Original Article

DOI: 10.5582/bst.2018.01150

Combined Biomarkers Composed of Environment and Genetic Factors in Stroke

Yingying Wang[1], Jianfeng Liu[2], Hongyan Wu[1,*], Yunpeng Cai[1,*]

[1] *Research Center for Biomedical Information Technology, Shenzhen Institutes of Advanced Technologies, Chinese Academy of Sciences, Shenzhen, China;*
[2] *Department of Neurology, The First Affiliated Hospital of Harbin Medical University, Harbin, China.*

Summary **It was widely accepted that stroke onset was the result of interactions between environment and genetic factors. However, the combined biomarkers covering environment and genetic factors and their interplay information in stroke were still lacking. In this study, we proposed a framework to identify the targeting or indicating role each factor played in the combined stroke biomarkers. A combined set of 36 biomarkers were identified based on evaluation and importance scores. Validations on three independent microarray data sets justified that the obtained markers were pervasively effective in discriminating stroke patients of different stages from healthy people on genetic levels. 8 and 3 genetic factors were identified as biomarkers in the acute and recovery phases of stroke, respectively. For example, the expression changing of *SERPINH1* only appeared in the acute phase of stroke showing its targeting role in the combined biomarker. Compared with this, 11 genetic factors such as *MMP9* were found to be differentially expressed in both acute and recovery phases of stroke showing their indicating roles in stroke. Functional analyses further revealed that the biomarkers could be grouped into 4 closely related processes of stroke including prevention, occurrence, processing, and recovery, respectively. These results indicated that the adoption of interactions between environment and genetic factors would be helpful in selecting robust and biologically relevant biomarkers, which cast a new insight for stroke biomarker identification.**

Keywords: Combined biomarker, genetic factor, environment factor, interaction, stroke

1. Introduction

Stroke, as one of the leading causes of death and disability in the world, has attracted lots of research efforts on the analyses of its underlying mechanisms on 'Omics levels continuously since the 20th century (*1*). Of which, the identification framework of biomarkers that measure stroke on different molecular levels has resulted in considerable enthusiasm for their wide usage in diagnosis and prognostication. A biomarker can be any measurement made on a biological system in theory, however, the stroke biomarkers typically refer to environment factors (EFs) and genetic factors (GFs). Many researches had been performed on EF(s) or GF(s) levels to explore the occurrence, development, and prognosis of stroke (*2,3*). For example, EFs such as ursolic acid (*4*) and GFs such as *ALOX5* (*5*) were shown to be closely related to the happening of stroke. Biomarkers can be generally classified into target factors and indicator factors, according to their causal relationship with the investigated disease. A target factor was one that played an important role in causing stroke, which meant that the changing of this factor would not only reflect the patient's condition, but also be a possible treatment target. In comparison, an indicator factor was one that distinguished stroke patients from healthy people, without necessarily being the cause of the disease. Usually, indicator factors were advantageous in the ability to discriminate potential

Released online in J-STAGE as advance publication August 29, 2018.

*Address correspondence to:
Dr. Hongyan Wu, and Dr. Yunpeng Cai, Shenzhen Institutes of Advanced Technologies, Chinese Academy of Sciences, 1068 Xueyuan Boulevard, University Town of Shenzhen, Xili Nanshan, Shenzhen 518055, China.
E-mail: hy.wu@siat.ac.cn; yp.cai@siat.ac.cn

patients and the ease of detection, but provided little useful information for disease treatment or prevention. On the contrary, target factors were favored for pathology or possible treatments, but may not be good for patient diagnosis or risk prediction. Ideally, a good biomarker set should contain the advantages of both types of factors and reflect the interaction(s) between them.

It was widely accepted during the past decades that the onset of stroke was the result of interactions between EFs and GFs (*6*). However, no combined biomarker covering both EFs and GFs on multiple omics levels were identified on stroke. The interplay(s) information between EFs and GFs under different conditions including several complex diseases were validated and stored in databases such as miREnvironment (*7*) and PEMDAM (*8*) with the development of bioinformatics which made the analyses on combined EF-GF biomarkers of stroke possible.

In this study, we proposed a framework to identify the targeting or indicating role each factor played in the combined stroke biomarkers selected based on bioinformatics analyses. The information of stroke related factors on both genetic level (including genes and miRNAs) and environment level were downloaded from several public databases. A combined set of 36 biomarkers were then chosen by an evaluation score for each candidate EF and an importance score calculated using the relationships between GFs. We further carried out validation experiments on three independent data sets with the selected biomarkers, which confirmed that the obtained markers were pervasively effective in discriminating stroke patients of different stages from healthy people. It was interesting to find out that the obtained biomarkers could further be grouped into 4 categories related to the prevention, occurrence, processing, and recovery of stroke, respectively. The classification of target and indicator factors in each combined biomaker made the relationships between environment and genetic factors clearer, which may provide a new sight for stroke prevention, treatment, and recovery.

2. Materials and Methods

2.1. *Materials*

Several raw data sets were built based on the data extracted from public databases as follows:

(a) Stroke-related diseases/symptoms: The following key words were used to perform the EF and GF searches according to the information from Comparative Toxicogenomics Database (CTD) (*9*): 'Stroke', 'Infarction, Middle Cerebral Artery', 'Cerebral Infarction', 'Brain Infarction', 'Lateral Medullary Syndrome', 'Brain Stem Infarctions', 'Infarction, Anterior Cerebral Artery', 'Infarction, Posterior Cerebral Artery', and 'Dementia, Multi-Infarct'.

(b) Stroke-related raw EFs: 4,833 stroke EFs extracted from CTD using stroke-related diseases/ symptoms (as listed in (a)) as key words.

