

# Joinville stroke biobank: study protocol and first year's results

## Biobanco de acidente cerebrovascular de Joinville: protocolo de estudos e resultados dos primeiros anos

Leslie Ecker Ferreira<sup>1</sup>, Paulo Henrique Condeixa de França<sup>1</sup>, Vivian Nagel<sup>1</sup>, Vanessa Venancio<sup>1</sup>, Juliana Safanelli<sup>1</sup>, Felipe Ibiapina dos Reis<sup>1</sup>, Luis Furtado<sup>2</sup>, Rui Kleber Martins<sup>3</sup>, Gustavo Weiss<sup>4</sup>, Elder Oda<sup>5</sup>, Iscia Lopes-Cendes<sup>6,7</sup>, Octavio Pontes-Neto<sup>3</sup>, Norberto Luiz Cabral<sup>1</sup>

### ABSTRACT

Aiming to contribute to studies that use detailed clinical and genomic information of biobanks, we present the initial results of the first Latin American Stroke Biobank. **Methods:** Blood samples were collected from patients included in the Joinville Stroke Registry and four Brazilian cities. Demographic socio-economic data, cardiovascular risk factors, Causative Classification System for Ischemic Stroke, Trial of Org 10172 in Acute Stroke Treatment and National Institutes of Health scores, functional stroke status (modified Rankin) and brain images were recorded. Additionally, controls from both geographic regions were recruited. High-molecular-weight genomic DNA was obtained from all participants. **Results:** A total of 2,688 patients and 3,282 controls were included. Among the patients, 76% had ischemic stroke, 12% transient ischemic attacks, 9% hemorrhagic stroke and 3% subarachnoid hemorrhage. Patients with undetermined ischemic stroke were most common according to the Trial of Org 10172 in Acute Stroke Treatment (40%) and Causative Classification System for Ischemic Stroke (47%) criteria. A quarter of the patients were under 55 years of age at the first-ever episode. **Conclusions:** We established the Joinville Stroke Biobank and discuss its potential for contributing to the understanding of the risk factors leading to stroke.

**Keywords:** stroke; genetics, biobank.

### RESUMO

Com o objetivo de contribuir para estudos que utilizam informações clínicas e genômicas de biobancos, apresentamos os resultados iniciais do primeiro Biobanco Latinoamericano em Acidente Vascular Cerebral (AVC). **Métodos:** Foram coletadas amostras de sangue de pacientes recrutados pelo Registro de AVC de Joinville e posteriormente de quatro cidades brasileiras. Foram registrados dados socioeconômicos demográficos, fatores de risco cardiovasculares, *Causative Classification System (CCS)*, *Trial of Org 10172 in Acute Stroke Treatment*, *National Institutes of Health*, estado funcional (Rankin modificado) e imagens cerebrais. Adicionalmente, foram recrutados controles das regiões geográficas correspondentes. Obteve-se DNA genômico de todos participantes. **Resultados:** Foram incluídos 2688 pacientes e 3282 controles. Entre os pacientes, 76% tiveram AVC isquêmico, 12% ataques isquêmicos transitórios, 9% AVC hemorrágico e 3% hemorragia subaracnóidea. Os casos indeterminados foram os mais frequentes e classificados de acordo com TOAST (40%) e CCS (47%). Um quarto dos pacientes tinham menos de 55 anos no primeiro evento. **Conclusões:** Estabelecemos o *Joinville Stroke Biobank*, e discutimos aqui seu potencial na compreensão dos fatores de risco do AVC.

**Palavras-chave:** acidente vascular cerebral; genética; biobancos.

According to the Global Burden of Disease Study, in 2013, there were almost 25.7 million stroke survivors (71% with ischemic stroke (IS)), 6.5 million deaths from stroke

(51% died from IS), 113 million disability-adjusted life years due to stroke (58% due to IS) and 10.3 million new strokes (67% IS) worldwide<sup>1</sup>. The INTERSTROKE study defined the

<sup>1</sup>Universidade da Região de Joinville, Joinville SC, Brasil;

<sup>2</sup>Universidade Federal do Ceará, Sobral CE, Brasil;

<sup>3</sup>Universidade de São Paulo, Ribeirão Preto SP, Brasil;

<sup>4</sup>Universidade Federal do Mato Grosso do Sul, Campo Grande MS, Brasil;

<sup>5</sup>Universidade Federal do Rio Grande do Sul, Porto Alegre RS, Brasil;

<sup>6</sup>Universidade de Campinas, Campinas SP, Brasil;

<sup>7</sup>Instituto Brasileiro de Neurociência e Neurotecnologia, Campinas SP, Brasil.

**Correspondence:** Leslie Ecker Ferreira; Rua Paulo Malschitzki, 10; 89219-710 Joinville SC; E-mail: [leslie.ferreira@univille.br](mailto:leslie.ferreira@univille.br)

**Conflict of interest:** There is no conflict of interest to declare.

**Support:** CNPq - (Grant Number: 402396-2013/08) and Coordination for the Improvement of Higher Education Personnel (CAPES) (PROCAD 183768).

