Online Data Supplement

Interleukin-1 Receptor-associated Kinase 3 Gene Associates with Susceptibility to Acute Lung Injury

María Pino-Yanes, Shwu-Fan Ma, Xiaoguang Sun, Paula Tejera, Almudena Corrales, Jesús Blanco, Lina Pérez-Méndez, Elena Espinosa, Arturo Muriel, Lluis Blanch, Joe G. N. Garcia, Jesús Villar, Carlos Flores

METHODS

Study design

Sample characteristics have been reported elsewhere (E1) consisting on unrelated cases and controls of Spanish descent. Cases fulfilled the international criteria for severe sepsis (E2) and basic demographic data, severity of illness scores and clinical information until discharge from the ICU, including source of infection and development of organ dysfunction, were recorded. All patients were followed prospectively for the development of ALI, as defined by the American-European Consensus Conference (E3). The median APACHE II score was 24 (IQR=18-27) and 184 patients (86%) developed ALI/ARDS. The overall ICU mortality was 45.5%. In agreement with previous studies (E4), no pathogens were identified in blood cultures as the causative microorganism for sepsis albeit having an identified site of infection for 50.8% of patients. Controls were randomly drawn from a population-based study of about 7000 samples from unrelated individuals as a representation of the general Spanish population (E5). A health survey consisting on a 2-hour personal interview by trained personnel aiming to characterize occupational and environmental risk factors for common acute and chronic health conditions (supported by personal or familiar medical records) in the population was obtained from all controls. The study sample size was based on *a priori* power calculation for the primary outcome, indicating 80% power to detect a risk of 1.8 at p=0.05 significance level assuming an allele frequency of 30%.

IRAK3 gene polymorphism identification

To study the variation in the *IRAK3* gene, 22.6 kilobases (kb) of non-repetitive gene regions in DNA samples were sequenced in 32 unrelated healthy Spanish individuals. The genomic sequence from the reference assembly (NC_000012.10) was filtered for interspersed repeats and low complexity DNA sequences using RepeatMasker (http://www.repeatmasker.org). Primers for DNA amplification of non-repetitive regions were designed using Primer3 (http://frodo.wi.mit.edu/cgi-bin/primer3/primer3_www.cgi). Primer pairs used are listed in Table E1. Sequencing was performed by both strands with BigDye® v3.1 Terminator Cycle Sequencing kit (Applied Biosystems, Foster City, CA, USA) under the manufacturer's recommendations using the amplification primers. Sequencing products were ethanol precipitated and run on an ABI PRISM 310 Genetic Analyzer (Applied Biosystems). Polymorphisms were identified by manual examination of traces. The sequence variants were confirmed by two independent authors and reported following established guidelines (E6).

tSNPs selection

In order to efficiently and accurately study common *IRAK3* gene variation in the population, tagging single nucleotide polymorphisms (tSNPs) were

selected using TagIT 3.03 using a SNP-dropping-with-re-sampling method (E7) and tested for their ability to impute untyped variants. Given a genomic region, the allelic frequencies of untyped variants can be imputed using a set of tSNPs with high information content of the variation in the particular region, by means of linear combinations of the tSNPs haplotype frequencies. TUNA software was used to derive such estimates (E8). TUNA also allows estimating the M_D parameter, representing a multilocus measure of linkage disequilibrium (LD) (E9). This value indicates the amount of information that a set of tSNPs contains for a given untyped SNP, for which a threshold ≥0.7 has been suggested for accurate predictions to avoid power loss to detect disease associations at untyped SNPs (E8).

In order to infer the association only at those untyped SNPs for which imputation was accurate, we evaluated the accuracy of the selected 7-tSNPs set (average haplotype r^2=0.55) in predicting allele frequencies for the remaining 51 untyped biallelic SNPs in the re-sequenced sample, obtaining an error rate from the comparison of observed vs. estimated minor allele frequencies (MAF) from:

$$Error(\%) = \frac{\left| MAF_{observed} - MAF_{estimated} \right|}{MAF_{observed}} \times 100$$

