly understand the genetic underpinnings of stroke. A good starting point could be replication in an independent South Asian stroke population of gene variants identified in large statistically powered GWAS studies conducted in Europeans. The Wellcome Trust Case Control Consortium 2 (WTCCC2) and the International Stroke Genetics Consortium (ISGC) have successfully identified novel genes associated with stroke risk (*PITX2*, *ZFXH3*, 9p21 locus and *HDAC9*) [99] while other GWAS studies have highlighted the subtype specific nature of these genetic effects [100]. The recent METASTROKE [100,101] meta-analysis which included 15 stroke cohorts comprising of 12,000 cases and 60,000 controls validated these findings but failed to identify any new genetic risk variants. Replication of SNPs from related cardiovascular GWAS studies found one novel association with gene *PHACTR1* which suggests that detection of any new genetic risk variants will rest on proper stroke subtyping. Though the effect sizes for these genes are small, it is likely that they are true associations as compared to genes identified in smaller underpowered candidate gene based studies.

Our work supports a genetic etiology of ischemic stroke in South Asians but the dataset is considerably smaller compared to those of European descent. We show no major differences in risk associations for previously studied stroke susceptibility genes between South Asians, Europeans and South East Asians. However it would be fallacious to assume a literal comparison between studies that are not statistically at power and hence there is a clear need for large prospective well powered GWAS studies in South Asians, as has been done in Europeans.

## Supporting Information

**Table S1 Summary table of gene polymorphisms associated with risk of ischemic stroke in South Asians. (DOCX)**

## Author Contributions

Conceived and designed the experiments: PS. Performed the experiments: SY NH TM. Analyzed the data: SY NH PB. Contributed reagents/

materials/analysis tools: MSK KP. Wrote the paper: SY PS PB NH TM KP MSK.

## References

- Ezzati M, Riboli E (2012) Can noncommunicable diseases be prevented? Lessons from studies of populations and individuals. *Science* 337: 1482–1487.
- Government of India (2008) Central Bureau of Health Intelligence, National Health Profile 2008. New Delhi: Directorate General of Health Sciences. 102.
- Banerjee TK (2006) Epidemiology of stroke in India. *Neurology Asia* 11: 1–4.
- Saha S (2003) Epidemiological study of neurological disorders in a rural population of Eastern India. *J Indian Med Assoc* 101: 299–304.
- TK B (2001) Stroke in the urban population of Calcutta: an epidemiological study. *Neuroepidemiology* 20: 201–207.
- Gourie-Devi M (2004) Prevalence of neurological disorders in Bangalore, India: a community-based study with comparison between urban and rural areas. *Neuroepidemiology* 23: 261–268.
- Das SK (2007) A Prospective Community-Based Study of Stroke in Kolkata, India. *Stroke* 38: 906–910.
- Rao S (1971) Aspects of epidemiology of stroke in south India. Proceedings of the First All India Conference on Stroke: 25–32.
- Bhattacharya S (2005) A 5 years prospective study of incidence, morbidity and mortality profile of stroke in a rural community of eastern India. *J Indian Med Assoc* 103: 655–659.
- Potluri R (2009) Increasing prevalence of haemorrhagic stroke among South Asian patients in the United Kingdom from 1997 to 2005. *Journal of Clinical Neuroscience* 16: 605–605.
- Baweja G, Nanda NC, Parikh N, Bhatia V, Venkataraman R (2004) Prevalence of stroke and associated risk factors in Asian Indians living in the state of Georgia, United States of America. *Am J Cardio* 93: 267–269.
- Biswas M (2009) Etiology and risk factors of ischemic stroke in Indian-American patients from a hospital-based registry in New Jersey, USA. *Neurology Asia* 14: 81–86.
- O'Donnell MJ, Xavier D, Liu L, Zhang H, Chin SL, et al. (2010) Risk factors for ischaemic and intracerebral haemorrhagic stroke in 22 countries (the INTERSTROKE study): a case-control study. *Lancet* 376: 112–123.
