ELECTRONIC LETTER

CLCA1 gene polymorphisms in chronic obstructive pulmonary disease

A E Hegab, T Sakamoto, Y Uchida, A Nomura, Y Ishii, Y Morishima, M Mochizuki, T Kimura, W Saitoh, H H Massoud, H M Massoud, K M Hassanein, K Sekizawa

J Med Genet 2004;**41**:e27 (http://www.jmedgenet.com/cgi/content/full/41/3/e27). doi: 10.1136/jmg.2003.012484

Chronic obstructive pulmonary disease (COPD), including chronic bronchitis and emphysema, is a major cause of morbidity and mortality in many countries. The most important risk factor for COPD is cigarette smoking, and nearly 90% of COPD patients are smokers.[1] However, only 15% of cigarette smokers develop clinically significant COPD.[2] COPD is known to aggregate in families,[3] and twin studies have shown that obstructive airway disease correlates directly with genetic similarities.[4] In contrast, there are differences in the prevalence of COPD between different ethnic groups.[5] All these results support the notion that genetic factors are involved in the pathogenesis of COPD. Recent segregation analysis has demonstrated that the genetic components determining forced expiratory volume in 1 second (FEV_1) consist of multiple genes, each of which has a small influence, rather than a single Mendelian gene.[6] An intensive search has been ongoing to find the genetic factors responsible for COPD development. Until now, more than 20 polymorphisms of candidate genes have been reported to have an association with COPD. They include genes for proteinases,[7] anti-proteinases,[8–10] anti-oxidants,[11] xenobiotic metabolising enzymes,[12 13] and inflammatory mediators.[14 15] For most of these loci, however, there have been contradictory results.[16] Although genetic heterogeneity among different ethnic groups under different environmental conditions could explain this inconsistency, it is still necessary to confirm the associations of the polymorphisms in different populations.

Recently the human *CLCA1* gene and its murine counterpart, m*CLCA3*, have been isolated.[17 18] To date, the human *CLCA* family consists of four homologous genes (*hCLCA1*, *hCLCA2*, *hCLCA3*, and *hCLCA4*), all clustered on the short arm of chromosome 1 (1p 22–31).[19] The region around 120 cM on chromosome 1 has been identified to show moderate linkage to FEV_1/forced vital capacity (FVC) in two independent genome scans for COPD.[20 21] In addition, the expression of *hCLCA1* and *mCLCA3* is strongly induced in the airway epithelium, especially in goblet cells, under asthmatic conditions, while they are not expressed in the normal lung.[22–25] Furthermore, *in vitro* transfection studies have demonstrated that both genes play a direct role in mucous production.[23 25] Mucous hypersecretion is a pathophysiological feature of both COPD and bronchial asthma. Therefore, we hypothesised that polymorphisms of the *CLCA1* gene would be related to the susceptibility to COPD.

In this study, we conducted a case–control study to assess the association of *CLCA1* gene polymorphisms with COPD in Japanese and Egyptian populations.

Key points

- *CLCA1* has been shown to regulate airway mucous production in inflammatory conditions. We hypothesised that polymorphisms of the *CLCA1* gene could have an important role in the pathogenesis of chronic obstructive pulmonary disease (COPD). We sought to identify the *CLCA1* gene polymorphisms and performed a case–control association study to evaluate the involvement of these variants in COPD in Japanese and Egyptian populations.
- We identified 22 novel single nucleotide polymorphisms. There was a significant difference in +5080 T/C genotypes between the patient and the control groups in the Egyptian population ($p = 0.024$). In the Japanese population, the distribution of +13924 T/A allele frequencies was significantly different between the patient and the control groups ($p = 0.042$). However, becuse multiple comparisons were made, these associations may represent type 1 error.
- In the Japanese population, the frequency of the haplotype +126 T: +13924 T: +25133 C: +31384 C was significantly higher in the COPD patients than in the controls ($p_{corr} = 0.0002$). In contrast, the frequencies of the haplotype +126 T: +13924 A: +25133 C: +31384 C and the haplotype +126 G: +13924 T: +25133 C: +31384 T were significantly higher in the controls than in the COPD patients ($p_{corr} = 0.0017$ and 0.0001, respectively) in the same population.
- These polymorphisms and haplotypes may be involved in the pathogenesis of COPD and may be useful for predicting the susceptibility to COPD.

METHODS

Subjects

The investigation is a case–control association study in two different ethnic groups, Japanese and Egyptian. The basic inclusion criterion was chronic heavy smoking. Clinically based ascertainment of subjects was used, based on past history, physical examination, and spirometric data (FEV_1/FVC ratio of <70%, according to the Global Initiative for Chronic Obstructive Lung Disease criteria[26]). The predicted values of pulmonary function were determined according to each subject's data (gender, age, weight, height, and ethnicity), using a ChestGraph Jr HI-101 spirometer (Chest, Tokyo, Japan) for Japanese subjects and a MultiSpiro-SX Spirometer (MultiSpiro Inc, CA, USA) for Egyptian subjects.[27] Controls were age matched healthy smokers. For both

Abbreviations: COPD, chronic obstructive pulmonary disease; FCV, forced vital capacity; FEV_1, forced expiratory volume in 1 second; RFLP, restriction fragment length polymorphism; SNPs, single nucleotide polymorphisms; SSCP, single strand conformational polymorphism

groups, those with borderline pulmonary function, with low Brinkman's index (the number of cigarettes/day × the number of years), or with other significant respiratory diseases such as bronchial asthma, bronchiectasis, and pulmonary tuberculosis were excluded. Furthermore, ethnic and geographic matching was considered to eliminate possible effect of population stratification. The Japanese subjects (88 COPD patients and 40 controls) were recruited from Tsukuba University Hospital. Most of the Japanese subjects were the same as those in a previous study.[10] The Egyptian subjects (106 COPD patients and 72 controls) were recruited from the Department of Chest Diseases and Tuberculosis at Cairo University Hospital and affiliated hospitals. Written informed consent was obtained from all the subjects, and the study was approved by the ethics committees of the hospitals involved.

