대한예방한의학회지 제27권 제2호(2023년 8월) doi : 10.25153/spkom.2023.27.2.003
Journal of Society of Preventive Korean Medicine 2023 ; 27(2) : 35-48 pISSN : 1226-7066 eISSN : 2713-9360

Identification of Potential Prognostic Biomarkers in lung cancer patients based on Pattern Identification of Traditional Korean Medicine
Running title: A biomarker based on the Korean pattern identification for lung cancer

Ji Hye Kim[1] · Hyun Sub Cheong[2] · Chunhoo Cheon[3] · Sooyeon Kang[1] · Hyun Koo Kim[4] · Hyoung Doo Shin[2] · Seong-Gyu Ko[1,3]*

[1] Laboratory of Clinical Biology and Pharmacogenomics, Department of Preventive Medicine, College of Korean Medicine, Kyung Hee University, 26 Kyungheedae-ro, Dongdaemun-gu, Seoul, 02447, Republic of Korea
[2] Department of Genetic Epidemiology, SNP Genetics, Inc., Seoul 04107, Republic of Korea
[3] Department of Preventive Medicine, College of Korean Medicine, Kyung Hee University, 26 Kyungheedae-ro, Dongdaemun-gu, Seoul 02447, Republic of Korea
[4] Department of Thoracic and Cardiovascular Surgery, Korea University Guro Hospital, Korea University College of Medicine, Seoul 08308, Republic of Korea

Abstract

Objective : We studied prognostic biomarkers discovery for lung cancer based on the pattern identification for the personalized Korean medicine.

Methods : Using 30 tissue samples, we performed a whole exome sequencing to examine the genetic differences among three groups.

Results : The exome sequencing identified among 23,490 SNPs germline variants, 12 variants showed significant frequency differences between Xu and Stasis groups ($P<0.0005$). As similar, 18 and 10 variants were identified in analysis for Xu vs. Gentleness group and Stasis vs. Gentleness group, respectively ($P<0.001$). Our exome sequencing also found 8,792 lung cancer specific variants and among the groups identified 6, 34, and 12 variants which showed significant allele frequency differences in the comparison groups; Xu vs. Stasis, Xu vs. Gentleness group, and Stasis vs. Gentleness group. As a result of PCA analysis, in germline data set, Xu group was divided from other groups. Analysis using somatic variants also showed similar result. And in gene ontology analysis using pattern identification variants, we found genes like as FUT3, MYCBPAP, and ST5 were related to tumorigenicity, and tumor metastasis in comparison between Xu and Stasis. Other significant SNPs for two were responsible for eye morphogenesis and olfactory receptor activity. Classification of somatic pattern identification variants showed close relationship in multicellular organism reproduction, anion-anion antiporter activity, and GTPase regulator activity.

Conclusions : Taken together, our study identified 40 variants in 29 genes in association with germline difference of pattern identification groups and 52 variants in 47 genes in somatic cancer tissues.

• 접수 : 2023년 7월 21일 • 수정접수 : 2023년 8월 5일 • 채택 : 2023년 8월 12일

*Funding : This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (No. 2020R1A5A2019413) and the Korean Medicine R&D Project of the Ministry of Health and Welfare (Grant number: HI18C2382).

*Corresponding author : Seong-Gyu Ko, M.D., Ph.D., M.P.H. Department of Preventive Medicine, College of Korean Medicine, Kyung Hee University, 26 Kyungheedae-ro, Dongdaemun-gu, Seoul, 02447, Republic of Korea
전화 : 82-2-961-0329, 팩스 : 82-2-966-1165, 전자우편 : epiko@khu.ac.kr

Key words : Pattern identification; Whole exome sequencing (WES); Lung cancer; Body constitution questionnaire (BCQ); Traditional Korean medicine (TKM).

Ⅰ. Introduction

Lung cancer is one of the refractory cancers. Since there is no effective prognostic method in early stage for lung cancer, symptomatic lung cancer is often in the late stage when diagnosed. Even progressed lung cancers are often asymptomatic. About 5–15% of lung cancer patients are diagnosed while asymptomatic, but the rest are diagnosed only after symptoms have manifested, which is already late for surgery and results in high mortality.

Whole-exome sequencing (WES) is a genome-wide testing approach that allows selective sequencing of the protein-coding regions of the genome, which are significantly enriched for disease-associated variants[1]. The identification of the variant of a rare form of inflammatory bowel disease in an infant was the first success to diagnose and inform a treatment for a human patient[2]. Exome sequencing in human medicine benefits from the availability of large databases of known single-nucleotide polymorphisms (SNPs), and known control genomes; Examples of diseases for which exome sequencing has been used to detect a causative variant include Leber congenital amaurosis[3], Alzheimer disease[4], immunodeficiency leading to infection with human herpes virus 8 causing Kaposi Sarcoma[5], and a lot of cancer predisposition mutations[6–9].

Application of traditional Korean medicine (TKM) as an adjuvant cancer therapy has been reported to help reduce adverse effects of medicines such as chemo- or/and radiotherapy and to enhance the efficacy of each[10]. Furthermore, the traditional herbal remedies use has recently been recognized as an important source for novel drug development, including anti-cancer drugs[11–12]. Therefore, modern western medicine practitioners and researchers are gradually open to exploring the potential of TKM to enhance conventional treatment of cancer patients[13–16]. The process of common treatment for lung cancer is diagnosis, surgery, chemotherapy and/or radiation therapy, targeted therapy which depend on the cancer's type. In comparison with western medicine, the medical treatment for cancer patients in TKM is significantly different from conventional treatment. "Pattern identification" is an essential intermediate step for diagnosis and medication in TKM. It is a concept where the therapeutic systems- determining the cause, the origin of the illness, the patient's physical condition, and the treatment are decided after various symptoms and signs in the patients are comprehensively observed. This has been used in Eastern Asia, China, Taiwan, Singapore, and so on. However, pattern identification has not been applied widely in clinical trials in which objective criteria are important. Because, it has a controversial issue because pattern identification relies on observation and clinical experience of TKM practitioners themselves, which may lack scientific objectivity. In order to overcome these limitations, many studies have attempted to standardize pattern identification-based diagnosis criteria with objective numerical values such as body mass index (BMI), heart rate variability (HRV), *etc*[17–18]. Among them, Body Constitution Questionnaire (BCQ) was developed in Taiwan to determine patients based on traditional Chinese medicine (TCM) from 2008 to 2012[19–22], symptoms of the patients are digitized by questionnaires composed of nineteen to twenty-three simple-to-answer questions. This criterion

may make pattern identification more objective.

