

ARTICLE

Received 25 Dec 2014 | Accepted 6 Jun 2015 | Published 15 Jul 2015

DOI: 10.1038/ncomms8754

OPEN

Three missense variants of metabolic syndrome-related genes are associated with alpha-1 antitrypsin levels

Kazuya Setoh¹, Chikashi Terao¹, Shigeo Muro², Takahisa Kawaguchi¹, Yasuharu Tabara¹, Meiko Takahashi¹, Takeo Nakayama³, Shinji Kosugi⁴, Akihiro Sekine⁵, Ryo Yamada⁶, Michiaki Mishima² & Fumihiko Matsuda^{1,6}

Alpha-1 antitrypsin (AAT) encoded by *SERPINA1* is an acute-phase inflammation marker, and AAT deficiency (AATD) is known as one of the common genetic disorders in European populations. However, no genetic determinants to AAT levels apart from the *SERPINA* gene clusters have been identified to date. Here we perform a genome-wide association study of serum AAT levels followed by a two-staged replication study recruiting a total of 9,359 Japanese community-dwelling population. Three missense variants of metabolic syndrome-related genes, namely, rs671 in *ALDH2*, rs1169288 in *HNF1A* and rs1260326 in *GCKR*, significantly associate with AAT levels ($P \leq 1.5 \times 10^{-12}$). Previous reports have shown the functional relevance of *ALDH2* and *HNF1A* to AAT. We observe a significant interaction of rs671 and alcohol consumption on AAT levels. We confirm the association between AAT and rs2896268 in *SERPINA1*, which is independent of known causative variants of AATD. These findings would support various AAT functions including metabolic processes.

¹Center for Genomic Medicine, Kyoto University Graduate School of Medicine, Shogoinkawaramachi 53, Sakyo-ku, Kyoto 606-8507, Japan. ²Department of Respiratory Medicine, Kyoto University Graduate School of Medicine, Shogoinkawaramachi 54, Sakyo-ku, Kyoto 606-8507, Japan. ³Department of Health Informatics, Kyoto University School of Public Health, Yoshidakonoemachi, Sakyo-ku, Kyoto 606-8501, Japan. ⁴Department of Medical Ethics and Medical Genetics, Kyoto University School of Public Health, Yoshidakonoemachi, Sakyo-ku Kyoto, 606-8501, Japan. ⁵EBM Research Center, Kyoto University Graduate School of Medicine, Shogoinkawaramachi 54, Sakyo-ku, Kyoto 606-8507, Japan. ⁶Statistical Genetics, Center for Genomic Medicine, Kyoto University Graduate School of Medicine, Shogoinkawaramachi 53, Sakyo-ku, Kyoto 606-8507, Japan. Correspondence and requests for materials should be addressed to C.T. (email: a0001101@kuhp.kyoto-u.ac.jp).

Alpha-1 antitrypsin (AAT) is an acute-phase inflammation marker synthesized predominantly in hepatocytes and serves as a protease inhibitor of proteolytic enzymes including elastase, trypsin, thrombin and bacterial proteases^{1,2}. Recent studies have revealed the wide range of functions for AAT^{3–5}. AAT is encoded by the *SERPINA1* gene on chromosome 14 (ref. 6). The reduction of AAT levels leads to inflammation in the lung, and is associated with the risk for early-onset chronic obstructive pulmonary disease (COPD)^{7–9}. AAT deficiency (AATD) is a common genetic disorder among Caucasian populations caused by mutations in *SERPINA1*. Patients with AATD present with emphysema due to imbalance of elastase and anti-elastase effect of AAT, and, in some cases, accumulating AAT affects hepatic functions and leads to liver cirrhosis and hepatocellular carcinoma^{10,11}. Many variants in *SERPINA1* were reported to cause AATD¹². The S variant (PI S, rs17580) and the Z variant (PI Z, rs28929474) are the major missense single-nucleotide polymorphisms (SNPs) leading to AATD by moderate and severe decline of AAT levels, respectively, through the conformational change of the AAT protein and the subsequent faulty secretion of AAT from hepatocytes¹². All other variants reported to cause AATD are rare variants in *SERPINA1* (ref. 12).

Genome-wide association studies (GWAS) have enabled us to detect susceptibility loci associated with quantitative traits as well as discrete phenotypes using an unbiased approach. To date, only one GWAS has been reported in an European population to detect susceptibility loci to serum AAT levels, and the authors identified markers in the *SERPINA* gene cluster on 14q32.13 (ref. 13). Thus, there are no other markers reported to be associated with serum AAT levels. Novel susceptibility loci would uncover mechanisms of AAT metabolism and molecular networks for regulation of inflammatory processes in the lung.

As genetic backgrounds differ between different populations, GWAS using populations other than European populations would lead to identify novel susceptibility loci to serum AAT levels. Here we perform GWAS and a two-staged replication study of serum AAT levels using 9,359 Japanese healthy individuals participating in the Nagahama Study and show a total of three missense variants, rs671, rs1169288 and rs1260326 in metabolic syndrome-related genes, *ALDH2*, *HNFI1A* and *GCKR*, respectively. We also show a significant interaction of rs671 and alcohol consumption on AAT levels. In addition, we show rs2896268 in *SERPINA1* as an AAT-associated variant independent of known causative variants of AATD.

Results

GWAS of serum AAT levels. We conducted genome scanning using 3,693 participants of the Nagahama Study, a prospective Japanese cohort of community-dwelling population¹⁴, with the use of Illumina Infinium arrays to identify genetic loci affecting serum AAT levels in the Japanese population. The serum AAT levels did not show a deviated distribution (Supplementary Fig. 1). We used generalized linear regression model (GLM) with covariates including age, sex, body-mass index (BMI), C-reactive protein (CRP), the recruiting years and smoking status (for details, see Methods). We performed genotype imputation by using the East Asian panel in the 1000 Genome Project¹⁵ as a reference. After quality control, 3,294 subjects and 6,569,727 SNPs were used for the association study (Table 1, for details, see Methods and Supplementary Figs 2 and 3). The quantile-quantile plots of the results of the genome scanning revealed no population structure in the current study (genomic inflation factor $\lambda = 1.01$, Fig. 1a). Thus, we did not apply genomic control¹⁶ into the association study. As a result, rs671 in *ALDH2* on chromosome 12 showed a significant association beyond the

GWAS significant level (GLM $P = 3.4 \times 10^{-11}$, Fig. 1a and Table 2). Three additional loci, namely, the *HNFI1A* and *KSR2* regions on chromosome 12 and the *GCKR* region on chromosome 2 showed suggestive associations with P values less than 1.0×10^{-6} (GLM $P = 8.5 \times 10^{-8}$, 9.8×10^{-7} and 7.1×10^{-7} , respectively, Fig. 1a and Table 2). All SNPs showing P values less than 1.0×10^{-6} are shown in Supplementary Table 1. rs1169288 on *HNFI1A* and rs11068574 on *KSR2* were 9.2 and 5.8 Mbp apart from rs671, respectively, and not in linkage disequilibrium (LD) with rs671 ($r^2 \leq 0.002$, Supplementary Fig. 4). We confirmed that these two associations were independent from rs671 by conditioning on rs671 (GLM $P < 2.0 \times 10^{-6}$, Fig. 1b). When we conditioned the associations in the four regions by each SNP (rs671, rs1169288, rs11068574 and rs1260326), no further association signals remained (GLM $P \geq 0.0072$, Supplementary Fig. 5). Thus, we selected these four markers to be genotyped in the replication studies as candidate markers affecting serum AAT levels.

Two replication studies and combined study. We conducted a replication study to genotype these four SNPs using 4,023 subjects in the Nagahama Study by Taqman assay (Table 1 and Supplementary Fig. 2). As a result, the associations between serum AAT levels and rs671, rs1169288 or rs1260326 were replicated with P values less than 2.7×10^{-4} (GLM $P = 2.2 \times 10^{-10}$, 2.7×10^{-4} and 2.3×10^{-7} , respectively, Table 2). The association between rs11068574 and serum AAT levels was not replicated (GLM $P = 0.22$, Table 2).

We further performed a second replication for these SNPs using 2,042 subjects by Taqman assay (Table 1 and Supplementary Fig. 2). As a result, the associations of rs671, rs1169288 or rs1260326 were again replicated and the overall associations in the combined study were beyond the GWAS significant level (GLM $P = 2.2 \times 10^{-23}$, 1.5×10^{-12} and 3.1×10^{-16} , respectively, Table 2 and Supplementary Fig. 6A–C). The association of rs11068574 was again not replicated (GLM $P = 0.093$, Table 2). All of the SNPs showed success rates of more than 0.96 and did not show departure from Hardy–Weinberg Equilibrium in each of the three studies (χ^2 -test $P \geq 0.0019$) and the combined study (χ^2 -test $P \geq 0.045$).

Explained effects by the combination of the significant SNPs.

All of the three markers showing significant associations are located in exonic regions of the genes and induce amino-acid alterations of the proteins (Fig. 2 and Table 2). We did not find heterogeneity between GWAS results and replication results for these three markers (Cochran's Q test $P \geq 0.14$). When we evaluated interactions between any pairs of these three SNPs, we did not find strong interactions on serum AAT levels (GLM $P \geq 0.038$). When we analysed the association between the number of susceptibility alleles of these three SNPs and serum AAT levels using GLM, the association showed a P value of 1.5×10^{-48} (Supplementary Fig. 6D). These three SNPs were evaluated to explain 1.95% of total variance of the serum AAT levels (Supplementary Table 2).

Functional annotation search for causative SNPs. We picked up all of the variants in strong LD ($r^2 > 0.8$) with the three SNPs and performed functional annotation search of these SNPs by HaploReg¹⁷ (Supplementary Table 3). None of the other missense mutations in the three genes were in LD with the three SNPs. The block of the five SNPs in strong LD with rs671 included *ACAD10* and *BRAP* as well as *ALDH2* (Supplementary Fig. 5 and Supplementary Table 3). *ALDH2* was the only gene among the three genes in the region whose functional relevance to AAT was

Table 1 The characteristics of study subjects.			
	GWAS	Replication 1	Replication 2
Number	3,294	4,023	2,042
Sex (female ratio) (%)	66.5	67.0	68.8
Age* (year)	52.3 ± 14.1	55.3 ± 12.6	53.1 ± 12.9
AAT* (mg dl ⁻¹)	133.0 ± 19.4	128.5 ± 17.6	135.4 ± 19.0
BMI* (kg m ⁻²)	22.3 ± 3.2	22.3 ± 3.3	22.3 ± 3.3
Brinkman index*	164.6 ± 319.8	167.2 ± 325.5	149.4 ± 307.6
Alcohol intake† (heavy/moderate/mild)	713/377/2,204	959/379/2,685	466/199/1,377
High-sensitive CRP* (ng ml ⁻¹)	922.1 ± 3261	937.6 ± 4108	786.5 ± 2548

AAT, Alpha-1 antitrypsin; BMI, body mass index; CRP, C-reactive protein; GWAS, genome-wide association studies.

*The mean ± s.d.

†Heavy: >3 days week⁻¹, moderate: 2-3 days week⁻¹, mild <2 days week⁻¹.

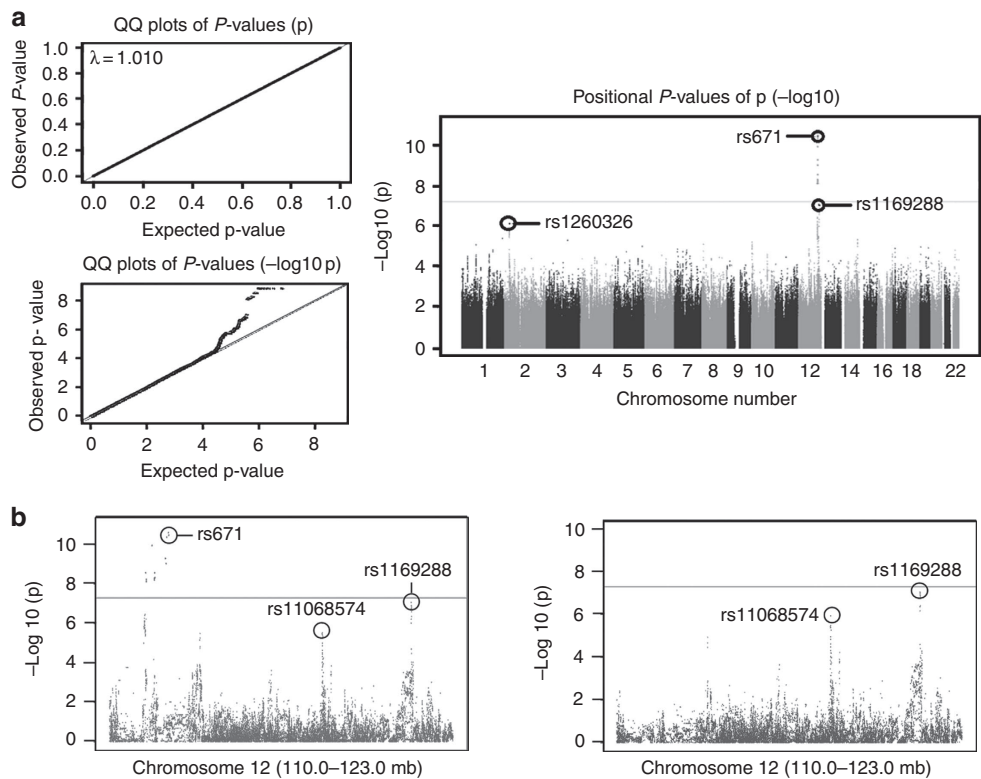


Figure 1 | The results of GWAS on the serum AAT levels ($n = 3,294$). (a) The quantile-quantile plots (left) and the Manhattan plot for the genome-wide scanning: The horizontal line indicates the genome-wide significant level, $P = 5.0 \times 10^{-8}$. (b) The regional plot for a part of chromosome 12 for nominal (left) and conditional (right) associations adjusted for rs671 genotypes. P values were calculated by generalized linear regression model.

reported^{18,19} and which contained SNPs with functional annotations in hepatic cells (Supplementary Table 3). *HNF1A* was shown to be an important expression regulator of AAT²⁰, and the two SNPs in strong LD with rs1169288 also resided in the *HNF1A* locus. The block of the 14 SNPs in strong LD with rs1260326 contained only the *GCKR* locus and was more than 46.5 kbp away from the adjacent gene, *C2orf16* (Supplementary Table 3 and Fig. 2). Thus, it is reasonable to regard *ALDH2*, *HNF1A* and *GCKR* as the responsible genes for the ssociations with levels of AAT. Notably, all of the three genes are expressed in the liver and are reported to be associated with metabolic phenotypes including impairment of insulin secretion²¹, lipid profiles^{22,23}, risk of type 2 diabetes mellitus²⁴ and cardiovascular diseases²⁵ and metabolic syndrome pathway²⁶.

