IDENTIFICATION OF GENES AND PATHWAYS INVOLVED

IN ALVEOLAR EPITHELIAL CELL DIFFERENTIATION

USING DNA MICROARRAY

By

ZHONGMING CHEN

Bachelor of Science
Sichuan University
Chengdu, P.R. China
1992

Master of Science
Shanghai Second Medical University
Shanghai, P.R. China
1998

Submitted to the Faculty of the
Graduate College of the
Oklahoma State University
in partial fulfillment of
the requirements for
the Degree of
DOCTOR OF PHILOSOPHY
December, 2005

IDENTIFICATION OF GENES AND PATHWAYS INVOLVED

IN ALVEOLAR EPITHELIAL CELL DIFFERENTIATION

USING DNA MICROARRAY

Dissertation Approved

Dr. Lin Liu
Dissertation Adviser

Dr. Robert Burnap

Dr. Guangping Chen

Dr. Nicholas Cross

Dr. Carey Pope

Dr. A. Gordon Emslie
Dean of the Graduate College

PREFACE[1]

The main function of the lung is gas exchange of oxygen and carbon dioxide. The gas exchange occurs in alveolar airspace and capillary blood vessels on the basolateral surface of alveoli. The alveolar epithelium is mainly composed of type I and type II alveolar epithelial cells (AEC I and AEC II). During inspiration and respiration, oxygen diffuses from alveolar airspace to capillary blood vessels, and carbon dioxide in an opposite direction driven by gas concentration gradients. Therefore, the knowledge of AEC I and AEC II is essential for the understanding of gas exchange and related lung diseases (e.g. lung injury induced by hyperoxia or air pollutants).

Currently, the functions of alveolar epithelial cells, especially AEC I are not well understood. AEC I are squamous and cover about 93% of alveolar surface. A number of AEC I marker genes were identified, including caveolin, T1 alpha, and aquaporin 5, but their functions in AEC I are not clear. Due to their large surface area, AEC I form a barrier, through which oxygen and carbon dioxide exchange. AEC I are believed to be terminally differentiated, and thus can not regenerate themselves. Compared with AEC I, AEC II are better understood, due to the fact that AEC II are easier to be isolated from the lung and studied *in vitro*. They are small and cuboidal, covering about 7% surface of alveoli. Various markers were identified in AEC II, including

---

[1] Only the most relevant references are listed herein. All other references in this section can be found in the following chapters.

surfactant proteins and LB180. Surfactant proteins A, B, C and D are proteins secreted by AEC II, and play roles in host defense and the formation of lipid-protein lattices, which prevent alveoli collapsing during inspiration and respiration. LB180 is important for cholesterol transport. AEC II are able to proliferate and differentiate, and are believed to be the progenitors of type I cells.   However, the proliferation and differentiation are poorly understood.

   Based on cellular kinetics studies of AEC, a model of AEC II differentiation was proposed over ten years ago (1). According to this model, most AEC II in normal lung are in resting status, G0 phase. In case of lung injury or mitogenic stimulation, a portion of AEC II can be activated by growth factors through MAPK signaling pathway.   KGF (FGF7) is one of the growth factors that are known to be required for AEC II proliferation and differentiation. Activated AEC II enter cell cycle and divide into daughter cells, which subsequently differentiate to AEC I and AEC II. Mature AEC I irreversibly lose mitogenic ability. However, it is not clear how this process occurs. Many genes, such as transcription factors, have been studied in AEC II differentiation since the proposing of the model, but the mechanisms involved are not completely understood (2). Furthermore, the proposed model can not explain how AEC markers, surfactant proteins and T1α, are differentially expressed in AEC I or AEC II.

The poor understanding of AEC differentiation is partially resulted from two fundamental problems, experimental cell differentiation models and gene study tools. There are *in vivo* (fetal lung development, hyperoxia, and hypoxia) and *in vitro* (AEC II culture) models, but none of them is effective. The limits behind *in vivo* the cell differentiation models are the complex of cell-cell

interaction, because there are over 40 cell types in the lung. Therefore, it is difficult to identify the factors controlling AEC II proliferation and differentiation from the interaction between AEC II and the other cell types in the lung. The *in vitro* AEC II culture has been used as a trans-differentiation model. AEC II lose surfactant proteins and the cuboidal shape, and gain T1α and "AEC I-like" squamous shape during culture in plastic dishes. This *in vitro* differentiation may be different from *in vivo* AEC II differentiation, due to the artifacts of culture environment, which is different from the lung alveoli. A common approach for AEC II differentiation is the combination of *in vivo* and *in vitro* models to eliminate the *in vivo* complex and *in vitro* artifacts. The second fundamental problem is gene study tools. In the past, individual genes were studied in AEC II differentiation, including the components of signaling pathways (e.g. KGF, IGF, ERK, MAPK and Bmp4), transcription factors (HNF, ssh and NF-kappaB), hormone (glucocorticoid) or cell proliferation control (e.g. p53 and rb). However, very few AEC-specific proliferation or differentiation genes have been identified. It is unclear if AEC II differentiation is controlled by a group of common genes or by some AEC-specific genes.

Facing with the two fundamental problems, I investigate AEC differentiation step by step in this dissertation. First, AEC differentiation models are developed and improved. An *in vitro* model with highly pure AEC II (Chapters 1, 4 and 5) and an *in vivo* hyperoxia model (Chapters 4 and 5) are developed for cell differentiation. In both models, a purity (about 95% for AEC II and 90% for AEC I) higher than previously reported results are achieved with modified cell isolation methods, and thus AEC-specific genes for cell differentiation can be identified. Second, high throughput gene study tools

based on DNA microarray are developed to identify genes involved in AEC II differentiation. A DNA microarray is a slide containing thousands of DNA probes, which capture fluorescently labeled cDNA from one or two biological samples (3). DNA microarray slides with 10,000 DNA probes are printed for the hybridization with 6 parallel samples, and tested by the identification of lung-prominent genes (Chapter 3). To increase the reliability of results from the data generated by DNA microarray hybridization, a bioinformatics software package, *RealSpot*, is written (Chapter 2). Last, the cell differentiation models and DNA microarray tools are applied to identify AEC-specific genes (Chapters 1 and 3), and AEC differentiation genes and pathways (Chapter 5).

The results from this dissertation provide both biological understanding of the lung, and technical tools for further study. Biologically, the Chapters 1 to 5 in this dissertation reveal clues of pulmonary functions from lung prominent genes, a novel AEC I function from AEC I-specific genes, and a fine-tuning model for balancing AEC II proliferation and differentiation. The biological knowledge will help to understand lung injury and repair during hyperoxia exposure, as well as other pathological conditions (e.g. asthma and edema). Technically, the improved AEC isolation methods and DNA microarray tools will be helpful for more detailed studies of AEC differentiation and functions of identified genes in AEC.

Reference List

1. Uhal BD. Cell cycle kinetics in the alveolar epithelium. Am J Physiol 272: L1031-L1045, 1997.

2. Warburton D, Wuenschell C, Flores-Delgado G and Anderson K. Commitment and differentiation of lung cell lineages. Biochem Cell Biol 76: 971-995, 1998.

3. Schena M, Shalon D, Davis RW and Brown PO. Quantitative monitoring of gene expression patterns with a complementary DNA microarray. Science 270: 467-470, 1995.

ACKNOWLEDGMENTS

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF SYMBOLS

μg	microgram
μl	microliter
^{o}C	centigrade
Σ	Sum

NOMENCLATURE

18S rRNA	18S ribosomal RNA
^{35}S TBPS	^{35}S t-Butylbicyclophosphorothionate
A260 & A280	light absorption at 260 nm & 280nm
aa-dUTP	5-[3-Aminoallyl]-2'-deoxyuridine 5'-triphosphate
acaa2	thiolase aa 1-397
actb	beta actin
AEC I & AEC II	type I & type II AEC
AEC	alveolar epithelial cell
Ager	advanced glycosylation end product-specific receptor
Alp	alkaline phosphatase
ANOVA	analysis of variable
Anp	atrial natriuretic peptide
Apo E	apolipoprotein E
AQP-5	aquaporin 5
ATP	adenosine Triphosphate
Bach	cytosolic peroxisome proliferator-induced acyl-coa thioesterase
BAL	bronchoalveolar lavage
BAX	Bcl2-associated X protein
BD-2	beta defensin-2

BMP	bitmap file
Bmp4	bone morphogenetic protein 4
BmpR	Bmp4 receptor
c/EBP	CCAAT/enhancer binding protein
Calb3	vitamin D-dependent calcium binding protein
CCSP	Clara cell phospholipid-binding protein
Cd37	leukocyte antigen cluster domain 37
Cdc2	Cell division cycle 2, also known as mitotic substrate of p34
cDNA	complementary DNA
Ces3	carboxylesterase 3
Cldn3	claudin 3
Cox6a2 & Cox8h	cytochrome c oxidase, subunit VI a2, & VIII h
CPU	central processing unit
Crhr1	corticotropin releasing hormone receptor
C_T	threshold cycle
Ctsd	preprocathepsin d
CV	coefficient of variation
Cy3 & Cy5	cyanide dyes 3 & 5
Ddr2	discoidin domain receptor family, member 2
DEP	disel exhaust particles
DIG	digoxygenin
Dlk1	delta-like homolog drosophila
DNA	deoxy-ribonucleic acid
DNase	deoxyl nuclease

EGF	epidermal growth factor
EMF	Windows Enhanced Metafile
ENT1	equilibrative nitrobenzylthioinosine-sensitive nucleoside transporter
ERK	Extracellular receptor kinase
EST	expression sequence tag
Fas	fatty-acid synthase
FBS	fetal bovine serum
Fdps	testis-specific farnesyl pyrophosphate synthetase
FDR	false discovery ratio
FGF	fibroblast growth factor receptor
FGFR1	fibroblast growth factor receptor-activating protein 1
FIZZ1	inflammatory zone 1
FRAG1	FGF receptor activating protein 1
Ftl1	ferritin light chain subunit 1
G0, G1, S, G2 & M	cell division cycle gap 0, gap 1, synthesis phase, gap 2 and mitosis, respectively
GABRP	gamma-aminobutyric acid receptor pi subunit
GAPDH	glycerol dehyde-3-phospate dehydrogenase
Gax	growth arrest-specific homeobox
GB	gegabytes
GEO	Gene Expression Omnibus
GHz	gegahetze
GO	gene ontology
GPR	GenePix result file

Gpt	guanine-hypoxanthine phosphoribosyltransferase
HEK-293	human embyonic kidney cell line 293
HRP	horseradish peroxidase
HSD	Tukey's honestly significant difference
Hspe1	heat shock 10 kd protein 1
HTML	hypertext markup language files
ICAM-1	intercellular adhesion molecule 1
IFN-IP	interferon inducible protein
Iga, Igm	immunoglobin A & M
IgG	immunoglobin G
IGF	Insulin-like growth factor
IGF1R	IGF isoform 1 receptor
IL-4, 13	interleukine 4 & 13
IS	image similarity
K19	keratin19
Kd, Kda	kilo-dalton
KGF	keratinocyte growth factor
LB180	lamellar body protein 180
LMD/SMD	Longhorn/Standford Microarray Database
LOWESS	local weighted scatter plot smooth
LPS	lipopolysaccharide
Mapk13	mitogen activated protein kinase 13
MB	megabytes
MEM	minimal essential medium
Meox2	mesenchyme homeobox 2

Mfge8	o-acetyltransferase milk fat globule membrane protein
Mg50	melanoma-associated antigen
MIAME	minimum information about microarray experiment
MMP9	matrix metalloproteinase 9
MPF	M-phase promoting factor
mRNA	message RNA
Mt2a & mt3	metallothionein 2a and 3
Nf2	neurofibromatosis 2
Nfyc	CCAAT binding transcription factor cbf subunit c
Nup155	nucleoporin 155kd
OSI	organ specificity index
p21$^{WAF/CIP}$	21KD cyclin-dependent kinase inhibitor 1A
P2X7	purinergic receptor P2X 7
P38δ	mitogen-activated protein kinase delta isoform
P450	cytochrome p450
PCA	principal component analysis
PCR	polymerase chain reaction
plaur	plasminogen activator receptor 2
PMT	photomultiplier tube
Pref-1	preadipocyte factor 1
QI	quality index
Retnla	resistin-like molecule alpha
Rgc-32	response gene to complement
RGD	Rat Genome Database
RNA	ribonucleic acid

S100a18	hornerin, also known as HRNR
SAGE	serial analysis gene expression
SAM	significant analysis of microarray
SBR	signal to background ratio
Scd1	stearyl-coa desaturase
SD	standard deviation
SD rats	Sprague-Dawley rats
SDS	sodium dodecyl sulfate
SE	standard error
Slc29a1	solute carrier family 29 transporters, member 1
Slc34a2	solute carrier family 34 sodium phosphate
Sp4	trans-acting transcription factor 4
SP-A, B, C & D	surfactant proteins A, B, C & D
SSC	saline sodium citrate
STAT6	signal transducer and activator of transcription 6,
T1α	type I AEC marker protein α
T7 & SP6	phage T7 & SP6
TGFβ	Transforming growth factor beta
TIFF	Tag Image Format File
TNFα	tumor necrosis factor α
Tnni2, tni3	troponin 2 & 3
Tpa66	tissue-type plasminogen activator
Vim	vimentin

Chapter 1

Identification of Two Novel Markers for

Alveolar Epithelial Type I and II Cells[1,2]

1.1   Abstract

Alveolar epithelial type I and type II cells (AEC I and II) are closely aligned in alveolar surface. There is much interest in the precise identification of AEC I and II in order to separate and evaluate functional and other properties of these two cells. This study aims to identify specific AEC I and AEC II cell markers by DNA microarray using the *in vitro* trans-differentiation of AEC II into AEC I-like cells as a model. Quantitative real time PCR confirmed five AEC I genes: fibroblast growth factor receptor-activating protein 1, aquaporin 5, purinergic receptor P2X 7 (P2X7), interferon-induced protein, and Bcl2-associated protein, and one AEC II gene: gamma-aminobutyric acid receptor pi subunit (GABRP). Immunostaining on cultured cells and rat lung tissue indicated that GABRP and P2X7 proteins were specifically expressed in AEC II and AEC I, respectively. *In situ* hybridization of rat lung tissue confirmed the localization of GABRP mRNA in type II cells. P2X7 and GABRP identified

---

[1] Reprinted from *Biochem Biophys Res Commun*, 319, Zhongming Chen, Nili Jin, Telugu Narasaraju, Jiwang Chen, Lucas R. McFarland, Mary Scott and Lin Liu, Identification of Two Novel Markers for Alveolar Epithelial Type I and II Cells, 774-80, Copyright (2004), with permission from Elsevier.
[2] Contributions of co-authors: Nili Jin, Lucas R. McFarland, Mary Scott performed real-time PCR; Telugu Narasaraju performed immunohistochemistry; Jiwang Chen performed in situ hybridization. Lin Liu is the principal investigator.

in this study could be used as potential AEC I and AEC II markers for studying lung epithelial cell biology and monitoring lung injury.

