Online Data Supplement

ANGPT2 Genetic Variant Is Associated with Trauma-Associated Acute Lung Injury and Altered Plasma Angiopoietin-2 Isoform Ratio

Nuala J. Meyer[1*], Mingyao Li[2], Rui Feng[2], Jonathan Bradfield[3], Robert Gallop[2], Scarlett Bellamy[2], Barry D. Fuchs[1], Paul N. Lanken[1], Steven M. Albelda[1], Melanie Rushefski[4], Richard Aplenc[2,4], Elena Abramova[1], Elena N. Atochina-Vasserman[1], Michael F. Beers[1], Carolyn S. Calfee[5], Mitchell J. Cohen[6], Jean-Francois Pittet[7], David C. Christiani[8], Grant E. O'Keefe[9], Lorraine B. Ware[10], Addison K. May[11], Mark M. Wurfel[12], Hakon Hakonarson[3], and Jason D. Christie[1,2]

DETAILED PATIENTS AND METHODS

Study Population: Stage I. We conducted a prospective cohort study at the Hospital of the University of Pennsylvania, which is a state-designated level I trauma center. From June 1999 - December 2002, as part of a Specialized Center of Research program (SCOR), we created a prospective cohort from all critically ill trauma patients presenting to our hospital's emergency department (ED) and admitted to a surgical intensive care unit (SICU) to determine clinical and genetic factors associated with the development of ALI (1). From 2005 to the present, with funding from the NHLBI (P01-HL079063), we reinitiated the cohort study using identical clinical criteria. All subjects over the age of 13 admitted to the SICU from the ED following major trauma were eligible if they had an injury severity score (ISS) ≥ 16, corresponding to severe trauma (2). Major exclusion criteria included discharge or death within 24 hours of admission, or the presence of an isolated head injury, defined as an Abbreviated Injury Score ≥ 3 for head/neck but ≤ 1 for all other body regions. A more complete list of inclusion and exclusion criteria has been previously published (3). Only African American subjects were analyzed in stage I of the study. Subjects from Penn of European descent were submitted to the Trauma ALI SNP Consortium (TASC), which formed the replication population in stage II. The characteristics of each population are

described in **Tables 1A** and **1B**. Race was genetically determined as described below under Statistical Analysiss. The institutional review board of the University of Pennsylvania and each TASC site approved the study protocols and granted a waiver of consent for participation in accordance with institutional and federal guidelines (3).

Determination of ALI and Collection of Clinical Data. All patients were followed prospectively from admission for the development of ALI during the first 5 days following admission for acute trauma. All blood gas, clinical laboratory, and ventilator data from admission through day 5 in the SICU were recorded using specific Case Report Forms designed for the trauma population. To be designated as having ALI, subjects had to meet all American-European Consensus Conference (AECC) definition criteria within a 24-hour period while tracheally intubated and receiving mechanical ventilation. Determination of ALI was made by two physician investigators who underwent training to standardize chest radiograph interpretation and independently reviewed all arterial blood gas and ventilator data. Determination of the presence or absence of ALI was completed prior to any genotyping.

Collection and Processing of Biological Samples. DNA was extracted from residual EDTA blood samples which were drawn for clinical purposes on admission to the ED. Blood was centrifuged, and buffy coat (pre-2004) or whole blood (post-2004) fractions were aliquoted and stored at -80 ºC until DNA extraction using the Qiagen Qiamp 96 blood kit (Qiagen, Valencia, CA). Negative controls were included with all DNA extraction runs. Extracted DNA from ALI and non-ALI subjects were plated together on each 96-well microplate, and lab personnel were unaware of the ALI status of each sample. Plasma samples were collected from EDTA vacutainers after centrifugation and stored at -80 °C until analysis. Plasma samples were collected only during the early Penn cohort (pre-2004) and were stored only for a subset of subjects. Of subjects with plasma available, 88% were drawn within 72 hours of ED triage.

Genotyping and Quality Control: Stage I. We used a custom 50K single nucleotide polymorphism (SNP) genotyping array, which was designed to assay SNPs in candidate genes and pathways affecting cardiovascular, pulmonary, inflammatory, and metabolic phenotypes, designed in collaboration by the Institute for Translational Medicine and Therapeutics at the University of Pennsylvania, the Broad Institute, and the National Heart Lung and Blood Institute (NHLBI)-supported Candidate-gene Association Resource (CARe) Consortium (4). The array is manufactured by llumina, Inc.® (San Diego, CA) and called the HumanCVD BeadChip™. The array was designed to tag all non-synonymous coding SNPs with minor allele frequencies (MAF) > 0.01, as well as provide coverage for a number of loci with MAF > 0.02 of potential import to cardiac, pulmonary, and metabolic phenotypes (4). Quality control thresholds for each SNP to be included in the analysis included genotyping call ≥ 95%; test of Hardy-Weinberg equilibrium (HWE) by chi-square testing on the whole population yielding a p value ≥ 10^{-6}; and MAF ≥ 0.01 overall. We planned a two-stage genotyping strategy using the Penn cohort to detect potential SNPs associated with ALI, and our validation cohort (TASC, stage II) to confirm associations (5). We used the African American cohort as the discovery population to allow for a more focused genomic association signal, and typed stage I with the HumanCVD BeadChip because it has superior coverage of high priority genes (**Table S1**) and superior coverage of African variation relative to the 610K platform (4). To verify BeadChip genotyping, positive associations were subjected to repeat genotyping by an alternative platform [rs1868554 (Taqman®, Applied Biosystems, Foster City, CA; r^2= 0.97) and rs2442598 (SNPlex®, Applied Biosystems, Foster City, CA; r^2= 1.0)].

Study Population: Stage II: Five academic trauma centers (Harvard University, University of California – San Francisco, University of Pennsylvania, University of Washington, and Vanderbilt University) together

with the Center for Applied Genomics at the Children's Hospital of Philadelphia (CHOP) conceived of and initiated the Trauma ALI SNP Consortium (TASC) in 2007. Descriptions of each center's trauma-associated ALI study population and procedures have been published (6-11). The control population included subjects recruited by CHOP clinicians and nursing staff within the CHOP Healthcare Network as previously described (12-14). The stage II Case and Control populations are described in **Table 1B** and **Figure 1.** The goal of TASC was to pool ALI cases and genetic material in order to perform a whole genome analysis. As of January 2009, over 1066 samples had been submitted, with over 90% of these being European ancestry. Analysis was thus restricted to subjects of European American descent as genetically determined (described below, under Statistical analysis: genetic ancestry determination).

