



# Sex-specific association of epidermal growth factor gene polymorphisms with acute respiratory distress syndrome

C.C. Sheu<sup>\*,#,§</sup>, R. Zhai<sup>\*,§</sup>, L. Su<sup>\*</sup>, P. Tejera<sup>\*</sup>, M.N. Gong<sup>†</sup>, B.T. Thompson<sup>+</sup>,  
F. Chen<sup>\*</sup> and D.C. Christiani<sup>\*,+</sup>

**ABSTRACT:** Epidermal growth factor (EGF) is involved in alveolar epithelial repair, lung fluid clearance and inflammation, and is regulated by sex hormones. An unmatched, nested case–control study was conducted to evaluate the associations of EGF variants with acute respiratory distress syndrome (ARDS) and the role of sex on the associations between EGF variants and ARDS.

Patients with ARDS risk factors upon intensive care unit admission were enrolled. Cases were 416 Caucasians who developed ARDS and controls were 1,052 Caucasians who did not develop ARDS. Cases were followed for clinical outcomes and 60-day mortality. One functional single nucleotide polymorphism (SNP), rs4444903, and six haplotype-tagging SNPs spanning the entire EGF gene were genotyped.

No individual SNP or haplotype was associated with ARDS risk or outcomes in all subjects. Sex-stratified analyses showed opposite effects of EGF variants on ARDS in males versus in females. SNPs rs4444903, rs2298991, rs7692976 and rs4698803, and haplotypes GGCGTC and ATCAAG were associated with ARDS risk in males. No associations were observed in females. Interaction analysis showed that rs4444903, rs2298991, rs7692976 and rs6533485 significantly interacted with sex for ARDS risk.

The present study suggests that associations of epidermal growth factor gene variants with acute respiratory distress syndrome risk are modified by sex. The current findings should be replicated in other populations.

**KEYWORDS:** Acute respiratory distress syndrome, epidermal growth factor, genetic susceptibility, haplotypes, lung injury, molecular epidemiology

**A**mong patients with sepsis, pneumonia, trauma and other triggering conditions, only a subset will develop acute respiratory distress syndrome (ARDS), and only ~60% of those developing ARDS will survive, suggesting that genetics may influence the susceptibility to and recovery from this syndrome. ARDS is characterised by diffuse damage to the alveolar barrier, which leads to increased permeability and influx of protein-rich oedema into the interstitial airspace. In addition to endothelial injury, epithelial damage also plays an important role in the development of and recovery from this disorder [1, 2]. The loss of epithelial integrity contributes to the formation of alveolar oedema, while the repair of epithelial injury and the restoration of alveolar epithelial fluid transport function facilitate the resolution of pulmonary oedema [2, 3].

Epidermal growth factor (EGF) is a key growth factor among the ligands of EGF receptors (EGFRs). The family of EGFs and receptors is important in regulating cell growth, maturation, function and maintenance in epithelial tissues [4]. It is suggested that acute lung injury elicits growth factor responses that trigger repair mechanisms to restore lung integrity [5]. Studies show that EGF regulates bronchial and alveolar epithelial repair after lung injury [6, 7]. Moreover, EGF decreases alveolar epithelial junctional permeability, upregulates alveolar epithelial Na<sup>+</sup>-K<sup>+</sup>-ATPase and increases lung fluid clearance [8–10]. Conversely, EGF has also been increasingly regarded as a pro-inflammatory mediator. In asthma, the EGFR pathway is involved not only in the bronchial epithelial repair [11] but also in lung inflammation [12]. EGF increases the

## AFFILIATIONS

<sup>\*</sup>Dept of Environmental Health, Harvard School of Public Health,  
<sup>+</sup>Pulmonary and Critical Care Unit, Dept of Medicine, Massachusetts General Hospital, Harvard Medical School, Boston, MA,  
<sup>†</sup>Division of Pulmonary and Critical Care Medicine, Mount Sinai School of Medicine, New York, NY, USA,  
<sup>#</sup>Division of Pulmonary and Critical Care Medicine, Kaohsiung Medical University Hospital, Kaohsiung Medical University, Kaohsiung, Taiwan.

<sup>§</sup>Both authors contributed equally to this article.

## CORRESPONDENCE

D.C. Christiani  
Dept of Environmental Health, Harvard School of Public Health, 665 Huntington Avenue, Room I-1407, Boston, MA 02115, USA.

Fax: 1 6174326981  
E-mail: dchris@hsph.harvard.edu

Received:  
June 16 2008  
Accepted after revision:  
October 13 2008

## SUPPORT STATEMENT

The present study was supported in part by grants from the National Institutes of Health (grants HL60710 and ES00002; Bethesda, MD, USA) and the Flight Attendant Medical Research Institute (grant no. 062459-YCSA; Miami, FL, USA).

STATEMENT OF INTEREST  
None declared.

European Respiratory Journal  
Print ISSN 0903-1936  
Online ISSN 1399-3003

This article has supplementary material accessible from [www.erj.ersjournals.com](http://www.erj.ersjournals.com)

production of interleukin-8, modulates the inflammatory effects of tumour necrosis factor- $\alpha$ , and enhances the neutrophil-mediated immunity [13–15]. Taken together, EGF might play a critical role in the pathogenesis of ARDS and, therefore, EGF gene polymorphisms could be potential risks for ARDS.