(c) Stroke-related raw genes: 287,171 interactions were found between the 4,833 stroke-related raw EFs and 23,472 genes. These genes were marked as 'stroke-related raw genes' in the following analyses.

(d) Stroke-related raw miRNAs (SRMI): 168 miRNAs were extracted from HMDD (*10*), PhenomiR (*11*), and PEMDAM using stroke-related diseases/ symptoms (as listed in (a)) as key words. Each miRNA was assigned a score calculated using the following equation:

$$W_m = S_h + S_{ph} + S_{pe} ,$$

Of which, if one miRNA was found to be related to stroke in HMDD or PEMDAM, the value of S_h or S_{pe} in Equation 1 was 1.0, otherwise, the value was assigned 0. Similarly, if one miRNA was found to be related to stroke in PhenomiR, S_{ph} was calculated as the product of the number of supportive literature and the number of validated types (the maximum number was 2.0, including over-regulation and down-regulation).

2.2. *Candidate EFs Selections based on GFs*

The relationships between EFs and miRNAs were downloaded from PEMDAM and MiREnvironment. The EFs in the 'stroke-related raw EFs' were left if at least one interaction was found. For these EFs, a score *Ep*, based on the interactions between GFs (including miRNAs and genes) was calculated using the following:

$$\mathrm{E_p} = \begin{cases} \max, p \in miRTarBase \\ \sum_{1}^{x} p_x, p \notin miRTarBase \end{cases}$$

Of which, if the interaction was validated in miRTarBase (*12*), *Ep* was assigned the maximum of p_x• p_x represented the times the interaction predicted by 10 different miRNA target computational prediction algorithms (*13*).

The EFs with at least one *Ep* score over 0 were selected as candidate EFs (C-EF).

2.3. *Evaluation Score for EF*

For any C-EF *i*, an evaluation score was calculated as follows:

$$EE(i) = ER(i) * ES(i) ,$$

A higher score indicated a closer relationship between the EF and stroke. Of which, the relation score *ER* (EF relation score) was calculated as follows:

$$\mathrm{ER}(i) = \sum_{i=1}^{n} (\sum_{i=1}^{n1} E_g \sum_{i=1}^{n2} E_m \sum_{i=1}^{n3} E_{\mathrm{p}}) \ ,$$

E_g was the score based on gene level, which was calculated using the interactions between EFs and genes. Of which, n_i represented the number of interactions, n_s represented the number of species, and n_l represented the number of supportive literature. E_m was the score based on miRNA level, which was calculated using the interactions between EFs and pre-miRNAs: t_n was the times each interaction between EF and miRNAs appears in the database of PEMDAM and MiREnvironment. w_m was calculated using the equation described above.

$$E_g = n_i * n_s * n_l \ ,$$

$$\mathrm{E_m} = \begin{cases} \mathrm{w}_m, i \in SRMI \\ \sum_{1}^{n} t_n, i \notin SRMI \end{cases}$$

For each C-EF i, a disease specificity score ES (EF specificity score) based on the biological network analyses was calculated as follows:

$$\mathrm{ES(i)} = \begin{cases} \mathrm{rank(d}_i) / \max(d_i), i \in C-EF \\ 1, i \notin C-EF \end{cases}$$

Of which, d_i represents the degrees of EF i in the EF-disease network built based on the relationships between EFs and diseases from PEMDAM and miREnvironment. If the candidate EF was not included in the C-EF data set, the maximum value 1.0 was assigned for it. A higher ES score indicated a higher specificity of the EF.

2.4. *Important Score for GFs*

For each EF in C-EF, a network was built using its related miRNAs and genes based on the targeting information (including validated and predicted interactions between miRNAs and genes) between them.

The R package 'igraph' was used to perform the Google PageRank analyses on the miRNA-mRNA network. The PageRank score for each node in the network was used to measure the importance based on counting the number and quality of links to a node. The miRNAs and mRNAs were ranked after each of them was assigned a calculated score .

2.5. *Validation based on Independent Data Sets*

To validate availability of the combined biomarkers, three independent microarray data sets from NCBI GEO were chosen to check the relationships between GFs and stroke (no such public data set for EFs were available currently) as listed in Table 1. All the samples of the three data sets were human blood. The Wilcoxon rank sum test was performed for each GF identified as a factor in any combined biomarker in this study. A GF with p-value not above the threshold we set was considered to be differentially expressed between the stroke patients and controls.

3. Results

3.1. *Combined biomarkers for Stroke*

There were 229 EFs in 'stroke-related raw EFs' selected as 'filtered stroke-related EFs' after the filter step based on stroke-related miRNAs. 36 EFs were finally defined as 'candidate EFs (C-EFs)'. One C-EF was thus considered as one factor of the combined biomarkers if it interacted with each C-EF with the maximum PageRank score (See Table 2 for details). The features of combined biomarkers were characterized from three aspects based on literature search results: (a) effect: the positive, negative, or dual effects on the processes of stroke; (b) structure: the role (indicator or target) EF/GF played in stroke; (c) mechanism: the interaction types (induce or inhibit) between the EF and GF(s) in the combined biomarkers. All the features were evaluated based on the target factor in each combined biomarker.

3.2. *Validation Results*

The expression changes of all the GFs in the combined biomarkers were checked. Any GF with at least one p-value not above 0.1 was considered as differentially expressed GFs (see details in Table 3). 8 and 3 genetic factors were identified as biomarkers in the acute and recovery phases of stroke, respectively. For example, the expression change of *SERPINH1* only appeared in the acute phase of stroke showing its targeting role in the combined biomarker. Compared with this, 11 genetic factors such as *MMP9* were found to be differentially expressed in both acute and recovery phases of stroke showing their indicated roles in stroke.