Screening the *CLCA1* gene for polymorphisms

The structure of *CLCA1* has been reported previously (Genbank accession no. AF039401).[17] We studied the fifteen exons with intronic junctions and the promoter region, 1617 bp upstream from the transcription start site. Twenty three pairs of primers were synthesised to make PCR DNA fragments covering these regions (table 1). The nucleotide positions in this study are given relative to the transcription start site. Genomic DNA was extracted from whole blood using a Qiagen DNA blood kit (Qiagen, Hilden, Germany). For 20 patients and 20 controls in each ethnic group, we screened all the PCR fragments for polymorphisms by single strand conformational polymorphism (SSCP) as previously described.[10] Following this, for samples showing variant band patterns, DNA sequencing was performed to locate the polymorphisms and design a method of genotyping them in the whole populations. Even when none of the screened samples had shown variant band pattern in SSCP analysis, we still performed DNA sequencing for eight samples randomly chosen from each ethnic group, to reduce the possibility of missing a common polymorphism that could not be detected by SSCP and to confirm the sequence of that portion by comparing it with the reference sequence.

Genotyping

It has been reported that SSCP has an average sensitivity of only 85–95% in detecting polymorphisms.[28] Therefore, when a polymorphism was detected in the screened subjects, we also confirmed its incidence in those subjects by using restriction fragment length polymorphism (RFLP), *Taq*Man allelic discrimination,[29] or direct sequencing. As rare polymorphisms are less likely to play a role in disease pathogenesis and need a very large number of subjects to show enough power, we

Table 1 Oligonucleotide primers used for screening the *CLCA1* gene polymorphisms

Name	Sequence	Position	Size (bp)	AT (℃)
CL.pro1.F	5′-AGAGTGAATCTGACCAGCTC-3′	−1614 to −1595	328	61
CL.pro1.R	5′-GAGAGAGGAAGGAAAAGCGT-3′	−1288 to −1307		
CL.pro2.F	5′-TTTCTAAGCGACTGCCGAAAG-3	−1394 to −1374	354	58
CL.pro2.R	5′-CCCTAAAAAACAGAGCCTA-3′	−1042 to −1062		
CL.pro3.F	5′-AACTCTCAAACTCTTAGCAC-3′	−1094 to −1073	230	58
CL.pro3.R	5′-GCAGTTAAAAGGGAAACACT-3′	−866 to −885		
CL.pro4.F	5′-TGGATCATCTGGACAACCAC-3′	−933 to −914	319	61
CL.pro4.R	5′-GCAATTGTGAATGGGAGTTC-3′	−616 to −635		
CL.pro5.F	5′-ACAGACAAACAGAGAGCC-3′	−663 to −646	350	48
CL.pro5.R	5′-TTGGTCTTCAGATGGTGA-3′	−315 to −332		
CL.pro6.F	5′-CAATGCCTTTCTTCACAGA-3′	−402 to −384	412	58
CL.pro6.R	5′-GCATTAGCAAGTGATGTAGA-3′	+9 to −11		
CL.ex1.F	5′-TGAGGAACTTGGAGTATGC-3′	−88 to −70	217	59
CL.ex1.R	5′-CTCCTGGTGGTCTTTACC-3′	+129 to +112		
CL.ex2A.F	5′-GTGCTTCCAAGGAACCGGATG-3′	+300 to +320	291	63
CL.ex2A.R	5′-CTGTGATTCCTCTCTTTGG-3′	+590 to +572		
CL.ex2B.F	5′-ACCTTCGTAACCCGCATTT-3′	+552 to +570	282	58
CL.ex2B.R	5′-CTCATGTTCTCTCTATCAC-3′	+833 to +815		
CL.ex3.F	5′-TCAGTCAATTGTTACGTATG-3′	+5012 to +5031	221	55
CL.ex3.R	5′-CAGGAGAAAGTGGAATTT-3′	+5232 to +5215		
CL.ex4.F	5′-GAAACTTACTCCAGTACGA-3′	+5284 to +5302	256	55
CL.ex4.R	5′-AGAGTATGAGAAGTATGGT-3′	+5539 to +5521		
CL.ex5.F	5′-TCCAGTGTGAGAAGGTAA-3′	+8014 to +8031	203	55
CL.ex5.R	5′-GTTTGGAACAAGTTAAATCA-3′	+8216 to +8197		
CL.ex6.F	5′-CACCTAAACATCTAACCTTCC-3′	+13805 to +13825	232	57
CL.ex6.R	5′-GGGGTGAGAACAAGGTACTTA-3′	+14036 to +14016		
CL.ex7.F	5′-TTCTACCACATAGTTTACCATT-3′	+16924 to +16945	293	58
CL.ex7.R	5′-AAGACCAAGACTCATTGAACA-3′	+17216 to +17196		
CL.ex8.F	5′-CCATCTAAACGTAACCTGTTT-3′	+18124 to +18144	291	58
CL.ex8.R	5′-CAATAGAAAAACACATCTCAC-3′	+18414 to +18394		
CL.ex9.F	5′-TTTGGCCACCCTAAGTCTGA-3′	+20588 to +20607	266	63
CL.ex9.R	5′-GTAGGAGAGACAATACCGGG-3′	+20834 to +20853		
CL.ex10.F	5′-GCTCTCTAAACAACAGCTAA-3′	+22827 to +22846	243	59
CL.ex10.R	5′-CTCCATCTTGGCTATGATTA-3′	+23069 to +23050		
CL.ex11.F	5′-CTCCAGAAAGTAAGAGCTG-3′	+24995 to +25014	292	54
CL.ex11.R	5′-TTATCTTGTTGAATTATGAGA-3′	+25286 to +25266		
CL.ex12.F	5′-GAAGTTATTCATTGGAAATGT-3′	+25790 to +25810	335	52
CL.ex12.R	5′-AAATAAATATCTTCCAATTTC-3′	+26124 to +26104		
CL.ex13.F	5′-CCCGTTACCTATTTATTAATTC-3′	+27110 to +27131	247	58
CL.ex13.R	5′-TTTGCACAAGCCAGGTATGTT-3′	+28953 to +28974		
CL.ex14.F	5′-ACCATTTAGTGATGTTGCTTA-3′	+30174 to +30194	337	58
CL.ex14.R	5′-TCTACTTGGCTGGTAAATCT-3′	+30510 to +30491		
CL.ex15A.F	5′-CAGAATTAAAGTCACATTCTT-3′	+31266 to +31286	345	56
CL.ex15A.R	5′-GAATATTAGGACAAGGAGC-3′	+31610 to +31592		
CL.ex15B.F	5′-CAGAGACACCTAGTCCCG-3′	+31542 to +31560	383	56
CL.ex15B.R	5′-TAAATTTTTGGGTTTGAATGG-3′	+31945 to +31925		