Genotyping Strategy: Stage II (Whole Genome Analysis): All DNA was shipped to Center for Applied Genomics (CAG) at CHOP. Lab personnel were unaware of the case status of each sample. DNA was genotyped using the Illumina Infinium™ II HumanHap610-quad BeadChip technology (Illumina®, San Diego, CA) (13, 15, 16). Quality control thresholds to be included in the analysis included: sample genotyping call rate ≥ 95%; SNP genotyping call rate ≥ 95%; HWE p-value ≥ 10^{-4}; and MAF ≥ 0.01. Imputation of HapMap SNPs (hg18 release 22) was performed using MACH 1.0.16 software http://www.sph.umich.edu/csg/abecasis/MACH/ using default parameters (r^2 0.30 to flag and remove unreliable markers, and posterior probability 0.90 to flag and remove unreliable genotyping calls). Confirmatory genotyping was performed for the validating SNPs. Taqman® genotying (Applied Biosystems, Foster City, CA) was employed for rs1868554 on all 600 cases given that the locus was imputed, whereas rs2442598, a genotyped marker on the Human610 quad, was typed with a SNPlex reaction for 10% of cases (Applied Biosystems, Foster City, CA).

Sequencing and in silico analysis. DNA from 24 AA and 24 EA stage I subjects, equally distributed between ALI and non-ALI status, were selected for sequencing of PCR fragments. We used 24 subjects of each race to provide 99% power to detect MAF ≥ 5% (17). To sequence the region bordering 10 kb around *ANGPT2* exon2, PCR primers were designed using PCRoverlap (University of Washington) to generate amplicons 600 - 800 bp that overlapped by at least 100 bp (**Table S2**). Primers were optimized, then DNA was amplified and sequenced in the forward and reverse direction using a 3730 automated sequencer (Applied Biosystems, Foster City CA). Sequencher 4.8 (Gene Codes, Ann Arbor MI) was used to facilitate secondary peak calls and to compare the sequence data to the NCBI reference sequence. Linkage disequilibrium (LD) between our tagging SNP (rs1868554) and sequenced polymorphisms were analyzed by race with Haploview (18).

In silico splice site enhancement prediction was performed using the SNP Analysis function of Human Splicing Finder 2.4.1 (19), a position weight matrix (PWM)-based package to predict the effect of mutations on consensus splicing signals. The effect of SNPs on consensus splice enhancers was also investigated with alternative PWM splice enhancer matrices ESE, RESCUE-ESE, PESE, and EIE (20-23). Each of these programs employs a distinct computational algorithm to predict splice enhancers or exon identity sequences. ESE, the first described exonic splice enhancer program, identified consensus sequence matrices for elements that bind and respond to proteins of the arginine – serine repeat (SR) family (23); RESCUE-ESE identifies hexamer sequences that are more associated with weak splice sites and which were functional when tested *in vitro* or disrupted by site-directed mutagenesis (20); PESE, named for putative exonic splicing enhancers, identifies octamers that distinguish constitutively spliced noncoding exons with those in pseudo exons and the 5′ untranslated regions (UTRs) of intronless genes, many of which have been validated by *in vitro* assays (21); and EIE (exon identity elements) predicts splice regulatory elements based on strand asymmetry patterns identifying introns and exons (22).

Plasma Protein Analysis: Plasma protein was not collected uniformly but was available from a subset of Stage I subjects (n=128). Blood was centrifuged at 8000 rpm and the plasma removed and stored at -80°C. Angiopoietin-2 protein (ANG2) was measured by sandwich enzyme-linked immunoabsorbant assay (ELISA, R&D Systems, Minneapolis, MN).

The 6 highest plasma ANG2 concentrations for each rs1868554 genotype were normalized based on ELISA ANG2 levels to a concentration of 4 ng/mL. Samples were precleared for immunoprecipitation (IP) by addition of 10% SDS, heating at 100°C for 4 minutes, dilution with 4 volumes of IP buffer (Immunopure IgG binding buffer, Pierce Scientific, Pittsburgh, PA) and centrifugation for 5 minutes at 11,000 X g in an Eppendorf 5415D micro centrifuge. The resulting supernatants were mixed with 100 µg of primary ANG2 antibody (R&D Systems monoclonal mouse, Minneapolis, MN) and incubated at 4 °C overnight. ANG2 –antibody complexes were captured by incubation with 100 µl of Protein A sepharose (Pierce Scientific, Pittsburgh, PA) at RT for 90 minutes. Pellets collected by microcentrifugation were washed three times with detergent-containing IP buffer and once with non-detergent IP buffer. To liberate captured proteins, washed beads were then treated with 2X SDS sample buffer and 2-mercaptoethanol, boiled for 2 minutes , and microcentrifuged.

Proteins in the resulting supernatants were separated by SDS/PAGE with 4-12% gradient gels under reducing conditions using MOPS buffer (24) run for 6 hours at constant voltage (150 v). Separated proteins were transferred to PDVF membranes by wet elution overnight and ANG2 isoforms detected by Western blotting using sequential incubations with primary monoclonal mouse anti-human ANG2 (1:200; Santa Cruz Biotechnology, Santa Cruz, CA) for 48 hours at 4 °C followed by HRP-conjugated secondary antibody (1:5000; GAM-HRP, Jackson Immunoresearch, West Grove, PA) for 90 minutes at RT.

Bands were visualized using enhanced chemiluminescence (ECL+, Amersham Inc, Pittscurgh, PA) and exposure to radiographic film or direct acquisition of chemiluminescence signal using a CCD camera (Kodak 440 Imaging System, New Haven, CT). Net band intensities were then quantified using Kodak 1D software.

Human lung microvascular endothelial cell lysate was used as a positive control for ANG2 staining and human A549 cell lysate was used as a negative control. To confirm ANG2 specific immunoreactivity, blots were developed with secondary antibody only, and bands corresponding to positive control were not visualized (data not shown).

Statistical Analysis:

Genetic Determination of Ancestry. In stage I, ancestry was inferred by multidimensional scaling (MDS) using all markers on the HumanCVD chip and adjusted for 2 principal components from MDS (**Figure S1**) as described previously (25, 26). In stage II, reported ethnicity was screened using the STRUCTURE package (27) and over 200 ancestry informative markers (AIMs) to cluster the TASC submissions with 90 HapMap individuals (CEU, Yoruban, and Chinese/Japanese). Samples with an inferred proportion of CEU ancestry < 90% were determined to be non-EA and excluded from stage II. To further minimize confounding due to population stratification, cases were matched to controls by 'genetic matching' with principal component matching as previously described (14, 28). Each case from TASC was matched to approximately 3 controls using 3 principal components and a distance threshold of 0.07.

SNP and Haplotype association testing. Alleles at each locus were tested for conformity to Hardy-Weinberg proportions. In the univariate analysis of individual SNPs, the incidence of ALI was calculated according to genotype, and the significance of odds ratio of ALI was determined using the χ^2 test. Tests for association were carried out using the software PLINK (25) for genotyped SNPs and SNPTEST (29) for

imputed SNPs (stage II only). Haplotypes were inferred using the standard expectation maximization algorithm in Haploview (18, 30) and the following confidence interval (CI) criteria: CI minima for strong LD: 0.7 – 0.98; upper CI maximum for strong LD: 0.98; fraction of strong LD in informative comparisons ≥ 0.95; and exclude markers with minor allele frequency (MAF) < 0.05. Haplotypes were tested for association with ALI first in a global association test, which performed contingency testing using all haplotypes of an LD block compared to no haplotypes, and then as individual haplotypes versus ALI coded in an additive fashion PLINK (25). Haplotype multiple testing was addressed by applying permutation tests (10,000 permutations). Significance of odds ratios was determined using the χ^2 test.