1.2   Introduction

Alveoli are the gas exchange units of the lung. The alveolar epithelium adapts to this functional role by developing two highly specialized alveolar epithelial cells, cuboidal type II (AEC II) and squamous type I (AEC I). AEC II consists of about 15% of the distal lung cells and occupies 5% of the alveolar surface. AEC II synthesize and secrete lung surfactant, a protein-lipid complex and surface-active material. Lung surfactant stabilizes alveoli by reducing the surface tension. AEC II also maintain the alveolar epithelium by cell proliferation and differentiation, minimize alveolar fluid by transport of sodium from the apical to the basolateral side, and alter the inflammatory process by the secretion of growth factors and cytokines. Surfactant proteins (SP)- A, B, C and D [1;2], and Maclura pomifera lectin [3] have been used as markers of AEC II. Recently, an apical membrane marker of AEC II and a lamellar body limiting membrane marker have also been identified [4;5].

AEC I contribute 7% of total lung cells and covers over 95% of the alveolar surface. This thin epithelium allows the easy diffusion of gases and forms a barrier against the indiscriminate leakage of fluid into alveolar spaces. It also regulates the exchange of physiologically important solutes and water between circulating blood and the alveolar space. Recent studies indicated the presence of sodium channel in AEC I, implying a role in fluid transport [6;7]. AEC I are terminally differentiated, lack mitotic activity, and are easily injured. When the lung is injured, AEC II proliferate and differentiate to AEC I to repair

the damaged alveolar epithelium. AEC I can be identified with marker genes including AQP-5 [8], caveolin-1 [9], and T1 α [10;11].

AEC I and II are closely aligned in alveolar surface and are difficult to separate for evaluating functional and other properties of the two epithelial cells. Cell markers for the precise identification of these cells are not only important for alveolar epithelial cell biology, but also could be used as biochemical markers for lung injury. In this study, we screened the genes differentially expressed in AEC I and AEC II using an *in vitro* culture model and a DNA microarray containing 1090 rat genes. We verified several genes that are specifically expressed in AEC I and AEC II using quantitative real time PCR. Two of the selected genes were also studied in cultured cells and lung tissue.

1.3 Materials and Methods

1.3.1 Materials

TRI reagents were from Molecular Research Center (Cincinnati, OH). DNA-free™ kit was from Ambion (Austin, TX). Atlas™ Glass Fluorescent Labeling kit and Glass™ Rat 1.0 microarray were from Clontech Laboratories (Palo Alto, CA). Cyanine fluorescence dyes, Cy3 and Cy5 were from Amersham-Pharmacia (Piscataway, NJ). QuantiTect™ SYBR® Green PCR kit was from Qiagen (Valencia, CA). GENECLEAN Turbo for PCR was from Qbiogene (Carlsbad, CA). Polyclonal goat anti-gamma-aminobutyric acid receptor pi subunit (anti-GABRP) and anti-purinergic receptor P2X 7 (anti-P2X7) antibodies were from Santa Cruz Biotechnology Inc. (Santa Cruz, CA). Monoclonal anti-LB180 antibodies were from Covance (Richmond CA). Monoclonal anti-T1α (E11) antibodies were a generous gift from Dr. Antoinette

Wetterwald (University of Berne, Switzerland) and Dr. Mary C. Williams (Boston University). FITC-conjugated anti-goat IgG and Cy3-conjugated anti-mouse IgG were from Jackson ImmunoRes (West Grove, PA).

1.3.2 Cell isolation and culture

AEC II were isolated from male Sprague-Dawley rats (180-200 grams) according to the method of Dobbs et al. [12] as previously described [13]. The isolated AEC II (5 x 10^6) were cultured in 100-mm plastic culture dishes containing MEM plus 10% FBS for 1-7 days. Total RNA was isolated from cultured cells using TRI reagents. The trace amount of genomic DNA contamination in the RNA samples was digested with 2U DNase I at 37^{0}C for 30 minutes using the DNA-free™ kit. The quality of RNA was assessed by A_{260}/A_{280}, agarose gel electrophoresis, and DNA contamination by PCR directly using RNA sample and primers of SP-A.

1.3.3 RNA labeling and DNA microarray hybridization

The RNA samples were labeled using the Atlas™ Glass Fluorescent Labeling kit. Because of difficulties in isolating AEC I cell preparation, we chose *in vitro* culture of isolated AEC II as a model of AEC I. It is well established that AEC II converts to AEC I-like cells when cultured on plastic dishes for 7 days [11;14]. In the present study, we chose 1-day and 7-day culture on plastics of isolated type II cells to represent AEC II and AEC I, respectively, for the microarray experiments. We used "AEC II" and "AEC I" for the 1-day and 7-day culture of AEC II in the rest of text. Two hybridizations were performed using independent RNA preparations, each pooled from two independent cell preparations. Briefly, 20 µg of total RNA isolated from "AEC II" an "AEC I" were reverse-transcribed into cDNA with the incorporation of

aminoallyl-dUTP (aa-dUTP), followed by the Cy3 or Cy5 coupling reaction. The labeled cDNAs were incubated with an Atlas Glass™ Rat 1.0 microarray at 50^0C overnight using the hybridization buffer and chamber provided with the kit. After hybridization, the microarray slide was washed and scanned with ScanArray 3000 (Packard BioChip Technologies, Billerica, MA).

1.3.4 Data analysis

The hybridization images of DNA microarray were analyzed with the software *GenePix Pro 4.0* (Axon Instruments, Union City, CA), which provides fluorescence intensity, standard deviation, and pixels of each spot and its adjacent background area. The raw data were imported into Microsoft™ *Excel* worksheets, as well as the respective genes' names, Genbank accession numbers, and function categories provided by the Atlas™ Glass Rat 1.0 microarray user manual. The fluorescence intensities and standard deviations of each gene were denoted as *"F650 mean"*, *"F650 SD"* for "AEC I", and *"F535 mean"*, *"F535 SD"*, for "AEC II". The pixels of each spot were marked as *"F pixels"*. The background information was similarly marked, except that *"B"* was used instead of *"F"*.

Weak spots were filtered if the mean fluorescence intensity of each spot was higher than that of the background as determined by one-tailed t-test. The

formula, $t = \dfrac{"F\,mean"-"B\,mean"}{\sqrt{\dfrac{"F\,SD"^2}{"F\,pixels"-1}+\dfrac{"B\,SD"^2}{"B\,pixels"-1}}}$, was adapted from the standard t-test

with a degree of freedom equal to $"F\,pixels"+"B\,pixels"-2$, p ≤ 0.01. Any spots with p>0.01 in both channels were filtered. Raw gene expression ratios of "AEC I" vs. "AECII" were calculated as

$$"raw\,ratio" = \frac{"F650\,mean" - "B650\,mean"}{"F535\,mean" - "B535\,mean"}$$. The raw ratios were normalized as

$$"ratio" = \frac{"raw\,ratio"}{"normalization\,coefficient"}$$. The *"normalization coefficient"* was the

median of the raw ratios of the genes after filtering. This normalization is based on the common assumption that normalized ratios would have a median of 1, which means most of the genes do not show changes of expression levels. The genes that consistently increased or decreased by 2-fold in both hybridizations were identified as differentially expressed genes.

1.3.5   Quantitative real-time PCR

Data validation of DNA microarray hybridization was performed by real-time quantitative PCR using QuantiTect™ SYBR® Green PCR kit following the manufacturer's instruction. Primers (Table 1,) were designed with either Primer 3 (http://www-genome.wi.mit.edu/cgi-bin/primer/primer3_www.cgi) or Primer Express 1.5 (Applied Biosystems, Foster City, CA). PCR Amplification was performed on ABI PRISM 7700 (Applied Biosystems, Foster City, CA). PCR products were purified with GENECLEAN Turbo for constructing standard curves ($10\sim10^9$ copies). A standard curve was plotted with the threshold cycle (C_T) versus the logarithmic value of the gene copy number. For most of the assays, gene copy number of unknown samples was generated directly from the standard curve by the software Sequence Detector 1.7. In some circumstances, logarithmic gene copy number was calculated with the best standard curve with formula (observed C_T – Y intercept)/slope. At least 2 duplications were run for each standard or unknown sample of three or more cell preparations. All gene copy numbers were normalized to 18S rRNA.

Table 1.1 Sequences of primer sets and thermal conditions used for real-time PCR. GABRP: gamma-aminobutyric acid receptor pi subunit, FRAG1: fibroblast growth factor receptor-activating protein 1, AQP5: aquaporin 5, P2X7: Purinergic receptor P2X 7, IFN-IP: Interferon-induced protein, BAX: Bcl2-associated protein. The Tm of the primers was set from 58°C to 60°C. Primers with similar GC content ($\leq$ 60%) were chosen. All the amplicons were ~ 100 bp. The Tm of amplicons was obtained from the dissociation analysis.

Gene name	Sequences of primer sets		Amplicon Tm (^{o}C)	Data acquisition Temperature (^{o}C)
GABRP	Forward	AAATTTCTGGCGACAATGTCAACTA	77	74
	Reverse	GTAATCAATGATTCTGCCGATCTTT		
FRAG1	Forward	GGGAACAAAGAGCTGCTAATAACCT	78	76
	Reverse	ACTGAGATAGGGAAGGCCAGAGT		
AQP-5	Forward	GGAGCAGAAGCCCATAATGG	83	76
	Reverse	TGGCAGACATTCTGCCTGATT		
P2X7	Forward	CTGCAAGATGTCAAAGGTCAAGAC	81	78
	Reverse	TGGACCTAGGAACCGCTTCTATC		
IFN-IP	Forward	CAGCATCCTCATGGTCATTATCAC	78	76
	Reverse	ACGTCAGGATCAGAAACCAGAATC		
BAX	Forward	TGGCAGCTGACATGTTTGCT	82	76
	Reverse	TTTAGTGCACAGGGCCTTGAG		
18S rRNA	Forward	TCCCAGTAAGTGCGGGTCATA	81	76
	Reverse	CGAGGGCCTCACTAAACCATC		

1.3.6   Immunocytochemistry and immunohistochemistry

AEC II fixed with 4% paraformaldehyde were stained with anti-GABRP with polyclonal goat anti-GABRP or anti-P2X7 antibodies (1:100 dilution), followed by incubation with FITC-conjugated anti-goat IgG (1:2000 dilution). Rat lung tissue sections were double-labeled with polyclonal goat anti-GABRP or anti-P2X7 and monoclonal anti-LB180 antibodies or anti-T1α (1:1000 dilution), followed by FITC-conjugated anti-goat IgG and Cy3-conjugated anti-mouse IgG (1:2500 dilution) as previously described [15]. The slides were examined using a Nikon Eclipse E600 fluorescence microscope.

1.3.7 *In situ* hybridization

Digoxygenin (DIG)-labeled sense and antisense probes were synthesized from PCR-amplified cDNA (371 bp) using T7 and SP6 RNA polymerase, respectively [16]. The primer sequences for cDNA amplification were 5'-TAATACGACTCACTATAGGGTCACGGTGGGAAACAGG-3' (forward) and 5'- ATTAGGTGACACTATAGAAGGAAGGTGGAAGGGACG-3' (reverse). The procedure used for *in situ* hybridization was adapted from the protocol [17]. The slide was examined with Nikon Eclipse 600 microscope (Nikon Corp., Tokyo. Japan).

1.4   Results

DNA microarray hybridizations were performed using Atlas Glass™ Rat 1.0 microarray containing 1090 genes to screen differentially expressed genes between "AEC II" and "AEC I" for the purpose of identifying cell markers. As the first step of data analysis, we filtered the spots where fluorescent intensities were not significantly higher than the background via one-tailed t-test ($p \leq 0.01$) using the raw data imported from *GenePix Pro*. As a result, 492 genes were filtered from two independent hybridizations, and 596 genes remained in both hybridizations. The positive controls, cDNA synthesis and coupling reaction controls, and hybridization control, showed high ratios of signal to background (>5) in both hybridizations, indicating the success of the respective reactions during the RNA labeling and hybridization. Based on a 2-fold ratio change threshold, we found that 18 genes were differentially expressed in AEC I and 42 genes in AEC II. Because only two hybridizations were carried out for screening purpose, we chose to not present the full set of microarray data at the present time.

To validate the differentially expressed genes with quantitative real-time RT-PCR, we first tested the precision and reproducibility of the protocol. All the selected genes were amplified with amplification efficiencies of >90%. The C_T at 10^8 copies ranged from 13 to 16 cycles. All the standard curves were linear in the range from 10 to 10^8 copies with correlation coefficients of >0.99. The intra-plate coefficient of variation (CV) was less than 2%. These data indicated that the quantitative real-time PCR was precise. The reproducibility was demonstrated by the inter-plate CV, which ranged in 1~6%, depending on genes and primers. The sensitivity of the protocol was about 10 copies. However, low copy numbers normally have higher deviations. We then verified 6 genes with a high-fold change from the microarray data, 5 for AEC I: fibroblast growth factor receptor-activating protein 1 (FGFR1), aquaporin 5 (AQP-5), purinergic receptor P2X 7 (P2X7), interferon inducible protein (IFN-IP), and Bcl2-associated X protein (BAX); 1 for AEC II: gamma-aminobutyric acid receptor (GABRP) (**Fig.** 1.1).
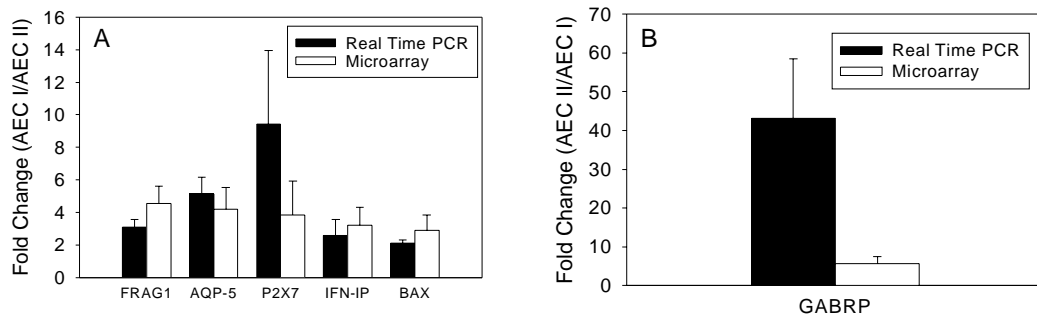
Fig. 1.1 **Comparison of gene expression fold changes by DNA microarray and real time PCR.** (A) "AEC I"-prominent. The fold changes are shown as "AEC I"/"AEC II". (B) "AEC II"-prominent. The fold changes are shown as "AEC II"/"AEC I". Data are expressed as mean ± SE or range. Microarray data represent two independent cell preparations for each hybridization, n=2 for two hybridizations. Real time PCR are in duplicate for each assay, n=3-4 for cell preparations.

