Genomics 104 (2014) 163–169

Contents lists available at ScienceDirect

Genomics

journal homepage: www.elsevier.com/locate/ygeno

Expression profile based gene clusters for ischemic stroke detection

Mateusz G. Adamski [a,b], Yan Li [a], Erin Wagner [a], Hua Yu [a], Chloe Seales-Bailey [a], Steven A. Soper [c,e], Michael Murphy [d], Alison E. Baird [a,*]

[a] Department of Neurology, SUNY Downstate Medical Center, 450 Clarkson Ave., Brooklyn, NY 11203, USA
[b] Department of Neurology, Jagiellonian University Medical College, ul. Botaniczna 3, Krakow 31-501, Poland
[c] Department of Biomedical Engineering, University of North Carolina, 220 E Cameron Ave., Chapel Hill, NC 27514, USA
[d] Department of Mechanical Engineering, Louisiana State University, 3357 Highland Rd., Baton Rouge, LA 70802, USA
[e] Department of Chemistry, University of North Carolina, 220 E Cameron Ave., Chapel Hill, NC27514, USA

ARTICLE INFO

Article history:
Received 2 April 2014
Accepted 7 August 2014
Available online 15 August 2014

Keywords:
Stroke
Gene expression
qPCR
Biomarker
Transcriptome

ABSTRACT

In microarray studies alterations in gene expression in circulating leukocytes have shown utility for ischemic stroke diagnosis. We studied forty candidate markers identified in three gene expression profiles to (1) quantitate individual transcript expression, (2) identify transcript clusters and (3) assess the clinical diagnostic utility of the clusters identified for ischemic stroke detection. Using high throughput next generation qPCR 16 of the 40 transcripts were significantly up-regulated in stroke patients relative to control subjects ($p < 0.05$). Six clusters of between 5 and 7 transcripts were identified that discriminated between stroke and control (p values between 1.01e−9 and 0.03). A 7 transcript cluster containing *PLBD1*, *PYGL*, *BST1*, *DUSP1*, *FOS*, *VCAN* and *FCGR1A* showed high accuracy for stroke classification (AUC = 0.854). These results validate and improve upon the diagnostic value of transcripts identified in microarray studies for ischemic stroke. The clusters identified show promise for acute ischemic stroke detection.

1. Introduction

Stroke is a leading cause of death and disability in the community and new diagnostics and therapeutics are greatly needed [1]. Inflammation and immune response after stroke impact significantly on tissue and clinical outcome [2,3]. Application of molecular and cellular approaches to study the immune system in stroke may offer new diagnostic and therapeutic approaches.

Using microarrays that contained between 22,000 and 54,000 oligonucleotide probes, genomic profiling has been applied to the circulating leukocytes of human stroke patients [4–7]. Peripheral blood mononuclear cells (PBMCs) [4,7] and whole blood samples [5,6] were used for these studies. In three independent analyses 22, 18 and 9 transcripts showed utility for stroke detection [4–6]. In these studies ribonucleic acid (RNA) was sampled between 3 and 72 h after stroke onset. Different microarrays from two companies (Affymetrix and Illumina) were used and therefore signal intensity was assessed differently for each study. Despite these methodological and experimental differences there was overlap among the transcripts identified and panels were able to be applied between the study cohorts [4–7].

These microarray studies raised the possibility of added diagnostic utility in stroke from genomic profiling of circulating leukocytes to clinical and neuroimaging information during the time window for thrombolytic therapy [8–11]. Expression changes were seen as early as 3 h post-stroke and persisted at 5 and 24 h[5]. However, further translation and application of these microarray results have been hindered by data normalization issues, cost, high turnaround time and the limited availability of arrays. While providing unprecedented coverage of the transcriptome, microarray data are also limited by low sensitivity and low accuracy for transcripts expressed at low levels [12,13].

The majority of these stroke-related transcripts were not validated with standard quantitative polymerase chain reactions (qPCR) — the gold standard for measuring gene expression. qPCR-based approaches are more likely than microarrays to be applied and developed for rapid assays and automated point of care systems that would be needed for early stroke diagnosis [14,15]. Compared to microarrays qPCR approaches are characterized by shorter assay turnaround times and high sensitivity, with a theoretical limit of detection of a single copy of messenger ribonucleic acid (mRNA) target [16]. Until now standard reverse transcription (RT)-qPCR has been feasible for studying 6 genes at most from typical clinical samples.