Analysis of HNF1A and GCKR conservations. While the functional importance of amino-acid alteration of *ALDH2* by

rs671 is well established (described later), functional impacts of amino-acid alteration by rs1169288 and rs1260326 on *HNF1A* and *GCKR* are not fully known except for some reports²⁷. Polyphen-2 software²⁸ did not suggest deleterious change brought about by the two SNPs. Thus, we analysed conservation of these amino-acid residues across vertebrates. As a result, isoleucine altered by rs1169288 was highly conserved in vertebrates (Supplementary Table 4), supporting the impact of rs1169288. Proline induced by rs1260326 rather than leucine was conserved among vertebrates (Supplementary Table 4), suggesting the importance of the C allele of rs1260326.

Exclusion of possibility of confounded by CRP. It might be noted that *HNF1A* and *GCKR* are known susceptibility loci to CRP levels²⁶, which are correlated with the levels of AAT. We confirmed no association between the residuals of AAT and serum CRP levels, indicating that the associations in these

Table 2 | Association studies on AAT serum level in the Nagahama Study.

SNP	Chr	Position	Gene	Ref/ var	AA alteration	GWAS (N = 3,294)			Replication 1 (N = 4,023)			Replication 2 (N = 2,042)			Combined (N = 9,359)		
						VAF	Effect (s.e.)	P	VAF	Effect (s.e.)	P	VAF	Effect (s.e.)	P	VAF	Effect (s.e.)	P
rs1260326	2	27730940	GCKR	T/C	L256P	0.425	2.15 (0.43)	7.1×10^{-7}	0.438	1.84 (0.36)	2.3×10^{-7}	0.430	2.27 (0.58)	9.6×10^{-5}	0.432	2.05 (0.25)	3.1×10^{-16}
rs671	12	112241766	ALDH2	G/A	E364K	0.270	3.28 (0.49)	3.4×10^{-11}	0.269	2.55 (0.40)	2.2×10^{-10}	0.267	2.65 (0.65)	4.4×10^{-5}	0.269	2.83 (0.28)	2.2×10^{-23}
rs11068574	12	118053982	KSR2	A/G	—	0.651	2.28 (0.46)	9.8×10^{-7}	0.650	0.48 (0.38)	0.22	0.667	-1.01 (0.60)	0.093	0.654	-0.76 (0.27)	4.3×10^{-3}
rs1169288	12	121416650	HNF1A	A/C	I27L	0.490	2.40 (0.45)	8.5×10^{-8}	0.479	1.29 (0.35)	2.7×10^{-4}	0.503	1.83 (0.58)	1.5×10^{-3}	0.487	1.79 (0.25)	1.5×10^{-12}

AA alteration, amino-acid alteration; AAT, alpha1-antitrypsin, Chr, chromosome, GWAS, genome-wide association study, P, P-value calculated by generalized linear regression model; ref/var, reference allele/variant allele; SNP, single-nucleotide polymorphism; VAF, variant allele frequency.

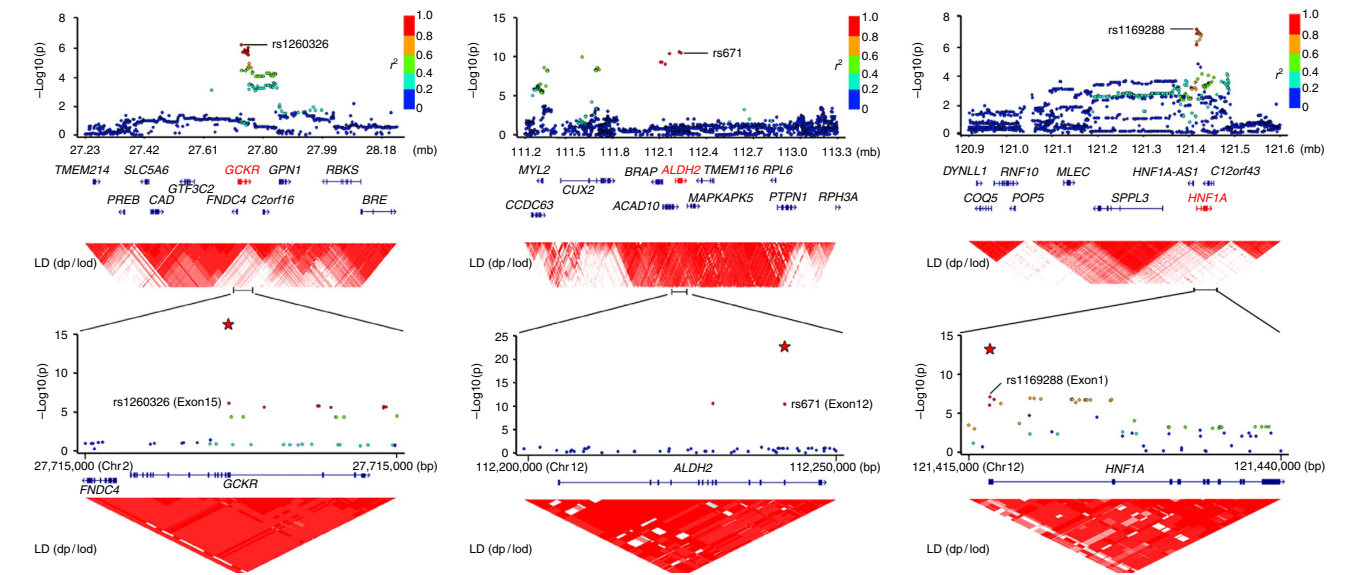


Figure 2 | Regional plots of the three regions significantly associated with the serum AAT levels. The chromosomal positions and P values for SNPs in the three significant regions are shown accordingly. Brightness of the red colour in LD blocks corresponds to the strength of LD. The red stars in the lower panels represent the associations in the combined study ($n = 9,359$).

two regions were not dependent on the associations between CRP levels and the two regions (GLM $P = 1.0$, Supplementary Fig. 7).

Expression correlations between *SERPINA1* and the three genes. We analysed whether these three gene expressions are associated with expressions of *SERPINA1* using gene expression database in the Japanese population²⁹ to infer the relationship between AAT and the three proteins encoded by these three genes. As a result, the expressions of *ALDH2*, *GCKR* and *HNF1A* were significantly associated with that of *SERPINA1* or the transcript variant of *SERPINA1* (Student's t distribution test $P = 5.6 \times 10^{-15}$, Spearman's rank sum coefficient (r_s) = 0.43, $P = 3.7 \times 10^{-6}$, $r_s = 0.26$ and $P = 1.7 \times 10^{-8}$, $r_s = 0.31$, respectively, Supplementary Fig. 8). These results may suggest a molecular network among these genes and support the association between AAT and *GCKR* as well as those between AAT and *ALDH2* or *HNF1A*.

Detailed analysis of *ALDH2*. Next, we concentrated on the associations observed in *ALDH2*, the most significant locus with previously reported functional analyses. *ALDH2* dehydrogenates acetaldehyde that is produced by various processes and inhibits the anti-elastase activity of AAT¹⁹. As ethanol is dehydrogenated into acetaldehyde, *ALDH2* is also involved with alcohol consumption. The A allele of rs671 is known to decrease *ALDH2* activity and lower alcohol tolerance^{18,23,30}. In fact, rs671 is very strongly associated with alcohol consumption in the current study (Student's t distribution test $P = 2.2 \times 10^{-248}$).

As alcohol consumption is known to affect AAT levels presumably via acetaldehyde³¹, the association of *ALDH2* might be explained by altering alcohol consumption due to rs671. In fact, conditioning on alcohol consumption greatly diminished the association of rs671 (after conditioned; GLM $P = 9.2 \times 10^{-8}$). We still found a substantial association after conditioning, suggesting further mechanisms aside from altering alcohol drinking by rs671. In contrast, conditioning on alcohol did not alter the associations of rs1260326 and rs1169288 (GLM $P = 3.0 \times 10^{-16}$ and 6.6×10^{-12} , respectively). We also confirmed that no novel susceptibility markers appeared in GWAS using alcohol consumption as an additional covariate.

Interaction between rs671 in *ALDH2* and alcohol consumption.

As rs671 has interactive effects with alcohol consumption on phenotypes including metabolic trait^{32,33}, we hypothesized that rs671 and alcohol would show an interactive effect on serum AAT levels as well. As a result, we identified a significant interaction between rs671 genotypes and alcohol intake (GLM $P = 2.2 \times 10^{-4}$, Fig. 3). rs671 A allele showed a trend of increase in its effect on AAT in those who were heavily drunk (Fig. 3). The results of other interactions are shown in Supplementary Table 5. While we also found an interaction between rs671 and smoking (GLM $P = 6.8 \times 10^{-4}$, Supplementary Table 5 and Supplementary Fig. 9), this seems to be explained by strong correlation between alcohol consumption and smoking ($r_s = 0.41$, Student's t distribution test $P = 7.2 \times 10^{-380}$). In fact, when the interaction between rs671 and smoking was evaluated by conditioning on the interaction between rs671 and alcohol intake, the effect was not

Table 3 | The associations between SNPs reported in the European GWAS and serum AAT levels in the Japanese population.

SNP	Chr	Position	Gene	Ref/var	Location	The European study				The current study			
						N	VAF	Effect	P	N	VAF	Effect	P
rs2736887	14	94812980	SERPINA2A/ SERPINA6	G/C	Intergenic	1,392	0.185	7.1	2.5×10^{-13}	3,294	0.576	0.34	0.44
rs926144	14	94813402	SERPINA2A/ SERPINA6	A/G	Intergenic	1,392	0.186	7.1	2.7×10^{-13}	3,294	0.549	0.29	0.51
rs4905179	14	94795492	SERPINA6	A/G	Upstream	1,392	0.180	6.8	1.2×10^{-12}	3,294	0.476	0.35	0.42
rs11621961	14	94769476	SERPINA6	C/T	Downstream	1,392	0.355	5.2	1.4×10^{-11}	3,294	0.310	0.37	0.44
rs2896268	14	94865708	SERPINA1	A/C	Upstream	5,569	0.495	1.3	4.1×10^{-5}	3,294	0.502	1.61	2.4×10^{-4}

Chr, chromosome; N, number of subjects; P, P value calculated by generalized linear regression model; ref/var: reference allele/variant allele; VAF, variant allele frequency.
Effects in European study are converted to mg dl⁻¹.

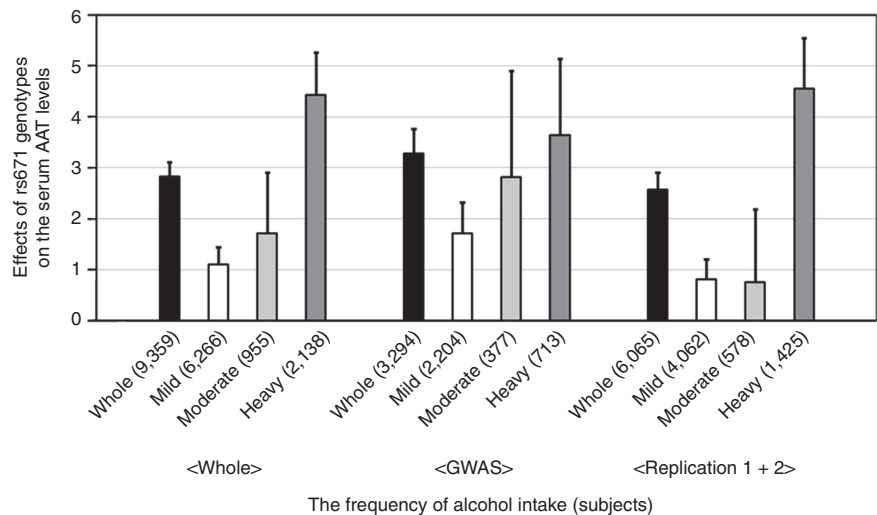


Figure 3 | The interaction between the rs671 genotypes and alcohol drinking on the serum AAT levels. The effects of rs671 on serum AAT levels in linear regression analysis are shown in the groups subdivided into three groups based on alcohol intake. Error bars represent standard errors.

significant (GLM $P=0.16$). When we assessed the interactions between rs1260326 in *GCKR* or rs1169288 in *HNF1A* and alcohol or smoking, no significant interactions were observed (Supplementary Table 6).

Association between FEV1.0 and the three SNPs. Next, we evaluated whether these genetic components associated with AAT levels in the current study are associated with lung function alone or in combination because patients with AATD presented with COPD in their youth. At first, we did not find a significant association between serum AAT levels and FEV1.0 (Supplementary Fig. 10a). While rs1260326 and rs1169288 did not show significant associations with FEV1.0, rs671 showed a negative association with FEV1.0 (GLM $P=4.5 \times 10^{-5}$, Supplementary Fig. 10b). We observed this association even after conditioning for AAT levels (GLM $P=6.1 \times 10^{-5}$, Supplementary Fig. 11a), suggesting the association of *ALDH2* unrelated to AAT. Therefore, the sum of susceptibility alleles of these three SNPs showing a suggestive negative association with FEV1.0 can be explained by the negative association of rs671 (GLM $P=0.0063$ and 0.0086 , before and after conditioning on AAT, Supplementary Figs 10C and 11B). As smoking has a big impact on the development of COPD in patients with AATD³⁴, we evaluated the association between FEV1.0 and rs671 based on smoking status. The negative association was found to be observed mainly in the non-smoking population (GLM $P=8.1 \times 10^{-5}$, Supplementary Fig. 12).

Comparison of associated SNPs between Japanese and Europeans. Finally, to evaluate the SNPs significantly associated with the serum AAT level in the European population, we searched the effects in the current study of the 14 SNPs whose significant or suggestive associations were reported in the European GWAS. Nine out of the 14 SNPs, including rs17580 (PI S) and rs28929474 (PI Z), showed no or very small minor allele frequencies (minor allele frequency (MAF) ≤ 0.006). The results of the remaining five SNPs are shown in Table 3. Although rs2896268 showed an association with the comparable effect size to the European study (GLM $P=2.4 \times 10^{-4}$, Table 3), the other four markers did not show a trend of association in spite of high minor allele frequencies (GLM $P \geq 0.42$, MAF ≥ 0.31 , Table 3).