AT, annealing temperature; bp, base pairs.

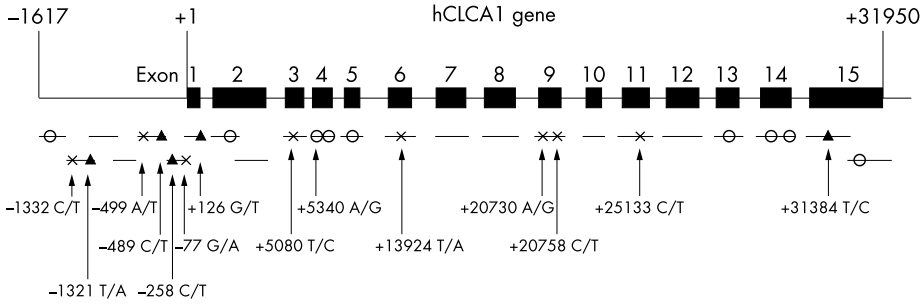

Figure 1 Schematic representation of the *CLCA1* gene showing the genomic structure and the location of the SNPs identified in this study. Filled boxes represent 15 exons. Short bars below show PCR DNA fragments. Symbols represent the SNPs. Open circles, crosses, and closed triangles are the SNPs confirmed by DNA sequencing, *Taq*Man allelic discrimination, and RFLP, respectively.

genotyped all the subjects only for polymorphisms showing allele frequencies of >10% in the initial screening step.

22 SNPs found in this study, 5 were determined by RFLP, 8 by *Taq*Man allelic discrimination and 9 by direct sequencing (fig 1). The restriction enzymes used for RFLP were *Ear*I for −1321 T/A, *Aci*I for +31384 T/C (both New England Biolabs, Beverly, MA, USA), *Mbo*I for −489 C/T, *Mva*I for +126 G/T (both Takara, Shiga, Japan), and *Hae*III for −258 C/T (Toyobo, Osaka, Japan). The same primers used for PCR-SSCP analysis were used to amplify the corresponding fragments for RFLP. The digested DNA fragments were resolved using 3% agarose gel containing ethidium bromide. The SNPs found at −1332 C/T, −499 A/T, −77 G/A, +5080 T/C, +13924 T/A, +20730 A/G, +20758 C/T, and +25133 C/T were analysed using *Taq*Man allelic discrimination. For each SNP, a pair of primers flanking the SNP and a pair of oligonucleotide probes, one homologous to the wild type labelled with the *Taq*Man FAM probe, and another homologous to the mutant type labelled with the VIC probe, were designed and synthesised by Applied Biosystems (Foster City, CA, USA) (table 2). The PCR was carried out on 20 ng of genomic DNA in a 25 µl reaction containing 50 to 900 nmol/l of each forward and reverse primer, 50 to 200 nmol/l of each FAM and VIC probe, and 1× TaqMan Universal PCR Master Mix (Applied Biosystems). PCR cycling conditions in the ABI PRISM 7000 (Applied Biosystems) were as follows: 50°C for 2 minutes and 95°C for 10 minutes, followed by 40 cycles of 95°C for 15 seconds and 60°C for 1 minute. The allelic discrimination was determined by the fluorescence intensity of FAM and VIC.

Table 2 Primers and probes used for *Taq*Man allelic discrimination

Target SNP		Primers and probes
−1332 C/T		
Primer	F	5′-GGATGACACAGCCAGGAAGTCT-3′
	R	5′-GCGTTTGGCTCCTTCTCATCT-3′
Probe		5′-[VIC]CTACAGGTGATCATC[MGB]-3′
		5′-[FAM]CTACAGATGATCATCTC[MGB]-3′
−499 A/T		
Primer	F	5′-CACTGCTCAAGGAAAAGAGGAT-3′
	R	5′-AAGAAAGGCATTGGTAGCTTGATG-3′
Probe		5′-[VIC]CATTTCATGTTCATGGAT[MGB]-3′
		5′-[FAM]AACATTTCATGCTCATGG[MGB]-3′
−77 G/A		
Primer	F	5′-GCAATAGAGAAAGCAGAAATCTGATG-3′
	R	5′-GGAGAGTAAAGATAAAATGTATATCTTA CTTTGC-3′
Probe		5′-[VIC]CATACTCCAAGTTCC[MGB]-3′
		5′-[FAM]AGCATACTTCAAGTTC[MGB]-3′
+5080 T/C		
Primer	F	5′-ACATGGTGACCCAGGCATCT-3′
	R	5′-TTCAGGAATCAAAATGGCAACAT-3′
Probe		5′-[VIC]AGCTTCAAACAGATACA[MGB]-3′
		5′-[FAM]CTTCAAGCAGATACAG[MGB]-3′
+13924 T/A		
Primer	F	5′-GCTGTTACACCAAAAGATGCACAT-3′
	R	5′-CGGGATTGGAGAACAAACTCA-3′
Probe		5′-[VIC]AGAGTCCTGTAACTTT[MGB]-3′
		5′-[FAM]AGTCCTGTTACTTTATT[MGB]-3′
+20730 A/G		
Primer	F	5′-CAACACTATAAGTGGGTGCTTTAACG-3′
	R	5′-TGCAGAGGGCCCCAAAG-3′
Probe		5′-[VIC]AGGTCAAACAAAGTGG[MGB]-3′
		5′-[FAM] AGGTCAGACAAAGTG [MGB]-3′
+20758 C/T		
Primer	F	5′-CAAAGTGGTGCCATCATCCA-3′
	R	5′-TTGGACAGCTCCTCTAGTTCTTGAG-3′
Probe		5′-[VIC]ACAGTCGCTTTGGG[MGB]-3′
		5′-[FAM]ACAGTGGCTTTGGG[MGB]-3′
+25133 C/T		
Primer	F	5′-CACCGTGGGAAAGGACACTT-3′
	R	5′-GGATCCCAGAGAAGGATTTGG-3′
Probe		5′-[VIC]TGGACAACGCAGCC[MGB]-3′
		5′-[FAM]CTGGACAATGCAGC[MGB]-3′

