ACTA MEDICA LITUANICA. 2007. VOLUME 14. No. 1. P. 24–29

Atherosclerosis: alterations in cell communication

Albertas Timinskas[1],

Zita Aušrelė Kučinskienė[2],

Vaidutis Kučinskas[2]

[1] *Institute of Biotechnology,*
V. A. Graičiūno 8,
Vilnius LT-02241, Lithuania

[2] *Faculty of Medicine,*
Vilnius University,
M. K. Čiurlionio 21,
LT-03101 Vilnius, Lithuania

Objective. We investigated the distribution of atherosclerosis-linked genes within known metabolic / signaling pathways in order to determine most vulnerable pathways during atherosclerosis pathogenesis as well as to compile a full list of genes predominant for the disease.

Materials and Methods. A search for genes associated with atherosclerosis within PubMed abstracts was carried out using modern data mining techniques available at SNPs3D database. Presence of data on metabolic pathways for each gene of interest was checked using KEGG and g-language tools. Association of atherosclerosis-linked genes within pathways was visualized and estimated using the g-language tool.

Results. Analysis of to-date PubMed abstracts resulted in 748 human genes associated with atherosclerosis. Further analysis has revealed that 134 of over 300 known to date metabolic / signaling pathways include at least one gene associated with atherosclerosis of the total of 436 genes. Reducing accident ascription pathways populated by at least 10 atherosclerosis-linked genes were only taken into account making final conclusions.

Conclusions. Taken together, our results demonstrate that the essence of atherosclerosis is the dysfunction of cell signaling and intercell communication cascades. We maintain an exclusively high role of cytokines and their receptors in atherosclerosis pathogenesis. A sufficiently high intersection of atherosclerosis and diabetes or several cancer species proposed an increased risk for further complications in these patients.

Keywords: atherosclerosis, linked genes, data mining, signaling pathways, cell communication, genomics

INTRODUCTION

Atherosclerosis, an inflammatory disease stemming from genetic and environmental factors, is the primary disease of coronary arteries (1–4). It is the number one killer in the United States (5) and every year claims more lives than the next five leading causes of death combined (5). The role of genetics in atherosclerosis pathophysiology has been recognized for some time: inheritance of risk factors was first shown in classical twin (6–9) and family history studies (10, 11). Diabetes (12, 13), hypercholesterolemia (14), hypertension (13), obesity (15), smoking (13), and physical inactivity (16) are also known risk factors for the disease. Although interventional cardiology procedures such as balloon angioplasty, stenting (17), and atherectomy (18) have been successful in combating local coronary artery disease (CAD), this has not been met by equivalent success in interrupting the underlying disease at the molecular level (pleiotrophic effects of statins and aspirin notwithstanding). There is, for example, no treatment currently on the market designed to target the molecular interactions of the disease process itself.

Traditional approaches to the molecular study of atherosclerosis target one molecule or gene thought to have an important role in the development of the disease and then manipulate it through knockout, knockdown, or transgenic technology (19, 20). Although much has been learned about atherosclerosis through these studies, there remains a lack of understanding of the disease as an integrated whole. With this in mind, we undertook a comprehensive gene level assessment of atherosclerosis taking a network, or pathway-based, approach to analysis. Statistical tools mined the resulting data for genes linked to disease, and the further study was conducted with currated pathways. Our findings could provide a good basis for the further modeling of atherosclerosis molecular pathway.

* Correspondence to: Dr. A. Timinskas, Institute of Biotechnology, V. A. Graičiūno 8, LT-02241 Vilnius, Lithuania. E-mail: timis@ibt.lt

MATERIALS AND METHODS

Genes linked to atherosclerosis were searched against all the Medline abstracts using the data mining techniques recently proposed by Yue and coworkers (21).

Literature dataset

The abstracts of all the Medline entries associated with each gene in the NCBI Gene database (22) are the source of words and terms. In the current version, there are 80,249 Medline references linked to 19,228 human genes. Word types are identified using SVMtagger (23). Keyterms are constructed from single nouns and adjectives, adjective / noun pairs, and continuous strings of words classified as adjectives or nouns. For example, the phrase 'intima media' occurring in an abstract would result in three keyterms: 'intima', 'media', and 'intima media'. Terms occurring only once are removed. There are currently a total of 266,337 keyterms.