But, it still causes a controversial. Therefore, standardized scientific technologies for TKM-based prognostic criteria are necessary for the medical treatment. The primary objective of this study was to determine whether WES can establish the standardized methodology with pattern identification, framework of TKM, for lung cancer patients. If so, in addition, this study is preliminary research for the discovering of biomarkers based on pattern identification for lung cancer patients.

2. Methods

1) Study design

The study was conducted at the Institutional Review Board (IRB) of Kyung Hee University (Seoul, Korea) and Korea University Guro Hospital (Seoul, Korea). The reference numbers are KUGH15010-002 for Korea University Guro Hospital and KHSIRB-15-005 (RA) for Kyung Hee University. The present study was conducted from May 2015 to June 2016. The subjects are 30 lung cancer patients and all of them are Korean males and females aged from 48 to 80. For thirty lung cancer patients, we recorded epidemiologic information, including gender, age, surgical history, smoking status, diagnosis (clinical TNM stage, pathological TNM stage, cancer subtypes) (Table 1) and BCQ. The current research is a cross-sectional study that the biomarkers from lung cancer patients is involved with the classification of WES by pattern identification in TKM.

Table 1. Information of patients

		Xu (N=4)	Stasis (N=11)	Gentleness (N=15)
Age - yr				
	Median	66.5	64.2	66.6
	Interquartile range	57-75	48-80	56-74
Sex - no. (%)				
	Male	3 (75)	9 (81.8)	6 (40)
	Female	1 (25)	2 (18.2)	9 (60)
Stage - no. (%)				
	IA(T1N0)	2 (50)	4 (36.4)	7 (46.7)
	IB(T2N0)	2 (50)	5 (45.5)	8 (53.3)
	T3	0	1 (9.1)	0
	T4	0	1 (9.1)	0
Tumor diameter - no. (%)				
	≤3 cm	2 (50)	5 (45.5)	10 (66.7)
	>3 cm	1 (25)	6 (54.5)	5 (33.3)
	Unknown	1 (25)	0	0
Histologic characteristics - no. (%)				
	Adenocarcinoma	2 (50)	6 (54.5)	9 (60)
	Squamous-cell	2 (50)	4 (36.4)	5 (33.3)
	Other	0	1 (9.1)	1 (6.7)
Smoking status - no. (%)				
	Current or former smoker	1 (25)	6 (54.5)	3 (20)
	Nonsmoker	3 (75)	5 (45.5)	12 (80)
	Unknown	0	0	0

2) Patients and Sample Collection

Under written informed consent, we collected fresh surgical specimens of primary lung cancer from patients who underwent major lung resection at Korea University Guro Hospital (Seoul, Korea). The tissue samples were transferred to a tube and stored in −80℃ deep freezer until further study. After the pathological examination, the tumor tissue samples (30) were determined with a cancer type.

3) Whole exome sequencing

Tissue samples were used to generate purified DNA using the Qiagen DNeasy kit with the Qiagen RNase (Qiagen, Hilden, Germany). A treatment option was in accordance with the manufacturer's instructions. DNA was assayed and quality controlled using the Qubit 2.0 Fluorimeter (Invitrogen) according to the manufacturer's instructions. Whole exome sequencing was performed using the Ion AmpliSeq™ Exome RDY kit (Life Technologies), which targets exons of whole human genome. Library construction was performed using the Ion AmpliSeq™ Library Kit 2.0 (Life Technologies) and library templates were prepared and barcoded for sequencing using the Ion One-Touch System as per manufacturer's instructions. The two barcoded samples were multiplexed for Ion PI Chip (Life Technologies) and sequenced on the Ion Proton Sequencer System (Life Technologies).

4) Sequence data analysis

Sequencing reads were processed using Ion Torrent Suite software v 4.0.2 (Life Technologies). Demultiplexed samples were assessed for sequencing quality and high quality sequencing reads were mapped to the complete hg19 human genome (UCSC version, February 2009). Variant discovery was performed using Torrent Variant Caller v 4.2 (Life Technologies), a software plug-in for the Ion Torrent Suite software. Significant variants were identified as those predicted to result in frameshift, missense, nonsense or essential splice site mutations predicted to having at least four impair protein function with The Genomic Evolutionary Rate Profiling (GERP)++, Scale-Invariant Feature Transform (SIFT), Likelihood-Ratio Test (LRT), Mutation Assessor, Mutation Taster, PhyloP and PolyPhen2 HDIV methods.

5) Body constitution questionnaire (BCQ) classification

The BCQ+, BCQ−, BCQs questionnaires were developed in order to determine Yang deficiency (Yang_Xu, YangX), Yin deficiency (Yin_Xu YinX), and Stasis patients based on TCM in objective manners in Taiwan from 2008 to 2012[19-22]. To classify the pattern identification of lung cancer patients, our group conducted the BCQ questionnaire to the subjects who participated in the study for reference to these 3 questionnaires. The diagnoses of Yang Xu were over 30.5 points by BCQ+ (5 scale, 19-item) and these of Yin Xu were over 29.5 points by BCQ− (5 scale, 19-item). We analyzed patients' information− epidemiologic information, including gender, age, surgical history, smoking status, diagnosis (clinical TNM stage, pathological TNM stage, cancer subtypes) (Table 1) and BCQ. In the WES analysis, subjects who correspond to either Yang deficiency or Yin deficiency are classified as "Xu", subjects who are diagnosed of Stasis through BCQ are categorized as "Stasis", and patients who belong to none of these categories are classified as "Gentleness" group.