MGB, minor groove binder.

Statistical analysis

Two sided Student's *t* test was used for checking significant differences in clinical data between the COPD patients and the control subjects in each ethnic group, with significance set at $p<0.01$. Hardy-Weinberg equilibrium was assessed using a goodness of fit χ^2 test for biallelic markers. The distribution of genotype and allele frequency was analysed with Fisher's exact test (2×3 and 2×2 tables), with significance set at $p<0.05$. Correction for multiple comparisons was performed by the Bonferroni method. Odds ratios (ORs) and 95% confidence intervals (CIs) were calculated to quantitatively assess the degree of association observed. The haplotype frequencies were estimated for the patients and the controls separately by the expectation–maximisation method using the SNPAlyze program (Dynacom, Mobara, Japan). Linkage disequilibrium between each pair of SNP loci was analysed using all the subjects including the patients and the controls with a likelihood ratio test. Lewontin's disequilibrium coefficient D′[30] was estimated from the haplotype frequencies. The distribution of haplotype frequency between the patient and control groups was analysed with Fisher's exact test.

RESULTS

Table 3 demonstrates the characteristics of subjects recruited in this study. All the subjects were heavy smokers. Age and Brinkman's index were not significantly different between patients and controls. All the patients in both ethnic groups had moderate to severe COPD.

Twenty two polymorphisms were identified as SNPs by PCR-SSCP analysis and DNA sequencing. All were novel; we registered them in the dbSNP database of the NCBI under accession numbers ss5607363 to ss5607365, ss6313704 to ss6313719, and ss7986610 to ss7986635. Eight of these 22 SNPs were found to be mutant homozygous in all the subjects examined in both Japanese and Egyptian populations compared with the reference sequence. They are: −1492 T/C (promoter), +582 G/A (exon 2), +5513 A/T (intron 4),

Table 3 Clinical features of study populations

	Japanese			Egyptians		
	COPD	Control	p	COPD	Control	p
Subjects (n)	88	40		106	72	
Sex, M/F (n)	85/3	40/0		106/0	72/0	
Age, years	66.9 (1.6)	72.9 (1.2)	NS	62.5 (0.9)	59.0 (1.0)	NS
Brinkman's index	1201.6 (70.1)	1071.1 (66.5	NS	1050.0 (60.3)	990.9 (69.6)	NS
FVC % pred	77.4 (2.2)	95.9 (3.8)	<0.001	56.3 (1.0)	92.3 (0.9)	<0.001
FEV_1 (L)	1.1 (0.1)	2.4 (0.1)	<0.001	0.9 (0.03)	2.8 (0.04)	<0.001
FEV_1 % pred	46.7 (1.5)	91.0 (2.3)	<0.001	30.3 (1.1)	85.9 (0.4)	<0.001
FEV_1/FVC, %	47.1 (1.3)	85.1 (1.6)	<0.001	44.3 (1.2)	78.3 (1.0)	<0.001

FVC, forced vital capacity; FEV1, forced expiratory volume in one second; NS, not significant. Brinkman's index: the number of cigarettes/day × the number of years.
Data are presented as the mean (SEM) unless otherwise indicated.

+8082 A/G (exon 5), +27216 G/A (exon 13), +30389 T/G (exon 14), +30472 -/C (intron 14), and +31579 C/T (exon 15). Two, +8082 A/G and +30389 T/G, are non-synonymous nucleotide substitutions. Accordingly, it seems likely that at least in Japanese and Egyptian populations the reference amino acid translation of the gene protein should be changed from lysine to arginine at amino acid number 152 and from asparagine to lysine at number 760. The SNP +582 G/A in exon 2 is located in the 5′ untranslated region. The SNPs +27216 G/A in exon 13 and +31579 C/T in exon 15 were coding but synonymous variants. Therefore, these SNPs have no influence on the reference amino acid sequence.