Received 20 May 2017; Received in final form 15 July 2017; Accepted 08 August 2017.



population-attributable risks for ischemic and hemorrhagic strokes in 22 countries, concluding that 10 classical cardiovascular risk factors are associated with 90% of the risk of stroke<sup>2</sup>. However, twins and familial aggregation studies suggest that the risk of stroke has a substantial genetic component<sup>3</sup>. Although genetics, and more recently genomics, play an increasingly large role in the practice of medicine, the daily clinical care of patients suffering from stroke has not significantly been affected by the advances in these fields. Therefore, one can assume that there is still much knowledge about stroke genetics and genomics needing to be translated into clinical practice.

Currently, there are tremendous efforts to understand the genetic basis of both rare and common cardiovascular and stroke disorders through strategies such as genome-wide association studies and next-generation sequencing studies<sup>4</sup>. In this scenario, it has been pointed out that well-established stroke biobanks with close collaboration between clinicians and geneticists are essential for supplying the biological and clinical information required for these large-scale studies<sup>4,5,6</sup>. To address these challenges specifically, recent cooperative efforts have been launched such as the Stroke Genetics Network and Bio-Repository of DNA in Stroke<sup>7,8</sup>.

As defined by Brazilian regulatory guidelines, a biobank represents an organized collection of human biological material and associated data that are prospectively collected and stored for research purposes, in accordance with predefined technical, ethical and operational standards, under institutional responsibility and management. We present the study protocol and current status of the Joinville Stroke Biobank (JSB), the first Latin American DNA biobank of stroke.

## METHODS

### Participants and samples

The current biobank originated from two sequential initiatives to study epidemiological aspects of stroke in recent years in Brazil<sup>9</sup>. The phenotypic data were extracted from the Joinville Stroke Registry, which is a population-based registry that has been ongoing since 2005. In 2013, the registry became supported by a municipal law<sup>10</sup>. The beginning of case and control data ascertainment, blood extraction and DNA storage took place in 2010. These samples and data, which are still being collected and processed in Joinville, constitute phase I of the JSB, which has no deadline for completion.

Phase II was launched in 2015, when the JSB started to receive blood samples and phenotypic data from four other Brazilian cities. This constitutes a task force of the Brazilian Consortium of Stroke Research, which is sponsored by the Brazilian Ministry of Health and the National Council for Scientific and Technological Development, to define stroke incidence trends, case-fatality proportions and the

prevalence of classic cardiovascular risk factors according to the WHO Steps criteria<sup>11</sup> in different geopolitical scenarios than Joinville, which has a higher Human Development Index than other cities. Data and samples will be collected until 2017.

### Central site and contributing sites

In JSB phase I, inpatients from five hospitals and outpatient data (mild strokes) from the town of Joinville (515,288 inhabitants, 2010 census) were included, whereas JSB phase II covered the following cities: Sobral, north-east region (three hospitals; 147,135 inhabitants), Campo Grande, central region (eight hospitals; 774,202 inhabitants), Sertãozinho, south-east region (two hospitals; 101,784 inhabitants) and Canoas, south region (three hospitals; 323,827 inhabitants), whose geographic locations are shown in the Figure. The central site, as defined by the International Stroke Genetics Consortium is located in the facilities at the University of the Region of Joinville.

### Inclusion and exclusion criteria

Since the beginning of both phases, data and blood samples have been collected from each case and related controls. Patients included transient ischemic attacks, subarachnoid hemorrhages, ischemic strokes and hemorrhagic strokes. We included first-ever or recurrent patients, regardless of age and sex, residing in one of the five cities. All patients had at least one cranial tomography. Stroke confirmation and diagnosis were performed as previously reported<sup>9</sup>. In brief, the phenotype of each stroke patient



**Figure.** Localization of cities, according to Brazilian mesoregions, corresponding to central and contributing sites of the Joinville stroke biobank.

includes demographic information, social class (according to the Brazilian Criteria of Economic Classification based on the National Household Sample Survey)<sup>12</sup>, years of education, type of work (manual or not), family history of stroke, cardiovascular risk factors, biochemical tests and functional status (modified Rankin scale)<sup>13</sup>. Furthermore, IS subtypes were classified according to the modified Trial of Org 10172 in Acute Stroke Treatment (TOAST) criteria and Causative Classification System for Ischemic Stroke (CCS) criteria<sup>14,15</sup>. Patients with IS of undetermined cause due to negative investigation or cryptogenic stroke were classified according to the Embolic Strokes of Undetermined Source international work group<sup>16</sup>. A neuroradiologist classified all hemorrhagic strokes as lobar and non-lobar subtypes.

Controls were matched by age and gender to patients, and individuals with a previous history of stroke or blood ties to patients were excluded. In cases of a positive family history of stroke, blood samples of all related patients and non-stroke relatives were searched for and nominated as nuclear cases.

### Phenotyping methods

Table 1 shows the diagnostic examination and the categories of data retrieved by questionnaires. After 2012, the CCS criteria for phenotyping IS patients were also employed. The neurologists of all cities in the contributing sites were trained by stroke neurologists from the Joinville Stroke Registry using the platform available on the CCS website<sup>15</sup>. After the TOAST diagnosis, the clinical history, physical examination findings and the results of diagnostic work-up were reviewed for determining the CCS classification. Ischemic stroke patients with incomplete diagnostic examination were classified as a possible IS subtype or received an IS diagnosis as undetermined with incomplete evaluation. All Joinville IS diagnoses were blinded for adjudication by two authors (FIR and NLC).