- Bentley P, Peck G, Smeeth L, Whittaker J, Sharma P (2010) Causal relationship of susceptibility genes to ischemic stroke: comparison to ischemic heart disease and biochemical determinants. *PLoS One* 5: e9136.
- Bevan S, Traylor M, Adib-Samii P, Malik R, Paul NL, et al. (2012) Genetic heritability of ischemic stroke and the contribution of previously reported candidate gene and genome-wide associations. *Stroke* 43: 3161–3167.
- Casas JP, Bautista LE, Smeeth L, Sharma P, Hingorani AD (2005) Homocysteine and stroke: evidence on a causal link from mendelian randomisation. *Lancet* 365: 224–232.
- DerSimonian R, Laird N (1986) Meta-analysis in clinical trials. *Control Clin Trials* 7: 177–188.
- Deeks JJ, Altman DG, Bradburn MJ (2008) Statistical Methods for Examining Heterogeneity and Combining Results from Several Studies in Meta-Analysis, in *Systematic Reviews in Health Care: Meta-Analysis in Context*. Egger M, Smith GD, Altman DG, editors. BMJ Publishing Group, London, UK Second Edition.
- Egger M, Davey Smith G, Schneider M, Minder C (1997) Bias in meta-analysis detected by a simple, graphical test. *BMJ* 315: 629–634.
- Wald DS, Law M, Morris J (2003) Serum homocysteine and the severity of coronary artery disease. *Thromb Res* 111: 55–57.
- Hozo SP, Djulbegovic B, Hozo I (2005) Estimating the mean and variance from the median, range, and the size of a sample. *BMC Med Res Methodol* 5: 13.
- Biswas A (2009) Homocysteine levels, polymorphisms and the risk of ischemic stroke in young Asian Indians. *J of stroke and CVD* 18: 103–110.
- Alluri (2005) MTHFR C677T gene mutation as a risk factor for arterial stroke: a hospital based study. *Eur J of Neurology* 12: 40–44.
- Panigrahi I (2006) Role of MTHFR C677T polymorphism in ischemic stroke. *Neurology India* 54.
- Somarajan BI, Kalita J, Mittal B, Misra UK (2011) Evaluation of MTHFR C677T polymorphism in ischemic and hemorrhagic stroke patients. A case-control study in a Northern Indian population. *J Neurol Sci* 15: 67–70.
- Mejia Mohamed EH, Tan KS, Ali JM, Mohamed Z (2011) TT genotype of the methylenetetrahydrofolate reductase C677T polymorphism is an important determinant for homocysteine levels in multi-ethnic Malaysian ischaemic stroke patients. *Ann Acad Med Singapore* 40: 186–186.
- Saleheen D (2005) Association of PDE4D gene with ischemic stroke in a Pakistani population. *Stroke* 36: 2270–2277.
- Munshi A (2009) Phosphodiesterase 4D (PDE4D) gene variants and the risk of ischemic stroke in South Indian population. *J of Neurological Sciences* 285: 142–145.
- Banerjee I (2008) Inflammatory system gene polymorphism and the risk of stroke: a case-control study in an Indian population. *Brain Res Bull* 75: 158–165.
- Majumdar V (2009) Association of ENOS gene polymorphisms with early-onset stroke in South Indians. *J of Atherosclerosis and thrombosis* 17.
- Munshi A (2010) VNTR polymorphism in intron 4 of the eNOS gene and the risk of ischemic stroke in a South Asian population. *Brain Res Bull*.
- Munshi A (2008) Angiotensin-converting enzyme Insertion/Deletion polymorphism and the risk of ischemic stroke in a south indian population. *J of Neurological Sciences* 272: 1320–1335.
- Kalita J, Somarajan BI, Kumar B, Mittal B, Misra UK (2011) A study of ACE and ADD1 polymorphism in ischemic and hemorrhagic stroke. *Clin Chim Acta* 412: 642–646.
