

**Literature Mining using MedGene and BioGene Database for Genes Related to:**

**A Genome-Wide Expression Analysis in Blood Identifies  
Pre-Elafin as a Biomarker in ARDS**

Zhaoxi Wang, Douglas Beach, Li Su, Rihong Zhai, and David C. Christiani

**Online Data Supplement**

## **ARDS**

MedGene and BioGene are literature mining projects sponsored by Harvard Institute of Proteomics (<http://www.hip.harvard.edu>) for the creation of human gene-to-disease co-occurrence network of all named human genes and all human diseases by automated analysis of MeSH indexes, title and abstracts in millions of Medline records (25). We used web-based interface of MedGene database (by 3/14/2007) hosted at Harvard Institute of Proteomics to conduct literature mining for genes associated with ARDS (MeSH disease term: Respiratory Distress Syndrome, Adult) in the Medline Database. We also searched BioGene database (by 3/14/2007) for genes associated with ARDS-related MeSH vocabularies, including neutrophil, leukotriene, and prostaglandin. The statistical method used to rank the gene list was the product of frequency. The database only allowed downloading maximum number of 100 top-ranked genes for each term of disease or MeSH vocabulary, without specific request for longer gene list. The numbers of top-ranked genes of MeSH mining are the combined lists of non-redundant genes of all sub-vocabularies available at BioGene. In addition, we used MedGene to sort the linkage of genes identified by paired *t*-tests and ARDS into four categories, including directly link (first-degree association), directly linked by gene family term, indirect link through other ARDS genes, and not previously associated with ARDS.

**Table E1. Characterization of Study Population for ELISA Assay**

	<b>ARDS</b>	<b>Control</b>	<b><i>p</i></b>
	<b>(n=40)</b>	<b>(n=23)</b>	<b>Value</b>
<b>General</b>			
Age, year (mean±SD)	60.5 ± 17.3	57.7 ± 9.2	0.407
Gender, (male/female)	12/11	24/16	0.547
APACHE III score, (mean±SD)	79.7 ± 24.3	69.7 ±15.4	0.052
Thrombocytopenia (platelets<80,000 mm <sup>-3</sup> ), N (%)	9 (22.5)	4 (17.4)	0.753
Creatinine >2.0 mg/dL, N (%)	16 (40.0)	6 (26.1)	0.265
Total bilirubin > 2.0 mg/dL, N (%)	8 (20.0)	3 (13.0)	0.732
Transfusion of PRBC, N (%)	29 (72.5)	9 (39.1)	0.009
Number of PRBC transfused	4.8 ± 6.3	3.0 ± 5.7	0.255
Liver disease, N (%)	4 (10.0)	1 (4.4)	0.644
End-stage renal disease, N (%)	3 (7.5)	1 (4.4)	1.000
Diabetes, N (%)	10 (25.0)	7 (30.4)	0.640
Recent steroid use, N (%)	5 (12.5)	4 (17.4)	0.713
<b>CLINICAL RISK FACTORS, N (%)</b>			
Direct lung injury	20 (50.0)	11 (47.8)	0.868
Septic shock	22 (55.0)	8 (34.8)	0.122
Sepsis	13 (32.5)	10 (43.5)	0.384
Trauma	2 (5.0)	1 (4.4)	1.000
Multiple transfusion	7 (17.5)	2 (8.7)	0.467