3.3. *Hiberarchy model of Combined bio-markers*

The 36 combined biomarkers could thus be divided into 8 different groups according to the three aspects mentioned above. The biomarkers in each group were divided into sub-groups according to their features. The mean EE scores were calculated for each sub-group as shown in Figure 1. The EE mean values of sub-groups and the statistical significance were calculated using student t-test. Results showed that the differences between 'induce' and 'inhibit' sub-groups on the 'mechanism' level was significant with a p-value of 0.09605 indicating to us that 'mechanism'

Table 1. List of validated data sets used in this study

GF type	Stroke Condition	NCBI GEO ID	Number of Stroke Patients (Case)	Number of Healthy People (Controls)
Gene	Acute Phase	GSE16561	39	24
Gene/miRNA	Recovery Phase	GSE22255	20	20
miRNA	Acute Phase	GSE55937	24	24

Table 2. List of combined biomarkers

Rank	C-EF	GF(s)	Effect	Structure (target-indicator)	Mechanism
1	Dexamethasone	hsa-mir-30e	Positive, Negative	EF-GF	Induce
2	Acetaminophen	hsa-mir-122	Positive	EF-GF	Induce
3	Vitamin E	hsa-mir-15b	Positive	Unknown	Unknown
4	Cisplatin	hsa-mir-642	Negative	GF-EF	Induce
5	Cocaine	hsa-let-7d	Negative	GF-EF	Induce
6	Cadmium	hsa-mir-146a	Negative	EF-GF	Inhibit
7	Bortezomib	hsa-mir-122	Positive	EF-GF	Induce
8	Gemcitabine	hsa-mir-149	Negative	Unknown	Unknown
9	Nicotine	hsa-mir-21	Positive	EF-GF	Induce
10	Metformin	hsa-mir-21	Positive	GF-EF	Inhibit
11	Ethanol	hsa-mir-21	Positive	EF-GF	Induce
12	Nitric Oxide	hsa-mir-155	Positive	GF-EF	Induce
13	DDT	*NOS2, STAT3*	Negative	EF-GF	Inhibit, Induce
14	Docetaxel	hsa-mir-100, hsa-mir-101-1, hsa-mir-126, hsa-mir-130a, hsa-mir-16-1/2, hsa-mir-194-1, hsa-mir-195, hsa-mir-212, hsa-mir-30a, hsa-mir-34a, hsa-mir-7-1	Positive	GF-EF	Inhibit
15	Hemin	hsa-mir-126, hsa-mir-130a, hsa-mir-18b	Positive	Unknown	Inhibit, Induce
16	Letrozole	hsa-let-7f-1/2	Negative	EF-GF	Induce
17	Curcumin	*ALOX5*	Positive	EF-GF	Inhibit
18	Bromocriptine	hsa-mir-550-1	Negative	EF-GF	Induce
19	Arsenic	hsa-mir-222	Negative	EF-GF	Induce
20	Sulindac sulfide	*ATF3*, *PTGS2(COX2)*, hsa-mir-17, hsa-mir-21	Positive	EF-GF	Inhibit, Induce
21	Imatinib Mesylate	hsa-mir-451	Positive	GF-EF	Inhibit
22	Ursolic acid	hsa-mir-21	Positive	EF-GF	Inhibit
23	Vitamin D (VitD)	hsa-mir-22	Positive	EF-GF	Induce
24	Bleomycin	hsa-mir-21	Negative	GF-EF	Induce
25	Oxaliplatin	hsa-mir-21	Negative	GF-EF	Inhibit
26	Glucose	hsa-mir-133a-1/2, hsa-mir-146a/b, hsa-mir-451	Negative	EF-GF, GF-EF	Unknown
27	Gefitinib	hsa-mir-222, hsa-mir-30b	Negative	EF-GF	Inhibit
28	Topotecan	hsa-mir-142, hsa-mir-34b	Positive	Unknown	Unknown
29	Lead	*ADORA1*, *IGF1*, *IL1B*, hsa-mir-146a, hsa-mir-21, hsa-mir-222	Negative	Unknown	Inhibit
30	Paroxetine	hsa-mir-30a	Positive	GF-EF	Inhibit
31	Genistein	hsa-mir-151, hsa-mir-27a	Positive	EF-GF	Inhibit
32	Decitabine	hsa-mir-145	Positive	EF-GF	Induce
33	Cytarabine	*EDN1*, *F2*, hsa-mir-29a, hsa-mir-30c-1	Negative	GF-EF	Inhibit
34	Fludarabine	*MMP9*	Positive	EF-GF	Induce
35	Polycyclic Aromatic Hydrocarbons (PAH)	*CCL2*	Negative	Unknown	Unknown
36	Sorafenib	hsa-mir-122, *SERPINH1*	Negative	GF-EF	Induce

could be used as the first analysis level. There were 16 biomarkers in the induce group and 12 in the inhibit group (8 combined biomarkers were not included in the following analyses since their mechanism were not clear).

To explore the second analysis level, we divided the biomarkers in one sub-group into sub-sub-group according to their 'effect' and 'structure' features. EE values and PageRank scores were used to test the differences between these sub-sub-groups. For the sub-groups containing combined biomarkers with induce mechanism, significant difference was found between two sub-sub-groups (positive and negative) according to the 'effect' feature with a *p*-value of 0.06137 using PageRank score. Compared with this, no significant difference was found in the 'inhibit' sub-group with *p*-value of 0.2826 (mean PageRank scores were 0.2588571 and 0.3178 for 'positive' and 'negative', respectively).

