Evidence-based mechanistic integrative modeling of stroke risk factors: A translational research study

Mahdi Habibi-Koolaee[1]* , Leila Shahmoradi[2] , Sharareh R. Niakan Kalhori[2] , Hossein Ghannadan[3], Erfan Younesi[4]

[1]Department of Education, Golestan University of Medical Sciences, Gorgan, Iran
[2]Department of Health Information Management, Faculty of Allied Medical Sciences, Tehran University of Medical Sciences, Tehran, Iran
[3]Clinical Research Development Unit (CRDU), Sayad Shirazi Hospital, Golestan University of Medical Sciences, Gorgan, Iran
[4]Information Technology for Translational Medicine, L-4362 Esch-sur-Alzette, Luxembourg, Luxembourg

Article Info	A B S T R A C T
	Introduction: According to global statistics, stroke is known as the main health problem in the world. Many clinical and molecular research, which are stored in the different repository with the various format have been conducted in the area of stroke domain. The heterogeneity of these research data does not make a comprehensive view of the disease. Recently, translational research has been developed to fill the gap between these studies. In this study, we used the integrative disease modeling method to model the underlying mechanism of stroke risk factors. **Material and Methods:** This study was conducted in three steps: data gathering, model construction, and mechanism discovery. First, using semantic and information retrieval tools, we extracted the cause and effect statement from the literature to create the mechanistic model, and the validated molecular data to evaluate the constructed model. Then, the integrative model was created and evaluated. Finally, we used Gene Set Enrichment Analysis to identify the main biological process and signaling pathways in the mechanism of the disease. **Results:** In the evidence-based information retrieval from the literature, 1837 causal statement was extracted. The initial network was created with 648 nodes (molecular, clinical, and environmental factors) and 1837 edges (interactions). Also, 51 genes/proteins and nine single nucleotide polymorphisms were matched with data in the model. The inflammatory response, response to lipid, regulation of body fluid levels, and regulation of response to stress, complement and coagulation cascades, and PPAR signaling pathway were the main biological processes and signaling pathways enriched in GSEA analysis. **Conclusion:** This study showed that we can identify the underlying mechanism of stroke risk factors and use a proper strategy to prevent it, using Integrative Disease Modeling.

INTRODUCTION

Stroke is the second cause of death worldwide [1] has more than 150 known causes and about 25%-30% of stroke cases are classified as heterogeneous [2]. Genetic factors, also are involved in both strokes with known and heterogeneous causes [3]. Therefore, a variety of environmental and genetic factors are involved in brain stroke.

The complexity of stroke is not just about the gene-environment interaction. The lack of access to brain tissue in live patients, the complexity of the anatomical structure of the brain, and the lack of biomarkers to predict stroke, all contribute to the complexity of the disease. In recent years, the macro pharmaceutical industry has invested more than $ 1 billion in discovering and producing new drugs for stroke, but they have not yet marketed a drug other

than tissue plasminogen activator (TPA), and most of these investments have failed [4, 5]. TPA has a specific list of indications and contraindication. The efficacy of this drug depends on the time consumed. Therefore, strategies based on primary or secondary prevention are prioritized [6]. These failures have led the pharmaceutical industries to reduce their activities in the production of neuroprotective drugs [3]. Neuroprotective drugs are drugs that prevent or reduce the progress of the disease. Interestingly, neuroprotective drugs in pre-clinical and in-vitro investigations have a very good and repeatable therapeutic effect but do not show a satisfactory performance when tested in humans. Therefore, the transfer of the therapeutic effects of new drugs from model animals to the human body requires a deep understanding of the molecular mechanisms involved in the stroke in the human brain. With deep knowledge and insight from the interaction of molecular and environmental factors, it is possible to identify the biological pathways that lead to stroke and target new drugs or preventive strategies.

Since stroke is a complex disease of the nervous system, identifying the causes of its occurrence, treatment, and prevention requires a comprehensive investigation of scientific findings at the pre-clinical, clinical, and post-clinical level. This disease not only has different phenotypes but also has very complex causative mechanisms and various risk factors [3]. One of the modern and expanding strategies, which recently proposed to overcome the complexity of the disease mechanism is translational research, which is done by integrating various data and modeling the mechanism of disease at all physio-pathological levels [7]. Various research groups and researchers have conducted different studies on the area of stroke domain. Each study is published and stored in different databases, with different and often heterogeneous formats. This heterogeneity does not result in a comprehensive view of the mechanism of the disease. In this study, we pointed out that the utilization of translational research techniques explains the underlying mechanism of stroke risk factors at the molecular level and translates into clinical practice.

