BRIEF COMMUNICATION

Intracerebral Hemorrhage and Ischemic Stroke of Different Etiologies Have Distinct Alternatively Spliced mRNA Profiles in the Blood: a Pilot RNA-seq Study

Cheryl Dykstra-Aiello[1] · Glen C. Jickling[1] · Bradley P. Ander[1] · Xinhua Zhan[1] ·
DaZhi Liu[1] · Heather Hull[1] · Miles Orantia[1] · Carolyn Ho[1] · Boryana Stamova[1,2]

Abstract Whole transcriptome studies have used 3′-biased expression microarrays to study genes regulated in the blood of stroke patients. However, alternatively spliced messenger RNA isoforms have not been investigated for ischemic stroke or intracerebral hemorrhage (ICH) in animals or humans. Alternative splicing is the mechanism whereby different combinations of exons of a single gene produce distinct mRNA and protein isoforms. Here, we used RNA sequencing (RNA-seq) to determine if alternative splicing differs for ICH and cardioembolic, large vessel and lacunar causes of ischemic stroke compared to controls. RNA libraries from 20 whole blood samples were sequenced to 200 M 2×100 bp reads using Illumina sequencing-by-synthesis technology. Differential alternative splicing was assessed using one-way analysis of variance (ANOVA), and differential exon usage was calculated. Four hundred twelve genes displayed differential alternative splicing among the groups (false discovery rate, FDR; $p < 0.05$). They were involved in cellular immune response, cell death, and cell survival pathways. Distinct expression signatures based on usage of 308 exons (292 genes) differentiated the groups ($p < 0.0005$; fold change $> |1.2|$). This pilot study demonstrates that alternatively spliced genes from whole blood differ in ICH compared to ischemic stroke and

✉ Boryana Stamova
  boryana.stamova@ucdmc.ucdavis.edu

[1] Department of Neurology, University of California at Davis, Sacramento, CA, USA

[2] Department of Neurology, MIND Institute Wet Labs, 2805 50th Street, Room 2417, Sacramento, CA 95817, USA

differ between different ischemic stroke etiologies. These results require validation in a separate cohort.

Keywords Ischemic stroke · Intracerebral hemorrhage · Alternative splicing · Biomarkers · RNA-seq · Exon

Introduction

Stroke is diagnosed based on patient history, neurological exam, and brain imaging. Differential diagnosis can be difficult particularly for distinguishing ischemic stroke (IS) from intracerebral hemorrhage (ICH) when imaging is unavailable in the acute setting. Thus, an accurate, inexpensive, and rapid blood test would be useful.

Blood transcriptomes show promise as diagnostic biomarkers and have provided insight into understanding the nature of the immune response following human stroke [1–6]. However, these studies have investigated only a portion of the protein coding transcriptome, since they have used 3′-biased microarrays to measure blood mRNA expression [1–6]. Though these studies demonstrated proof-of-principle, most of the stroke transcriptome which is comprised of all alternatively spliced isoforms remains unstudied in stroke. The importance of alternative splicing is supported by evidence implicating it in the pathogenesis of many diseases [7, 8] .

Alternative splicing is the process whereby exons from a single gene are included or excluded in the final mRNA transcript (Supplementary Figure 1). A single gene can produce several alternatively spliced isoforms which have specific functions in different cells, tissues, developmental stages, and disease states. Thus, the ~20,000 known genes code for >250,000 different mRNAs and proteins. Differential alternative splicing (DAS) is alternative splicing that differs between groups. We hypothesized that DAS would vary for different

causes of IS (cardioembolic, large vessel, and lacunar) and for ICH when compared to each other and to controls.

RNA-seq is a new technology that allows for estimation of expression of each splice variant (Supplementary Figure 1), a significant advance over previous technologies. Because there have been no studies of alternative splicing related to stroke etiology, or for IS versus ICH either in humans or in animal models, we performed this pilot RNA-seq study to examine DAS in whole blood following IS and ICH in humans.