Adapting the methods of Satagopan (31) for 2-stage designs, we carried forward the SNPs with the strongest ALI association from our discovery population into the replication stage (stage II). We used a p-value of 10^{-4} to pass stage I (32), rather than 10^{-6} (0.05/50,000 SNPs), given the candidate gene design of the HumanCVD chip and its dense genotyping of covered loci (32). Consensus is lacking for the appropriate significance threshold when using an array containing thousands of hypothesis-driven, densely covered loci. Previous reports using this array have used alpha thresholds of 1×10^{-5}, 5×10^{-5}, and 1×10^{-6} (26, 32, 33), with at least one of these studies reporting no replication (32). We used a more relaxed stage I threshold to balance the concerns of power adequacy with the potential for false positives and considered independent replication the most reliable measure of true association (31). Polymorphisms meeting this cutoff were tested for replication as long as they were present in the EA population with MAF ≥ 0.05. We used logistic regression to adjust for potential confounding by clinical factors (stage I). We used a replication stage to reduce the risk of type I error and applied Bonferroni correction for the number of SNPs tested in replication, such that $\alpha < 0.0167$ was the threshold for significance. Replication at the SNP level was tested using χ^2 statistics assuming an additive model.

Imputed genotypes were determined using MACH 1.0 software (26, 27). To confirm imputation quality, at least 5% of stage II samples were subjected to alternative genotyping methodology (Taqman, SNPlex).

Power Calculations: We had 227 AA trauma subjects for stage I (222 were confirmed AA with < 10% admixture by multidimensional scaling). Given an ALI incidence of approximately 30% in this population (3, 34), we calculated that we had 80% power to detect variants with MAF ≥ 0.25 with a minimum relative risk (RR) of 2.0, at an alpha level of 0.0001. Rarer genetic variants would require a stronger effect size to detect in our derivation populations; a MAF of 0.10 would only be detectable with RR ≥ 2.6. In stage II, with 600 cases and approximately 2000 controls, we calculated 80% power to detect variants with MAF ≥ 0.15 at $p \leq 0.05$. All power calculations were carried out using the Power for Genetic Association analysis package (35). We also performed power calculations for the joint analysis, estimating that we had approximately 75% power to detect variants with MAF = 0.10 and 90% power to detect variants with MAF = 0.40 assuming that we would genotype 10% of samples in stage I and 10% of markers in stage II with a genotype RR of 1.7 (36).

RESULTS

Genotyping Quality Control and Filtering: Stage I: Of 49,094 SNPs on the HumanCVD BeadChip, 352 SNPs failed completely (0.72%). A further 1820 SNPs were monomorphic, leaving 45,102 SNPs for analysis. Filtering criteria resulted in the loss of 797 SNPs with genotype call rate < 95%; 44 SNPs with HWE p-value < 10-4; and 585 SNPs with MAF < 0.01, leaving 42,801 SNPs (**Table S3)**. DNA from 474 subjects was genotyped. Samples were dropped due to genotype call rate < 97.5% (n=4); extreme heterozygosity (n=6); cryptic relatedness (n=3) based on a pi_hat (measure of identity by descent) > 0.3 as implemented in PLINK (25), or if genetic ancestry analysis classified the subjects as non-European non-African (n=26).

This left 435 subjects, with 222 of African ancestry (Stage I) and 213 of European ancestry who were added to the stage II population .

Stage I: Haplotype Associations with ALI

With 2 SNPs in *ANGPT2* demonstrating significant association with ALI, we investigated whether *ANGPT2* haplotypes, or combinations of alleles at different loci, also manifested ALI association. We constructed 9 haplotype blocks assuming a minimum haplotype frequency > 1% (18, 30). As shown in **Figure 2**, 3 of the blocks (blocks 3, 4, and 5) demonstrated considerable linkage between them. The association between LD blocks and ALI was first tested in an omnibus fashion, including all observed haplotypes of a block, and then at the level of individual haplotype with ALI. In omnibus testing, block 1 showed a marginal association with ALI while blocks 3, 4, and 5 each demonstrated significant associations with ALI (**Tables 3 and S4**). Block 4 TCA, in which the first allele is rs1868554T, showed the strongest association with ALI. In comparing haplotype phases between blocks, also called the crossover matrix, carriers of block 4 TCA were almost uniformly carriers of block 5 AAC (99%), while no carriers of block 3 GG haplotype, associated with a decreased risk of ALI, were also carriers of block 4 TCA haplotype (**Table S5**). As shown in the geneview in **Figure 2**, these LD blocks span the end of the first intron and into the second intron of *ANGPT2*. A mini-Manhattan plot of the region demonstrates a total of 4 SNPs, with r^2 ranging from 0.33 to 0.58 with rs1868554, in *ANGPT2* associated with ALI at $p < 0.005$ (Supplementary figure S2).

Sensitivity Analysis of the ALI definition. We have previously reported an improved ability to detect ALI risk factors by excluding patients with an "equivocal" or "difficult to classify" ALI diagnosis (37). When this approach was taken in our stage I population, the number of non-ALI individuals decreased from 161 to 95. **Table S6** displays the association of the two *ANGPT2* SNPs with ALI. The odds ratios, 95%

confidence intervals, and p-values remained significant, with a minimal attenuation of odds ratio despite losing more than a third of the control population.

Site-stratified analysis for stage II. To evaluate for heterogeneity among TASC sites for the rs1868554 – ALI association, we stratified results by site. A formal test for interaction could not be performed as the control group was not enrolled at each site; rather, a single control group was used for all sites. This likewise made any meta-analysis problematic. As displayed in **Table S7**, the direction of effect was consistent across all sites.

Sequencing and in silico splice site analysis. The results of sequencing in 24 individuals of AA and EA ancestry is displayed in **Table S8**. Sequencing was inadequate for approximately 8 individuals of each race, decreasing our effective sample size to 16 EA and 16 AA individuals. We identified 87 novel SNPs. No coding SNPs or large structural variations (insertion / deletions) were identified. We performed linkage analysis in Haploview for each ancestry to determine the LD of sequenced variants with rs1868554.

Splice site analysis of the *ANGPT2* sequence 1000 bp upstream and downstream of exon 2 suggested numerous splice site alterations – predicted creation of novel sites, novel enhancers, or disrupted splice sites or disrupted enhancers – for individual SNPs, as depicted in **Table S9.**

Functional Assessment. Plasma ANG2 ELISA levels were higher in ALI subjects: mean 9292 ± 8832 ng/ml versus 4814 ± 2409 ng/ml, p=0.0041 by Wilcoxon ranksum test (**Figure S3A**). ANG2 did not vary according to rs1868554 genotype (**Figure S3B**), with a wide range of values for each genotype (p=0.84 by Kruskal Wallis test).