- Luthra K, Prasad K, Kumar P, Dwivedi M, Pandey RM, et al. (2002) Apolipoprotein E gene polymorphism in cerebrovascular disease: a case-control study. *Clin Genet* 62: 39–44.
- Chowdhury AH, Yokoyama T, Kokubo Y, Zaman MM, Haque A, et al. (2001) Apolipoprotein E genetic polymorphism and stroke subtypes in a Bangladeshi hospital-based study. *J Epidemiol* 11: 131–138.
- Dindagur N, Kruthika-Vinod TP, Christopher R (2007) Factor V gene A4070G mutation and the risk of cerebral veno-sinus thrombosis occurring during puerperium. *Thromb Res* 119: 497–500.
- Dindagur N, Kruthika-Vinod TP, Christopher R (2006) Thrombophilic gene polymorphisms in puerperal cerebral veno-sinus thrombosis. *J Neurol Sci* 249: 25–30.
- Saleheen D, Bukhari S, Haider SR, Nazir A, Khanum S, et al. (2005) Association of phosphodiesterase 4D gene with ischemic stroke in a Pakistani population. *Stroke* 36: 2275–2277.
- Munshi A, Babu MS, Kaul S, Shafi G, Anila AN, et al. (2009) Phosphodiesterase 4D (PDE4D) gene variants and the risk of ischemic stroke in a South Indian population. *J Neurol Sci* 285: 142–145.
- Munshi A, Rajeshwar K, Kaul S, Al-Hazzani A, Alshatwi AA, et al. (2010) Interleukin-10-1082 promoter polymorphism and ischemic stroke risk in a South Indian population. *Cytokine* 52: 221–224.
- Sultana S, Jeedigunta Y, Penagaluru PK, Joshi S, Rani PU, et al. (2011) Tumour necrosis factor alpha and interleukin 10 gene polymorphisms and the risk of ischemic stroke in south Indian population. *J Genet* 90: 361–364.
- Biswas A, Ranjan R, Meena A, Akhter MS, Yadav BK, et al. (2009) Homocysteine levels, polymorphisms and the risk of ischemic stroke in young Asian Indians. *J Stroke Cerebrovasc Dis* 18: 103–110.
- Biswas A, Tiwari AK, Ranjan R, Meena A, Akhter MS, et al. (2009) Prothrombotic polymorphisms, mutations, and their association with pediatric non-cardioembolic stroke in Asian-Indian patients. *Ann Hematol* 88: 473–478.
- Munshi A, Aliya N, Jyothy A, Kaul S, Alladi S, et al. (2009) Prothrombin gene G20210A mutation is not a risk factor for ischemic stroke in a South Indian Hyderabad Population. *Thromb Res* 124: 245–247.
- Banerjee I, Gupta V, Ahmed T, Faizaan M, Agarwal P, et al. (2008) Inflammatory system gene polymorphism and the risk of stroke: a case-control study in an Indian population. *Brain Res Bull* 75: 158–165.
- Munshi A, Sharma V, Kaul S, Rajeshwar K, Babu MS, et al. (2010) Association of the -344C/T aldosterone synthase (CYP11B2) gene variant with hypertension and stroke. *J Neurol Sci* 296: 34–38.
- Munshi A, Sharma V, Kaul S, Al-Hazzani A, Alshatwi AA, et al. (2010) Estrogen receptor alpha genetic variants and the risk of stroke in a South Indian population from Andhra Pradesh. *Clin Chim Acta* 411: 1817–1821.
- Munshi A, Rajeshwar K, Kaul S, Al-Hazzani A, Alshatwi AA, et al. (2011) Association of tumor necrosis factor-alpha and matrix metalloproteinase-3 gene variants with stroke. *Eur J Neurol* 18: 1053–1059.
- Munshi A, Sharma V, Kaul S, Al-Hazzani A, Alshatwi AA, et al. (2012) Association of 1347 G/A cytochrome P450 4F2 (CYP4F2) gene variant with hypertension and stroke. *Mol Biol Rep* 39: 1677–1682.