MATERIAL AND METHODS

This study was done in three steps, including data gathering, model construction, and mechanism discovery.

Step 1: Data gathering

Selection of the main stroke risk factors

In this study, we used the integrative modeling of risk factors mechanism underlying stroke. Since, stroke has many risk factors, to avoid the complexity of the model, we have to select one or two main risk factors.

To do this, we conducted a hospital-based study. In the study population, hypertension, dyslipidemia, and diabetes significantly increased the risk of stroke. For model construction, we selected dyslipidemia and diabetes. The result of this study was published in [8] and [9].

Information retrieval from biomedical literature

Since the mechanism of the disease is usually a cause-and-effect relationship, it was necessary to access this information to model the mechanisms of stroke. We used text mining and semantic tools to retrieve information and extract knowledge from the literature. First, we integrated Stroke Ontology (STO) (https://bioportal.bioontology.org/ontologies/STO-DRAFT) with KNIME, a text-mining tool [10]. Then using STO root class terms and based on two main stroke risk factors (dyslipidemia and diabetes) we conducted a search strategy on PubMed abstract; accessed 25.02.2017. We manually filtered 545 retrieved abstracts based on the relevancy of their content, of which 157 abstracts were selected. We extract the casual statement underlying stroke from the full text of these selected papers. The process of information retrieval and knowledge extraction illustrated in Fig 1.

Molecular data gathering for model evaluation

Furthermore, we need experimental omics data (gene and protein expression and Single Nucleotide Polymorphisms (SNPs) data) to biologically evaluate the constructed model. Gene expression data gathered from the Gene Expression Omnibus (GEO) database. The GEO database accession number was GSE43618 (PMID: 23559260), GDS4521 (PMID: 22453632), GSE55937 (PMID: 24911610), and GSE37587 (PMID: 25124890). Also, we found some experimental data from (PMID: 15630028), (PMID: 16395289), (PMID: 17997827), and (PMID: 27407070). The human protein atlas was used to gather protein expression data. We used the GWAS catalog and Array Express database to gather SNPs related to stroke. The full list of gene expression, protein expression, and SNPs data could be found in supplementary file 1, 2, and 3, respectively.

Step 2: Model construction and validation

In this study, we create a computational cause and effect network of all molecular and non-molecular data of stroke pathophysiology using the Cytoscape tool [11]. This model illustrated the interaction between various factors extracted from the literature. This network contains nodes (clinical, molecular, and environmental data) and relations between them which shows causal relations of various factors.

The created model was based on published literature, which may be shown in the elusive biological process.

To realize the constructed model, we evaluated the model using curated molecular data (explained in Step 1). Using the Cytoscape tool, we matched our model with these data.
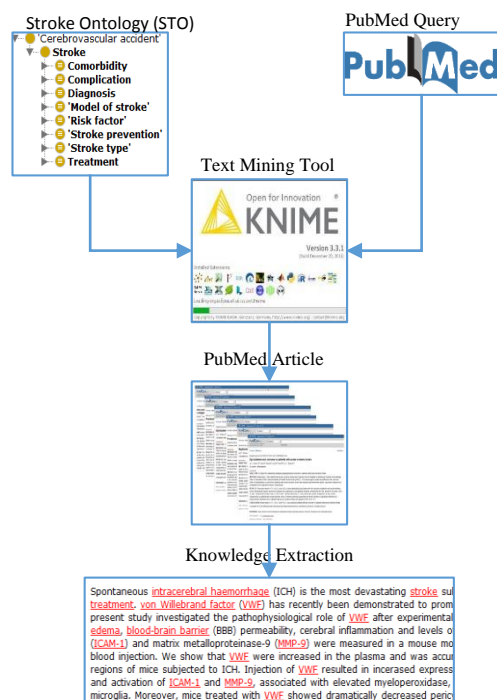

Fig 1: The process of extraction of the cause and effect statements from the literature, using ontology and data mining tools. As the process showed, we firstly integrated the stroke ontology (STO) with KNIME. After that, using the PubMed query, we retrieved evidence related to dyslipidemia and diabetes in stroke. we manually select related evidence. Finally, we extract a casual statement for model creation.