**Table E2. Patient characteristics for subjects with plasma samples for ELISA assay**

Patient #	Sample Group	Age	Gender	Etiology of ARDS	APACHE III*	Outcome
<b>ARDS patients</b>						
1	pre-diagnosis	71	male	sepsis/septic shock/multiple transfusion	72	died
2	pre-diagnosis	89	male	sepsis/septic shock/pneumonia	79	alive
3	pre-diagnosis	64	male	sepsis/septic shock/pneumonia/aspiration	83	died
4	pre-diagnosis	77	female	sepsis/septic shock/pneumonia	86	died
5	pre-diagnosis	55	male	pneumonia	88	died
6	pre-diagnosis	47	male	sepsis/septic shock	90	alive
7	pre-diagnosis	57	male	sepsis/septic shock/pneumonia	92	died
8	pre-diagnosis	70	male	sepsis	92	alive
9	pre-diagnosis	82	male	sepsis/septic shock/pneumonia	105	died
10	pre-diagnosis	71	male	sepsis/septic shock/pneumonia	105	died
11	pre-diagnosis	69	male	sepsis/septic shock/pneumonia	105	alive
12	pre-diagnosis	73	male	sepsis/septic shock/multiple transfusion	120	died
13	day of diagnosis	78	female	sepsis/pneumonia	50	alive
14	day of diagnosis	62	female	multiple transfusion	63	alive
15	day of diagnosis	66	female	sepsis	68	died
16	day of diagnosis	52	female	sepsis/septic shock/pneumonia	80	alive
17	day of diagnosis	84	female	sepsis/pneumonia	89	died
18	day of diagnosis	67	female	sepsis/septic shock/pneumonia	92	alive
19	day of diagnosis	82	female	sepsis/septic shock/pneumonia/aspiration	99	died
20	day of diagnosis	22	male	sepsis/septic shock/pneumonia/aspiration	101	died
21	day of diagnosis	93	male	sepsis/septic shock/pneumonia/aspiration	116	died
22	day of diagnosis	63	female	sepsis/septic shock	123	alive
23	post-diagnosis	45	female	trauma	35	alive
24	post-diagnosis	51	female	multiple transfusion	42	alive
25	post-diagnosis	62	female	sepsis/septic shock	43	died
26	post-diagnosis	38	male	sepsis/pneumonia	46	alive
27	post-diagnosis	45	male	sepsis/pneumonia	48	alive
28	post-diagnosis	42	male	sepsis/pneumonia	56	alive
29	post-diagnosis	48	male	sepsis/multiple transfusion	58	alive
30	post-diagnosis	45	female	sepsis/pneumonia	62	alive
31	post-diagnosis	70	male	sepsis	66	died
32	post-diagnosis	34	female	sepsis/pneumonia	69	alive
33	post-diagnosis	76	male	sepsis/pneumonia	73	died
34	post-diagnosis	68	female	sepsis/septic shock/pneumonia	73	alive
35	post-diagnosis	20	male	pneumonia/trauma/multiple transfusion	74	alive
36	post-diagnosis	47	male	sepsis/septic shock/pneumonia	74	alive
37	post-diagnosis	74	male	sepsis/septic shock/multiple transfusion	85	alive
38	post-diagnosis	45	male	sepsis/septic shock/pneumonia	106	died
39	post-diagnosis	63	female	sepsis/septic shock	123	alive
40	post-diagnosis	51	female	sepsis/pneumonia	132	died

Patient #	Sample Group	Age	Gender	Etiology of ARDS	APACHE III*	Outcome
<b>At-risk patients</b>						
1	control	44	male	sepsis/pneumonia	34	alive
2	control	75	male	sepsis/septic shock/pneumonia	50	alive
3	control	45	female	sepsis/pneumonia	53	alive
4	control	43	male	sepsis/pneumonia	54	alive
5	control	45	female	sepsis	55	alive
6	control	56	male	multiple transfusion	59	alive
7	control	56	male	multiple transfusion	60	alive
8	control	59	female	sepsis/pneumonia	62	alive
9	control	54	female	sepsis	64	alive
10	control	55	male	sepsis/pneumonia	67	alive
11	control	54	female	sepsis	67	alive
12	control	66	male	aspiration	71	alive
13	control	62	female	sepsis/septic shock	74	alive
14	control	51	female	pneumonia/trauma	75	alive
15	control	68	male	sepsis/septic shock/pneumonia/aspiration	76	alive
16	control	58	female	sepsis/septic shock	77	alive
17	control	64	male	sepsis/septic shock/pneumonia	79	alive
18	control	68	male	sepsis/septic shock/pneumonia	79	alive
19	control	66	female	sepsis/pneumonia	82	alive
20	control	51	male	sepsis/septic shock/pneumonia	83	alive
21	control	70	female	sepsis/pneumonia	86	alive
22	control	48	male	sepsis/septic shock	95	alive
23	control	67	female	aspiration	101	alive

\* APACHE III score was calculated for each patient within 24 hours of ICU admission.