Discussion

The current GWAS and the following two-staged replication study identified a total of three missense SNPs significantly associated with serum AAT levels in the *ALDH2* and *HNF1A* loci on chromosome 12 and the *GCKR* locus on chromosome 2. *ALDH2* and *HNF1A* have been reported to be functionally associated with AAT. Significant correlations of gene expression between *SERPINA1* and all of the three genes support functional relevance of these genes to AAT.

While the SNPs in LD with rs671 contained two other genes and *BRAP* was associated with metabolic syndrome³⁵ as well as *ALDH2*, the deleterious effect of rs671, the strongest association of the *ALDH2* region and the functional involvement of *ALDH2*

with AAT activity suggest *ALDH2* as a responsible gene. rs671 is known to strongly affect alcohol consumption by making people get drunk easily and be light drinkers through decreasing *ALDH2* activity^{23,30}. The decrease of *ALDH2* activity is attributed to the impaired binding affinity to coenzyme, NAD⁺ (ref. 36) and triggers the facial flushing caused by alcohol intake^{37,38} and local inflammatory diseases^{39–41}. The heterozygous rs671 shows 6% of the normal activity, and homozygous variant shows almost negative activity¹⁸. Although the alteration of alcohol consumption by rs671 seemed to explain more than half of the association in the *ALDH2* region, the substantial association remained after conditioning on alcohol consumption. The significant interaction between rs671 and alcohol intake revealed the stronger association of rs671 in heavy drinkers. Functional annotation analysis revealed DNaseI hypersensitivity site (DHS) and enhancer activity in liver-derived cells at rs4646776 in intron 8 of *ALDH2*, in almost complete LD with rs671 (Supplementary Table 3). Thus, (1) alteration of alcohol consumption by rs671, (2) interaction between rs671 and alcohol consumption and (3) potential transcriptional alteration of *ALDH2* induced by a haplotype of rs671 and rs4646776 may explain the association in this region.

While functional importance of rs1169288 on *HNF1A* protein is not known and *in silico* analysis using Polyphen-2 suggests a benign change brought about by this SNP, conserved search suggested the importance of this SNP. In fact, rs1169288 has been reported for its association with impairment of insulin secretion²¹ and risk of type 2 diabetes mellitus²⁴ and cardiovascular diseases²⁵ in community-dwelling populations. The functional annotation search revealed active promoter effects in liver-derived cells at rs1169288 and the other two SNPs in LD with rs1169288 (Supplementary Table 3). It also showed DHS and binding of transcription factor in hepatic cells at rs1169288 and one of the two SNPs. A conformational or transcriptional alteration of *HNF1A* via rs1169288 and/or SNPs in LD with rs1169288 would lead to modulation of the regulatory effects of *HNF1A* on *SERPINA1* expression and explain the association in the *HNF1A* region.

While *in silico* analysis using Polyphen-2 did not result in damaging effect brought about by rs1260326, a recent *in vitro* analysis reported that the C allele of rs1260326 leads to increased glucokinase activity in the liver and resultant increase of triglyceride and decrease of glucose levels²⁷. Functional search revealed enhancer activity and DHS in liver-derived cells at rs1260326 (Supplementary Table 3). It also showed that three other SNPs in LD with rs1260326 have strong enhancer activity and transcription factor binding in hepatic cells. rs1260326 was reported to be associated with metabolic traits including HOMA-IR⁴², lipid profiles²², NAFLD and NASH with fibrosis⁴³. Conservation of proline across vertebrates may suggest advantage of increased glucokinase activity and AAT levels. An increased glucokinase activity and potential transcriptional alteration in the liver due to rs1260326 and other SNPs in LD with rs1260326 would explain the association between AAT and the *GCKR* region.

It should be noted that all of the three genes are associated with metabolic phenotypes and are expressed in the liver. As AAT is predominantly produced in the liver, the associations between these three genes and AAT suggest that these genes modulate AAT levels mainly in the liver. When we analysed whether the three SNPs display *cis* associations with expressions of *ALDH2*, *GCKR* and *HNF1A*, respectively, or *trans* associations with expression of *SERPINA1*, we did not observe significant associations. Since functional annotation analysis suggested enhancer activity and/or DHS at the three SNPs and/or variants in strong LD with the SNPs in liver-derived cells, we might detect

significant *cis* or *trans* effects of the SNPs in expression data from liver-derived cells. As the three variants are missense variants of genes predominantly expressed in the liver in which AAT is mainly produced, it is also likely that these three SNPs influence serum AAT levels through conformational change derived from amino-acid alterations, especially rs671.

While AAT is an enzyme controlling inflammation by inhibiting proteases like the neutrophil elastase, recent studies have revealed a broad range of AAT functions including inflammation, apoptosis, cellular senescence, lipid metabolism and proteostasis. Thus, the associations between AAT and the three loci related to metabolic phenotypes suggest the involvement of AAT with the metabolic syndrome-related pathway beyond its function as an acute inflammatory marker.

While patients with AATD present with COPD, we did not find a significant association between AAT levels and FEV1.0. In addition, we did not find significant associations between FEV1.0 and genotypes except for rs671. As the A allele of rs671 is associated with increased AAT levels and decreased FEV1.0, the association between rs671 and FEV1.0 does not match with the relationship between AATD and COPD. We did not observe a significant positive association between rs671 and FEV1.0 even in the smoking population (Supplementary Fig. 12). Further replication study is necessary to confirm the negative association of rs671 A allele with FEV1.0 independent of AAT. The poor translation of genetic associations with AAT into lung function was also observed in the European study¹³. As lung functions are finely regulated by multiple complexed mechanisms including smoking and systemic or local inflammation, alterations of lung function derived from these three genetic components associated with serum AAT levels might be limited and difficult to identify. Another possibility is that serum AAT levels do not precisely represent AAT levels in the lung. Uncovering mechanisms regulating AAT levels in the lung and appropriate modelling for lung AAT levels would lead to elucidation of the importance of genetic components associated with AAT levels on lung function.

The three SNPs were not reported in the previous European GWAS¹³. As the A allele of rs671 was frequently observed in Asian populations and is rarely seen in European populations, the lack of the association between rs671 and AAT in the European study seems appropriate. In addition, all of the SNPs in strong LD with rs671 in the Japanese population showed very small allele frequencies in European population (Supplementary Table 3). In contrast, rs1169288 and rs1260326 have comparable allele frequencies in Asians and Europeans (0.539 versus 0.283 and 0.490 versus 0.600, respectively, Hapmap phase II data). There are several possibilities to explain the discrepancy of the two associations between the two populations. First, the lack of power ($n = 1,392$) or overrepresentation of asthmatic patients in the previous European GWAS may explain the lack of significant associations in the Europeans. Second, there might be other causative variants that are in LD with the two SNPs that can specifically be found in the Japanese and not in the European population due to the different LD structures between populations. However, no SNPs in strong LD with the two SNPs in Japanese showed highly different allele frequencies from the two SNPs in European population (Supplementary Table 3). As the two SNPs were associated with multiple phenotypes in European populations and the functional importance of rs1169288 was experimentally shown, the first possibility seems more likely. It is reasonable to perform quantitative trait linear regression analysis of AAT levels in a large-scale healthy European population.

The four SNPs in the *SERPINA* gene clusters showing significant associations in the European study, namely,

rs2736887, rs926144, rs4905179 and rs11621961, did not show evidence of associations in the current study. As the associations in the European population were shown to largely depend on PI Z and PI S and these two variants were not found in the current participants, the lack of associations of these four SNPs in the current study seems reasonable. We found a suggestive association of rs2896268 with a comparable effect size to the European population. rs2896268 showed an association independent of PIS and PIZ in the Europeans, suggesting a common mechanism in the regulation of *SERPINA1* expression beyond ethnicity that is independent of PI S and PI Z.

As our results showed an association of the Asian-specific variant and associations of the common variants in *GCKR* and *HNF1A* with serum AAT levels, further replication studies and functional experiments recruiting other Asian populations and Europeans are necessary.

The increase of serum CRP levels is an independent risk factor of developing metabolic syndrome⁴⁴, diabetes mellitus⁴⁵, stroke and cardiac infarction⁴⁶, and therefore it is reasonable to quantify AAT levels in a prospective manner to test AAT as the possible independent risk factor for metabolic syndrome, diabetes mellitus or stroke. In addition, Mendelian randomization approach^{47,48} to test the association between these three markers and developing metabolic syndrome, diabetes mellitus or stroke would clarify the direct or indirect function of AAT beyond acute phase inflammation marker.

Methods

Study subjects. This study was performed as a part of The Nagahama Prospective Genome Cohort for Comprehensive Human Bioscience (the Nagahama Study)¹⁴. The participants of The Nagahama Study were recruited from the general population living in Nagahama City, a largely rural city of 125,000 inhabitants in Shiga Prefecture, located in the centre of Japan. Persons aged 30 to 74 years, living independently in the community and with no physical impairment or dysfunction, were enrolled and detailed questionnaires were answered by the participants. Among 9,804 participants recruited from 2008 to 2010, 9,761 subjects were registered for our analyses excluding pregnant women ($n = 43$). All study procedures were approved by the ethics committee of Kyoto University Graduate School of Medicine and the Nagahama Municipal Review Board. Written informed consent was obtained from all participants.

Genotyping. Among the total of 9,761 samples, genome scanning was conducted on 3,693 samples by using one of or combinations of human hap610K quad array (hap610k), human omni 2.5M-4 array (2.5M-4), human omni 2.5M-8 array, (2.5M-8) human omni 2.5s array (2.5M-s), human exome (Exome) and human core exome (CoreExome) (Illumina, San Diego, CA, USA). The detailed distribution of samples for genome-scanning arrays are shown in Supplementary Table 7. The remaining 4,023 and 2,042 DNA samples were used for the replication studies 1 and 2, respectively. The details of sample selection are written in Supplementary Note 1 and Supplementary Fig. 2. The genotyping of the SNPs in the replication studies was conducted by TaqMan assay (Applied Biosystems).

Quality control for the genotyping data. In the genome scanning of 3,693 subjects, 57 individuals were excluded with a low call rate less than 0.90, 0.95 and 0.99 in the hap610k, 2.5M-4 and other plat forms, respectively. Seven individuals were excluded due to outliers from Asian clusters by principal component analysis with HapMap Phase 2 release 28 JPT data set as reference (EIGENSTRAT ver. 2.0, Supplementary Fig. 3)⁴⁹ and a total of 316 subjects were excluded due to high degrees of kinship (Pi-hat greater than 0.35, PLINK ver. 1.07)⁵⁰. Nineteen samples were excluded because they lack information of smoking status.

SNPs with a call rate less than 0.99, MAF below 0.01 or showing departure from HWE (χ^2 -test $P < 1.0 \times 10^{-6}$) were excluded from the analyses. As a result, 267,882 SNPs remained for the following genotype imputation. The details of these quality check procedures are shown in Supplementary Fig. 2.

Genotype imputation. Genotype imputation was conducted using a standard procedure with MACH ver. 1.0.16 software⁵¹ using East Asian panel in the 1000 Genome project as a reference. Imputed SNPs with MAF below 0.01 or R-square (RSQ) below 0.3 were removed from the subsequent association analysis. In total, 6,569,727 SNPs were used in the following GWAS on the serum AAT level.

Quantification of AAT and CRP levels. The serum levels of AAT and high-sensitive CRP (hs-CRP) were measured by nephelometry on BN II (R) Nephelometer (Siemens Healthcare Diagnostics, Munchi, Germany) using N-antisera to human alpha1-antitrypsin kit and N-latex CRP II kit (the same company), respectively. These assays were performed at SRL Laboratories (Tokyo, Japan) and were commercially available. The intra- and inter- assay coefficients of variation of the AAT levels were 1.07–1.67% and 1.02–1.31%, respectively.

Evaluation of LD. LD between markers were evaluated by PLINK⁵⁰ software or Haploview software⁵².

Heterogeneity of the genetic studies. Heterogeneity among GWAS and the two replication studies was assessed by Cochran's Q test for the three SNPs showing significant associations beyond the GWAS significant level in the combined study.

Variance explained by the significantly associated SNPs. Variance that was explained by the three SNPs significantly associated with AAT levels was calculated based on the following formula:

$$V_e = E^2 \times 2 \times (1 - MAF) \times MAF \times V_{tot}^{-1} \quad (1)$$

Where V_e is an explained variance by an SNP, E is an effect size of the SNP, MAF is minor allele frequency of the SNP and V_{tot} is total variance of AAT.

Functional annotation and conservation of amino acids. We used HaploReg¹⁷ and analysed whether SNPs in strong LD with the top SNPs have functional effects including promoter or enhancer activity, DHS and binding of transcription factor. *In silico* analysis of assessing effects of amino-acid alteration induced by the associated markers were performed by Polyphen-2 software²⁸. Conservation of amino-acid sequences including amino acids altered by the associated markers was analysed by using UCSC Genome Browser on Human Feb. 2009 (GRCh37/hg19) Assembly.

Gene expression analysis. The expression data were obtained by the database of whole peripheral blood in the Japanese population²⁹ (Human Genetic Variation Browser: <http://www.genome.med.kyoto-u.ac.jp/SnpDB/>). We used the expression data of *SERPINA1* (Target ID: NM_001002236), a transcriptional variant of *SERPINA1* (*SERPINA1_Var*) (Target ID: ENST_00000402629), *GCKR* (Target ID: NM_001486), *ALDH2* (Target ID: NM_000690) and *HNF1A* (Target ID: NM_000545). We assessed the correlation between gene expressions by Spearman's rank correlation coefficient. The P values for the significance of the coefficients were calculated by Student's t distribution test. The associations between gene expressions and genotypes were analysed by generalized linear regression model adjusted by sex, age, BMI, smoking status, the experimental day and the logarithm of hs-CRP. The smoking status consists of three categories, namely, never-smoker, ever-smoker and current-smoker.

The interactions between correlates on the serum AAT level. The interactions between SNPs were analysed in linear regression model with the use of the same covariates described in the Statistical analysis section. The interactions between each of the three SNPs and alcohol consumption consisting of three categories (Heavy: > 3 days week⁻¹, Moderate: 2–3days week⁻¹, Mild: < 2 days week⁻¹) or Brinkmann index, defined by smoking pack day⁻¹ multiplied by year which reflects the lifetime smoking exposure⁵³, were also analysed. We also analysed interaction between rs671 and age, sex, BMI or CRP.

