



Contents lists available at ScienceDirect

# Atherosclerosis

journal homepage: [www.elsevier.com/locate/atherosclerosis](http://www.elsevier.com/locate/atherosclerosis)



Short communication

## Replication of the *CELSR1* association with ischemic stroke in a Portuguese case-control cohort

Liliana O. Gouveia<sup>a,b</sup>, João Sobral<sup>a,c,d,e</sup>, Astrid M. Vicente<sup>c,d,e</sup>, José M. Ferro<sup>a,b</sup>, Sofia A. Oliveira<sup>a,c,\*</sup>

<sup>a</sup> Clinical Neurology Research Unit, Instituto de Medicina Molecular, Faculdade de Medicina da Universidade de Lisboa, Lisboa, Portugal

<sup>b</sup> Serviço de Neurologia, Hospital de Santa Maria, Lisboa, Portugal

<sup>c</sup> Instituto Gulbenkian de Ciência, Oeiras, Portugal

<sup>d</sup> Departamento Promoção da Saúde e Doenças Crónicas, Instituto Nacional de Saúde Dr. Ricardo Jorge, Lisboa, Portugal

<sup>e</sup> Center for Biodiversity, Functional & Integrative Genomics (BIOFIG), Lisboa, Portugal

### ARTICLE INFO

#### Article history:

Received 5 January 2011  
Received in revised form 11 March 2011  
Accepted 16 March 2011  
Available online xxx

#### Keywords:

CELSR1  
Association  
Ischemic stroke  
GWAS

### ABSTRACT

**Objectives:** Replication of GWAS association findings remains the gold standard for results validation. Our aim was to test the association of four polymorphisms (rs1671021 in *LLGL2*, rs753307 in *RUVBL2*, rs6007897 and rs4044210 in *CELSR1*) previously identified as ischemic stroke (IS) risk factors in a phased GWAS performed on 6341 Japanese individuals [1].

**Methods:** These polymorphisms were genotyped in a Portuguese sample of 566 IS cases and 525 controls, and their allele, genotype and haplotype associations were assessed.

**Results:** rs6007897 and rs4044210 in *CELSR1* were associated with stroke risk individually (OR[95%CI] = 1.43[1.13–1.81],  $p = 0.003$  and 1.38[1.09–1.74],  $p = 0.007$ , respectively), and in combination as a haplotype. These associations remain after correction for multiple testing and in a meta-analysis with the original findings. The other polymorphisms were not associated.

**Conclusions:** Our study independently confirmed for the first time the association between IS and *CELSR1*. This finding and the mechanisms by which these genetic variants exert their effects on stroke pathogenesis warrant further replication and investigation.

© 2011 Published by Elsevier Ireland Ltd.

### 1. Introduction

Stroke is the leading cause of disability worldwide and the third cause of mortality. It is a complex disease resulting from a multifactorial etiology. The classic risk factors rapidly gave rise to the understanding that the genetic background of an individual also influences the risk of stroke and the age in which it manifests. Genome-wide association studies (GWAS) are currently the preferred method for uncovering novel genetic risk variants and have brought novel insights into the biological and genetic underpinnings of many complex disorders. Replication of their most significant findings in independent cohorts remains the *sine qua non* for validating susceptibility genes.

Several stroke GWAS have been published to date [1–4] and have led to the identification of potential IS risk genes. The most recent stroke GWAS [1] involved a total of 6341 Japanese individuals from three independent datasets. This GWAS was performed in

subject panel A (131 IS cases and 135 controls) and 100 SNPs were selected for further examination in subject panel B (705 IS cases and 3,426 controls). rs1671021 (T→C, Phe479Leu) of *LLGL2* (lethal giant larvae homolog 2), rs9615362 of *CELSR1* (cadherin, epidermal growth factor laminin A G-type repeats seven-pass G-type receptor 1), and rs753307 of *RUVBL2* (RuvB-like 2) were significantly associated with IS in both panel A and B. Upon fine-mapping of linkage disequilibrium blocks containing these SNPs, rs1671021 of *LLGL2*, as well as rs6007897 (A→G, Thr2268Ala) and rs4044210 (A→G, Ile2107Val) of *CELSR1*, were found associated in the full dataset through multivariable logistic regression analysis.

Our goal was to validate these findings by testing in an independent sample the association with IS of SNPs rs1671021 in *LLGL2*, rs753307 in *RUVBL2*, rs6007897 and rs4044210 in *CELSR1*.