6) Statistical analysis

Fisher's exact tests for allele-frequency diffe-

rences in variants were conducted in the SVS8 software (Golden Helix Inc., Bozeman, MT, USA). The pattern identification stratification using principal components analysis (PCA) was also examined using SVS8. Gene ontology enrichment was performed by the gene ontology (GO) database using the WebGestaltusing GOTM (http://bioinfo.vanderbilt.edu/gotm/), which was based on hypergeometric distribution to show the over-represented gene ontology categories ($p < 0.05$). P-value was calculated using BINOMDIST function on the basis of overrepresentation of gene ontology categories when compared to all genes on genome ($p < 0.05$).

3. Results

The objectives of this preliminary research are biomarkers discovery on the pattern identification. It will be performed to identify variants which can be used as the prognostic element for patient-specific TKM for lung cancer patients. The patients' samples are divided into three groups based on pattern identification (Xu, Stasis, and Gentleness) in TKM and each group is individually explored. The identified variants from each group are categorized and compared one another to see the reliability of the experiments.

In the present study, we recruited a total of 30 subjects including 4 Xu, 11 Stasis, and 15 Gentleness group. Using the tissue sample of the subjects, we performed a WES to examine the genetic differences among three groups. Subjects were composed of normal and cancer tissues to identify cancer specific polymorphisms. A total of seven in silico analyses (GERP++, SIFT, LRT, Mutation Assessor, Mutation Taster, PhyloP and PolyPhen2 HDIV) were performed to predict functional role of significant SNPs.

The exome sequencing identified a total of 27,844 of frameshift, missense, nonsense or essential splice site polymorphisms. Among the polymorphisms, 23,490 SNPs were germline variants and showed functional changes (damaging) in at least four *in silico analyses*. Using the variants, we performed Fisher's exact tests to isolate markers showing frequency differences of three groups. Our results revealed that 12 variants showed significant frequency differences between Xu and Stasis groups ($P < 0.0005$). As similar, 18 and 10 variants were identified in analysis for Xu vs. Gentleness group and Stasis vs. Gentleness group, respectively ($P < 0.001$). Of these, there were two novel variants, Val1207Leu in *C4B_2* and Phe613Ser in *MEGF11*. Detailed minor allele frequencies of the variants were displayed in Table 2 with their amino acid changes, locations, and genes. Our exome sequencing also found 8,792 lung cancer specific variants and statistical analyses among the groups identified 6, 34, and 12 variants which showed significant allele frequency differences in the comparison groups; Xu vs. Stasis, Xu vs. Gentleness group, and Stasis vs. Gentleness group (Table 3). Among them, 15 variants were novel, e.g. Gly2260Arg in *HECTD4* and Glu663Asp in *MYCBPAP*.

To explore whether the pattern identification specific variants could differentiate each group, PCA analysis was performed. As a result, in germline data set, Xu group was divided from other groups. However, Stasis and Gentleness group were not clearly separated each other (Figure 1A). Analysis using somatic variants also showed similar result (Figure 1B).

To investigate the potential function of the pattern identification specific variants, gene ontology analysis was performed. Ontology analysis for germline variants showed that 6 genes (FUT3, PPP1R3G, MEGF11, TULP1, OR52N1, and OR52N2) showed most significance and were related with the carbohydrate biosynthetic process, eye morphogenesis, and olfactory receptor activity(Table 4). Among the genes, FUT3 was demonstrated that the gene was responsible for tissue differentiation

Table 2. Germline variants and genes showing significant frequency differences (P <0.001) between pattern identification groups