To compare the 7-tSNPs set utilized for the study with an optimal tagging alternative (i.e. with an average haplotype r^2=1.0), we also selected an additional 15-tSNPs set using the same methodology described. For both sets, the M_D measures, observed and estimated MAFs, and error values for all untyped SNPs were derived. A high correlation (R^2>0.99) between observed and estimated MAFs for the two tSNPs sets was found. Average errors for MAF estimates were similar between both tSNPs sets (6.7% for 15-tSNPs and 7.5% for 7-tSNPs). However, average errors were reduced by more than 20% among the untyped SNPs with M_D≥0.7 (5.3% for 15-tSNPs and 4.6% for 7-tSNPs) and even more (by more than 60%) among the untyped SNPs with MAF≥10% (1.7% for 15-tSNPs and 2.7% for 7-tSNPs). Thus, using a threshold of MAFs≥10% rather than the M_D≥0.7, reduced MAF error estimates at least by two-fold. This is supported by recent observations across diverse populations indicating that genotype imputation errors produced the greatest power reduction for SNPs with MAFs≤10% (E10). To reduce the MAF error rate as low as possible, we considered the overlapping SNPs of both thresholds for the association study (with MAF≥10% and M_D≥0.7). MAF error estimates derived from these two tSNPs sets were minimal (0.7% for 15-tSNPs and 2.6% for 7-tSNPs), still allowing for the accurate imputation of more than 80% of SNPs with MAF≥10% (27 SNPs for 15-tSNPs and 24 SNPs for 7-tSNPs). Worth noting, when a set of 6 tSNPs was selected to provide an average haplotype r^2=1.0 for the Utah residents with ancestry from Northern and Western Europe (CEU) from HapMap Phase II (E11) and was tested against our re-sequencing sample, MAF error estimates were five-fold larger on average irrespective of MAF or M_D values (not shown).

Taken together, these data support that the utilized set of 7 tSNPs allowed accurate imputing of a similar number of untyped SNPs in the *IRAK3* gene as a set of 15 tSNPs with maximum theoretical coverage (haplotype r^2=1.0). In addition, we adopted a conservative approach by imputing and studying the association of those 17 untyped SNPs with MAF≥10% and M_D≥0.7 to reduce errors to the minimum.

Genotyping quality control

Genotyping was blind to case and control status. Approximately 9% of the samples were genotyped in duplicate to monitor genotyping quality. Samples used for polymorphism identification were also genotyped with this platform to test allele-

calling reliability. Genotypes were assigned using all data simultaneously. Among all 550 samples analyzed in this study, completion rate and the average confidence of automatic genotype calls were over 90% for all tSNPs assessed. Among duplicated samples, no discrepancies were found with the original genotypes for any of the SNPs, giving an estimated overall discordance rate of 0.0%. Only 3 alleles out of the 268 genotypes obtained by both sequencing and iPLEX Gold assays were discordant (1.11%; 95% CI=0.23-3.23%). Taken together, these data are indicative of accurate genotyping methods.

Statistical analysis

Departures from Hardy-Weinberg equilibrium (HWE) were assessed using the Pearson´s goodness-of-fit χ^2-test.

Logistic regression model was constructed by backwards elimination, starting with a model including all the variables showing significant differences between septic patients with ALI and septic patients without ALI from Table 3, and maintaining in the model the variables showing *p*-values ≤0.1. The final model excluded APACHE II, and considered as covariates the SNP of interest, gender (male), source of infection (extra-pulmonary) and ICU mortality.

To adjust for the multiple comparisons, a permutation procedure was also performed. The permuted *p*-value was calculated by testing the association of 7 tSNPs over permuted samples, obtained from 1000 swappings of the ALI and non-ALI labels, and saving best chi-square value overall loci for each single sample. This resulted in an empirical distribution of association values considering the multiple tests done. P-values adjusted for the multiple tests being performed were then obtained by comparing the observed chi-square of each tSNP in this distribution (E12). Because the permutation scheme preserves the LD structure, this provides a less stringent correction for multiple testing in comparison to the commonly used Bonferroni correction, which assumes that all tests are independent.

Information contained in the tSNPs set was used to impute untyped SNPs in ALI and non-ALI patients and to infer the association of these SNPs by means of TUNA ver. 1.1 (E8). For this aim, PHASE 2.1 (E13) was used to estimate the haplotypes from sequencing data that were employed by TUNA software as a reference for the underlying LD across all gene variants in the population. To maintain an extra level of confidence in the imputation of untyped variation, only common well-tagged variants (i.e. with multilocus LD values≥0.7) (E8) with MAF≥10% were considered for this analysis. The justification for this decision is explained above.

The patterns of LD, in terms of r^2 values, were explored using Haploview 3.32 (E14) and SNAP (E15). The latter was applied to plot LD across the 250 kb region centered on *IRAK3* gene for the HapMap CEU sample.