Based on the availability of antibodies, we chose GABRP and P2X7 for verification at the protein level in cultured cells and lung tissue. Isolated AEC II cells were cultured on plastic dishes for 1 and 7 days, and immunostained with anti-GABRP or anti-P2X7 antibodies. As shown in **Fig.** 1.2, GABRP antibodies stained day 1 cultured cells, but not day 7 cultured cells, whereas P2X7 antibodies recognized day 7 cultured cells, but not day 1 cultured cells. The control cells in which primary antibodies were omitted did not generate signals (data not shown).
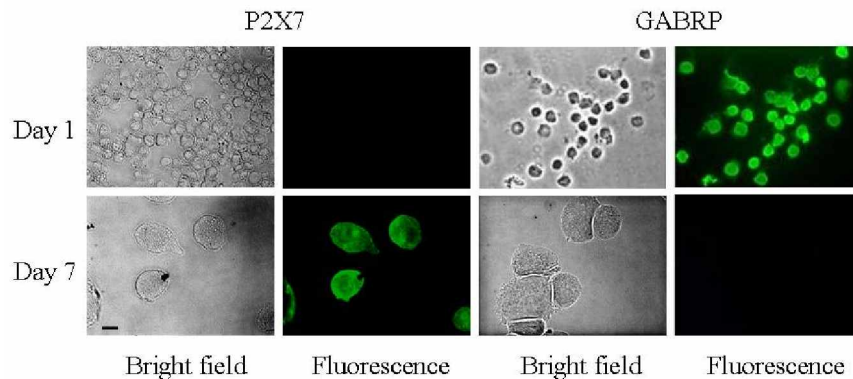
Fig. 1.2 Immunostaining of cultured type II cells for GABRP and P2X7. Isolated AEC II were cultured on plastics dishes for 1 day and 7 days. The 7-day cultured cells were trypsinized and cytospun to a slide. After fixed with 4% pararformaldehyde, the cells were stained with anti-GABRP or anti-P2X7 antibodies, followed by FITC-conjugated anti-goat IgG. Scale bar: 10 μm.

When rat lung tissue was double-labeled with anti-GABRP and anti-LB180 antibodies (a lamellar body limited membrane protein as AEC II marker), co-localization of GABRP and LB-180 was observed, indicating that GABRP was specifically expressed in AEC II, but not in AEC I (**Fig.** 1.3). However, GABRP also stained the bronchial epithelial cells (data not shown). Similarly, P2X7 was co-localized with the known AEC I marker, T1α (**Fig.** 1.3). The

control slides in which primary antibodies were omitted did not show fluorescence (data not shown).
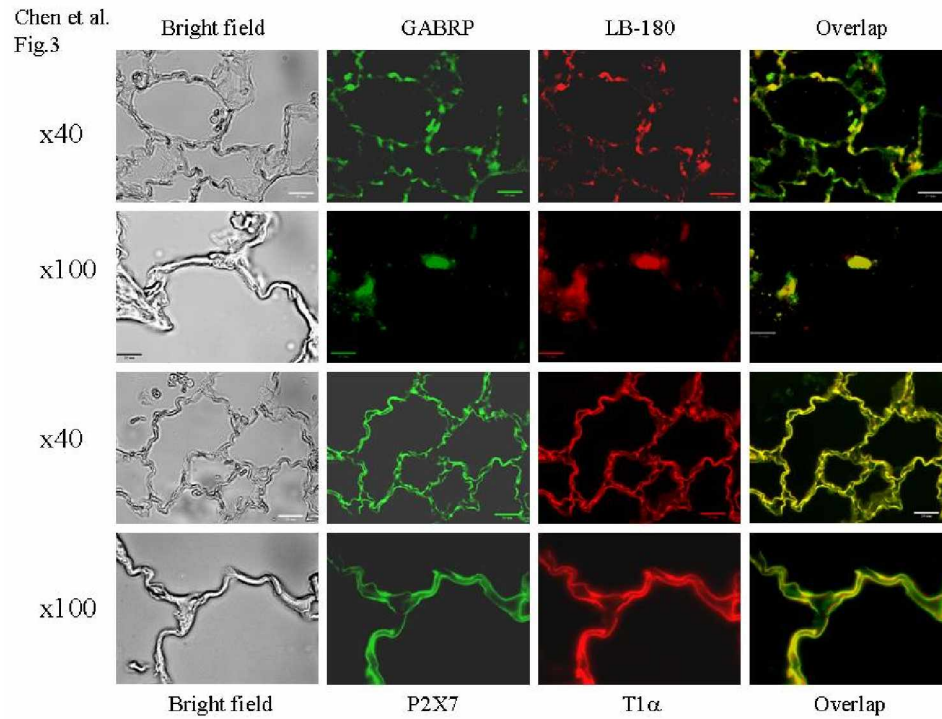
Fig. 1.3 Dual-labeling of rat lung tissue for GABRP and LB-180 or P2X7 and T1α. Rat lung tissue was fixed with 4% formaldehyde and incubated with anti-GABRP and anti-LB180 antibodies or anti-P2X7 and T1α, followed by FITC-conjugated anti-goat IgG and Cy3-conjugated anti-mouse IgG. Two different magnifications (x 40 and x100) are shown. Scale bar: 20 μm.
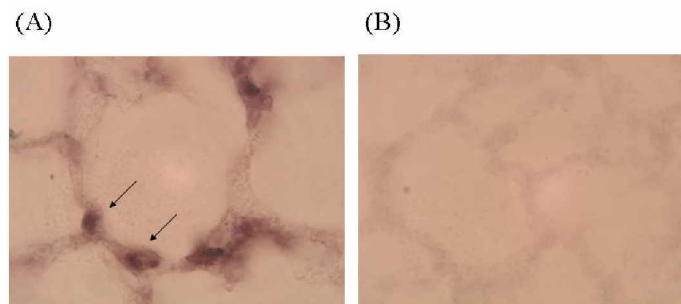
Fig. 1.4 In situ hybridization of GABRP in rat lung tissue. Rat lung tissue was formaldehyde-fixed and paraffin-embedded. The slide was hybridized with DIG-labeled antisense (A) or sense (B) probe and detected by DIG Detection kit. Arrows point to staining of type II cells.

To confirm that GABRP mRNA is expressed in type II cells *in vivo*, we performed *in situ* hybridization using rat lung tissue. A GABRP fragment (371 bp) containing T7 (sense strand) or SP6 (anti-sense strand) RNA polymerase promoter was generated by PCR. Digoxigenin (DIG) labeled anti-sense- and sense cRNA probes were produced *in vitro* using SP6 or T7 RNA polymerase. As shown in **Fig.** 1.4, the anti-sense GABRP probe binds to the cells located in the corner, suggesting its location of type II cells. The sense probe did not generate signals.

1.5   Discussion

The regeneration of squamous alveolar epithelium following lung injury occurs by proliferation of AEC II with subsequent trans-differentiation into AEC I. Isolated AEC II, when cultured on plastic dishes for several days acquires AEC I-like phenotypic characteristics. Utilizing this unique property of AEC II, the present study was designed to identify the genes that can be used as markers for AEC I and AEC II by DNA microarray containing 1090 rat genes, real time PCR, immunohistochemistry and *in situ* hybridization.

We have verified one AEC II gene (GABAP) and five AEC I genes (FRAG1, AQP5, P2XP7, IFN-IP and BAX) using real time PCR. GABRP, a pi subunit of the GABA "A" receptor family, was specifically expressed in AEC II but not in AEC I cells as revealed by DNA microarray, real time PCR and immunostaining and in situ hybridization on rat lung tissue. Although it exists in brain, uterus, and some extent in lung, this is the first report to demonstrate its localization in AEC II. The expression of GABRP has been directly correlated with uterine contractility [18]. GABRP is expressed throughout pregnancy and

decreases considerably during labor. Transfection of GABRP in HEK-293 did not express binding sites for the GABA receptor ligand (^{35}S TBPS) and failed to elicit chloride currents in response to GABA or glycine. However, transfection along with GABA receptor beta 3 showed the change in the sensitivity of the receptors to pregnanolone, suggesting that the GABRP-GABA receptor beta 3 complex form a active channel [19]. The function of GABRP in AEC II is not known. Because GABA receptors are Cl$^-$ channels, GABRP may regulate fluid transport in lungs.

Among the genes expressed in AEC I, AQP-5 is a known marker of AEC I and functions as a water channel. AQP-5 is expressed in adult rat lung and participates in fluid clearance [20]. AQP-5 $^{-/-}$ mice showed a higher hyper-responsive level than did AQP-5 $^{+/+}$ mice to intravenously administered acetylcholine, indicating that AQP-5 influences bronchoconstriction [21]. FRAG1, discovered as a cancer-promoting gene, may activate fibroblast growth factor receptor by C-terminal fusion through chromosomal rearrangement and stimulate the transforming activity and autophosphorylation of FGF [22;23].

Purinergic receptors are composed of P2Y and P2X, and are activated by extracellular nucleotides. Binding of ATP to P2Y receptor activates the G-protein coupled to phospholipase C, raises inositol-1,4,5-triphosphate and intracellular Ca^{2+}, and, thus, increases surfactant secretion in AEC II [24]. P2X receptors are ATP-gated ion channels. Isoforms of P2X receptors were found in respiratory epithelium [25;26]. P2X4 has been detected in AEC I [27]. The activation of P2X stimulates secretory Cl$^-$ transport across epithelia. Recent studies have indicated that AEC I express Na$^+$ channels and AEC I may

contribute to fluid transport in the lung [6;7]. The present finding that P2X7 was expressed higher in AEC I suggested that P2X7 may regulate Cl⁻ transport.

Interferons induce gene expressions of various proteins including antiviral proteins, chemokines, heat shock proteins, etc. [28]. IFN-IP has 70% similarity to the human alpha-IFN-induced gene, 9-27 [29]. The latter is a cell surface protein, leu-13 and mediates antiproliferation and cell adhesion [30]. The function of IFN-IP in AEC I is yet to be determined.

BAX is a pro-apoptotic regulatory protein. Activation of BAX leads to its relocation from cytosol to mitochondria, causes the release of cytochrome c, and promotes apoptosis [31]. A higher expression of BAX in AEC I than AEC II indicates that AEC I may undergo active apoptosis.

In summary, the present study identified several genes specifically expressed in AEC II or AEC I. Some of those genes could be used as potential markers for these cell types.

1.6   Acknowledgement

1.7 REFERENCES

[1] M. Kalina, R.J. Mason, J.M. Shannon, Surfactant protein C is expressed in alveolar type II cells but not in Clara cells of rat lung, Am. J. Respir. Cell. Mol. Biol. 6 (1992) 594-600.

[2] H. Shimizu, J.H. Fisher, P. Papst, B. Benson, K. Lau, R.J. Mason, D.R. Voelker, Primary structure of rat pulmonary surfactant protein D. cDNA and deduced amino acid sequence, J. Biol. Chem. 267 (1992) 1853-1857.

[3] L.G. Dobbs, M.C. Williams, A.E. Brandt, Changes in biochemical characteristics and pattern of lectin binding of alveolar type II cells with time in culture, Biochim. Biophys. Acta. 846 (1985) 155-166.

[4] G.M. Boylan, J.G. Pryde, L.G. Dobbs, M.C. McElroy, Identification of a novel antigen on the apical surface of rat alveolar epithelial type II and Clara cells, Am. J. Physiol. Lung Cell. Mol. Physiol. 280 (2001) L1318-L1326.

[5] K. Zen, K. Notarfrancesco, V. Oorschot, J.W. Slot, A.B. Fisher, H. Shuman, Generation and characterization of monoclonal antibodies to alveolar type II cell lamellar body membrane, Am. J. Physiol. 275 (1998) L172-L183.

[6] M.D. Johnson, J.H. Widdicombe, L. Allen, P. Barbry, L.G. Dobbs, Alveolar epithelial type I cells contain transport proteins and transport

sodium, supporting an active role for type I cells in regulation of lung liquid homeostasis, Proc. Natl. Acad. Sci. U. S. A. 99 (2002) 1966-1971.

[7] Z. Borok, J.M. Liebler, R.L. Lubman, M.J. Foster, B. Zhou, X. Li, S.M. Zabski, K.J. Kim, E.D. Crandall, Na transport proteins are expressed by rat alveolar epithelial type I cells, Am. J. Physiol. Lung Cell. Mol. Physiol. 282 (2002) L599-L608.

[8] S. Raina, G.M. Preston, W.B. Guggino, P. Agre, Molecular cloning and characterization of an aquaporin cDNA from salivary, lacrimal, and respiratory tissues, J. Biol. Chem. 270 (1995) 1908-1912.

[9] G.R. Newman, L. Campbell, C. von Ruhland, B. Jasani, M. Gumbleton, Caveolin and its cellular and subcellular immunolocalisation in lung alveolar epithelium: implications for alveolar epithelial type I cell function, Cell Tissue Res. 295 (1999) 111-120.

[10] R.F. Gonzalez, L.G. Dobbs, Purification and analysis of RTI40, a type I alveolar epithelial cell apical membrane protein, Biochim. Biophys. Acta 1429 (1998) 208-216.

[11] Z. Borok, S.I. Danto, R.L. Lubman, Y. ao, M.C. Williams, E.D. Crandall, Modulation of t1alpha expression with alveolar epithelial cell phenotype *in vitro*, Am. J. Physiol. 275 (1998) L155-L164.

[12] L.G. Dobbs, R. Gonzalez, M.C. Williams, An improved method for isolating type II cells in high yield and purity, Am. Rev. Respir. Dis. 134 (1986) 141-145.

[13] L. Liu, M. Wang, A.B. Fisher, U.J.P. Zimmerman, Involvement of annexin II in exocytosis of lamellar bodies from alveolar epithelial type II cells, Am. J. Physiol. 270 (1996) L668-L676.

[14] J.M. Cheek, M.J. Evans, E.D. Crandall, Type I cell-like morphology in tight alveolar epithelial monolayers, Exp. Cell Res. 184 (1989) 375-387.

[15] T.A. Narasaraju, N. Jin, C.R. Narendranath, Z. Chen, D. Gou, L. Liu, Protein nitration in rat lungs during hyperoxia exposure: a possible role of myeloperoxidase, Am. J. Physiol. Lung Cell. Mol. Physiol. 285 (2003) L1037-L1045.

[16] I.D. Young, L. Ailles, K. Deugau, R. Kisilevsky, Transcription of cRNA for in situ hybridization from polymerase chain reaction-amplified DNA, Lab. Invest. 64 (1991) 709-712.