### Data and blood sample collection

Patient selection and blood collection were performed by neurologists and nurses, respectively. Research nurses invited the controls to enroll from among the accompanying persons willing to participate who fulfilled the inclusion criteria. After clarifying the aims of the JSB and the corresponding roles of all parties involved, blood sampling (phase I: 12 mL by venous puncture employing common EDTA-containing vacuum tubes; phase II: finger puncture followed by drying and storage on an FTA Elute Card<sup>†</sup>, Whatman, Kent, United Kingdom) of patients and controls was performed by research nurses. Baseline data included self-declared skin color, collected in strict accordance with the mandatory procedures established by the Brazilian Institute of Geography and Statistics for purposes of population census. Demographics, socio-economic data,

cardiovascular risk factors, National Institutes of Health Stroke Scale (clinical stroke severity) and functional outcome can be extracted from the Joinville Stroke Registry, which also includes causes and dates of death. The follow-up of patients is being carried out by phone calls by previously-trained research nurses in one, three, six, nine and 12 months in the first year and once a year for the following four years after stroke diagnosis. Patients are asked about regular visits to a general physician, blood pressure control, results of glycated hemoglobin and cholesterol levels, and smoking habits.

An identification code, which comprises numbers that denote the date of recruitment, contributing site and entry order, was assigned to each patient enrolled. Controls were identified using numbers that keep their relationship with each matched patient. These identification codes link clinical data and demographics with deposited DNA samples and only the biobank managers or authorized personnel have access to the key linking samples and phenotypic data.

### Processing of samples, DNA extraction and quality control

Blood samples from phase I were maintained at 4°C until DNA extraction. Whole blood samples were centrifuged for 10 minutes at 2,500 x g for separation of the buffy coat, which contains most of the white blood cells. Thereafter, genomic DNA was obtained by the classic “phenol-chloroform” procedure for total DNA extraction. The DNA was resuspended in 200 µL of buffer TE (10 mM Tris-HCl; 1 mM EDTA; pH 8.0). For phase II, we adopted the procedure for large-amount DNA extraction, starting with three paper punches of 3 mm circles, as recommended by the manufacturer of the FTA Elute Card<sup>†</sup>. The purity and yield of DNA extracted were evaluated by spectrophotometry at 260 and 280 nm (Epoch, Biotek Instruments, Winooski, USA). Long-term storage was considered for  $A_{260}/A_{280}$  ratios in the range 1.8–2.0. Finally, DNA was stored in aliquots at -80°C.

### Ethical considerations

The JSB development plan and the informed consent forms were both approved by the Institutional Review Board of the University of the Region of Joinville and by the Brazilian National Commission for Ethics in Research (protocol 25000.142907/2013-07), which holds the final approval prerogative for biobanks in Brazil. The adopted standards and procedures strictly follow the ethical rules that came into force in 2011 and are described in Resolution 441/11 of the National Health Council and in Ordinance 2.201 of the Brazil Ministry of Health, which correspond to the national guidelines for biorepositories and biobanks of human biological material for research purposes. The JSB also follows the international ethical regulations. All volunteers (patients and controls) are enrolled after written informed consent has been given freely by the participants

**Table 1.** Demographic and diagnostic examination data available in the Joinville Stroke Biobank.

Category	Description	Cases	Controls
Sociodemographic	Date of birth/age	Y	Y
	Gender	Y	Y
	Skin color	Y	Y
	City of residence	Y	Y
	Address	Y	Y
	Social status	Y	N
	Education level	Y	N
Clinical and biochemical data	First-ever event?	Y	N
	Date of enrollment	Y	N
	Hour of enrollment	Y	N
	Creatinine	Y	N
	Triglyceride	Y	N
	Uric acid	Y	N
	Blood glucose	Y	N
	Cholesterol (hdl/ldl)	Y	N
	National Institutes of Health Stroke Scale (admission)	Y	N
	Barthel Index	Y	N
	Rankin Index	Y	N
	Oxfordshire Community Stroke Project classification (Banford)	Y	N
	TOAST classification	Y	N
	CCS classification	Y	N
Premorbid risk factors	Transient ischemic attack	Y	Y
	Hypertension	Y	Y
	Diabetes mellitus	Y	Y
	Dyslipidemia	Y	Y
	Acute myocardial infarction	Y	Y
	Congestive heart failure	Y	N
	Atrial fibrillation	Y	N
	Smoking	Y	Y
	Alcohol consumption **	Y	N
	Drug consumption	Y	N
Physical activity	Y	N	
Diagnostic examination	Brain CT	Y	N
	Brain MRI	Y	N
	Digital angiography	Y	N
	Extracranial vessels evaluation		
	Ultrasound Doppler	Y	N
	CT – angiography	Y	N
	MR – angiography	Y	N
	Intracranial vessels evaluation		
	Transcranial Doppler	Y	N
	CT – angiography	Y	N
	MR – angiography	Y	N
	Embolic sources		
	ECG	Y	N
	Transthoracic echocardiography	Y	N
Transesophageal echocardiography	Y	N	
24 h Holter	Y	N	
Outcome measures	Rankin scale (30 days, 6 months, once a year up to 5 years)	Y	N
	Death (cause, date) by death certificate	Y	N