Step 3: mechanism discovery

In this step, we used Gene Set Enrichment Analysis (GSEA) to identify the main signaling pathway and biological process. For enrichment analysis, we used Gene Ontology (GO) biological process gene set and Kyoto Encyclopedia for Gene and Genome (KEGG) gene set from the Molecular Signature Database (MSigDB). After adjustment enriched processes and signaling pathways with our constructed model, the underlying mechanism of dyslipidemia and diabetes in stroke patients was explained.

RESULTS

Knowledge extraction

About 1837 statement filtered after STO-supported manually retrieval of a casual statement from the literature. Of each statement, we extract the cause, effect, and their relation to save in a separate file for using in the next step.

Model construction and validation

The initial network of different factors underlying stroke comprises 648 nodes (molecular, clinical, and environmental factors) and 1837 edge (interactions). Using this casual model, we aimed to identify the underlying mechanism which increases the risk of stroke. Fig 2 illustrates the causal model of selected risk factors of stroke.
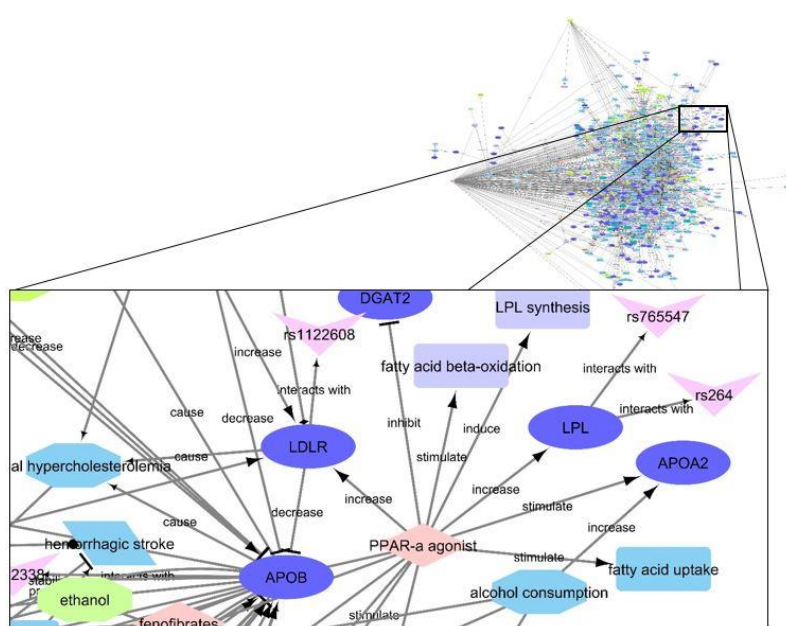

Fig 2: Cause and effect model of stroke risk factors. The initial model of stroke underlying mechanism consists of causal factors (nodes) and the relation (edges) between them.

To evaluate the constructed model, curated molecular data (see Method) were matched with the model. As shown in Table 1, 51 expressed Gene and Protein were matched within the model.

Table 1: The count and name of the expressed gene and protein were matched within the model.

Expression data	Count	Description
Gene	21	TFPI; F3; LCN2; TBXA2R; ITGA2B; MIR223; CXCL8; MIR15A; MIR21; SELP; MIR126; MMP9; MIR20B; MAPK1; APOE; CD36; TNF; FABP4; IL10; PDGFA; VEGFA
Protein	36	APOB; ADIPOQ; VCAM1; PON1; APOA1; ICAM1; APOC3; FGA; MMP8; LCN2; SAA1; SERPINA3; TGFB1; LPA; CCL2; CRP; CDC42; FABP4; F13B; MPO; PROC; F2; AGT; CD36; SERPINC1; VWF; F7; F12; AOC3; ITGA2B; APOA2; PLTP; CHIT1; PLA2G7; MMP9; APOE

Also, nine SNPs were found within the model. Table 2 shows the list of SNPs with their related gene/protein name.