Abbreviations: HT RT-qPCRs, high throughput reverse transcription quantitative polymerase chain reactions; IS, ischemic stroke; NIHSS, National Institutes of Health Stroke Scale; CBCs, complete blood counts; WBC, white blood cell; FDR, false discovery rate.

* Corresponding author at: Department of Neurology, SUNY Downstate Medical Center, 450 Clarkson Ave., Box 1274, Brooklyn, NY 11203, USA. Fax: +1 718221 5761.
E-mail address: alison.baird@downstate.edu (A.E. Baird).

http://dx.doi.org/10.1016/j.ygeno.2014.08.004
0888-7543/

Recently next generation microfluidic high throughput qPCR approaches have become available. These methods, known as high throughput RT-qPCR (HT RT-qPCR) or nanofluidic qPCR, permit the rapid quantification of multiple transcripts using small sample volumes [17,18] with very high sensitivity. Plates can contain up to 96 samples in which 96 transcripts can be simultaneously studied in 9216 reactions. We have applied HT RT-qPCR to forty candidate markers identified in the three prior gene expression profiling studies to (1) quantitate individual transcript expression, (2) identify transcript clusters and (3) assess the clinical diagnostic utility of the clusters identified for ischemic stroke detection.

2. Methods

2.1. Study subjects

Peripheral blood samples were obtained from 18 ischemic stroke patients admitted to the University Hospital of Brooklyn at SUNY Downstate Medical Center and at Long Island College Hospital and 15 gender and race matched control subjects recruited from the local community. The median time of blood draw was 36 h post stroke onset. Stroke was diagnosed according to World Health Organization stroke criteria. The Institutional Review Board at the State University of New York (SUNY) approved the study and all study participants or their authorized representatives gave full and signed informed consent.

The study inclusion criteria were: over 18 years of age and acute ischemic stroke. The exclusion criteria were: current immunological diseases, taking steroid or immunosuppressive therapies, severe allergies, acute infection and severe anemia. The following clinical data were recorded: age, gender, race, self-reported risk factors, National Institutes of Health Stroke Scale (NIHSS) score in the stroke subjects and complete blood counts (CBCs), including total white blood cell count and white cell differential counts. Hypertension was defined as a prior (at any time in the past) diagnosis of hypertension by the subject's physician or currently receiving treatment for hypertension. Diabetes was defined as a past medical history of known diabetes mellitus. Coronary artery disease was defined as a physician-diagnosed past history of ischemic heart disease or angina. Hyperlipidemia was defined as a past history of documented elevation in total cholesterol (>200 mg/dl). Smoking was defined as current or prior smoking. Atrial fibrillation was defined as a past or current history of physician-diagnosed atrial fibrillation.

Table 1
Clinical and laboratory characteristics of patients and controls.

	All (n = 33)	Stroke (n = 18)	Control (n = 15)	p
Factor				
Age	65.4 ± 14.3	71.6 ± 13.0	58.1 ± 12.3	0.004
Gender — male	14 (42)	7 (39)	7 (47)	0.9
Race — black	30 (91)	17 (94)	13 (87)	0.9
Risk factors				
Hypertension	28 (85)	17 (94)	11 (73)	0.2
Diabetes	15 (45)	8 (39)	7 (53)	0.6
Coronary artery disease	8 (24)	5 (28)	3 (20)	0.9
Smoking history	7 (21)	5 (28)	2 (13)	0.6
Atrial fibrillation	4 (12)	4 (22)	0 (0)	0.2
Hyperlipidemia	16 (48)	8 (44)	8 (53)	0.9
Medications				
Diuretics	9 (27)	6 (33)	3 (15)	0.6
ACEIs/ARBs	9 (27)	7 (39)	2 (13)	0.2
Beta blockers	21 (64)	14 (78)	7 (47)	0.1
Calcium channel blockers	8 (24)	5 (28)	3 (20)	0.9
Anti-thrombotics	18 (54)	10 (55)	8 (53)	1.0
Statins	14 (42)	7 (39)	7 (47)	0.9
WBC count (10^9 cells/l)	6.9 ± 2.4	7.45 ± 2.2	6.18 ± 2.6	0.2
Stroke-related				
Time of blood draw (hours)	N/A	36.0 (23.0, 48.0)	N/A	N/A
Infarct volume (mm^3)	N/A	5404.0(1207.0, 22,870.0)	N/A	N/A
NIHSS score	N/A	7.5 (4.2, 10.0)	N/A	N/A

Results are mean ± SD and median (interquartile range) for continuous factors and numbers (%) for categorical factors. ACEI — angiotensin converting enzyme inhibitor, ARB — angiotensin receptor blocker, WBC — white blood cell, N/A — not applicable, NIHSS — National Institutes of Health Stroke Scale.