### 2. Materials and methods

The ascertainment and collection criteria for study participants, the DNA extraction and genotyping protocols, as well as the statistical analyses have been described in detail previously [5]. The research protocol was approved by the Ethics Committees of participating institutions. All participants were informed of the study and provided informed consent. Briefly, SNPs were genotyped

\* Corresponding author at: Instituto de Medicina Molecular, 1649-028 Lisboa, Portugal. Tel.: +351 217 999 411; fax: +351 217 999 412.

E-mail addresses: [aaoliveira@fm.ul.pt](mailto:aaoliveira@fm.ul.pt), [soliveira@igc.gulbenkian.pt](mailto:soliveira@igc.gulbenkian.pt) (S.A. Oliveira).

**Table 1**  
Characteristics of the study subjects. An unpaired Student's *t* test and a  $\chi^2$  test were used to compare quantitative and qualitative clinical and demographic data, respectively, between cases and controls.

Characteristic	Controls	Cases	<i>p</i>
No. of subjects	525	566	
Gender (males), <i>n/N</i> (%)	242/525 (46.1%)	361/566 (63.8%)	<10 <sup>-4</sup>
Age (mean $\pm$ SD, years)	62.9 $\pm$ 6.9	52.4 $\pm$ 9.3	<10 <sup>-4</sup>
Age at stroke onset (mean $\pm$ SD, years)	–	51.6 $\pm$ 9.3	–
Risk factors, <i>n/N</i> (%)			
Hypertension (>140–85 mmHg)	193/513 (37.6%)	289/507 (57.0%)	<10 <sup>-4</sup>
Diabetes	59/501 (11.8%)	95/539 (17.6%)	0.071
Hypercholesterolemia (>200 mg/dL)	328/520 (63.1%)	328/528 (62.1%)	0.992
Hypertriglyceridemia (>200 mg/dL)	68/438 (15.5%)	42/228 (18.4%)	0.823
Smoking	147/512 (28.7%)	271/557 (48.7%)	<10 <sup>-4</sup>
Drinking	218/505 (43.2%)	326/559 (58.3%)	<10 <sup>-4</sup>

using Sequenom's iPLEX assays (Sequenom, USA) following manufacturer's instructions. The primer sequences were designed using Sequenom's MassArray Assay Design 3.0 software and are indicated in [Supplementary Table 1](#). All genotype determinations were performed blinded to affection status, and an extensive quality control was performed (e.g. eight HapMap controls of diverse ethnic affiliation, sample duplication within and across plates, negative controls scattered on the plaques, Hardy-Weinberg equilibrium  $p > 0.05$  in controls, genotyping success rate >99%). Association results were considered significant below the conventional level of 0.05. Correction for multiple testing was performed by permutation analysis (10,000 SNP and haplotype permutations) in Haploview 4.1 [6]. A fixed-effects (Mantel-Haenszel) meta-analysis was performed using the *rmeta* package and the *meta.MH* command in R.2.7.2. Power analyses were performed using CaTS [7] with a disease prevalence of 2.6% [8] and a multiplicative genetic model.

### 3. Results

A total of 1091 participants were included with the main characteristics described in [Table 1](#). Gender and age were not considered in the multivariate logistic regression analyses because controls were selected to be older than the case subjects to minimize misclassification as "stroke-free". Gender was also not included because it is highly correlated with smoking in our dataset.

All four SNPs passed every genotyping quality control. While no association of the SNPs in *LLGL2* and *RUVBL2* with stroke risk was detected in any of the tests performed, our study confirmed that *CELSR1* is a susceptibility gene for stroke. SNPs rs6007897 and rs4044210 are associated with ischemic stroke in allelic, unadjusted and adjusted (for hypertension, diabetes, and ever smoking) genotypic tests of association ([Table 2](#)), with an odds ratio and 95%CI for the minor allele (A) of 0.76[0.61–0.94] and 0.78[0.63–0.96], respectively. As observed in the Japanese population ( $r^2 = 0.96$ ) [1], these two SNPs were in very high, but not complete, linkage disequilibrium (LD) in our dataset ( $r^2 = 0.91$ ). Furthermore, we also confirmed the association of the two most common haplotypes defined by these two polymorphisms ([Table 2](#)). The very strong LD among these two SNPs and the presence of only two major haplotypes formed by their combination indicates that it is not possible to distinguish which one is the causative SNP in these populations, but it may be possible in other populations (e.g.  $r^2 = 0.58$  in the HapMap YRI samples from Yoruba in Ibadan, Nigeria). The associations of rs6007897 and of the GG haplotype withstand correction for multiple testing by permutation analysis ([Table 2](#)). rs6007897 is also associated with stroke in codominant, dominant and recessive logistic regression models, adjusted and unadjusted for co-variables ([Supplementary Table 2](#)).