Methods

Stroke patients and control subjects were randomly selected from those recruited at the University of California Davis Medical Center between 2008 and 2012. Stroke patients were chosen to represent the major IS etiologies (cardioembolic, large vessel atherosclerotic, lacunar) or had ICH. IS diagnosis and causes were assessed as described previously [5, 9]. ICH patients had deep ICH (basal ganglia, deep white matter, or thalamus) confirmed by CT and/or MRI brain scans and were associated with hypertension without evidence of vascular malformation, tumor, or aneurysm. Control subjects were selected to match stroke subjects for age, race, sex, and vascular risk factors and had no history of previous stroke or cardiovascular events. Blood from all subjects was drawn into PAXgene tubes between 5.8 and 101.2 h following IS or ICH. RNA from whole blood was isolated as previously described [3]. The UCD Institutional Review Board approved this study and all subjects provided informed consent.

Whole blood RNA was used to prepare mRNA libraries using the TruSeq RNA Sample Prep v2 kit and protocol (Illumina). Two hundred million PE 100-bp RNA-seq reads were obtained from each mRNA library using Illumina Solexa sequencing by synthesis on the Illumina HiSeq 2000. TopHat v2.0.7 (Bowtie v2.0.6) was used with default parameters to map reads to a reference genome (Hg19) and generate *bam* files for analysis [10]. RNA transcript quantification was performed using Hg19 AceView transcripts in the Partek Genomics Suite 6.6 RNA-seq workflow.

The raw reads for genes displaying DAS are shown in Supplementary Table 2 and the raw reads for genes displaying differential exon usage are shown in Supplementary Table 6. They were generated from aligned *bam* files using *featureCounts* against AceView (NCBI 37) [11] with options allowing for any and multiple overlaps [12]. However, they were not used directly for the statistical analysis. Instead, raw aligned reads were normalized, and differential alternatively spliced transcript expression and exon expression quantification were performed using the expectation/maximization (E/M) algorithm (briefly described below) as implemented in Partek Genomics Suite [13]. DAS was determined with one-way ANOVA on Group (Benjamini-Hochberg false discovery rate, FDR; $p<0.05$), and differential exon usage was assessed between each two groups ($p<0.0005$, fold change$>|1.2|$).

Principal components analysis (PCA) and hierarchical clustering were performed in Partek Genomics Suite. Ingenuity Pathway Analysis (IPA®) and DAVID identified regulated pathways and processes as described previously [6].

Results

Subject Demographics

Subject demographics and clinical characteristics are presented in Table 1. Only Caucasian males were studied because of the small group sizes. Age, time since event for IS or ICH, and vascular risk factors were not significantly different between groups. Coverage of a wide range of post-stroke biology was obtained by selecting patients with early (5.8 h) through late (101.2 h) blood draw times after IS and ICH. However, the means were similar between the stroke groups. Cardioembolic IS post-event blood draw times were, on average, 33.7 h; large vessel averaged 47.4 h; lacunar averaged 34.6 h; and ICH averaged 29.4 h (Table 1).