pattern identification group	Variants (chr:position)	rs#	Gene	AA Change	Alleles	Minor allele frequency		P-value
						Xu (n=4)	Stasis (n=11)	
Xu vs. Stasis	19:55671337	rs890872	DNAAF3	Asp412Asn	T>C	1.000	0.091	7.7E-06
	17:41133071	rs1708875	RUNDC1	Trp160Arg	C>T	1.000	0.091	7.7E-06
	6:5086558	rs436556	PPP1R3G	Pro280Gln	A>C	1.000	0.136	2.8E-05
	6:31324531	rs1131204	HLA-B	Ala93Thr	T>C	1.000	0.136	2.8E-05
	6:46679303	rs1805018	PLA2G7	Ile198Thr	A>G	0.750	0.000	4.7E-05
	3:13670536	rs9843344	FBLN2	Thr854Ala	G>A	0.750	0.000	4.7E-05
	6:31964320	.	C4B_2	Val1207Leu	C>G	1.000	0.182	8.5E-05
	19:5844649	rs812936	FUT3	Arg68Trp	A>G	1.000	0.182	8.5E-05
	15:65157482	rs2010875	PLEKHO2	Pro290Ser	T>C	0.625	0.000	0.0004
	6:31324528	rs1071817	HLA-B	Gln94Lys	G>T	0.000	0.727	0.0005
	15:65621441	rs12907128	IGDCC3	Val751Leu	A>C	1.000	0.273	0.0005
	15:93588336	rs4238485	RGMA	Asp415Glu	C>A	1.000	0.273	0.0005
						Xu (n=4)	Gentleness (n=15)	
Xu vs. Gentleness	6:35479574	rs7764472	TULP1	Thr67Arg	C>G	1.000	0.200	6.1E-05
	17:41133071	rs1708875	RUNDC1	Trp160Arg	C>T	1.000	0.200	6.1E-05
	19:36243089	rs170758	LIN37	Val16Ala	C>T	0.750	0.067	0.0003
	3:13670536	rs9843344	FBLN2	Thr854Ala	G>A	0.750	0.067	0.0003
	16:88872145	rs507329	CDT1	Cys234Arg	C>T	0.750	0.067	0.0003
	19:55671337	rs890872	DNAAF3	Asp412Asn	T>C	1.000	0.267	0.0003
	19:52223121	rs7248778	HAS1	Cys14Arg	G>A	1.000	0.267	0.0003
	19:5844649	rs812936	FUT3	Arg68Trp	A>G	1.000	0.267	0.0003
	19:52941618	rs1366258	ZNF534	Glu302Ala	C>A	0.875	0.167	0.0004
	7:92760738	rs10282508	SAMD9L	Asn1516Thr	T>G	0.875	0.167	0.0004
	19:52938361	rs11084161	ZNF534	Thr57Asn	A>C	0.875	0.167	0.0004
	19:52938459	rs11084163	ZNF534	Thr90Ala	G>A	0.875	0.167	0.0004
	19:52941540	rs1366257	ZNF534	Gly276Glu	A>G	0.875	0.167	0.0004
	17:8046772	rs2585405	PER1	Ala962Pro	G>C	1.000	0.300	0.0005
	19:52938417	rs11084162	ZNF534	Ile76Val	G>A	0.875	0.200	0.0009
	3:129281980	rs2713625	PLXND1	Ser1542Asn	T>C	0.500	0.000	0.0009
	15:66214795	.	MEGF11	Phe613Ser	A>G	0.500	0.000	0.0009
	9:131085373	rs3003601	COQ4	Gly50Ala	C>G	0.500	0.000	0.0009
						Stasis (n=11)	Gentleness (n=15)	
Stasis vs. Gentleness	15:90213229	rs6496589	PLIN1	Pro194Ala	C>G	0.818	0.133	7.9E-07
	11:6231731	rs10769671	C11orf42	Pro242Ser	C>T	0.091	0.700	1.3E-05
	20:1895950	rs1135200	SIRPA	Asp95Glu	G>C	0.818	0.267	0.0002
	20:1895963	rs17855613	SIRPA	Asn100Asp	G>A	0.818	0.267	0.0002
	16:3293922	rs1231123	MEFV	Asp424Glu	A>T	0.636	0.133	0.0003
	11:5842310	rs8181529	OR52N2	Ser249Ala	T>G	0.045	0.500	0.0006
	20:1895965	rs17855614	SIRPA	Asn100Lys	A>C	0.773	0.267	0.0006
	19:50926264	rs1673030	SPIB	Leu84Pro	T>C	0.545	0.100	0.0007
	19:687142	rs8102982	PRSS57	Pro143Leu	A>G	0.182	0.667	0.0007
	11:5809308	rs7934670	OR52N1	Phe247Ile	T>A	0.409	0.033	0.001

Note: Fisher's exact tests for allele-frequency differences in variants were conducted in the SVS8 software.

Table 3. Somatic variants and genes showing significant frequency differences ($P<0.05$) between pattern identification groups

pattern identification group	Variants (chr:position)	rs#	Gene	AA Change	Alleles	Minor allele frequency		*P*-value
						Xu (n=4)	Stasis (n=11)	
Xu vs. Stasis	12:112654894	.	HECTD4	Gly2260Arg	C>G	0.375	0.000	0.01
	16:4033436	rs2230739	ADCY9	Ile772Met	T>C	0.375	0.000	0.01
	19:22155918	rs10425763	ZNF208	Lys640Glu	T>C	0.375	0.000	0.01
	22:19119751	rs1052763	TSSK2	Thr280Met	C>T	0.375	0.000	0.01
	16:1279717	rs201728868	TPSB2	Arg28Gln	C>T	0.375	0.045	0.05
	17:48603319	.	MYCBPAP	Glu663Asp	G>C	0.375	0.045	0.05
						Xu (n=4)	Gentleness (n=15)	
Xu vs Gentleness	16:4033436	rs2230739	ADCY9	Ile772Met	T>C	0.375	0.000	0.007
	19:22155918	rs10425763	ZNF208	Lys640Glu	T>C	0.375	0.000	0.007
	22:19119751	rs1052763	TSSK2	Thr280Met	C>T	0.375	0.000	0.007
	15:43820245	.	MAP1A	Gly2192Arg	G>C	0.375	0.033	0.02
	2:85895338	rs2077079	SFTPB	His2Pro	T>G	0.375	0.033	0.02
	8:92330489	.	SLC26A7	Thr175Ser	A>T	0.375	0.033	0.02
	7:64291987	rs35742014	ZNF138	Gly97Ser	G>A	0.250	0.000	0.04
	7:150761314	rs2303929	SLC4A2	Gly26Glu	G>A	0.250	0.000	0.04
	6:31617055	.	BAG6	Pro115Arg	G>C	0.250	0.000	0.04
	8:381344	rs28438773	FBXO25	Ile46Met	C>G	0.250	0.000	0.04
	12:11461798	rs77336955	PRB4	Arg40Pro	C>G	0.250	0.000	0.04
	12:71533534	rs3763978	TSPAN8	Gly73Ala	C>G	0.250	0.000	0.04
	13:42465713	rs9562362	VWA8	Arg165His	C>T	0.250	0.000	0.04
	13:52676275	rs34756139	NEK5	Lys255Gln	T>G	0.250	0.000	0.04
	15:66214795	.	MEGF11	Phe613Ser	A>G	0.250	0.000	0.04
	16:84035446	rs2271298	NECAB2	Leu353Val	C>G	0.250	0.000	0.04
	19:6755155	rs12608960	SH2D3A	Asp223Gly	T>C	0.250	0.000	0.04
	19:52520607	rs61574510	ZNF614	Gly82Arg	C>T	0.250	0.000	0.04
	20:3675498	rs709012	SIGLEC1	His919Pro	T>G	0.250	0.000	0.04
	1:228467076	.	OBSCN	Ser2443Pro	T>C	0.250	0.000	0.04
	2:64863734	.	SERTAD2	Gln91Leu	T>A	0.250	0.000	0.04
	6:32064098	rs204896	TNXB	Arg511His	C>T	0.250	0.000	0.04
	7:149511985	rs10952230	SSPO	Asn3512Ser	A>G	0.250	0.000	0.04
	9:98691137	rs2274654	ERCC6L2	Val592Ala	T>C	0.250	0.000	0.04
	9:101765841	rs10988532	COL15A1	Thr391Met	C>T	0.250	0.000	0.04
	9:139973820	rs7037849	UAP1L1	Ala319Val	C>T	0.250	0.000	0.04
	10:64927823	rs1935	JMJD1C	Glu2298Asp	C>G	0.250	0.000	0.04
	11:49227620	rs202676	FOLH1	Tyr75His	A>G	0.250	0.000	0.04
	12:9317764	.	PZP	Val820Ile	C>T	0.250	0.000	0.04
	17:74288418	.	QRICH2	Val631Asp	A>T	0.250	0.000	0.04
	19:17337882	rs891203	OCEL1	Ala109Gly	C>G	0.250	0.000	0.04
	20:17600357	rs11960	RRBP1	Ser766Leu	G>A	0.250	0.000	0.04
	22:22869742	rs361959	ZNF280A	Lys71Asn	C>A	0.250	0.000	0.04
	22:24622648	rs2275984	GGT5	Lys330Arg	T>C	0.250	0.000	0.04

