Association analysis of SNPs with CT image-based phenotype of emphysema progression in heavy smokers

Hidenobu Suzuki*[a], Mikio Matsuhiro[a], Yoshiki Kawata[a], Noboru Niki[a],
Issei Imoto[b], Yasutaka Nakano[c], Masahiko Kusumoto[d], Masahiro Kaneko[e]
[a] Graduate School of Technology, Industrial and Social Sciences, Tokushima University, Tokushima, Japan;
[b] Division of Molecular Genetics, Aichi Cancer Center Research Institute, Aichi, Japan;
[c] Division of Respiratory Medicine, Shiga University of Medical Science, Shiga, Japan;
[d] Department of Diagnostic Radiology, National Cancer Center Hospital, Tokyo, Japan;
[e] Tokyo Health Service Association, Tokyo, Japan

ABSTRACT

Chronic obstructive pulmonary disease (COPD) is predicted to become the third leading cause of death worldwide by 2030. Smoking is a well-known risk factor in the development of COPD. Association between COPD genes and smoking have been studied. This paper presents an association analysis of single nucleotide polymorphisms (SNPs) with a CT image-based phenotype of emphysema progression in heavy smokers. The emphysema progression was quantitatively represented by the annual increment of low attenuation volume (LAV) on CT scans for five years. 10 candidate SNPs were selected from 316 SNPs in 125 papers of genetic studies of COPD and lung cancer. The genotypes were determined by real-time polymerase chain reaction (PCR) using deoxyribonucleic acid (DNA) extracted from saliva samples. The association analysis was performed by Fisher's exact test and logistic regression analysis. This method was applied to a dataset with 144 participants (71 smokers, 61 past smokers, and 12 non-smokers). The results showed that the genotypes of rs3923564 and rs13180 SNPs were candidate SNPs associated with the CT image based-emphysema progression.

Keywords: Emphysema, single nucleotide polymorphism, computed tomography, longitudinal follow-up study

1. INTRODUCTION

Chronic obstructive pulmonary disease (COPD) is a major public health problem that is predicted to become the third leading cause of death worldwide by 2030 [1]. Smoking is a well-known risk factor in the development of COPD [2]. Association between COPD genes and smoking have been studied [3]. Our previous studies presented a quantitative assessment method of emphysema progression from low-dose CT images [4][5]. This paper presents an association analysis of single nucleotide polymorphisms (SNPs) with a CT image-based phenotype of emphysema progression in heavy smokers. The procedure of this association analysis is as follows; (1) thoracic organs and low attenuation volume (LAV) segmentation from low-dose CT images, (2) data selection using CT image-based features, (3) evaluation of CT image-based phenotype of emphysema progression, and (4) association analysis of SNPs with emphysema progression.

2. MATERIALS AND METHODS

2.1 CT scan, saliva sample collection, and DNA extraction

This study used a low-dose CT dataset which was collected in Tokyo health service association. The collection and analysis of data were approved by the Institutional Review Board. The scans which were acquired from 2002 to 2009. The participants were 71 smokers, 61 past smokers, and 12 non-smokers. CT dataset was acquired on Aquilion scanner with 30 mA at 120 kVp, plane resolution: 0.625 or 0.781 mm, reconstruction matrix: 512 x 512, slice thickness: 2.0 mm, convolution kernel: FC01, and reconstruction interval: 1.0 mm. Saliva samples were collected by using Oragene saliva collection kit (OG-500, DNA Genotek Inc., Canada) from the participants of the dataset. DNA was extracted from saliva samples using the prepIT.L2P (PT-L2P-5, DNA Genotek Inc.) and precipitated with ethanol according to manufacturer's instructions. All of the participants were Japanese.

Medical Imaging 2020: Computer-Aided Diagnosis, edited by Horst K. Hahn,
Maciej A. Mazurowski, Proc. of SPIE Vol. 11314, 113142D · © 2020 SPIE
CCC code: 1605-7422/20/$21 · doi: 10.1117/12.2549431

2.2 Thoracic organs and low attenuation volume (LAV) segmentation from low-dose CT images

The thoracic organs (body tissue, bone, trachea and bronchi, lungs, pulmonary blood vessels, mediastinum, thoracic wall) were automatically segmented on the basis of the intensity and anatomical features [6]-[8]. The loss of lung tissue associated with emphysema is represented by low attenuation volume (LAV) in CT images. In this study, the LAV percentage (LAV%) was defined as percentage of voxels which has CT value below -950 HU [9][10]. The segmentation results of LAV for two smokers who have over 40.0 pack-years were shown in Figure 1. The LAV% of a smoker had increased to 8.5 % from 6.0 % for 2.5 years as shown in (a) and (b) in Figure 1. On the other smoker, the LAV% had increased to 0.5 % from 0.3 % for 5.5 years as shown in (c) and (d) in Figure 1.