Similarly to the above analyses, we checked all the possible third analyses levels using statistical test. For the 8 combined biomarkers in 'Mechanism (induce)-Effect (positive)', only 1 biomarker was shown to have the structure of 'GF-EF'. The *p*-value of

Table 3. *p*-value of GFs in combined biomarkers

GF Name	*P*-value in Acute Phase (GSE16561, GSE55937)	*P*-value in Recovery Phase (GSE22255)
ADORA1	3.64E-10	2.76E-10
ALOX5	0.007267241	0.000472032
ATF3	1.43E-15	3.36E-05
CCL21	2.86E-15	0.006144956
EDN1	5.83E-06	1.69E-07
IGF1	7.16E-16	5.41E-09
IL1B	6.92E-11	0.040175085
MMP9	6.08E-10	1.34E-07
NOS2	7.16E-16	1.45E-11
PTGS2	7.97E-08	0.383413282
SERPINH1	0.000431237	0.120699706
STAT3	0.001638866	0.00513072
F2	--	0.09132
hsa-mir-145	0.000921373	--
hsa-mir-122	0.014428574	--
hsa-mir-550	0.019369559	--
hsa-mir-642	0.066476228	--
hsa-mir-30e	0.071189216	--
hsa-mir-21	0.077797772	0.06035
hsa-mir-16-2	0.092850883	--
hsa-mir-22	0.675265662	0.04298
hsa-mir-142	0.645619886	0.0675

-- Some of the GFs were not detected in all the three data sets due to the original platform, as a result, no p-value could be calculated.

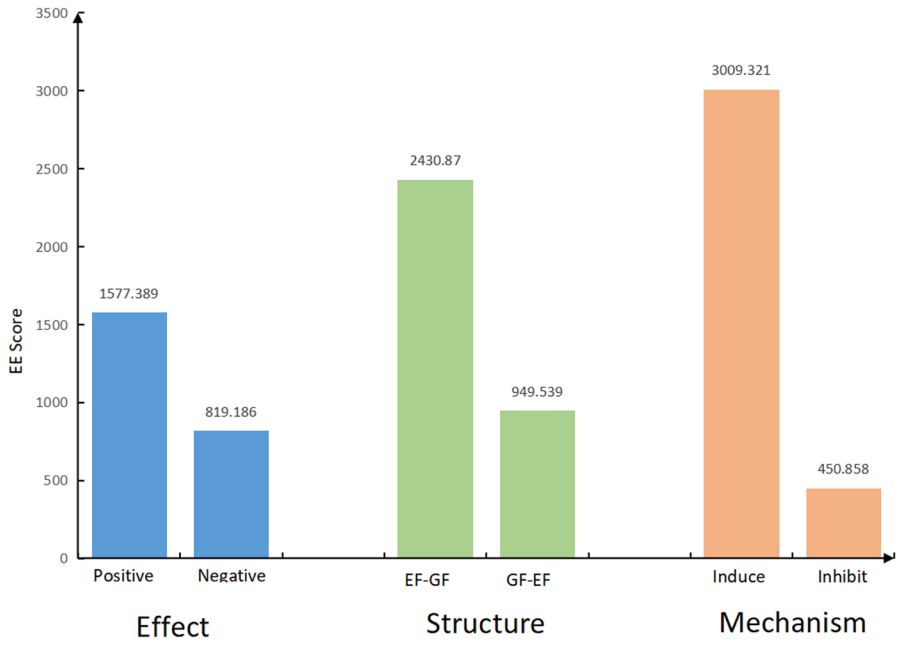

Figure 1. Comparison of mean EE scores on different levels. The mean EE scores were calculated for each sub-group.

binomial distribution analysis was 0.01563 indicating a significant difference. Analysis of the 8 combined biomarkers in 'Mechanism (induce)-Effect (negative)' showed no significant difference on structure level. Taken together, the analyses hierarchy model of the three features were then fixed as 'Mechanism-Effect'.

It was interesting to find that the significant differences were found on induce and positive level, which may indicated these combined biomarkers might play their roles in a forward way rather than feedback.

3.5. *Function of Combined biomarkers*

3.5.1. *Combined biomarkers for Stroke Prevention and Damages Mitigation*

Seven combined biomarkers in the 'Mechanism (inhibit)-Effect (positive)' model were shown to play a role in protecting against stroke or mitigating stroke-induced damages. All the EFs including curcumin, ursolic acid, genistein, metformin, docetaxel, imatinib mesylate, and

paroxetine functioned as stroke prevention or damage mitigation factors (See Table S1 of Supplemental Data for details, *http://www.biosciencetrends.com/action/getSupplementalData.php?ID=27*).

Hsa-mir-21, hsa-mir-27a, hsa-mir-151, hsa-mir-100, hsa-mir-101-1,hsa-mir-130a, hsa-mir-16-1/2, hsa-mir-194-1, hsa-mir-195, hsa-mir-212, hsa-mir-30a, hsa-mir-451, and hsa-mir-34a were shown to be over-regulated; while hsa-mir-126 and hsa-mir-7-1 were shown to be down-regulated in young stroke patients (*14,15*).

The relationships between EF and GF in each biomarker was inhibit, which meant that the expression or activation of indicator factor was inhibited by the target factor. For example, curcumin was shown to perform the damage mitigation role by inhibiting the catalytic activities of *ALOX5* (*16*), which may attenuate neuro-protection following focal cerebral ischemia (*5*). hsa-mir-21 was shown to be suppressed by ursolic acid in human glioblastoma cell lines U251 (*17*). Genistein was shown to play its roles through inhibition of mir-27a and mir-151 in different diseases (*18,19*). Metformin was shown to improve skeletal muscle insulin resistance by inhibiting mir-21 expression (*20*). Over-regulation of mir-100 could prevent docetaxel chemoresistance in patients with lung adenocarcinoma (*21*). Docetaxel resistance was associated with increased expression of mir-34a, and decreased expression of mir-100, mir-7, mir-16, mir-30a, and mir-126 in human breast cancer cells (*22*). mir-195 was a negative regulator in the resistance of *DU145/DOC* cells to docetaxel (*23*). hsa-mir-451 was observed to be down-regulated in imatinib-resistant chronic myeloid leukemia patients (*24*). mir-30a may limit the effects of paroxetine by targeting *BDNF* (*25*).