Table 2
Comparison of 41 transcripts between stroke and control subjects.

Transcript	Cellular source	Fold change	p value	Adjusted p value[a]	Adjusted p value[b]
CD163	PBMC[4]	2.22	0.069	0.14	1.0
PLBD1	PBMC[4]	3.18	0.0034	0.03	0.14
ADM	PBMC[4]	1.85	0.0066	0.03	0.27
KIAA0146	PBMC[4]	1.21	0.43	0.52	1.0
APLP2	PBMC[4]	1.08	0.56	0.62	1.0
NPL	PBMC[4], WB[5]	1.67	0.094	0.16	1.0
FOS	PBMC[4]	2.64	0.043	0.10	1.0
TLR2	PBMC[4]	1.37	0.57	0.62	1.0
NAIP	PBMC[4]	1.71	0.24	0.34	1.0
CD36	PBMC[4]	2.11	0.29	0.10	1.0
DUSP1	PBMC[4]	2.89	0.033	0.10	1.0
ENTPD1	PBMC[4]	2.03	0.039	0.10	1.0
VCAN	PBMC[4], WB[6]	2.36	0.058	0.13	1.0
CYBB	PBMC[4]	2.61	0.0083	0.04	0.34
IL13RA1	PBMC[4]	1.58	0.10	0.16	1.0
LTA4H	PBMC[4]	1.61	0.20	0.30	1.0
ETS2	PBMC[4], WB[5]	2.86	0.017	0.07	0.70
CD14-1	PBMC[4]	1.93	0.065	0.14	1.0
CD14-2	PBMC[4]	1.39	0.74	0.78	1.0
BST1	PBMC[4]	6.42	0.0035	0.03	0.14
CD93	PBMC[4]	2.11	0.00086	0.02	0.03
PILRA	PBMC[4]	1.29	0.56	0.62	1.0
FCGR1A	PBMC[4]	3.28	0.076	0.14	1.0
CKAP4	WB[5]	1.93	0.0040	0.03	0.14
S100A9	WB[5]	3.84	0.0014	0.02	0.06
MMP9	WB[5,6]	2.21	0.10	0.16	1.0
S100P	WB[5]	2.67	0.0399	0.10	1.0
F5-1	WB[5]	2.14	0.034	0.10	1.0
FPR1	WB[5]	1.79	0.07	0.14	1.0
S100A12	WB[5,6]	2.93	0.000593	0.02	0.02
RNASE2	WB[5]	1.06	0.84	0.86	1.0
ARG1	WB[5,6]	1.34	0.34	0.42	1.0
CA4	WB[5,6]	1.74	0.17	0.27	1.0
LY96	WB[5,6]	1.41	0.27	0.36	1.0
SLC16A6	WB[5]	1.64	0.23	0.34	1.0
HIST2H2AA3	WB[5]	1,48	0.25	0.34	1.0
BCL6	WB[5]	0.97	0.58	0.62	1.0
PYGL	WB[5]	2.55	0.0059	0.03	0.24
CCR7	WB[6]	0.995	0.96	0.96	1.0
IQGAP1	WB[6]	1.67	0.04	0.10	1.0
ORM1	WB[6]	1.28	0.31	0.40	1.0

Wilcoxon rank sum tests and t tests used for analyses. The superscripts in the Cellular source column refer to References [4], [5] and [6] from which the three expression profiles and 40 transcripts had been identified.
[a] FDR method.
[b] Bonferroni method.

2.2. Primer selection and development

40 transcripts identified in 3 previously published studies [4–6] were selected for analysis (Supplementary Table S1 and Supplementary Fig. S1). The 3 studies had identified 9, 18 and 22 genes within panels with some overlap among the studies. *Hox 1.11*, the transcript identified in Tang et al.'s study [5], was not studied because it is a non-coding RNA sequence. Hypothetical protein FLJ22662 Laminin A motif from the Moore list [4] is now termed phospholipase B domain containing 1 (*PLBD1*) according to current nomenclature. Two variants of *CD14* were studied to give a total of 41 transcripts that were tested. The