TOAST: Trial of Org 10172 in Acute Stroke Treatment criteria<sup>16</sup>; CCS: Causative Classification System criteria<sup>17</sup>. CT: computed tomography; MRI: magnetic resonance imaging. Y: yes (available); N: no (not available). \*Self-declared skin color; \*\*current moderate or heavy drinker was defined as 50 g/day, which is equivalent to 500 mL or two glasses of wine, 1000 mL of beer or 5 units of spirits, or being intoxicated at least once a week.

or their legal representatives. The freedom to refuse participation or withdraw at any time is guaranteed, and the confidentiality of personal data is ensured in all circumstances. Free access to information and associated implications is guaranteed to participants, thereby respecting the expression of individual will.

## RESULTS

Since 2010, a total of 5,970 DNA samples have been registered belonging to patients (2,688) and controls (3,282). The mean ages of patients and controls were  $64.6 \pm 19.3$  and  $56.9 \pm 14.9$  years, respectively. Men were more represented among the patients (64%; 1,731/2,688), while women were more prevalent among controls (68%; 2,226/3,282).

Table 2 shows the distribution of patients and controls per city, as well as their self-declared skin color. White

individuals predominated in the south (88%) and south-east (55%) regions and brown in the north-east (53%) and central (41%) regions.

Table 3 shows the distribution of stroke diagnoses of patients in the different cities. Most of the patients (76%; 2,031/2,688) were IS, followed by 12% (319/2,668) transient ischemic attacks, 9% (243/2,668) hemorrhagic stroke and 3% (95/2,688) subarachnoid hemorrhage. A quarter of all stroke patients were aged  $\leq 55$  years.

The distribution of IS cases by TOAST criteria (Table 4) was: 20% (426/2,031) cardioembolic; 22% (440/2,031) small-artery occlusion; 16% (317/2,031) large-artery atherosclerosis; 40% undetermined (807/2,031) and 2% (41/2,031) other determined cause. After 2012, patients were also classified by CCS criteria (56%; 1,503/2,688). As expected, among IS patients, the most common IS subtype was undetermined, not only by TOAST criteria (40%; 807/2,031), but also by CCS criteria (47%; 718/1,503). Out of those 807 undetermined IS patients

**Table 2.** Baseline characteristics of stroke patients and controls in the Joinville Stroke Biobank database.

City	Campo Grande	Canoas	Joinville	Sertãozinho	Sobral	All
Demography						
Cases (n)	507	187	1619	175	200	2688
Mean age ( $\pm$ SD) (years)	65.2 (14.9)	64.5 (14.3)	64.3 (14.1)	64.9 (15.1)	67.1 (16.1)	64.6 (19.3)
Male [n(%)]	290 (57.2)	100 (53.5)	1185 (73.2)	56 (32)	100 (50)	1731 (64.4)
Skin color [% (n/N)]						
White	46.9 (238/507)	49.7 (93/187)	88.2 (1428/1619)	55.4 (97/175)	29.5 (59/200)	71.2 (1915/2688)
Black	7.7 (39/507)	16.0 (30/187)	5.4 (87/1619)	13.1 (23/175)	16.0 (32/200)	7.8 (211/2688)
Brown	41.4 (210/507)	20.9 (39/187)	6.0 (97/1619)	29.1 (51/175)	53.5 (107/200)	18.8 (504/2688)
Native American	2.0 (10/507)	2.7 (5/187)	0.2 (4/1619)	1.1 (2/175)	0.5 (1/200)	0.8 (22/2688)
Yellow	1.4 (7/507)	0.5 (1/187)	0.1 (2/1619)	1.1 (2/175)	0.5 (1/200)	0.5 (13/2688)
Not-declared	0.6 (3/507)	10.2 (19/187)	0.1 (1/1619)	0.0	0.0	0.9 (23/2688)
Controls (n)	180	24	2586	276	216	3282
Mean age ( $\pm$ SD) years)	50.2 (14.6)	64.56 (13.1)	68.3 (14.7)	50.2 (13.9)	66.3 (15.5)	56.9 (14.9)
Male [n(%)]	61 (33.8)	14 (58.3)	865 (33.5)	65 (23.5)	51 (23.6)	1056 (32.2)
Skin color [% (n/N)]						
White	53.9 (97/180)	62.5 (15/24)	90.4 (2338/2586)	46.7 (129/276)	32.9 (71/216)	80.7 (2650/3282)
Black	0.6 (1/180)	25.0 (6/24)	2.7 (71/2586)	6.9 (19/276)	6.5 (14/216)	3.4 (111/3282)
Brown	40.0 (72/180)	8.3 (2/24)	6.4 (166/2586)	44.2 (122/276)	47.7 (103/216)	14.2 (465/3282)
Native American	0.6 (1/180)	4.2 (1/24)	0.0 (1/2586)	0.4 (1/276)	0.9 (2/216)	0.2 (6/3282)
Yellow	0.6 (1/180)	0.0	0.1 (3/2586)	1.4 (4/276)	0.0	0.2 (8/3282)
Not-declared	4.4 (8/180)	0.0	0.3 (7/2586)	0.4 (1/276)	12.0 (26/216)	1.3 (42/3282)