3.5.2. *Combined biomarkers for Causing Stroke*

There were 8 combined biomarkers proved to be related to the occurrence of stroke. It was interesting to find out that all these biomarkers were in the model of 'Mechanism (induce)-Effect (negative)', of which 4 target factors were EFs while the other 4 target factors were GFs.

At least one of the EF or GF in each biomarker were proved to be a risk factor causing stroke including dexamethasone, letrozole, bromocriptine, arsenic, cisplatin, cocaine, bleomycin, sorafenib, and *SERPINH1* (*26*) (See Table S1 of Supplemental Data for details, *http://www.biosciencetrends.com/action/getSupplementalData.php?ID=27*).

Of all the GFs, previous studies showed that hsa-mir-30e, hsa-mir-550-1, hsa-mir-222, hsa-mir-642, hsa-mir-let-7d, hsa-mir-21, hsa-mir-122 were up-regulated while hsa-mir-7f-1, hsa-mir-7f-2, and hsa-mir-15b were down-regulated in young stroke patients (*14,27,28*).

The increased expression of indicator factors may be induced by the target factor in the same combined biomarker. One study showed that dexamethasone-induced *IEC-6* cells differentiation caused a 2.5-fold increase in mir-30e expression, and upon beta-catenin siRNA transfection, mir-30e increased 1.3-fold (*29*). After letrozole treatment for 48 hours, all let-7 subtypes showed a trend toward increased expression (*29*). mir-550 was confirmed to be significantly up-regulated between the group of bromocriptine-treated and untreated prolactinomas (*30*). mir-222 was up-regulated in arsenic-transformed human lung epithelial *BEAS-2B* cells indicating its role in arsenic-induced tumor growth (*31*). The increased expression of mir-642 could increase the sensitivity of cisplatin in cell lines and advanced bladder cancer (*28*). Cocaine up-regulated let-7d in zebrafish embryos (*32*). The repressing of mir-21 could attenuate bleomycin-induced pulmonary fibrosis (*33*). mir-122 was shown to be up-regulated during apoptosis induced by bortezomib and sensitized hepatocellular carcinoma cells to sorafenib (*27,34*).

3.5.3. *Combined biomarkers for Stroke Processes*

Four combined biomarkers in the model of 'Mechanism (inhibit)-Effect (negative)' showed adverse effects on stroke. Gefitinib, cadmium, oxaliplatin, cytarabine, hsa-mir-146a (*35*), and *EDN1* (*36*) were proved to be negative factors of stroke (See Table S1 of Supplemental Data for details, *http://www.biosciencetrends.com/action/getSupplementalData.php?ID=27*).

hsa-mir-222, hsa-mir-21, hsa-mir-29a and hsa-mir-30c-1 were shown to be up-regulated in young stroke patients (*14*), while hsa-mir-30b, and hsa-mir-146a were shown to be down-regulated in young stroke patients (*14,35*).

mir-30b and mir-222 were shown to be down-regulated by gefitinib (*37*). The expression of mir-146a was negatively correlated with exposure to cadmium (*38*). The over-expression of mir-21 could protect CRC cells from oxaliplatin-induced apoptosis and increase the proliferative capacity (*39*). The deregulated expression of mir-29a and mir-30c was shown to contribute to the sensitivity to cytarabine (*40*).

3.5.4. *Combined biomarkers for Stroke Recovery*

There were 8 combined biomarkers in the 'Mechanism (induce)-Effect (positive)' model. All the GFs including hsa-mir-122, hsa-mir-21, hsa-mir-22, hsa-mir-145, hsa-mir-155, and *MMP9* in the 7 biomarkers (hsa-mir-122 and hsa-mir-21 were involved in two bio-markers) were proved to have increased expression in stroke patients (*14,15,41*).

Four EFs including bortezomib, nicotine, vitamin D (VitD), and fludarabine were proved to contribute to neuroprotection after stroke. The other 4 EFs including acetaminophen, ethanol, decitabine, and nitric oxide

were proved to be associated with better outcome for stroke. For example, acetaminophen was indicated to improve outcome in patients with stroke and fever without dramatically lowering body temperature in one clinical trial study. Low to moderate levels of ethanol could not only decrease the risk of stroke, but also reduce post-ischemic sequelae. Decitabine was widely used in sickle cell anemia which was closely related to stroke. Treatment with nitric oxide was shown to improve functional recovery after stroke (See Supplemental Data for details).

Most of the GFs were considered to be induced by EFs in this model based on their relationships in other conditions since their relationships were still lacking in stroke-related fields. Mir-122 was reported as a novel biomarker of acetaminophen toxicity and was shown to be up-regulated during apoptosis induced by bortezomib (*27,42,43*). Besides, the up-regulation of mir-21 induced by nicotine could promote EMT transforming growth factor beta (*TGF-β*) dependently in human esophageal cancer (*44*). Chronic ethanol feeding was shown to enhance mir-21 induction during liver regeneration while inhibiting proliferation in rats (*45*). Mir-22 was induced by VitD and contributed to its antiproliferative, antimigratory and gene regulatory effects in colon cancer cells (*46*). Decitabine was shown to induce the expression of mir-145 (*47*). *MMP9* was shown to be involved in chronic lymphocytic leukemia cell response to fludarabine (*48*). Knockdown of mir-155 could significantly decrease the production of nitric oxide (*49*), which was the indicated factor.