Table 1 Subject demographics and clinical characteristics

	Ischemic stroke			Intracerebral hemorrhage	Controls
	Cardioembolic	Large vessel	Lacunar		
Subjects (total $n=20$)	4	4	4	4	4
Age, years (mean±SD)	62.3±9.6	61.0±8.2	58.9±9.0	60.1±2.3	60.8±9.2
Time since event, h (mean±SD)	33.7±18.9	47.4±47.8	34.6±23.7	29.4±15.5	N/A
Hypertension	4	3	2	3	3
Diabetes	2	2	0	0	1
Hyperlipidemia	3	2	2	0	2
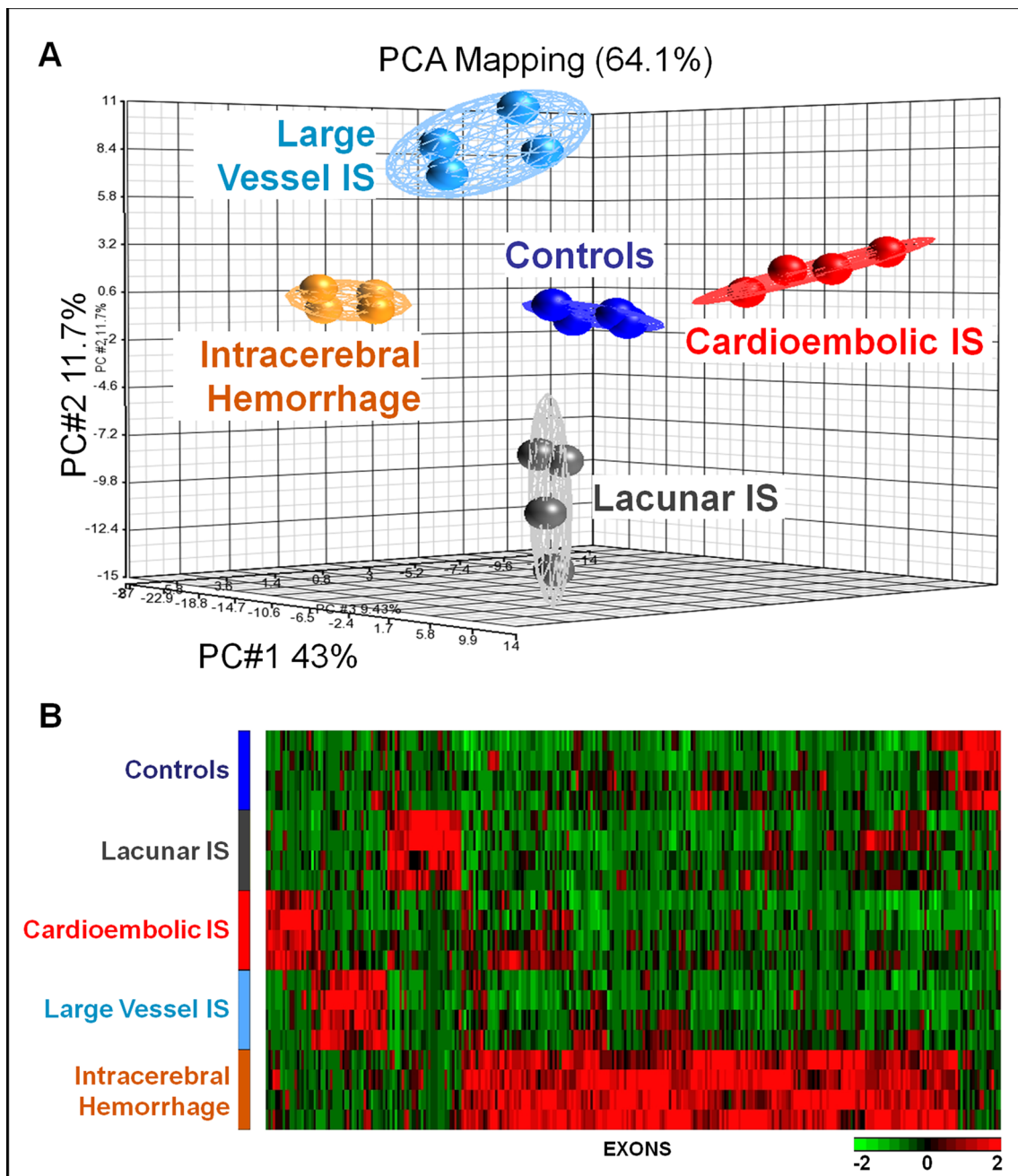
Fig. 1 Principal components analysis (PCA) (Fig. 1a) and unsupervised hierarchical clustering (Fig. 1b) of the 308 exons (292 genes) with differential exon usage among intracerebral hemorrhage (*n*=4), ischemic strokes (IS) (cardioembolic, large vessel, and lacunar) (*n*=12), and control subjects (*n*=4). In Fig. 1a, the expression of the 308 exons is compressed on to three axes in the PCA plot. The three principal components on the PCA plot account for 64.1 % of the variance. In Fig. 1b, exon expression is shown on the *X*-axis and subjects are shown on the *Y*-axis. *Each row on the Y-axis* represents a single individual, with five individuals per group. The dendrograms were removed from this figure. *Red* indicates increased expression. *Green* indicates decreased expression