The EGF gene is located on chromosome 4 (4q25), spans ~99 kb and has 24 exons and 23 introns. Genetic variation within the EGF gene has been studied in relation to the EGF phenotype and disease susceptibility [16]. However, most of these studies have focused on the single nucleotide polymorphism (SNP) rs4444903, the GG genotype of which is associated with a higher secretion of EGF protein than the AA genotype [17]. Recently, it has been recognised that a haplotype-based tagging SNP approach can comprehensively scan the common variation of an entire gene and provide greater power than single-marker tests for genetic disease association [18]. In addition, it is evident that the EGFR signalling pathways are regulated by sex hormones [19–21]. Studies have shown sex differences for salivary and tear EGF levels [22, 23], as well as the EGF effects in ulcer healing [24]. Moreover, sex-specific associations of EGF polymorphisms with phenotypes have been found in schizophrenia [25, 26]. In the current study, it was hypothesised that common genetic variation of EGF is associated with the risk and outcomes of ARDS, and that such an association is modified by sex. A hospital-based, unmatched, modified, nested case-control study of patients at risk for ARDS was conducted, and a haplotype-tagging SNP approach was used to test the hypotheses.

## METHODS

### Study design and subjects

The current study is part of an ongoing molecular epidemiology project investigating the influences of genetic factors on the development and outcomes of ARDS. Details of the study have been described previously [27]. Briefly, study subjects were selected from patients admitted to the intensive care units (ICU) at Massachusetts General Hospital (Boston, MA, USA) from September 1999 to November 2006. Patients with clinical risk factors for ARDS such as sepsis, septic shock, trauma, pneumonia, aspiration or multiple transfusions were eligible for inclusion (see supplementary material table 1). Exclusion criteria included age <18 yrs, diffuse alveolar haemorrhage, chronic lung diseases other than chronic obstructive pulmonary disease or asthma, directive to withhold intubation, immunosuppression not secondary to corticosteroid, and treatment with granulocyte colony-stimulating factor. Baseline characteristics and Acute Physiology and Chronic Health Evaluation (APACHE) III scores were recorded on ICU admission. The enrolled patients who fulfilled the American-European Consensus Committee (AECC) criteria for ARDS upon ICU admission or during the daily follow-up were considered as ARDS cases, whereas at-risk patients who did not meet the criteria for ARDS were considered as controls. All enrolled patients with ARDS were then followed for clinical outcomes and all-cause 60-day mortality after the development of ARDS. To reduce the potential confounding from ethnic backgrounds, only Caucasian (non-Hispanic white) patients were analysed. A schematic of study design and patient selection is illustrated in figure 1. The present study was

approved by the Human Subjects Committees of the Massachusetts General Hospital and the Harvard School of Public Health (Boston). Written informed consent was obtained from all subjects or surrogates.

### SNP selection

The haplotype-tagging SNPs in the EGF gene were selected based on the HapMap data [28] for the CEU population (Utah residents with ancestry from northern and western Europe). The multimarker tagging algorithm was used with criteria of  $r^2 > 0.8$  and minor allele frequency  $\geq 0.1$ . The entire EGF gene was covered, including 5 kb on each side of the gene encompassing the promoter and 3' untranslated region (UTR). In addition, a functional SNP rs4444903 (+61 A>G), which has not been included in the HapMap data, was also used.

### Genotyping

Genomic DNA was extracted from whole blood using the Puregene DNA Isolation Kit (Gentra Systems, Minneapolis, MN, USA). The selected SNPs of EGF were genotyped using TaqMan® SNP Genotyping Assay (Applied Biosystems, Foster City, CA, USA). Primers and probes were ordered from Applied Biosystems. All PCR amplifications were performed in a 384-well format on GeneAmp® PCR Systems 9700 (Applied Biosystems). The fluorescence of PCR products was detected using the ABI Prism® 7900HT Sequence Detection System (Applied Biosystems). Genotyping was performed by laboratory personnel blinded to case-control status. A random 10% of samples were inserted in different 384-well plates as duplicates for quality-control purposes. Two investigators reviewed all genotyping results independently. The concordance rate for the duplicate samples was >99% and the overall genotyping success rate was 98.8%. Samples not yielding the genotypes of all SNPs were excluded from analysis.

### Statistical analyses

The homogeneity of baseline characteristics between the two groups was tested by Fisher's exact test for categorical variables and by an unpaired t-test for continuous variables. The differences of genotype distributions between groups were compared using a Chi-squared test. SAS/Genetics was used to calculate the allele frequencies, test the deviation from the Hardy-Weinberg equilibrium (HWE), and estimate pairwise  $D'$  and  $r^2$  values for linkage disequilibrium (LD).

Haplotype frequencies were estimated from the unphased genotype data in the combined population (cases and controls), using the expectation maximisation algorithm as implemented in SAS/Genetics (SAS Institute Inc., Cary, NC, USA). The associations between EGF haplotypes and the risk and survival of ARDS were analysed using the expectation-substitution approach as implemented in the SAS macro Haplotype Scoring for Generalized Linear Modeling and Haplotype-Disease Association Tests (HAPPY) [29, 30]. This approach treats subject-specific expected haplotype indicators, calculated by an additive mode, as observed covariates for regression models. Haplotypes with a frequency  $\geq 5\%$  in the total population were considered to be common and the most common haplotype was used as the referent in the regression model to assess the haplotype-specific risk for ARDS. All other haplotypes were pooled into a separate rare haplotypes category.