The number of occurrences of each keyterm 'KW' in all the abstracts ('Total_count(KW)' is retained, as well as the number of occurrences of each keyterm in the abstracts associated with each gene 'G', 'Count(G, KW)', and the fraction of all occurrences of each keyterm that are associated with each gene is calculated as

$$F1(G, KW) = \frac{Count(G, KW)}{Total\_Count(KW)}. \quad (1)$$

Compilation of a candidate gene list for atherosclerosis disease

Given a disease name, a list of candidate genes is generated as follows:

A: The subset of abstracts relevant to the disease is identified.

B: A keyterm profile is generated for the disease, using the selected abstracts. All keyterms are ranked by the fraction of disease abstracts that contain them:

$$Rank(KW) = \frac{Count\_abstracts(D, KW)}{[Total\_abstracts(KW) + 50]} \quad (2)$$

where 'Count_abstracts(D, KW)' is the number of abstracts for disease 'D' containing the keyterm 'KW', and 'Total_abstracts(KW)' is the total number of abstracts containing the keyterm. A pseudo count of 50 is added to reduce noise. The top ranking 40 keyterms are selected, provided Rank(KW) is at least 0.1.

C: The overlap of the disease keyterms with those of each gene is calculated:

1. The number of times each selected keyterm 'KW' occurs in the abstracts associated with the disease 'D', 'Count(D, KW)', is determined, and the relative frequency is calculated as

$$F2(D, KW) = \frac{Count(D, KW)}{Total\_Count(KW)} \quad (3)$$

2. The strength of association of the disease 'D' with a gene 'G' is calculated as the dot product of the relative frequencies of the disease keyterms with the relative frequencies of those same keyterms in that gene:

$$SD(D, G) = \Sigma_{KW} F1(G, KW) \cdot F2(D, KW), \quad (4)$$

where the sum is only over the up to 40 keywords selected as the keyterm set of disease 'D'. The association strength is deliberately biased towards the keyterms most strongly associated with the disease, as opposed to those associated with particular genes.

D: Finally, all genes with a non-zero score are returned as candidates.

Distribution of atherosclerosis candidate genes within metabolic pathways

Presence of atherosclerosis candidate genes within KEGG pathways was checked by searching the KEGG pathway database using the *bget* mode (25) for the gene under analysis. The method is available on KEGG web page. Genes involved in the KEGG pathway were taken for further analysis. Atherosclerosis-linked genes within each pathway were marked using the *g-language* tool (26, 27). The density of genes associated with disease within each pathway scheme was calculated by examining schemes provided by the tool.

RESULTS

Search of PubMed abstracts for '*atherosclerosis*' by means of SNPs3D data mining techniques revealed 40 more or less associated keyterms (see Table 1). These keyterms are most specific for abstracts containing 'atherosclerosis'. One could see the main features of atherosclerosis represented here. All keywords could be grouped into five groups: 'athero*', 'tunica intima-media thickness (imt)', 'heart carotid coronary artery disease (cad)', 'foam cell', 'oxidised low density lipoprotein (ldl)'.

The compiled list of keyterms was used for the further mining of atherosclerosis-linked genes. The resulting list of 748 candidate genes was obtained. Using the KEGG *bget* tool it has been found that 436 genes of the list are involved in 134 various metabolic / signaling pathways. No data on pathways for 312 genes are yet available in KEGG database. The population of atherosclerosis-associated genes within the 134 pathways was investigated using the *g-language* tool. Most populated pathways are listed in Table 2 (more than 10 genes associated with atherosclerosis).

Analysis of pathway schemes has been carried out investigating the density of candidate genes in each scheme to decide whether the level or density of disease-linked genes is essential. Results on density are given in Table 3. One can see that the same pathways are present on both lists showing no marginal difference between the two criteria.