4. Discussion

The identification of complex diseases' biomarkers were considered to be one of the traditional topics in cardiovascular related fields. For example, the identification and evaluation of genetic biomarkers such as IL-6, TNF-α which were measured using blood as samples for heart failure had been performed which were considered to be of great importance since these results had improved the diagnosis and treatments greatly in the clinic. Compared with this, similar work performed on stroke had always been questioned since most of the samples were human blood, which may not reflect the changes of the ischemic regions due to the existence of the blood brain barrier (BBB). Considering this, the stroke related EFs attracted more and more attention since their roles could be both risk factors and/or biomarkers. These EFs could be divided into two different groups according to the their features as follows: (a) risk factor (clinical environment factor): cardiovascular risk factors widely accepted in clinic such as hypertension, diabetes mellitus, smoking, alcohol consumption, and air pollution, *etc*. (b) biomarker (toxicology environment factor): chemicals, drugs, and small molecules, *etc*. However, it was widely accepted that the occurrence of stroke was caused by both genetic and environment reasons. In recent years, research began to exploit the interplay(s) between EF(s) and GF(s). For example, one study showed that the combined effects of the *MTHFR* 3'-UTR polymorphisms and tHcy/folate levels might contribute to stroke prevalence (*50*). However, the relationship analyses on a systematic scale for stroke combined biomarkers were still lacking especially on multiple omics levels.

It was widely believed that genetic factors may coordinately support the influence of macro or micro environment factors. Exploring this type of genetic biomarkers would be beneficial in understanding the mechanism of disease onset and aiding the development of therapy or prevention schemes. Many mRNAs were identified as biomarkers of stroke in former studies using different frameworks such as differential analyses, network analyses, *etc*. Besides mRNAs, miRNAs were widely considered as genetic factors since they could play important roles in stroke through regulating their target genes. One miRNA might regulate hundreds to thousands of mRNAs, which made the parallel analyses performed using both mRNAs and miRNAs data sets possible. Based on these concerns, many researchers including us had constructed the combined biomarkers of stroke containing both miRNAs and mRNAs on the genetic level in previous studies. However, the close relationships between miRNAs and EFs had not been fully integrated in such studies.

In this study we demonstrated that by exploiting the environment-genetic interactions we do achieve a set of biomarkers that were robust across different data sets and with clear biological relevance. This suggested that knowledge about environment risk factors and environment-genetics would serve as a good guideline for exploring new combined biomarkers and the framework proposed in this study can be a useful tool for this purpose.

Acknowledgements

This work was supported by grants from the Basic Research Program of Shenzhen, [JCYJ20160229203627477] [JCYJ20160229193120432], National Natural Science Foundation of China, [61702496], and the Application Research and Development of Guangdong, [2015B010129012].

References

1. Markus HS, Bevan S. Mechanisms and treatment of ischaemic stroke – insights from genetic associations. Nat Rev Neurol. 2014; 10:723-730.
2. Wang Y, Cai Y. Obtaining human ischemic stroke gene expression biomarkers from animal models: a cross-species validation study. Sci Rep. 2016; 6:29693.
3. Peters JL, Perlstein TS, Perry MJ, McNeely E, Weuve J.