**Table 3.** Major stroke diagnosis of patients in the Joinville Stroke Biobank database per city [% (n/N)].

Variable	City					
	Campo Grande	Canoas	Joinville	Sertãozinho	Sobral	All
IS	16.5 (335/2031)	7.5 (152/2031)	60.9 (1238/2031)	5.6 (112/2031)	9.5 (194/2031)	75.5 (2031/2688)
TIA	20.6 (66/319)	1.6 (5/319)	73.1 (233/319)	4.7 (15/319)	0	11.9 (319/2688)
HS lobar	-	-	11.7 (28/243)	0	-	-
HS non-lobar	-	-	7.8 (19/243)	0	-	-
HS non-classified	34.6 (84/243)	11.1 (27/243)	20.1 (49/243)	13.9 (34/243)	0.8 (2/243)	9.1 (243/2688)
SAH	23.1 (22/95)	3.5 (3/95)	54.6 (52/95)	14.6 (14/95)	4.2 (4/95)	3.5 (95/2688)
Young adult (≤ 55 years old)	4.3 (118/2688)	0.4 (13/2668)	15.7 (424/2688)	1.9 (52/2688)	1.9 (53/2688)	24.5 (660/2688)
Nuclear cases	0	0	0.4 (7/1619)	0	0	0.4 (7/2688)
All cases	18.9 (507/2688)	6.9 (187/2688)	60.2 (1619/2688)	6.5 (175/2688)	7.5 (200/2688)	2688

IS: ischemic stroke; TIA: transient ischemic attack; HS: hemorrhagic stroke; SAH: subarachnoid hemorrhage.

classified by TOAST criteria, 55% (446/807) were patients with an incomplete evaluation, 42% (340/807) were cryptogenic and 3% (21/807) were undetermined with two or more possible causes. Out of those 718 undetermined IS subtypes classified by CCS criteria, 25% (370/1,503) were IS patients with incomplete evaluation, 17% (262/1,503) were unknown other cryptogenic, 8% (61/1,503) were unknown cryptogenic embolism and 1% (25/1,503) were unclassified.

## DISCUSSION

Up to October 2016, DNA from 5,970 blood samples distributed among patients (2,688) and controls (3,282) had been extracted and stored. Since it is believed that most of the genetic variants identified in stroke studies will have small, individual effects on disease risk, efforts aimed at the successful discovery of the impact of genes have required large sample sizes (usually involving tens or hundreds of thousands of cases and controls) to achieve sufficient statistical power<sup>10</sup>. In addition, as with any case-control study, controls should be representative of cases<sup>17,18</sup>.

With this in mind, the JSB comprises well-defined phenotypes and samples of high-quality DNA. The priority recruitment of spouses or partners as case-matched controls represents a suitable strategy for minimizing differences due to environmental effects to which the participants (patients and controls) have been exposed. As controls are collected in the same geographic region as patients, it is reasonable to expect a homogeneous distribution of ethnic backgrounds between patients and controls. On the other hand, despite having inflated influences due to shared variants and environmental effects between patients and controls, the JSB may also provide relevant information on familial aggregation analysis. It is also noteworthy to clarify that the JSB has a forecast of

unlimited continuity as, among other factors, the Joinville Stroke Registry is supported by a municipal law. This gives rise to the fact that if there are cases of stroke in individuals previously categorized as a control (since stroke events may happen later in life), their biobank status changes, but maintains the traceability of the change. This allows the possibility of evaluation of the impact on results of any studies already performed with the data and samples of those individuals who no longer belong to the control group. As far as we know, this is the first stroke biobank in Latin America, a region with a massive race mixture.

Under TOAST, IS cases are assigned to one of three main subtypes (small-artery occlusion, large-artery atherosclerosis, cardioembolic); to a rare cause (e.g., carotid or vertebral artery dissection); or remain undetermined (because of more than one potential cause, incomplete investigation or no apparent cause despite complete investigation/cryptogenic stroke). The proportion in the undetermined category varies, but can be substantial (up to 40% or more), and is usually excluded from ischemic subtype-specific analyses<sup>19,20</sup>. In our cohort, between 22% (TOAST criteria) and 25% (CCS criteria) were undetermined IS due to incomplete investigation. All patients with incomplete investigation were assigned this diagnosis. In fact, the highly diverse sociodemographic scenarios at sites contributing to the JSB, including difficulties related to health-care system access and imaging diagnostic workup, might have influenced significantly. Nevertheless, independent raters may perform blinded readjudications of all JSB IS diagnoses whenever considered necessary.

