A Genome Wide-Association Study for Lung Function in the Korean Population Using an Exome Array

강원대학교 병원 내과

*이규선, 김우진

Background/Aims: Lung function is an objective indicator of diagnosis and prognosis of respiratory diseases. Many common genetic variants have been associated with lung function in multiple ethnic populations, many of which are known to play a role in susceptibility to respiratory diseases including COPD. We looked for coding variants associated with FEV1 and FEV1/FVC in the Korean general population. **Methods:** We carried out exome-array analysis and lung function measurements of the FEV1 and FEV1/FVC in 7,524 individuals of the Korean population. We evaluated single variants with minor allele frequency greater than 0.5%. We performed look-ups for candidate coding variants associations in the UK biobank, Spirometer, and CHARGE consortia. **Results:** We identified coding variants in the SMIm^{2}9 (C6orf1) (p=1.2 × 10-5) and HMGA1 locus on chromosome 6p21 and the ARHGEF40 (p=9.9 × 10-5) locus on chromosome 14q11 as having a significant association with lung function (FEV1). We also confirmed a previously reported association with lung function and COPD in the FAM13A (p=4.54 × 10-6) locus on chromosome 4q22, in TNXB (p=1.30 × 10-6) and in AGER (p =1.09 × 10-8) locus on chromosome 6p21. **Conclusions:** Our exome array analysis identified that several protein coding variants were associated with lung function in the Korean population. Common coding variants in SMIm^{2}9 (C6orf1), HMGA1, FAM13A, TNXB, AGER and low frequency variant in ARHGEF40 potentially affect lung function, which warrant further study.

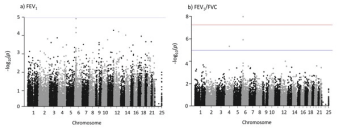

Fig.1 Manhattan plots of association results for FEV_1 and FEV_1/FVC. The Manhattan plots for FEV_1 (a) and FEV_1/FVC (b) are ordered by chromosome position. SNPs for which $-\log_{10}$ P>5 are indicated in blue line.

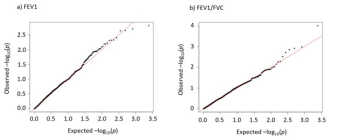

Fig.2 Quantile-Quantile (QQ) plots show $-\log_{10}(p)$ of observed genome-wide association results against expected association results for FEV_1 and FEV_1/FVC. Genomic control inflation factors (λGC) before genomic control was 0.90 for FEV_1 and 0.84 for FEV_1/FVC.

Table 1. Top nine coding-variants associated with FEV_1, FEV_1/FVC and results of Look-up analyses.

	KoGES							UK biobank		CHARGE consortium		Spirometer consortium	
Trait	SNP ID	Chr:Pos	Gene	MAF	Odd Ratio or beta (QT)	p-value		Odd Ratio or beta (QT)	p-value	Odd Ratio or beta (QT)	p-value	Odd Ratio or beta (QT)	p-value
FEV_1	rs7742369	6:34,165,721	–	0.171	3.75E-02	$7.11X10^{-5}$		0.028	4.00E-31	–	–	–	–
	rs2780226	6:34,199,092	–	0.165	3.94E-02	$3.86X10^{-5}$		-0.035	1.15E-26	–	–	–	–
	rs1150781	6:34,214,322	C6orf1 (SMIM29)	0.181	4.04E-02	$1.20X10^{-5}$		-0.033	5.65E-24	–	–	–	–
	11:130,247,700	11:130,247,700	–	0.464	-2.82E-02	$5.36X10^{-5}$		–	–	–	–	–	–
	12:110,390,979	12:110,390,979	GIT2/TCHP	0.078	-5.31E-02	$6.55X10^{-5}$		–	–	–	–	–	–
	rs114591848	14:21,550,212	ARHGEF40	0.014	1.17E-01	$9.95X10^{-5}$		–	–	–	–	–	–
	*rs7143633	14:21,547,088	ARHGEF40	–	–	–		2.33	2.36E-03	–	–	–	–
FEV_1/FVC	rs7671167	4:89,883,979	FAM13A	0.493	4.74E-01	$4.54X10^{-6}$		–	–	7.61E-02	6.30E-07	–	5.87E-06
	rs2239688	6:32,054,212	TNXB	0.175	7.03E-01	$1.30X10^{-6}$		–	–	–	–	–	–
	krs433061	6:32,047,051	TNXB	–	–	–		–	–	1.27E-01	2.00E-04	–	–
	krs1150754	6:32,050,758	TNXB	–	–	–		–	–	1.11E-01	3.00E-03	–	–
	rs2070600	6:32,151,443	AGER	0.158	8.75E-01	$1.09X10^{-5}$		–	–	1.86E-01	1.50E-08	–	2.56E-02

Table 2. Protein and mRNA expression profiles of implicated genes from single association analyses.

	Protein Expression				RNA profile			
	Human Protein Atlas				Human Protein Atlas		GTEx	
Gene	Nasopharynx	Bronchus	Lung macrophages	Lung pneumocytes	TPM	Category	RPKM	Category
SMIM29(C6orf1)	high	not detected	medium	medium	21.2	medium	14.8	medium
HMGA1	medium	low	medium	not detected	25.9	medium	27.1	medium
GIT2	high	low	high	low	25	medium	8.5	low
TCHP	medium	medium	low	low	11	medium	5.8	low
ARHGEF40	N/A	N/A	N/A	N/A	10.2	medium	15.2	medium
FAM13A	N/A	medium	medium	medium	17.8	medium	3.4	low
TNXB	N/A	N/A	N/A	N/A	16.4	medium	20.3	medium
AGER	not detected	not detected	high	medium	880.2	medium	424.4	high

Table 3. SIFT/PolyPhen predictions for rs1150781 and rs114591848

SNP ID	Chr:Pos	Allele	Consequence	Gene	Codons / Amino acid	SIFT consequence(score)	PolyPhen consequence(score)
rs1150781	6:34,214,322	G	missense variant	*C6orf1 (SMIM29)*	GGG(Gly150)-->GCG(Ala)	deleterious(0)	benign(0.015)
rs114591848	14:21,550,212	G	missense variant	*ARHGEF40*	CGG(Arg1062)-->CAG(Gln)	tolerated(0.08)	possibly damaging (0.679)