Cadmium exposure in association with history of stroke and heart failure. Environ Res. 2010; 110:199-206.
4. Li L, Zhang X, Cui L, Wang L, Liu H, Ji H, Du Y. Ursolic acid promotes the neuroprotection by activating Nrf2 pathway after cerebral ischemia in mice. Brain Res. 2013; 1497:32-39.
5. Singh DP, Chopra K. Flavocoxid, dual inhibitor of cyclooxygenase-2 and 5-lipoxygenase, exhibits neuroprotection in rat model of ischaemic stroke. Pharmacol Biochem Behav. 2014; 120:33-42.
6. Szolnoki Z, Melegh B. Gene-gene and gene-environment interplay represent specific susceptibility for different types of ischaemic stroke and leukoaraiosis. Curr Med Chem. 2006; 13:1627-1634.
7. Yang Q, Qiu C, Yang J, Wu Q, Cui Q. miREnvironment database: providing a bridge for microRNAs, environmental factors and phenotypes. Bioinformatics. 2011; 27:3329-3330.
8. Li J, Wu Z, Cheng F, Li W, Liu G, Tang Y. Computational prediction of microRNA networks incorporating environmental toxicity and disease etiology. Sci Rep. 2014; 4:5576.
9. Davis AP, Grondin CJ, Johnson RJ, Sciaky D, King BL, McMorran R, Wiegers J, Wiegers TC, Mattingly CJ. The comparative toxicogenomics database: update 2017. Nucleic Acids Res. 2017; 45:972-978.
10. Li Y, Qiu C, Tu J, Geng B, Yang J, Jiang T, Cui Q. HMDD v2.0: a database for experimentally supported human microRNA and disease associations. Nucleic Acids Res. 2014; 42:1070-1074.
11. Ruepp A, Kowarsch A, Schmidl D, Buggenthin F, Brauner B, Dunger I, Fobo G, Frishman G, Montrone C, Theis FJ. PhenomiR: a knowledgebase for microRNA expression in diseases and biological processes. Genome Biol. 2010; 11:R6.
12. Hsu SD, Tseng YT, Shrestha S, *et al*. miRTarBase update 2014: an information resource for experimentally validated miRNA-target interactions. Nucleic Acids Res. 2014 ;42:78-85.
13. Li J, Zhang Y, Wang Y, Zhang C, Wang Q, Shi X, Li C, Zhang R, Li X. Functional combination strategy for prioritization of human miRNA target. Gene. 2014; 533:132-141.
14. Tan KS, Armugam A, Sepramaniam S, Lim KY, Setyowati KD, Wang CW, Jeyaseelan K. Expression profile of microRNAs in young stroke patients. PloS One. 2009; 4:e7689.
15. Maitrias P, Metzinger-Le Meuth V, Massy ZA, M'Baya-Moutoula E, Reix T, Caus T, Metzinger L. MicroRNA deregulation in symptomatic carotid plaque. J Vasc Surg. 2015; 62:1245-1250.
16. Hong J, Bose M, Ju J, Ryu JH, Chen X, Sang S, Lee MJ, Yang CS. Modulation of arachidonic acid metabolism by curcumin and related beta-diketone derivatives: effects on cytosolic phospholipase A(2), cyclooxygenases and 5-lipoxygenase. Carcinogenesis. 2004; 25:1671-1679.
17. Wang J, Li Y, Wang X, Jiang C. Ursolic acid inhibits proliferation and induces apoptosis in human glioblastoma cell lines U251 by suppressing TGF-β1/miR-21/PDCD4 pathway. Basic Clin Pharmacol Toxicol. 2012; 111:106-112.
18. Xia J, Cheng L, Mei C, Ma J, Shi Y, Zeng F, Wang Z, Wang Z. Genistein inhibits cell growth and invasion through regulation of miR-27a in pancreatic cancer cells. Curr Pharm Des. 2014; 20:5348-5353.
19. Chiyomaru T, Yamamura S, Zaman MS, Majid S, Deng G, Shahryari V, Saini S, Hirata H, Ueno K, Chang I, Tanaka Y, Tabatabai ZL, Enokida H, Nakagawa M, Dahiya R. Genistein suppresses prostate cancer growth through inhibition of oncogenic microRNA-151. PloS One. 2012; 7:e43812.
20. Wang J, Gao Y, Duan L, Wei S, Liu J, Tian L, Quan J, Zhang Q, Liu J, Yang J. Metformin ameliorates skeletal muscle insulin resistance by inhibiting miR-21 expression in a high-fat dietary rat model. Oncotarget. 2017; 8:98029-98039.
21. Feng B, Wang R, Chen LB. MiR-100 resensitizes docetaxel-resistant human lung adenocarcinoma cells (SPC-A1) to docetaxel by targeting Plk1. Cancer Lett. 2012; 317:184-191.
22. Kastl L, Brown I, Schofield AC. miRNA-34a is associated with docetaxel resistance in human breast cancer cells. Breast Cancer Res Treat. 2012; 131:445-454.
23. Ma X, Zou L, Li X, Chen Z, Lin Q, Wu X. MicroRNA-195 regulates docetaxel resistance by targeting clusterin in prostate cancer. Biomed Pharmacother. 2018; 99:445-450.
24. Soltani I, Douzi K, Gharbi H, Benhassine I, Teber M, Amouri H, Ben Hadj Othman H, Farrah A, Ben Lakhel R, Abbes S, Menif S. Downregulation of miR-451 in Tunisian chronic myeloid leukemia patients: potential implication in imatinib resistance. Hematology. 2017; 22:201-207.
25. Angelucci F, Croce N, Spalletta G, Dinallo V, Gravina P, Bossu P, Federici G, Caltagirone C, Bernardini S. Paroxetine rapidly modulates the expression of brain-derived neurotrophic factor mRNA and protein in a human glioblastoma-astrocytoma cell line. Pharmacology. 2011; 87:5-10.
26. Yagil C, Hubner N, Monti J, Schulz H, Sapojnikov M, Luft FC, Ganten D, Yagil Y. Identification of hypertension-related genes through an integrated genomic-transcriptomic approach. Circ Res. 2005; 96:617-625.
27. Manfe V, Biskup E, Rosbjerg A, Kamstrup M, Skov AG, Lerche CM, Lauenborg BT, Odum N, Gniadecki R. miR-122 regulates p53/Akt signalling and the chemotherapy-induced apoptosis in cutaneous T-cell lymphoma. PloS One. 2012; 7:e29541.
28. Nordentoft I, Birkenkamp-Demtroder K, Agerbaek M, Theodorescu D, Ostenfeld MS, Hartmann A, Borre M, Orntoft TF, Dyrskjot L. miRNAs associated with chemo-sensitivity in cell lines and in advanced bladder cancer. BMC Med Genomics. 2012; 5:40.
29. Cho S, Mutlu L, Zhou Y, Taylor HS. Aromatase inhibitor regulates let-7 expression and let-7f-induced cell migration in endometrial cells from women with endometriosis. Fertil Steril. 2016; 106:673-680.
30. Wang C, Su Z, Sanai N, Xue X, Lu L, Chen Y, Wu J, Zheng W, Zhuge Q, Wu ZB. microRNA expression profile and differentially-expressed genes in prolactinomas following bromocriptine treatment. Oncol Rep. 2012; 27:1312-1320.
31. Wang M, Ge X, Zheng J, Li D, Liu X, Wang L, Jiang C, Shi Z, Qin L, Liu J, Yang H, Liu LZ, He J, Zhen L, Jiang BH. Role and mechanism of miR-222 in arsenic-transformed cells for inducing tumor growth. Oncotarget. 2016; 7:17805-17814.
32. Gao Y, Lu J, Zhang Y, Chen Y, Gu Z, Jiang X. Baicalein attenuates bleomycin-induced pulmonary fibrosis in rats

