Genome-Wide Expression Profiles for Ischemic Stroke: A Meta-Analysis

Carlos E. Moreno-Ramírez, * Eulogia Gutiérrez-Garzón, * George E. Barreto, MSc, PhD,† and Diego A. Forero, MD, PhD*

Background: Genome-wide expression studies (GWES), using microarray platforms, have allowed a deeper understanding of the molecular factors involved in the pathophysiology of ischemic stroke (IS), one of the main global causes of mortality and disability. *Methods:* In the current work, we carried out a meta-analysis of available GWES for IS. Bioinformatics and computational biology analyses were applied to identify enriched functional categories and convergence with other genomic datasets for IS. *Results:* Three primary datasets were included and in the meta-analyses for GWES and IS, 41 differentially expressed (DE) genes were identified using a random effects model. Thirteen of these genes were downregulated and 28 were upregulated. An analysis of functional categories found a significant enrichment for the Gene Ontology Term "Inflammatory Response" and for binding sites for the PAX2 transcription factor. *Conclusions:* The list of DE genes identified in this meta-analysis of GWES for IS is useful for future genetic and molecular studies, which would allow the identification of novel mechanisms involved in the pathophysiology of IS. Several of the DE genes found in this meta-analysis have known functional roles related to mechanisms involved in the pathophysiology of IS. It is recognized the role of the inflammatory response in the pathophysiology of IS.

Key Words: Ischemic stroke—neurogenetics—genome-wide expression—bioinformatics—meta-analysis—computational biology

Introduction

Ischemic stroke (IS) is currently defined as an acute episode of focal dysfunction in the central nervous system that lasts longer than 24 hours, or of any duration if imaging methods or autopsy confirms focal infarction.[1] Stroke is one of the main causes of mortality, after ischemic cardiomyopathy, and it is the third global cause of disability, with an incidence (in 2010) of 176 per 100,000 person-years and a mortality rate of 42 per 100,000 person-years.[2] The main risk factors for stroke include advanced age, arterial hypertension, hypercholesterolemia, coronary heart disease, auricular fibrillation, diabetes mellitus, and obesity, among others.[3,4]

In order to discover the molecular basis of IS, several genomic approaches, such as Genome-Wide Association Studies (GWAS) have been carried out.[5] Genome-wide expression studies (GWES), using microarray platforms with tens of thousands of probes,[6,7] have allowed a deeper understanding of the factors molecular factors involved in the pathophysiology of stroke, as it has been used for other neurological disorders.[8] However, as the sample sizes for the individual GWES for IS performed until now have been limited, it highlights the importance of carrying out meta-analyses for available GWES.[9] Meta-analyses of GWES have been carried out for an important number of major neurological diseases and psychiatric

From the *Laboratory of Neuropsychiatric Genetics, Biomedical Sciences Research Group, School of Medicine, Universidad Antonio Nariño, Bogotá, Colombia; and †Departamento de Nutrición y Bioquímica, Facultad de Ciencias, Pontificia Universidad Javeriana, Bogotá, Colombia.

Received November 28, 2017; revision received July 7, 2018; accepted July 22, 2018.

Financial disclosure: DAF is supported by research grants from Colciencias (Grant No. 823-2015) and VCTI-UAN (Grant No. 2016220).

Conflict of Interest: The authors declare that they have no conflict of interest.

Address correspondence to Diego A. Forero, MD, PhD, Laboratory of NeuroPsychiatric Genetics, School of Medicine, Universidad Antonio Nariño, Bogotá, Colombia. E-mailes: diego.forero@uan.edu.co, daforerog@gmail.com

1052-3057/$ - see front matter

https://doi.org/10.1016/j.jstrokecerebrovasdis.2018.07.035

disorders.[9-12] An enrichment analysis of functional categories for the lists of differentially expressed (DE) genes can be used to identify biological pathways and mechanisms related to the pathophysiology of IS.[13] In the current work, we carried out a meta-analysis of available GWES for IS.