[17] P. Komminoth, J. Roth, S. Schroder, P. Saremaslani, P.U. Heitz, Overlapping expression of immunohistochemical markers and synaptophysin mRNA in pheochromocytomas and adrenocortical carcinomas. Implications for the differential diagnosis of adrenal gland tumors, Lab. Invest. 72 (1995) 424-431.

[18] E. Fujii, S.H. Mellon, Regulation of uterine gamma-aminobutyric acid(A) receptor subunit expression throughout pregnancy, Endocrinology 142 (2001) 1770-1777.

[19] E. Hedblom, E.F. Kirkness, A novel class of GABAA receptor subunit in tissues of the reproductive system, J. Biol. Chem. 272 (1997) 15346-15350.

[20] Y. Song, N. Fukuda, C. Bai, T. Ma, M.A. Matthay, A.S. Verkman, Role of aquaporins in alveolar fluid clearance in neonatal and adult lung, and in oedema formation following acute lung injury: studies in transgenic aquaporin null mice, J. Physiol. 525 (2000) 771-779.

[21] C.M. Krane, C.N. Fortner, A.R. Hand, D.W. McGraw, J.N. Lorenz, S.E. Wert, J.E. Towne, R.J. Paul, J.A. Whitsett, A.G. Menon, Aquaporin 5-deficient mouse lungs are hyperresponsive to cholinergic stimulation, Proc. Natl. Acad. Sci. U. S. A. 98 (2001) 14114-14119.

[22] M.V. Lorenzi, Y. Horii, R. Yamanaka, K. Sakaguchi, T. Miki, FRAG1, a gene that potently activates fibroblast growth factor receptor by C-terminal fusion through chromosomal rearrangement, Proc. Natl. Acad. Sci. U. S. A. 93 (1996) 8956-8961.

[23] M.V. Lorenzi, P. Castagnino, D.C. Aaronson, D.C. Lieb, C.C. Lee, C.L. Keck, N.C. Popescu, T. Miki, Human FRAG1 encodes a novel membrane-spanning protein that localizes to chromosome 11p15.5, a region of frequent loss of heterozygosity in cancer, Genomics 62 (1999) 59-66.

[24] A.M. Gilfillan, S.A. Rooney, Functional evidence for involvement of P2 purinoceptors in the ATP stimulation of phosphatidylcholine secretion in type II alveolar epithelial cells, Biochim. Biophys. Acta 959 (1988) 31-37.

[25] J. Leipziger, Control of epithelial transport via luminal P2 receptors, Am. J. Physiol. Renal Physiol. 284 (2003) F419-F432.

[26] A.L. Taylor, L.M. Schwiebert, J.J. Smith, C. King, J.R. Jones, E.J.

Sorscher, E.M. Schwiebert, Epithelial P2X purinergic receptor channel expression and function, J. Clin. Invest. 104 (1999) 875-884.

[27] R. Qiao, B. Zhou, J.M. Liebler, X. Li, E.D. Crandall, Z. Borok, Identification of three genes of known function expressed by alveolar epithelial cells, Am. J. Respir. Cell. Mol. Biol. 29 (2003) 95-105.

[28] G.C. Sen, Novel functions of interferon-induced proteins, Semin. Cancer Biol. 10 (2000) 93-101.

[29] D.J. Hayzer, E. Brinson, M.S. Runge, A rat beta-interferon-induced mRNA: sequence characterization, Gene 117 (1992) 277-278.

[30] G.A. Deblandre, O.P. Marinx, S.S. Evans, S. Majjaj, O. Leo, D. Caput, G.A. Huez, M.G. Wathelet, Expression cloning of an interferon-inducible 17-kDa membrane protein implicated in the control of cell growth, J. Biol. Chem. 270 (1995) 23860-23866.

[31] L.M. Sutherland, Y.S. Edwards, A.W. Murray, Alveolar type II cell apoptosis, Comp. Biochem. Physiol. A Mol. Integr. Physiol. 129 (2001) 267-285.

Chapter 2

***RealSpot*: Software Validating Results from**

DNA Microarray Data Analysis with Spot Images[3,4]

2.1 ABSTRACT

The spot images from DNA microarray highly affect the discovery of biological knowledge from gene expression data. However, results from quality analysis, normalization, differential expression, and cluster analysis are rarely validated with spot images in current data analysis methods or software packages. We designed *RealSpot*, a software package, to validate the results by directly associating spot quality and data with spot images in a spreadsheet table. *RealSpot* splits hybridization images into individual spots stored in a spreadsheet table. It subsequently associates microarray data with spot images and performs data validation through the standard table operation such as sorting, searching, and editing. *RealSpot* has several built-in functions to facilitate data validation, including spot quality analysis, data organization, one-way ANOVA, gene ontology association, verification, import, and export. We used *RealSpot* to evaluate 77 slides (30,000 features each) from real hybridization experiments, and to validate results from each step of data

---

[3] Reprinted from *Physiol Genomics*, 21, Zhongming Chen and Lin Liu, *RealSpot*: Software Validating Results from DNA Microarray Data Analysis with Spot Images, 284-91, Copyright (2005), with permission from American Physiological Society.
[4] Contributions of co-authors: Liu L is the principal investigator.

analysis. It took about 10 min to validate results of spot quality after initial evaluation and correct ~ 0.3% of falsely assigned qualities of 10K spots. We validated 1,641 of 2,110 differentially expressed genes identified by SAM analysis in about half an hour by comparing each gene with its respective spot image. Furthermore, we found 6 of 48 genes in one cluster from k-mean clustering method showed inconsistent trends of spot images. *RealSpot* is efficient for validating microarray results and thus helpful for improving the reliability of the whole microarray experiment for experimentalists.

2.2 Introduction

DNA microarray is a high throughput technique for the investigation of mRNA abundance (5). Gene probes on one slide are hybridized simultaneously with cDNA or cRNA samples labeled with fluorescence dyes, yielding images containing thousands of spots. The mRNA abundance is obtained by extracting quantification information from all spots on the images ("raw data"). According to the minimum information about microarray experiment (MIAME) guidelines, a microarray experiment is organized as gene probes, print layout, samples, hybridization images, raw data, and data normalization (2).

Raw data from the hybridization images are the fundamental information for further data analysis including data normalization (8), statistical inference (10), cluster analysis (4), principal component analysis (PCA) (9), pathway construction (3; 11) and data interpretation. The results of a microarray experiment depend on the quality of hybridization images and the respective raw data sets. However, results from each analysis step are rarely validated with spot images in current data analysis methods or software packages. A

typical data analysis procedure of microarray is as following: extracting quantification data from images, filtering data with chosen standards (e.g. background-to-noise ratio >2), normalizing $\log_2$ ratios, identifying differentially expressed genes (e.g. 2-fold change of expression or statistical inference with p=5), clustering genes (e.g. hierarchic or k-mean clustering), and selecting target genes for further study. Microarray data analyses confront with challenges of diverse methods, standards and software packages. An example is spot quality evaluation discussed below.

The image quality varies from spot to spot due to printing, sample quality and hybridization. Ideally, all the spots with a poor quality should be filtered before further data analysis. There are two main approaches for filtering poor quality spots: manually flagging spots and automatically filtering genes (1; 6; 12; 13). Manually flagging spots is time-consuming due to the large number of spots in a microarray slide. Automatic methods based on quantification information are fast and efficient. These methods calculate composite scores from spot size, intensity, signal to background ratio (SBR), and/or circularity coefficient (area-to-perimeter ratio). Generally, the composite scores represent several aspects of spot images. These methods may fail at spots with irregular morphology, e.g., donuts, black holes ("ghost images"), and tiny-dust contained spots. Although they are suitable for large scale data analysis, different methods or parameters frequently generate non-trivial different results from an identical raw data set. In such a case, to validate results and choose an appropriate method is not straightforward. The association of spot images with derived data may identify irregular spots. Some of software packages such as Acuity and Longhorn/Standford Microarray Database (LMD/SMD) can

locate each spot image on a single slide (7). Acuity locates a spot image on a scanned slide image using *GenePix* results, while LMD can show a spot image as well as other retrieved data in a data query report. In both cases, only a single spot image can be retrieved to validate the derived data from one slide.

Here, we report a software package, *RealSpot*, for validating results from dual-color DNA microarray hybridizations. *RealSpot* evaluates spot quality and validates it with spot images. *RealSpot* links the images and raw data of each spot side-by-side and organizes them in a spreadsheet table. By standard table operation such as sorting, searching, and editing, a user can directly compare the spot images, raw data, and processed data in an efficient and reliable way. Furthermore, *RealSpot* provides tools for one-way ANOVA, gene ontology, and web page export, which are also helpful in choosing target genes for further study and thus improving the reliability of the whole microarray experiment. It is freely available for academic use, and can be obtained at http://www.cvm.okstate.edu/research/Facilities/LungBiologyLab/*RealSpot*.htm or via e-mail request (liulin@okstate.edu).

2.3   Implementation

2.3.1   General workflow

The basic workflow of *RealSpot* is composed of 5 modules: data import, quality evaluation, data organization, data verification, and data export (**Fig. 2.1**). The data import module converts raw data, images, and sample information into a spreadsheet table. The quality evaluation module evaluates data quality using spot images and raw data from each slide and stored as a binary slide file. The data organization module organizes slide files as an experiment, visualizes data in a meta-table based on the sample information,

performs one-way ANOVA, and associates gene ontology information (http://www.geneontology.org) with gene list in a microarray experiment. The data verification module validates the results by searching a group of genes obtained from down-stream data analysis, e.g. K-mean cluster, and then comparing them with raw data and spot images. The data export module exports spot images, quality index, and normalized data as images, web pages, and text files for data presentation, internet data communication, and further analysis. Each component is implemented as a module to guide a user in finishing the respective operations step by step. The following is the detailed description of these components.

2.3.2 Data import

An import module imports raw data and image files, and collects sample information of the respective slide. The raw data files such as *GenePix* result (GPR) file are tab-delimited flat text files containing gene ID, gene name, geometry of subarray grids (block or meta-row and meta-column), and spots in each subarray (row and column). The GPR file also holds spot geometry (x, y, and diameter) for image splitting, and gene expression data (fluorescence intensity, background, and ratio). The image files are 16-bit TIFF (Tag Image Format File, 16-bit image pixel: 0~65,531) files directly generated by a scanner, e.g. ScanArray™ Express. During data import, each 16-bit TIFF image is split into spot images using the spot geometry x, y, and diameter from the respective raw data file. The spot images are then linearly transformed into 8-bit for visualization. The linear transformation is based on the whole image or individual sub-grid. By default, the lowest 5 percent image data are converted to 0, the highest 5% to 255, and the rest between 0 and 255, calculated from

$$f_8 = \frac{F_{16} - P_5}{P_{95} - P_5}$$, where f_8 is the intensity of a transformed 8-bit image pixel

(0~255 for image visualization), F_{16} is the original 16-bit fluorescence intensity

of each pixel, and P_5 and P_{95} are the 16-bit intensities at the 5th and 95th

percentiles from a slide image or a sub-grid, respectively.
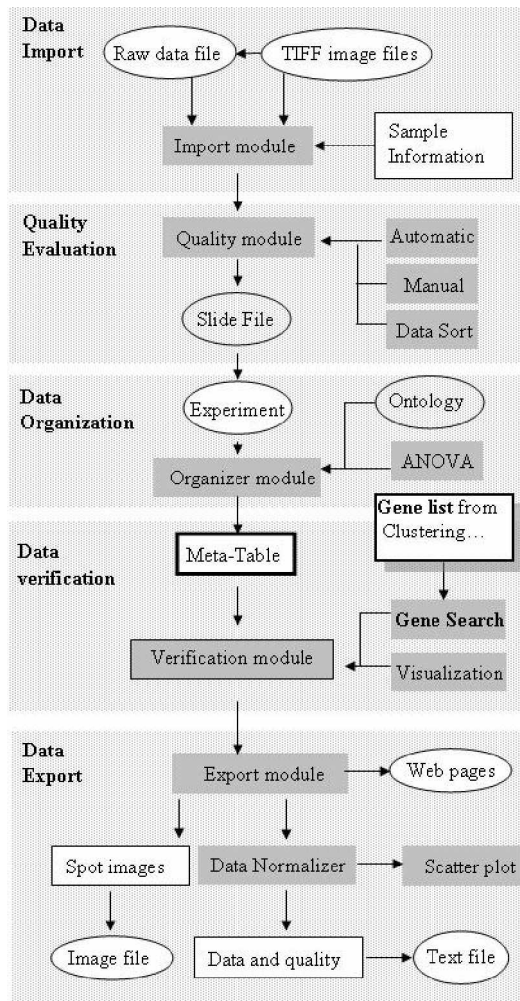
Fig. 2.1 Overview of *RealSpot*. There are five components shown as gray boxes in *RealSpot*. Import module imports raw data file and TIFF image files into *RealSpot*. Quality module evaluates spot images and data, and saves them as a slide file. Data sort function facilitates the quality evaluation. Oraganizer module organizes multiple slides in a Meta-table. Verification module confirms the final results with spot images and raw data by searching selected genes. Export module exports spot images or data quality and gene expression data of selected genes. Data Normalizer normalizes raw data and draws a scatter plot. The solid text boxes are the functional modules in *RealSpot*, the rectangle and ovary white text boxes are sample and data information, and the files used by *RealSpot*, respectively. The arrows indicate the workflows of data or module calling.

During data import, a user is also asked to import sample information

including the sample names and the respective dye channels (green or red). In

a *RealSpot* spreadsheet, each row represents a gene and each column

contains gene probes (gene ID and name), information of array print layout

such as block, and sub-array, spot images, and raw data such as fluorescence intensity and background (**Fig.** 2.2). Additional columns are added to the table, e.g., quality index, 16-bit spot signal, and SBR calculated directly from 16-bit TIFF images (see below).

Table 2.1 Quality index and icon

Icon	Meaning	Default threshold*	Quality Index
	Empty spot	<30%	0
	Weak and ambiguous spot		1
	Middle and clear spot**		2
	Strong spot		3
	Saturated spot	>=95%	4
	Contaminated spot	SBR<2	5

* Signal percentiles of 30% and 95% were obtained from a plot of intensity-gene rank percentage. SBR: Signal-to-background ratio. See the text for details.