RNA Sequencing Alignments

RNA sequencing alignment statistics for all samples among the five groups are presented in Supplementary Table 1. Cardioembolic stroke samples had on average, 1.60E+08 alignments; large vessel had 1.65E+08 alignments; lacunar stroke had 1.64E+08 alignments; and ICH and control groups each averaged 1.59E+08 alignments. These data show that there is no bias in the numbers of alignments for any of the five groups.

Distinct Alternative Splicing of Genes in Whole Blood of Stroke Patients and Controls

A total of 412 genes displayed differential alternative splicing (DAS) in the whole blood transcriptomes of the five groups of

patients with ischemic stroke (cardioembolic, large vessel, and lacunar), intracerebral hemorrhage (ICH) and controls (FDR $p<0.05$; Supplementary Table 2, raw reads; Supplementary Table 3, ANOVA results). These 412 genes are those predicted using the E/M algorithm [13] as implemented in Partek, to have DAS for IS (cardioembolic, large vessel, lacunar) versus ICH versus controls. The E/M algorithm probabilistically assigns reads to known isoforms/exons of a gene [13]. Partek then uses a log-likelihood ratio test to identify genes with DAS across samples [13, 14]. The 412 significant genes displaying DAS across IS, ICH, and controls are involved in cellular immunity, cytokine signaling, and cell death and survival pathways (Supplementary Table 4).

Pathways highly over-represented with differentially alternative spliced genes between the five groups included CD28 signaling in T helper cells, CDC42 signaling, Nur77 signaling in T lymphocytes, fMLP signaling in neutrophils, and interferon signaling (Supplementary Table 4). Molecular and cellular functions most highly associated with the differentially alternatively spliced genes were cell death and survival of immune cells, cell-cell signaling, activation and recruitment of leukocytes, antigen-presenting cells, activation of T lymphocytes, adhesion of vascular endothelial cells, and immune response of neutrophils (Supplementary Table 5).

Specific Exon-Usage Profiles for ICH and Different Ischemic Stroke Etiologies

A total of 308 exons from 292 genes were differentially expressed for the three causes of IS (cardioembolic, large vessel, lacunar), ICH, and controls ($p<0.0005$, fold change $>|1.2|$; Supplementary Table 6, raw reads; Supplementary Table 7, ANOVA results). These exons separated the five groups, including the three causes of IS (cardioembolic, large vessel, lacunar), ICH, and controls on principal components analysis (PCA) plots (Fig. 1a) and using unsupervised hierarchical clustering (Fig. 1b). Given that the E/M algorithm uses counts of the numbers of reads on each exon [13, 14], the differential expression of exons across the five groups represents differential exon usage across the five groups. These results are relevant to DAS because DAS results from differential exon usage.

Biological functions and networks represented by genes with differentially expressed exons in each group (Fig. 1b) are summarized in Supplementary Table 8. Cardioembolic stroke genes with differential exon usage were involved in ion binding/transport and cellular assembly/organization. Large-vessel stroke genes were associated with cell death, transcription, and chromatin remodeling. Lacunar stroke genes were associated with cellular compromise, cell cycle, cell death and survival. ICH genes were involved with protein transport and localization (Supplementary Table 8).

Discussion

Although differential alternative splicing (DAS) is implicated in many human diseases, this is the first study to show that DAS differs between intracerebral hemorrhage (ICH), ischemic stroke, and control subjects. In addition, it is the first study to show that DAS differs between different etiologies of ischemic stroke including cardioembolic, large vessel, and lacunar causes. Identification of DAS in RNA from whole blood for specific stroke etiologies and ICH suggests the immune response varies for each condition. This will be important for understanding the pathogenesis of each condition and will be important for developing biomarkers to differentiate ischemic stroke from ICH and for developing biomarkers to differentiate the different causes of ischemic stroke.