ACE (†)	IL1B	NQO1 (†)	SFTPD
ANGPT2 (†)	IL1RN	NRF2 (†)	STAT1
CC16	IL6 (†)	PBEF1 (†)	STAT3
CD14	IL8 (†)	*PI3* (†)	SOD3 (†)
CXCL2	IL10 (†)	PLAU (†)	TLR1 (†)
EGF (†)	MBL2 (†)	PAI1 (SEPINE1) (†)	TNFA (†)
F5 (†)	MIF (†)	PROC	LTA
FAS (†)	MYLK (†)	PROCR (EPCR)	VEGFA (†)
FTL (†)	NFKB1 (†)	SFTPA	VWF
HMOX2 (†)	NFKBIA (†)	*SFTPB* (†)	

Table S1: Coverage of ALI candidate genes on the HumanCVD BeadChip. Out of 39 candidate genes with stong evidence for lung injury, inflammation, or vascular regulation, 4 (10%) listed in italics and underlined had no coverage on the Chip. Twenty-six of these genes, indicated with (†), have been associated with ALI susceptibility or outcome in published reports, and the chip genotyped 22 of these genes. With the exception of *MYLK,* which was genotyped only for nonsynonymous or known functional variants with minor allele frequency (MAF) ≥ 0.01, all genes present on the HumanCVD chip were genotyped by loci selected to capture at least minor allele frequency (MAF) > 0.05 with genetic coverage (r^2) ≥ 0.50. Candidate genes and pathways were described in 2 recent reviews (38, 39) as well as a 9/2/2010 Pubmed search for “genetic variation” and “acute lung injury / ARDS”. While this list is comprehensive for the published associations with ALI as of 9/2/2010, there may be additional candidate genes without published ALI associations that are not listed here.

Table S2:

Forward Strand	Reverse Strand
5'-TTGTCGAGAGGGAGTGTTCC-3'	5'-TATTTGGCACAGACCCCATT-3'
5'-CCTGACCTCAAGGGATTCAC-3'	5'-CTTGGCAACAATGCTGACTG-3'
5'-GATTCTCTAGCGCCTGGTTG-3'	5'-GCAAGATTCTCCAAGCCTGT-3'
5'-TTTGGCCTACGTCTTCTTTGA-3'	5'-GCGTGCTTGTGTGCACTTAT-3'
5'-GGTGGCACCATATCTGTTCA-3'	5'-CATGCCCCACCTTTTGATAC-3'
5'-AAGGTTTGCATTGCTGCTCT-3'	5'-GGTTGGACCAAAAGTAACTGC-3'
5'-AGGTTGCAGCAAGTCGAGAT-3'	5'-CCGCTTTCTGACTCCTCATC-3'
5'-ATGGAGTGGGGATTGTTTCA-3'	5'-GTTGGGCCAGACAAGTGATT-3'
5'-CTCAGGTGATCCACCCATCT-3'	5'-CCACCACAGGAGAAAAGGAA-3'
5'-CACAGCCGTCTGGTTCTGTA-3'	5'-GGACCAAGGTGCGATTAAAA-3'
5'-GTACGTGGGAAGAGCCTTTG-3'	5'-GAGGCAGCCTTAGTGACCAG-3'
5'-GATGAGCAGGATGCATGAGA-3'	5'-AGACAGTGCCTCTGGGTGAC-3'
5'-GCACCCCAGCCAAGATAATA-3'	5'-CTCCCAAAGTGCTGGGATTA-3'
5'-ATGGCTCATGCCTGTAATCC-3'	5'-CTGGGCTTTTCTGGTGGTAA-3'
5'-CCAATTCCAGGCTTCACAAT-3'	5'-GTGCCTACAGATCACGCTCA-3'

Table S2: Primers used for sequencing 5 kb upstream and downstream of the *ANGPT2* 2nd exon.

Filtering Criterion	**SNPs**	**Remaining SNPs of 49,094 on BeadChip**
SNP failure	352	48,742
Monomorphic	1820	46,922
Genotype call rate < 95%	797	46,125
HWE p-value < 10E-04	44	46,081
MAF < 0.01	1460	44,621
Filtering Criterion	**Samples**	**Remaining Samples of 474**
Genotype call rate < 97.5%	4	470
Extreme heterozygosity	6	464
Cryptic Relatedness/Duplicates	3	461
PCA non-African Non-European	26	435
PCA European	213	222

Table S3: Filtering results of stage I genotyping. For the stage I analysis, 44,621 SNPs were analyzed for 222 African ancestry (AA) subjects. *HWE* Hardy Weinberg equilibrium; *PCA* principal components analysis.

LD Block	SNPS (in order)	Haplotype	Case freq	Control freq	Chi square (df)†	p-value	Permutation p-value
1	rs17077194/ rs2442468/ rs2515416/ rs6559166	Omnibus	NA	NA	9.87 (4)	0.043	
		CCAC	0.20	0.32	6.61	0.010	0.20
		GGGC	0.42	0.30	5.68	0.017	0.28
		CCAG	0.17	0.13	1.07	0.30	1.0
		GCGC	0.16	0.19	0.41	0.52	1.0
		CCGC	.05	.06	0.11	0.74	1.0
2	rs2515428/ 2442621	Omnibus	NA	NA	4.35 (2)	0.11	
		GA	0.03	0.09	4.05	0.044	0.58
		CG	0.93	0.86	3.55	0.059	0.69
		CA	0.04	0.05	0.18	0.67	1.0
3	rs2515466/ rs2442621	Omnibus	NA	NA	10.45 (2)	**0.0054**	
		GG	0.40	0.57	10.4	**0.0013**	**0.031**
		GA	0.42	0.30	6.37	0.012	0.21
		AA	0.18	0.13	1.46	0.023	1.0
4	rs1868554/ rs1050337 / rs1264822	Omnibus	NA	NA	17.39 (2)	**0.0017**	
		TCA	0.50	0.28	17.29	**$3.2\ e^{-5}$**	**0.0013**
		ACA	0.27	0.41	6.59	0.010	0.17
		AGC	0.23	0.31	2.64	0.10	0.84
5	rs1984857/ rs17077419/ rs2959813	Omnibus	NA	NA	17.31 (3)	**0.00061**	
		AAC	0.52	0.32	16.28	**$5.5\ e^{-5}$**	**0.0016**
		GAC	0.17	0.30	7.47	0.0063	0.13
		GAG	0.11	0.15	1.15	0.28	1.0
		AGC	0.20	0.24	0.83	0.36	1.0
6	rs13255574/ rs1988762	Omnibus	NA	NA	0.16 (2)	0.92	
		GG	0.55	0.53	0.15	0.70	1.0
		GA	0.34	0.36	0.07	0.79	1.0
		CG	0.11	0.11	0.05	0.83	1.0
7	rs2515488/ rs2442595	Omnibus	NA	NA	0.58 (2)	0.0054	
		CG	0.52	0.55	0.50	0.48	1.0
		AA	0.28	0.24	0.45	0.50	1.0
		AG	0.21	0.20	0.02	0.88	1.0
8	rs6991221/ rs2959764	Omnibus	NA	NA	4.11 (2)	0.13	
		GC	0.45	0.35	3.96	0.047	0.60
		AG	0.50	0.58	2.42 (2)	0.12	0.95