- Sharma V, Kaul S, Al-Hazzani A, Prabha TS, Rao PP, et al. (2011) Association of C3435T multi drug resistance gene-1 polymorphism with aspirin resistance in ischemic stroke and its subtypes. *J Neurol Sci*.
- Babu MS, Prabha TS, Kaul S, Al-Hazzani A, Shafi G, et al. (2012) Association of genetic variants of fibrinolytic system with stroke and stroke subtypes. *Gene*.
- Chandra G, Pal S, Gangopadhyay PK (2006) Cystathionine beta-synthase T833C/844ins68 polymorphism and stroke. *Neurology India* 54: 446.
- Majumdar V, Nagaraja D, Christopher R (2010) Association of the functional KL-VS variant of Klotho gene with early-onset ischemic stroke. *Biochem Biophys Res Commun* 403: 412–416.
- Kirti Kain JY, Bavington J, Bamford J, Catto AJ (2005) Coagulation factor XIII Bsubunit antigen, FXIIIVal34Leu genotype and ischaemic stroke in South Asians. *Thrombosis and Haemostasis* 93: 394–395.
- Somarajan BI, Kalita J, Misra UK, Mittal B (2010) A study of alpha1 antichymotrypsin gene polymorphism in Indian stroke patients. *J Neurol Sci* 290: 57–59.

56. Kalita J, Srivastava R, Bansal V, Agarwal S, Misra UK (2006) Methylene-tetrahydrofolate reductase gene polymorphism in Indian stroke patients. *Neurol India* 54: 260–263.
57. Narang AP (2009) Homocysteine-Risk Factor for ischemic stroke. *Indian J Physiol Pharmacol* 53.
58. Modi M (2005) Hyperhomocysteinemia as a risk factor for ischemic stroke: An Indian scenario. *Neurology India* 53: 297–301.
59. Kalita J (2009) Relationship of homocysteine with other risk factors and outcome of ischemic stroke. *Clinical Neurology and Neurosurgery* 111: 364–367.
60. Dhamija RK (2009) Homocysteine and Lipoprotein (a) correlation in Ischemic stroke patients. *Journal of the Neurological Sciences* 281: 64–68.
61. Mamatha SN, Nagaraja D, Philip M, Christopher R (2011) Asymmetric dimethylarginine as a risk marker for early-onset ischemic stroke in Indian population. *Clin Chim Acta* 412: 139–142.
62. Dhillon VS, Shahid M, Husain SA (2007) Associations of MTHFR DNMT3b 4977 bp deletion in mtDNA and GSTM1 deletion, and aberrant CpG island hypermethylation of GSTM1 in non-obstructive infertility in Indian men. *Mol Hum Reprod* 13: 213–222.
63. Chambers JC, Thompson E, Reilly P, Obeid OA, Refsum H, et al. (2000) Methylene-tetrahydrofolate Reductase 677 C>T Mutation and Coronary Heart Disease Risk in UK Indian Asians. *Arterioscler Thromb Vasc Biol* 20: 2448–2452.
64. Kumar J, Das SK, Sharma P, Karthikeyan G, Ramakrishnan L, et al. (2005) Homocysteine levels are associated with MTHFR A1298C polymorphism in Indian population. *J Hum Genet* 50: 655–663.
65. Liew SC, Das-Gupta E, Wong SF, Lee N, Safdar N, et al. (2012) Association of Methylene-tetrahydrofolate Reductase (MTHFR) 677 C>T gene polymorphism and homocysteine levels in psoriasis vulgaris patients from Malaysia: a case-control study. *Nutr J* 11: 1.
66. Perwaiz Iqbal M, Fatima T, Parveen S, Yousuf FA, Shafiq M, et al. (2005) Lack of association of methylene-tetrahydrofolate reductase 677C>T mutation with coronary artery disease in a Pakistani population. *Journal of Molecular and Genetic Medicine* 1: 26–32.