This study identified several pathways, molecular functions, and genes previously reported in human ischemic stroke using 3′-biased microarrays [6, 15]. These included actin cytoskeleton signaling, CCR5 signaling in macrophages, NF-κB activation, α-adrenergic signaling, cellular growth and proliferation, cell death and survival, cell morphology, hematopoiesis, hematological system development, and inflammatory response [4, 5, 16, 17]. Moreover, a number of the pathways implicated in different etiologies of ischemic stroke in our previous microarray studies were confirmed in these RNA-seq studies [4, 5, 16, 17].

This study is the first to describe genes with DAS and pathways unique for ICH. Among the genes that differentiated ICH from IS were INPP5D (inositol polyphosphate-5-phosphatase) and ITA4 (integrin alpha 4). The INPP5D enzyme regulates myeloid cell proliferation and programming, and its expression correlates with hemorrhagic transformation of ischemic stroke [18]. ITA4 is involved in leukocyte recruitment after intracerebral hemorrhage [19], and leukocytes are intimately associated with ICH. For example, leukocytes are involved in clotting and interact with injured vessels and brain following ICH [15].

Other genes with DAS associated with ICH in this study included NAV1 (neuron navigator 1), PDGFC (platelet derived growth factor C), and CCM2 (cerebral cavernous malformation 2) which participate in vascular endothelial growth factor (VEGF) signaling, which predisposes the brain to hemorrhage because of new vessel formation [20]. Of interest, mutations of CCM2 cause cerebral cavernous malformations which can lead to intracerebral hemorrhage [21]. Other genes with DAS associated with ICH included EXOSC1 (exosome component 1) and EXOSC9 (exosome component 9) which code for core components of the exosome complex [22]. Although exosomes have been implicated in neuroinflammation, neurodegeneration, and cancer, they have not previously been associated with ICH [23, 24]. Lastly, another gene with DAS associated with ICH included DGCR8 (DiGeorge syndrome critical region 8, a microprocessor complex subunit)

which is involved in the biogenesis of microRNAs [25], which could suggest that miRNAs are involved with differential alternative splicing following ICH.

Study Limitations

Sample sizes in this pilot study were small. Thus, we cannot rule out splicing changes due to vascular risk factors. However, hierarchical clustering of the differentially expressed exons demonstrates separation on diagnosis and not on vascular risk factors (Supplementary Figure 2). Validation of these findings in a separate cohort is needed to confirm the present results. These results are important because they provide evidence for differential alternative splicing in the pathophysiology of the immune response to ischemic stroke and intracerebral hemorrhage and also might provide novel biomarkers for ICH and different causes of IS.

References

1. Moore DF, Li H, Jeffries N, Wright V, Cooper Jr RA, Elkahloun A, et al. Using peripheral blood mononuclear cells to determine a gene expression profile of acute ischemic stroke: a pilot investigation. Circulation. 2005;111(2):212–21.

2. Tang Y, Xu H, Du X, Lit L, Walker W, Lu A, et al. Gene expression in blood changes rapidly in neutrophils and monocytes after ischemic stroke in humans: a microarray study. J Cereb Blood Flow Metab. 2006;26(8):1089–102.

3. Stamova B, Xu H, Jickling G, Bushnell C, Tian Y, Ander BP, et al. Gene expression profiling of blood for the prediction of ischemic stroke. Stroke. 2010;41(10):2171–7.

4. Jickling GC, Xu H, Stamova B, Ander BP, Zhan X, Tian Y, et al. Signatures of cardioembolic and large-vessel ischemic stroke. Ann Neurol. 2010;68(5):681–92.

5. Jickling GC, Stamova B, Ander BP, Zhan X, Tian Y, Liu D, et al. Profiles of lacunar and nonlacunar stroke. Ann Neurol. 2011;70(3): 477–85.