**TABLE 1** Baseline characteristics of study population

	ARDS development			ARDS 60-day mortality		
	ARDS	Controls	p-value	Nonsurvivors	Survivors	p-value
<b>Subjects n</b>	416	1052		170	246	
<b>Age yrs</b>	59±18	63±17	<0.001	67±15	53±18	<0.001
<b>Male</b>	246 (59.1)	641 (60.9)	0.554	100 (58.8)	146 (59.4)	0.920
<b>APACHE III score</b>	68±23	61±21	<0.001	79±21	60±20	<0.001
<b>ARDS risk factor</b>						
Sepsis	112 (26.9)	383 (36.4)	<0.001	43 (25.3)	69 (28.1)	0.575
Septic shock	249 (59.8)	461(43.8)	<0.001	111 (65.3)	138 (56.1)	0.067
Pneumonia	284 (68.3)	455 (43.3)	<0.001	123 (72.4)	161 (65.5)	0.164
Aspiration	41 (9.9)	88 (8.4)	0.359	19 (11.2)	22 (8.9)	0.505
Multiple transfusion	37 (8.9)	126 (12.0)	0.097	17 (10.0)	20 (8.1)	0.600
Trauma	31 (7.5)	81 (7.7)	0.914	2 (1.2)	29 (11.8)	<0.001
<b>Comorbidity</b>						
Post-operative	23 (5.5)	77 (7.3)	0.251	11 (6.5)	12 (4.9)	0.518
Diabetes	73 (17.6)	285 (27.1)	<0.001	32 (18.8)	41 (16.7)	0.601
End-stage renal disease	25 (6.0)	55 (5.2)	0.610	15 (8.8)	10 (4.1)	0.058
Liver cirrhosis/failure	29 (7.0)	41 (3.9)	0.020	21 (12.4)	8 (3.3)	<0.001
Metastatic cancer	12 (2.9)	51 (4.9)	0.115	9 (5.3)	3 (1.2)	0.018
<b>History of steroid use</b>	44 (10.6)	94 (8.9)	0.323	27 (15.9)	17 (6.9)	0.005
<b>History of alcohol abuse</b>	60 (14.4)	105 (10.0)	0.017	27 (15.9)	33 (13.4)	0.482

Data are presented as mean±SD or n (%), unless otherwise stated. ARDS: acute respiratory distress syndrome; APACHE: Acute Physiology and Chronic Health Evaluation.

A multivariate logistic regression was used to estimate the genotype- and haplotype-specific odds ratio (OR) and 95% confidence interval (CI) for ARDS risk. To evaluate the associations of individual genotypes and haplotypes with ARDS survival, the Cox proportional hazard model was used to estimate the hazard ratio (HR) and 95% CI. Covariates for the logistic regression and Cox models included age, sex (not included in sex-stratified analyses), APACHE III score and the potential risks for ARDS development and mortality based on univariate analysis. Global tests for the associations between haplotypes and ARDS risk and survival were carried out using the likelihood ratio test. For statistically significant associations, adjusted p-values were calculated to correct for multiple comparisons, using the false discovery rate (FDR) procedure of BENJAMINI and HOCHBERG [31]. The gene–sex interactions were examined by sex-stratification and their strength was evaluated in multivariate logistic regression models including an interaction term. A two-sided p-value <0.05 was considered to be statistically significant.

## RESULTS

### Study population

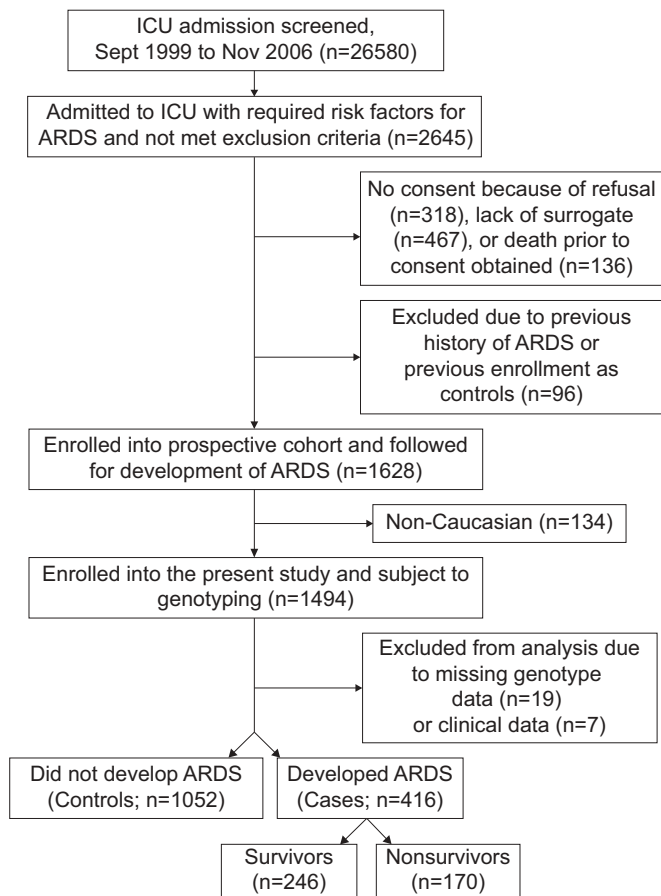
The flow diagram of patient selection in the present study is illustrated in figure 1. The baseline characteristics of the current study population, including 416 ARDS cases and 1052 at-risk controls, are shown in table 1. Patients who developed ARDS were younger and had higher APACHE III scores than those who did not develop ARDS. Among patients with ARDS, the survivors were younger and had lower APACHE III scores than the nonsurvivors. There were no

differences in sex distributions between ARDS patients and controls, or between ARDS survivors and nonsurvivors. The comparisons of ARDS risk factors and comorbidities between ARDS patients and controls, and between ARDS survivors and nonsurvivors, are also shown in table 1. The comparisons of clinical characteristics between males and females are provided in the online supplementary material table 2.

### SNP selection and genotype frequencies

Seven SNPs of the *EGF* gene were selected for the current study. In the order of 5' to 3', they were: the functional SNP rs4444903 and six tagging SNPs, rs2298991, rs11568993, rs6850557, rs7692976, rs4698803 and rs6533485 (table 2). All SNPs conformed to HWE, with the exception of rs6850557 (FDR adjusted p<0.05), which was then excluded from further analyses. Pairwise LD analysis revealed that all alleles among selected SNPs are in high LD (online supplementary material figure 1).