Statistical analysis. GWAS and replication studies on the serum AAT levels were conducted by GLM adjusting for the recruiting year, age, sex, BMI, Brinkman index and the common logarithm of hs-CRP. These covariates used in the analyses were selected from non-genetic factors considered in another GWAS report (age, sex and the smoking status)¹³ and based on the perceptions on AAT as an acute-phase inflammation marker². Analysis adding alcohol consumption was also performed. The cigarette smoking and alcohol consumption history was obtained from the questionnaire. When we conducted conditional GWAS to find other candidate SNPs associated with serum AAT level independent of each of the associated SNPs, we added the SNP genotype into the former liner model as a covariate. Correlation between smoking or rs671 and alcohol consumption was assessed by Spearman's rank sum correlation coefficient.

All statistical calculations were conducted using the free software R (<http://www.r-project.org/>)⁵⁴ or PLINK. P values less than 5.0×10^{-8} were considered to satisfy the genome-wide significance in the combined study. Otherwise, P value less than 0.05 after Bonferroni's correction was regarded as significant.

References

1. Asofsky, R. & Thorbecke, G. J. Sites of formation of immune globulins and of a component of c(3): ii. Production of immunoelectrophoretically identified serum proteins by human and monkey tissues in vitro. *J. Exp. Med.* **114**, 471–483 (1961).

2. Gabay, C. & Kushner, I. Acute-phase proteins and other systemic responses to inflammation. *N. Engl. J. Med.* **340**, 448–454 (1999).
3. Hunt, J. M. & Tuder, R. Alpha 1 anti-trypsin: one protein, many functions. *Curr. Mol. Med.* **12**, 827–835 (2012).
4. Lai, E. C., Kao, F. T., Law, M. L. & Woo, S. L. Assignment of the alpha 1-antitrypsin gene and a sequence-related gene to human chromosome 14 by molecular hybridization. *Am. J. Hum. Genet.* **35**, 385–392 (1983).
5. Schroeder, W. T., Miller, M. F., Woo, S. L. & Saunders, G. F. Chromosomal localization of the human alpha 1-antitrypsin gene (PI) to 14q31–32. *Am. J. Hum. Genet.* **37**, 868–872 (1985).
6. Cox, D. W., Markovic, V. D. & Teshima, I. E. Genes for immunoglobulin heavy chains and for alpha 1-antitrypsin are localized to specific regions of chromosome 14q. *Nature* **297**, 428–430 (1982).
7. Poller, W. *et al.* A leucine-to-proline substitution causes a defective alpha 1-antichymotrypsin allele associated with familial obstructive lung disease. *Genomics* **17**, 740–743 (1993).
8. Stoller, J. K. *et al.* American Thoracic Society/European Respiratory Society Statement: Standards for the diagnosis and management of individuals with alpha-1 antitrypsin deficiency. *Pneumologie* **59**, 36–68 (2005).
9. Poller, W. *et al.* Mis-sense mutation of alpha 1-antichymotrypsin gene associated with chronic lung disease. *Lancet* **339**, 1538 (1992).
10. Lomas, D. A. Loop-sheet polymerization: the structural basis of Z alpha 1-antitrypsin accumulation in the liver. *Clin. Sci.* **86**, 489–495 (1994).
11. Lomas, D. A., Evans, D. L., Finch, J. T. & Carrell, R. W. The mechanism of Z alpha 1-antitrypsin accumulation in the liver. *Nature* **357**, 605–607 (1992).
12. Zorzetto, M. *et al.* SERPINA1 gene variants in individuals from the general population with reduced alpha1-antitrypsin concentrations. *Clin. Chem.* **54**, 1331–1338 (2008).
13. Thun, G. A. *et al.* Causal and synthetic associations of variants in the SERPINA gene cluster with alpha1-antitrypsin serum levels. *PLoS Genet.* **9**, e1003585 (2013).
14. Yoshimura, K. *et al.* Prevalence of postmicturition urinary incontinence in Japanese men: Comparison with other types of incontinence. *Int. J. Urol.* (2013).
15. Consortium, G. P. A map of human genome variation from population-scale sequencing. *Nature* **467**, 1061–1073 (2010).
16. Devlin, B. & Roeder, K. Genomic control for association studies. *Biometrics* **55**, 997–1004 (1999).
17. Ward, L. D. & Kellis, M. HaploReg: a resource for exploring chromatin states, conservation, and regulatory motif alterations within sets of genetically linked variants. *Nucleic Acids Res.* **40**, D930–D934 (2012).
18. Enomoto, N., Takase, S., Yasuhara, M. & Takada, A. Acetaldehyde metabolism in different aldehyde dehydrogenase-2 genotypes. *Alcohol Clin. Exp. Res.* **15**, 141–144 (1991).
19. Brecher, A. S., Thevananther, S. & Franco-Saenz, R. Acetaldehyde inhibits the anti-elastase activity of alpha 1-antitrypsin. *Alcohol* **11**, 181–185 (1994).
20. Rollini, P. & Fournier, R. E. The HNF-4/HNF-1alpha transactivation cascade regulates gene activity and chromatin structure of the human serine protease inhibitor gene cluster at 14q32.1. *Proc. Natl Acad. Sci. USA* **96**, 10308–10313 (1999).
21. Florez, J. C. Newly identified loci highlight beta cell dysfunction as a key cause of type 2 diabetes: where are the insulin resistance genes? *Diabetologia* **51**, 1100–1110 (2008).
22. Shen, Y. *et al.* Common genetic variants associated with lipid profiles in a Chinese pediatric population. *Hum. Genet.* **132**, 1275–1285 (2013).
23. Takeuchi, F. *et al.* Confirmation of ALDH2 as a Major locus of drinking behavior and of its variants regulating multiple metabolic phenotypes in a Japanese population. *Circ. J.* **75**, 911–918 (2011).
24. Holmkvist, J. *et al.* Common variants in maturity-onset diabetes of the young genes and future risk of type 2 diabetes. *Diabetes* **57**, 1738–1744 (2008).
25. Reiner, A. P. *et al.* Common coding variants of the HNF1A gene are associated with multiple cardiovascular risk phenotypes in community-based samples of younger and older European-American adults: the Coronary Artery Risk Development in Young Adults Study and The Cardiovascular Health Study. *Circ. Cardiovasc. Genet.* **2**, 244–254 (2009).
26. Ridker, P. M. *et al.* Loci related to metabolic-syndrome pathways including LEPR, HNF1A, IL6R, and GSKR associate with plasma C-reactive protein: the Women's Genome Health Study. *Am. J. Hum. Genet.* **82**, 1185–1192 (2008).
27. Beer, N. L. *et al.* The P446L variant in GSKR associated with fasting plasma glucose and triglyceride levels exerts its effect through increased glucokinase activity in liver. *Hum. Mol. Genet.* **18**, 4081–4088 (2009).
28. Adzhubei, I. A. *et al.* A method and server for predicting damaging missense mutations. *Nat. Methods* **7**, 248–249 (2010).
29. Narahara, M. *et al.* Large-scale East-Asian eQTL mapping reveals novel candidate genes for LD mapping and the genomic landscape of transcriptional effects of sequence variants. *PLoS ONE* **9**, e100924 (2014).
30. Higuchi, S. *et al.* The relationship between low Km aldehyde dehydrogenase phenotype and drinking behavior in Japanese. *J. Stud. Alcohol* **53**, 170–175 (1992).
31. Senn, O. *et al.* Circulating alpha1-antitrypsin in the general population: determinants and association with lung function. *Respir. Res.* **9**, 35 (2008).
32. Tan, A. *et al.* A genome-wide association and gene-environment interaction study for serum triglycerides levels in a healthy Chinese male population. *Hum. Mol. Genet.* **21**, 1658–1664 (2012).
33. Matsuo, K. *et al.* Gene-environment interaction between an aldehyde dehydrogenase-2 (ALDH2) polymorphism and alcohol consumption for the risk of esophageal cancer. *Carcinogenesis* **22**, 913–916 (2001).
34. American Thoracic, S. & European Respiratory, S. American Thoracic Society/European Respiratory Society statement: standards for the diagnosis and management of individuals with alpha-1 antitrypsin deficiency. *Am. J. Respir. Crit. Care Med.* **168**, 818–900 (2003).
35. Wu, L. *et al.* The single nucleotide polymorphisms in BRAP decrease the risk of metabolic syndrome in a Chinese young adult population. *Diab. Vasc. Dis. Res.* **10**, 202–207 (2013).
36. Larson, H. N., Weiner, H. & Hurley, T. D. Disruption of the coenzyme binding site and dimer interface revealed in the crystal structure of mitochondrial aldehyde dehydrogenase "Asian" variant. *J. Biol. Chem.* **280**, 30550–30556 (2005).
37. Wolff, P. H. Ethnic differences in alcohol sensitivity. *Science* **175**, 449–450 (1972).
38. Harada, S., Agarwal, D. P. & Goedde, H. W. Aldehyde dehydrogenase deficiency as cause of facial flushing reaction to alcohol in Japanese. *Lancet* **2**, 982 (1981).
39. Hira, A. *et al.* Variant ALDH2 is associated with accelerated progression of bone marrow failure in Japanese Fanconi anemia patients. *Blood* **122**, 3206–3209 (2013).
40. Wang, Q. *et al.* ALDH2 rs671 Polymorphism and coronary heart disease risk among Asian populations: a meta-analysis and meta-regression. *DNA Cell Biol.* **32**, 393–399 (2013).
41. Au Yeung, S. L. *et al.* Aldehyde dehydrogenase 2-a potential genetic risk factor for lung function among southern Chinese: evidence from the Guangzhou Biobank Cohort Study. *Ann. Epidemiol.* **24**, 606–611 (2014).
42. Mente, A. *et al.* Causal relationship between adiponectin and metabolic traits: a Mendelian randomization study in a multiethnic population. *PLoS ONE* **8**, e66808 (2013).
43. Tan, H. L. *et al.* Association of glucokinase regulatory gene polymorphisms with risk and severity of non-alcoholic fatty liver disease: an interaction study with adiponutrin gene. *J. Gastroenterol.* **49**, 1056–1064 (2014).
44. Han, T. S. *et al.* Prospective study of C-reactive protein in relation to the development of diabetes and metabolic syndrome in the Mexico City Diabetes Study. *Diabetes Care* **25**, 2016–2021 (2002).
45. Pradhan, A. D., Manson, J. E., Rifai, N., Buring, J. E. & Ridker, P. M. C-reactive protein, interleukin 6, and risk of developing type 2 diabetes mellitus. *JAMA* **286**, 327–334 (2001).
46. Ridker, P. M., Hennekens, C. H., Buring, J. E. & Rifai, N. C-reactive protein and other markers of inflammation in the prediction of cardiovascular disease in women. *N. Engl. J. Med.* **342**, 836–843 (2000).
47. Timpson, N. J. *et al.* C-reactive protein and its role in metabolic syndrome: mendelian randomisation study. *Lancet* **366**, 1954–1959 (2005).
48. Brunner, E. J. *et al.* Inflammation, insulin resistance, and diabetes--Mendelian randomization using CRP haplotypes points upstream. *PLoS Med.* **5**, e155 (2008).
49. Price, A. L. *et al.* Principal components analysis corrects for stratification in genome-wide association studies. *Nat. Genet.* **38**, 904–909 (2006).
50. Purcell, S. *et al.* PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**, 559–575 (2007).
51. Scott, L. J. *et al.* A genome-wide association study of type 2 diabetes in Finns detects multiple susceptibility variants. *Science* **316**, 1341–1345 (2007).
52. Barrett, J. C., Fry, B., Maller, J. & Daly, M. J. Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics* **21**, 263–265 (2005).
53. Brinkman, G. L. & Coates, E. O. Jr. The effect of bronchitis, smoking, and occupation on ventilation. *Am. Rev. Respir. Dis.* **87**, 684–693 (1963).
54. Ihaka, R. & Gentleman, R. R. A language for data analysis and graphics. *J. Comput. Graph. Stat.* **5**, 299–314 (1996).

Acknowledgements

We are grateful to Non-profit Organization Zero-ji Club, and the staffs of the Nagahama City Office for their help in conducting the Nagahama Study. We thank all the members of the Nagahama Study. This study was supported by a University Grants and a Grants-in-Aid for Scientific Research from the Ministry of Education, Culture, Sports, Science & Technology in Japan; the Programme for Enhancing Systematic Education in Graduate

School from the Japan Society for the Promotion of Science; a research grant from the Takeda Science Foundation.

Author contributions

K.S., C.T., S.M., T.K., Y.T., M.T., T.N., S.K., A.S., R.Y., M.M. and F.M. conceived and designed the study. K.S., C.T., S.M., T.K., Y.T., M.T., T.N., S.K., A.S., R.Y., M.M. and F.M. acquired samples and data. K.S., C.T. and M.T. performed the experimental work. K.S., C.T., T.K. and Y.T. analysed and interpreted the data. K.S., C.T. and M.T. wrote the paper and all authors revised and approved it to be published. None of the funding sources was related to design and conduct of the study; collection, management, analysis and interpretation of the data; and preparation, review or approval of the manuscript; and decision to submit the manuscript for publication.