The remaining 14 SNPs are shown in fig 1. Six were identified in the promoter: −1332 C/T, −1321 T/A, −499 A/T, −489 C/T, −258 C/T, and −77 G/A. Six were in the exons: +5080 T/C (exon 3; non-synonymous, Phe/Leu), +13924 T/A (exon 6; synonymous), +20730 A/G (exon 9; non-synonymous, Lys/Arg), +20758 C/T (exon 9; synonymous), +25133 C/T (exon 11; non-synonymous, Thr/Met) and +31384 T/C (exon 15; synonymous). Two were in the introns: +126 G/T (intron 1) and +5340 A/G (intron 3). Only four SNPs were found to have >10% incidence in each population during the initial screening step, therefore they were genotyped for the whole population (table 4). The genotypic frequencies for the control populations were consistent with Hardy-Weinberg equilibrium. Of these four SNPs, +5080 T/C in exon 3 tended to show a significant difference between the patient group and the control group in Egyptians only ($p = 0.024$, $p_{corr} = 0.096$). Furthermore, the allele frequency for T in Egyptians was significantly higher in the COPD group than in the control group (94% *v* 86%; $p = 0.013$, $p_{corr} = 0.052$; OR = 2.69; 95% CI 1.27 to 5.69). The +13924 T/A polymorphism was found in 48% of Japanese patients, but only 34% of the control group ($p = 0.042$, $p_{corr} = 0.168$; OR = 1.79; 95% CI 1.03 to 3.11). Therefore, it is possible that Japanese COPD patients tend to have T at the +13924 locus but because of multiple comparisons, the possibility of type 1 error cannot be excluded. In Egyptians, the distributions of +13924 T/A genotypes and alleles were not significantly different between the patients and the controls. No other significant

Table 4 Genotype and allele frequencies of *CLCA1* gene polymorphisms in Japanese and Egyptians

	Japanese					Egyptian				
	Genotype			Allele		Genotype			Allele	
+126 locus										
	G/G	G/T	T/T	G	T	G/G	G/T	T/T	G	T
COPD	17 (19)	51 (58)	20 (23)	85 (48)	91 (52)	23 (22)	66 (62)	17 (16)	112 (53)	100 (47)
Control	11 (27)	25 (63)	4 (10)	47 (59)	33 (41)	21 (29)	40 (56)	11 (15)	82 (57)	62 (43)
p	NS			NS		NS			NS	
+13924 locus										
	T/T	T/A	A/A	T	A	T/T	T/A	A/A	T	A
COPD	17 (19)	50 (57)	21 (24)	84 (48)	92 (52)	18 (17)	55 (52)	33 (31)	91 (43)	121 (57)
Control	2 (5)	23 (57)	15 (38)	27 (34)	53 (66)	14 (19)	35 (49)	23 (32)	63 (44)	81 (56)
p	0.051 (p_{corr}, 0.204)			0.042 (p_{corr}, 0.168)		NS			NS	
+31384 locus										
	T/T	T/C	C/C	T	C	T/T	T/C	C/C	T	C
COPD	26 (31)	37 (44)	21 (25)	89 (53)	79 (47)	67 (67)	29 (29)	4 (4)	163 (82)	37 (18)
Control	10 (28)	21 (58)	5 (14)	41 (57)	31 (43)	45 (64)	20 (28)	6 (8)	110 (77)	32 (23)
p	NS			NS		NS			NS	
+25133 locus										
	C/C	C/T	T/T	C	T					
COPD	69 (78)	19 (22)	0 (0)	157 (89)	19 (11)					
Control	35 (87)	5 (13)	0 (0)	75 (94)	5 (6)					
p	NS			NS						
+5080 locus										
						T/T	T/C	C/C	T	C
COPD						94 (89)	12 (11)	0 (0)	200 (94)	12 (6)
Control						55 (76)	14 (20)	3 (4)	124 (86)	20 (14)
p						0.024 (p_{corr}, 0.094)			0.013 (p_{corr}, 0.052)	

Data presented as n (%). NS, not significant.

Table 5 Pairwise linkage disequilibrium (D′ above diagonal) and statistical significance (p value below diagonal) in the Japanese populations

	+126 G/T	+13924 T/A	+31384 T/C	+25133 C/T
+126 G/T		0.300	0.297	0.514
+13924 T/A	<0.001		0.116	0.259
+25133 C/T	0.016	0.298	0.833	
+31384 T/C	<0.001	0.115		<0.001

Table 6 Pairwise linkage disequilibrium (D′ above diagonal) and statistical significance (p value below diagonal) for the Egyptian populations

	+126 G/T	+13924	+25133 C/T	+31384 T/C
+126 G/T		0.300	0.514	0.297
+13924 T/A	<0.001		0.259	0.116
+25133 C/T	0.016	0.298		0.833
+31384 T/C	<0.001	0.115	<0.001	

differences were detected in the genotypic and allele frequencies of the remaining SNPs independently in either ethnic group.

The results of the pairwise linkage disequilibrium analysis for Japanese and Egyptians are summarised in tables 5 and 6. In Japanese, +13924 T/A was in strong linkage disequilibrium with +126 G/T ($D' = 0.300$, $\chi^2 = 18.92$, $p<0.001$). The polymorphism +126 G/T was also in linkage disequilibrium with +25133 C/T and +31384 T/C. In Egyptians, +5080 T/C was in significant linkage disequilibrium with +126 G/T ($D' = 0.52$, $\chi^2 = 12.33$, $p<0.001$) and +13924 T/A ($D' = 1.00$, $\chi^2 = 26.81$, $p<0.001$).

Next, haplotype frequencies among these SNPs were estimated separately for the patient group and for the control group, then the distributions of these haplotype frequencies were tested for association with COPD using Fisher's exact test (table 7). In the Japanese population, the frequency of the haplotype +126 T: +13924 T: +25133 C: +31384 C (haplotype 1) was significantly higher in the COPD patient group than in the control group. Twenty three percent of Japanese COPD patients displayed that haplotype, but only 2% of the controls ($p_{corr} = 0.0002$). In contrast, the haplotype +126 T: +13924 A: +25133 C: +31384 C (haplotype 5) and the haplotype +126 G: +13924 T: +25133 C: +31384 T (haplotype 6) showed a significant increase in frequency in Japanese controls compared with the COPD patients. The frequencies of haplotype 5 were 25% in controls *v* 7% in patients ($p_{corr} = 0.0017$) and the frequencies of haplotype 6 were 26% in controls *v* 6% in patients ($p_{corr} = 0.0001$). In Egyptians, there was no significant difference in the distribution of the haplotype frequencies between the COPD patients and the controls.