Materials and Methods

Search and Inclusion of Primary Datasets

Our meta-analyses of microarray data followed published recommendations for conducting this type of work, which involves the analysis of genome-wide expression data from multiple individual studies.[9] GWES, with available raw data, for IS, using the case-control design, were searched in NCBI GEO and Array Express databases.[14,15] Later, the search was complemented with a search in PubMed for articles describing GWES for IS (using "stroke," "IS," and "cardioembolic stroke" as keywords) and in reference lists of key review articles.[16,17] In case of published articles without associated records in microarray repositories, the corresponding authors were contacted by email to ask for availability of raw data. Inclusion criteria were studies that analyzed genome-wide expression data in IS patients and controls and that had available raw data in microarray repositories. Exclusion criteria were studies that focused on analysis of miRNA expression or DNA methylation or that only provided raw data for patients.

Data Processing and Meta-Analysis of GWES

Demographic data were extracted and tabulated from the different studies that were selected and included. The preprocessed expression data from each microarray dataset were downloaded from NCBI GEO database and used for the subsequent analyses (for the Stamova et al[18] study, the data from patients before treatment was employed). In these datasets, the background correction and normalization using the Robust Microarray Analysis algorithm were carried out.[18-20] Annotation files for the different microarray platforms were downloaded from the NCBI GEO database. The NetworkAnalyst online program[21] was used for the meta-analysis of genome-wide expression data, under a random effects model, following published recommendations for this type of study (including the annotation matching, mapping probes to genes).[22] A false discovery rate of .05 was used to correct for multiple testing.

Functional Annotation and Enrichment Analysis

The Database for Annotation Visualization and Integrated Discovery 6.8 and Babelomics 5.0 online tools[23,24] were used for the enrichment analysis of functional categories: Gene Ontology, Gene Expression, Chromosomal Location, Kyoto Encyclopedia of Genes and Genomes Pathways, Interpro domains, Transcription Factor Binding Sites, and miRNA targets,[13] employing the default parameters. The option of a comparison against the entire genome was chosen and a false discovery rate approach was also used for adjustment for multiple testing.

In order to explore the convergence of DE genes identified in this meta-analysis and top candidate genes from meta-analyses for GWAS for IS, top Single Nucleotide Polymorphism (SNPs) from the GWAS for IS published by Ikram et al[25] and Kilarski et al[26] were extracted, and the BioMart online tool[27] was used for the identification of the respective genes. The ToppGene server[28] was used for carrying out a candidate gene prioritization (employing the default parameters),[29] including top genes derived from GWAS as a training set and the genes identified in this meta-analysis as the testing set, incorporating multiple lines of available biological evidence.

Experimentally validated protein-protein interaction (PPI) data were retrieved from the Human Interactome Project (Center for Cancer Systems Biology, Harvard University, MA). This resource incorporates results from several available datasets: HI-II-14 and Lit-BM-13,[30] HI-I-05[31]; Venkatesan-09[32] and Yu-11.[33] For the analysis of PPI networks, the Cytoscape 3.2.1[34] program was used and a subnetwork of highly connected proteins (>15 connections) was generated in Cytoscape to facilitate the visualization of PPI data (to avoid generation of PPI networks with too many nodes and edges).[10]

Results

Three available NCBI GEO datasets were used (GSE16561, GSE22255, and GSE58294), describing GWES in RNA extracted from blood cells of healthy controls and patients after the IS. Other several published studies did not deposit the respective raw data in public repositories. These datasets represented a total sample of 149 unique subjects, with 82 cases and 67 controls, showing inter-study differences in several clinical parameters, including time from onset of the IS to blood sampling (Table 1). Barr et al[20] used the Illumina HumanRef-8 v3.0 expression beadchip (GPL6883), Krug et al[19] used the Affymetrix Human Genome U133 Plus 2.0 Array (GPL570), and Stamova et al[18] used the Affymetrix Human Genome U133 Plus 2.0 Array (GPL570).

In the meta-analyses for GWES and IS, derived from genome-wide expression data from 82 IS cases and 67 controls, 41 DE genes were identified using a random effects model (Table 2). Thirteen of these genes were downregulated and 28 were upregulated. An analysis of functional categories found a significant enrichment for the Gene Ontology Term "Inflammatory Response" and for binding sites for the transcription factor PAX2 (Table 3). Other functional categories, such as Gene Ontology Cellular Components or Molecular Functions, Kyoto