Additional information

Supplementary Information accompanies this paper at <http://www.nature.com/naturecommunications>

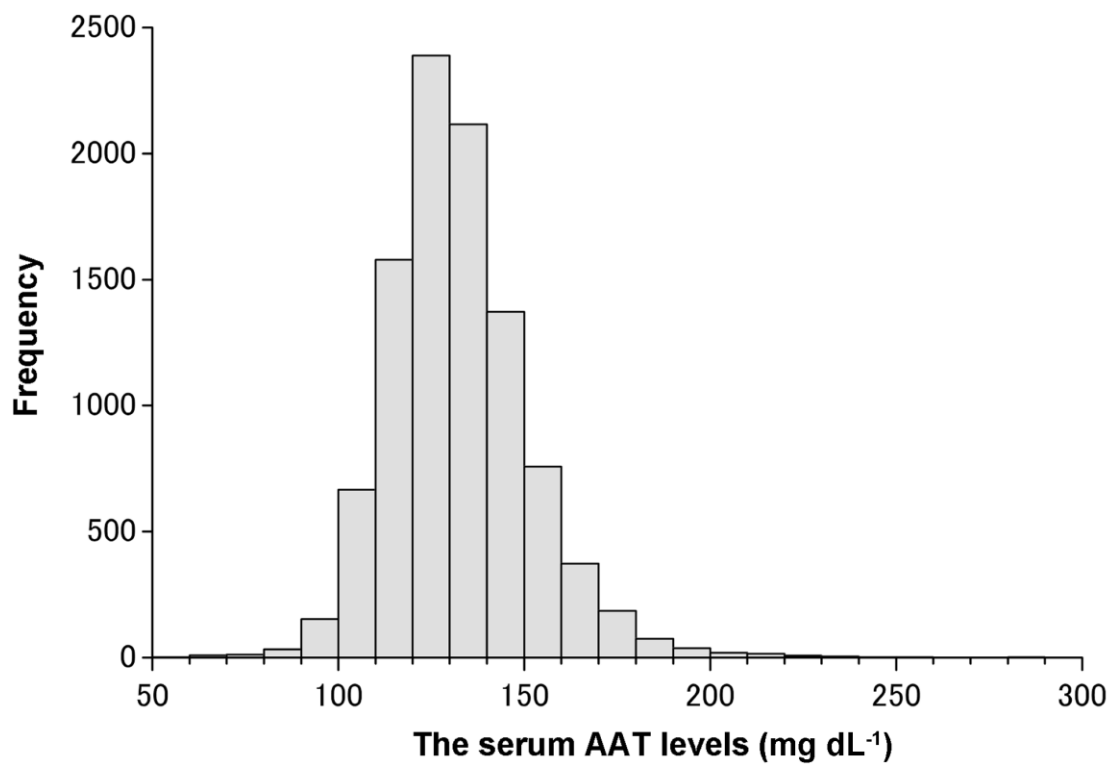
Competing financial interests: The authors declare no competing financial interests.

Reprints and permission information is available online at <http://npg.nature.com/reprintsandpermissions/>

How to cite this article: Setoh, K. *et al.* Three missense variants of metabolic syndrome-related genes are associated with alpha-1 antitrypsin levels. *Nat. Commun.* 6:7754 doi: 10.1038/ncomms8754 (2015).



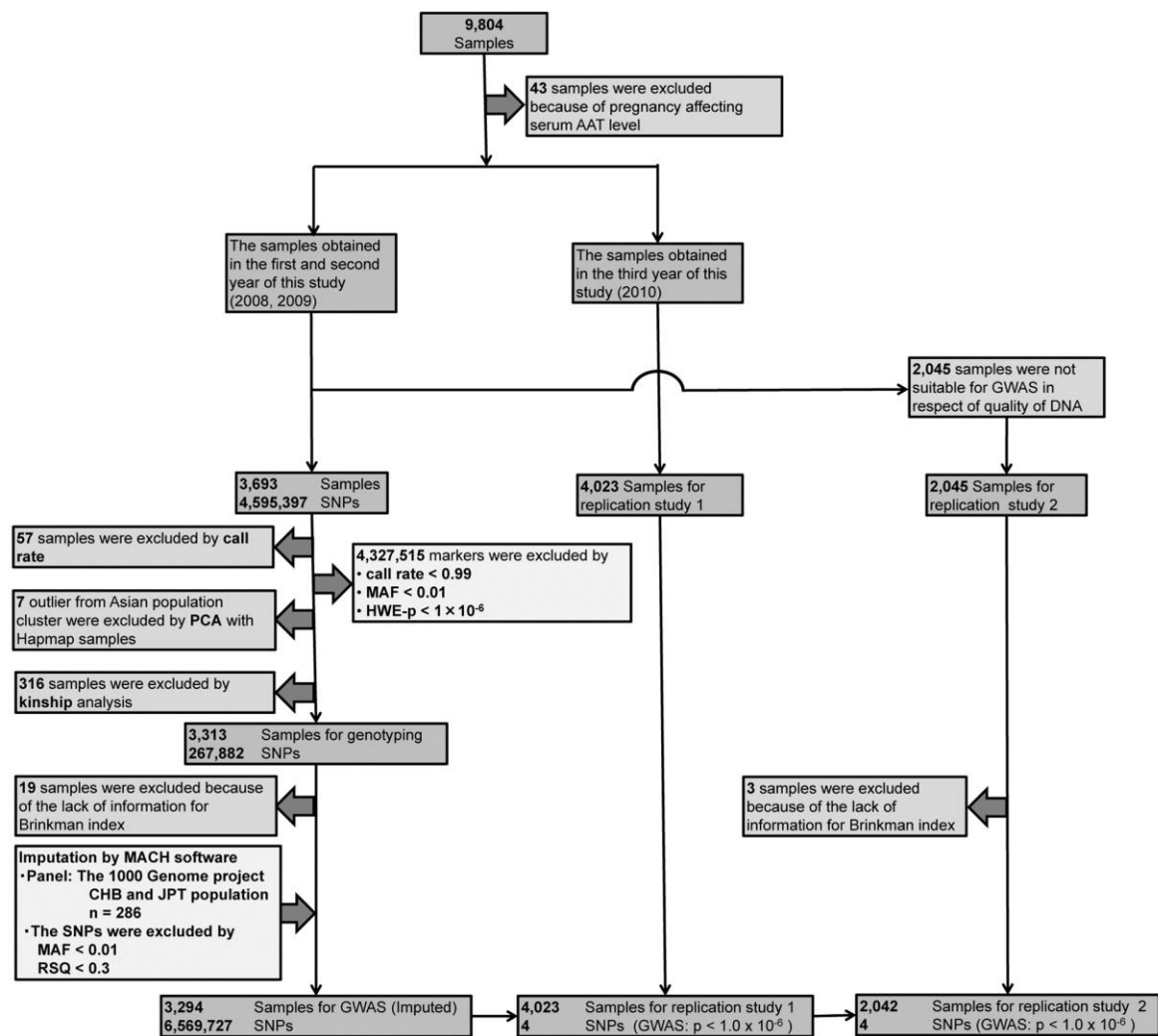
This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>



1

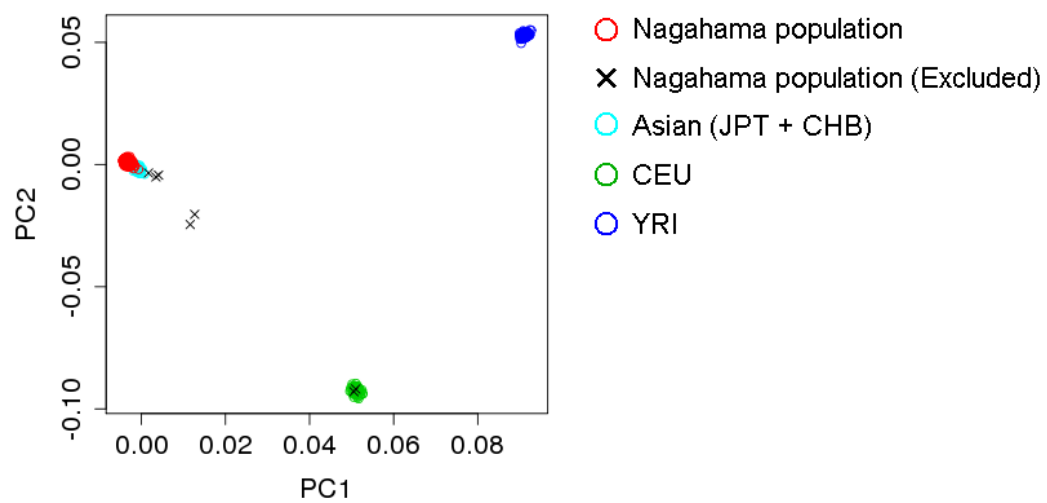
2 **Supplementary Figure 1 | Distribution of serum AAT levels**

3 A histogram of serum AAT levels is indicated.



Supplementary Figure 2 | A flowchart of sample selection and quality control in the current study

A schematic view of sample selection and quality control of samples and SNPs in the current study is indicated.



1

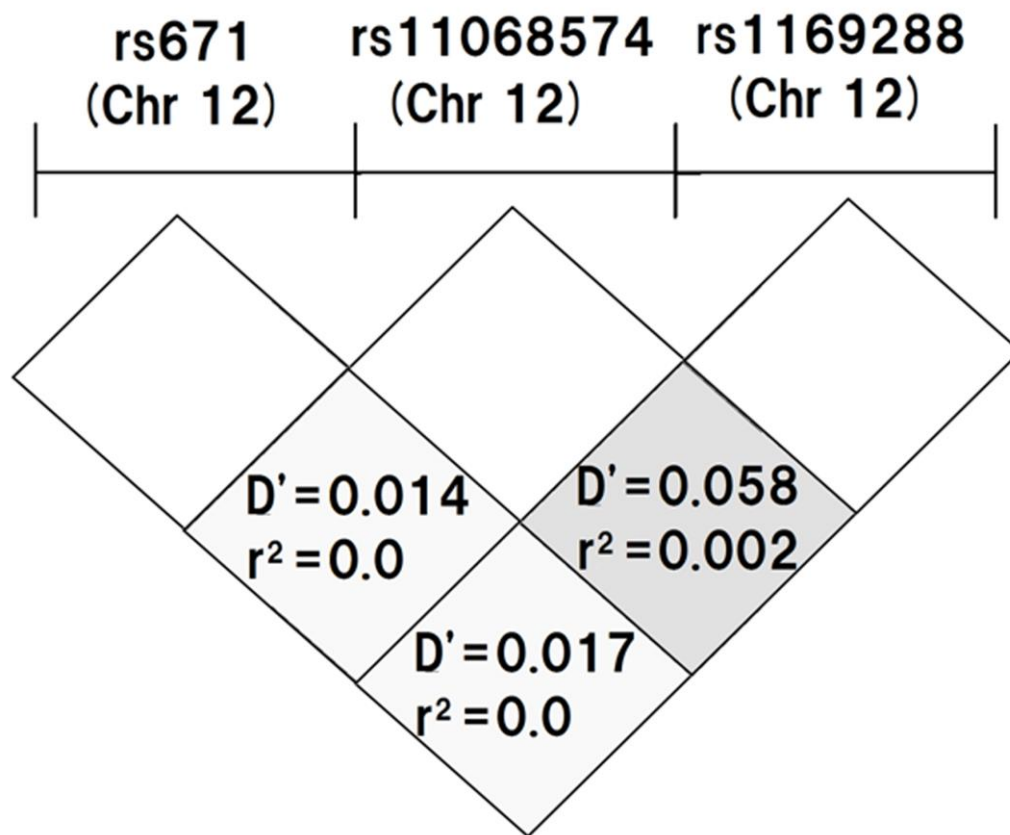
2 **Supplementary Figure 3 | A result of principal component analysis of GWAS data**

3 A result of principal component analysis is indicated according to principal components

4 1 and 2. The 7 samples shown as black crosses were excluded as outliers.