Table 1. **Top40 PubMed keyterms associated with 'atherosclerosis'**

No.	Keyterm	No.	Keyterm
1.	Atherosclerosis	21.	Thickness imt
2.	Imt	22.	Coronary
3.	Carotid	23.	Coronary heart
4.	Coronary atherosclerosis	24.	Coronary artery
5.	Atherosclerotic	25.	Human atherosclerotic
6.	Carotid atherosclerosis	26.	Artery disease
7.	Arteriosclerosis	27.	Tunica intima
8.	Atherosclerotic lesions	28.	Coronary arteriosclerosis
9.	Foam	29.	Foam cell formation
10.	Intima-media	30.	Intimal
11.	Intima-media thickness	31.	Lipoproteins, ldl
12.	Foam cells	32.	Coronary disease
13.	Carotid artery diseases	33.	Artery
14.	Carotid arteries	34.	Oxidized low density lipoprotein
15.	Foam cell	35.	Thickness
16.	Atherogenesis	36.	Arteries
17.	Atherosclerotic plaques	37.	Coronary artery disease
18.	Atherogenic	38.	Atherosclerosis risk
19.	Premature atherosclerosis	39.	Tunica media
20.	Carotid artery	40.	Cad

Table 2. **KEGG pathways with highest count (N) of atherosclerosis candidate genes**

No.	KEGG ID	Pathway description	N	Pathway group
1	hsa04060	Cytokine–cytokine receptor interaction	75	3.3 Signaling Molecules and Interaction
2	hsa04080	Neuroactive ligand–receptor interaction	53	3.3 Signaling Molecules and Interaction
3	hsa04010	MAPK signaling pathway	52	3.2 Signal Transduction
4	hsa04510	Focal adhesion	51	4.3 Cell Communication
5	hsa04810	Regulation of actin cytoskeleton	41	4.1 Cell Motility
6	hsa04670	Leukocyte transendothelial migration	30	4.5 Immune System
7	hsa04610	Complement and coagulation cascades	29	4.5 Immune System
8	hsa04020	Calcium signaling pathway	29	3.2 Signal Transduction
9	hsa03320	PPAR signaling pathway	27	4.4 Endocrine System
10	hsa04350	TGF-beta signaling pathway	26	3.2 Signal Transduction
11	hsa04514	Cell adhesion molecules (CAMs)	24	3.3 Signaling Molecules and Interaction
12	hsa04640	Hematopoietic cell lineage	24	4.5 Immune System
13	hsa05212	Pancreatic cancer	24	5.4 Cancers
14	hsa04620	Toll-like receptor signaling pathway	23	4.5 Immune System
15	hsa04512	ECM-receptor interaction	23	3.3 Signaling Molecules and Interaction
16	hsa04210	Apoptosis	22	4.2 Cell Growth and Death
17	hsa04650	Natural killer cell mediated cytotoxicity	22	4.5 Immune System
18	hsa04630	Jak-STAT signaling pathway	22	3.2 Signal Transduction
19	hsa04920	Adipocytokine signaling pathway	20	4.4 Endocrine System
20	hsa05214	Glioma	19	5.4 Cancers
21	hsa04910	Insulin signaling pathway	19	4.4 Endocrine System
22	hsa04540	Gap junction	19	4.3 Cell Communication
23	hsa04730	Long-term depression	18	4.6 Nervous System
24	hsa04370	VEGF signaling pathway	18	3.2 Signal Transduction
25	hsa05210	Colorectal cancer	18	5.4 Cancers
26	hsa04660	T cell receptor signaling pathway	17	4.5 Immune System
27	hsa04360	Axon guidance	16	4.8 Development
28	hsa05120	Epithelial cell signaling in *Helicobacter pylori* infection	15	5.2 Infectious Diseases
29	hsa04530	Tight junction	15	4.3 Cell Communication
30	hsa04912	GnRH signaling pathway	15	4.4 Endocrine System