The main model of stroke was filtered by these genes/proteins and their first neighbors. The new network contains validated multiple causal factors which have 209 nodes and 819 edges.

Table 2: The Single nucleotide polymorphisms(SNPs) and related gene/protein were matched within the model.

SNP ID	Gene/Protein	Gene/Protein Description
rs2592902	CRP	C-reactive protein
rs2048327 rs10455872	LPA	Lipoprotein(a)
rs1799963	F2	Coagulation factor II, thrombin
rs2022309	F3	Coagulation factor III, tissue factor
rs9326246	APOA1	Apolipoprotein A1
rs6041 rs555212	F7	Coagulation factor VII
rs562338	APOB	Apolipoprotein B

Gene Set Enrichment Analysis

Using GSEA analysis of GO (biological process) on our gene/protein list (retrieved from the literature), we identify a list of the enriched biological process involved in the mechanism of dyslipidemia and diabetes in stroke (Table 3). In accordance with the stroke model, biological process of the inflammatory response, response to lipid, regulation of body fluid level, and regulation of stress response enriched.

Table 3: The top biological process from Gene Set Enrichment Analysis (GSEA) of Gene Ontology (GO): biological process data set

Gene Set Name	# Genes in Gene Set (K)	# Genes in Overlap (k)	k/K	p-value	FDR q-value
REGULATION_OF_RESPONSE_TO_WOUNDING	413	20	0.0484	4.47E-28	1.98E-24
RESPONSE_TO_EXTERNAL_STIMULUS	1821	29	0.0159	1.18E-27	2.63E-24
REGULATION_OF_RESPONSE_TO_EXTERNAL_STIMULUS	926	23	0.0248	8.78E-26	1.3E-22
INFLAMMATORY_RESPONSE	454	19	0.0419	1.97E-25	2.19E-22
RESPONSE_TO_LIPID	888	22	0.0248	1.4E-24	1.15E-21
REGULATION_OF_BODY_FLUID_LEVELS	506	19	0.0375	1.56E-24	1.15E-21
REGULATION_OF_PLASMA_LIPOPROTEIN_PARTICLE_LEVELS	45	11	0.2444	9.74E-24	6.17E-21
RESPONSE_TO_WOUNDING	563	19	0.0337	1.18E-23	6.54E-21
REGULATION_OF_RESPONSE_TO_STRESS	1468	24	0.0163	1.07E-22	5.25E-20
RESPONSE_TO_OXYGEN_CONTAINING_COMPOUND	1381	23	0.0167	7.2E-22	3.2E-19

GSEA pathway analysis resulted in a list of the enriched significant pathway (Table 4). Complement and coagulation cascades and PPAR signaling pathway corresponded with the stroke model.

Fig 3 illustrates the mechanisms by which oxidative stress led to inflammatory responses and increases the risk of stroke.

Table 4: The top pathway from Gene Set Enrichment Analysis (GSEA) of KEGG data set