Note: Fisher's exact tests for allele-frequency differences in variants were conducted in the SVS8 software.

Table 3. Somatic variants and genes showing significant frequency differences ($P<0.05$) between pattern identification groups (continued)

pattern identification group	Variants (chr:position)	rs#	Gene	AA Change	Alleles	Minor allele frequency		*P*-value
						Stasis (n=11)	Gentleness (n=15)	
Stasis vs Gentleness	17:74288595	.	QRICH2	Arg572His	C>T	0.227	0.000	0.01
	12:11461271	.	PRB4	Pro216Ser	G>A	0.227	0.000	0.01
	4:3495095	rs9684786	DOK7	Gly461Asp	G>A	0.182	0.000	0.03
	2:49189921	rs6166	FSHR	Ser680Asn	C>T	0.182	0.000	0.03
	2:166789527	.	TTC21B	Met251Leu	T>G	0.182	0.000	0.03
	3:119128387	.	ARHGAP31	Pro564Ala	C>G	0.182	0.000	0.03
	5:54404053	rs2270627	GZMA	Trp153Leu	G>T	0.182	0.000	0.03
	8:623435	.	ERICH1	Pro306Leu	G>A	0.182	0.000	0.03
	11:8751889	rs3794153	ST5	Lys316Asn	C>G	0.182	0.000	0.03
	17:6493198	rs1443417	KIAA0753	Gln896Arg	T>C	0.182	0.000	0.03
	22:51015838	rs3213445	CPT1B	Ile66Val	T>C	0.182	0.000	0.03
	9:7799653	rs1127430	TMEM261	Pro28Thr	G>T	0.000	0.200	0.03

Note: Fisher's exact tests for allele-frequency differences in variants were conducted in the SVS8 software.

and tumor metastasis (27453266) in comparison between Xu and Stasis. Other significant SNPs for two groups (Xu vs. Gentleness group and Stasis vs. Gentleness group) were responsible for eye morphogenesis and olfactory receptor activity. Analysis using somatic variants also identified 6 significant genes (MYCBPAP, TSSK2, SLC4A2, SLC26A7, ST5, and ARHGAP31) and showed close relationship in multicellular organism reproduction, anion-anion antiporter activity, and GTPase regulator activity (Table 5).

4. Discussion

In TKM, Xu refers to a deficiency of qi and fluid, and Stasis refers to blood stagnation[4]. Both suggest a state of imbalance and abnormal functioning of the whole body[23-26]. Imbalanced fluid and blood stream of body results in solid tumor like as lung cancer in organs. Human body

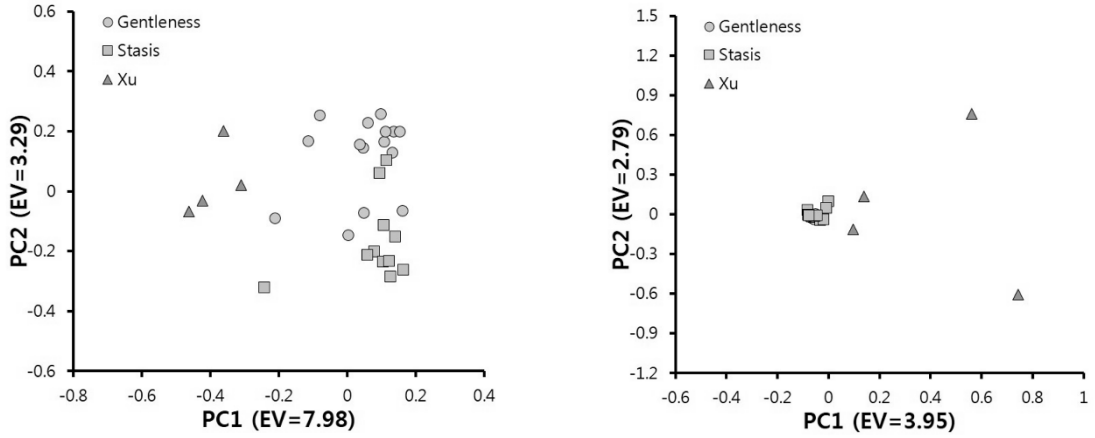

Figure 1. Principal component analysis of pattern identification subjects: First and second principal components of our pattern identification samples based on significantly different germline variants (A) and somatic variants (B) (circle: Gentleness, triangle: Xu, square: Stasis).