67. Misra UK, Kalita J, Srivastava AK, Agarwal S (2010) MTHFR gene polymorphism and its relationship with plasma homocysteine and folate in a North Indian population. *Biochem Genet* 48: 229–235.
68. Rajneesh Tripathi ST, Singh PK, Agarwal S (2010) Association of homocysteine and methylene tetrahydrofolate reductase (MTHFR C677T) gene polymorphism with coronary artery disease (CAD) in the population of North India. *Genetics and Molecular Biology* 33: 224–228.
69. Shazia Micheal AA, Qamar R, Akhtar F, Khan MI, Khan WA (2009) MTHFR gene C677T and A1298C polymorphisms and homocysteine levels in primary open angle and primary closed angle glaucoma. *Molecular Vision* 15: 2268–2278.
70. Sukla KK, Raman R (2012) Association of MTHFR and RFC1 gene polymorphism with hyperhomocysteinemia and its modulation by vitamin B12 and folic acid in an Indian population. *Eur J Clin Nutr* 66: 111–118.
71. Radha Rama DA, Govindaiah V, Ramakrishna G, Naushad SM (2004) Prevalence of methylene tetrahydrofolate reductase polymorphism in South Indian population. *Curr Sci* 86.
72. Bland JM, Altman DG (2000) Statistics notes. The odds ratio. *BMJ* 320: 1468.
73. 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: 327–333.
74. Conti M, Richter W, Mehats C, Livera G, Park JY, et al. (2003) Cyclic AMP-specific PDE4 phosphodiesterases as critical components of cyclic AMP signaling. *J Biol Chem* 278: 5493–5496.
75. Gretarsdottir S, Thorleifsson G, Reynisdottir ST, Manolescu A, Jonsdottir S, et al. (2003) The gene encoding phosphodiesterase 4D confers risk of ischemic stroke. *Nat Genet* 35: 131–138.
76. Rosand J, Bayley N, Rost N, de Bakker PI (2006) Many hypotheses but no replication for the association between PDE4D and stroke. *Nat Genet* 38: 1091–1092; author reply 1092–1093.
77. Bevan S, Dichgans M, Gschwendtner A, Kuhlenbaumer G, Ringelstein EB (2008) Variation in the PDE4D Gene and Ischemic Stroke Risk. A Systematic Review and Metaanalysis on 5200 Cases and 6600 Controls. *Stroke* 39: 1966–1971.
78. Matsushita T, Kubo M, Yonemoto K, Ninomiya T, Ashikawa K, et al. (2009) Lack of association between variations of PDE4D and ischemic stroke in the Japanese population. *Stroke* 40: 1245–1251.
79. Worrall BB, Mychaleckyj JC (2006) PDE4D and stroke: a real advance or a case of the Emperor's new clothes? *Stroke* 37: 1955–1957.
80. Yoon D, Park SK, Kang D, Park T, Park JW (2011) Meta-Analysis of Homogeneous Subgroups Reveals Association between PDE4D Gene Variants and Ischemic Stroke. *Neuroepidemiology* 36: 213–222.
81. Reich D, Thangaraj K, Patterson N, Price AL, Singh L (2009) Reconstructing Indian population history. *Nature* 461: 489–494.
82. Sacco RL (2001) Race-Ethnic Disparities in the Impact of Stroke Risk Factors, The Northern Manhattan Stroke Study. *Stroke* 32:1725.
83. Hall S (2000) Frequency and allelic association of common variants in the lipoprotein lipase gene in different ethnic groups: The Wandsworth Heart and Stroke Study. *Genetic Epidemiology* 18: 203–216.
84. Tao HM, Shao B, Chen GZ (2009) Meta-analysis of the ACE gene polymorphism in cerebral infarction. *Can J Neurol Sci* 36: 20–25.