**Table E3. Complete list of 126 genes with changed expression in peripheral blood between the date of ARDS diagnosis and the date of ICU discharge (paired *t*-test, *p* < 0.05)**

Symbol	Probe Set	Gene Name	Gene ID	Fold Change*	MedGene Category †
<i>PI3</i>	203691_at	peptidase inhibitor 3, skin-derived (SKALP)	5266	-2.98	1
	41469_at			-2.65	
<i>IL8</i>	202859_x_at	interleukin 8	3576	-2.93	1
	205592_at			-1.79	
<i>unknown</i>	211781_x_at	unknown (gb:BC006164.1)		-2.22	
<i>MME</i>	203435_s_at	membrane metallo-endopeptidase (neutral endopeptidase, enkephalinase)	4311	-2.12	3
<i>PTGS2</i>	204748_at	prostaglandin-endoperoxide synthase 2 (prostaglandin G/H synthase and cyclooxygenase)	5743	-2.03	1
<i>SGK</i>	201739_at	serum/glucocorticoid regulated kinase	6446	-2.03	3
<i>BNIP3L</i>	221478_at	BCL2/adenovirus E1B 19kDa interacting protein 3-like	665	-1.84	3
<i>POLB</i>	203616_at	polymerase (DNA directed), beta	5423	-1.83	3
<i>STAT1</i>	200887_s_at	signal transducer and activator of transcription 1, 91kDa	6772	-1.78	3
<i>FGL2</i>	204834_at	fibrinogen-like 2	10875	-1.68	3
<i>GPR177</i>	221958_s_at	G protein-coupled receptor 177	79971	-1.58	3
<i>PIGF</i>	205077_s_at	phosphatidylinositol glycan anchor biosynthesis, class F	5281	-1.56	3
<i>CLEC7A</i>	221698_s_at	C-type lectin domain family 7, member A	64581	-1.56	3
<i>C14orf159</i>	218298_s_at	chromosome 14 open reading frame 159	80017	-1.56	4

<b>Symbol</b>	<b>Probe Set</b>	<b>Gene Name</b>	<b>Gene ID</b>	<b>Fold Change*</b>	<b>MedGene Category †</b>
<i>TBCC</i>	202495_at	tubulin folding cofactor C	6903	-1.56	4
<i>LY75</i>	205668_at	lymphocyte antigen 75	4065	-1.55	3
<i>BTRC</i>	216091_s_at	beta-transducin repeat containing	8945	-1.55	3
<i>DUSP6</i>	208892_s_at	dual specificity phosphatase 6	1848	-1.52	3
<i>ZYG11BL</i>	202448_s_at	zyg-11 homolog B (C. elegans)-like	10444	-1.50	4
<i>ZNF589</i>	210062_s_at	zinc finger protein 589	51385	-1.48	3
<i>ZNF133</i>	216960_s_at	zinc finger protein 133	7692	-1.46	3
<i>ZCCHC10</i>	221193_s_at	zinc finger, CCHC domain containing 10	54819	-1.42	4
<i>XK</i>	206698_at	X-linked Kx blood group (McLeod syndrome)	7504	-1.42	4
<i>WARS</i>	200629_at	tryptophanyl-tRNA synthetase	7453	-1.42	3
<i>VPS13C</i>	218396_at	vacuolar protein sorting 13 homolog C (S. cerevisiae)	54832	-1.41	4
<i>VDAC1</i>	217140_s_at	voltage-dependent anion channel 1	7416	-1.41	3
<i>UBE2L6</i>	201649_at	ubiquitin-conjugating enzyme E2L 6	9246	-1.41	4
<i>TUBB2A</i>	204141_at	tubulin, beta 2A	7280	-1.40	4
<i>TRGC2 /</i>	215806_x_at	T cell receptor gamma constant 2 / T cell	445347	-1.40	3
<i>TRGV9 /</i>		receptor gamma variable 9 / TCR gamma	/ 6967 /		
<i>TARP</i>		alternate reading frame protein	6983		
	209813_x_at			-1.39	
<i>TRBV21-1 /</i>	211796_s_at	T cell receptor beta variable 21-1 / T cell	28566 /	-1.39	4
<i>TRBV19 /</i>		receptor beta variable 19 / T cell receptor beta	28568 /		
<i>TRBV7-2 /</i>		variable 7-2 / T cell receptor beta variable 5-4 /	28596 /		
<i>TRBV5-4 /</i>		T cell receptor beta variable 3-1 / T cell receptor	28611 /		
<i>TRBV3-1 /</i>		beta constant 1	28619 /		
<i>TRBC1</i>			28639		