		AC	0.05	0.07	0.61	0.43	1.0
9	rs10108504/ rs2515509	Omnibus	NA	NA	4.88 (2)	0.087	
		AA	0.52	0.40	4.70	0.03	0.44
		AG	0.38	0.45	1.88	0.17	0.99
		GA	0.11	0.15	1.32	0.25	1.0

Supplementary Table S4: *ANGPT2* Haplotype blocks and ALI association. The 9 haplotype blocks present in the stage I population with greater than 1% frequency are shown. The SNPs defining each block are listed in the same order as the alleles in column 3. For each haplotype block, we first performed a global test of association with ALI which included all specific haplotypes of the block compared to a model without haplotypes. These were chi square tests with the degree of freedom equal to the number of haplotypes – 1, which is shown in parentheses for each block. For individual haplotypes, the chi square test of association assumed 1 degree of freedom. The final column displays the result of permutation testing based on 10,000 permutations of the haplotype blocks to control for multiple comparisons. Three haplotypes remain significant after permutation testing; block 4 TCA, block 5 AAC, and block 3 GG. These 3 haplotypes are not independent; of carriers with the TCA haplotype (block 4), 99% carry haplotype AAC (block 5) and 0% carry the haplotype associated with decreased risk of ALI, haplotype GG (block 3).

	Block 5:AAC	Block 5:GAC	Block 5:AGC	Block5: GAG	Freq
Block 4:ACA	0.034	0.203	0.0	0.135	0.372
Block 4: TCA	**0.337**	0.0	0.0	0.002	0.339
Block 4: AGC	0.0	0.06	0.225	0.0	0.285
Frequency	0.371	0.263	0.225	0.137	

Table S5A: Stage I haplotype crossover matrix for Blocks 4 and 5 (A) and 3 and 4 (B). For each haplotype block, the population frequency is shown at the end of the column or row, shaded in grey. The rest of the table displays the crossover percentages for carriers of the two blocks which intersect at that cell. For example, block 4 TCA occurs in 33.9% of the population overall. For blocks 4 and 5, 33.7% of the population have the haplotype combination TCA – AAC and 0.2% are TCA – GAG. Haplotypes associated with ALI in this population are displayed in **bold**. For blocks 3 and 4, 0% of the population carry GG – TCA. The block 4 TCA haplotype is observed only with GA or AA carriers. Block 3 GG was associated with a decreased risk of ALI (OR 0.50, 95% confidence interval (0.32 – 0.78), whereas carriers of block 4 TCA haplotype had an increased risk of ALI (OR 2.43 95% confidence interval (1.54 – 3.82). Note: population frequencies sum to slightly less than 100% due to the exclusion of haplotypes with population frequency < 0.01.

Freq: population frequency.

	Block 3:GG	Block 3:GA	Block 3:AA	Freq
Block 4:ACA	0.337	0.015	0.021	0.372
Block 4: TCA	**0.0**	0.287	0.052	0.339
Block 4: AGC	0.192	0.026	0.067	0.285
Frequency	0.529	0.328	0.14	

Table S5B: Stage I haplotype crossover matrix for Blocks 3 and 4. Note: population frequencies sum to slightly less than 100% due to the exclusion of haplotypes with population frequency < 0.01.

Freq: population frequency.

SNP	Gene	Case Genotypes	Non-ALI Genotypes	HWE p	OR (95% CI)	p-value (additive)
rs1868554	ANGPT2	21/19/20	4/48/43	0.37	2.40 (1.47, 3.90)	0.00043
rs2442598	ANGPT2	15/21/24	2/38/55	0.30	2.49 (1.50, 4.13)	0.00043

Table S6: Sensitivity analysis of the ALI non-case definition on the association of *ANGPT2* SNPs with ALI. Both stage I *ANGPT2* SNPs remain significantly associated with ALI after removal of approximately 40% of the "equivocal" controls.

	U Washington	Penn	Vanderbilt	Harvard	UCSF	CHOP
MAF	0.320	0.385	0.375	0.308	0.400	0.286
Genotypes	44 / 198 /205	10 / 30 /25	8 / 16 / 26	2 / 12 / 12	1 / 6 / 3	171/ 873/ 1083
Odds ratio (95% CI)	1.18 (1.01, 1.38)	1.57 (1.09, 2.25)	1.34 (0.88, 1.99)	1.11 (0.61, 2.01)	1.67 (0.68, 4.12)	Reference
p-value	0.026	0.015	0.18	0.73	0.26	--

Table S7: Stage II rs1868554 results stratified by TASC site. For each site, the rs1868554 genotype count is displayed in the following order, with T being the minor allele: TT / AT/ AA. The odds ratio, 95% confidence interval, and p-value assume an additive model of risk. TASC sites are abbreviated as follows: Washington (University of Washington); Penn (University of Pennsylvania); Vanderbilt (Vanderbilt University); Harvard (Harvard University / Massachusetts General Hospital); UCSF (University of California, San Francisco). *MAF:* minor allele frequency.