** Compared with weak and ambiguous spot, the spot is stronger and its shape is clear from background.

2.3.3   Quality evaluation

The main purpose of *RealSpot* is the quality evaluation of spot images and raw data, and the verification of the final data analysis results with the spot images. *RealSpot* first evaluates spot quality based on the signal intensity and SBR, and assigns a quality index (QI) to each spot. QI 0-4 indicate empty, weak, middle, strong and saturated spots, respectively (Table 2.1). By default, QI 0 and 4 are assigned to the empty and saturated spots, whose intensities are less than 30% and greater than 95%, respectively. QI 1-3 is calculated, based on the intensity of 16-bit spot signals, as: $QI_{ij} = round(\frac{I_{ij} - I_0}{I_1 - I_0} * 4)$, where

QI$_{ij}$ is the quality index of spot j on slide i and I_{ij} the intensity of the spot j on

slide i. By default, I_0 is the intensity at 30th percentile, and I_1 at 95th percentile of the plot (intensity vs. gene rank percentage) of the slide image. These settings may be adjusted based on visual estimation (**Fig.** 2.3). A QI of 5 is assigned to a contaminated or bad spot based on SBR. By default, any spots with a SBR of <2.0 are given a QI of 5. QI is visualized as an icon (Table 2.1). The QI 0~4 are shown as columns, the shorter height the weaker intensity. A QI of 5 is shown as a prohibiting cross.
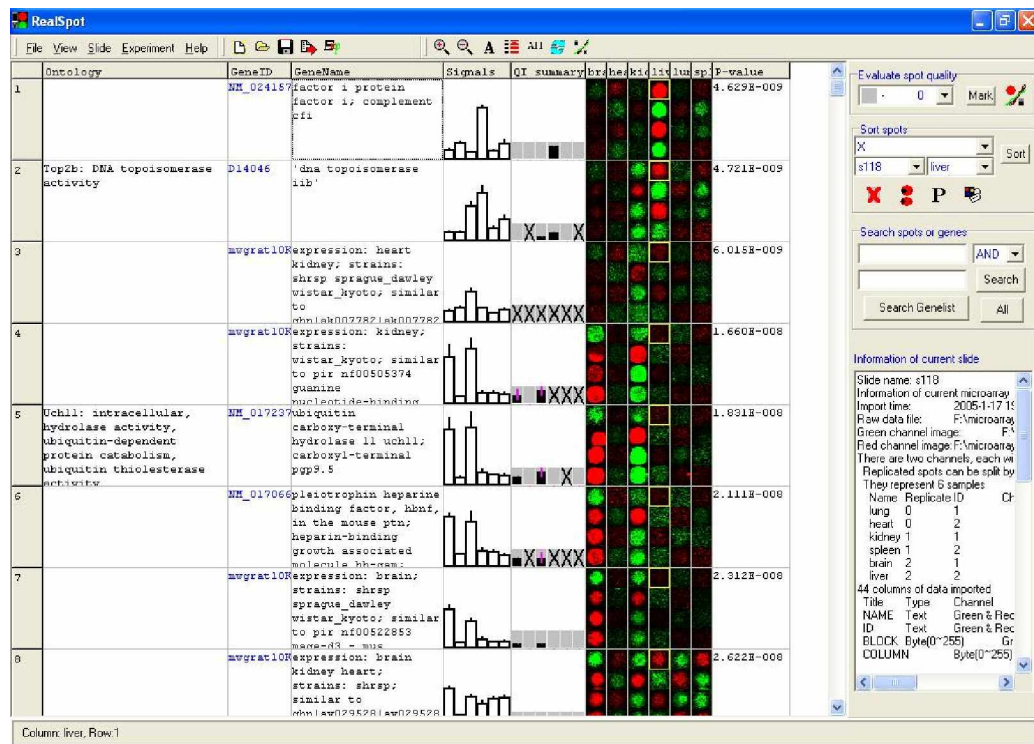
Fig. 2.2 Data visualized in a Meta-table. Tables from four slides are grouped as a meta-table, in which spot image cells contain sub-rows and sub-columns. Each row represents the information of a gene. The columns from left to right are gene ontology, gene ID, gene name, signal summary, quality index (QI) summary, spot images of each sample from each slide, and data column (P-value) for sorting the table. The yellow frames on spot images highlight current sample (brain) and table (slide 3). The error bars of signal summary and QI summary are standard deviations. Significantly differentially expressed genes are highlighted with thick lines (p-value <0.05).

A small portion of spots (<0.3%, typically 10 out of 30000 features) may be

falsely assigned a QI by the initial evaluation based on signal intensity and SBR, and thus a manual correction is required (see Results section). *RealSpot* facilitates the manual correction in three ways. First, as an option, a user may choose an additional data column from upstream software, e.g. "flags" column from GPR files. If such a column is provided, *RealSpot* compares QI with this column. If a spot is marked as "bad" (SBR<2.0) by *RealSpot* but as "good" by *GenePix* (flag= -100), *RealSpot* ignores SBR criteria and reassign a quality index. Second, some good spots with a very large diameter may be found to have a QI of 5. Such "bad" spots frequently have a much larger diameter than an average spot of a whole slide. *RealSpot* marks these spots with a question marker "?" to prompt a user to manually check and correct. These spots are assigned a temporary QI (=21) to distinguish them from other spots. This QI is either replaced by a user-corrected QI or restored to 5 after re-open the file if a user ignores these spots. Third, *RealSpot* offers tools to facilitate the manual correction by sorting similar quality spots together using QI, spot diameter, signal intensity, or spot image, and correcting them at once. Some irregular and contaminated spots may be assigned a QI of 1-4. A user may select these spots, and assign a right QI to them using an annotation tool. *RealSpot* also offers another tool to identify the spots with a similar shape by comparing the images of a selected spot (e.g., a contaminated spot) and all other spots in this channel, and calculating an image similarity (*IS*) for each spot based on Pearson's correlation coefficient:

$$IS_f = \frac{(\sum FiCi - \sum Fi \sum Ci / n)^2}{(\sum Fi^2 - (\sum Fi)^2 / n)(\sum Ci^2 - (\sum Ci)^2 / n)}$$

Where, IS_f is the image similarity of a spot image *F*, which has *n* pixels

marked as *Fi* (i=1~n), to the selected spot *C*, marked as *Ci* (i=1~n). The summations are based on index i and calculated from i=1 to n. *IS* value is ranged from 1 (identical spots) to 0 (entirely different spots). *RealSpot* then sorts the spots by *IS*, so that spots with similar images are arranged together. After sorting, the selected spot moves to the first row of the top, followed by other similar spots. A user may manually check these similar spots and correct QI accordingly, since these spots has similar morphology.
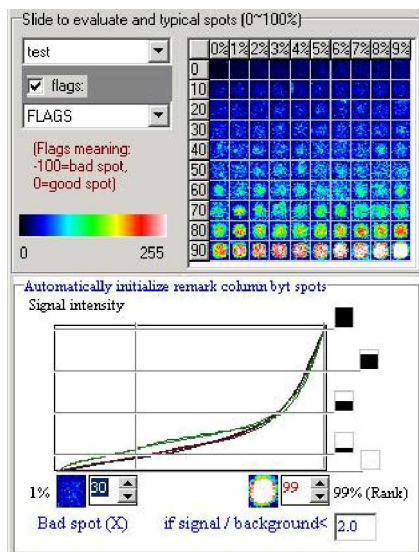
Fig. 2.3 A plot for defining quality index. The chart was plotted as signal fluorescence intensity vs gene rank percentage (1~99%). By default, quality index (QI) of 0 and 4 was assigned to the genes whose intensities were <30% and >95%, respectively. These percentages can be adjusted based on the pseudo-color spots in the top grid. QI 1 - 3 are calculated as described in the text. A QI of 5 for bad spots was determined by signal-to-background ratio (<2 by default). The check box, "flags" from upstream image quantification (e.g. GPR file from *GenePix*) is an optional column for QI evaluation.

2.3.4 Data organization

When there are more than one hybridizations or slides in one DNA microarray experiment, *RealSpot* organizes the evaluated slides as an experiment for calculating quality index summary, spot signal summary, performing one-way ANOVA, associating gene ontology information with each gene, and retrieving data to verify results. *RealSpot* uses sample information of each slide and aligns slides by sample names. A user can import multiple slide files into a meta-table at the same time by selecting the file names. A

column for summarizing quality index of each spot from multiple slides is added into the meta-table. This column is calculated as following: the contaminated or bad spots are first removed; the mean and SD of the QIs are calculated from spots with a QI of 0, 1, 2, 3 and 4; the mean QI is round to an integer and shown as an icon as shown in Table 2.1 and the SD is shown as an error bar (**Fig.** 2.2). A column of spot signal summary from multiple slides is calculated as following. The 16-bit signal intensities of each channel are scaled to an arbitrary range (1~1,000 in *RealSpot*). The data scaling serves as a global normalization, so that the gene expression data from different slides are comparable. The data scaling is based on an assumption that the lowest 5% genes are not expressed (i.e., negative spots) and converted to 1, and the highest 95% spots are highly expressed (i.e., saturated spots, typically from housekeeping genes) and converted to 1,000. The rest spot signals are linearly scaled to 1~1,000. The mean and SD of scaled spot signals for each sample group are calculated, and visualized in the column as bar plots. A one-way ANOVA is performed based on the above globally normalized signals, if 3 or more slides are used in an experiment. Prior to ANOVA, normalized signals are logarithm transformed, which can improve the homogeneity of standard deviations among sample groups (a pre-requirement of ANOVA). A p-value of each gene is obtained from ANOVA. *RealSpot* highlights the bar plots of significant genes (p-value< significance level, default=0.05) with thick lines. This indicates that there is a significant difference of gene expression for at lease two samples. *RealSpot* also accepts gene ontology association files (tab-delimited ext files with columns of GO ID, gene symbol, gene ID, GO term and GO part). If a gene ontology file is read, a column of ontology is added to

display the functions of known genes (**Fig.** 2.2). In the meta-table, some columns are the same for all the slides such as gene ID and name, printing layout information, and summary quality index. Other columns such as quality index and spot images are specific for each slide and are organized as a sub-table within the respective row of each gene. A sorting column is used for showing information for sorting, e.g. "P-value" column in **Fig.** 2.2.

2.3.5 Data verification

The data verification module directly compares DNA microarray data with spot images, providing an additional step for quality control, and more importantly, a method to validate data analysis results. After data quality evaluation, the filtered data set can be used for further data analysis such as cluster analysis or hunting differentially expressed genes. From the down-stream analysis, a researcher may obtain a list of genes and associated data, such as the genes with similar normalized ratios, differential expression, expressed pattern across a serial of conditions. Before further analysis or functional studies, a user may compare the final genes with the respective spot images by clicking "search genelist" button in **Fig.** 2.2 to search and group these genes, and optionally, the associated data. *RealSpot* shows all the found genes on the top of the meta-table. It is relatively easy for a human being to identify a few distinct spots from other spots showing a similar pattern. Consequently, search by genelist in *RealSpot* is efficient for identifying the inconsistence of data analysis results and spot images. The inconsistent genes may be eliminated from further analysis, or an alternative method may be chosen to analysis the same data set and consistent results achieved.

2.3.6 Data export

Spot images, raw data as well as quality index can be exported by selecting interested genes or items. An ExportModule guides a user exporting the respective information. For image export, the spot images of selected genes are exported as a Windows Enhanced Metafile (EMF), bitmap file (BMP), and web pages (hypertext markup language files, HTML). The EMF file is the default format. It contains the instructions for drawing the text and spot images, and has a very high resolution. It is best for printing high quality image. The file formats can be read by most image and word processing software packages. *RealSpot* also exports the summary quality index and gene expression ratio of two samples. *RealSpot* can export hundreds or thousands of genes as a HTML file. A user may directly post it on internet for data communication among internal lab members or external DNA microarray communities. Before data export, *RealSpot* provides tools for data normalization and scatter plotting.   A user may select two samples and filter spots by quality index or directly select spots from the Meta-table. The scatter plot visualizes the global distribution of the signal intensity of the selected genes in two samples. Global or intensity-dependent normalization methods are provided. LOWESS normalization (15) based on print tip is the default. A user may select an appropriate normalization method based on the scatter plot. *RealSpot* exports the gene ID and name, quality index and normalized expression ratio of select genes and samples as a tab-delimited text file.

2.4   Results

2.4.1   Environment requirement

RealSpot is an executable under Windows 98 or above operation system. The

minimal hardware configuration is 250 MHz Pentium III CPU, 64 MB memory and 200 MB free hard disk space for storing data. For experiments with multiple slides, a current main stream computer with 2.6 GHz Pentium IV CPU, 512MB memory and 2GB free hard disk space is recommended. The following performance results are based on the recommended computer.

2.4.2   Performance

It took about 10s for *RealSpot* to import raw data and two images of a slide with 30,000 features. The table file created by *RealSpot* was around 20 MB or 1/3 of the total size of the imported raw data and image files (60 – 70 MB). *RealSpot* evaluates one slide based on the intensity and SBR in about 500 ms. It takes about 5 - 10 min for a user to semi-automatically correct spot quality index. Current version *RealSpot* can manage an experiment with hundreds of slides. For loading 77 slide files (30,000 spots each) from a whole experiment, *RealSpot* only spent about 10 s because of the compact size and binary format of table files. The slowest performance of *RealSpot* was data exporting, due to the intensity-dependent LOWESS normalization. *RealSpot* spent about 10 s to normalize a table, or 3 min for a whole experiment of 20 slides.

2.4.3   Quality evaluation

During the above performance test, about 0.3% of the spots (100 of 30K) of each slide were semi-automatically corrected after initial quality evaluation based on visual and subjective observation of spot images. By sorting, these spots were grouped at the end of the table. Most of these spots were extremely big and were falsely identified as bad spots. It took 5-10 min to correct these

spots for each slide. This is a substantial time saving, compared with *GenePix*, by which we normally spent several hours on manually evaluating the location and quality of individual spots through a whole slide.

To assess the data quality after the quality evaluation, we compared the scatter plots after data filtering using different ranges of QIs. As shown in **Fig. 2.4**, most of bad spots and empty spots were in the lower intensity end. After filtering these spots (QI=0 or 5), more consistent results were obtained.

Table 2.2   False-counts of negative and positive probes from *RealSpot* and *GenePix*

Control		*GenePix*	*RealSpot*
Negative	False-positive	23.5±35.0	19.5±10.9
(169)	True-negative	145.5±37.7	146.3±84.6
Positive	False-negative	12.2±17.4	0.4±1.9
(86)	True-positive	73.9±19.3	85.5±4.0

Results were based on 169 negative control probes and 86 highly abundant genes (e.g. ribosomal proteins and GAPDH) in a 10 K rat DNA microarray. There were 24 replicated spots for each gene probe. Spot counts were average ± standard deviation. False-positive counts in negative control probes were based on a flag of ≥0 in *GenePix* and a QI of ≥ 1 in *RealSpot*. False-negative counts in positive control probes were based on a flag of <0 in *GenePix* and a QI of 0 or 5 in *RealSpot*.