Table 4. Gene ontology analysis of the pattern identification associated genes with germline variants

Comparison	Gene Ontology Category		Genes	#Reference genes in category	#Observed genes in category	#Expected genes in category	Ratio of enrichment	*P*-value
Xu vs. Stasis	Biological process	Carbohydrate biosynthetic process	FUT3, PPP1R3G	167	2	0.09	21.9	0.004
		Response to external stimulus	HLA-B, PLA2G7, RGMA	1323	3	0.72	4.15	0.03
		Oxidation-reduction process	PLA2G7, PPP1R3G	558	2	0.31	6.55	0.03
		Single-organism carbohydrate metabolic process	FUT3, PPP1R3G	592	2	0.32	6.18	0.04
		Chemotaxis	PLA2G7, RGMA	595	2	0.33	6.15	0.04
	Molecular function	Phospholipid binding	PLA2G7, PLEKHO2	490	2	0.19	10.4	0.01
	Cellular component	Golgi membrane	FUT3, HLA-B	553	2	0.26	7.71	0.03
Xu vs Gentleness	Biological process	Eye morphogenesis	MEGF11, TULP1	126	2	0.1	19.35	0.005
		Response to light stimulus	PER1, TULP1	203	2	0.17	12.01	0.01
		Anatomical structure formation involved in morphogenesis	PLXND1, MEGF11, TULP1, DNAAF3	1594	4	1.31	3.06	0.03
		Cell projection organization	PLXND1, TULP1, DNAAF3	934	3	0.77	3.92	0.04
		Cell adhesion	HAS1, MEGF11, FBLN2	954	3	0.78	3.83	0.04
		Regulation of cell migration	HAS1, PLXND1	396	2	0.32	6.16	0.04
	Molecular function	Transferase activity, transferring hexosyl groups	FUT3, HAS1	189	2	0.11	17.97	0.005
Stasis vs Gentleness	Biological process	Immune system process	SPIB, MEFV, SIRPA	1792	3	0.61	4.9	0.02
		Response to wounding	MEFV, SIRPA	1109	2	0.38	5.28	0.05
	Molecular function	Olfactory receptor activity	OR52N1, OR52N2	419	2	0.19	10.42	0.01
	Cellular component	Microtubule cytoskeleton	SPIB, MEFV	863	2	0.35	5.65	0.05

Note: Fisher's exact tests for allele-frequency differences in variants were conducted in the SVS8 software.

Table 5. Gene ontology analysis of the pattern identification associated genes with somatic variants

Comparison	Gene Ontology Category		Genes	#Reference genes in category	#Observed genes in category	#Expected genes in category	Ratio of enrichment	*P-value
Xu vs. Stasis	Biological process	Multicellular organism reproduction	MYCBPAP, TSSK2	638	2	0.26	7.64	0.03
		Synaptic transmission	MYCBPAP, ADCY9	651	2	0.27	7.49	0.03
		Protein metabolic process	HECTD4, ADCY9, TSSK2, TPSB2	3896	4	1.6	2.5	0.05
	Biological process	Bicarbonate transport	SLC4A2, SLC26A7	28	2	0.05	37.32	0.001
		Cell adhesion	SSPO, MEGF11, SIGLEC1, COL15A1, TNXB	954	5	1.83	2.74	0.03
		Cell-matrix adhesion	SIGLEC1, TNXB	149	2	0.29	7.01	0.03
		Cell-cell adhesion	MEGF11, SIGLEC1, TNXB	404	3	0.77	3.88	0.04
		JNK cascade	SH2D3A, TNXB	170	2	0.33	6.15	0.04
Xu vs Gentleness	Molecular function	Anion:anion antiporter activity	SLC4A2, SLC26A7	19	2	0.04	53.63	0.0006
		Peptidase inhibitor activity	SSPO, PZP	172	2	0.34	5.92	0.04
		Guanyl-nucleotide exchange factor activity	OBSCN, SH2D3A	183	2	0.36	5.57	0.05
	Cellular component	Basolateral plasma membrane	SLC4A2, MEGF11, SLC26A7	167	3	0.29	10.21	0.003
		Collagen	COL15A1, TNXB	89	2	0.16	12.77	0.01
		Extracellular space	SSPO, PZP, SFTPB, COL15A1, TNXB	856	5	1.51	3.32	0.02
		Extracellular region	SSPO, PZP, SFTPB, COL15A1, TNXB, PRB4, SIGLEC1, VWA8	2140	8	3.76	2.12	0.03
		Extracellular matrix part	COL15A1, TNXB	185	2	0.33	6.15	0.04
Stasis vs Gentleness	Molecular function	GTPase regulator activity	ST5, ARHGAP31	446	2	0.2	9.79	0.02
	Cellular component	Cell junction	DOK7, ARHGAP31	737	2	0.35	5.78	0.04

Note: Fisher's exact tests for allele-frequency differences in variants were conducted in the SVS8 software

is influences by genetically and environmentally, so we checked germline and somatic mutation variants in lung cancer patients based on the pattern identification. We were able to identify 40 variants in 29 genes in association with germline difference of pattern identification group and 52 variants in 47 genes in somatic cancer tissues. Our ontology analysis for germline variants showed that 6 genes (FUT3, PPP1R3G, MEGF11, TULP1, OR52N1, and OR52N2) showed most significance and were related with the carbohydrate biosynthetic process, eye morphogenesis, and olfactory receptor activity. Among the genes, FUT3 was demonstrated that the gene was responsible for tissue differentiation and tumor metastasis (27453266). In details, in germline data set, we found that Stasis group is different from SPIB, MEFV, and SIRPA genes related to immune system process and response to wounding comparison to Gentleness group. This result suggests that Stasis group which has stagnated blood may affect on abnormal immune system process in comparison to another group. And Xu group is different from HAS1, MEGF11, FBLN2, and PLXND1 genes, which are related to cell migration and adhesion, when comparing to Gentleness group. The deficiency of fluid or blood of body blocks migration and adhesion of cells when cells need to move proper organ. Xu group has different FUT3 gene compared to Stasis group. As mentioned above, it is documented that FUT3 gene was responsible for tumor metastasis. Therefore, it is believed that both 2 groups' patient may suffer severer than Gentleness group owing to possibility of tumor metastasis if the disease relapse.