through inhibition of miR-21. Pulm Pharmacol Ther. 2013; 26:649-654.
33. Lopez-Bellido R, Barreto-Valer K, Sanchez-Simon FM, Rodriguez RE. Cocaine modulates the expression of opioid receptors and miR-let-7d in zebrafish embryos. PloS One. 2012; 7:e50885.
34. Bai S, Nasser MW, Wang B, Hsu SH, Datta J, Kutay H, Yadav A, Nuovo G, Kumar P, Ghoshal K. MicroRNA-122 inhibits tumorigenic properties of hepatocellular carcinoma cells and sensitizes these cells to sorafenib. J Biol Chem. 2009; 284:32015-32027.
35. Li SH, Chen L, Pang XM, Su SY, Zhou X, Chen CY, Huang LG, Li JP, Liu JL. Decreased miR-146a expression in acute ischemic stroke directly targets the Fbxl10 mRNA and is involved in modulating apoptosis. Neurochem Int. 2017; 107:156-167.
36. Zhang L, Sui R. Effect of SNP polymorphisms of EDN1, EDNRA, and EDNRB gene on ischemic stroke. Cell Biochem Biophys. 2014; 70:233-239.
37. Garofalo M, Romano G, Di Leva G, *et al*. EGFR and MET receptor tyrosine kinase-altered microRNA expression induces tumorigenesis and gefitinib resistance in lung cancers. Nat Med. 2011; 18:74-82.
38. Bollati V, Marinelli B, Apostoli P, Bonzini M, Nordio F, Hoxha M, Pegoraro V, Motta V, Tarantini L, Cantone L, Schwartz J, Bertazzi PA, Baccarelli A. Exposure to metal-rich particulate matter modifies the expression of candidate microRNAs in peripheral blood leukocytes. Environ Health Perspect. 2010; 118:763-768.
39. Bullock MD, Pickard KM, Nielsen BS, Sayan AE, Jenei V, Mellone M, Mitter R, Primrose JN, Thomas GJ, Packham GK, Mirnezami AH. Pleiotropic actions of miR-21 highlight the critical role of deregulated stromal microRNAs during colorectal cancer progression. Cell Death Dis. 2013; 4:e684.
40. Russ AC, Sander S, Luck SC, Lang KM, Bauer M, Rucker FG, Kestler HA, Schlenk RF, Dohner H, Holzmann K, Dohner K, Bullinger L. Integrative nucleophosmin mutation-associated microRNA and gene expression pattern analysis identifies novel microRNA - target gene interactions in acute myeloid leukemia. Haematologica. 2011; 96:1783-1791.
41. Laterza OF, Lim L, Garrett-Engele PW, Vlasakova K, Muniappa N, Tanaka WK, Johnson JM, Sina JF, Fare TL, Sistare FD, Glaab WE. Plasma MicroRNAs as sensitive and specific biomarkers of tissue injury. Clin Chem. 2009; 55:1977-1983.
42. Antoine DJ, Dear JW, Lewis PS, Platt V, Coyle J, Masson M, Thanacoody RH, Gray AJ, Webb DJ, Moggs JG, Bateman DN, Goldring CE, Park BK. Mechanistic biomarkers provide early and sensitive detection of acetaminophen-induced acute liver injury at first presentation to hospital. Hepatology. 2013; 58:777-787.
43. Yang X, Salminen WF, Shi Q, Greenhaw J, Gill PS, Bhattacharyya S, Beger RD, Mendrick DL, Mattes WB, James LP. Potential of extracellular microRNAs as biomarkers of acetaminophen toxicity in children. Toxicol Appl Pharmacol. 2015; 284:180-187.
44. Zhang Y, Pan T, Zhong X, Cheng C. Nicotine upregulates microRNA-21 and promotes TGF-β-dependent epithelial-mesenchymal transition of esophageal cancer cells. Tumour Biol. 2014; 35:7063-7072.
45. Dippold RP, Vadigepalli R, Gonye GE, Hoek JB. Chronic ethanol feeding enhances miR-21 induction during liver regeneration while inhibiting proliferation in rats. Am J Physiol Gastrointest Liver Physiol. 2012; 303:733-743.
46. Alvarez-Diaz S, Valle N, Ferrer-Mayorga G, Lombardia L, Herrera M, Dominguez O, Segura MF, Bonilla F, Hernando E, Munoz A. MicroRNA-22 is induced by vitamin D and contributes to its antiproliferative, antimigratory and gene regulatory effects in colon cancer cells. Hum Mol Genet. 2012; 21:2157-2165.
47. Chinnathambi S, Wiechert S, Tomanek-Chalkley A, Winter MC, Bickenbach JR. Treatment with the cancer drugs decitabine and doxorubicin induces human skin keratinocytes to express Oct4 and the OCT4 regulator mir-145. J Dermatol. 2012; 39:617-624.
48. Amigo-Jimenez I, Bailon E, Ugarte-Berzal E, Aguilera-Montilla N, Garcia-Marco JA, Garcia-Pardo A. Matrix metalloproteinase-9 is involved in chronic lymphocytic leukemia cell response to fludarabine and arsenic trioxide. PloS One. 2014; 9:e99993.
49. Zhang RL, Zhang ZG, Chopp M. Targeting nitric oxide in the subacute restorative treatment of ischemic stroke. Expert Opin Investig Drugs. 2013; 22:843-851.
50. Kim JO, Park HS, Ryu CS, Shin JW, Kim J, Oh SH, Kim OJ, Kim NK. Interplay between 3'-UTR polymorphisms in the methylenetetrahydrofolate reductase (MTHFR) gene and the risk of ischemic stroke. Sci Rep. 2017; 7:12464.

(Received July 2, 2018; Revised August 18, 2018; Accepted August 24, 2018)