Furthermore, we compared *RealSpot* with *GenePix* for quality evaluation by counting false-positive spots from negative control probes or false-negative spots from positive control probes (Table 2.2). There are 169 *Arabidopsis* probes used as negative controls in our 10K rat DNA microarray. The false-positive spots identified by *RealSpot* from these negative control probes (a QI of 1) were slightly lower than those by *GenePix* (a flag of ≥ 0). It is noteworthy that the false-positive spots from *RealSpot* had a QI of 1, which means weak

or ambiguous spots. A user may filter such weak spots in a particular experiment, and these spots might not be false-positive in such a circumstance. For positive control probes, we choose 86 highly abundant genes such as ribosomal proteins and GAPDH. There was a significantly lower false-negative spot counts in *RealSpot* than these in *GenePix* (p<5).
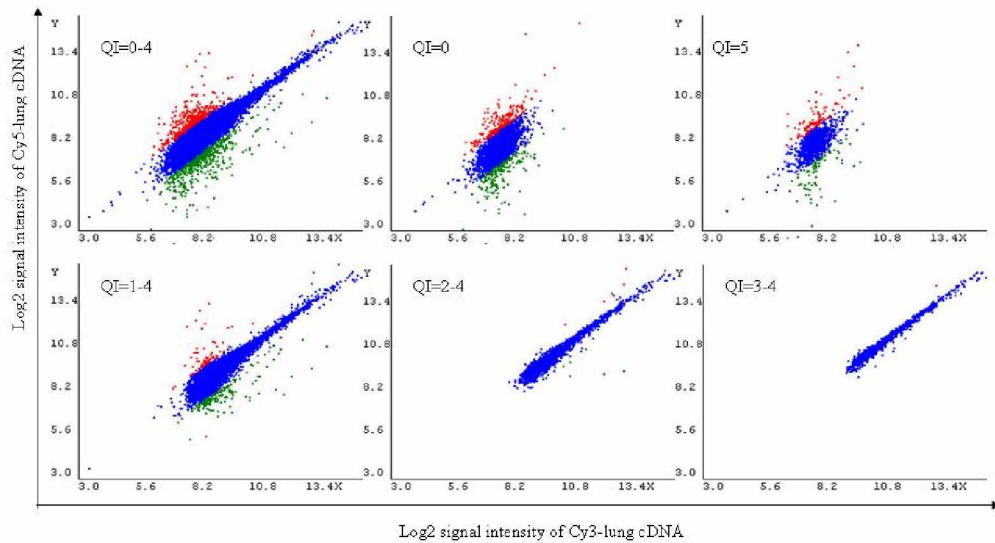
Fig. 2.4 Scatter plots after data filtering based on quality index. X and Y-axis are background subtracted signal intensity from a lung-lung self-hybridization image after quality evaluation. Data were plotted to show spots within various QI ranges.

2.4.4 Result Validation

Using *RealSpot*, we also compared spot images with the results from different stages of data analysis, such as data filtering, data normalization, differential gene expression and cluster analysis (**Fig.** 2.5). The spots shown in **Fig.** 2.5A had an identical flag of -100 for both red and green channels, generated by *GenePix*. These spots were identified as bad spots by *GenePix*. However, some of spots obviously are good spots as evaluated by spot images and QIs generated by *RealSpot* (**Fig.** 2.5A). We used two methods to

normalize data: $\log_2$ transformed ratios from background-subtracted signals (e.g., "F535 mean – B535" data column in *GenePix* raw data) were normalized by LOWESS based on print-tip (**Fig.** 2.5B), or by globally adjusting $\log_2$ ratio median to 0 (**Fig.** 2.5C). The genes shown had a 2-fold increase of expression (heart vs lung). Compared with spot images, it appears that the log ratios from global normalization was more consistent with spot images than these from LOWESS normalization in this particular example. Based on the normalized $\log_2$ ratios, we identified a set of differentially expressed genes by SAM (**Fig.** 2.5D) or ANOVA (**Fig.** 2.5E). The genes shown in **Fig.** 2.5D or 2.5E should be expressed significantly higher in lung than in heart, and were generally consistent with spot images. However, the first gene in **Fig.** 2.5 D or **Fig.** 2.5E might be considered to be eliminated from further studies, since there were two strong replicated spots in heart column (2) for the former and the gene expression level appeared to be very weak for the latter. We validated 1,641 out of 2,110 differentially expressed genes identified by SAM test in about half an hour in this way. About 22% differential genes were eliminated because of inconsistence between log ratio and spot images or very low gene expression levels (weak spots). Using K-mean cluster analysis (4), we identified 10 clusters from 6 tissue hybridizations (lung, heart, kidney, liver, spleen, and brain). One of the clusters is spleen-specific genes (**Fig.** 2.5E and 2.5F) The spot images were generally consistent with expression patterns from cluster analysis, except the second gene, which apparently was a false-positive result and should be eliminated from further study. We found 6 out of 48 genes in a whole cluster showed inconsistent trends between gene expression levels (normalized signal in arbitrary unit) and the respective spot images.
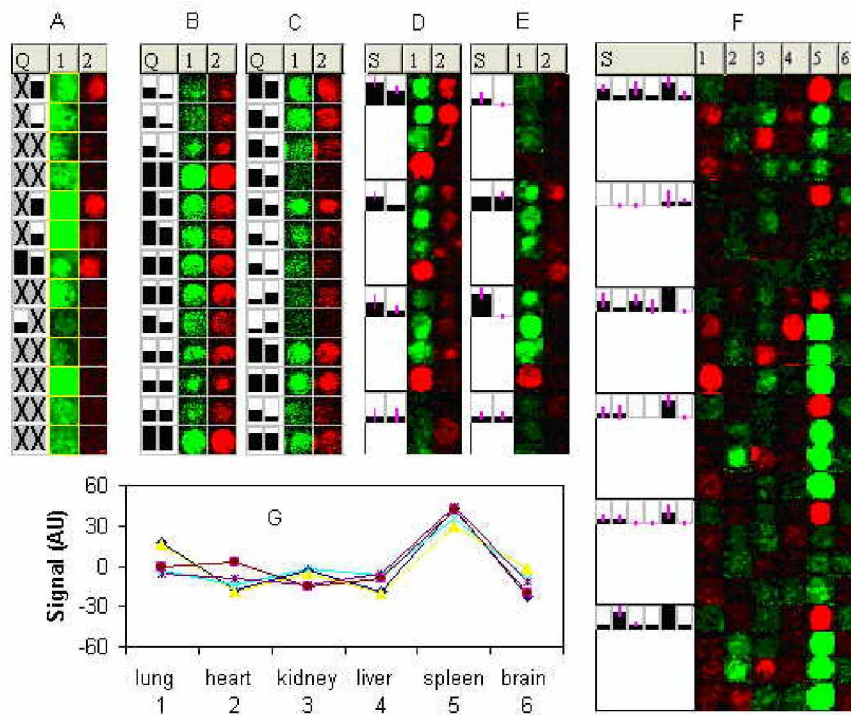
Fig. 2.5 Validation of the results at different stages of microarray data analysis. **(A)** Spots with a spot flag of -100 (bad spot) generated by *GenePix* from lung (column 1) – lung (column 2) self-hybridization. The two icons in each row of column Q were the QI of column "1" and "2" from *RealSpot*, respectively. **(B, C)** All of the spots have a normalized $\log_2$ ratio of 1.0 (lung/heart, column 1/2) by LOWESS (B) or global (C) normalization. **(D, E)** The genes with 4 replicated spots were differentially expressed in lung (column 1) vs heart (column 2) identified by SAM (q-value=0.05) (D) or ANOVA (p-value=0.05) (E). Column S indicates QI summary of the respective spots. **(F, G)** One cluster genes identified by k-mean cluster analysis (k=10) from tissue hybridizations: lung (1), heart (2), kidney (3), liver (4), spleen (5) and brain (6). Column S in Panel F represents QI summary. The signals in the y-axis of Panel G were relative gene expression levels of genes among the six organs in arbitrary unit (AU).

2.5   Discussion

We designed a software package, *RealSpot* to validate the results from DNA microarray experiments at any stages of data analysis by directly associating spot quality and results with spot images.   *RealSpot* evaluates

37

DNA microarray data quality by assigning a quality index to each spot based on signal intensity and SBR, followed by directly comparing spot images, raw data, and analysis results. The reliability of quality evaluation is dependent on both spot images and the respective raw data. The direct comparison of spot images and raw data makes the quality evaluation more reliable than using either of them separately. *RealSpot* also has several built-in functions including search, sort, data organization, one-way ANOVA, gene ontology, data normalization, plotting and web page generation. Furthermore, step-by-step Modules make *RealSpot* easy to use.

The spot images are directly imported from scanned microarray slides and linearly transformed to 0-255, representing original image information. This linear transformation trims the extreme signals, i.e., the lowest and the highest 5% of image signals, since the former usually is background noise, and the latter the saturated signal. Trimming these signals does not lose much image information, but makes 90% pixels visible without adjusting brightness and contrast. This method is similar to Affymetrix data normalization, which linearly transforms fluorescence intensity to an arbitrary range, e.g. 0 ~ 10,000. The advantage is that different slides are comparable after transformation. We noticed that some slides show differences among printing tips. We therefore added an option in *RealSpot* for a separate transformation of each sub-array or block from an identical printing tip to compensate such differences, similar to the printing tip-based LOWESS normalization (8). We generated quantitative data (signal intensity and signal-to-background ratio) by *RealSpot* from original 16-bit spot images located by geometric data from *GenePix* (x, y, and diameter) for quality evaluation and verification of the original raw data

from image analysis software packages. Signal intensity and SBR were affected by spot alignment algorithms but not spot segmentation algorithms (14). In *RealSpot*, the image of each spot was split from a whole slide at the respective spot center (x,y). The size of each spot image was identical, i.e., the average distance between two adjacent spots. The signal intensity was the average intensity of the whole transformed image of a spot, and the SBR was estimated from the center ¼ area (for signal) and the four corners (for background). The problem in term of assigning a quality index to large spots was associated with the estimation of SBR in *RealSpot*. The problematic spots usually had a diameter larger than the average distance between two adjacent spots. The four corners of these spot images were largely occupied by the spots. Consequently, the estimated SBR was lower than true SBR and thus a larger spot may be falsely assigned a QI for a contaminated spot. This problem would be even worse in the slide area where large spots were clustered together. A potential solution would be the estimation of SBR using global background through a whole slide, since this problem was resulted from the estimation of local background. Currently, we do not test global background, since local background works well with 99.7% spots for identifying low quality spots in our results.   However, in *RealSpot*, these falsely assigned QI can be manually corrected in a quick table operation style: sorting, selecting, and editing. Comparing the time-consuming manual spot correction in a whole slide as does in *GenePix* , the quick manual correction in a table as does in *RealSpot* results in a higher efficiency.

The data imported from raw data and transformed images are useful for manual evaluation. For instance, the contaminated spots move together when

the table from a slide is sorted by "flags", "SNR", or "B535 mean" column (from GPR file). These contaminated spots may be manually selected and remarked simultaneously. Sorting table also helps to correct errors, in particularly, weak, noisy or irregular spots. On the other hand, some spots may be marked as bad spots based on SBR, but they are good spots by visual assessment. Most of the reported automatic methods were based on raw data, using some criteria such as signal-to-noise ratio, SBR, circularity coefficient, and composite score (1; 6; 12). We found that these criteria were sometimes inconsistent with spot images. For instance, many weak spots have high SBRs because both their signals and backgrounds are closed to 0. Some spots with a good morphology and intensity may be contaminated with a few tiny dusts. These spots are good spots if manually marked, but may be identified as bad spots because their SNRs are low. These spots can be corrected by sorting the quality index column, followed by spot diameter column. In *RealSpot*, the quality indexes of all the spots can be corrected in 5 ~ 10 minutes for 30,000 spots, and thus the mistakes that are unavoidable for automatic tools are minimized.

The one-way ANOVA implemented in *RealSpot* simplifies multiple statistical factors (e.g. dye, slide, and sample treatment) to one factor (sample treatment). The original 16-bit gene expression data from each slide are globally scaled to an identical range (1~1000) and logarithm-transformed for calculating p-values. The p-values from this simplified one-way ANOVA can be used for fast monitoring differentially expressed genes. For instance, a gene with a lower p-value than a significance cutoff (p-value=0.05) may be considered as a significantly differentially expressed gene for further investigation. The significance cutoff may be adjusted, e.g. Bonferroni

adjustment, which sets p-value cutoff=$\frac{0.05}{n}$, (n=total gene number). The adjusted cutoff may decrease type I error (false-positive).

The gene ontology information of known genes is helpful for understanding gene functions. In a typical ontological annotation file, a gene is assigned multiple GO IDs, reflecting the molecular function, biological process, and cellular location. *RealSpot* summarizes all the ontological annotations of each gene, and thus provides a user with comprehensive information of a gene. Based on the p-value from one-way ANOVA and ontological annotation, a user can quickly find interesting genes and their potential roles in a particular experiment.

RealSpot is efficient. It evaluates and assigns a QI to each spot immediately after data import. A user also can identify incorrectly evaluated spots, by sorting similar spots together using spot images and raw data. Another feature is the icons used for quality indices. *RealSpot* uses standard scores 0-4 and represents them as bar plot-like icons to visualize spot quality (Table 2.1). It is easier for a user to compare a spot with the respective icon than with a number. It is also helpful for visualizing the trends of gene expression level when several slides or samples are grouped together.

RealSpot is flexible and easy to use. First, the images from many DNA micrarroay scanners, and raw data from commonly used image analysis software packages can be directly imported. The import module guides a user importing data step by step with detailed help information. *RealSpot* skips the description of some raw data files, e.g. *GenePix* GPR files, and imports user selected columns of raw data. By organizing raw data and spot images in a table, the user interface of *RealSpot* is similar to Microsoft™ *Excel* worksheets.

Under such operation environment, a user can focus on data evaluation without learning new instructions for operating software. The ExportModule exports the table as a text file for importing to database or data analysis software. Images can be exported in bitmap or metafile formats. These formats are most popularly supported in Windows operation systems. Organizing data by samples and slides clearly displays microarray experimental designs such as a loop or reference design, and helps a user to interpret data from biological samples.

RealSpot is designed for quality evaluation of raw data and spot images, not for data analysis, although some simple data process tools such as data normalization are included. Further improvement of *RealSpot* may include image transformation, a direct link to database, and data analysis tools. In the current version of *RealSpot*, a 16-bit image is linearly transformed to an 8-bit image for display. A square root transformation may be used to strengthen weak spots. A direct link of *RealSpot* with database will help a user manage microarray data. More powerful sort and search tools may be implemented in Meta-table. Another limitation of the current *RealSpot* version is that it can only work with images from dual color hybridization and this issue should be addressed in the future version. In summary, the software package, *RealSpot* is efficient for validating microarray results and thus helpful for improving the reliability of the whole microarray experiment. The improvement is resulted from the association of micorarray data with the respective spot images.