Symbol	Probe Set	Gene Name	Gene ID	Fold Change*	MedGene Category †
<i>TRBV19</i> /	210915_x_at	T cell receptor beta variable 19 / T cell receptor	28568 /	-1.39	4
<i>TRBC1</i>		beta constant 1	28639		
<i>TRAK2</i>	202124_s_at	trafficking protein, kinesin binding 2	66008	-1.38	3
<i>TRAK1</i>	202079_s_at	trafficking protein, kinesin binding 1	22906	-1.38	3
<i>TRAC</i>	209670_at	T cell receptor alpha constant	28755	-1.38	3
<i>TRA@</i>	211902_x_at	T cell receptor alpha locus	6955	-1.37	3
<i>TRA@</i> /	210972_x_at	T cell receptor alpha locus / T cell receptor delta	28517 /	-1.37	4
<i>TRDV2</i> /		variable 2 / T cell receptor alpha variable 20 / T	28663 /		
<i>TRAV20</i> /		cell receptor alpha joining 17 / T cell receptor	28738 /		
<i>TRAJ17</i> /		alpha constant	28755 /		
<i>TRAC</i>			6955		
<i>TNS1</i>	221748_s_at	tensin 1	7145	-1.37	3
<i>TNFSF10</i>	202688_at	tumor necrosis factor (ligand) superfamily, member 10	8743	-1.36	3
<i>TNFAIP2</i>	202510_s_at	tumor necrosis factor, alpha-induced protein 2	7127	-1.36	3
<i>TMEM158</i>	213338_at	transmembrane protein 158	25907	-1.36	4
<i>TLE4</i>	216997_x_at	transducin-like enhancer of split 4 (E(sp1) homolog, Drosophila)	7091	-1.34	3
<i>THBS1</i>	215775_at	Thrombospondin 1	7057	-1.34	1
<i>TAF9</i>	202168_at	TAF9 RNA polymerase II, TATA box binding protein (TBP)-associated factor, 32kDa	6880	-1.33	3
<i>STK19</i>	36019_at	serine/threonine kinase 19	8859	-1.31	2
<i>SRPRB</i>	218140_x_at	signal recognition particle receptor, B subunit	58477	-1.31	4
<i>SOD2</i>	215223_s_at	superoxide dismutase 2, mitochondrial	6648	-1.31	1
<i>SMCHD1</i>	212569_at	structural maintenance of chromosomes flexible hinge domain containing 1	23347	-1.30	4



Symbol	Probe Set	Gene Name	Gene ID	Fold Change*	MedGene Category †
	212577_at			-1.30	
	212579_at			-1.30	
<i>SLC16A5</i>	213590_at	solute carrier family 16, member 5 (monocarboxylic acid transporter 6)	9121	-1.30	4
<i>SETDB1</i>	214197_s_at	SET domain, bifurcated 1	9869	-1.30	3
<i>SELENBP1</i>	214433_s_at	selenium binding protein 1	8991	-1.29	3
<i>SCO2</i>	205241_at	SCO cytochrome oxidase deficient homolog 2 (yeast)	9997	-1.29	3
<i>RUNX3</i>	204197_s_at	runt-related transcription factor 3	864	-1.29	3
	204198_s_at			-1.28	
<i>RP3-377H14.5</i>	222279_at	hypothetical protein FLJ35429	285830	-1.28	4
<i>RARRES3</i>	204070_at	retinoic acid receptor responder (tazarotene induced) 3	5920	-1.27	3
<i>RAD9A</i>	204828_at	RAD9 homolog A (S. pombe)	5883	-1.27	3
<i>RAD17</i>	207405_s_at	RAD17 homolog (S. pombe)	5884	-1.26	3
<i>PSME2</i>	201762_s_at	proteasome (prosome, macropain) activator subunit 2 (PA28 beta)	5721	-1.26	3
<i>PSME1</i>	200814_at	proteasome (prosome, macropain) activator subunit 1 (PA28 alpha)	5720	-1.26	3
<i>PPP2R1B</i>	202884_s_at	protein phosphatase 2 (formerly 2A), regulatory subunit A (PR 65), beta isoform	5519	-1.26	3
<i>POLS</i>	202466_at	polymerase (DNA directed) sigma	11044	-1.26	3
<i>PLA2G7</i>	206214_at	phospholipase A2, group VII (platelet-activating factor acetylhydrolase, plasma)	7941	-1.26	2