5 PC:principal component, JPT:Japanese in Tokyo, CHB:Chinese in Beijing, CEU:Utah

6 residents with Northern and Western European ancestry, YRI:Yoruba in Ibadan



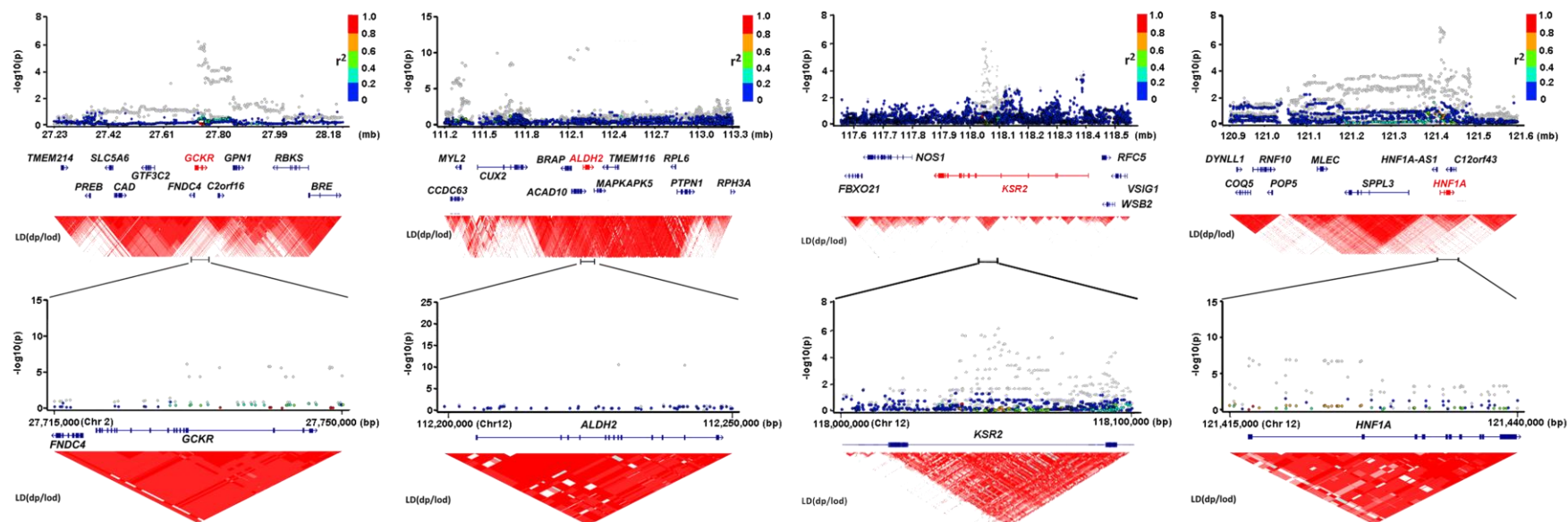
1

2 **Supplementary Figure 4 | Linkage disequilibrium among the three SNPs on**

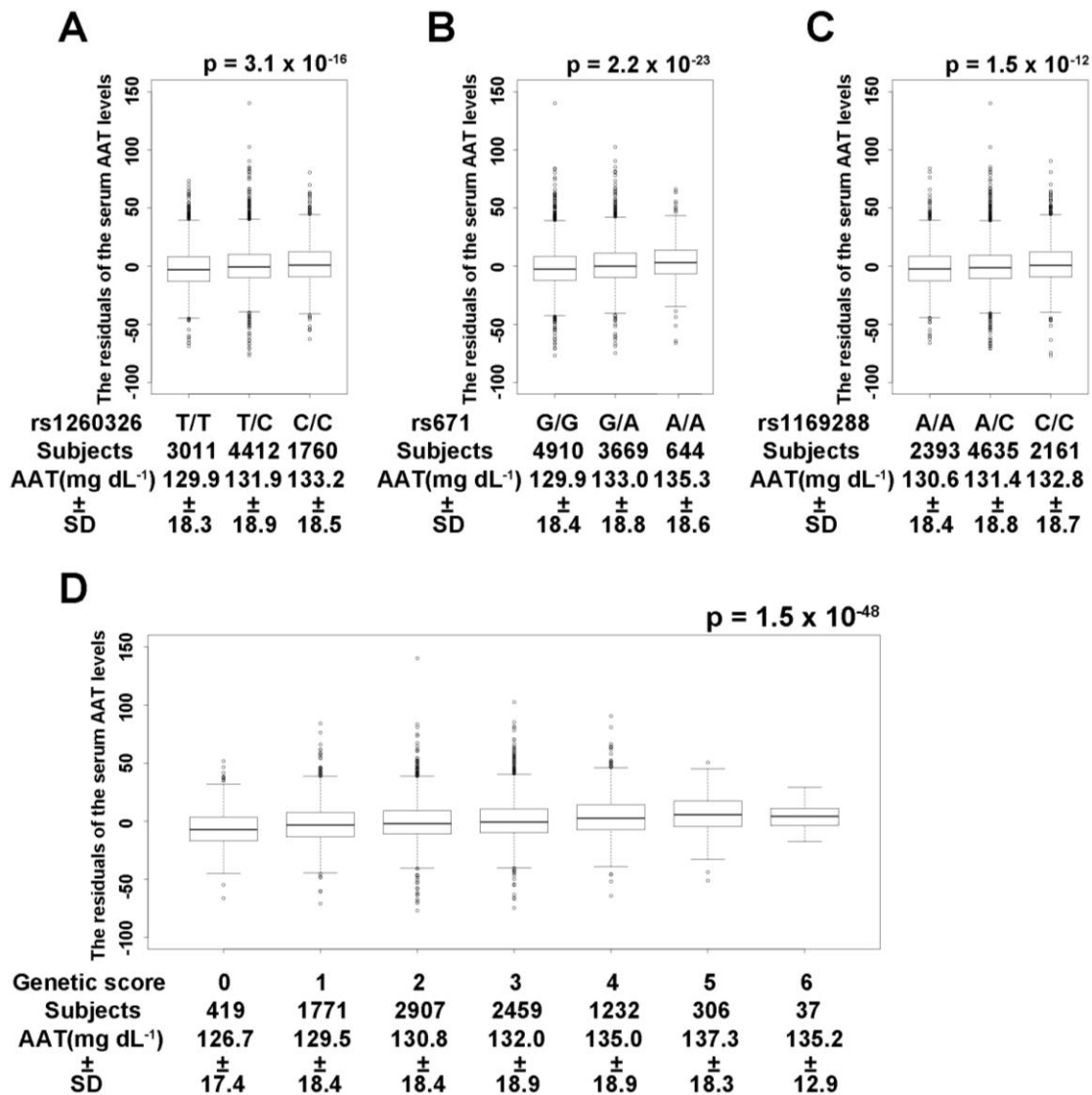
3 **chromosome 12**

4 The LD structure of the three SNPs on chromosome 12 showing associations with AAT

5 in GWAS is indicated.



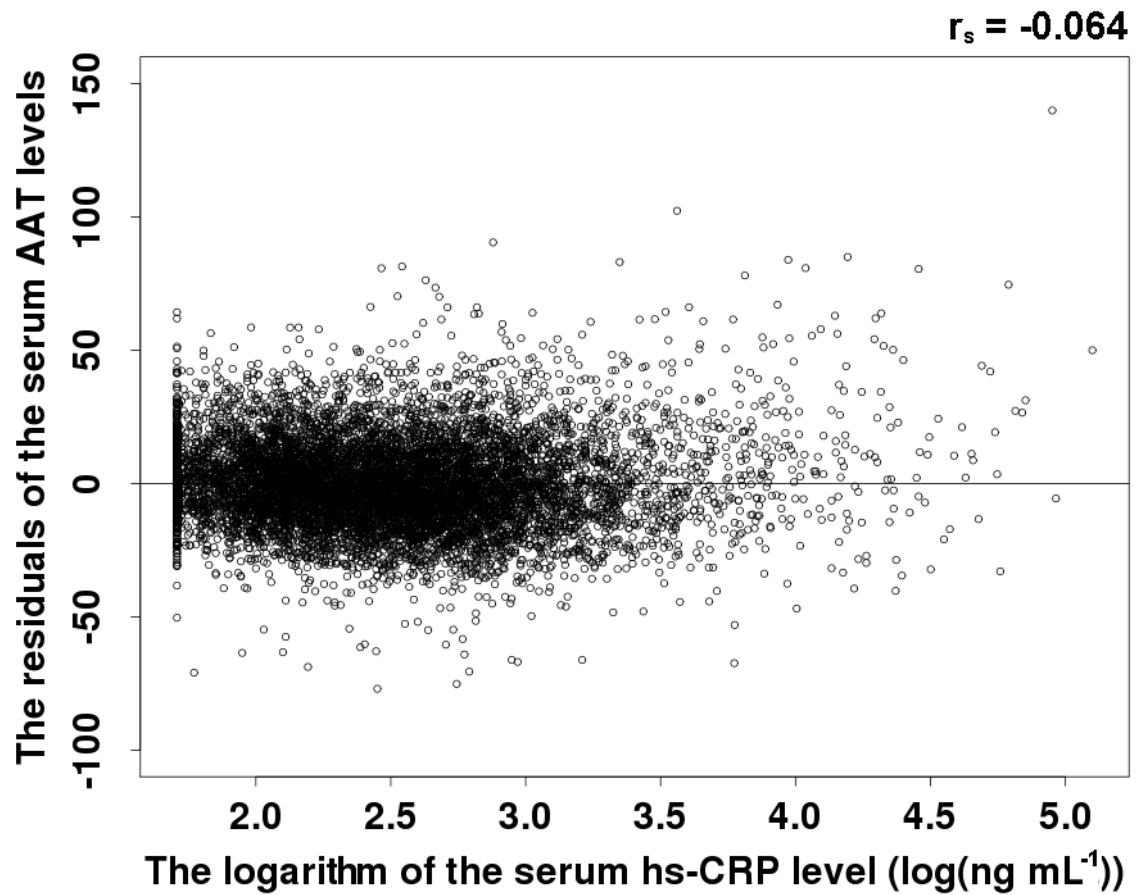
1
2 **Supplementary Figure 5 | Conditional analyses reveal single association in each of the four regions selected for the replication**
3 **studies**
4 Each association is plotted according to its position (X axis) and its strength of association (Y axis). Gray and colored circles in each
5 region indicate associations before and after conditioning on the SNP selected for replication studies, respectively.



Supplementary Figure 6 | The associations between serum AAT levels and the three SNPs alone or in combination

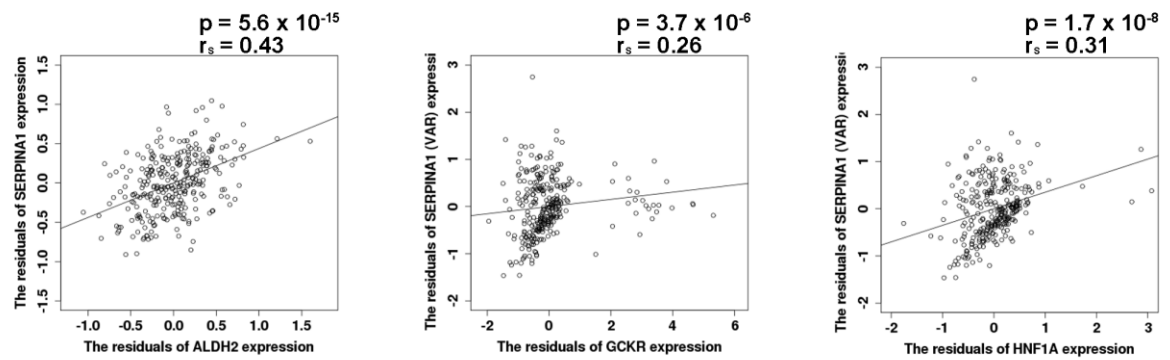
The box plots showing the associations between AAT and A)rs1260326, B)rs671, C)rs1169288 or D) the number of the risk alleles of the three SNPs are indicated. The upper and lower boundaries of the boxes represent the third and first quartiles, respectively. The horizontal lines represent the medians. The whiskers represent more or less than 1.5 times of the interquartile range.

P: p-value calculated by generalized linear regression model.



Supplementary Figure 7 | Lack of confounding associations between CRP levels and residuals of generalized linear regression analysis

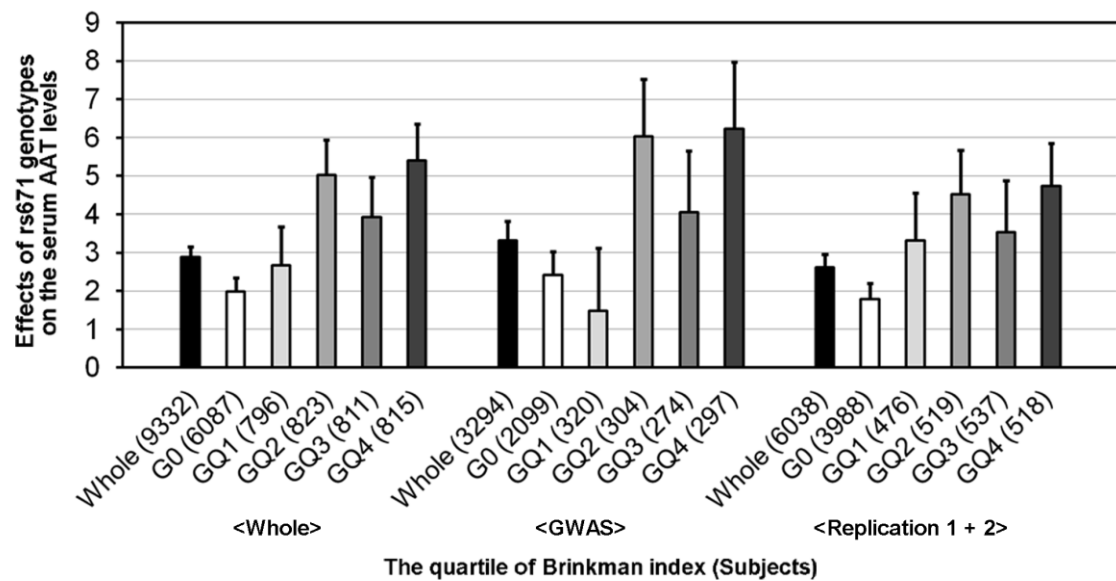
The association between hs-CRP and the residuals of AAT in the generalized linear regression model is indicated.



Supplementary Figure 8 | The associations of gene expressions between *SERPINA1* and *ALDH2*, *GCKR* or *HNF1A*

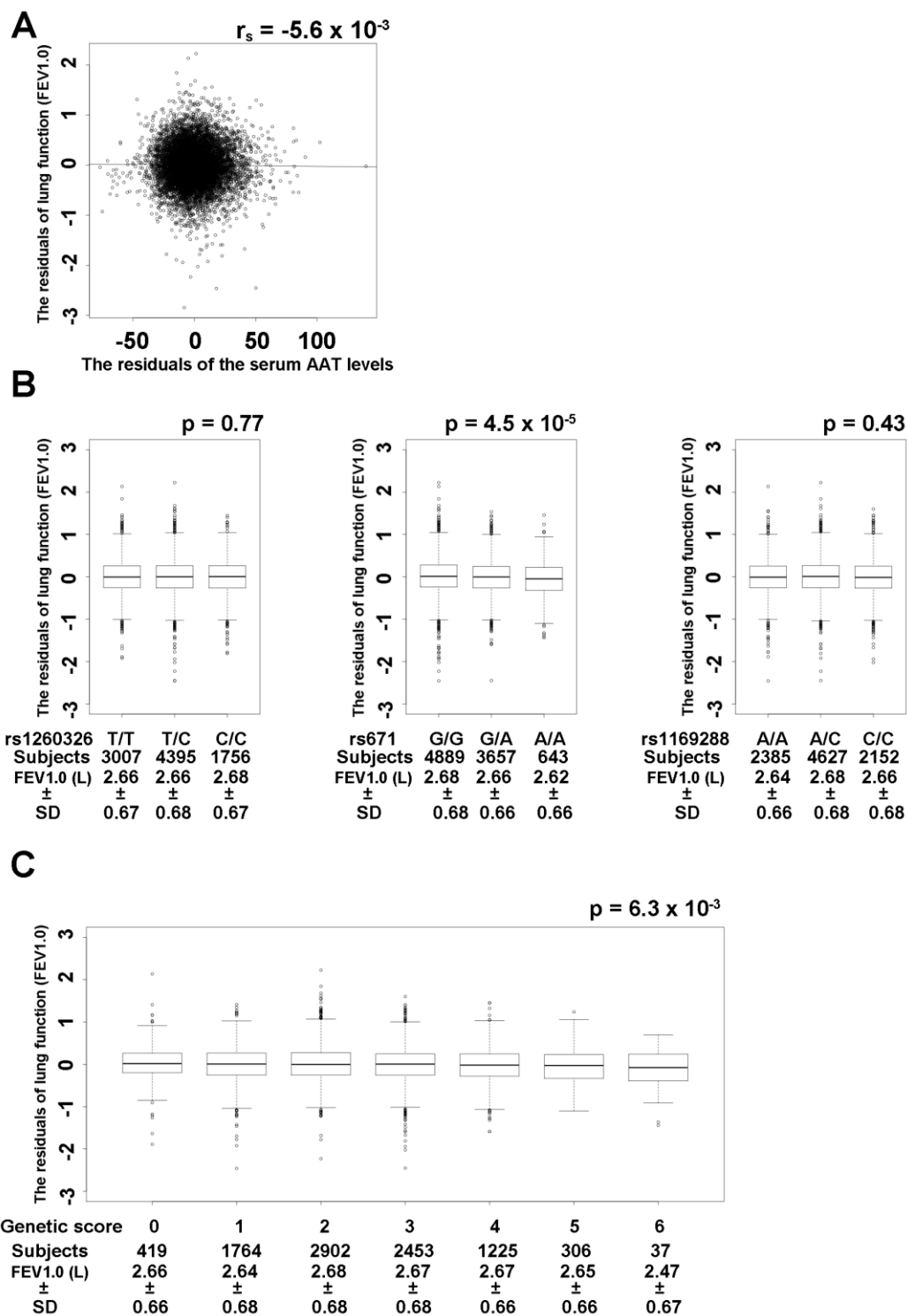
The data is based on the data of the Human Genetic Variation Browser.