Globally, large DNA biobanks have been contributing to the understanding of the different genetic architectures and pathophysiological aspects of stroke. In this regard, the UK Biobank (European and British Asian), Brains-SA (India and Sri Lanka), Bio-Repository of DNA in Stroke – Middle East (Qatar)<sup>21</sup>, China Kadoorie Biobank<sup>22</sup> and other

**Table 4.** Ischemic stroke subtypes diagnosis of patients in the Joinville Stroke Biobank database [% (n/N)].

Variable	City					
	Campo Grande	Canoas	Joinville	Sertãozinho	Sobral	All
IS subtype by TOAST						
Large-artery atherosclerosis	6.9 (22/317)	5.1 (16/317)	81.8 (259/317)	2.8(9/317)	3.4 (11/317)	15.8 (317/2031)
Cardioembolic	12.2 (52/426)	3.2 (14/426)	78.3 (333/426)	2.8(12/426)	3.5 (15/426)	20.2 (426/2031)
Small-artery occlusion	17.1 (75/440)	4.3 (19/440)	6.9 (307/440)	0.9(4/440)	7.9 (35/440)	21.7 (440/2031)
Other determined cause	17.1(7/41)	0	82.9 (34/41)	0	0	2.2(41/2031)
Undetermined						
Cryptogenic/ESUS	5.6 (19/340)	16.2 (55/340)	51.8 (176/340)	2.6 (9/340)	23.8 (81/340)	16.8 (340/2031)
Incomplete investigation	35.6(159/446)	10.8 (48/446)	24.6 (110/446)	17.5 (78/446)	11.4 (51/446)	22.1 (446/2031)
Two or more determined causes	4.7 (1/21)	0	90.6 (19/21)	0	4.7 (1/21)	1.2 (21/2031)
IS subtype by CCS						
Large-artery atherosclerosis						
Evident	5.3 (8/150)	10.7 (16/150)	76.7 (115/150)	0	7.3 (11/150)	9.9 (150/1503)
Probable	0	0	92.1(35/38)	7.9(3/38)	0	2.4 (38/1503)
Possible	12(3/25)	0	64 (16/25)	24 (6/25)	0	1.5 (25/1503)
Cardioembolic						
Evident	11.7 (18/154)	7.1 (11/154)	75.3 (116/154)	3.2 (5/154)	2.6 (4/154)	10.2 (154/1503)
Probable	0	0	100 (40/40)	0	0	2.5 (40/1503)
Possible	15.7 (8/51)	0	49.1 (25/51)	13.6(7/51)	21.6 (11/51)	3.3 (51/1503)
Small-artery occlusion						
Evident	6.8(12/175)	10.3 (18/175)	6.5 (113/175)	5.7(1/175)	17.7 (31/175)	11.5 (175/1503)
Probable	15.1 (10/66)	0	8.2(54/66)	3.1 (2/66)	0	4.3 (66/1503)
Possible	2.1(1/48)	0	87.5 (42/48)	2.1(1/48)	8.3(4/48)	3.1 (48/1503)
Other determined cause						
Evident	11.1 (4/36)	0	86.2(31/36)	0	2.7 (1/36)	2.3 (36/1503)
Probable	0	0	100(2/2)	0	0	1.2 (2/1503)
Possible	0	0	0	0	0	0
Undetermined						
Unknown cryptogenic embolism	0	0	80.3(49/61)	0	19.7 (12/61)	4.1 (61/1503)
Unknown other cryptogenic	6.1 (16/262)	20.2(53/262)	44.6 (117/262)	26.7(7/262)	26.3 (69/262)	17.4 (262/1503)
Incomplete evaluation	27.6 (102/370)	12.9 (48/370)	24.3 (90/370)	21.3 (79/370)	13.8 (51/370)	24.6 (370/1503)
Unclassified	4(1/25)	24(6/25)	64 (16/25)	4 (1/25)	4 (1/25)	1.7 (25/1503)
Total IS by TOAST (since 2010)	16.5 (335/2031)	7.5 (152/2031)	60.9 (1238/2031)	5.6 (112/2031)	9.5 (194/2031)	75.5 (2031/2688)
Total IS by CCS (since 2012)	12.2 (183/1503)	10.1 (152/1503)	57.3 (861/1503)	74.5 (112/1503)	12.9 (195/1503)	72.0(1503/2100)

IS: ischemic stroke; TOAST: Trial of Org 10172 in Acute Stroke Treatment criteria<sup>16</sup>; ESUS: Embolic Strokes of Undetermined Source; CCS: Causative Classification System criteria<sup>17</sup>.

collaborative initiatives, such as the International Stroke Genetics Consortium, have contributed to the identification of several loci associated with IS subtypes.

Genome-wide association study approaches have been adopted for investigating many complex diseases, such as stroke, and have been effective in identifying new genetic variants associated with the risk of disease. However, for stroke studies, genome-wide association studies have so far been applied principally in European, North American and Japanese populations<sup>3,23</sup>.

In particular, four loci (*PITX2*, *HDAC9*, *ZFHX3* and 12q24.2) were convincingly implicated by genome-wide association studies as associated with IS in European-ancestry patients<sup>24,25</sup>. Recently, a novel locus (rs12122341) was identified at 1p13.2 (near *TSPAN2*), associated with large-artery atherosclerosis stroke, and the locus 12q24 (near *ALDH2*) was associated with small-artery occlusion<sup>26</sup>. In addition, the gene *TRPV3* was associated with cardioembolic stroke after replication of exome sequencing analysis, demonstrating that two polymorphisms were associated with cardioembolic stroke risk in two cohorts,

with rs151091899 being the most significant<sup>27</sup>. Therefore, replication studies in other populations are determinants in understanding the significance of such polymorphisms as stroke risk factors. The Brazilian population is one of the most heterogeneous populations in the world, comprising an admixture of Native Americans, Europeans and Africans<sup>28</sup>. The admixture process occurred through different means in Brazilian geographic regions. For instance, the Native American contribution is more pronounced in northern Brazil, the African contribution is more elevated in the north-east, and the south features a European predominance with few Native American and African influences<sup>29</sup>. Consequently, the JSB, unlike other biobanks<sup>8,22,23</sup>, comprises an ideal admixed population with great ethnic variability linked to epidemiological and clinical data, which may allow confirmation of polymorphisms found

in specific populations, contributing to global understanding of genetic mechanisms involved in stroke.