Symbol	Probe Set	Gene Name	Gene ID	Fold Change*	MedGene Category †
<i>PDE4B</i>	203708_at	phosphodiesterase 4B, cAMP-specific (phosphodiesterase E4 dunce homolog, Drosophila)	5142	-1.25	3
	222326_at			-1.25	
<i>PDCD4</i>	212594_at	programmed cell death 4 (neoplastic transformation inhibitor)	27250	-1.25	3
<i>PCAF</i>	203845_at	p300/CBP-associated factor	8850	-1.25	3
<i>PBLD</i>	219543_at	phenazine biosynthesis-like protein domain containing	64081	-1.25	3
<i>PARVB</i>	37965_at	parvin, beta	29780	-1.24	4
<i>PAIP1</i>	213754_s_at	poly(A) binding protein interacting protein 1	10605	-1.24	3
<i>P2RY5</i>	218589_at	purinergic receptor P2Y, G-protein coupled, 5	10161	-1.24	3
<i>OPTN</i>	202074_s_at	optineurin	10133	-1.24	3
<i>ODC1</i>	200790_at	ornithine decarboxylase 1	4953	-1.24	3
<i>OCLM</i>	208274_at	oculomedin	10896	-1.24	4
<i>NPC2</i>	200701_at	Niemann-Pick disease, type C2	10577	-1.24	3
<i>NOV</i>	214321_at	nephroblastoma overexpressed gene	4856	-1.23	3
<i>NOD2</i>	220066_at	nucleotide-binding oligomerization domain containing 2	64127	-1.23	3
<i>NFATC2IP</i>	217527_s_at	nuclear factor of activated T-cells, cytoplasmic, calcineurin-dependent 2 interacting protein	84901	-1.23	4
<i>NBPF1 /</i>	215434_x_at	neuroblastoma breakpoint family, member 1 /	440673	-1.23	3
<i>NBPF10</i>		neuroblastoma breakpoint family, member 10	/ 55672		
<i>NAPA</i>	206491_s_at	N-ethylmaleimide-sensitive factor attachment protein, alpha	8775	-1.23	3

<b>Symbol</b>	<b>Probe Set</b>	<b>Gene Name</b>	<b>Gene ID</b>	<b>Fold Change*</b>	<b>MedGene Category †</b>
<i>NAP1L4</i>	201414_s_at	nucleosome assembly protein 1-like 4	4676	-1.22	3
<i>MX1</i>	202086_at	myxovirus (influenza virus) resistance 1, interferon-inducible protein p78 (mouse)	4599	-1.22	3
<i>MULK</i>	222132_s_at	multiple substrate lipid kinase	55750	-1.22	3
<i>MRPS18A</i>	218385_at	mitochondrial ribosomal protein S18A	55168	-1.22	4
<i>MMD</i>	203414_at	monocyte to macrophage differentiation-associated	23531	-1.22	3
<i>METTL4</i>	219698_s_at	methyltransferase like 4	64863	-1.22	4
<i>MBOAT2</i>	213288_at	membrane bound O-acyltransferase domain containing 2	129642	-1.21	4
<i>MBD4</i>	214048_at	methyl-CpG binding domain protein 4	8930	-1.21	3
<i>MARCH8</i>	221824_s_at	membrane-associated ring finger (C3HC4) 8	220972	-1.21	3
<i>MAN2A2</i>	219999_at	mannosidase, alpha, class 2A, member 2	4122	-1.21	4
<i>LOC646912</i>	217092_x_at	similar to 60S ribosomal protein L7	646912	-1.21	4
<i>LOC54103</i>	222150_s_at	hypothetical protein LOC54103	54103	-1.21	4
<i>CLEC2B /</i>	209732_at	C-type lectin domain family 2, member B /	94158 /	1.21	4
<i>CDRT15P</i>		CMT1A duplicated region transcript 15 pseudogene	9976		
<i>CLCN5</i>	206704_at	chloride channel 5 (nephrolithiasis 2, X-linked, Dent disease)	1184	1.22	3
<i>CLC</i>	206207_at	Charcot-Leyden crystal protein	1178	1.22	3
<i>CIR</i>	209571_at	CBF1 interacting corepressor	9541	1.23	3
<i>CHRNE</i>	215916_at	cholinergic receptor, nicotinic, epsilon	1145	1.23	4
<i>CECRI</i>	219505_at	cat eye syndrome chromosome region, candidate	51816	1.23	3