Luciferase activity assay

PupaSuite was used to predict the putative functionality of associated SNPs, including effects on microRNA binding sites, transcription factor binding sites, splicing sites, exonic splicing enhancers, and exonic splicing silencers, using the default parameters of the on-line tool (E16). The putative *IRAK3* promoter region was identified by Cap-Analysis of Gene Expression (CAGE) and Genomatix database (http://www.genomitix.de). Human pulmonary artery endothelial cells (HPAEC) (Cambrex, Walkersville, MD, USA), which are known to expresses *IRAK3* (E17), were cultured as previously described (E18) and transfected with FuGENE 6 Transfection Reagent (Roche Applied Science, Indianapolis, IN) in Opti-MEM medium for 2 h. A plasmid with renilla luciferase gene (phRL-TK) was co-transfected with pGL3 fused constructs as transfection efficacy control. Cells were

harvested 48 h post-transfection and lysed in Passive Lysis Buffer. Luciferase activity was measured by Dual-Luciferase Assay Kits (Promega). The relative construct activities were expressed as the ratio of firefly luciferase in pGL3 to renilla luciferase in phRL-TK (RLU) following the Dual-Glo luciferase reporter assay protocols recommended by the manufacturer. Four independent transfections and duplicate luciferase assays were performed for each condition, along with pGL3-basic vector used as a control. Values were normalized to the activity of the construct containing the reference allele rs1732887_A (-1464A). A Student *t*-test was used for comparisons with a two-tailed statistical significance set at $p<0.05$.

REFERENCES

E1. Villar J, Perez-Mendez L, Flores C, Maca-Meyer N, Espinosa E, Muriel A, Sanguesa R, Blanco J, Muros M, Kacmarek RM. A *CXCL2* polymorphism is associated with better outcomes in patients with severe sepsis. *Crit Care Med* 2007;35:2292-2297.

E2. Levy MM, Fink MP, Marshall JC, Abraham E, Angus D, Cook D, Cohen J, Opal SM, Vincent JL, Ramsay G. 2001 SCCM/ESICM/ACCP/ATS/SIS international sepsis definitions conference. *Crit Care Med* 2003;31:1250-1256.

E3. Bernard GR, Artigas A, Brigham KL, Carlet J, Falke K, Hudson L, Lamy M, Legall JR, Morris A, Spragg R. The American-European consensus conference on ARDS. Definitions, mechanisms, relevant outcomes, and clinical trial coordination. *Am J Respir Crit Care Med* 1994;149:818-824.

E4. Baudouin SV, Saunders D, Tiangyou W, Elson JL, Poynter J, Pyle A, Keers S, Turnbull DM, Howell N, Chinnery PF. Mitochondrial DNA and survival after sepsis: A prospective study. *Lancet* 2005;366:2118-2121.

E5. Cabrera de Leon A, Rodriguez Perez Mdel C, Almeida Gonzalez D, Dominguez Coello S, Aguirre Jaime A, Brito Diaz B, Gonzalez Hernandez A, Perez Mendez LI. [presentation of the "CDC de Canarias" Cohort: Objectives, design and preliminary results]. *Rev Esp Salud Publica* 2008;82:519-534.

E6. Antonarakis SE. Recommendations for a nomenclature system for human gene mutations. Nomenclature working group. *Hum Mutat* 1998;11:1-3.

E7. Ahmadi KR, Weale ME, Xue ZY, Soranzo N, Yarnall DP, Briley JD, Maruyama Y, Kobayashi M, Wood NW, Spurr NK, Burns DK, Roses AD, Saunders AM, Goldstein DB. A single-nucleotide polymorphism tagging set for human drug metabolism and transport. *Nat Genet* 2005;37:84-89.

E8. Wen X, Nicolae DL. Association studies for untyped markers with tuna. *Bioinformatics* 2008;24:435-437.

E9. Nicolae DL. Quantifying the amount of missing information in genetic association studies. *Genet Epidemiol* 2006;30:703-717.

E10. Huang L, Wang C, Rosenberg NA. The relationship between imputation error and statistical power in genetic association studies in diverse populations. *Am J Hum Genet* 2009;85:692-698.