Table 2, continuation

No.	KEGG ID	Pathway description	N	Pathway group
31	hsa04310	Wnt signaling pathway	14	3.2 Signal Transduction
32	hsa01430	Cell communication	14	4.3 Cell Communication
33	hsa04520	Adherens junction	14	4.3 Cell Communication
34	hsa04930	Type II diabetes mellitus	13	5.3 Metabolic Disorders
35	hsa04664	Fc epsilon RI signaling pathway	13	4.5 Immune System
36	hsa00590	Arachidonic acid metabolism	12	1.3 Lipid Metabolism
37	hsa04150	mTOR signaling pathway	11	3.2 Signal Transduction
38	hsa04662	B cell receptor signaling pathway	10	4.5 Immune System
39	hsa04940	Type I *diabetes mellitus*	10	5.3 Metabolic Disorders
40	hsa05130	Pathogenic *Escherichia coli* infection – EHEC	10	5.2 Infectious Diseases
41	hsa05131	Pathogenic *Escherichia coli* infection – EPEC	10	5.2 Infectious Diseases

Table 3. **Density of atherosclerosis candidate genes within KEGG pathway schemes**

No.	Pathway	Density, %	Number of genes in scheme
1	ECM-receptor interaction	75.9	116
2	PPAR signaling pathway	51.9	52
3	Type II diabetes mellitus	48.0	25
4	Long-term depression	47.1	34
5	Gap junction	46.5	43
6	Adipocytokine signaling pathway	43.2	37
7	Leukocyte transendothelial migration	41.8	79
8	Toll-like receptor signaling pathway	39.4	66
9	TGF-beta signaling pathway	38.2	55
10	Cell adhesion molecules	35.2	233
11	Focal adhesion	33.3	60
12	JAK-STAT signaling pathway	33.3	27
13	Axon guidance	32.9	79
14	Apoptosis	31.7	60
15	Complement and coaguliation cascades	30.5	59
16	MAPK signaling pathway	30.2	126
17	mTOR signaling pathway	30.0	30
18	Epithelial cell signaling in *Helicobacter pylori* infection	28.0	50
19	Cytokine–cytokine receptor interaction	27.9	301
20	Adherens junction	27.4	73
21	Regulation of actin cytoskeleton	25.0	72
22	Neuroactive ligand-receptor interaction	24.3	136
23	Insulin signaling pathway	24.2	62
24	Cell communication	23.5	17
25	Natural killer cell mediated cytotoxicity	23.2	82
26	Calcium signaling pathway	22.9	48
27	Fc epsilon RI signaling pathway	22.5	40
28	Arachidonic acid metabolism	22.2	54
29	Hematopoietic cell lineage	20.0	140
30	Tight junction	19.6	56
31	T cell receptor signalig pathway	18.6	59
32	Wnt signaling pathway	12.7	71

DISCUSSION

A review of 41 pathways (see Tables 2 and 4) highly populated by atherosclerosis-linked genes has revealed that most affected are pathways ascribed to sections 3 and 4 (KEGG classification scheme) and especially 3.2, 3.3, 4.3, 4.4 and 4.5. These five subsections comprise 2/3 of 41 best representatives (Table 2). The only pathway, namely arachidonic acid metabolism, represents the largest section of pathways – metabolism. No representatives are in the 2nd section, what could mean that genetic information processing isn't affected in the case of atherosclerosis.

Table 4. **Distribution of atherosclerosis candidate genes within KEGG pathway sections**

Pathway section	Number of pathways	Number of atherosclerosis--linked representatives
1. Metabolism	**141**	**1**
2. Genetic information processing	**15**	**0**
3. Environmental information processing	**17**	**11**
3.1 Membrane Transport	2	0
3.2 Signal Transduction	11	7
3.3 Signaling Molecules and Interaction	4	4
4. Cellular Processes	**31**	**21**
4.1 Cell Motility	3	1
4.2 Cell Growth and Death	2	1
4.3 Cell Communication	5	5
4.4 Endocrine System	5	4
4.5 Immune System	9	8
4.6 Nervous System	2	1
4.7 Sensory System	2	0
4.8 Development	2	1
4.9 Behavior	1	0
5. Human Diseases	**16**	**8**
5.1 Neurodegenerative Disorders	6	0
5.2 Infectious Diseases	4	3
5.3 Metabolic Disorders	3	2
5.4 Cancers	3	3

CONCLUSIONS

On the basis of the obtained results we conclude that the essence of atherosclerosis is dysfunction of signal transfer / cell communication. Due to a wide variety of communication forms affected, a variety of atherosclerosis phenotypes could be predicted.