Symbol	Probe Set	Gene Name	Gene ID	Fold Change*	MedGene Category †
<i>CDC2L1</i> /	211289_x_at	cell division cycle 2-like 1 (PITSLRE proteins) /	984 /	1.23	3
<i>CDC2L2</i>		cell division cycle 2-like 2 (PITSLRE proteins)	985		
<i>CD74</i>	209619_at	CD74 molecule, major histocompatibility complex, class II invariant chain	972	1.24	3
<i>CD52</i>	34210_at	CD52 molecule	1043	1.24	3
<i>CD3E</i>	205456_at	CD3e molecule, epsilon (CD3-TCR complex)	916	1.26	3
<i>CD36</i>	209555_s_at	CD36 molecule (thrombospondin receptor)	948	1.26	3
<i>CD2</i>	205831_at	CD2 molecule	914	1.26	1
<i>CD14</i>	201743_at	CD14 molecule	929	1.27	1
<i>CCL5</i>	204655_at	chemokine (C-C motif) ligand 5	6352	1.30	3
<i>CASP9</i>	203984_s_at	caspase 9, apoptosis-related cysteine peptidase	842	1.30	3
<i>C1orf50</i>	62212_at	chromosome 1 open reading frame 50	79078	1.31	4
<i>BTN3A3</i>	38241_at	butyrophilin, subfamily 3, member A3	10384	1.32	4
<i>BLVRB</i>	202201_at	biliverdin reductase B (flavin reductase (NADPH))	645	1.33	3
<i>BCL2L1</i>	215037_s_at	BCL2-like 1	598	1.34	3
<i>BAG1</i>	202387_at	BCL2-associated athanogene	573	1.36	3
<i>ATP2B1</i>	209281_s_at	ATPase, Ca <sup>++</sup> transporting, plasma membrane 1	490	1.37	3
<i>ATG4B</i>	204903_x_at	ATG4 autophagy related 4 homolog B (S. cerevisiae)	23192	1.38	3
<i>ASCC2</i>	215684_s_at	activating signal cointegrator 1 complex subunit 2	84164	1.39	4
<i>ARL4C</i>	202207_at	ADP-ribosylation factor-like 4C	10123	1.39	3
<i>APOBEC3G</i>	204205_at	apolipoprotein B mRNA editing enzyme, catalytic polypeptide-like 3G	60489	1.40	3

Symbol	Probe Set	Gene Name	Gene ID	Fold Change*	MedGene Category †
<i>APBB1IP</i>	219994_at	amyloid beta (A4) precursor protein-binding, family B, member 1 interacting protein	54518	1.40	3
<i>AMPD2</i>	212360_at	adenosine monophosphate deaminase 2 (isoform L)	271	1.46	3
<i>ABLIM1</i>	210461_s_at	actin binding LIM protein 1	3983	1.47	3
<i>ABCA1</i>	203505_at	ATP-binding cassette, sub-family A (ABC1), member 1	19	1.48	3
<i>HIP2</i>	202346_at	huntingtin interacting protein 2	3093	1.55	3
<i>OSBPL1A</i>	208158_s_at	oxysterol binding protein-like 1A	114876	1.58	4
<i>STXBP2</i>	209367_at	syntaxin binding protein 2	6813	1.62	4
<i>IDI1</i>	204615_x_at	isopentenyl-diphosphate delta isomerase 1	3422	1.58	3
	208881_x_at			1.65	
<i>HSPA1A /</i>	200800_s_at	heat shock 70kDa protein 1A / heat shock	3303 /	1.66	1
<i>HSPA1B</i>		70kDa protein 1B	3304		
<i>GALNT2</i>	217788_s_at	UDP-N-acetyl-alpha-D-galactosamine:polypeptide N-acetylgalactosaminyltransferase 2 (GalNAc-T2)	2590	1.83	3
<i>PDGFC</i>	218718_at	platelet derived growth factor C	56034	2.08	3
<i>HPGD</i>	203914_x_at	hydroxyprostaglandin dehydrogenase 15-(NAD)	3248	2.00	3
	203913_s_at			2.29	
<i>GADD45A</i>	203725_at	growth arrest and DNA-damage-inducible, alpha	1647	2.38	3