Analysis using somatic variants also identified 6 significant genes (MYCBPAP, TSSK2, SLC4A2, SLC26A7, ST5, and ARHGAP31). Previous study reported that MYCBPAP was related to breast cancer invasion and metastasis (19416474). In addition, ST5 showed suppress ability of the tumorigenicity of Hela cells in nude mice (1390339). Several studies also reported association between the other genes and various cancers, but functional role of the genes was not clear. In details, we found that Stasis group has differences in ST5, DOK7, and ARHGAP31 genes related to cell junction when comparing to Gentleness group. This result suggests that Stasis group which has stagnated blood may affect on cell to cell interaction in comparison to another group. Xu group has different from MYCBPAP gene compared to Stasis group. As previously stated, this gene is well known to responsible for tumor metastasis and invasion. Both 2 groups' patients may have possibility to suffer severer than Gentleness group due to chance of tumor metastasis and invasion if the disease recurrence likewise germline variants.

A limitation of the present study is small number of subjects. Especially, only four lung cancer patients of Xu group were used for the statistical analyses. Therefore, further studies were required to confirm the significance of our variants. However, in our PCA analysis, Xu group was clearly separated from other groups in both germline and somatic variants although Stasis and Gentleness group were not separated in somatic variants. These results suggest the use of the variants as possible prognostic markers in clinical trials.

The identification of biomarkers based on pattern identification considering individual patient's symptom could help determine which patients might avoid unnecessary therapy or may provide that information could outweigh the risk of treatment-related toxicity[27-28]. For example, Jin group has reported that the integrated therapy of Chinese herbal remedy, which was prescribed different herbal drugs for patients having a different symptom, and radiotherapy for treatment of abdominal malignant tumor showed valuable results like as improving living quality and prolonging survival time[28]. Until now pattern identification, itself, has not been applied widely

in clinical trials because of relying on observation and clinical experience of TKM practitioners themselves without scientific objectivity. However, identifying biomarkers, standardized scientific tools, for TKM-based prognostic criteria through WES make western medicine practitioners and researchers to exploring the potential of TKM to enhance conventional treatment of cancer patients.

5. Conclusions

Overall, our study suggests that this is the first attempt to explore the specific variants discovery in whole exome regions for prognostic biomarkers of lung cancer based on pattern identification of Korean Traditional Medicine. Our study identified 40 variants in 29 genes in association with germline difference of pattern identification groups and 52 variants in 47 genes in somatic cancer tissues. It is likely to provide valuable information during the cancer patient care process to assist in understanding the pattern identification mechanism and in the selection of the most appropriate patient population to achieve predictable outcomes.

List of abbreviations

Body constitution questionnaire, BCQ; Body mass index, BMI; The Genomic evolutionary rate profiling, GERP; Gene ontology, GO; Heart rate variability, HRV; Institutional Review Board, IRB; Likelihood-ratio test, LRT; Principal component analysis, PCA; Scale-invariant feature transform, SIFT; single nucleotide polymorphism, SNP; traditional Chinese medicine, TCM; traditional Korean medicine, TKM; Whole exome sequencing, WES;

Declarations

Ethics approval and consent to participate

The study was conducted at the Institutional Review Board (IRB) of Kyung Hee University (Seoul, Korea) and Korea University Guro Hospital (Seoul, Korea). The reference numbers are KUGH15010-002 for Korea University Guro Hospital and KHSIRB-15-005 (RA) for Kyung Hee University. Prior to undertaking any study-related procedures, written informed consent will be collected from all participants.

Consent for publication

Not applicable

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

JHK contributed to the design of this study, collected the specimens, and mainly participated in the drafting the manuscript. HSC contributed to the data collection and performed the analysis. CC contributed to the design of this study, to analyze the BCQ questionnaire and classification, and helped to draft the manuscript. SK contributed to the collection of the specimens. HKK contributed to the providing the specimens. SGK, and HDS contributed to the funding, is the general supervisors for this research and participated in both the study design.

Acknowledgements

The authors thank the families, patients, and control volunteers who participated in this research.

The author thanks Mr. MJP for English editing support.

References

1. Warr A, Robert C, Hume D, Archibald A, Deeb N, Watson M: Exome Sequencing: Current and Future Perspectives. *G3 (Bethesda)* 2015, 5(8): 1543-1550.
2. Worthey EA, Mayer AN, Syverson GD, Helbling D, Bonacci BB, Decker B, Serpe JM, Dasu T,