2.3 Data selection using CT image-based features

While the CT images are generally obtained during deep inspiration breath-hold, the LAV% extracted by the thresholding is fluctuated according to the inspiration level. To reduce the fluctuation, data selection was performed on the basis of the following three variables: lung volume, peak CT value of lungs, and LAV%.

1) Data selection using total lung volume: Average of total lung volume $\overline{TLV}$ for a participant was calculated using lung volumes segmented from each scan. To exclude the scans with lower or higher inspiration level, threshold values for lung volume were set to $\overline{TLV} \pm 3\%$.

2) Data selection using peak CT value of lungs: While the scans were selected using lung volume, there were difference in CT value histograms of lungs. Through excluding these scans, the standard deviation of peak CT value of lungs was adjusted to less than 5 HU.

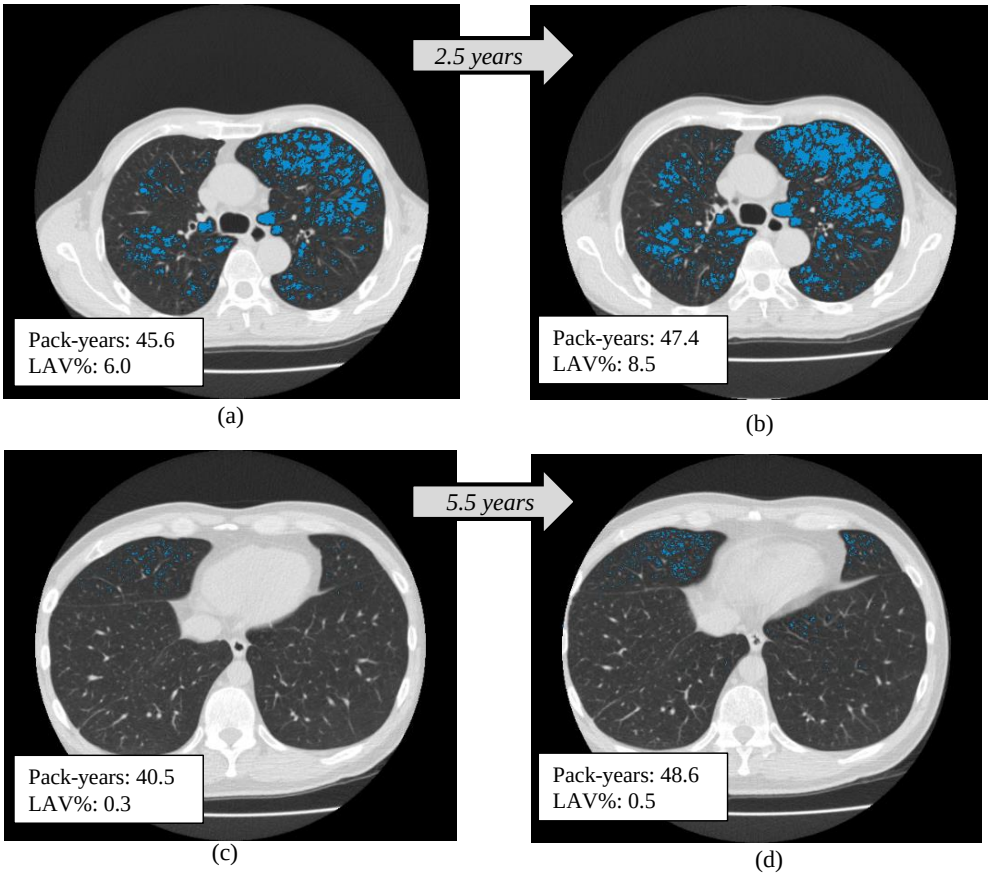

Figure 1. Segmentation results of LAV for two smokers who have over 40 pack-years. The LAV% of a smoker (a)(b) increased rapidly for 2.5 years. On the other smoker (c)(d), the LAV% have increased slightly for 5.5 years.

3) Data selection using LAV%: We assumed that the LAV% of an emphysema patient increase linearly with progression of the disease. The association of LAV% with follow-up time was represented by linear regression model. The outlier scans against to the linear regression model were excluded, and the residual variance was adjusted to less than 0.4.

If the number of scans was less than three time by above selections, these participants were excluded. Through this procedure, 99 participants were selected. The participant characteristics were shown in Table 1. The association analysis of SNPs with emphysema progression was performed using scans of the current and past smokers.