The alleles, locations, chromosome positions and minor allele frequencies of these seven SNPs are presented in table 2. The minor allele frequencies for ARDS and controls were not different in all subjects for all SNPs, but were significantly different in males for rs7692976 and rs4698803 (p=0.030 and p=0.024, respectively). The genotype distributions are shown in the online supplementary material table 3. No differences in genotype frequencies were found between ARDS and controls in all subjects, between males and females, or between ARDS survivors and nonsurvivors. However, the sex-stratified analyses showed that the genotype frequencies of rs4444903, rs7692976 and rs4698803 were significantly different between



**FIGURE 1.** Schematic of study design and patient selection. ICU: intensive care unit; ARDS: acute respiratory distress syndrome.

ARDS and controls in males ( $p=0.029$ ,  $p=0.037$  and  $p=0.032$ , respectively; online supplementary material table 4).

#### Associations between EGF variants and ARDS risk

Considering all subjects as a whole, none of the individual SNPs were significantly associated with ARDS risk (table 3). Upon stratification by sex, the variant genotypes of rs4444903

(OR 1.64, 95% CI 1.17–2.31;  $p=0.005$ ), rs2298991 (OR 1.50, 95% CI 1.07–2.11;  $p=0.019$ ) and rs7692976 (OR 1.64, 95% CI 1.17–2.31;  $p=0.005$ ) were found to be associated with increased risks, while the variant genotypes of rs4698803 (OR 0.67, 95% CI 0.48–0.96;  $p=0.025$ ) were associated with reduced risks of developing ARDS in males. In females, none of the associations between EGF variants and ARDS risk were significant. However, the effects of variant genotypes on ARDS risk were observed to be opposite from those in males for five of the six SNPs.

There were five common haplotypes (frequencies 30.6, 29.0, 14.0, 8.8 and 7.5%) inferred from the six polymorphisms analysed. Similar with the results from genotypes analyses, a significant association between haplotypes and ARDS risk was observed only in the male subgroup, in which the global test for association was significant (likelihood ratio test  $p=0.005$ ; table 3). In addition, Hap2 (GGCGTC) was associated with an increased ARDS risk (OR 1.35, 95% CI 1.00–1.81;  $p=0.048$ ), whereas Hap3 (ATCAAG) was associated with a reduced ARDS risk (OR 0.64, 95% CI 0.44–0.94;  $p=0.022$ ). The haplotype-specific associations were assessed in one regression model, thus correcting for multiple comparisons is not necessary.

#### Associations between EGF variants and clinical outcomes of ARDS

The clinical outcomes of ARDS patients among different genotype groups are shown in the online supplementary material table 5. In 28 days since ARDS diagnosis, no significant differences in ICU-free days, ventilator-free days, successful extubation rates and mortality rates were observed between genotypes of all SNPs. The 60-day mortality rates were also not different between genotypes. In analyses stratified by sex (data not shown), male ARDS patients with variant genotypes of rs4698803 had a higher successful extubation rate (68.7 versus 54.2%;  $p=0.043$ ) and more ventilator-free days ( $11.9 \pm 9.8$  versus  $8.5 \pm 9.7$  days;  $p=0.014$ ) than those with wildtype homozygote. There were no differences in clinical outcomes between genotypes in females.

None of the EGF genotypes or haplotypes were significantly associated with ARDS 60-day survival in all subjects, or in male and female subgroups (table 4). Only Hap3 was marginally

**TABLE 2** Summary of seven single nucleotide polymorphisms (SNPs)

SNP	Allele	Location	Position <sup>#</sup>	Minor allele frequency (ARDS/controls)			HWE p-value <sup>†</sup>
				All subjects	Male	Female	
rs4444903	A>G	5' UTR	111053559	0.436/0.416	0.441/0.393	0.429/0.453	0.140
rs2298991	T>G	Intron 11	111111461	0.433/0.421	0.439/0.400	0.424/0.453	0.169
rs11568993	C>T	Exon 13	111116764	0.096/0.095	0.083/0.094	0.115/0.097	0.592
rs6850557	G>A	Intron 18	111130464	0.359/0.347	0.364/0.333	0.353/0.367	<0.001 <sup>‡</sup>
rs7692976	A>G	Intron 18	111131016	0.440/0.417	0.455/0.399 <sup>§</sup>	0.418/0.447	0.134
rs4698803	T>A	Exon 19	111133876	0.168/0.192	0.152/0.199 <sup>§</sup>	0.191/0.181	0.026
rs6533485	G>C	Intron 22	111147012	0.499/0.491	0.504/0.481	0.509/0.494	0.630

ARDS: acute respiratory distress syndrome; HWE: Hardy–Weinberg equilibrium; UTR: untranslated region. <sup>#</sup>: chromosome positions based on the National Center for Biotechnology Information Build 36; <sup>†</sup>: Chi-squared test for deviation from the HWE equilibrium in all controls; <sup>‡</sup>: false discovery rate adjusted p-value <0.05; <sup>§</sup>: p-value by Chi-squared test <0.05.

**TABLE 3** Associations between epidermal growth factor gene variants and acute respiratory distress syndrome (ARDS) risk