| Gene Set Name	# Genes in Gene Set (K)	# Genes in Overlap (k)	k/K	p-value	FDR q-value
COMPLEMENT_AND_COAGULATION_CASCADES	69	10	0.1449	3.58E-19	6.66E-17
PPAR_SIGNALING_PATHWAY	69	7	0.1014	1.38E-12	1.28E-10
PATHWAYS_IN_CANCER	328	8	0.0244	3.02E-9	1.87E-7
CYTOKINE_CYTOKINE_RECEPTOR_INTERACTION	267	7	0.0262	1.92E-8	8.94E-7
FOCAL_ADHESION	201	6	0.0299	9.93E-8	3.69E-6
BLADDER_CANCER	42	4	0.0952	1.46E-7	4.52E-6
NOD_LIKE_RECEPTOR_SIGNALING_PATHWAY	62	4	0.0645	7.15E-7	1.9E-5
PANCREATIC_CANCER	70	4	0.0571	1.17E-6	2.41E-5
RENAL_CELL_CARCINOMA	70	4	0.0571	1.17E-6	2.41E-5
LEISHMANIA_INFECTION	72	4	0.0556	1.31E-6	2.43E-5
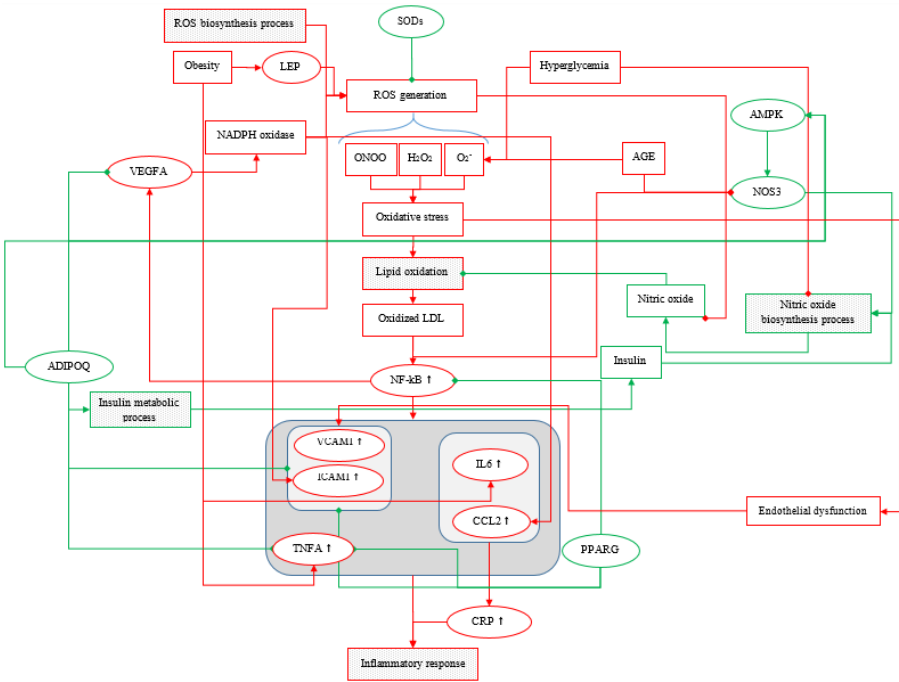
Fig 3: Mechanism of oxidative stress in stroke. This mechanism filtered from the constructed model.

DISCUSSION

In this study, employing integrative disease modeling methods, two common important risk factors for stroke (dyslipidemia and diabetes) were mechanistically modeled. GSEA showed that the biological process and pathway of the inflammatory response, response to lipid, regulation of body fluid levels, regulation of response to stress, complement, and coagulation cascades, and PPAR signaling pathway are involved in the stroke mechanism. In each of these biological pathways, there are several factors involved in which inhibition or stimulation can affect these pathways. Modeling has shown that various factors such as Insulin, ADIPOQ, PPARG, NOS3, and HDL-C have an effect on these pathways and in many threatening processes such as lipid oxidation and ROS biosynthesis process.

Inflammation is common in cardiovascular diseases (CVD) and increases the risk of CVD and diabetes [12]. The presence of inflammation, which somehow indicates the stimulation of the immune system, is followed by damage to the nerve cells and the impact of various risk factors, such as diabetes and obesity [13, 14]. Also, some markers, such as CRP, TNFA, and IL6, are present in the mechanism of inflammation that increases in an ischemic obstruction [15].

Dyslipidemia may also be involved in stroke. For example, reducing High-Density Lipoprotein (HDL) cholesterol leads to an increase in inflammation, which increases the risk of CVD and ischemic stroke [16], especially in diabetic patients [17]. As our model shows, Reactive Oxygen Species (ROS) play an important role in the production of CRP through NfKB activation [18], nitric oxide reduction [19], and increased oxidative stress [20], and TNF [21]. The model also showed that some cytokines, such as ICAM1 and VCAM1 [22], also contribute to inflammatory responses. According to model findings, factors such as PPARG and ADIPOQ inhibit the underlying factors of inflammation. For example, PPARG reduces CRP production and thereby inflammation by restricting cytokines or inhibiting TNF [22].