P: p-value calculated by Student's t-distribution test. r_s : Spearman's rank sum coefficient.



Supplementary Figure 9 | Interaction between smoking and rs671

Interactive effect sizes between smoking and rs671 are shown in the combined study, GWAS and replications 1 and 2. Error bars indicate standard errors. G0: non-smokers, GQ1-4: smokers quartile according to Brinkman index.



Supplementary Figure 10 | Associations between lung function and serum AAT levels or the three variants associated with serum AAT levels

A) The lack of association between AAT and FEV1.0 in the current study is indicated.

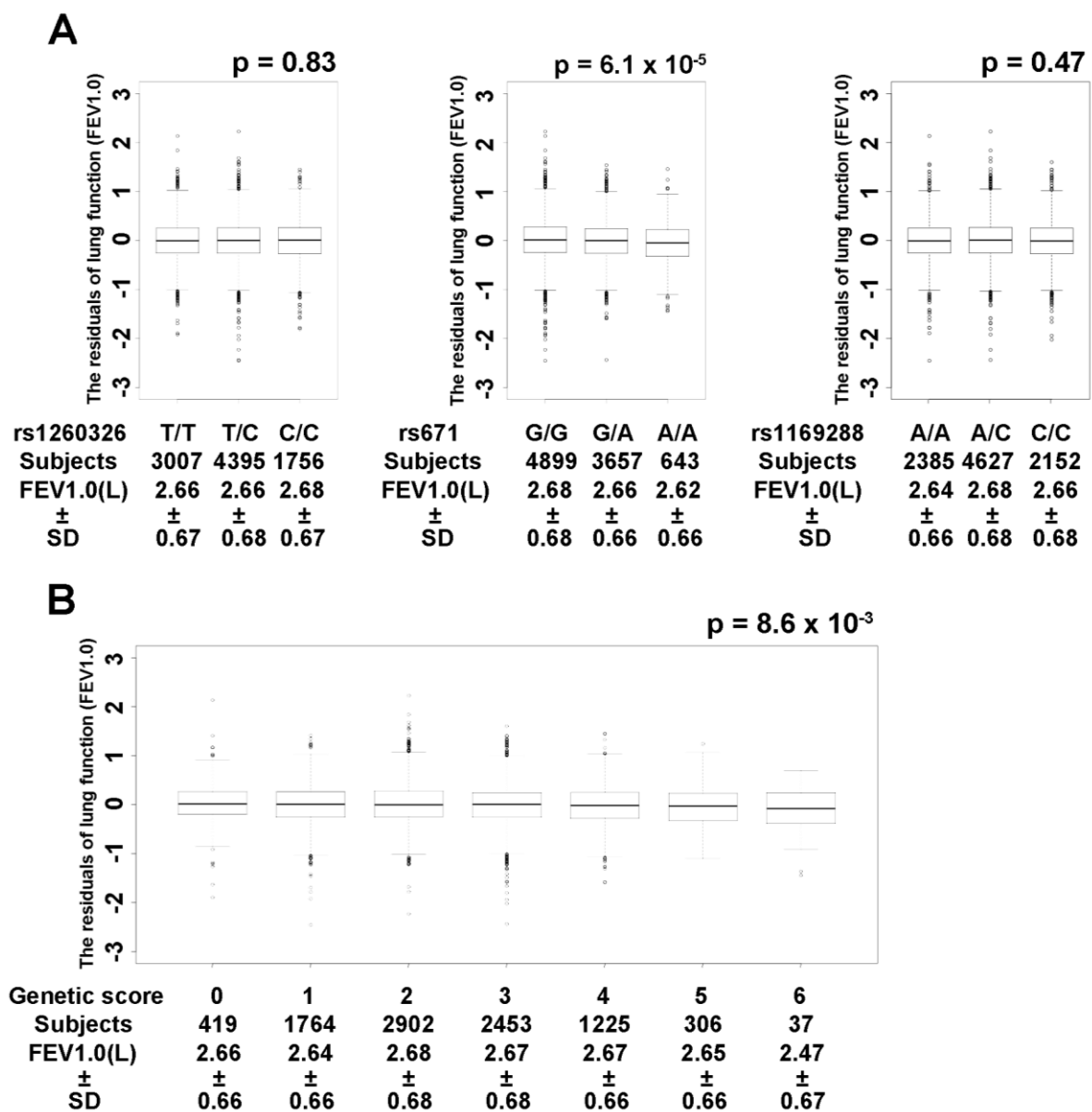
1 The associations between AAT and the three SNPs (B) or the number of risk alleles of
2 the three SNPs(C) are indicated.

3 r_s : Spearman's rank sum coefficient. P: p-value calculated by generalized linear
4 regression model.

5 The upper and lower boundaries of the boxes represent the third and first quartiles,
6 respectively. The horizontal lines represent the medians. The whiskers represent more or
7 less than 1.5 times of the interquartile range.

8

9



Supplementary Figure 11 | Associations between lung function and the three variants after conditioning on serum AAT levels

The associations between AAT and the three SNPs(A) or the number of risk alleles of the three SNPs(B) after conditioning on AAT levels are indicated.

P: p-value calculated by generalized linear regression model.

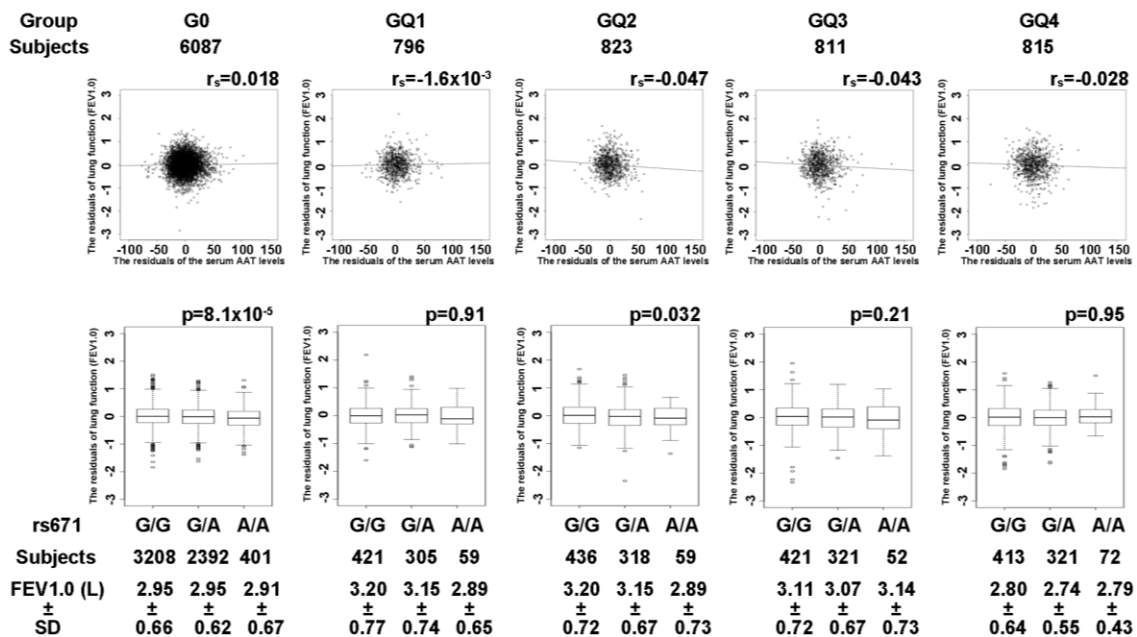
The upper and lower boundaries of the boxes represent the third and first quartiles,

1 respectively. The horizontal lines represent the medians. The whiskers represent more or

2 less than 1.5 times of the interquartile range.

3

4



Supplementary Figure 12 | An association between rs671 and FEV1.0 according to smoking status

The association between rs671 and FEV1.0 is shown based on the subgroups according to smoking status. G0: non-smokers, GQ1-4: smokers quartile according to Brinkman Index.

P: p-value calculated by generalized linear regression model. r_s : Spearman's rank sum coefficient.

The upper and lower boundaries of the boxes represent the third and first quartiles, respectively. The horizontal lines represent the medians. The whiskers represent more or less than 1.5 times of the interquartile range.

1 Supplementary Table 1 | The SNPs showing p-values less than 1.0×10^{-6} in the

2 GWAS

SNP	CHR	Position	Gene	Ref/Var	Location	VAF	RSQ	HWE-p	Effect (SE)	P
rs1260326	2	27730940	GCKR	T/C	nonsyn	0.425	0.997	1.90×10^{-3}	2.15 (0.43)	7.07×10^{-7}
rs7316287	12	111321512	CCDC63	T/C	intronic	0.220	0.925	0.777	2.62 (0.53)	8.12×10^{-7}
rs76024719	12	111321724	CCDC63	G/A	intronic	0.220	0.92	0.707	2.64 (0.53)	7.60×10^{-7}
rs202051532	12	111322994	CCDC63	AAGAG/A	intronic	0.218	0.895	0.671	2.63 (0.53)	8.53×10^{-7}
rs7311323	12	111323939	CCDC63	G/A	intronic	0.781	0.883	0.328	-2.69 (0.54)	5.43×10^{-7}
rs10849914	12	111325437	CCDC63	T/A	intronic	0.240	0.859	0.540	2.55 (0.52)	8.09×10^{-7}
rs192395428	12	111328450	CCDC63	C/T	intronic	0.237	0.909	0.458	2.58 (0.52)	7.30×10^{-7}
rs11065749	12	111329227	CCDC63	G/A	intronic	0.237	0.916	0.444	2.57 (0.52)	7.60×10^{-7}
rs139183566	12	111330204	CCDC63	A/AC	intronic	0.237	0.914	0.444	2.57 (0.52)	7.60×10^{-7}
rs11065750	12	111331016	CCDC63	G/A	intronic	0.237	0.936	0.444	2.57 (0.52)	7.60×10^{-7}
rs11065751	12	111331156	CCDC63	T/A	intronic	0.237	0.937	0.444	2.57 (0.52)	7.60×10^{-7}
rs11065752	12	111331165	CCDC63	G/A	intronic	0.237	0.938	0.490	2.57 (0.52)	7.78×10^{-7}
rs149021839	12	111333947	CCDC63	AATTT/A	intronic	0.217	0.789	0.211	2.74 (0.54)	3.79×10^{-7}
rs75295329	12	111344621	CCDC63	G/T	intronic	0.206	0.697	0.082	2.81 (0.55)	3.45×10^{-7}
rs12227162	12	111367244	MYL2/LOC1000131138	C/T	intergenic	0.208	0.641	0.015	3.29 (0.55)	2.64×10^{-9}
rs145558108	12	111377776	LOC100131138/CUX2	C/CT	intergenic	0.265	0.579	1.37×10^{-5}	2.59 (0.52)	5.89×10^{-7}
rs2188380	12	111386127	LOC100131138/CUX2	T/C	intergenic	0.207	0.702	0.032	3.21 (0.55)	6.05×10^{-9}
rs149607519	12	111389437	LOC100131138/CUX2	C/G	intergenic	0.205	0.675	0.050	3.20 (0.55)	7.36×10^{-9}
rs148177611	12	111390454	LOC100131138/CUX2	TAGAA/T	intergenic	0.205	0.681	0.055	3.22 (0.55)	5.92×10^{-9}
rs3809297	12	111609727	CUX2	G/T	intronic	0.261	0.546	0.047	3.28 (0.51)	1.13×10^{-10}
rs2339717	12	111696528	CUX2	T/C	intronic	0.626	0.976	0.600	-2.67 (0.45)	4.94×10^{-9}
rs6490029	12	111698457	CUX2	G/A	intronic	0.626	0.993	0.581	-2.66 (0.45)	5.57×10^{-9}
rs916682	12	111699146	CUX2	A/G	intronic	0.626	0.997	0.531	-2.65 (0.46)	6.13×10^{-9}
rs3858704	12	111705893	CUX2	A/G	intronic	0.654	0.987	0.908	-2.75 (0.46)	2.71×10^{-9}
rs4766566	12	111706877	CUX2	C/T	intronic	0.654	0.996	0.908	-2.75 (0.46)	2.71×10^{-9}
rs79105258	12	111718231	CUX2	C/A	intronic	0.254	0.721	0.770	2.96 (0.50)	4.72×10^{-9}
rs3782886	12	112110489	BRAP	T/C	syn	0.286	0.995	0.767	3.02 (0.48)	4.98×10^{-10}
rs11066001	12	112119171	BRAP	T/C	intronic	0.286	0.864	0.767	3.02 (0.48)	4.98×10^{-10}
rs11066008	12	112140669	ACAD10	A/G	intronic	0.288	0.523	0.769	2.96 (0.48)	8.96×10^{-10}
rs11066015	12	112168009	ACAD10	G/A	intronic	0.270	0.632	0.814	3.27 (0.49)	4.17×10^{-11}
rs4646776	12	112230019	ALDH2	G/C	intronic	0.269	0.841	0.783	3.31 (0.49)	2.49×10^{-11}
rs671	12	112241766	ALDH2	G/A	nonsyn	0.270	0.997	0.814	3.28 (0.49)	3.38×10^{-11}
rs11068574	12	118053982	KSR2	A/G	intronic	0.651	0.995	0.423	-2.28 (0.46)	9.77×10^{-7}
rs1169289	12	121416622	HNF1A	C/G	syn	0.454	0.898	0.705	2.16 (0.44)	9.60×10^{-7}
rs1169288	12	121416650	HNF1A	A/C	nonsyn	0.490	0.703	0.077	2.40 (0.45)	8.45×10^{-8}
rs2244608	12	121416988	HNF1A	A/G	intronic	0.471	0.872	0.636	2.30 (0.44)	1.74×10^{-7}
rs63470963	12	121417536	HNF1A	G/GACTC	intronic	0.465	0.856	0.424	2.19 (0.44)	6.02×10^{-7}
rs1169284	12	121419926	HNF1A	T/C	intronic	0.472	0.939	0.366	2.32 (0.44)	1.24×10^{-7}
rs7979473	12	121420260	HNF1A	A/G	intronic	0.527	0.932	0.358	-2.31 (0.44)	1.31×10^{-7}
rs7979478	12	121420263	HNF1A	A/G	intronic	0.527	0.943	0.358	-2.31 (0.44)	1.31×10^{-7}
rs1183910	12	121420807	HNF1A	G/A	intronic	0.472	0.954	0.404	2.30 (0.44)	1.59×10^{-7}
rs11065384	12	121423285	HNF1A	T/C	intronic	0.528	0.956	0.509	-2.29 (0.44)	1.81×10^{-7}
rs7970695	12	121423376	HNF1A	G/A	intronic	0.527	0.968	0.501	-2.30 (0.44)	1.58×10^{-7}
rs11065385	12	121423386	HNF1A	A/G	intronic	0.528	0.960	0.509	-2.29 (0.44)	1.81×10^{-7}
rs9738226	12	121423659	HNF1A	A/G	intronic	0.521	0.949	0.442	-2.22 (0.44)	4.35×10^{-7}
rs2393791	12	121423956	HNF1A	C/T	intronic	0.528	0.981	0.384	-2.28 (0.44)	2.06×10^{-7}