	All subjects <sup>#</sup>		Male <sup>†</sup>		Female <sup>‡</sup>	
	OR (95% CI) <sup>§</sup>	p-value	OR (95% CI) <sup>§</sup>	p-value	OR (95% CI) <sup>§</sup>	p-value
<b>Genotype<sup>¶</sup></b>						
rs4444903	1.22 (0.94–1.59)	0.134	1.64 (1.17–2.31)	0.005 <sup>**†</sup>	0.74 (0.48–1.13)	0.165
rs2298991	1.13 (0.87–1.46)	0.367	1.50 (1.07–2.11)	0.019 <sup>**†</sup>	0.68 (0.45–1.04)	0.077
rs11568893	0.88 (0.64–1.21)	0.435	0.86 (0.56–1.32)	0.488	0.90 (0.55–1.48)	0.678
rs7692976	1.21 (0.94–1.58)	0.145	1.64 (1.17–2.31)	0.005 <sup>**†</sup>	0.73 (0.48–1.11)	0.140
rs4698803	0.83 (0.64–1.09)	0.178	0.67 (0.48–0.95)	0.025 <sup>**†</sup>	1.10 (0.72–1.69)	0.650
rs6533485	1.06 (0.80–1.39)	0.707	1.39 (0.96–2.02)	0.080	0.66 (0.42–1.03)	0.068
<b>Haplotype</b>						
Global test <sup>##</sup>	5.78 <sup>++</sup>	0.328	16.99 <sup>++</sup>	0.005	2.75 <sup>++</sup>	0.738
Hap1 (ATCATG)	Reference		Reference		Reference	
Hap2 (GGCGTC)	1.10 (0.88–1.38)	0.403	1.35 (1.00–1.81)	0.048	0.76 (0.53–1.10)	0.147
Hap3 (ATCAAG)	0.80 (0.60–1.06)	0.118	0.64 (0.44–0.94)	0.022	1.02 (0.65–1.60)	0.941
Hap4 (GGTGTC)	0.91 (0.65–1.26)	0.550	0.87 (0.55–1.36)	0.538	0.91 (0.55–1.49)	0.705
Hap5 (ATCATC)	0.93 (0.64–1.34)	0.681	0.88 (0.55–1.42)	0.605	1.00 (0.54–1.87)	0.993
All others (frequency <5%)	1.11 (0.81–1.52)	0.528	1.22 (0.81–1.84)	0.332	0.85 (0.50–1.43)	0.532

OR: odds ratio; CI: confidence interval. <sup>#</sup>: n=1,468; <sup>†</sup>: n=887; <sup>‡</sup>: n=581. <sup>§</sup>: adjusted for age, sex, Acute Physiology and Chronic Health Evaluation (APACHE) III score, all risk factors for ARDS (table 1), diabetes, liver cirrhosis/failure and alcohol abuse. APACHE III score was revised to remove age and arterial oxygen tension/inspiratory oxygen fraction, to avoid colinearity. Sex was removed from the model during sex-stratified analyses. <sup>¶</sup>: wildtype homozygote of each single nucleotide polymorphism was considered as reference. <sup>##</sup>: likelihood ratio test with five degrees of freedom. <sup>\*\*†</sup>: false discovery rate (FDR) adjusted p-value <0.05. <sup>++</sup>: Chi-squared.

associated with decreased mortality in male ARDS patients (HR 0.57, 95% CI 0.31–1.05; p=0.069).

### Gene–sex interaction for ARDS risk

The interaction analyses showed that the variant genotypes of rs4444903, rs2298991, rs7692976 and rs6533485 significantly interacted with sex for ARDS risk (p-values for interaction 0.010, 0.010, 0.007 and 0.021, respectively; table 5). Because sex hormone activities decrease with age, the current authors further stratified the interaction analysis by age. In order to have the largest statistical power, the median age was used for stratification. The gene–sex interactions were found to be even stronger in the younger group (age <65 yrs), but were not significant in the elder group (age ≥65 yrs), whose sex hormone effects were expected to be less active.

### DISCUSSION

In the present study, the associations of *EGF* variants with ARDS risk and outcomes, and the role of sex in the association between *EGF* variants and ARDS were comprehensively evaluated. The common *EGF* variants were found to be associated with ARDS risk in a sex-specific manner. Variant genotypes of rs4444903, rs2298991, rs7692976 and rs4698803 and haplotypes GGCGTC and ATCAAG were associated with ARDS risk in males. Although no significant associations were found in females, the current authors observed that the associations of *EGF* variants on ARDS development in females were mostly opposite to those observed in males. Such gene–sex interaction was further supported by the results from interaction analyses. Conversely, *EGF* variants did not significantly influence the ARDS outcomes and survival in the present study.

*EGF* was first discovered in submaxillary glands of adult male mice, while human *EGF* ( $\beta$ -urogastrone) was first isolated from urine. The 6-kDa human *EGF* consists of 53 amino acids. It is initially synthesised as a transmembrane glycoprotein precursor (prepro-*EGF*) of 1,207 amino acids, and is then processed through a pro-*EGF* stage to a mature *EGF* protein [32, 33]. The large prepro-*EGF* polypeptide contains the *EGF* subunit and eight additional *EGF*-like subunits, the biological significance of which is unknown [32]. The heparin-binding 160-kDa pro-*EGF*, isolated from human urine, has been shown to be biologically active [33]. However, there seem to be increasing reports of molecular heterogeneity in mature *EGF*. The biological properties and physiological significance of the *EGF* precursors and the heterogeneity of mature *EGF* remain to be elucidated [34].

The *EGF* gene in humans is located on chromosome four. As for functional polymorphisms, the variant *G* allele of rs4444903 has been associated with a higher secretion of *EGF* protein than the *A* allele. The mechanism by which *EGF* levels are modulated has been recently proposed in a study of hepatocellular carcinoma risk [35]. Transcripts from the *G* allele exhibit a longer half-life than those from the *A* allele, thus the *EGF* mRNA and *EGF* protein are increased in cell lines with more copies of *G* in the genotype. In the current study, the variant genotypes (*AG/GG*) of this polymorphism predisposed at-risk male patients to develop ARDS but were not associated with better clinical outcomes, suggesting that higher *EGF* levels in the early phase might contribute to the pathogenesis of ARDS, possibly through the pro-inflammatory properties of *EGF*. Conversely, the proposed function of *EGF* to facilitate epithelial repair and alveolar fluid reabsorption

**TABLE 4** Cox proportional hazard analysis of associations between epidermal growth factor gene variants and 60-day survival in patients with acute respiratory distress syndrome (ARDS)