The analysis of cause and effect mechanisms involved in increasing the risk of stroke showed that oxidative stress plays an important role in various mechanisms of stroke, including endothelial dysfunction and lipid oxidation. As the model showed, among free radical producing factors, VEGFA increases the production of ROS by activating NADPH oxidase and obesity due to the high amount of Leptin (LEP) in obese people. On the other hand, hyperglycemia also increases the amount of superoxide (O_2^-) and AGE production. AGEs are products that inhibit NOS3 enzymes. According to our findings, NOS3 inhibition inhibits the production of nitric oxide, resulting in increased oxidative stress. Oxidative stress causes LDL oxidation, which itself activates inflammatory cells and increases factors such as NfKB [18]. Following this, the increase in some inflammatory markers, such as IL6, CRP, and TNFA, and some adhesion molecules [23], provides the basis for inflammatory response. The role of some factors in increasing the biosynthesis of nitric oxide in the model is well demonstrated. ADIPOQ plays an important role in this process in two ways. First, stimulating the metabolic process of insulin, increases the biosynthesis of nitric oxide. Also, through the second way, increasing the expression of AMPK, and subsequently increasing the NOS3, facilitates the production of nitric oxide.

CONCLUSION

Integrative Modeling showed that using a variety of molecular and clinical data which are stored in different databases, can increase our understanding of disease mechanisms. Forasmuch as this method is based on evidence, so using this method provides a better understanding of the disease mechanism. Since the proper understanding of the role of the various factors in the mechanism of diseases is contributing to the development and proposal of new therapies, we must look for solutions to interfere with these biological pathways.

In this study, only the risk factors of dyslipidemia and diabetes were examined mechanistically, while other factors such as hypertension were not studied. Therefore, the proposed model was only for two mentioned risk factors without considering other factors.

AUTHOR'S CONTRIBUTION

The authors agree on this final form of the manuscript, and attested that all authors contributed in the final draft of the manuscript.

CONFLICTS OF INTEREST

The authors declare no conflicts of interest regarding the publication of this study.

FINANCIAL DISCLOSURE

No financial interests related to the material of this manuscript have been declared.

REFERENCES

1. Naghavi M, Alemu Abajobir A, Abbafati C, Abbas KM, Abd-Allah F, Ferede Abera S, et al. Global, regional, and national age-sex specific mortality for 264 causes of death, 1980-2016: A systematic analysis for the Global Burden of Disease Study 2016. Lancet. 2017/09/19. 2017; 390(10100):1151–210. PMID: 28919116 DOI: 10.1016/S0140-6736(17)32152-9 [PubMed]

2. Amarenco P, Bogousslavsky J, Caplan LR, Donnan GA, Hennerici MG. Classification of stroke subtypes. Cerebrovasc Dis. 2009; 27(5): 493-501. PMID: 19342825 DOI: 10.1159/000210432 [PubMed]

3. Buczek J, Czlonkowska A. Stroke and genetics. Periodicum Biologorum. 2012; 114(3): 259–66.

4. Feuerstein GZ, Chavez J. Translational medicine for stroke drug discovery the pharmaceutical industry perspective. Stroke. 2009; 40(3 suppl 1): S121–5. PMID: 19064772 DOI: 10.1161/STROKEAHA.108.535104 [PubMed]

5. Caffes N, Kurland DB, Gerzanich V, Simard JM. Glibenclamide for the treatment of ischemic and hemorrhagic stroke. Int J Mol Sci. 2015; 16(3): 4973–84. PMID: 25749474 DOI: 10.3390/ijms16034973 [PubMed]

6. Audebert HJ, Sobesky J. Stroke: 'Time is brain' after stroke, regardless of age and severity. Nat Rev Neurol. 2014; 10(12): 675–6. PMID: 25330727 DOI: 10.1038/nrneurol.2014.194 [PubMed]

7. Woolf SH. The meaning of translational research and why it matters. JAMA. 2008; 299(2): 211-3. PMID: 18182604 DOI: 10.1001/jama.2007.26 [PubMed]

8. Habibi-Koolaee M, Shahmoradi L, Niakan Kalhori SR,

Ghannadan H, Younesi E. Prevalence of stroke risk factors and their distribution based on stroke subtypes in Gorgan: A retrospective hospital-based study - 2015-2016. Neurol Res Int. 2018; 2018: 2709654. PMID: 30147952 DOI: 10.1155/2018/2709654 [PubMed]

9. Habibi-Koolaee M, Shahmoradi L, Niakan Kalhori SR, Ghannadan H, Hosseini A, Younesi E. Lipid profile and the risk of stroke: A study from North of Iran. Journal of Research in Medical and Dental Science. 2018; 6(1): 343–9.