Table S8: Variants detected via direct sequencing

rs Number	Location	Chromosome Position (dbSNP Build 131)	Reference Allele	Base Change	Number of Variants	LD (r^2) with rs1868554: EA	LD (r^2) with rs1868554: AA
	intron	6385190	T	A/T	1		
	intron	6385192	T	A/T	1		
	intron	6385435	G	G/T	1		
rs71535978	intron	6385589.1	:	:/T	1		
rs71525734	intron	6385618	A	A/T	4	1.0	
	intron	6385658	T	G/T	1		
	intron	6385659	T	G/T	1		
	intron	6385660	C	G/C	1		
	intron	6385661	A	A/G	1		
	intron	6385662	G	G/C	1		
rs2442610	intron	6385679	T	A/T	28		
	intron	6385720	T	C/T	1		
	intron	6385721	T	C/T	1		
rs58191715	intron	6385779	C	G/C	2		
rs2515464	intron	6385839	T	G/T	39		
rs2515465	intron	6385862	G	G/C	42		
	intron	6385905	A	A/G	1		
	intron	6386081	G	A/G	1		
	intron	6386095	G	G/T	1		
rs74344164	intron	6386110	C	C/T	1		
rs2922887	intron	6386127	G	G/C	31		
	intron	6386131	C	C/T	5		
rs6990020	intron	6386204	C	C/T	35		
rs3824310	intron	6386250	T	C/T	4	0.40	
	intron	6386275.1	:	:/C	1		
rs2515466	intron	6386286	A	A/G	30		0.471
rs2515467	intron	6386358	G	A/G	5		
rs3824312	intron	6386441	G	A/G	16	0.425	
rs2515468	intron	6386483	C	G/C	6		
	intron	6386539	G	A/G	1		
	intron	6386541	A	A/C	1		
	intron	6386542	C	C/T	1		
rs2922886	intron	6386584	G	G/C	22		
rs2442609	intron	6386593	T	C/T	24	0.536	
rs2515469	intron	6386599	T	G/T	22	0.486	
	intron	6386613	G	A/G	2		
rs2442608	intron	6386620	T	C/T	26	0.486	
rs59017488	intron	6386743	G	A/G	2		
rs1868554	intron	6386747	T	A/T	29		
rs7825407	intron	6386780	C	G/C	16		
rs1868553	intron	6386824	A	A/C	10		
rs75140394	intron	6386974	C	C/T	6		
	intron	6387019	C	C/G	1		
	intron	6387028	G	A/G	2		
rs13268979	intron	6387042	T	C/T	3		
rs1868552	intron	6387118	T	C/T	6		
	intron	6387266	C	C/T	1		

	intron	6387288	C	G/C	1		
rs2442607	intron	6387308	T	C/T	1		
	intron	6387353	C	C/T	2		
	intron	6387402	C	G/C	1		
	intron	6387504	A	A/G	1		
rs1807209	intron	6387556	T	C/T	7		
rs2515471	intron	6387602	G	A/G	12		
	intron	6387758	G	A/G	1		
rs2256628	intron	6387845	G	A/G	31	1.0	0.633
	intron	6387922	C	C/T	2		
rs2442606	intron	6387923	G	A/G	16		
rs2922885	intron	6387938	G	G/T	5		
rs2515472	intron	6387953	C	C/T	14		
	intron	6387991	C	C/T	33	1.0	0.636
	intron	6388011	T	C/T	4		
	intron	6388012	G	A/G	4		
	intron	6388013	C	G/C	4		
rs71213314	intron	6388049	A	A/G	1		
	intron	6388086	A	A/T	1		
rs3045059	intron	6388095.1	:	:/A	7		
	intron	6388095.2	:	:/A	7		
rs34288136	intron	6388096.1	:	:/T	7		
rs734703	intron	6388155	A	A/G	8	1.0	
rs746073	intron	6388184	G	A/G	15	1.0	1.0
rs734702	intron	6388213	T	C/T	14	1.0	0.455
rs734701	intron	6388247	G	A/G	14	1.0	0.455
	intron	6388249	C	G/C	1		
rs734704	intron	6388306	A	A/G	14	1.0	0.455
	intron	6388375	G	A/G	1		
rs2515475	intron	6388439	C	C/T	1		
rs2442605	intron	6388463	T	A/T	1		
rs2515477	intron	6388640	C	C/T	2		
rs73507122	intron	6388707	G	A/G	3		
	intron	6388708	C	A/C	1		
	intron	6388723	T	G/T	1		
rs10503371	intron	6388782	G	G/C	15	0.72	0.60
	intron	6388814	C	C/T	1		
	intron	6388817.1	:	:/C	1		
	intron	6388818	T	C/T	3		
	intron	6388924	C	G/C	1		
	intron	6388962	A	A/C	3		
	intron	6388963	C	A/C	1		
	intron	6388967	A	A/T	1		
	intron	6388971	G	G/T	1		
	intron	6388973	C	A/C	2		
	intron	6388975	A	A/G	1		
	intron	6388979	T	A/T	1		
rs2515478	intron	6388980	G	G/T	21	0.583	0.294
	intron	6388981	T	C/T	1		
rs2442603	intron	6389083	A	A/G	14		
	intron	6389085	T	C/T	1		
	intron	6389086	T	A/T	1		
rs2515479	intron	6389120	G	G/T	25		

	intron	6389163	T	C/T	1		
rs12674822	intron	6389216	T	G/T	20	0.743	0.206
	intron	6389244	A	A/G	2		
rs2515480	intron	6389280	G	A/G	22		0.267
rs5889176	intron	6389323	A	A/:	13		
rs1984860	intron	6389383	A	A/T	14		0.238
rs1984859	intron	6389424	T	C/T	15		0.238
rs17077416	intron	6389495	A	A/G	4		
	intron	6389599	G	A/G	1		
	intron	6389714	T	C/T	1		
rs2515481	intron	6390027	A	A/G	7		
	intron	6390028	G	A/G	1		
	intron	6390028.1	:	:/G	1		
rs2515482	intron	6390079	G	A/G	28	1.0	
	intron	6390113.1	:	:/A	1		
rs66589203	intron	6390142.1	:	:/C	3		
	intron	6390156	C	A/C	1		
rs1031304	intron	6390162	T	C/T	5		
	intron	6390278	G	G/C	1		
rs1031303	intron	6390337	G	G/T	30	1.0	
	intron	6390514	A	A/G	1		
rs17077419	intron	6390530	A	A/G	15	0.491	
	intron	6390966	A	A/G	1		
rs2959813	intron	6391182	C	G/C	5		
	intron	6391263	C	C/T	2		
rs2959812	intron	6391302	A	A/G	33	1.0	
	intron	6391391	A	A/T	1		
	intron	6391395	T	C/T	1		
rs2959810	intron	6391726	T	A/T	5	1.0	
	intron	6391727	A	A/T	1		
	intron	6391736	A	A/:	1		
rs2922881	intron	6391767	T	C/T	18		0.474
	intron	6391771	A	A/T	1		
rs12115093	intron	6391785	C	G/C	1		
	intron	6391787	A	A/C	1		
	intron	6391797	T	G/T	1		
rs12550255	intron	6391950	C	G/C	6		
rs12676103	intron	6392186	A	A/G	7		
	intron	6392236	A	A/G	1		
rs2442601	intron	6392243	G	A/G	4		
rs2515483	intron	6392269	G	G/C	16		0.325
rs17623064	intron	6392298	T	A/T	6		
rs75195933	intron	6392359	C	G/C	7		
	intron	6392398.1	:	:/A	1		
rs2442599	intron	6394151	G	A/G	33	1.0	
rs17552444	intron	6394179	A	A/G	14		
	intron	6394344	G	G/T	1		
	intron	6394345	A	A/T	1		
	intron	6394346	A	A/T	1		
	intron	6394349	A	A/T	1		
	intron	6394350	G	G/T	3		
	intron	6394366	T	C/T	1		
	intron	6394371	C	A/C	1		

	intron	6394374	G	A/G	1		
	intron	6394389	T	A/T	1		
	intron	6394430	A	A/C	1		
rs12550622	intron	6394494	A	A/G	2		1.0
	intron	6394515	A	A/C	1		
	intron	6394523	A	A/G	1		
	intron	6394548	C	G/C	1		
	intron	6394687	G	G/T	1		

Table S8: *ANGPT2* variants detected by direct sequencing in the 10 kb surrounding exon 2. The Haploview-determined LD with rs1868554 (r^2) is shown for European (EA) and African (AA) ancestry. When no reference sequence (refseq, rs) number is listed, the polymorphism has not previously been described and is novel to this sequencing attempt. All novel SNPs will be submitted to dbSNP. Eighty-seven novel SNPs were discovered, most with low frequency (0-3 occurrences). The one notable exception was for the novel SNP at bp (underlined), which was observed in 33 individuals and exhibited significant LD with rs1868554 in both EA and AA populations.