E11. Frazer KA, Ballinger DG, Cox DR, Hinds DA, Stuve LL, Gibbs RA, Belmont JW, Boudreau A, Hardenbol P, Leal SM, Pasternak S, Wheeler DA, Willis TD, Yu F, Yang H, Zeng C, Gao Y, Hu H, Hu W, Li C, Lin W, Liu S, Pan H, Tang X, Wang J, Wang W, Yu J, Zhang B, Zhang Q, Zhao H, Zhao H, Zhou J, Gabriel SB, Barry R, Blumenstiel B, Camargo A, Defelice M, Faggart M, Goyette M, Gupta S, Moore J, Nguyen H, Onofrio RC, Parkin M, Roy J, Stahl E, Winchester E, Ziaugra L, Altshuler D, Shen Y, Yao Z, Huang W, Chu X, He Y, Jin L, Liu Y, Shen Y, Sun W, Wang H, Wang Y, Wang Y, Xiong X, Xu L, Waye MM,

Tsui SK, Xue H, Wong JT, Galver LM, Fan JB, Gunderson K, Murray SS, Oliphant AR, Chee MS, Montpetit A, Chagnon F, Ferretti V, Leboeuf M, Olivier JF, Phillips MS, Roumy S, Sallée C, Verner A, Hudson TJ, Kwok PY, Cai D, Koboldt DC, Miller RD, Pawlikowska L, Taillon-Miller P, Xiao M, Tsui LC, Mak W, Song YQ, Tam PK, Nakamura Y, Kawaguchi T, Kitamoto T, Morizono T, Nagashima A, Ohnishi Y, Sekine A, Tanaka T, Tsunoda T, Deloukas P, Bird CP, Delgado M, Dermitzakis ET, Gwilliam R, Hunt S, Morrison J, Powell D, Stranger BE, Whittaker P, Bentley DR, Daly MJ, de Bakker PI, Barrett J, Chretien YR, Maller J, McCarroll S, Patterson N, Pe'er I, Price A, Purcell S, Richter DJ, Sabeti P, Saxena R, Schaffner SF, Sham PC, Varilly P, Altshuler D, Stein LD, Krishnan L, Smith AV, Tello-Ruiz MK, Thorisson GA, Chakravarti A, Chen PE, Cutler DJ, Kashuk CS, Lin S, Abecasis GR, Guan W, Li Y, Munro HM, Qin ZS, Thomas DJ, McVean G, Auton A, Bottolo L, Cardin N, Eyheramendy S, Freeman C, Marchini J, Myers S, Spencer C, Stephens M, Donnelly P, Cardon LR, Clarke G, Evans DM, Morris AP, Weir BS, Tsunoda T, Mullikin JC, Sherry ST, Feolo M, Skol A, Zhang H, Zeng C, Zhao H, Matsuda I, Fukushima Y, Macer DR, Suda E, Rotimi CN, Adebamowo CA, Ajayi I, Aniagwu T, Marshall PA, Nkwodimmah C, Royal CD, Leppert MF, Dixon M, Peiffer A, Qiu R, Kent A, Kato K, Niikawa N, Adewole IF, Knoppers BM, Foster MW, Clayton EW, Watkin J, Gibbs RA, Belmont JW, Muzny D, Nazareth L, Sodergren E, Weinstock GM, Wheeler DA, Yakub I, Gabriel SB, Onofrio RC, Richter DJ, Ziaugra L, Birren BW, Daly MJ, Altshuler D, Wilson RK, Fulton LL, Rogers J, Burton J, Carter NP, Clee CM, Griffiths M, Jones MC, McLay K, Plumb RW, Ross MT, Sims SK, Willey DL, Chen Z, Han H, Kang L, Godbout M, Wallenburg JC, L'Archevêque P, Bellemare G, Saeki K, Wang H, An D, Fu H, Li Q, Wang Z, Wang R, Holden AL, Brooks LD, McEwen JE, Guyer MS, Wang VO, Peterson JL, Shi M, Spiegel J, Sung LM, Zacharia LF, Collins FS, Kennedy K, Jamieson R, Stewart J. A second generation human haplotype map of over 3.1 million SNPs. *Nature* 2007;449:851-861.

E12. Westfall PH, Young SS. Resampling-based multiple testing; examples and methods for p-value adjustment. NY: John Wiley & Sons; 1993.

E13. Stephens M, Smith NJ, Donnelly P. A new statistical method for haplotype reconstruction from population data. *Am J Hum Genet* 2001;68:978-989.

E14. Barrett JC, Fry B, Maller J, Daly MJ. Haploview: Analysis and visualization of LD and haplotype maps. *Bioinformatics* 2005;21:263-265.

E15. 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;24:2938-2939.