DISCUSSION

Our study was designed to search for genetic polymorphisms in *CLCA1* and to perform a case–control association study in order to reveal loci involved in COPD susceptibility. It has been demonstrated that a case–control association study with a candidate gene can be a very powerful approach for identifying genetic causes of complex diseases such as COPD.[31] Silverman *et al* proposed the major criteria to evaluate a case–control association study, including selection of candidate genes, population stratification, Hardy-Weinberg equilibrium and correction for multiple comparisons.[32]

From this study, it is reasonable to assume that *CLCA1* is a candidate gene for COPD susceptibility. Although *CLCA1* and its murine counterpart, *mCLCA3*, are virtually undetectable in the normal lung, they are strongly expressed in the bronchial epithelium in response to inflammatory stimulation.[21–25] Furthermore, it has been demonstrated that *CLCA1* regulates airway mucous production under inflammatory conditions.[25] These findings suggest that a high level of continuous *CLCA1* expression results in the overproduction of mucus and may be related to the pathogenesis of COPD, especially of chronic bronchitis. Conversely, as the mucous layer is a defensive barrier against environmental stimuli, the failure of mucus production under inflammatory conditions caused by mutations of the *CLCA1* gene may result in destruction of the parenchyma, leading to emphysema.

As for population stratification, the genetic background of the Japanese population is considered to be homogeneous because Japan is a single racial nation. For the Egyptian population, ethnic and geographic matching were stressed in order to increase homogeneity. All the Egyptian subjects and their parents had to be born in the region of recruitment and to have grandparents who were born inside the country. The present study was designed to improve the power of the case–control approach and decrease the chance of ascertainment bias. By selecting only those individuals who had been sufficiently heavy smokers to develop COPD, we attempted to ensure that both the cases and the controls had a high exposure to the most important risk factor for developing COPD. All the patients and the controls in both ethnic groups

Table 7 Haplotype frequencies of the *CLCA1* gene polymorphisms in Japanese subjects

Hap	+126	+13924	+25133	+31384	COPD	Control
1	T	T	C	C	40 (23)*	2 (2)
2	G	A	C	T	30 (17)	8 (10)
3	G	A	C	C	26 (15)	14 (18)
4	T	T	C	T	22 (13)	5 (6)
5	T	A	C	C	12 (7)**	20 (25)
6	G	T	C	T	10 (6)***	21 (26)
7	T	A	C	T	8 (5)	5 (7)
8	G	A	T	T	8 (4)	4 (5)
9	G	T	T	T	8 (4)	0 (0)
10	T	A	T	T	6 (3)	0 (0)
11	G	T	C	C	4 (2)	0 (0)
12	T	A	T	C	2 (1)	1 (1)

Numbers represent haplotype carriage frequencies with percentages in parentheses.
*$p_{corr} = 0.0002$, **$p_{corr} = 0.0017$, ***$p_{corr} = 0.0001$.
Hap, haplotype

were matched except for symptoms and pulmonary functions.

All the genotypic frequencies for the control populations in both ethnic groups were consistent with Hardy-Weinberg equilibrium; it is noteworthy that only the genotype data at +126 G/T in the Egyptian COPD group were out of equilibrium. Given the hypothesis that *CLCA1* is involved in the pathogenesis of COPD, the genotypic frequency could be altered in the patient group. We ruled out the possibility of genotyping errors at this locus by using both DNA sequencing and RFLP. Although there are many pairwise linkage disequilibriums between the SNPs studied, these linkages are not complete. Therefore, we adopted the Bonferroni correction for multiple comparisons.

In this study, 22 polymorphisms in the *CLCA1* gene were identified, eight of which were mutant homozygous in all the samples in both populations. Of the remaining 14 SNPs, only four were found to have more than 10% incidence in each population. When tested independently, the distributions of two SNPs, +5080 T/C and +13924 T/A, tended to show significant difference in the Egyptian and the Japanese groups, respectively. As +5080 T/C in exon 3 is a non-synonymous nucleotide substitution, this amino acid change may have an influence on the function of *CLCA1* protein. Because the minor allele frequency of this SNP, +5080 C, is higher in the controls than in the patients, the nucleotide substitution T→C can be thought to play a protective role in susceptibility to development of COPD. The other SNP, +13924 T/A in exon 6, is synonymous. It is possible, however, that this nucleotide substitution exerts an influence on the function of *CLCA1* protein by changing the rate of translation by modification of the mRNA stability and/or the ribosome binding. It has been demonstrated that codon usage can affect the general expression level of a heterologous gene, and that prevalent codons can result in a substantial increase in expression efficiency.[33] It is also possible that one or more of these SNPs are in linkage disequilibrium with an original causal polymorphism in nearby *CLCA* genes that share structural and functional similarities with *CLCA1*.[19]

The result obtained by analysing only individual SNPs is insufficient. Haplotype analysis, testing associations using several polymorphisms, sometimes demonstrates genetic influences that are not detected by the analysis of single polymorphisms.[34]

In this study, the associations of the individual SNPs with COPD were different between the two populations. Additionally, there was an association between COPD and some haplotypes in the Japanese population; no such association was found for any haplotype in Egyptians. As the susceptibility to COPD is considered to be influenced by multiple genetic causes and genotype–environment interactions,[35] it is also possible that different polymorphisms in different ethnic groups cause the same COPD phenotype. Another possibility is that the component of COPD was not the same between the Japanese and the Egyptians, although there was no population stratification in each ethnic group independently.