SNP	HSF potential splice sites	ESE	RESCUE-ESE	PESE	EIE
rs2442602	Site broken: 40%	New site	----	----	----
rs17077419	Site broken: 36%	Site broken	----	----	----
rs12681141	Site broken: 30%	----	----	----	Site broken
rs1031303	New site: 60%	Site broken	----	New site	Site broken
rs1301304	New site: 78%	----	New site	----	----
rs28460445	New site: 64%	New site	Site broken	Site broken	Site broken
rs2515482	No variation	----	Site broken	Site broken	New site
rs2515481	Minor variation 10%	Site broken	----	----	New site
rs6559167	Minimal variation <1%	Site broken	----	----	Site broken
rs1984857	New site: 70%	----	New site	New site	New site
rs1984858	Minimal variation <1%	Site broken	----	----	Site broken
rs17077416	Minimal variation <1%	New site	Site broken	----	Site broken
rs1984859	No variation	New site	New site	New site	New site
rs1984860	Minimal variation <5%	Site broken	New site	----	New site
rs2515480	Minimal variation <1%	----	----	----	----
rs12674822	Minor variation 13%	New site	----	----	----
rs2515479	Minor variation 8%	Site broken	----	----	Site broken
rs2442603	Minimal variation <5%	New site	----	----	New site
rs2515478	New site: 70%	----	New site	New site	New site

Table S9: *In silico* SNP analysis of predicted splice site alteration. Using HSF 2.4.1, a web-enabled open access resource, the reference ANGPT2 isoform a sequence (ENST00000325203) was searched 1000 bp upstream and downstream of the 2nd exon. Nineteen SNPs in this region were returned for SNP analysis. HSF reports the variation caused by each SNP with reference to the consensus values (CV), based on a similarity matrix, to the natural consensus 9-mer 5' and 14-mer 3' (donor and acceptor, respectively) splice sites. Sites with strong positive variation relative to the reference sequence are shaded green for "new splice site," whereas those with strong negative variation are shaded red for "site broken." The last 4 columns report the results of position weight matrix programs which grade similarity to splice enhancer elements, rather than intron – exon boundaries (20-23). Details of each matrix are presented in the Supplemental methods section. SNPs shaded in grey were not observed in our sequencing, while SNPs in yellow demonstrated LD with rs1868554 in EA or both populations (rs2515478 only).

Figure S1: Principal components analysis of all stage I subjects following multidimensional scaling. Results were adjusted for 2 principal components after the MDS procedure.

Figure S2: Regional association plot depicting the association between the *ANGPT2* gene in stage I and the development of ALI. Association is depicted on the *y*-axis as the negative log (p-value) by χ^2 test assuming an additive model. Four SNPs including rs1868554 and rs2442598 were associated with ALI at $p< 10^{-3}$. Each ALI-associated SNP demonstrates at least marginal LD ($r^2 > 0.3$) with rs1868554 in this population, as indicated by the intensity of red color. In blue, the background recombination rate for the HapMap YRI population is depicted. Figure was created using the SNP Annotation and Proxy (SNAP) program (40).

Figure S3A: Plasma ANG2 is increased in ALI subjects. Consistent with published reports, we found ALI subjects to have significantly higher plasma ANG2 levels by ELISA (p=0.0041).

Figure S3B: Plasma ANG2 concentration is unrelated to rs1868554 genotype (p=0.85).

REFERENCES:

1. Christie JD, Gaughan C, Gallop R. Clinical risk factors of development of ards in a cohort study of patients with major trauma. *Am J Respir Crit Care Med* 2003;167:A740.

2. Civil ID, Schwab CW. The abbreviated injury scale, 1985 revision: A condensed chart for clinical use. *J Trauma* 1988;28(1):87-90.

3. Shah CV, Localio AR, Lanken PN, Kahn JM, Bellamy S, Gallop R, Finkel B, Gracias VH, Fuchs BD, Christie JD. The impact of development of acute lung injury on hospital mortality in critically ill trauma patients. *Crit Care Med* 2008;36(8):2309-2315.

4. Keating BJ, Tischfield S, Murray SS, Bhangale T, Price TS, Glessner JT, Galver L, Barrett JC, Grant SF, Farlow DN, et al. Concept, design and implementation of a cardiovascular gene-centric 50 k snp array for large-scale genomic association studies. *PLoS ONE* 2008;3(10):e3583.

5. Chanock SJ, Manolio T, Boehnke M, Boerwinkle E, Hunter DJ, Thomas G, Hirschhorn JN, Abecasis G, Altshuler D, Bailey-Wilson JE, et al. Replicating genotype-phenotype associations. *Nature* 2007;447(7145):655-660.

6. Gong MN, Zhou W, Williams PL, Thompson BT, Pothier L, Christiani DC. Polymorphisms in the mannose binding lectin-2 gene and acute respiratory distress syndrome. *Crit Care Med* 2007;35(1):48-56.

7. Su L, Zhai R, Sheu CC, Gallagher DC, Gong MN, Tejera P, Thompson BT, Christiani DC. Genetic variants in the angiopoietin-2 gene are associated with increased risk of ards. *Intensive Care Med* 2009.

8. Wang Z, Beach D, Su L, Zhai R, Christiani DC. A genome-wide expression analysis in blood identifies pre-elafin as a biomarker in ards. *Am J Respir Cell Mol Biol* 2008;38(6):724-732.

9. Cohen MJ, Brohi K, Calfee CS, Rahn P, Chesebro BB, Christiaans SC, Carles M, Howard M, Pittet JF. Early release of high mobility group box nuclear protein 1 after severe trauma in humans: Role of injury severity and tissue hypoperfusion. *Crit Care* 2009;13(6):R174.

10. Shalhub S, Junker CE, Imahara SD, Mindrinos MN, Dissanaike S, O'Keefe GE. Variation in the tlr4 gene influences the risk of organ failure and shock posttrauma: A cohort study. *Journal of Trauma-Injury Infection & Critical Care* 2009;66(1):115-123.