2.6 Availability and requirements:

2.6.1 Project name: *RealSpot*

2.6.2 Project home page:

2.6.3   Operating system(s): Windows 98 or above operation system.

2.6.4   Programming language: Delphi (version 6)

2.6.5   Other requirements: None

2.7   Acknowledgments

2.8   References

1. **Bozinov D and Rahnenfuhrer J**. Unsupervised technique for robust target separation and analysis of DNA microarray spots through adaptive pixel clustering. *Bioinformatics* 18: 747-756, 2002.

2. **Brazma A, Hingamp P, Quackenbush J, Sherlock G, Spellman P, Stoeckert C, Aach J, Ansorge W, Ball CA, Causton HC, Gaasterland T, Glenisson P, Holstege FC, Kim IF, Markowitz V, Matese JC, Parkinson H, Robinson A, Sarkans U, Schulze-Kremer S, Stewart J, Taylor R, Vilo J and Vingron M**. Minimum information about a microarray experiment (MIAME)-toward standards for microarray data. *Nat Genet* 29: 365-371, 2001.

3. **Dahlquist KD, Salomonis N, Vranizan K, Lawlor SC and Conklin BR**. GenMAPP, a new tool for viewing and analyzing microarray data on biological pathways. *Nat Genet* 31: 19-20, 2002.

4. **Eisen MB, Spellman PT, Brown PO and Botstein D**. Cluster analysis

and display of genome-wide expression patterns. *Proc Natl Acad Sci U S A* 95: 14863-14868, 1998.

5. **Forster T, Roy D and Ghazal P**. Experiments using microarray technology: limitations and standard operating procedures. *J Endocrinol* 178: 195-204, 2003.

6. **Jain AN, Tokuyasu TA, Snijders AM, Segraves R, Albertson DG and Pinkel D**. Fully automatic quantification of microarray image data. *Genome Res* 12: 325-332, 2002.

7. **Killion PJ, Sherlock G and Iyer VR**. The Longhorn Array Database (LAD): an open-source, MIAME compliant implementation of the Stanford Microarray Database (SMD). *BMC Bioinformatics* %20;4: 32, 2003.

8. **Smyth GK and Speed T**. Normalization of cDNA microarray data. *Methods* 31: 265-273, 2003.

9. **Tamames J, Clark D, Herrero J, Dopazo J, Blaschke C, Fernandez JM, Oliveros JC and Valencia A**. Bioinformatics methods for the analysis of expression arrays: data clustering and information extraction. *J Biotechnol* 98: 269-283, 2002.

10. **Tusher VG, Tibshirani R and Chu G**. Significance analysis of microarrays applied to the ionizing radiation response. *Proc Natl Acad Sci U S A* 98: 5116-5121, 2001.

11. **van Someren EP, Wessels LF, Backer E and Reinders MJ**. Genetic network modeling. *Pharmacogenomics* 3: 507-525, 2002.

12. **Wang X, Ghosh S and Guo SW**. Quantitative quality control in microarray image processing and data acquisition. *Nucleic Acids Res* 29: E75, 2001.

13. **Yang MC, Ruan QG, Yang JJ, Eckenrode S, Wu S, McIndoe RA and She JX**. A statistical method for flagging weak spots improves normalization and ratio estimates in microarrays. *Physiol Genomics* 7: 45-53, 2001.

14. **Yang YH, Buckley MJ and Speed TP**. Analysis of cDNA microarray images. *Brief Bioinform* 2: 341-349, 2001.

15. **Yang YH, Dudoit S, Luu P, Lin DM, Peng V, Ngai J and Speed TP**. Normalization for cDNA microarray data: a robust composite method addressing single and multiple slide systematic variation. *Nucleic Acids Res* 30: e15, 2002.

Chapter 3

Identification of rat lung - prominent genes by a parallel DNA microarray hybridization [5,6]

3.1 ABSTRACT

Background-The comparison of organ transcriptomes is an important strategy for understanding gene functions. In the present study, we attempted to identify lung-prominent genes by comparing the normal transcriptomes of rat lung, heart, kidney, liver, spleen, and brain. To increase the efficiency and reproducibility, we first developed a novel parallel hybridization system, in which 6 samples could be hybridized onto a single slide at the same time. **Results-**We identified the genes prominently expressed in the lung (147) or co-expressed in lung-heart (23), lung-liver (37), lung-spleen (203), and lung-kidney (98). The known functions of the lung-prominent genes mainly fell into 5 categories: ligand binding, signal transducer, cell communication, development, and metabolism. Real-time PCR confirmed 13 lung-prominent genes, including 5 genes that have not been investigated in the lung, vitamin

---

[6] Our microarray dataset has been submitted to GEO and can be accessed by following steps:
1.   Go to webpage http://www.ncbi.nlm.nih.gov/projects/geo/;
2.   Go to the bottom of this page;
3.   Login to GEO; a. user: **zhongmingchen**; b. password:**123456**; c. Click   button "Login";
4.   Go to our data set; a.   GEO accession: GPL1699; b.   Click button "GO";
5.   View dataset by click hyperlinked items of "Sample id")

D-dependent calcium binding protein (Calb3), mitogen activated protein kinase 13 (Mapk13), solute carrier family 29 transporters, member 1 (Slc29a1), corticotropin releasing hormone receptor (Crhr1), and lipocalin 2 (Lcn2).

Conclusions-The lung-prominent genes identified in this study may provide an important clue for further investigation of pulmonary functions.

3.2 BACKGROUND

With the completion of genome projects of human and other model species, functional studies on a genomic scale are coming to a frontier. The investigation of transcriptome reveals gene expression of organs and cells from normal and diseased animals and humans. By comparing transcriptomes of multiple organs, physiological functions in different organs can be further explored. Identifying the genes expressed prominently in the lung may reveal its unique physiological functions in the respiratory system.

The expression of some individual genes in the lung and other organs may be found in literature and public databases. In literature, newly discovered genes have been tested in various organs at the mRNA level with Northern blotting and RT-PCR and at the protein level with Western blotting. In public databases, gene expression are compiled from literature, cDNA library (e.g. UniGene) and high throughput tools such as serial analysis gene expression (SAGE) and DNA microarray (e.g., GEO) [1]. Several studies using DNA microarray have been reported for profiling differential gene expression among normal human and mouse organs, but very little information is available for the rat [2-6].

Dual color hybridizations are commonly used for differential expression of thousands of genes between two samples [7]. For three or more

samples, a reference or loop design has to be employed to adapt dual color hybridization [8,9]. In the reference design, several samples are hybridized onto different slides separately with a common reference, which is prepared by pooling all the samples or using genomic DNA [10]. In the loop design, samples are paired in a loop pattern for hybridization and each sample is hybridized twice. However, the efficiency and reproducibility of both designs are poor for the identification of organ-prominent genes. Only two samples are hybridized on one slide, and the hybridization on different slides is known to have high variations due to slide printing and hybridization conditions [7]. For instance, there are 15 pair-wise combinations among 6 distinct organs. Consequently, 15 co-hybridizations between samples are required for a single replication and 60 slides for an experiment with 4 biological replications.

To eliminate these problems, we developed a parallel hybridization system in which 6 samples can be hybridized onto one single slide. This technique simplifies the investigation of multiple samples, reduces experimental errors and improves experimental efficiency. Using this system, we investigated lung-prominent genes by comparing gene expression profiles among rat lung, heart, kidney, liver, spleen, and brain. The lung-prominent genes identified in the present study may provide a clue for further exploration of pulmonary functions.

3.3   MATERIAL AND METHODS

3.3.1   Microarray Preparation

The DNA microarray slides used in this study were in-house printed on epoxy-coated glass slides with 50-mer aminated oligonucleotides, Pan Rat 10K Oligonucleotide Set (MWG Biotech Inc., High Point, NC). It contains 6,221

known rat genes, 3,594 rat ESTs, and 169 Arabidopsis negative controls. The oligonucleotides were suspended in 3x SSC at 25 μM and printed on epoxy-coated slides (CEL Associates, Pearland, Texas) with an OmniGrid 100 arrayer (GeneMachine, San Carlos, CA) per manufacturer's instructions. Each oligonucleotide was spotted in triplicate on three identical 18 x 18 mm blocks: A, B, and C (Table 3.1). The total spots on one slide were 30,000 including 186 blank spots. The spot-spot distance was 180 μm and the space between blocks was 4 mm. The printed slides were incubated in 65% humidity overnight at room temperature. The slides were then dried and stored in room temperature. Prior to hybridization, the slides were washed one time with 0.2% SDS, four times with water, and dried by centrifugation. The 3 blocks on a slide were separated by two 2 x 25 x 1 mm thermostatic transparent tape stripes during the hybridization. The stripes were removed after hybridization to wash and scan slides.

Table 3.1:   Slide layout and hybridization design
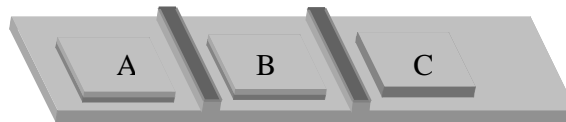
Slide	Block A Green	Block A Red	Block B Green	Block B Red	Block C Green	Block C Red
1	Lung	Heart	Liver	Brain	Kidney	Spleen
2	Heart	Kidney	Liver	Lung	Spleen	Brain
3	Kidney	Liver	Lung	Brain	Spleen	Heart
4	Heart	Brain	Kidney	Lung	Liver	Spleen
5	Heart	Liver	Lung	Spleen	Brain	Kidney

The array has three identical blocks, A, B, and C; each containing 9,984 spots representing 6,221 known rat genes, 3,594 ESTs, and 169 Arabidopsis negative controls. The three blocks are separated with thermal plastic rings.   Three paired Cy3-(green) and Alexa 647 (red)-labeled cDNA samples were hybridized onto three blocks, A, B, and C of slide 1-5. Dye and sample assignments are random for each slide. Five slides represent technical replications.

3.3.2   Sample collection and hybridization

Six organs, the lung, heart, kidney, liver, spleen, and brain of male Sprague-Dawley rats (200 g, Charles River Laboratories, Inc., Wilmington, MA) were dissected. The organs were briefly washed with deionized water and immediately homogenized in 10 ml TRI reagents (Molecular Research Center, Cincinnati, OH). Total RNA was subsequently extracted according to the manufacturer's protocol. RNA quality and quantity were assessed by spectrophotometer (NanoDrop Technologies, Inc, Rockland, DE) and agarose gel electrophoresis. Total RNA samples were aliquoted (20 µg each) for cDNA synthesis and 2-step microarray hybridization with 3DNA 50 Expression kit (Genisphere Inc., Hatfield, PA). Briefly, total RNA was reverse-transcribed with Cy3- or Alexa 647-specific primers. The cDNA products were purified with the Microcom YM-30 columns (Millipore, Billerica, MA) and mixed with 2x formamide hybridization buffer (50% formamide, 6x SSC, 0.2% SDS). The DNA microarray slides were hybridized with the cDNA samples at 42°C for 48 hours. The slides were washed and re-hybridized with Cy3- and Alexa 647-specific capture reagents at 42°C for 2 hours. In our experiments, the concentration of purified cDNA samples were normalized to 0.5 - 0.6 µg/µl before hybridization. The cDNA aliquots from 6 organs of the same rat were randomly paired and independently hybridized onto one of 3 blocks on a glass slide. Each sample was repeated 20 times: 4 biological replications and 5 technical replications. The arrangement of samples, fluorescence dyes, and blocks for one of the biological replications is shown in Table 3.1. The other 3 biological replications were similarly arranged in a style of random block design. Each hybridized slide was scanned twice by a laser confocal scanner,

ScanArray Express (PerkinElmer Life and Analytical Sciences, Boston, MA). The first scanning was used for quantification and performed with 90% laser power and 70~80% PMT so that about 5% spots were saturated. The second scanning was used for spot alignment and was carried out with 90% laser power and 95% PMT. Hybridization images were analyzed with GenePix pro 4 (Axon Instruments, Inc. Union City, CA).

3.3.3 Data analysis

Hybridization reproducibility: The reproducibility was assessed by Pearson correlation coefficients of spot signals from self-self hybridizations. The spot signals were background-subtracted fluorescence intensity extracted from hybridization images by GenePix. To estimate the variations among 6 paired organs, accumulated errors of log ratios were calculated. The log ratios between two samples were assessed from the respective spot signals, normalized by local weighted scatter plot smooth (LOWESS) based on print-tip. The accumulated error of ratios of each gene was assessed as (1)

$$e = \left(\log' \frac{s1}{s2} + \log' \frac{s2}{s3} + \log' \frac{s3}{s4} + \log' \frac{s4}{s5} + \log' \frac{s5}{s6} + \log' \frac{s6}{s1} \right)^2 \qquad (1)$$

Where e is the accumulated error, and $\log' \dfrac{si}{sj}$ are normalized log ratios between samples s1~s6, the 6 organs arranged in a loop design. The e was calculated in 2 groups, within-slide group and among-slide group. The log ratios of within-slide group were obtained from one slide with 6 samples, and those of among-slide group were from 6 slides comprised of a loop design for 6 samples.

Identification of lung-prominent Genes: To identify differentially expressed genes among 6 organs, we first globally normalized 16-bit mean

fluorescence intensity of each gene from original images using the software *RealSpot* developed in our laboratory [12] (freely available for download for academic usage, http://www.lungmicroarray.org). The global normalization converted the weakest 5% fluorescence intensities to 0 (background) and the strongest 5% fluorescence intensities to 1,000 (saturated spots, reflecting normally scanned images). The other fluorescence intensities were scaled to the range of 0 to 1,000. This transformation makes different slides and different channels comparable. It is similar to Affymetrix single channel data normalization. The transformed images and intensities were used for data quality filters, statistics tests, and direct confirmation of the data analysis results with spot images.

For spot quality evaluation, a quality index (QI) was assigned to each spot based on signal intensity and signal-to-noise ratio. QI 0-4 indicate empty, weak, middle, strong, and saturated spots, respectively. By default, QI 0 and 4 were assigned to the empty and saturated spots, whose intensities were less than 30% and greater than 95%, respectively. QI 1-3 was calculated, based on the intensity of spot signals, as: $QI_{ij} = round(\frac{I_{ij} - I_0}{I_1 - I_0} * 4)$, where QI_{ij} is the quality index of spot j on slide i and I_{ij} the intensity of the spot j on slide i. QI_{ij} is rounded to an integer. By default, I_0 is the intensity at 30th percentile, and I_1 at 95th percentile of the plot (intensity vs gene rank percentage) of the slide image. A QI of 5 was assigned to a contaminated or bad spot based on signal background ratio (SBR). By default, any spots with a SBR of <2.0 were given a QI of 5. A mean quality index was calculated from the replicated spots of a gene from multiple slides, excluding bad spots (QI=5). Data were filtered if a

mean quality index was 1.0 or less.