COPD has been recognised as a heterogeneous disease that includes chronic bronchitis and emphysema.[36] In the present study, most of the Egyptian COPD patients presented with chronic productive cough and a diagnosis of chronic bronchitis was made for them. In contrast, more than half of the Japanese patients did not complain of a productive cough but of progressive dyspnoea on exertion. Plain chest radiographs and computed tomography images of most Japanese patients showed findings suggestive of emphysema such as hyperinflation, hyperlucency, and markedly low attenuation areas.[37 38] It has also been demonstrated that the prevalence of COPD in smokers is lower in Japanese than in white populations[5] and that mortality rates for COPD are higher in whites than in non-whites.[39] Therefore, it seems likely that different ethnic populations have different components of COPD. This may be one of the reasons why prevalences of risk alleles of candidate genes for COPD differ greatly between different studies analysing different ethnic groups. The main drawback of our study is the relatively small sample size, which reduces the power of detection of true association, but even with a larger sample, the functional and biological impact of the described polymorphisms and haplotypes would require further study.

In conclusion, we have performed for the first time a comprehensive investigation of *CLCA1* gene polymorphisms. We have identified 22 SNPs and characterised the associations between these SNPs and susceptibility to COPD. The SNPs and the haplotypes showing significant differences may be useful for predicting COPD susceptibility, thus preventing the progression of the disease by early intervention. Further functional characterisation of these SNPs and haplotypes is required to clarify the significance of *CLCA1* in the pathogenesis of COPD. These results would also be useful for further evaluation of the *CLCA1* gene in other complex respiratory diseases such as bronchial asthma.

ACKNOWLEDGEMENTS

The authors thank all participants in the study, especially staff of the participating hospitals who helped in patient recruitment and sample collection.

Authors' affiliations

A E Hegab, T Sakamoto, Y Uchida, A Nomura, Y Ishii, Y Morishima, M Mochizuki, T Kimura, W Saitoh, K Sekizawa, Department of Pulmonary Medicine, Institute of Clinical Medicine, University of Tsukuba, Ibaraki, Japan
H H Massoud, H M Massoud, K M Hassanein, Department of Chest Diseases and Tuberculosis, Faculty of Medicine, Cairo University, Cairo, Egypt

Correspondence to: Dr K Sekizawa, Department of Pulmonary Medicine, Institute of Clinical Medicine, University of Tsukuba, 1-1-1 Tennoudai, Tsukuba, Ibaraki 305-8575, Japan; kiyo-se@md.tsukuba.ac.jp

Received 24 September 2003
Accepted 25 September 2003

REFERENCES

1 **Snider GL**. Chronic obstructive pulmonary disease: risk factors, pathophysiology and pathogenesis. *Annu Rev Med* 1989;**40**:411–29.
2 **Bascom R**. Differential susceptibility to tobacco smoke: possible mechanisms. *Pharmacogenetics* 1991;**1**:102–6.
3 **Lebowitz MD**, Knudson RJ, Burrows B. Family aggregation of pulmonary function measurements. *Am Rev Respir Dis* 1984;**129**:8–11.
4 **Redline S**, Tishler PV, Rosner B, Lewitter FI, Vandenburgh M, Weiss ST, Speizer FE. Genotypic and phenotypic similarities in pulmonary function among family members of adult monozygotic and dizygotic twins. *Am J Epidemiol* 1989;**129**:827–36.
5 **Marcus EB**, Buist AS, Curb JD, MacLean CJ, Reed DM, Johnson LR, Yano K. Correlates of FEV1 and prevalence of pulmonary conditions in Japanese-American men. *Am Rev Respir Dis* 1988;**138**:1398–404.
6 **Givelber RJ**, Couropmitree NN, Gottlieb DJ, Evans JC, Levy D, Myers RH, O'Connor GT. Segregation analysis of pulmonary function among families in the Framingham study. *Am J Respir Crit Care Med* 1998;**157**:1445–51.
7 **Joos L**, He JQ, Shepherdson MB, Connett JE, Anthonisen NR, Pare PD, Sandford AJ. The role of matrix metalloproteinase polymorphisms in the rate of decline in lung function. *Hum Mol Genet* 2002;**11**:569–76.
8 **Poller W**, Meisen C, Olek K. DNA polymorphisms of the alpha 1-antitrypsin gene region in patients with chronic obstructive pulmonary disease. *Eur J Clin Invest* 1990;**20**:1–7.
9 **Poller W**, Faber JP, Scholz S, Weidinger S, Bartholome K, Olek K, Eriksson S. Mis-sense mutation of alpha 1-antichymotrypsin gene associated with chronic lung disease. *Lancet* 1992;**339**:1538.
10 **Hirano K**, Sakamoto T, Uchida Y, Morishima Y, Masuyama K, Ishii Y, Nomura A, Ohtsuka M, Sekizawa K. Tissue inhibitor of metalloproteinases-2 gene polymorphisms in chronic obstructive pulmonary disease. *Eur Respir J* 2001;**18**:748–52.