E16. Conde L, Vaquerizas JM, Dopazo H, Arbiza L, Reumers J, Rousseau F, Schymkowitz J, Dopazo J. PupaSuite: Finding functional single nucleotide polymorphisms for large-scale genotyping purposes. *Nucleic Acids Res* 2006;34:W621-625.

E17. Karoly ED, Li Z, Dailey LA, Hyseni X, Huang YC. Up-regulation of tissue factor in human pulmonary artery endothelial cells after ultrafine particle exposure. *Environ Health Perspect* 2007;115:535-540.

E18. Garcia JG, Liu F, Verin AD, Birukova A, Dechert MA, Gerthoffer WT, Bamberg JR, English D. Sphingosine 1-phosphate promotes endothelial cell barrier integrity by EDG-dependent cytoskeletal rearrangement. *J Clin Invest* 2001;108:689-701.

SUPPLEMENTARY RESULTS

Table E1. Primers used for polymorphism discovery in *IRAK3* gene

Fragment	Forward primer	Reverse primer	Size*
1	ATCGAAATTCTTGCCCACAC	GAAAAGACACCAAATCAGCGTA	538
2	AGTTCGCACTCTGCTTGTC	CTGCATCCAACCCAGGAA	531
3	CAGCGTGAAAGCATCTTCTG	GGCCTTCTGCTAGTCCTGGT	562
4	GCCTGCTGGAAGTGTTGAGT	CCCCATTCACCAACATTAAAC	530
5	AGGACCACCCAAATGAGAA	ATGACCACCTGCACAATGAA	566
6	TTCATTGTGCAGGTGGTCAT	AAGAAGCTGGACCCTCAACC	550
7	CCAGTGTGTGCAGGATAAGG	GGTGTCAGCTGCTCCAAATA	550
8	CCTTGGTGTTGAGGGTGAAT	TGCAGAAATCACGTCTCCAT	543
9	CTGAAAGGGGACACGTAAGG	TCCTGTGACTTGGCTGAGAA	533
10	GCTACTATTGCCACCCGATG	CGTTCCTGCTACCCTTTGAA	522
11	CCTAAACTTTCCCTTTGCTGAA	GCATGAAGACAGGCCTCTAAC	558
12	TGGCATTTGGATAGTCTTTGC	TGCCTCTGTGTCTCTTCCAG	539
13	AACGAAAAATTGGGCATGG	CCAATGGGAGACTTCAAAGG	551
14	CATTAGCAGAAGGCAGGTGA	CACAAGAAAATGTGAAGCTGGA	519
15	AGGGACTTTCCAGCTTCACA	TGGAACAAACAGAACAGGTGA	563
16	TGGGTCTGTGTCATGCCTTA	TTTCACTTGCTCAGTCAACTTCA	555
17	TTCCCTTTGTCTTTTGCACAT	TTGAAACATCAACTCAATCAATTACA	528
18	GCCACCGTGGACAGATTTAT	GGGCTGTATGGGCACTAAAA	531
19	CGTAATGTTGCATGGGAATG	TGCAAAACACCAACCAAAAA	515
20	TTGGTTGGTGTTTTGCATGT	TTAGCCAATGGAGCCTTAACA	591
21	TTAAAATAACTTCAATTTGATTGACAC	TTGGCTATGTCCTTCAGAACC	553
22	AAGGCTTTGGTTCTTCAACA	TAGAGAAAGGCCCAGGAACC	568
23	CTCTGTGGAATGGTGGGAAC	CTGGGGATGGAAAGCATTTA	548
24	TAGTGCTAGCCTGCCCTGAT	TTGCTATCTTTTATGTGCTTCTGTG	518
25	GGGGCTAATGGATTGTTGAA	GGAAAGGAGGAAATGGTGAAA	558
26	GGCTATTCCTGGGGTACCAT	CCTCGAGAAAGGAGGGAAAA	523
27	CCTGGCTTTTGGCTTAACCT	CTGGTTGGTTCTCCTGCAAC	579
28	ATTCCCCATCCCACCTTACT	TTGATGCCAACTCACTGGAG	531
29	TCACCCCACCTTGCTATCTA	TGGGGATACTCAAAGTATGATTTCT	600
30	TGGGGACAGAGGGTCACTAA	GGCTAAAGATGGCTCACAGC	559
31	AAGAGGAGGCCAGTTCTTCC	ACAGGCAGGGACACATTGAT	389
32	CAGGCATGTGTCAGGGTATG	TCCATTGTTCACTCCTCCAA	648
33	CCGCGTCCTCCGATATTATT	ATTCACACTTTGGGCACCAG	544
34	TCCCTCCTAGCCCTTAGTGA	TTGCATTCCCTGTAGGCTTT	534
35	CCACCACTCTAATCCCCAAC	AGGCAGGTCTGAGAGACACA	596
36	AAGGTAATGGACATAGGATGGA	CCAGAGATAAAGAGGACCCAAC	600
37	AGCCTGCAGGTTCTAGGAGA	CTGAATCCCCCAAACAAATG	550
38	AGGGGATGGCCTCTCTGTTA	GGCCATCAACAAGATGTGAA	545
39	CCAAAGTTGAAGGGCAGAAG	GGTGGCAAGGGAATTGTAAA	545
40	GGAATTGCATGCCTTTCTGT	AGCGCAGAGCCACTAACTTG	564
41	GCTAACAGGAAATCAATGAGTCAA	GCTAACAGGAAATCAATGAGTCAA	545
42	TTGACTCATTGATTTCCTGTTAGC	AGATTTTCCATGGTTCTGTGG	548
43	GGTGGCAAATCTGAGGCTTA	AGAAGAACTGGGCATGTTGC	560
44	CCTTCTGATGAAGGCCTGAG	AATTGAGCGTGGATTTGGTC	558
45	TCCAAATTTTCCTGGGATGA	CAAGCAGGCTAGTCAGGGATT	578
46	GGCCATCTCTTCAGAGCCTA	GTTCTCCCAACTGCTTACCG	532
47	AGTCCTCCCATCAGGTCCTT	AGCAAAGTGAGGTCCACCAT	589