85. Rao R (2009) Ischaemic Stroke Subtypes and Their Genetic Basis: A Comprehensive Meta- Analysis of Small and Large Vessel Stroke. *Eur Neurol* 61: 76–86.
86. Gunarathne A (2009) Ischemic Stroke in South Asians A Review of the Epidemiology, Pathophysiology, and Ethnicity-Related Clinical Features. *Stroke* 40: e415–e423.
87. Rigat B, Hubert C, Alhenc-Gelas F, Cambien F, Corvol P, et al. (1990) An insertion/deletion polymorphism in the angiotensin I-converting enzyme gene accounting for half the variance of serum enzyme levels. *J Clin Invest* 86: 1343–1346.
88. Chang LT, Yuen CM, Liou CW, Lu CH, Chang WN, et al. (2010) Link between interleukin-10 level and outcome after ischemic stroke. *Neuroimmunomodulation* 17: 223–228.
89. Dzedzic T, Bartus S, Klimkowicz A, Motyl M, Slowik A, et al. (2002) Intracerebral hemorrhage triggers interleukin-6 and interleukin-10 release in blood. *Stroke* 33: 2334–2335.
90. Kelemen LE (2004) Associations of plasma homocysteine and the methylene-tetrahydrofolate reductase C677T polymorphism with carotid intima media thickness among South Asian, Chinese and European Canadians. *Atherosclerosis* 176: 361–370.
91. Holmes MV, Newcombe P, Hubacek JA, Sofat R, Ricketts SL, et al. (2011) Effect modification by population dietary folate on the association between MTHFR genotype, homocysteine, and stroke risk: a meta-analysis of genetic studies and randomised trials. *Lancet* 378: 584–594.
92. Hughes K, Ong CN (2000) Homocysteine, folate, vitamin B12, and cardiovascular risk in Indians, Malays, and Chinese in Singapore. *J Epidemiol Community Health* 54: 31–34.
93. Chandalia M, Abate N, Cabo-Chan AV Jr, Devaraj S, Jialal I, et al. (2003) Hyperhomocysteinemia in Asian Indians living in the United States. *J Clin Endocrinol Metab* 88: 1089–1095.
94. Casas JP, Hingorani AD, Bautista LE, Sharma P (2004) Meta-analysis of genetic studies in ischemic stroke: thirty-two genes involving approximately 18,000 cases and 58,000 controls. *Arch Neurol* 61: 1652–1661.
95. Bissoli L, Di Francesco V, Ballarin A, Mandragona R, Trespido R, et al. (2002) Effect of vegetarian diet on homocysteine levels. *Ann Nutr Metab* 46: 73–79.
96. Lin BK, Clyne M, Walsh M, Gomez O, Yu W, et al. (2006) Tracking the epidemiology of human genes in the literature: the HuGE Published Literature database. *Am J Epidemiol* 164: 1–4.
97. Ioannidis JP, Panagiotou OA (2011) Comparison of effect sizes associated with biomarkers reported in highly cited individual articles and in subsequent meta-analyses. *JAMA* 305: 2200–2210.
98. Burgess S, Thompson SG (2011) Avoiding bias from weak instruments in Mendelian randomization studies. *Int J Epidemiol* 40: 755–764.
99. International Stroke Genetics C, Wellcome Trust Case Control C, Bellenguez C, Bevan S, Gschwendtner A, et al. (2012) Genome-wide association study identifies a variant in HDAC9 associated with large vessel ischemic stroke. *Nat Genet* 44: 328–333.
100. Holliday EG, Maguire JM, Evans TJ, Koblar SA, Jannes J, et al. (2012) Common variants at 6p21.1 are associated with large artery atherosclerotic stroke. *Nat Genet* 44: 1147–1151.
101. Traylor M, Farrall M, Holliday EG, Sudlow C, Hopewell JC, et al. (2012) Genetic risk factors for ischaemic stroke and its subtypes (the METASTROKE Collaboration): a meta-analysis of genome-wide association studies. *Lancet Neurol* 11: 951–962.