rs2393776	12	121424406	HNF1A	G/A	intronic	0.528	0.985	0.384	-2.28 (0.44)	2.06 x 10 ⁻⁷
rs2243458	12	121424490	HNF1A	C/T	intronic	0.472	0.985	0.384	2.28 (0.44)	2.06 x 10 ⁻⁷
rs2393775	12	121424574	HNF1A	G/A	intronic	0.528	0.994	0.384	-2.28 (0.44)	2.06 x 10 ⁻⁷
rs7310409	12	121424861	HNF1A	A/G	intronic	0.528	0.999	0.384	-2.28 (0.44)	2.06 x 10 ⁻⁷
rs1169292	12	121426478	HNF1A	C/T	intronic	0.478	0.949	0.466	2.26 (0.44)	2.51 x 10 ⁻⁷
rs1169294	12	121426594	HNF1A	G/A	intronic	0.479	0.946	0.364	2.28 (0.44)	1.90 x 10 ⁻⁷

1 CHR: chromosome, Ref/Var: Reference allele/Variant allele, VAF: Variant allele
2 frequency, HWE: Hardy–Weinberg equilibrium, SE: standard error, syn/nonsyn:
3 synonymous/nonsynonymous. RSQ indicates a quality of imputation. P: p-value
4 calculated by generalized linear regression model.

5 The applied reference panel: NCBI build 37.

6

7

8

9

10

11

12

13

1 **Supplementary Table 2 | Variances of AAT levels explained by the three SNPs**

SNP	Variance explained by SNPs (%)
rs1260326	0.59
rs671	0.90
rs1169288	0.46
Total	1.95

1 **Supplementary Table 3 | Functional annotation for the associated SNPs and those in high LD ($r^2>0.8$)**

SNP	CHR	Position	Gene	Location	r ²		Ref /Var	VAF		P (GWAS)	Promoter		Enhancer			DNase	Proteins bound by ChIP	
					Nagahama (Index SNP)	European		NAG	EUR		Active promoter	Weak promoter	Strong enhancer	Weak enhancer	others		Cell_type	Protein
rs1260326	2	27730940	GCKR	Nonsyn	-	-	T / C	0.43	0.59	7.1 x 10 ⁻⁷					LIV.A (Roadmap, TxEnhG1)	HepG2		
rs6547692	2	27734972	GCKR	Intronic	0.928 (rs1260326)	0.80	G / A	0.43	0.55	2.3 x 10 ⁻⁶								
rs780096	2	27741072	GCKR	Intronic	0.914 (rs1260326)	0.81	C / G	0.43	0.56	1.6 x 10 ⁻⁶			HepG2				HepG2	RXRA
rs780095	2	27741105	GCKR	Intronic	0.914 (rs1260326)	0.81	A / G	0.43	0.56	1.6 x 10 ⁻⁶			HepG2	K562			HepG2	FOXA2 , RXRA
rs780094	2	27741237	GCKR	Intronic	0.914 (rs1260326)	0.91	T / C	0.43	0.59	1.6 x 10 ⁻⁶			HepG2	K562			HepG2	FOXA2, MAFK, RXRA
rs780093	2	27742603	GCKR	Intronic	0.903 (rs1260326)	0.91	T / C	0.43	0.59	2.5 x 10 ⁻⁶					LIV.A (Roadmap, TxEnhG1)	GM1289, CD4_Th0_A		
rs1313566	2	27748904	GCKR/ C2orf16	intergenic	0.901 (rs1260326)	0.81	G / A	0.43	0.56	2.2 x 10 ⁻⁶								
rs1260334	2	27748597	GCKR/ C2orf16	intergenic	0.900 (rs1260326)	0.72	C / A	0.43	0.57	2.9 x 10 ⁻⁶								
rs1260333	2	27748624	GCKR/ C2orf16	intergenic	0.898 (rs1260326)	0.80	A / G	0.43	0.55	1.8 x 10 ⁻⁶								
rs199682409	2	27750543	GCKR/ C2orf16	intergenic	0.892 (rs1260326)	0.55	A / AAG	0.42	0.50	1.8 x 10 ⁻⁶				K562				
rs200747666	2	27750544	GCKR/ C2orf16	intergenic	0.891 (rs1260326)	0.55	A / AG	0.42	0.50	1.6 x 10 ⁻⁶				K562				
rs2950835	2	27750545	GCKR/ C2orf16	intergenic	0.891 (rs1260326)	0.55	A / G	0.42	0.50	1.6 x 10 ⁻⁶				K562				
rs11127048	2	27752463	GCKR/ C2orf16	intergenic	0.891 (rs1260326)	0.72	G / A	0.42	0.57	1.1 x 10 ⁻⁶								
rs2911711	2	27750546	GCKR/ C2orf16	intergenic	0.890 (rs1260326)	0.55	T / A	0.42	0.50	2.1 x 10 ⁻⁶				K562				
rs6753534	2	27752871	GCKR/ C2orf16	intergenic	0.831 (rs1260326)	0.79	C / T	0.42	0.56	3.4 x 10 ⁻⁶								

SNP	CHR	Position	Gene	Location	r ²		Ref /Var	VAF		P (GWAS)	Promoter			Enhancer		DNase	Proteins bound by ChIP	
					Nagahama (Index SNP)	European		NAG	EUR		Active promoter	Weak promoter	Strong enhancer	Weak enhancer	others		Cell_type	Protein
rs671	12	112241766	ALDH2	nonsyn	-	-	G / A	0.27	0	3.4 x 10 ⁻¹¹						Chorion, GM19239, LNCaP (AT), CD4_Th0_A		
rs4646776	12	112230019	ALDH2	intronic	0.998 (rs671)	-	G / C	0.27	0	2.5 x 10 ⁻¹¹			K562		LIV.A (Roadmap, TxEnhG1), IMR90 (Roadmap, TxEnhG1)	HepG2, Caco-2		
rs11066015	12	112168009	ACAD10	intronic	0.995 (rs671)	-	G / A	0.27	0	4.2x 10 ⁻¹¹						GM19239		
rs3782886	12	112110489	BRAP	syn	0.916 (rs671)	-	T / C	0.29	0	5.0 x 10 ⁻¹⁰								
rs11066001	12	112119171	BRAP	intronic	0.916 (rs671)	-	T / C	0.29	0.0026	5.0 x 10 ⁻¹⁰								
rs11066008	12	112140669	ACAD10	intronic	0.911 (rs671)	-	A / G	0.29	0	9.0 x 10 ⁻¹⁰								
rs1169288	12	121416650	HNFI1A	nonsyn	-	-	A / C	0.49	0.33	8.5 x 10 ⁻⁸	HepG2, LIV.A (Roadmap)					LNCaP, iPS	HepG2	POL2, TAF1
rs1169289	12	121416622	HNFI1A	syn	0.846 (rs1169288)	0.57	C / G,T	0.45	0.46	9.6 x 10 ⁻⁷	HepG2, LIV.A (Roadmap)					LNCaP, iPS	HepG2	POL2, TAF1
rs2244608	12	121416988	HNFI1A	intronic	0.804 (rs1169288)	0.95	A / G	0.47	0.34	1.7 x 10 ⁻⁷	HepG2, LIV.A (Roadmap)							

1

2

3

4

5

CHR: chromosome, syn/nonsyn: synonymous/nonsynonymous. Ref/Var: Reference allele/Variant allele, VAF: Variant allele frequency,
 NAG: Nagahama population, EUR: European, P: p-value calculated by generalized linear regression model. LIV.A: Adult Liver, HepG2:
 Cell line (Hepatocellular carcinoma), K562: Cell line (Chronic myelogenic leukemia), GM1289: B-lymphocyte lymphoblastoid,
 CD4_Th0_A: CD4+ cells isolated from human blood and enriched for Th0 populations (Adult), IMR90: Cell line (Fetal lung fibroblast),

1 Caco2: Cell line (Colorectal adenocarcinoma), Chorion: Chorion cells (Outermost of two fetal membranes), GM19239: B-lymphocyte
2 lymphoblastoid, LNCaP: Cell line (Prostate adenocarcinoma), AT: Androgen treated, iPS: Induced pluripotent stem cell derived from
3 skin fibroblast, TxEnhG1: Transcription Enhancer-like. The functional annotations shown above are results of ENCODE project. When
4 functional annotations in liver or lung-derived cells were found in Roadmap project, these are also indicated.
5 The applied reference panel: NCBI build 37.

6

7

1 **Supplementary Table 4 | Amino acid conservation for HNF1A and GCKR**

2 Rs1169288

Species	Amino acid sequence																				
Human	L	E	S	G	L	S	K	E	A	L	I	Q	A	L	G	E	P	G	P	Y	L
Rhesus	L	E	S	G	L	S	K	E	A	L	I	Q	A	L	G	E	P	E	P	Y	L
Mouse	L	E	S	G	L	S	K	E	A	L	I	Q	A	L	G	E	P	G	P	Y	L
Dog	L	H	S	G	L	S	K	E	A	L	I	Q	A	L	G	E	P	G	P	Y	L
Elephant	L	E	S	G	L	T	K	E	A	L	I	Q	A	L	G	E	P	G	P	Y	L
Opossum	L	E	S	G	L	T	K	E	T	L	I	R	A	L	G	E	T	R	P	Y	D
Chicken	L	E	S	G	L	T	K	E	T	L	I	K	A	L	S	E	A	E	P	Y	V
Zebrafish	L	D	S	G	V	T	K	D	V	L	L	Q	A	L	E	D	L	D	P	S	P

3 Rs1260326

Species	Amino acid sequence																				
Human	L	A	H	S	T	V	G	Q	T	L	L	I	P	L	K	K	L	F	P	S	I
Rhesus	L	A	H	S	T	V	G	Q	N	L	P	I	P	L	K	K	L	F	P	S	V
Mouse	L	V	H	S	T	V	G	Q	S	L	P	A	P	L	K	K	L	F	P	S	L
Dog	L	A	H	S	T	V	G	Q	S	L	P	T	P	L	K	K	L	F	P	S	I
Elephant	L	A	H	S	T	V	G	Q	S	L	P	T	L	L	K	K	L	F	P	S	I
Opossum	L	A	H	S	T	V	G	Q	Y	L	P	S	P	L	K	K	L	F	P	S	I
Chicken	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Zebrafish	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-

4 Data source: UCSC Genome Browser on Human Feb. 2009 (GRCh37/hg19) Assembly

5 -: No homologous sequences

6

Supplementary Table 5 | Interactions between rs671 and covariates on serum AAT levels

Interaction	Effect (SE)	P
rs671 x Sex	-1.53 (0.60)	0.011
rs671 x Age	-0.019 (0.021)	0.38
rs671 x BMI	-0.058 (0.085)	0.50
rs671 x log (hs-CRP)	0.22 (0.54)	0.68
rs671 x Alcohol intake	1.59 (0.43)	2.2×10^{-4}
rs671 x Brinkman index	2.9×10^{-3} (8.7×10^{-4})	6.8×10^{-4}

P: p-value calculated by generalized linear regression model.

Supplementary Table 6 | The lack of interactions between the two genetic variants and alcohol consumption or smoking on serum AAT levels

Interaction	Effect (SE)	P
rs1260326 x Alcohol intake	0.21 (0.30)	0.48
rs1169288 x Alcohol intake	-0.87 (0.30)	0.77
rs1260326 x Brinkman index	-3.8×10^{-3} (7.8×10^{-4})	0.62
rs1169288 x Brinkman index	2.6×10^{-4} (7.8×10^{-4})	0.74

P: p-value calculated by generalized linear regression model.

1 **Supplementary Table 7 | The number of subjects according to GWAS arrays**

Platform	Sample number
Hap610K	1,718
2.5M-4	1,024
2.5M-8	186
Hap610K + 2.5M-4 + CoreExome + 2.5M-s	112
2.5M- 4 + 2.5M-s + Exome	480
2.5M-8 + 2.5M-s + Exome	192

2 Hap610K: human hap610K quad array, 2.5M-4: human omni 2.5M -4 array, 2.5M-8:
3 human omni 2.5M -8 array, 2.5M-s: human omni 2.5s array, Exome: human exome,
4 CoreExome: human core exome (Illumina, San Diego, CA, USA).

5
6
7
8
9

Supplementary Note

Selection of the samples in the current study

Samples for genome-scanning in the Nagahama Study were selected from the DNA samples of the participants recruited in 2008 or 2009. Since high quality of DNA samples is an essential factor to obtain excellent GWAS results especially in terms of success rate, we put strict quality control for sample selection of GWAS. As a result, we excluded 2,405 subjects from samples that showed any signs of fragmented DNA. We performed genome-scanning using the remaining subjects (Supplementary Fig. 1 and Supplementary Table 7). The 4,023 DNA samples from the participants recruited in 2010 were subjected for replication study 1. Since we used Taqman assay in the replication studies which does not always require excellent quality of DNA, the 2,405 subjects that were not used for GWAS were subjected for replication study 2. Therefore, 2,402 subjects out of the 2,405 whose smoking information were available were genotyped for replication 2 (Supplementary Fig. 1). We used a total of 9,359 subjects for the current study to maximize the power to detect significant variations.