A combined analysis of our significant results and those of Yamada et al. [1] in their complete dataset reveals a protective effect of the major allele A of rs6007897 (com-

bined OR[95%CI] = 0.69[0.58–0.83], heterogeneity  $p = 0.072$ ) and rs4044210 (0.71[0.59–0.85], heterogeneity  $p = 0.067$ ) in *CELSR1* (the forest plots and summary tables are shown in [Supplementary Figure](#)).

### 4. Discussion

This report constitutes the first replication study of the Yamada et al. [1] GWAS and confirmed that *CELSR1* is a stroke risk factor in a Portuguese IS case-control dataset. Interestingly, even though the characteristics of the Japanese and Portuguese cohorts, as well as the rs6007897 and rs4044210 frequencies, are very different (e.g. G allele of rs6007897 has a frequency of 1.54% in Japanese [1] and 20.1% in Portuguese), the association is verified in both populations with the same allele and similar odds ratio. These findings lend further support to the true involvement of *CELSR1* in ischemic stroke, but additional large-scale independent replication efforts are required to firmly establish this genotype–phenotype correlation.

Power calculations show that our dataset had 71% power to detect the association we observed in the *CELSR1* SNPs (risk allele frequencies [RAF] = 0.18, OR = 1.3), but power amounted to 85% and 88% to detect effects with similar genetic magnitude of those reported in Yamada et al. [1] with our allele frequencies for rs1671021 in *LLGL2* (RAF = 0.62, OR = 1.3) and rs753307 in *RUVBL2* (RAF = 0.62, OR = 1.3), respectively. It is therefore unlikely that we did not detect an association in these last two markers due to lack of power, but we cannot exclude that other polymorphisms in these genes (e.g. rs1062708 which is in low LD with rs753307 [ $r^2 = 0.28$ ] and was associated with IS in the overall Japanese dataset [1]) are associated with IS in our dataset. In fact, rs1671021 and rs753307 were found associated with IS in the Japanese panel A and B, but only rs1671021 was reported to be associated in panel C and in the overall dataset [1]. The discrepancies between the original report and our results can be accounted for by several reasons, including chance of false-positive and -negative findings, differences in study design (e.g. older Japanese cohort, exclusion of cardioembolic stroke, cohort versus case-control design), unstudied causal variant, and genetic heterogeneity.

The modest effect size of these common variants on IS risk precludes their practical use in the prediction of individual risk but opens new avenues for the discovery of novel biological pathways underlying polygenic stroke. *CELSR1* is a large 35-exon gene mapping to chromosome 22q13.31 and encoding for several transcripts and multiple protein isoforms [9]. *Celsr1* is a cadherin with an atypical combination of seven-pass transmembrane domains, extracellular epidermal growth factor-like and laminin-G like domains crowned by nine protocadherin repeats. This unusual grouping of motifs generates a plasma membrane molecule capable of signaling after self-recognition in neighboring cell:cell interfaces or through heterophilic ligand/receptor inter-

**Table 2**  
SNP and haplotype association results. Significant p values are bolded.

Gene	SNP/haplotype	Genotype/alleles	Controls n (%)	Cases n (%)	HWE p (controls)	p (allele) Allele: OR [95% CI]	Permutation p	p (unadjusted) <sup>a</sup> Allele: OR [95% CI]	p (adjusted) <sup>b</sup> Allele: OR [95% CI]
LLGL2	rs1671021	CC	74 (14.2%)	97 (17.1%)	0.853	0.449	0.898	0.368	0.965
		CT	249 (47.7%)	258 (45.6%)					
		TT	199 (38.1%)	211 (37.3%)					
RUVBL2	rs753307	CC	128 (24.6%)	130 (23.0%)	0.430	0.795	1	0.862	0.123
		CT	251 (48.2%)	295 (52.1%)					
		TT	142 (27.3%)	141 (24.9%)					
CELSR1	rs6007897	GG	12 (2.3%)	27 (4.8%)	0.230	0.011	0.047	0.009	0.003
		AG	162 (31.0%)	197 (34.9%)					
		AA	348 (66.7%)	341 (60.4%)					
CELSR1	rs4044210	GG	15 (2.9%)	26 (4.6%)	0.559	0.021	0.094	0.019	0.007
		AG	162 (31.0%)	201 (35.5%)					
		AA	346 (66.2%)	339 (59.9%)					
CELSR1	rs6007897-rs4044210	AA	81.1%	77.1%	0.108	0.024	0.108	1.29 [1.04–1.60]	1.38 [1.09–1.74]
		GG	17.3%	21.6%					

<sup>a</sup> p-Value in the log-additive model.