*Amplification product in base pairs.

Table E2. Summary of *IRAK3* gene re-sequencing in population controls (tSNPs are in bold and underlined)

#	Position*	Location (effect)	rs/ss#	5'†	3'†	MAF‡	Val§
1	64867853	-1494T>C	rs1732888	TGCCAGTTTT	AGATACATTT	0.28	‖
2	64867883	-1464A>G	rs1732887	ATTTTCGCTG	CAAAACCATT	0.25	‖
3	64868152	-1195C>T	rs2701653	TTCAAAACAA	AATTTTACTT	0.11	‖
4	64870002	c.133+523T>G	rs11465939	TTTACCCAAT	TCGTTGAATG	0.02	‖
5	64870029	c.133+550A>G	rs1732886	TATGCTAAAG	GTGGCTTTTT	0.28	‖
6	64870071	c.133+592G>A	rs11465940	CTTCAGAACT	TATTTTATGC	0.02	††
7	64871319	c.133+1841_1853delGGAAGTGGTCAGT	ss289570834	CAAAAATTGA	AATTATTCAA	0.30	§§
8	64871396	c.133+1917C>G	rs1625441	GCTGCCTGGG	TAGTTGCTAC	0.50	‖
9	64872005	c.133+2526T>A	rs2701654	GCAGCCTATA	GATATGACCA	0.06	‖
10	64872107	c.133+2628C>T	rs2576183	ACTCCTAGAG	TGCTGATTCA	0.07	**
11	64872409	c.133+2931T>C	rs56156535	TTATGGCTTT	AGTGAAAGCA	0.39	**
12	64872503	c.133+3024C>T	rs11176084	GGAGACAACT	ATCTGATTAC	0.03	‖
13	64876748	c.134-7010T>C	rs1168772	ATATATTAAC	AGTAAGTGTT	0.06	‖
14	64877754	c.134-6004A>G	rs1168774	CTTAGATGTG	TGCAAACTCA	0.44	‖
15	64877849	c.134-5909G>C	rs4762088	CTGAATCTCA	AAATGTGCTT	0.37	‖
16	64878292	c.134-5466C>T	rs12823414	GTCTGTCAAT	AGCAAAAATA	0.06	‖
17	64878509	c.134-5249C>A	rs17767286	CCTAGTTAAA	CTTACCCCAC	0.36	‖
18	64879508	c.134-4250A>G	rs17767298	CTTGTTTTCA	TCTGAAACTG	0.34	‖
19	64880383	c.134-3375T>C	rs1168757	GCAAACTCTC	TCCATGTAGT	0.08	‖
20	64880401	c.134-3357C>G	rs1168758	AGTGACTCAG	ATTTCAGGCT	0.06	‖
21	64880943	c.134-2815T>C	rs1185253	CTGCAACCTC	GCCTCCTGGG	0.08	††
22	64881027	c.134-2731T>G	rs1732875	CTGGCTAATT	TTGTATTTTT	0.06	††
23	64883686	c.134-72T>A	rs1185630	AATAAAACAT	AGGTACACAA	0.09	‖
24	64883968	c.316+28C>T	rs1882200	TAATGTGGCT	TTAATCTGTA	0.28	‖
25	64884556	c.316+616A>G	rs1617632	ATTTAGAACA	CATGATCAAT	0.02	‖
26	64884954	c.316+1014T>C	rs2289134	TATTAGTTCC	AATAGACTTC	0.34	‖
27	64885411	c.316+1471T>C	rs1732877	AAGACTTTAC	GCCCTAGTAC	0.39	‖
28	64886149	c.316+2209delT	rs11465951	TTTTCAATCT	CCATTTTATA	0.08	‖
29	64889666	c.381+99C>T	rs11465955	CTGGATTTGC	GGCACTTTTT	0.31	‖
30	64890247	c.435A>G (p.Lys145Lys)	rs56001649	ATAATGAAAA	GGTATGAAAA	0.03	‖