In conclusion, the clinical, environmental and epidemiological issues related to the different types of stroke, combined with the large genetic variability captured by the JSB, may increase the chance of identifying new and relevant factors predisposing individuals to stroke (and recovery) in Brazil. Our aim is to continue patient assessment and blood collection, as well as begin genotyping samples in the near future. Some initial studies are already being conducted in the Laboratory of Molecular Genetics at the University of Campinas with the support of the Brazilian Institute of Neuroscience and Neurotechnology and the use of a genome-wide association genotyping platform (Affymetrix™ 6.0).

## References

1. Feigin VL, Krishnamurthi RV, Parmar P, Norrving B, Mensah GA, Bennett DA et al. Update on the global burden of ischemic and hemorrhagic stroke in 1990–2013: the GBD 2013 study. *Neuroepidemiology*. 2015;45(3):161-76. <https://doi.org/10.1159/000441085>
2. O'Donnell MJ, Xavier D, Liu L, Zhang H, Chin SL, Rao-Melacini P et al. Risk factors for ischemic and intracerebral haemorrhagic stroke in 22 countries (the INTERSTROKE study): a case-control study. *Lancet*. 2010;376(9735):112-23. [https://doi.org/10.1016/S0140-6736\(10\)60834-3](https://doi.org/10.1016/S0140-6736(10)60834-3)
3. Ikram MA, Seshadri S, Bis JC, Fornage M, DeStefano AL, Aulchenko YS et al. Genomewide association studies of stroke. *N Engl J Med*. 2009;360(17):1718-28. <https://doi.org/10.1056/NEJMoa0900094>
4. Musunuru K, Hickey KT, Al-Khatib SM, Delles C, Fornage M, Fox CS et al. Basic concepts and potential applications of genetics and genomics for cardiovascular and stroke clinicians: a scientific statement from the American Heart Association. *Circ Cardiovasc Genet*. 2015;8(1):216-42. <https://doi.org/10.1161/HCG.000000000000020>
5. Dichgans M. Genetics of ischaemic stroke. *Lancet Neurol*. 2007;6(2):149-61. [https://doi.org/10.1016/S1474-4422\(07\)70028-5](https://doi.org/10.1016/S1474-4422(07)70028-5)
6. Hewitt J, Walters M, Padmanabhan S, Dawson J. Cohort profile of the UK Biobank: diagnosis and characteristics of cerebrovascular disease. *BMJ Open*. 2016;6(3):e009161. <https://doi.org/10.1136/bmjopen-2015-009161>
7. Meschia JF, Arnett DK, Ay H, Brown RD Jr, Benavente OR, Cole JW et al. Stroke Genetics Network (SiGN) study: design and rationale for a genome-wide association study of ischemic stroke subtypes. *Stroke*. 2013;44(10):2694-702. <https://doi.org/10.1161/STROKEAHA.113.001857>
8. Yadav S, Schanz R, Maheshwari A, Khan MS, Stark J, Silva R et al. Bio-repository of DNA in stroke (BRAINS): a study protocol. *BMC Med Genet*. 2011;12:34. <https://doi.org/10.1186/1471-2350-12-34>
9. Cabral NL, Cougo-Pinto PT, Magalhaes PS, Longo AL, Moro CH, Amaral CH et al. Trends of stroke incidence from 1995 to 2013 in Joinville, Brazil. *Neuroepidemiology*. 2016;46(4):273-81. <https://doi.org/10.1159/000445060>
10. Prefeitura Municipal de Joinville. Lei Nº 7.448, de 12 de junho de 2013. Institui o banco de dados Registro de Acidente Vascular Cerebral – AVC de Joinville e dá outras providências. *Jornal Município*. 28 jun 2013.
11. World Health Organization – WHO. WHO Steps Stroke Manual (V2.0): the WHO STEPwise approach to stroke surveillance.. Geneva: World Health Organization; 2005.
12. Paiva GFS, Silva DBN, Feijó CA. Consumption and socioeconomic classification in Brazil: a study based on the Brazilian family expenditure survey. In: IARIW-IBGE Special Conference on "Income, Wealth and Well-being in Latin America"; 2013 Sep 11-14; Rio de Janeiro, Brazil.
13. Sulter G, Steen C, De Keyser J. Use of the Barthel index and modified Rankin scale in acute stroke trials. *Stroke*. 1999;30(8):1538-41. <https://doi.org/10.1161/01.STR.30.8.1538>
14. Adams HP Jr, Bendixen BH, Kappelle LJ, Biller J, Love BB, Gordon DL et al. Classification of subtype of acute ischaemic stroke. Definitions for use in a multicenter clinical trial. TOAST. Trial of Org 10172 in Acute Stroke Treatment. *Stroke*. 1993;24(1):35-41. <https://doi.org/10.1161/01.STR.24.1.35>
15. Ay H, Benner T, Arsava EM, Furie KL, Singhal AB, Jensen MB et al. A computerized algorithm for etiologic classification of ischemic stroke: the Causative Classification of Stroke System. *Stroke*. 2007;38(11):2979-84. <https://doi.org/10.1161/STROKEAHA.107.490896>
16. Hart RG, Diener HC, Coutts SB et al. Embolic strokes of undetermined source: the case for a new clinical construct. *Lancet Neurol*. 2014;13(4):429-38. [https://doi.org/10.1016/S1474-4422\(13\)70310-7](https://doi.org/10.1016/S1474-4422(13)70310-7)
17. Jayasinha SR, Mishra A, Van Daal A, Kwan E. Genetics and cardiovascular disease: design and development of a DNA biobank. *Exp Clin Cardiol*. 2009;14(3):33-7.
18. McCarty CA, Chisholm RL, Chute CG, Kullo IJ, Jarvik GP, Larson EB et al. The eMERGE Network: a consortium of biorepositories linked to electronic medical records data for conducting genomic studies. *BMC Med Genomics*. 2011;4(1):13. <https://doi.org/10.1186/1755-8794-4-13>
19. Wolfe C, Rudd A, Howard R et al. Incidence and case fatality rates of stroke subtypes in a multiethnic population: the South London Stroke Register. *J Neurol Neurosurg Psychiatry*. 2002;72:211-6.
20. Feigin VL, Lawes CM, Bennett DA, Barker-Collo SL, Parag V. Worldwide stroke incidence and early case fatality reported in 56 population-based studies: a systematic review. *Lancet Neurol*. 2009;8(4):355-69. [https://doi.org/10.1016/S1474-4422\(09\)70025-0](https://doi.org/10.1016/S1474-4422(09)70025-0)
21. Cotlarciuc I, Khan MS, Maheshwari A, Yadav S, Khan FY, Al-Hail H et al. Bio-repository of DNA in stroke: a study protocol of three ancestral populations. *JRSM Cardiovasc Dis*. 2012;1(4):cvd.2012.12019. <https://doi.org/10.1258/cvd.2012.012019>