\* A negative number of fold change means the gene was down-regulated during the acute-stage of ARDS, and a positive number of fold change means the gene was up-regulated during the acute-stage.

† The gene list was sorted in MedGene database by the linkage of genes and ARDS. 1: first-degree association, genes that have been directly linked to ARDS by gene term

search; 2: first-degree associations by family term, genes that have been directly linked to this disease by family term search; 3: second degree associations, genes that have never been co-cited with ARDS, but have been linked to at least one first-degree gene; and 4: Genes new to ARDS, genes that have not been previously associated with this disease.

**Table E4. Literature mining of ARDS-related genes in Medline Database against 28 altered genes identified by microarray analysis**

	Top Ranked Genes ‡	Genes Identified from Microarray Analysis <sup>§</sup>		
		Total	Down-regulated at acute stage	Up-regulated at acute stage
<b>Disease*</b>				
ARDS	100	7	<i>IL8, PI3</i>	<i>CD14</i>
<b>MeSH Vocabulary †</b>				
Neutrophil	538	22	<i>IL8, LY75, MBD4, PCAF, PDE4B, PI3, SELENBP1, STAT1, THBS1, VDAC1, WARS</i>	<i>ABCA1, CCL5, CD14, CD36, CD74</i>
Leukotriene	419	17	<i>IL8, PI3, PTGS2</i>	<i>ABCA1, BCL2LI, CCL5, CD14, HPGD</i>
Prostaglandin	754	22	<i>FGL2, IL8, MMD, OPTN, PDCD4, PDE4B, PIGF, POLB, PTGS2, STAT1, STK19, TNFSF10</i>	<i>ABCA1, CD2, CD36, GADD45A, HPGD, PDGFC</i>

\* Text mining through MedGene project at Harvard Institute of Proteomics (<http://www.hip.harvard.edu>, by 3/14/2007) for genes associated with ARDS (MeSH term: Respiratory Distress Syndrome, Adult) in the Medline Database.

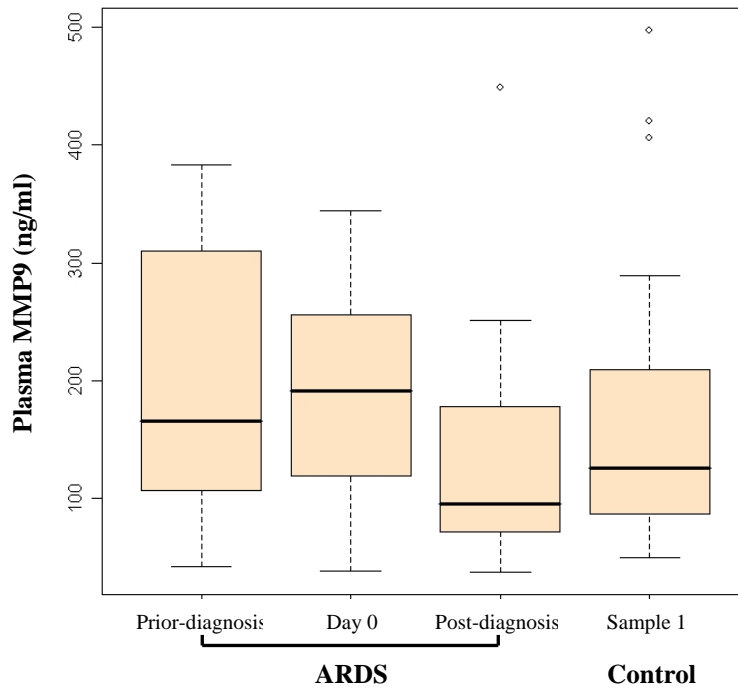
† Text mining through BioGene project at Harvard Institute of Proteomics (<http://www.hip.harvard.edu>, by 3/14/2007) for genes associated with ARDS-related MeSH vocabularies in the Medline Database.