31	64891273	c.437-220A>T	rs2293657	GGTAAAAATC	CAATGCTGCT	0.31	‖
32	64891495	c.439G>A (p.Val147lle)	rs1152888	TTCCTAAGGA	TACTGCTTAA	0.10	‖
33	64892495	c.588+851C>T	rs1821777	CACACACAAA	ACAGAAGAGC	0.31	‖
34	64904334	c.654-2436C>T	rs55715236	CCTGAATCAT	TTCTCAGATG	0.30	**
35	64904449	c.654-2321G>T	rs1623665	CTAATTTAAG	GATAAGATGA	0.39	‖
36	64904483	c.654-2287G>A	rs1624395	ACAGAGGAGG	AATGATAACT	0.38	‖
37	64904584	c.654-2186G>A	rs1370127	AATATTAAGG	AATTACCTTA	0.38	‖
38	64904905	c.654-1865C>T	rs1370128	AACAAGGCAC	CAGAACATTT	0.39	‖
39	64908730	c.887+313delA	rs10716217	GATTTTTTTT	CTTTACAATG	0.39	‖
40	64912239	c.887+3822G>T	rs17102235	GCTGACAATT	TTGCCTTTCT	0.02	‖
41	64912278	c.887+3861A>G	ss289570833	AAGGACTTCA	TGTAACCCAT	0.02	§§
42	64912424	c.887+4007G>A	rs726302	TCCCTTCCAC	GCCATTGAAC	0.02	‖
43	64912865	c.887+4448A>G	rs17102243	TAATTGGCTG	GAAAAGGACA	0.02	‖
44	64915771	c.887+7354T>A	rs1152908	AGCGTATAGT	TATAATTATA	0.38	‖

45	64919287	c.888-5246T>C	rs1152911	AACAGTCTTT	ATACAAGGGA	0.08	‖
46	64919431	c.888-5102G>A	rs17826057	CAAAAGACTA	AAAAACTGAG	0.07	‖
47	64919952	c.888-4581T>C	rs11176095	TGTCTTAAAA	GTCTGGAAAA	0.08	‖
48	64920175	c.888-4358G>A	rs1152912	TGGCCAGCCT	TATCCTTTAA	0.42	‖
49	64920682	c.888-3851T>A	rs1152913	CAGACAGCAC	TGTGTGCTGT	0.44	‖
50	64920772	c.888-3761T>C	rs1152914	CCCTCTTCCC	AGGTCCTTCT	0.05	**
51	64920868	c.888-3652TA(6_8)	ss289570835	TATACCCACA	AAATATATCG	0.05	§§
52	64920925	c.888-3608A>C	rs1152915	GTATATATAT	TCTCTCTCTT	0.26	**
53	64921571	c.888-2962A>T	rs1152916	TTAAGGTTGA	GCAGGGGAAA	0.45	‖
54	64922569	c.888-1964C>T	rs1152918	TGGGGGTGTG	GAGGCCCAGG	0.05	‖
55	64923752	c.888-781C>T	rs3782348	CTGAGGCTTA	ATGGCGAAGG	0.05	‖
56	64924916	c.1087-79A>G	rs11176097	AACTATATTC	TAGTCATGGT	0.03	‖
57	64928242	c.*21A>G	rs7135413	ATAAAGAAAA	AGCAAGTATT	0.03	‖
58	64930378	*2157T>C	rs10506481	GTTAATCATA	TCACATTAGA	0.09	‖
59	64932466	*4245G>T	rs7972963	GTCATTTGTT	ATTCCACAGA	0.08	‖

*According to NCBI build 36; †Flanking sequences; ‡Minor allele frequency; §Validation status as October 2009: frequency available, ‖; 1000 genomes Project,**; unknown or by 2-hit as in dbSNP build 129, ††; described here for the first time,§§.