	All ARDS <sup>#</sup>		Male <sup>†</sup>		Female <sup>‡</sup>	
	HR (95% CI) <sup>§</sup>	p-value	HR (95% CI) <sup>§</sup>	p-value	HR (95% CI) <sup>§</sup>	p-value
<b>Genotype<sup>f</sup></b>						
rs4444903	0.93 (0.67–1.29)	0.662	0.84 (0.54–1.32)	0.452	1.09 (0.66–1.83)	0.731
rs2298991	0.86 (0.62–1.20)	0.378	0.76 (0.48–1.19)	0.229	1.06 (0.63–1.77)	0.829
rs11568893	0.76 (0.49–1.17)	0.213	0.95 (0.53–1.68)	0.849	0.58 (0.28–1.20)	0.142
rs7692976	0.88 (0.63–1.23)	0.460	0.73 (0.46–1.16)	0.179	1.12 (0.67–1.86)	0.664
rs4698803	0.88 (0.62–1.24)	0.470	0.70 (0.43–1.15)	0.137	1.11 (0.66–1.89)	0.689
rs6533485	0.75 (0.53–1.06)	0.102	0.68 (0.42–1.10)	0.112	0.86 (0.51–1.46)	0.577
<b>Haplotype</b>						
Global test <sup>###</sup>	2.60 <sup>**†</sup>	0.761	5.64 <sup>**†</sup>	0.343	6.93 <sup>**†</sup>	0.226
Hap1 (ATCATG)	Reference		Reference		Reference	
Hap2 (GGCGTC)	1.03 (0.77–1.38)	0.840	0.79 (0.52–1.20)	0.267	1.32 (0.85–2.05)	0.222
Hap3 (ATCAAG)	0.93 (0.63–1.38)	0.734	0.57 (0.31–1.05)	0.069	1.37 (0.79–2.39)	0.268
Hap4 (GGTGTC)	0.79 (0.51–1.22)	0.287	0.91 (0.48–1.70)	0.759	0.62 (0.31–1.22)	0.166
Hap5 (ATCATC)	0.84 (0.52–1.34)	0.462	0.83 (0.45–1.53)	0.545	0.89 (0.42–1.92)	0.774
All others (frequency <5%)	1.13 (0.76–1.67)	0.548	1.28 (0.78–2.11)	0.336	0.95 (0.47–1.89)	0.872

HR: hazard ratio; CI: confidence interval. <sup>#</sup>: n=416; <sup>†</sup>: n=246; <sup>‡</sup>: n=170. <sup>§</sup>: adjusted for age, sex, Acute Physiology and Chronic Health Evaluation (APACHE) III score, septic shock, pneumonia, trauma, end-stage renal disease, liver cirrhosis/failure, metastatic cancer, recent use of steroid and treatment with activated protein C. APACHE III score was revised to remove age to avoid colinearity. Sex was removed from the model during sex-stratified analyses. <sup>f</sup>: wildtype homozygote of each single nucleotide polymorphism considered as reference. <sup>###</sup>: likelihood ratio test with five degrees of freedom. <sup>\*\*†</sup>: Chi-squared.

might not be decisive in the recovery from ARDS. Hap2 (GGCGTC), carrying the variant alleles of rs4444903, rs2298991, rs7692976 and rs6533485, was identified as a common haplotype associated with increased ARDS risk in males. The variant genotypes of these four SNPs were all associated with increased ARDS risk ( $p < 0.05$  for the first three SNPs;  $p = 0.08$  for rs6533485). Based on the established functional significance of rs4444903, the positions of SNPs (rs4444903 is in the 5' UTR; the other three are in introns) and the high degree of pairwise LD within these SNPs ( $D' > 0.85$ ), the present authors inferred that the associations of rs2298991 and rs7692976 with ARDS observed in genotype analysis were mediated by their LD with rs4444903.

SNP rs4698803 is a missense  $T > A$  polymorphism located in exon 19 of the *EGF* gene, causing an amino acid substitution (V920E). The *A* allele of rs4698803 was identified as a potentially protective allele against ARDS. Based on the current authors' tagging algorithm, this SNP did not tag any other SNPs in the *EGF* gene. Hap3 (ATCAAG), carrying the variant allele of rs4698803 and the wildtype alleles of all other SNPs, was also identified as a potentially protective haplotype against ARDS. The present results from genotype and haplotype analyses suggest that rs4698803 is independently associated with ARDS. Future studies are needed to understand whether and how this single amino acid mutation, V920E,

**TABLE 5** Genotype–sex interactions for acute respiratory distress syndrome (ARDS) risk

Genotype–sex interaction	All subjects <sup>#</sup>		Age <65 yrs <sup>†</sup>		Age ≥65 yrs <sup>‡</sup>	
	OR (95% CI) <sup>§</sup>	p-value <sup>f</sup>	OR (95% CI) <sup>§</sup>	p-value <sup>f</sup>	OR (95% CI) <sup>§</sup>	p-value <sup>f</sup>
rs4444903·sex	2.01 (1.18–3.43)	0.010 <sup>###</sup>	2.50 (1.18–5.33)	0.017 <sup>###</sup>	1.55 (0.72–3.33)	0.261
rs2298991·sex	2.02 (1.18–3.43)	0.010 <sup>###</sup>	2.67 (1.26–5.69)	0.011 <sup>###</sup>	1.44 (0.67–3.07)	0.350
rs11568893·sex	0.99 (0.52–1.89)	0.973	1.20 (0.49–2.96)	0.688	0.69 (0.26–1.85)	0.464
rs7692976·sex	2.07 (1.22–3.54)	0.007 <sup>###</sup>	3.08 (1.45–6.57)	0.004 <sup>###</sup>	1.32 (0.62–2.82)	0.473
rs4698803·sex	0.64 (0.37–1.09)	0.100	0.68 (0.32–1.48)	0.336	0.61 (0.28–1.30)	0.200
rs6533485·sex	1.96 (1.11–3.46)	0.021 <sup>###</sup>	2.46 (1.09–5.56)	0.031 <sup>###</sup>	1.46 (0.66–3.26)	0.353