6. Sharp FR, Jickling GC, Stamova B, Tian Y, Zhan X, Liu D, et al. Molecular markers and mechanisms of stroke: RNA studies of blood in animals and humans. J Cereb Blood Flow Metab. 2011;31(7):1513–31.

7. Gamazon ER, Stranger BE. Genomics of alternative splicing: evolution, development and pathophysiology. Hum Genet. 2014;133(6):679–87.

8. Poulos MG, Batra R, Charizanis K, Swanson MS. Developments in RNA splicing and disease. Cold Spring Harb Perspect Biol. 2011;3(1):a000778.

9. Jickling GC, Stamova B, Ander BP, Zhan X, Liu D, Sison SM, et al. Prediction of cardioembolic, arterial, and lacunar causes of cryptogenic stroke by gene expression and infarct location. Stroke. 2012;43(8):2036–41.

10. Trapnell C, Pachter L, Salzberg SL. TopHat: discovering splice junctions with RNA-Seq. Bioinformatics. 2009;25(9):1105–11.

11. Thierry-Mieg D, Thierry-Mieg J. AceView: a comprehensive cDNA-supported gene and transcripts annotation. Genome Biol. 2006;7 Suppl 1:S12.1–4.

12. Liao Y, Smyth GK, Shi W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics. 2014;30(7):923–30.

13. Xing Y, Yu T, Wu YN, Roy M, Kim J, Lee C. An expectation-maximization algorithm for probabilistic reconstructions of full-length isoforms from splice graphs. Nucleic Acids Res. 2006;34(10):3150–60.

14. Marioni JC, Mason CE, Mane SM, Stephens M, Gilad Y. RNA-seq: an assessment of technical reproducibility and comparison with gene expression arrays. Genome Res. 2008;18(9):1509–17.

15. Sharp FR, Jickling GC. Whole genome expression of cellular response to stroke. Stroke. 2013;44(6 Suppl 1):S23–5.

16. Tian Y, Stamova B, Jickling GC, Liu D, Ander BP, Bushnell C, et al. Effects of gender on gene expression in the blood of ischemic stroke patients. J Cereb Blood Flow Metab. 2012;32(5):780–91.

17. Xu H, Tang Y, Liu DZ, Ran R, Ander BP, Apperson M, et al. Gene expression in peripheral blood differs after cardioembolic compared with large-vessel atherosclerotic stroke: biomarkers for the etiology of ischemic stroke. J Cereb Blood Flow Metab. 2008;28(7):1320–8.

18. Jickling GC, Ander BP, Stamova B, Zhan X, Liu D, Rothstein L, et al. RNA in blood is altered prior to hemorrhagic transformation in ischemic stroke. Ann Neurol. 2013;74(2):232–40.

19. Hammond MD, Ambler WG, Ai Y, Sansing LH. alpha4 integrin is a regulator of leukocyte recruitment after experimental intracerebral hemorrhage. Stroke. 2014;45(8):2485–7.

20. Jeney V, Balla G, Balla J. Red blood cell, hemoglobin and heme in the progression of atherosclerosis. Front Physiol. 2014;5:379.

21. Kar S, Samii A, Bertalanffy H. PTEN/PI3K/Akt/VEGF signaling and the cross talk to KRIT1, CCM2, and PDCD10 proteins in cerebral cavernous malformations. Neurosurg Rev. 2014.

22. Shen V, Kiledjian M. A view to a kill: structure of the RNA exosome. Cell. 2006;127(6):1093–5.

23. Gupta A, Pulliam L. Exosomes as mediators of neuroinflammation. J Neuroinflammation. 2014;11:68.

24. Candelario KM, Steindler DA. The role of extracellular vesicles in the progression of neurodegenerative disease and cancer. Trends Mol Med. 2014;20(7):368–74.

25. Gregory RI, Yan KP, Amuthan G, Chendrimada T, Doratotaj B, Cooch N, et al. The microprocessor complex mediates the genesis of microRNAs. Nature. 2004;432(7014):235–40.