Table E3. Summary of quality control measures for genotyped *IRAK3* gene tSNPs

rs/ss#	Location (effect)	Position*	CR† (%)	CAC‡ (%)	HWE§ controls	HWE§ cases
ss289570834	Intron 1	64871319	95	98	0.295	0.295
rs11465955	Intron 3	64889666	91	98	0.295	0.295
rs1152888	Exon 5 (Val147Ile)	64891495	94	99	0.355	0.303
rs1624395	Intron 6	64904483	95	98	0.295	0.295
rs1370128	Intron 6	64904905	95	97	0.295	0.295
rs1152912	Intron 8	64920175	95	94	0.063	0.295
rs10506481	3’ flanking region	64930378	94	91	0.295	0.063

*According to NCBI build 36; †Completion rate; ‡Average confidence of automatic calls; §Hardy-Weinberg equilibrium *p*-values.

Table E4. Genotype counts of genotyped SNPs associated with ALI.

		ALI			non-ALI		
rs#	Ref. allele/Risk allele	1/1	1/2	2/2	1/1	1/2	2/2
rs1732887	A/G	103	63	12	21	9	0
rs1732886	A/G	102	63	13	21	9	0
rs10506481	C/T	2	28	139	4	6	20

Table E5. Final regression models used to adjust *IRAK3* SNP associations with ALI.

			95% CI	
	p-value	OR	Lower	Upper
(A) Regression model for rs10506481 (T as risk)				
rs10506481	**0.021**	2.50	1.15	5.47
Gender	0.073	2.38	0.92	6.17
ICU mortality	**0.002**	8.46	2.25	31.81
Source of infection	**0.011**	0.07	0.10	0.54
(B) Regression model for rs1732887 (G as risk)				
rs1732887	0.053	2.37	0.99	5.65
Gender	0.074	2.37	0.92	60.09
ICU mortality	**0.001**	8.48	2.34	30.76
Source of infection	**0.017**	0.08	0.01	0.63
(C) Regression model for rs1732886 (G as risk)				
rs1732886	**0.033**	2.57	1.02	6.10
Gender	0.063	2.46	0.95	6.38
ICU mortality	**0.001**	8.82	2.42	32.11
Source of infection	**0.016**	0.08	0.01	0.62

Significant *p*-values in bold.

Table E6. Association of imputed SNPs of *IRAK3* gene with ALI.

		MAF[†]		
rs#	Position[*]	non-ALI	ALI	*p*-value
rs1732888	64867853	0.141	0.269	**0.009**
rs1732887	64867883	0.138	0.265	**0.010**
rs1732886	64870029	0.165	0.299	**0.012**
rs56156535	64872409	0.410	0.389	0.713
rs4762088	64877849	0.373	0.396	0.703
rs17767286	64878509	0.378	0.359	0.772
rs17767298	64879508	0.362	0.343	0.759
rs1882200	64883968	0.331	0.314	0.785
rs2289134	64884954	0.377	0.363	0.791
rs1732877	64885411	0.460	0.416	0.520
rs2293657	64891273	0.325	0.312	0.827
rs1821777	64892495	0.338	0.327	0.868
rs55715236	64904334	0.325	0.312	0.812
rs1623665	64904449	0.462	0.406	0.387
rs1370127	64904584	0.448	0.390	0.387
rs10716217	64908730	0.448	0.400	0.507
rs1152908	64915771	0.428	0.385	0.470

[*]According to NCBI build 36; [†]Minor allele frequency. Significant *p*-values in bold.

Table E7. Association of *IRAK3* SNPs with quantitative measures of clinical severity.

rs#	Location (effect)	Position[*]	Days in ICU[‡]	Organs failing[‡]	Apache II[‡]
rs1732887	-1464A>G	64867883	0.063	0.102	0.771
rs1732886	c.133+550A>G	64870029	0.051	0.107	0.751
rs10506481	3' flanking region	64930378	0.388	0.374	0.702

[*]According to NCBI build 36; [‡]*p*-value from a linear regression model.