Tschannen MR, Veith RL et al: Making a definitive diagnosis: successful clinical application of whole exome sequencing in a child with intractable inflammatory bowel disease. *Genet Med* 2011, 13(3):255-262.
3. Wang X, Wang H, Cao M, Li Z, Chen X, Patenia C, Gore A, Abboud EB, Al-Rajhi AA, Lewis RA *et al* : Whole-exome sequencing identifies ALMS1, IQCB1, CNGA3, and MYO7A mutations in patients with Leber congenital amaurosis. *Hum Mutat* 2011, 32(12):1450-1459.
4. Sassi C, Guerreiro R, Gibbs R, Ding J, Lupton MK, Troakes C, Lunnon K, Al-Sarraj S, Brown KS, Medway C et al: Exome sequencing identifies 2 novel presenilin 1 mutations (p.L166V and p.S230R) in British early-onset Alzheimer's disease. *Neurobiol Aging* 2014, 35(10):2422 e2413-2426.
5. Byun M, Abhyankar A, Lelarge V, Plancoulaine S, Palanduz A, Telhan L, Boisson B, Picard C, Dewell S, Zhao C et al: Whole-exome sequencing-based discovery of STIM1 deficiency in a child with fatal classic Kaposi sarcoma. *J Exp Med* 2010, 207(11):2307-2312.
6. Yan XJ, Xu J, Gu ZH, Pan CM, Lu G, Shen Y, Shi JY, Zhu YM, Tang L, Zhang XW *et al*: Exome sequencing identifies somatic mutations of DNA methyltransferase gene DNMT3A in acute monocytic leukemia. *Nat Genet* 2011, 43(4):309-315.
7. Greif PA, Dufour A, Konstandin NP, Ksienzyk B, Zellmeier E, Tizazu B, Sturm J, Benthaus T, Herold T, Yaghmaie M et al: GATA2 zinc finger 1 mutations associated with biallelic CEBPA mutations define a unique genetic entity of acute myeloid leukemia. *Blood* 2012, 120(2):395-403.
8. Snape K, Ruark E, Tarpey P, Renwick A, Turnbull C, Seal S, Murray A, Hanks S, Douglas J, Stratton MR et al: Predisposition gene identification in common cancers by exome sequencing: insights from familial breast cancer. *Breast Cancer Res Treat* 2012, 134(1):429-433.
9. Cai J, Li L, Ye L, Jiang X, Shen L, Gao Z, Fang W, Huang F, Su T, Zhou Y et al: Exome sequencing reveals mutant genes with low penetrance involved in MEN2A-associated tumorigenesis. *Endocr Relat Cancer* 2015, 22 (1):23-33.
10. Chakraborty R, Savani BN, Litzow M, Mohty M, Hashmi S: A perspective on complementary/ alternative medicine use among survivors of hematopoietic stem cell transplant: Benefits and uncertainties. *Cancer* 2015, 121(14):2303-2313.
11. Efferth T, Li PC, Konkimalla VS, Kaina B: From traditional Chinese medicine to rational cancer therapy. *Trends Mol Med* 2007, 13(8): 353-361.
12. Kim JH, Shin YC, Ko SG: Integrating traditional medicine into modern inflammatory diseases care: multitargeting by Rhus verniciflua Stokes. *Mediators Inflamm* 2014, 2014: 154561.
13. Kim K-S KS-H, Eo W-K, Cheon S-H, Eom S-K, Jo H-J: Clinical research methodology for Traditional Korean Medicine treatment of lung cancer: Evidence-based approach. *Journal of Korean Medical Classics* 2010, 23 (4):39-62.
14. Park B-k LJ-h, Cho C-k, Shin H-k, Eom S-k, Yoo H-s. : Systemic review of clinical studies about Oriental medical treatment of cancer in Korea. *Korean Journal of Oriental Internal Medicine* 2008, 29(4):1061-1074.
15. Swarup AB, Barrett W, Jazieh AR: The use of complementary and alternative medicine by cancer patients undergoing radiation therapy. *Am J Clin Oncol* 2006, 29(5):468-473.
16. Hunt KJ, Coelho HF, Wider B, Perry R, Hung SK, Terry R, Ernst E: Complementary and alternative medicine use in England: results from a national survey. *Int J Clin Pract* 2010,

64(11):1496-1502.
17. Adler SR, Fosket JR: Disclosing complementary and alternative medicine use in the medical encounter: a qualitative study in women with breast cancer. *J Fam Pract* 1999, 48(6):453-458.
18. Gansler T, Kaw C, Crammer C, Smith T: A population-based study of prevalence of complementary methods use by cancer survivors: a report from the American Cancer Society's studies of cancer survivors. *Cancer* 2008, 113(5):1048-1057.
19. Chen LL, Lin JS, Lin JD, Chang CH, Kuo HW, Liang WM, Su YC: BCQ+: a body constitution questionnaire to assess Yang-Xu. Part II: Evaluation of reliability and validity. F*orsch Komplementmed* 2009, 16(1):20-27.
20. Lin JD, Chen LL, Lin JS, Chang CH, Huang YC, Su YC: BCQ-: a body constitution questionnaire to assess Yin-Xu. Part I: establishment of a provisional version through a Delphi process. *Forsch Komplementmed* 2012, 19(5):234-241.
21. Lin JS, Chen LL, Lin JD, Chang CH, Huang CH, Mayer PK, Su YC: BCQ-: A Body Constitution Questionnaire to assess Yin-Xu. Part II: evaluation of reliability and validity. *Forsch Komplementmed* 2012, 19(6):285-292.
22. Su YC, Chen LL, Lin JD, Lin JS, Huang YC, Lai JS: BCQ+: a body constitution questionnaire to assess Yang-Xu. Part I: establishment of a first final version through a Delphi process. *Forsch Komplementmed* 2008, 15(6):327-334.
23. Chen Z, Wang P: Clinical Distribution and Molecular Basis of Traditional Chinese Medicine ZHENG in Cancer. *Evid Based Complement Alternat Med* 2012, 2012:783923.
24. Dai J, Fang J, Sun S, Chen Q, Cao H, Zheng N, Zhang Y, Lu A: ZHENG-Omics Application in ZHENG Classification and Treatment: Chinese Personalized Medicine. *Evid Based Complement Alternat Med* 2013, 2013:235969.
25. Berle CA, Cobbin D, Smith N, Zaslawski C: A novel approach to evaluate Traditional Chinese Medicine treatment outcomes using pattern identification. *J Altern Complement Med* 2010, 16(4):357-367.
26. Jiang M, Zhang C, Zheng G, Guo H, Li L, Yang J, Lu C, Jia W, Lu A: Traditional chinese medicine zheng in the era of evidence-based medicine: a literature analysis. *Evid Based Complement Alternat Med* 2012, 2012:409568.
27. Shi X, Tian L, Zhu XD, Wang HM, Qin H: Effect of Chinese drugs combining with chemotherapy on quality of life in 146 children with solid tumor. *Chin J Integr Med* 2011, 17(1):31-34.
28. Hong J, Xiangwei W, Yanping C, Qionghui L, Wen L: An analysis of the long-term therapeutic effect of the integrated therapy of traditional Chinese medicine and radiotherapy on abdominal malignant tumor. *J Tradit Chin Med* 2005, 25(2):125-128.