ACKNOWLEDGEMENT

This work was supported by the Lithuanian National Science and Studies Foundation grant U-04002.

Received 31 January 2007
Accepted 21 February 2007

References

1. Libby P, Ridker PM, Maseri A. Inflammation and atherosclerosis. Circulation 2002; 105: 1135–43.
2. Poulter N. Coronary heart disease is a multifactorial disease. Am J Hypertens 1999; 12: 92S–95S.
3. Rosal MC, Ockene JK, Luckmann R, Zapka J, Goins KV, Saperia G, Mason T, Donnelly G. Coronary heart disease multiple risk factor reduction; providers' perspectives. Am J Prev Med 2004; 27: 54–60.
4. Ross R. Atherosclerosis – an inflammatory disease. N Engl J Med 1999; 340: 115–26.
5. American Heart Association. Heart Disease and Stroke Statistics–2004 Update. Dallas, TX: American Heart Association; 2003.
6. Evans A, Van Baal GC, McCarron P, DeLange M, Soerensen TI, De Geus EJ, Kyvik K, Pedersen NL, Spector TD, Andrew T, Patterson C, Whitfield JB, Zhu G, Martin NG, Kaprio J, Boomsma DI. The genetics of coronary heart disease: the contribution of twin studies. Twin Res 2003; 6: 432–41.
7. Hong Y, de Faire U, Heller DA, McClearn GE, Pedersen N. Genetic and environmental influences on blood pressure in elderly twins. Hypertension 1994; 24: 663–70.
8. Iliadou A, Lichtenstein P, Morgenstern R, Forsberg L, Svensson R, de Faire U, Martin NG, Pedersen NL. Repeated blood pressure measurements in a sample of Swedish twins: heritabilities and associations with polymorphisms in the rennin–angiotensin–aldosterone system. J Hypertens 2002; 20: 1543–50.
9. Larkin JE, Frank BC, Gaspard RM, Duka I, Gavras H, Quackenbush J. Cardiac transcriptional response to acute and chronic angiotensin II treatments. Physiol Genomics 2004; 18: 152–66.
10. Romaldini CC, Issler H, Cardoso AL, Diament J, Forti N. Risk factors for atherosclerosis in children and adolescents with family history of premature coronary artery disease. J Pediatr 2004; 80: 135–40.
11. Valentine RJ, Guerra R, Stephan P, Scoggins E, Clagett GP, Cohen J. Family history is a major determinant of subclinical peripheral arterial disease in young adults. J Vasc Surg 2004; 39: 351–56.
12. Beckman JA, Creager MA, Libby P. Diabetes and atherosclerosis: epidemiology, pathophysiology, and management. JAMA 2002; 287: 2570–81.
13. Willerson JT and Ridker PM. Inflammation as a cardiovascular risk factor. Circulation 2004; 109: II2–II10.
14. Libby P. Inflammation in atherosclerosis. Nature 2002; 420: 868–74.
15. Correia ML and Haynes WG. Leptin, obesity and cardiovascular disease. Curr Opin Nephrol Hypertens 2004; 13: 215–23.
16. Armen J, Smith BW. Exercise considerations in coronary artery disease, peripheral vascular disease, and diabetes mellitus. Clin Sports Med 2003; 22: 123–33, VIII.
17. Presbitero P, Zavalloni D, Scatturin M, Marisco F, Pagnotta P, Boccuzzi G. Procedural and long-term results of sirolimus-eluting stent in patients at high

risk for restenosis. Minerva Cardioangiol 2004; 52: 189–94.