<sup>b</sup> p-Value in the log-additive model adjusted for hypertension, diabetes, and ever smoking.

actions [9]. Celsr1 is a key component of the planar cell polarity pathway which directs the re-structuring of epithelial tissue during development of several tissues [10]. The rs4044210 SNP is located in the hormone receptor domain of Celsr1, an extracellular domain containing four conserved cysteine residues likely form disulfide bridges and which may function as a ligand-binding domain. The functional impact of non-synonymous rs6007897 (Thr2268Ala) and rs4044210 (Ile2107Val) SNPs in Celsr1 function are unknown.

It is theoretically possible that our association findings are caused by hidden population stratification. This is however unlikely given that the Portuguese population is fairly homogeneous and the association was only replicated in one out of the three genes tested. Furthermore, since we only genotyped two polymorphisms in *CELSR1*, we cannot exclude the possibility that other SNPs in that gene show a stronger association with IS, and therefore a more thorough association study of this gene should be undertaken. Additional genetic approaches for completing our understanding of *CELSR1*'s role in IS include the study of its structural variants (e.g. copy number variants), deep-sequencing of the whole gene and adjacent regions, a search for gene-gene and gene-environment interactions, and epigenetic modifications. In parallel, the integration of complementary data from multiple levels and sources (e.g. *CELSR1* proteomics, biochemical assays, *in vivo* models) may contribute to novel insights into stroke pathophysiology and ultimately lead to new diagnostic approaches and to the development of novel therapies in susceptible individuals.

### Acknowledgments

We are deeply grateful to all study participants and to the genotyping unit at the Instituto Gulbenkian de Ciência. This work was supported by the PTDC/SAU-GMG/64426/2006 grant and a Ciência 2008 contract (SAO) from the Portuguese Fundação para a Ciência e a Tecnologia (FCT), and by a fellowship from Fundação AstraZeneca/Faculdade de Medicina de Lisboa (LOG).

### Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.atherosclerosis.2011.03.022.

### References

- [1] Yamada Y, Fuku N, Tanaka M, et al. Identification of *CELSR1* as a susceptibility gene for ischemic stroke in Japanese individuals by a genome-wide association study. *Atherosclerosis* 2009;207:144–9.
- [2] Matarin M, Brown WM, Scholz S, et al. A genome-wide genotyping study in patients with ischaemic stroke: initial analysis and data release. *Lancet Neurol* 2007;6:414–20.
- [3] Kubo M, Hata J, Ninomiya T, et al. A nonsynonymous snp in *PRKCH* (protein kinase c eta) increases the risk of cerebral infarction. *Nat Genet* 2007;39:212–7.
- [4] Ikram MA, Seshadri S, Bis JC, et al. Genomewide association studies of stroke. *N Engl J Med* 2009;360:1718–28.
- [5] Krug T, Manso H, Gouveia L, et al. Kalirin: a novel genetic risk factor for ischemic stroke. *Hum Genet* 2010;127:513–23.
- [6] Barrett JC, Fry B, Maller J, Daly MJ. Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics* 2005;21:263–5.
- [7] Skol AD, Scott LJ, Abecasis GR, Boehnke M. Joint analysis is more efficient than replication-based analysis for two-stage genome-wide association studies. *Nat Genet* 2006;38:209–13.
- [8] American Heart Association. Heart disease and stroke statistics — 2008 update. Dallas, Texas: American Heart Association; 2008. ©2008, American Heart Association.
- [9] Formstone CJ, Moxon C, Murdoch J, et al. Basal enrichment within neuropithelia suggests novel function(s) for Celsr1 protein. *Mol Cell Neurosci* 2010;44:210–22.
- [10] Yates LL, Schnatwinkel C, Murdoch JN, et al. The PCP genes Celsr1 and Vangl2 are required for normal lung branching morphogenesis. *Hum Mol Genet* 2010;19:2251–67.