OR: odds ratio; CI: confidence interval. Females with wildtype homozygote of each single nucleotide polymorphism were used as a reference. <sup>#</sup>: n=1,468; <sup>†</sup>: n=736; <sup>‡</sup>: n=732; <sup>§</sup>: for interaction in multivariate logistic regression models, adjusted for age, sex, Acute Physiology and Chronic Health Evaluation (APACHE) III score, all risk factors for ARDS (table 1), diabetes, liver cirrhosis/failure and alcohol abuse. APACHE III score was revised to remove age from the score, to avoid colinearity. <sup>f</sup>: for interaction. <sup>###</sup>: false discovery rate adjusted p-value <0.05.

might alter the biochemical function of the EGF precursor and mature EGF protein.

A sex difference similar to that found in the current study has also been observed in the association studies of *EGF* with schizophrenia. The *G* allele at SNP rs4444903 is associated with the age of onset in male patients with schizophrenia, but not in females [25, 26]. The sex difference in the genetic influence of *EGF* on ARDS might be explained by the sex- and tissue-specific regulation of EGFR signalling pathways by sex hormones. In the present study, the gene–sex interaction was not significant in the elder group, whose sex hormone effects are expected to be less active. It is evident that there is crosstalk between sex hormone receptors and EGFR pathways. In lung development, the expression and activity of EGFR appears to be sex specific and cell specific [21]. Androgen treatment has been shown to decrease EGFR density and EGF-induced autophosphorylation of EGFR in foetal rabbit lung [19]. Conversely, oestrogen upregulates the expression of EGFR, whereas progesterone upregulates the expression of 133- and 71-kDa immunoreactive EGF (the prepro-EGF-like proteins) in uterine leiomyoma cells [20]. Indeed, the impacts of sex and sex hormones on acute lung injury have been studied in animal models. Androgens appear to be detrimental while oestrogens tend to be protective in the pathogenesis of acute lung injury [36].

To the current authors' knowledge, the present study is the first to use the haplotype-tagging SNP approach to investigate genetic susceptibility to ARDS. A major advantage of this approach is that it allows a cost-effective identification of common susceptibility alleles across the entire gene region. In addition, the current study is so far the largest report for genetic epidemiology of ARDS and, thus, provided more statistical power to detect genetic associations with ARDS, particularly in sex-stratified analyses. Another major strength of the present study is its study design. The AECC definition was used for ARDS diagnosis, in order to clearly define the phenotype prospectively. The at-risk critically ill patients were selected as controls, to reduce the possible confounding from associations between candidate polymorphisms and predisposing conditions for ARDS. Furthermore, the analysis was restricted to a single ethnic group, thus minimising false results due to population stratification.

One of the limitations in the current study is that neither EGF nor sex hormone levels were determined, thus the functional significance of the *EGF* genotypes and haplotypes on ARDS and their interactions with sex hormones remain to be further defined. Although this is so far the largest population available for ARDS association study, the present authors might not have had adequate power to detect the association of *EGF* variants on ARDS survival with 416 ARDS cases, particularly when it was expected to be small because the resolution of lung oedema and injury contributes only partially to surviving ARDS. Since the current study included only a single cohort, the present findings need to be validated in other independent populations. Finally, the current results are based on Caucasian subjects and additional studies in other ethnic groups will be needed.

In conclusion, the present study demonstrated that genetic associations of epidermal growth factor gene variants with

acute respiratory distress syndrome risk were modified by sex. The variant genotypes of rs4444903, rs2298991, rs7692976 and rs4698803, and haplotypes GGCGTC and ATCAAG were significantly associated with ARDS development in at-risk males. The current findings should be replicated in other cohorts. The present results also warrant future basic research to understand the role of epidermal growth factor, as well as interactions with sex hormones, in the pathogenesis of acute respiratory distress syndrome.

## REFERENCES

- 1 Piantadosi CA, Schwartz DA. The acute respiratory distress syndrome. *Ann Intern Med* 2004; 141: 460–470.
- 2 Ware LB. Pathophysiology of acute lung injury and the acute respiratory distress syndrome. *Semin Respir Crit Care Med* 2006; 27: 337–349.
- 3 Matthay MA, Wiener-Kronish JP. Intact epithelial barrier function is critical for the resolution of alveolar edema in humans. *Am Rev Respir Dis* 1990; 142: 1250–1257.
- 4 Bazley LA, Gullick WJ. The epidermal growth factor receptor family. *Endocr Relat Cancer* 2005; 12: S17–S27.
- 5 Desai TJ, Cardoso WV. Growth factors in lung development and disease: friends or foe? *Respir Res* 2002; 3: 2.
- 6 Ryan RM, Mineo-Kuhn MM, Kramer CM, Finkelstein JN. Growth factors alter neonatal type II alveolar epithelial cell proliferation. *Am J Physiol* 1994; 266: L17–L22.
- 7 Van Winkle LS, Isaac JM, Plopper CG. Distribution of epidermal growth factor receptor and ligands during bronchiolar epithelial repair from naphthalene-induced Clara cell injury in the mouse. *Am J Pathol* 1997; 151: 443–459.
- 8 Borok Z, Hami A, Danto SI, Lubman RL, Kim KJ, Crandall ED. Effects of EGF on alveolar epithelial junctional permeability and active sodium transport. *Am J Physiol* 1996; 270: L559–L565.
- 9 Danto SI, Borok Z, Zhang XL, *et al.* Mechanisms of EGF-induced stimulation of sodium reabsorption by alveolar epithelial cells. *Am J Physiol* 1998; 275: C82–C92.
- 10 Sznajder JI, Ridge KM, Yeates DB, Ilekis J, Olivera W. Epidermal growth factor increases lung liquid clearance in rat lungs. *J Appl Physiol* 1998; 85: 1004–1010.
- 11 Puddicombe SM, Polosa R, Richter A, *et al.* Involvement of the epidermal growth factor receptor in epithelial repair in asthma. *FASEB J* 2000; 14: 1362–1374.
- 12 Hamilton LM, Torres-Lozano C, Puddicombe SM, *et al.* The role of the epidermal growth factor receptor in sustaining neutrophil inflammation in severe asthma. *Clin Exp Allergy* 2003; 33: 233–240.
- 13 Subauste MC, Proud D. Effects of tumor necrosis factor- $\alpha$ , epidermal growth factor and transforming growth factor- $\alpha$  on interleukin-8 production by, and human rhinovirus replication in, bronchial epithelial cells. *Int Immunopharmacol* 2001; 1: 1229–1234.
- 14 Lewkowicz P, Tchorzewski H, Dytnerka K, Banasik M, Lewkowicz N. Epidermal growth factor enhances TNF- $\alpha$ -induced priming of human neutrophils. *Immunol Lett* 2005; 96: 203–210.
- 15 Uddin M, Seumois G, Lau LC, Rytala P, Davies DE, Djukanovic R. Enhancement of neutrophil function by the bronchial epithelium stimulated by EGF. *Eur Respir J* 2008; 31: 714–724.