Table 1. Participant characteristics

	Nonsmokers	Past smokers	Current smokers
The number of participants	6	43	50
Age[year]	68.5±12.8	66.9±8.3	64.5±7.8
Smoking[year]	-	37.4±9.5	44.6±8.3
Pack-year	-	60.7±19.5	66.6±31.0
Smoking cessation[year]	-	9.9±8.2	-
Follow-up[year]	4.4±1.7	4.3±1.3	3.5±1.6
The number of scan for a participant	5.7±1.0	5.2±1.6	5.2±1.8

2.4 Evaluation of CT image-based phenotype of emphysema progression

The emphysema progression was quantified by linear regression of LAV% and follow-up time [5]. The slope means the annual increment of LAV%. We employed the annual increment of LAV% as CT image-based emphysema progression. The current and past smokers were classified into high- and low-risk groups of emphysema progression using the annual increment of LAV%. The threshold value was experimentally set to 0.3 %/year.

2.5 Association analysis of SNPs with emphysema progression

10 candidate SNPs were selected from 316 SNPs in 125 papers of genetic studies for COPD and lung cancer; rs7733088 in HTR4 [11], rs7671167 in FAM13A [12], rs13118928 in HHIP [13], rs13180 in IREB2 [14], rs3923564 in SFTPD [15], rs7937 in EGLN2 [16], rs2736100 in TERT [17], rs401681 in CLPTM1L [18], rs1333040 in CDKN2B-AS1 [19], and rs10849605 in RAD52 [20]. The selection procedure is as follows. In the first step, 103 papers (256 SNPs) with impact factor over 3.0 were selected. In the second step, 109 SNPs with minor allele frequency for Japanese over 0.3 were selected. In the third step, 87 SNPs were selected on the basis of the linkage disequilibrium. The minor allele frequency and the linkage disequilibrium were confirmed by referring Ensembl database [21]. Finally, the selection was performed in accordance with the comments of a researcher of human genetics. SNP genotyping was performed using TaqMan SNP Genotyping Assays [22]. The genotypes were determined by real-time polymerase chain reaction (PCR) using DNA extracted from saliva samples. The associations of the genotypes of 10 SNPs with the CT image-based emphysema progression were statistically assessed by Fisher's exact test and logistic regression analysis.

3. RESULTS

The measurement results of the annual increment of LAV% are shown in Figure 2. Fisher's exact test results are shown in Table 3. The lowest *p*-value was observed in the frequency of genotype at rs3923564 (*p*-value=0.030). The results of the logistic regression analysis were shown in Table 4. The lowest *p*-values were observed in genetic models at rs13180 (*p*-values=0.028) and rs3923564 (*p*-values=0.016). High odds ratio over 3.0 was observed at rs13180. The results showed that the emphysema progressions of the smokers with AG or GG genotypes at rs3923564 were slow, and the emphysema progressions of the smokers with TT genotype at rs13180 were rapid.

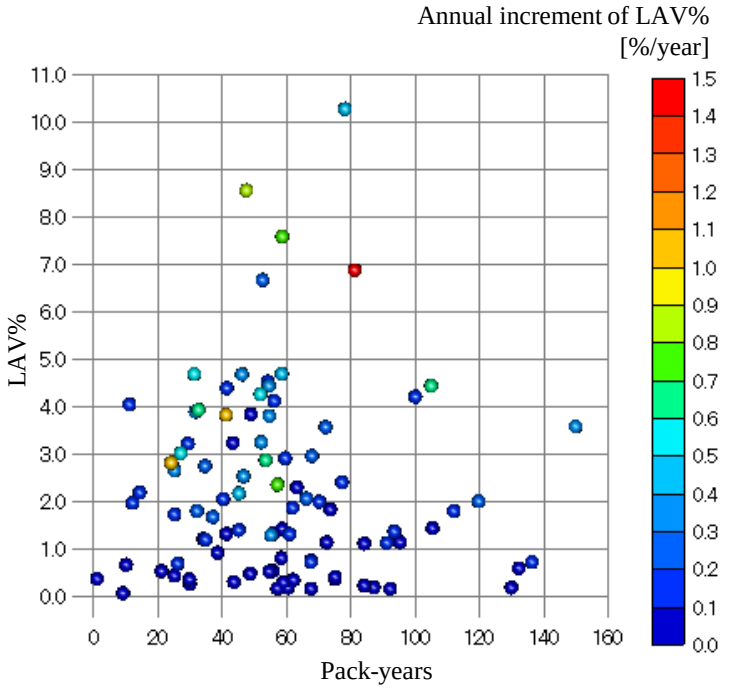

Figure 2. Relations between LAV% at baseline, pack-years, and annual increment of LAV%. The colors represent the annual increment of LAV%.