18. Rubartelli P, Niccoli L, Alberti A, Giachero C, Ettori F, Missiroli B, Bernardi G, Maiello L, Reimers B, Cernigliaro C, Sardella G, Bramucci E. Coronary rotational atherectomy in current practice: acute and mid-term results in high- and low-volume centers. Catheter Cardiovasc Interv 2004; 61: 463–71.
19. Franz WM, Frey N, Muller O, Kubler W, Katus HA. A transgenic animal model: new possibilities for cardiovascular research. Z Kardiol 84, Suppl 1995; 4: 17–32.
20. Ohashi R, Mu H, Yao Q, Chen C. Cellular and molecular mechanisms of atherosclerosis with mouse models. Trends Cardiovasc Med 2004; 14: 187–90.
21. Yue P, Melamud E, Moult J. SNPs3D: Candidate gene and SNP selection for association studies. BMC Bioinformatics 2006; 7: 166.
22. Pruitt KD, Katz KS, Sicotte H, Maglott DR. Introducing RefSeq and LocusLink: curated human genome resources at the NCBI. Trends Genet 2000; 16: 44–7.
23. Rauchova H, Vokurkova M, Koudelova J. Developmental changes of erythrocyte catalase activity in rats exposed to acute hypoxia. Physiol Res 2005; 54(5): 527–32.
24. Minoru Kanehisa, Susumu Goto, Shuichi Kawashima, Yasushi Okuno, Masahiro Hattori. The KEGG resource for deciphering the genome. Nucl Acids Res 2004; 32: D277–80.
25. http://www.genome.jp/kegg/genes.html
26. http://www.g-language.org/data/marray/
27. Arakawa K, Kono N, Yamada Y, Mori H, Tomita M. KEGG-based pathway visualization tool for complex omics data. In Silico Biol 2005; 5(4): 419–23.

Albertas Timinskas, Zita Aušrelė Kučinskienė, Vaidutis Kučinskas

ATEROSKLEROZĖ: LĄSTELIŲ KOMUNIKACIJŲ POKYČIAI

Santrauka

Tikslas. Mes tyrėme su ateroskleroze siejamų genų pasiskirstymą žinomuose metaboliniuose / signaliniuose keliuose siekdami nustatyti labiausiai pažeidžiamus metabolinius kelius aterosklerozės patogenezėje. Kitas šio darbo tikslas buvo sudaryti pilną su šia liga susijusių genų sąrašą.

Medžiagos ir metodai. Su ateroskleroze siejamų genų paieška PubMed santraukose buvo vykdyta pasitelkus šiuolaikinius SNPs3D duomenų bazės duomenų gavybos įrankius. Suformuoto genų sąrašo kiekvieno atstovo dalyvavimas žinomuose metaboliniuose keliuose buvo patikrintas KEGG ir ‘g-language’ įrankiais. ‘G-language’ įrankis buvo pasitelktas vizualizuojant bei vertinant aterosklerozės kandidatinių genų tarpusavio sąsajas metaboliniuose keliuose.

Rezultatai. Išanalizavus PubMed santraukas, identifikuoti 748 su ateroskleroze susiję žmogaus genai. Tolesnė analizė rodo, kad 436 šio sąrašo atstovai dalyvauja 134-iuose iš daugiau nei 300 žinomų metabolinių / signalinių kelių. Siekiant didesnio rezultatų patikimumo galutinės išvados buvo formuojamos apibendrinant tik metabolinių kelių, turinčių ne mažiau kaip 10 su ateroskleroze siejamų genų, analizę.

Išvados. Išanalizavus genų kandidatų paplitimą atskiruose metaboliniuose / signaliniuose keliuose, galima teigti, kad aterosklerozės esmė glūdi ląstelių signalinių sistemų ir tarpląstelinės komunikacijos sutrikimuose. Ypač išryškėjo citokinų ir jų receptorių genų reikšmė signaliniuose keliuose. Kai kurie genai kandidatai yra bendri kancerogenezėje ir aterogenezėje.

Raktažodžiai: aterosklerozė, susiję genai, duomenų gavyba, signaliniai keliai, ląstelių komunikacija, genomika