- 16 Araujo A, Ribeiro R, Azevedo I, *et al.* Genetic polymorphisms of the epidermal growth factor and related receptor in non-small cell lung cancer – a review of the literature. *Oncologist* 2007; 12: 201–210.
- 17 Shahbazi M, Pravica V, Nasreen N, *et al.* Association between functional polymorphism in EGF gene and malignant melanoma. *Lancet* 2002; 359: 397–401.
- 18 Akey J, Jin L, Xiong M. Haplotypes *vs* single marker linkage disequilibrium tests: what do we gain? *Eur J Hum Genet* 2001; 9: 291–300.
- 19 Klein JM, Nielsen HC. Androgen regulation of epidermal growth factor receptor binding activity during fetal rabbit lung development. *J Clin Invest* 1993; 91: 425–431.
- 20 Maruo T, Ohara N, Wang J, Matsuo H. Sex steroidal regulation of uterine leiomyoma growth and apoptosis. *Hum Reprod Update* 2004; 10: 207–220.
- 21 Rosenblum DA, Volpe MV, Dammann CE, Lo YS, Thompson JF, Nielsen HC. Expression and activity of epidermal growth factor receptor in late fetal rat lung is cell- and sex-specific. *Exp Cell Res* 1998; 239: 69–81.
- 22 Guh JY, Chen HC, Chuang LY, Yang CY, Tsai JH, Lai YH. Significance of salivary epidermal growth factor in peptic ulcer disease in hemodialysis patients. *Nephron* 2001; 87: 134–138.
- 23 Nava A, Barton K, Monroy DC, Pflugfelder SC. The effects of age, gender, and fluid dynamics on the concentration of tear film epidermal growth factor. *Cornea* 1997; 16: 430–438.
- 24 Wingren U, Brown TH, Watkins BM, Larson GM. Delayed gastric ulcer healing after extirpation of submandibular glands is sex-dependent. *Scand J Gastroenterol* 1989; 24: 1102–1106.
- 25 Anttila S, Illi A, Kampman O, Mattila KM, Lehtimäki T, Leinonen E. Association of EGF polymorphism with schizophrenia in Finnish men. *Neuroreport* 2004; 15: 1215–1218.
- 26 Hanninen K, Katila H, Anttila S, *et al.* Epidermal growth factor a61g polymorphism is associated with the age of onset of schizophrenia in male patients. *J Psychiatr Res* 2007; 41: 8–14.
- 27 Gong MN, Thompson BT, Williams P, Pothier L, Boyce PD, Christiani DC. Clinical predictors of and mortality in acute respiratory distress syndrome: potential role of red cell transfusion. *Crit Care Med* 2005; 33: 1191–1198.
- 28 International HapMap Project. www.hapmap.org release 22, on NCBI Build 36 assembly, dbSNP build 126.
- 29 Kraft P. Haplotype Scoring for Generalized Linear Modeling, Haplotype-Disease Association Tests. <http://www.hsph.harvard.edu/faculty/kraft/soft.htm> Date last updated: March 20, 2007. Date last accessed: June 16, 2008.
- 30 Kraft P, Cox DG, Paynter RA, Hunter D, De Vivo I. Accounting for haplotype uncertainty in matched association studies: a comparison of simple and flexible techniques. *Genet Epidemiol* 2005; 28: 261–272.
- 31 Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc Series B* 1995; 51: 289–300.
- 32 Burgess AW. Epidermal growth factor and transforming growth factor alpha. *Br Med Bull* 1989; 45: 401–424.
- 33 Parries G, Chen K, Misono KS, Cohen S. The human urinary epidermal growth factor (EGF) precursor. Isolation of a biologically active 160-kilodalton heparin-binding pro-EGF with a truncated carboxyl terminus. *J Biol Chem* 1995; 270: 27954–27960.
- 34 Aybay C, Karakus R, Yucel A. Characterization of human epidermal growth factor in human serum and urine under native conditions. *Cytokine* 2006; 35: 36–43.
- 35 Tanabe KK, Lemoine A, Finkelstein DM, *et al.* Epidermal growth factor gene functional polymorphism and the risk of hepatocellular carcinoma in patients with cirrhosis. *JAMA* 2008; 299: 53–60.
- 36 Carey MA, Card JW, Voltz JW, Germolec DR, Korach KS, Zeldin DC. The impact of sex and sex hormones on lung physiology and disease: lessons from animal studies. *Am J Physiol Lung Cell Mol Physiol* 2007; 293: L272–L278.