11 **Yamada N**, Yamaya M, Okinaga S, Nakayama K, Sekizawa K, Shibahara S, Sasaki H. Microsatellite polymorphism in the heme oxygenase-1 gene promoter is associated with susceptibility to emphysema. *Am J Hum Genet* 2000;**66**:187–95.
12 **Smith CA**, Harrison DJ. Association between polymorphism in gene for microsomal epoxide hydrolase and susceptibility to emphysema. *Lancet* 1997;**350**:630–3.
13 **Baranova H**, Perriot J, Albuisson E, Ivaschenko T, Baranov VS, Hemery B, Mouraire P, Riol N, Malet P. Peculiarities of the GSTM1 0/0 genotype in French heavy smokers with various types of chronic bronchitis. *Hum Genet* 1997;**99**:822–6.
14 **Huang SL**, Su CH, Chang SC. Tumor necrosis factor-alpha gene polymorphism in chronic bronchitis. *Am J Respir Crit Care Med* 1997;**156**:1436–9.
15 **Joos L**, McIntyre L, Ruan J, Connett JE, Anthonisen NR, Weir TD, Pare PD, Sandford AJ. Association of IL-1beta and IL-1 receptor antagonist haplotypes with rate of decline in lung function in smokers. *Thorax* 2000;**56**:863–6.
16 **Joos L**, Pare PD, Sandford AJ. Genetic risk factors of chronic obstructive pulmonary disease. *Swiss Med Wkly* 2002;**132**:27–37.
17 **Gruber AD**, Elble RC, Ji H-L, Schreur KD, Fuller CM, Pauli BU. Genomic cloning, molecular characterization and functional analysis of human hCLCA1, the first human member of the family of Ca^{2+}-activated Cl^{-} channels. *Genomics* 1998;**54**:200–14.
18 **Komiya T**, Tanigawa Y, Hirohashi S. Cloning and identification of the gene gob-5, which is expressed in intestinal goblet cells in mice. *Biochem Biophys Res Commun* 1999;**255**:347–51.
19 **Pauli BU**, Abdel-Ghany M, Cheng HC, Gruber AD, Archibald HA, Elble RC. Molecular characteristics and functional diversity of CLCA family members. *Clin Exp Pharmacol Physiol* 2000;**27**:901–5.
20 **Silverman EK**, Palmer LJ, Mosley JD, Barth M, Senter JM, Brown A, Drazen JM, Kwiatkowski DJ, Chapman HA, Campbell EJ, Province MA, Rao DC, Reilly JJ, Ginns LC, Speizer FE, Weiss ST. Genomewide linkage analysis of quantitative spirometric phenotypes in severe early-onset chronic obstructive pulmonary disease. *Am J Hum Genet* 2002;**70**:1229–39.
21 **Wilk JB**, DeStefano AL, Arnett DK, Rich SS, Djousse L, Crapo RO, Leppert MF, Province MA, Cupples LA, Gottlieb DJ, Myers RH. A genome-wide scan of pulmonary function measures in the National Heart, Lung, and Blood Institute Family Heart Study. *Am J Respir Crit Care Med* 2003;**167**:1528–33.
22 **Toda M**, Tulic MK, Levitt RC, Hamid Q. A calcium-activated chloride channel (HCLCA1) is strongly related to IL-9 expression and mucus production in bronchial epithelium of patients with asthma. *J Allergy Clin Immunol* 2002;**109**:246–50.
23 **Nakanishi A**, Morita S, Iwashita H, Sagiya Y, Ashida Y, Shirafuji H, Fujisawa Y, Nishimura O, Fujino M. Role of gob-5 in mucus overproduction and airway hyperresponsiveness in asthma. *Proc Natl Acad Sci USA* 2001;**98**:5175–80.
24 **Zhou Y**, Dong Q, Louahed J, Dragwa C, Savio D, Huang M, Weiss C, Tomer Y, McLane MP, Nicolaides NC, Levitt RC. Characterization of a calcium-activated chloride channel as a shared target of Th2 cytokine pathways and its potential involvement in asthma. *Am J Respir Cell Mol Biol* 2001;**25**:486–91.
25 **Hoshino M**, Morita S, Iwashita H, Sagiya Y, Nagi T, Nakanishi A, Ashida Y, Nishimura O, Fujisawa Y, Fujino M. Increased expression of the human Ca^{2+}-activated Cl^{-} channel 1 (CaCC1) gene in the asthmatic airway. *Am J Respir Crit Care Med* 2002;**165**:1132–6.
26 **Pauwels RA**, Buist AS, Calverley PM, Jenkins CR, Hurd SS, GOLD Scientific Committee. Global strategy for the diagnosis, management, and prevention of chronic obstructive pulmonary disease. NHLBI/WHO Global Initiative for Chronic Obstructive Lung Disease (GOLD) Workshop summary. *Am J Respir Crit Care Med* 2001;**163**:1256–76.
27 **American Thoracic Society**. Lung function testing: selection of reference values and interpretative strategies. ATS Statement. *Am Rev Respir Dis* 1991;**144**:1202–18.
28 **Vidal-Puig A**, Moller DE. Comparative sensitivity of alternative single–strand conformation polymorphism (SSCP) methods. *Biotechniques* 1994;**17**:490–2, 494, 496.
29 **Kwok PY**. SNP genotyping with fluorescence polarization detection. *Hum Mutat* 2002;**19**:315–23.
30 **Lewontin RC**. On measures of gametic disequilibrium. *Genetics* 1988;**120**:849–52.
31 **Risch N**, Merikangas K. The future of genetic studies of complex human diseases. *Science* 1996;**273**:1516–17.
32 **Silverman EK**, Palmer LJ. Case-control association studies for the genetics of complex respiratory diseases. *Am J Respir Cell Mol Biol* 2000;**22**:645–8.
33 **Kim CH**, Oh Y, Lee TH. Codon optimization for high-level expression of human erythropoietin (EPO) in mammalian cells. *Gene* 1997;**199**:293–301.
34 **Corbex M**, Poirrier O, Fumeron F, Betoulle D, Evance A, Ruidavets JB, Arveiler D, Luc G, Tiret L, Cambien F. Extensive association analysis between the CETP gene and coronary heart phenotypes reveals several putative functional polymorphisms and gene-environment inteaction. *Genet Epidemiol* 2000;**19**:64–80.
35 **Lander ES**, Schork NJ. Genetic dissection of complex traits. *Science* 1994;**265**:2037–48.
36 **Mannino DM**. COPD: epidemiology, prevalence, morbidity and mortality, and disease heterogeneity. *Chest* 2002;**121**:121–6S.
37 **Saunders C**. The radiographic diagnosis of emphysema. *Radiol Clin North Am* 1991;**29**:1019–30.
38 **Gevenois PA**, De Vuyst P, de Maertelaer V, Zanen J, Jacobovitz D, Cosio MG, Yernault JC. Comparison of computed density and microscopic morphometry in pulmonary emphysema. *Am J Respir Crit Care Med* 1996;**154**:187–92.
39 **Higgins MW**, Thom T. Incidence, prevalence, and mortality: intra- and inter-county differences. In: Hensley MJ, Saunders NA, editors. *Clinical epidermiolgy of chronic obstructive pulmonary disease*. New York: Marcel Dekker, 1989:23–43.