10. Berthold MR, Cebron N, Dill F, Gabriel TR, Kötter T, Meinl T, et al. KNIME: The Konstanz information miner. In: Preisach C, Burkhardt H, Schmidt-Thieme L, Decker R (eds). Data analysis, machine learning and applications. Studies in Classification, Data Analysis, and Knowledge Organization. Springer; 2007.

11. Smoot ME, Ono K, Ruscheinski J, Wang P-L, Ideker T. Cytoscape 2.8: New features for data integration and network visualization. Bioinformatics. 2011; 27(3): 431–2. PMID: 21149340 DOI: 10.1093/bioinformatics/btq675 [PubMed]

12. Micha R, Mozaffarian D. Saturated fat and cardiometabolic risk factors, coronary heart disease, stroke, and diabetes: A fresh look at the evidence. Lipids. 2010; 45(10): 893-905. PMID: 20354806 DOI: 10.1007/s11745-010-3393-4 [PubMed]

13. DeFronzo RA, Abdul-Ghani M. Assessment and treatment of cardiovascular risk in prediabetes: impaired glucose tolerance and impaired fasting glucose. Am J Cardiol. 2011; 108(3 Suppl): 3B-24B. PMID: 21802577 DOI: 10.1016/j.amjcard.2011.03.013 [PubMed]

14. Sowers JR. Diabetes in the elderly and in women: cardiovascular risks. Cardiol Clin. 2004; 22(4): 541-51. PMID: 15501622 DOI: 10.1016/j.ccl.2004.06.008 [PubMed]

15. Wiseman S, Marlborough F, Doubal F, Webb DJ, Wardlaw J. Blood markers of coagulation, fibrinolysis, endothelial dysfunction and inflammation in lacunar stroke versus non-lacunar stroke and non-stroke: Systematic review and meta-analysis. Cerebrovasc Dis. 2014; 37(1): 64-75. PMID: 24401164 DOI: 10.1159/000356789 [PubMed]

16. Demarin V, Lisak M, Morovic S, Cengic T. Low high-density lipoprotein cholesterol as the possible risk factor for stroke. Acta Clin Croat. 2010; 49(4): 429–39. PMID: 21830454 [PubMed]

17. Tkac T. Pharmacological treatment of diabetic patients with respect to prevention of macrovascular disease. Acta Diabetol. 2003; 40 (Suppl 2): S338-42. PMID: 14704865 DOI: 10.1007/s00592-003-0116-9 [PubMed]

18. Yamaoka-Tojo M, Tojo T, Takahira N, Masuda T, Izumi T. Ezetimibe and reactive oxygen species. Curr Vasc Pharmacol. 2011; 9(1): 109–20. PMID: 21044015 DOI: 10.2174/157016111793744652 [PubMed]

19. Idris I, Thomson GA, Sharma JC. Diabetes mellitus and stroke. International Journal of Clinical Practice. 2006; 60(1): 48–56.

20. Orr JD. Statins in the spectrum of neurologic disease. Curr Atheroscler Rep. 2008; 10(1): 11–8. PMID: 18366980 DOI: 10.1007/s11883-008-0003-5 [PubMed]

21. Mehta SL, Li PA. Neuroprotective role of mitochondrial uncoupling protein 2 in cerebral stroke. J Cereb Blood Flow Metab. 2009; 29(6): 1069–78. PMID: 19240738 DOI: 10.1038/jcbfm.2009.4 [PubMed]

22. Culman J, Zhao Y, Gohlke P, Herdegen T. PPAR-gamma: Therapeutic target for ischemic stroke. Trends Pharmacol Sci. 2007; 28(5): 244–9. PMID: 17416424 DOI: 10.1016/j.tips.2007.03.004 [PubMed]

23. Walcher D, Marx N. Advanced glycation end products and C-peptide-modulators in diabetic vasculopathy and atherogenesis. Semin Immunopathol. 2009; 31(1): 103–11. PMID: 19347338 DOI: 10.1007/s00281-009-0144-9 [PubMed]