‡ The statistical method used to rank the gene list is product of frequency. The project allows downloading maximum number of 100 top-ranked genes for each disease or MeSH vocabulary, without specific request for longer gene list. The

numbers of top-ranked genes of MeSH mining are the combined lists of non-redundant genes of all sub-vocabularies available at BioGene (by 3/14/2007).

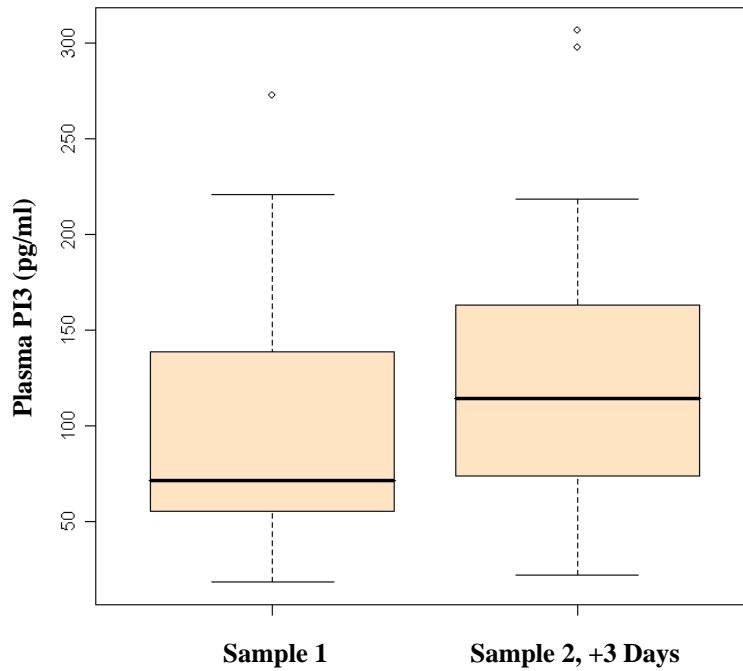
§ Genes with altered expression in microarray analysis were found in ARDS-related text mining.





	n	Mean (ng/μl)	95% CI
<b>ARDS</b>			
Pre-diagnosis (Day -5 to -2)	11	192.1	120.4, 263.8
Day of diagnosis (Day 0)	11	192.1	134.7, 249.4
Post-diagnosis (Day 1 to 3)	15	141.1	86.6, 195.5
<b>Control</b>			
Sample 1	21	194.2	139.5, 249.0
Sample 2, +3 day	18	203.0	132.9, 273.2

**Figure E1. Levels of MMP9 in plasma of ARDS patients and critical ill patients who did not develop ARDS (controls).** Each ARDS case provides one plasma sample. Based on the date of sample collection relative to the ARDS diagnosis date, including pre-diagnosis group (Day -5 to Day -1), day of diagnosis group (Day 0), and post-diagnosis group (Day 1 to Day 3). Sample 1 of control was collected during the first two days of ICU admission. There was no statistically significant difference in baseline characteristics between ARDS cases and controls, except that ARDS cases more frequently received transfusion ( $p = 0.009$ ). There is no statistically significant difference existed in plasma levels of MMP-9 between ARDS cases and at-risk controls.



**Figure E2. Levels of pre-elafin (PI3) in plasma of critical ill patients who did not develop ARDS (controls).** All controls provided paired plasma samples. Sample 1 was collected during the first two days of ICU admission; and Sample 2 was collected three days after Sample 1. Significant increase of plasma PI3 levels (mean ratio = 1.53, 95% CI, 1.28 – 1.78;  $p = 0.0007$ ) in Sample 2 (three days after the first 24 hours of ICU admission) was observed using paired  $t$ -test.