11. Fremont RD, Koyama T, Calfee CS, Wu W, Dossett LA, Bossert FR, Mitchell D, Wickersham N, Bernard GR, Matthay MA, et al. Acute lung injury in patients with traumatic injuries: Utility of a panel of biomarkers for diagnosis and pathogenesis. *J Trauma* 2009.

12. Glessner JT, Wang K, Cai G, Korvatska O, Kim CE, Wood S, Zhang H, Estes A, Brune CW, Bradfield JP, et al. Autism genome-wide copy number variation reveals ubiquitin and neuronal genes. *Nature* 2009;459(7246):569-573.

13. Hakonarson H, Grant SF, Bradfield JP, Marchand L, Kim CE, Glessner JT, Grabs R, Casalunovo T, Taback SP, Frackelton EC, et al. A genome-wide association study identifies kiaa0350 as a type 1 diabetes gene. *Nature* 2007;448(7153):591-594.

14. Sleiman PM, Flory J, Imielinski M, Bradfield JP, Annaiah K, Willis-Owen SA, Wang K, Rafaels NM, Michel S, Bonnelykke K, et al. Variants of dennd1b associated with asthma in children. *N Engl J Med*;362(1):36-44.

15. Gunderson KL, Steemers FJ, Lee G, Mendoza LG, Chee MS. A genome-wide scalable snp genotyping assay using microarray technology. *Nat Genet* 2005;37(5):549-554.

16. Steemers FJ, Chang W, Lee G, Barker DL, Shen R, Gunderson KL. Whole-genome genotyping with the single-base extension assay. *Nat Methods* 2006;3(1):31-33.

17. Kruglyak L, Nickerson DA. Variation is the spice of life. *Nat Genet* 2001;27(3):234-236.

18. Barrett JC, Fry B, Maller J, Daly MJ. Haploview: Analysis and visualization of ld and haplotype maps. *Bioinformatics* 2005;21(2):263-265.

19. Desmet FO, Hamroun D, Lalande M, Collod-Beroud G, Claustres M, Beroud C. Human splicing finder: An online bioinformatics tool to predict splicing signals. ***Nucleic Acids Res*** **2009;37(9):e67.**
20. Fairbrother WG, Yeh RF, Sharp PA, Burge CB. Predictive identification of exonic splicing enhancers in human genes. ***Science*** **2002;297(5583):1007-1013.**
21. Zhang XH, Chasin LA. Computational definition of sequence motifs governing constitutive exon splicing. ***Genes Dev*** **2004;18(11):1241-1250.**
22. Zhang C, Li WH, Krainer AR, Zhang MQ. Rna landscape of evolution for optimal exon and intron discrimination. ***Proc Natl Acad Sci U S A*** **2008;105(15):5797-5802.**
23. Cartegni L, Wang J, Zhu Z, Zhang MQ, Krainer AR. Esefinder: A web resource to identify exonic splicing enhancers. ***Nucleic Acids Res*** **2003;31(13):3568-3571.**
24. Atochina-Vasserman EN, Gow AJ, Abramova H, Guo C-J, Tomer Y, Preston AM, Beck JM, Beers MF. Immune reconstitution during pneumocystis lung infection: Disruption of surfactant component expression and function by s-nitrosylation. ***J Immunol*** **2009;182(4):2277-2287.**
25. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Maller J, Sklar P, de Bakker PI, Daly MJ, et al. Plink: A tool set for whole-genome association and population-based linkage analyses. ***Am J Hum Genet*** **2007;81(3):559-575.**
26. Cappola TP, Li M, He J, Ky B, Gilmore J, Qu L, Keating B, Reilly M, Kim CE, Glessner J, et al. Common variants in hspb7 and frmd4b associated with advanced heart failure. ***Circ Cardiovasc Genet*****;3(2):147-154.**
27. Falush D, Stephens M, Pritchard JK. Inference of population structure using multilocus genotype data: Linked loci and correlated allele frequencies. ***Genetics*** **2003;164(4):1567-1587.**
28. Luca D, Ringquist S, Klei L, Lee AB, Gieger C, Wichmann HE, Schreiber S, Krawczak M, Lu Y, Styche A, et al. On the use of general control samples for genome-wide association studies: Genetic matching highlights causal variants. ***Am J Hum Genet*** **2008;82(2):453-463.**
29. Marchini J, Howie B, Myers S, McVean G, Donnelly P. A new multipoint method for genome-wide association studies by imputation of genotypes. ***Nat Genet*** **2007;39(7):906-913.**
30. Gabriel SB, Schaffner SF, Nguyen H, Moore JM, Roy J, Blumenstiel B, Higgins J, DeFelice M, Lochner A, Faggart M, et al. The structure of haplotype blocks in the human genome. ***Science*** **2002;296(5576):2225-2229.**
31. Satagopan JM, Elston RC. Optimal two-stage genotyping in population-based association studies. ***Genet Epidemiol*** **2003;25(2):149-157.**
32. Talmud PJ, Drenos F, Shah S, Shah T, Palmen J, Verzilli C, Gaunt TR, Pallas J, Lovering R, Li K, et al. Gene-centric association signals for lipids and apolipoproteins identified via the humancvd beadchip. ***The American Journal of Human Genetics*** **2009;85(5):628-642.**
33. Clarke R, Peden JF, Hopewell JC, Kyriakou T, Goel A, Heath SC, Parish S, Barlera S, Franzosi MG, Rust S, et al. Genetic variants associated with lp(a) lipoprotein level and coronary disease. ***N Engl J Med*** **2009;361(26):2518-2528.**
34. Netzer G, Shah CV, Iwashyna TJ, Lanken PN, Finkel B, Fuchs B, Guo W, Christie JD. Association of rbc transfusion with mortality in patients with acute lung injury. ***Chest*** **2007;132(4):1116-1123.**
35. Menashe I, Rosenberg PS, Chen BE. Pga: Power calculator for case-control genetic association analyses. ***BMC Genet*** **2008;9:36.**
36. Skol AD, Scott LJ, Abecasis GR, Boehnke M. Optimal designs for two-stage genome-wide association studies. ***Genet Epidemiol*** **2007;31(7):776-788.**
37. Shah CV, Lanken PN, Localio AR, Gallop R, Bellamy S, Ma SF, Flores C, Kahn JM, Finkel B, Fuchs BD, et al. An alternative method of acute lung injury classification for use in observational studies. ***Chest*****.**

38. Flores C, Pino-Yanes Mdel M, Villar J. A quality assessment of genetic association studies supporting susceptibility and outcome in acute lung injury. *Crit Care* 2008;12(5):R130.
39. Gao L, Barnes KC. Recent advances in genetic predisposition to clinical acute lung injury. *Am J Physiol Lung Cell Mol Physiol* 2009;296(5):L713-725.
40. Johnson AD, Handsaker RE, Pulit SL, Nizzari MM, O'Donnell CJ, de Bakker PI. Snap: A web-based tool for identification and annotation of proxy snps using hapmap. *Bioinformatics* 2008;24(24):2938-2939.