4. CONCLUSIONS

This paper presented an association analysis of SNPs with a CT image-based phenotype of emphysema progression in heavy smokers. The results showed two candidate SNPs associated with CT image-based emphysema progression in heavy smokers; rs3923564 in SFTPD and rs13180 in IREB2.

REFERENCES

[1] World Health Organization (WHO), World Health Statistics (2008).

[2] Global Initiative for Chronic Obstructive Lung Disease (GOLD), "Global strategy for the diagnosis, management and prevention of chronic obstructive pulmonary disease: 2018 report (2018).

[3] Wain, L. V., Shrine, N., Miller, S., et al., "Novel insights into the genetics of smoking behavior, lung function, and chronic obstructive pulmonary disease (UK BiLEVE): a genetic association study in UK Biobank", Lancet Respir. Med. 3(10), 769-781 (2015).

[4] Suzuki, H., Matsuhiro, M., Kawata, Y., Niki, N., Nakano, Y., Ohmatsu, H., Kusumoto, M., Tsuchida, T., Eguchi, K., Kaneko, M., and Moriyama, N., "Longitudinal follow-up study of smoking-induced emphysema progression in low-dose CT screening of lung cancer", Proc. SPIE 9035, 90352M (2014).

[5] Suzuki, H., Mizuguchi, R., Matsuhiro, M., Kawata, Y., Niki, N., Nakano, Y., Ohmatsu, H., Kusumoto, M., Tsuchida, T., Eguchi, K., Kaneko, M., and Moriyama, N., "Quantitative assessment of smoking-induced emphysema progression in longitudinal CT screening for lung cancer", Proc. SPIE 9414, 94142O (2015).

[6] Kanazawa, K., Kawata, Y., Niki, N., Satoh, H., Ohmatsu, H., Kakinuma, R., Kaneko, M., Moriyama, N., and Eguchi, K., "Computer-aided diagnosis for pulmonary nodules based on helical CT images," Comput. Med. Imaging Graph. 22(2), 157-167 (1998)

Table 3. Fisher's exact test results of 10 candidate SNPs

SNP	Group	Genotype			*p-value*
rs7733088		AA	AG	GG	
	Control	15	37	13	0.473
	Case	6	10	7	
rs7671167		CC	CT	TT	
	Control	10	33	19	0.793
	Case	2	13	8	
rs13118928		AA	AG	GG	
	Control	28	26	7	0.706
	Case	11	8	4	
rs13180		CC	CT	TT	
	Control	20	35	9	0.081
	Case	6	8	8	
rs3923564		AA	AG	GG	
	Control	23	29	13	0.030
	Case	15	14	4	
rs7937		CC	CT	TT	
	Control	8	37	22	1.000
	Case	3	12	8	
rs2736100		AA	AC	CC	
	Control	22	22	11	0.886
	Case	7	7	5	
rs401681		CC	CT	TT	
	Control	27	27	11	0.851
	Case	11	9	3	
rs1333040		CC	CT	TT	
	Control	11	25	32	0.721
	Case	3	11	9	
rs10849605		CC	CT	TT	
	Control	6	26	32	0.875
	Case	2	11	10	

Table 4. Logistic regression analysis results of 10 candidate SNPs

SNP	enetic models	*p-value*	*OR*
rs7733088			
	Dominant (AA+AG)vs.GG	0.771	0.850
	Recessive AAvs.(AG+GG)	0.308	1.750
rs7671167			
	Dominant (CC+CT)vs.TT	0.716	0.828
	Recessive CC vs.(CT+TT)	0.389	0.495
rs13118928			
	Dominant (AG+GG)vs.AA	0.875	0.926
	Recessive GG vs.(AG+AA)	0.477	1.620
rs13180			
	Dominant (CT+TT)vs.CC	0.726	1.210
	Recessive TTvs.(CT+CC)	0.028	3.490
rs3923564			
	Dominant (AG+GG)vs.AA	0.016	0.292
	Recessive GGvs.(AG+AA)	0.786	0.842
rs7937			
	Dominant (CC+CT)vs.TT	0.864	0.917
	Recessive CCvs.(CT+TT)	0.889	1.110
rs2736100			
	Dominant (AC+CC)vs.AA	0.808	1.140
	Recessive CCvs.(AC+AA)	0.565	1.430
rs401681			
	Dominant (CT+TT)vs.CC	0.601	0.775
	Recessive TTvs.(CT+CC)	0.663	0.736
rs1333040			
	Dominant (CC+CT)vs.TT	0.510	1.380
	Recessive CC vs.(CT+TT)	0.719	0.777
rs10849605			
	Dominant (CC+CT)vs.TT	0.592	1.300
	Recessive CC vs.(CT+TT)	0.923	0.921