22. Chen Z, Chen J, Collins R, Guo Y, Peto R, Wu F et al. China Kadoorie Biobank of 0.5 million people: survey methods, baseline characteristics and long-term follow-up. *Int J Epidemiol*. 2011;40(6):1652-66. <https://doi.org/10.1093/ije/dyr120>
23. Matsushita T, Umeno J, Hirakawa Y, Yonemoto K, Aschikawa K, Amitani H et al. Association study of the polymorphisms on chromosome 12p13 with atherothrombotic stroke in the Japanese population. *J Hum Genet*. 2010;55(7):473-6. <https://doi.org/10.1038/jhg.2010.45>
24. Bellenguez C, Bevan S, Gschwendtner A, Spencer CC, Burgess AI, Pirinen M et al. Genome-wide association study identifies a variant in HDAC9 associated with large vessel ischemic stroke. *Nat Genet*. 2012;44(3):328-33. <https://doi.org/10.1038/ng.1081>
25. Traylor M, Farrall M, Holliday EG, Sudlow C, Hopewell JC, Cheng YC et al. Genetic risk factors for ischaemic stroke and its subtypes (the METASTROKE Collaboration): a meta-analysis of genome-wide association studies. *Lancet Neurol*. 2012;11(11):951-62. [https://doi.org/10.1016/S1474-4422\(12\)70234-X](https://doi.org/10.1016/S1474-4422(12)70234-X)
26. NINDS Stroke Genetics Network (SiGN); International Stroke Genetics Consortium (ISGC). Loci associated with ischaemic stroke and its subtypes (SiGN): a genome-wide association study. *Lancet Neurol*. 2016;15(2):174-84. [https://doi.org/10.1016/S1474-4422\(15\)00338-5](https://doi.org/10.1016/S1474-4422(15)00338-5)
27. Carrera C, Jiménez-Conde J, Derdak S et al. Whole exome sequencing analysis reveals TRPV3 as a risk factor for cardioembolic stroke. *Thromb Haemost*. 2016;116(6):1165-71. <https://doi.org/10.1160/TH16-02-0113>
28. Amador MA, Cavalcante GC, Santos NP et al. Distribution of allelic and genotypic frequencies of IL1A, IL4, NFKB1 and PAR1 variants in Native American, African, European and Brazilian populations. *BMC Res Notes*. 2016;9(1):101. <https://doi.org/10.1186/s13104-016-1906-9>
29. Santos NP, Ribeiro-Rodrigues EM, Ribeiro-dos-Santos AK, Pereira R, Gusmão L, Amorim A et al. Assessing individual interethnic admixture and population substructure using a 48-Insertion-deletion (INSEL) ancestry-informative marker (AIM) panel. *Hum Mutat*. 2010;31(2):184-90. <https://doi.org/10.1002/humu.21159>