[7] Yoneda, K., Matsuhiro, M., Suzuki, H., Kawata, Y., Niki, N., Nakano, Y., Ohmatsu, H., Kusumoto, M., Tsuchida, T., Eguchi, K., and Kaneko, M. "Computer-aided diagnosis for osteoporosis using chest 3D CT images," Proc. of SPIE 9785, 97853A (2016)

[8] Matsuhiro, M., Suzuki, H., Kawata, Y., Niki, N., Nakano, Y., Ohmatsu, H., Kusumoto, M., Tsuchida, T., Eguchi, K., and Kaneko, M., "Peripleural lung disease detection based on multi-slice CT images," Proc. of SPIE, 9414, 94142W (2015)

[9] Gevenosis, P. A., de Maertelaer, V., de Vuyst, P., Zanen, J., and Yernault, J., "Comparison of Computed Density and Macroscopic Morphometry in Pulmonary Emphysema," Am. J. Respir. Crit. Care. Med. 152, 653-657 (1995).

[10] Parr, D. G., Stoel, B. C., Stolk, J., Nightingale, P. G., and Stockley, R. A., "Influence of Calibration on Densitometric Studies of Emphysema Progression Using Computed Tomography," Am. J. Respir. Crit. Care Med. 170(8), 883-890 (2004).
[11] Wilk, J. B., Shrine, N. R. G., Loehr, L. R., et al., "Genome-Wide Association Studies Identify CHRNA5/3 and HTR4 in the Development of Airflow Obstruction", Am. J. Respir. Crit. Care Med. 186(7), 622-632 (2012).
[12] Cho, M. H., Boutaoui, N., Klanderman, B. J., et al., "Variants in FAM13A are associated with chronic obstructive pulmonary disease", Nat. Genet. 42(3), 200-202 (2010).
[13] Kim, W. J., Oh, Y. M., Lee, J. H., et al., "Genetic variants in HHIP are associated with FEV1 in subjects with chronic obstructive pulmonary disease", Respirology 18(8), 1202-1209 (2013).
[14] Zhou, H., Yang, J., Li, D., et al., "Association of IREB2 and CHRNA3/5 polymorphisms with COPD and COPD-related phenotypes in a Chinese Han population", Journal of Human Genetics 57, 738-746 (2012).
[15] Chen, W., Brehm, J. M., Manichaikul, A., et al., "A Genome-Wide Association Study of Chronic Obstructive Pulmonary Disease in Hispanics", Annals of American Thoracic Society, 12(3), 340-348 (2015).
[16] Castaldi, P. J., Cho, M. H., Zhou, X., et al, "Genetic control of gene expression at novel and established chronic obstructive pulmonary disease loci", Human Molecular Genetics, 24(4), 1200-1210 (2014).
[17] Wei, R., Cao, L., Pu, H., et al., "TERT Polymorphism rs2736100-C Is Associated with EGFR Mutation–Positive Non–Small Cell Lung Cancer", Clinical Cancer Research, 21(22), 5173-5180 (2015).
[18] Wang, Y., Broderick, P., Webb, E., et al., "Common 5p15.33 and 6p21.33 variants influence lung cancer risk", Nat. Genet. 40(12), 1407-1409 (2008).
[19] Liu, K. Y., Muehlschlegel, J. D., Perry, T. E., et al., "Common genetic variants on chromosome 9p21 predict perioperative myocardial injury after coronary artery bypass graft surgery", The Journal of Thoracic and Cardiovascular Surgery, 139(2), 483-488.e2 (2010).
[20] Timofeeva, M. N., Hung, R. J., Rafnar, T., et al., "Influence of common genetic variation on lung cancer risk: meta-analysis of 14 900 cases and 29 485 controls", Human Molecular Genetics 21(22), 4980-4995 (2012).
[21] Zebino, D. R., Achuthan, P., Akanni, W., et al., "Ensembl 2018," Nucleic acids research 46(D1), 754-761 (2018).
[22] Francisco, M. D. L. V., Lazaruk, K. D., Rhodes, M. D., et al., "Assessment of two compatible SNP genotyping platforms: Taqman SNP Genotyping Assays and the SNPlex Genotyping System," Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis 573, 1-2 (2005).