For the genes that passed the quality index filter, statistical tests were performed. The genes with a significantly differential expression among 6 organs for at least one organ-pair were identified by a software package, *SAM,* (Significant Analysis of Microarray, http://www-stat.stanford.edu/~tibs/SAM/) [11]. The median false discovery ratio (FDR) cutoff for a multiple class response test by SAM was set to 5%. The genes with a minimal FDR (q-value) of >5% were discarded. The genes that passed the SAM test were further classified into organ-prominent genes or co-expressed genes in two organs by pair-wise multiple comparisons with Tukey's honestly significant difference (HSD) at an overall confidence level of 95%. Organ-prominent genes were defined as the genes that were expressed significantly higher in one particular organ than in other organs (p<0.05). Similarly, co-expressed genes in two organs were defined as the genes that were expressed in the two organs than the other 4 organs.

To determine the relative specificity of a gene among organs, an organ specificity index (OSI) was defined as the correlation coefficient of gene expression levels between a gene and a putative gene. The expression levels of a putative gene were 1,000 in prominent organs and 0 in other organs. For example, the expression level of a putative gene prominent in the lung will be (from left to right are lung, heart, kidney, liver, spleen, and brain) 1,000, 0, 0, 0, 0, 0. The OSI is calculated as

$$OSI = \frac{\sum_{i=1}^{n}(Xi*Pi) - \dfrac{\sum_{i=1}^{n}Xi*\sum_{i=1}^{n}Pi}{n}}{\sqrt{\sum_{i=1}^{n}(Xi^2) - \dfrac{(\sum_{i=1}^{n}Xi)^2}{n}}\sqrt{\sum_{i=1}^{n}(Pi^2) - \dfrac{(\sum_{i=1}^{n}Pi)^2}{n}}} \qquad (2)$$

Where Xi and Pi are the mean gene expression levels of each organ of a gene and the putative gene, respectively, in organ i. N is the total number of organs (n=6 in this study). A higher correlation coefficient indicates a higher tendency of a gene for expression in a particular organ.

Finally, the gene expression data were directly compared with the respective spot images. The spot images of the genes in each sorted data set were searched and organized by *RealSpot*. The genes with visual consistence between differential gene expression and spot images were marked as highly prominent genes for the organ(s). The functional categories of these highly prominent genes were assessed based on gene ontology annotation from Rat Genome Database gene association file (RGD, http://rgd.mcw.edu) and gene ontology definitions (GO, http://www.geneontology.org).

3.3.4 Real-time PCR

Selected lung-specific genes were validated by SYBR Green I based real-time PCR (QIAGEN, Foster City, CA) as previously described [33]. Total RNA (5 µg) was reverse-transcribed into cDNA with 0.2 µg/µl dT$_{17}$, 0.3 µg/µl random hexamer primer, and MMLV reverse transcriptase (Invitrogen Inc., Carlsbad, CA). The primer pairs were as follows ("_F": forward, "_R": reverse): beta defensin-2, BD-2_F, AAT CAC ATG CCT GAC CAA AGGA; BD-2_R, GGA GCA AAT TCT GTT CAT CCCA; keratin19, K19_F, CCA GGT CGC TGT CCA

CAC TAC; K19_R, CCT TCC AGG GCA GCT TTC AT; vitamin D-dependent

calcium-binding protein, Calb3_F, CAG CAC TCA CTG ACA GCA AGCA,

Calb3_R, TCC TCC TTG GAC AGC TGG TTT; surfactant protein D, SP-D_F,

TTC TCT CCA TGC TTG TCC TGC T, SP-D_R, GAC TAG GGT GCA CGT

GTT GGT T; intercellular adhesion molecule 1, ICAM-1_F, GGA GTC TCA

TGC CCG TGA AAT, ICAM-1_R, GTG CCT ACC CTC CCA CAA CA; mitogen

activated protein kinase 13, Mapk13_F, CCC AGC AGC CAT TTG ATG AT,

Mapk13_R, CAC TGC AGC TTC ATC CCA CTT; corticotropin releasing

hormone receptor, Crhr1_F, GGT CTC CAG GGT CGT CTT CAT C, Crhr1_R,

ACG CCA CCT CTT CCG GAT AG; solute carrier family 29 transporters,

member 1, Slc29a1_F, GGA CAA TGG TCT CTG ACG GAC A; Slc29a1_R,

CCT GGA ACA GGC ACA GAA GAA A; advanced glycosylation end

product-specific receptor, Ager_F, TCC GGT GTC GGG CAA CTA, Ager_R,

GGG ACA TTG GCT GTG AGT TCAG; solute carrier family 34 sodium

phosphate, member 2, Slc34a2_F, GCC CAT AGG TGT GAG CCT TTC,

Slc34a2_R, CCC CAT TCA CTC CAT CCT AGG A; lipocalin 2, Lcn2_F, TCT

GGG CCT CAA GGA TAA CAAC, Lcn2_R, AGA CAG GTG GGA CCT GAA

CCA; matrix metalloproteinase 9, MMP9_F, TGG GCA TTA GGG ACA GAG

GAAT, MMP9_R, GGG CTG TTT CCC CTG TGA GT; nucleoporin 155kd,

Nup155_F, AAG TGG ATC AAA ACC GAG TTCG, Nup155_R, TCG CTG CTG

CAG TGA AAT TTC; discoidin domain receptor family, member 2, Ddr2_F,

AAC CAA GCA CCG ACC ATC CTT, Ddr2_R,ATG TGG CTG AGC GGT

AGG TCT T; trans-acting transcription factor 4, Sp4_F, TTG TCA CAG TTG

CCG CCA TT, Sp4_R,TGA CCA GCC CAT TTC CAG ATT T;

melanoma-associated antigen, Mg50_F, TGC CAC ATC AGT CAC CCA TGA,

Mg50_R, AGC CGA GAC TCC AGG CTG TTT A;18S rRNA_F: TCC CAG TAA

GTG CGG GTC ATA, 18S rRNA_R: CGA GGG CCT CAC TAA ACC ATC. The

real-time PCR thermal conditions for all 14 genes listed above were 95°C 15

min, followed by 40 cycles of 95°C for 30 sec, 60°C for 30 sec, 72°C for 30 sec,

and 77°C for 35 sec. To eliminate experimental variations, all genes were

amplified in the same plate, each with 6 organ cDNA samples from one rat

(totally 84 wells for organ samples, other wells for negative controls). Three

plates were used for the three biological replications. Data were analyzed

using relative real-time PCR quantification based on the delta delta Ct method

[34]. The endogenous reference gene was 18S rRNA, and the control organ

was lung. One-way ANOVA tests were performed for statistical significance

($p < 0.05$).

3.4   RESULTS

3.4.1   Reproducibility and efficiency of parallel hybridization

Our parallel hybridization system consists of three identical blocks: A,

B, and C, on a single slide (Table 3.1). Each block contains ~10,000 50-mer

oligonucleotides (6,221 known rat genes, 3,594 rat ESTs, and 169 Arabidopsis

negative controls). Six labeled cDNA samples (3 Cy3 and 3 Alexa 647) were

combined into 3 green-red pairs and hybridized onto each block of one slide.

During the hybridization step, the blocks were separated by thermostatic tapes.

The latter was removed during the washing and scanning steps. To examine

whether there was cross-contamination among blocks, blocks A and C on the

same slide were hybridized simultaneously for 3 days with Alexa 647-labeled

lung cDNA. No signals were detected in block B (data not shown), indicating

no cross contaminations among blocks.
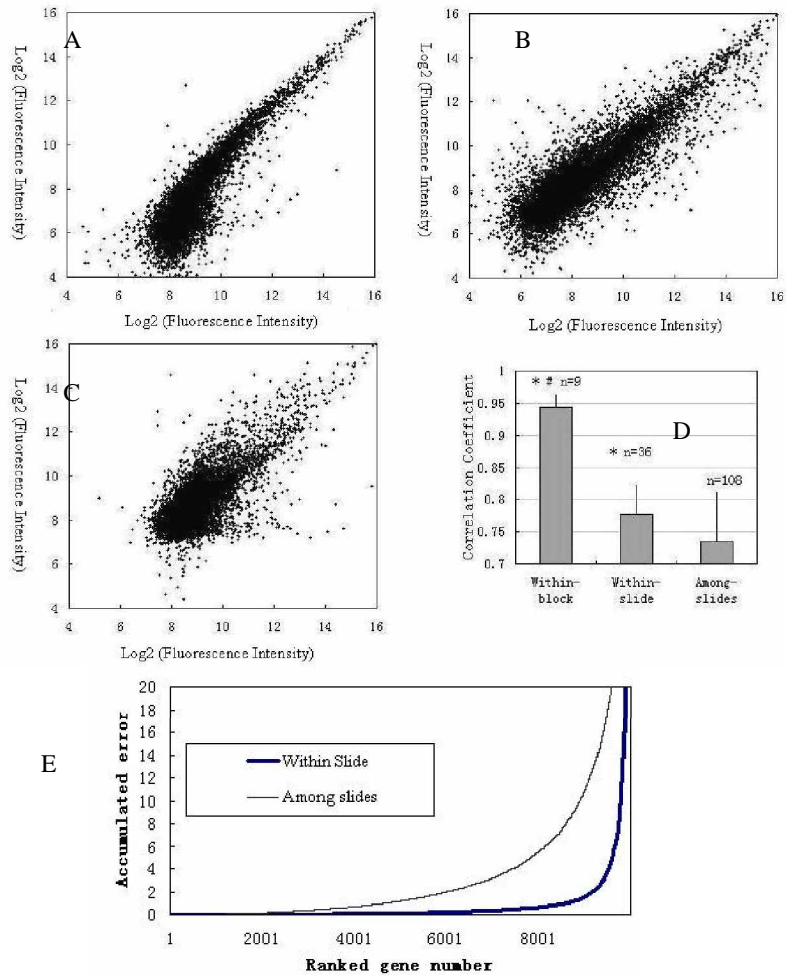
Fig. 3.1 Reproducibility of hybridizations. (A-C): typical scatter plots of self-self hybridization of lung cDNAs between two channels within a block (within-block, panel A), two different blocks in one slide (within-slide, panel B), and among slides (among-slide, panel C), respectively. The cDNAs from an identical lung tissue were labeled with Cy3 or Alexa 647, and hybridized to each block of the slides. The numbers on x- and y-axis were background-subtracted fluorescence intensities of each spot with $\log_2$ transformation. (D) A comparison of correlation coefficients from replicated hybridizations. The results were expressed as means ± SE. *P<0.01 vs among-slide; [#]P<0.01 vs within-slide. (E) Comparison of accumulated errors between within-slide and among-slide groups. For the within-slide group, the log ratios were from parallel hybridization on a single slide. For the among-slides, the log ratios were from different slides. The accumulated errors were calculated as described in *Materials and Methods*.

Self-self hybridizations were performed on three slides to assess the reproducibility of hybridizations using Cy3- and Alexa 647-labeled lung cDNA samples. We observed the highest correlation coefficient between two samples co-hybridized in one block (within-block group, **Fig.** 4.1A), and the lowest one between two samples hybridized in two different blocks on two separate slides (among-slide group, **Fig.** 4.1C). The within-slide group (two samples in two distinct blocks on one slide, **Fig.** 4.1B) possessed a significantly higher reproducibility than the among-slide group, but lower than the within-block group (**Fig.** 4.1D, $p<0.01$). The lower reproducibility of the among-slide group may be due to the experimental variations among slides, such as hybridization temperature fluctuation, washing, and scanning. These conditions were identical for the within-block and within-slide groups, in which samples were hybridized in a single slide.

Next, we investigated the relative gene expression levels in 6 rat organs: lung, heart, kidney, brain, spleen, and liver. The hybridization of each organ was repeated 20 times: 4 biological replications (rats), each with 5 technical replications (slides). Six samples from each of four rats were split into 5 aliquots for hybridization on 5 slides. The labeling dyes, the sample pairing, and the hybridization blocks on a slide were randomly assigned for each biological replication. This minimized the variations among biological and technical replications, including animals, fluorescence dyes, sample combinations, blocks on a slide, slides, and experimental conditions (Table 3.1). Statistically, each slide was a random block containing 6 samples. There were 60 sample-sample hybridizations performed on 20 slides (60 Alexa 647-cDNA and 60 Cy3-cDNAs) in this experiment. To achieve similar statistical

results, a traditional reference design requires 120 slides for co-hybridizations of sample and reference. Alternatively, in a loop design, 60 slides are required for co-hybridization of sample-sample.

The difference of fluorescence intensity between the parallel hybridization and traditional dual-color hybridization was evaluated. We first compared the difference of log ratios between the traditional and parallel hybridization systems by SAM [11]. The samples of lung and heart were used as an example. The log ratios of fluorescence intensity between lung and heart were normalized with the print-tip based LOWESS [7]. The traditional log ratios were from 4 slides, in which lung and heart were paired and co-hybridized onto the same block of each slide. The parallel log ratios were from 4 other slides, in which lung and heart were hybridized onto two different blocks of each slide. The 2-class SAM test identified no genes that showed a significant difference between the traditional co-hybridization group and the parallel hybridization group (false discovery ratio<0.047, q-value>0.05). Other organ pairs showed similar results. These results demonstrated that the log ratios of two samples from two different blocks in the parallel hybridization were not significantly different from that of the traditional two sample co-hybridization. Consequently, any two of the six samples hybridized onto one slide in the parallel hybridization can be directly compared as if these samples were pair-wise combined and co-hybridized onto one traditional slide.

We also tested the accumulated error of the $\log_2$ ratios among 6 organs. In a traditional loop design, the sum of log ratios along the loop should be zero, but frequently fluctuating. Therefore, the square sum of log ratios can be adapted to assess the accumulated error of each gene or the data

fluctuation in one experiment. We selected one block from each of the six different slides and simulated the traditional loop design. The 6 blocks formed a loop as if they were 6 traditional co-hybridization slides. In another group, a loop was formed from a single parallel hybridization slide. The slides for both groups were randomly selected. The accumulated errors were calculated as described in the *Materials and Methods*, followed by being sorted ascending, and plotted against ranked genes. We found that 21% of the genes showed an accumulated error of >5 in the traditional hybridization group, but only 4% in the parallel hybridization group (**Fig.** 4.1E). A paired t-test of the accumulated errors between the two groups revealed that the fluctuation of the traditional co-hybridization was significantly higher than that of the parallel hybridization ($p < 0.05$).

3.4.2 Prominent genes expressed in the lung

Lung-prominent genes were identified through quality filter, statistics filter, and image confirmation. Several steps of data analysis were followed (see Materials and Methods for details): (i) After hybridization, we first checked the qualities of whole hybridization images and excluded the images from poor slides (one out of 20 slides was discarded); (ii) We filtered 2,829 low quality spots based on a mean quality index of <1 as our quality filter; (iii) Statistics test using SAM analysis revealed that the expression levels of 3,576 genes were significantly different among 6 organs (false-positive ratio <5%, and median false discovery ratio <0.05); (iv) In order to identify organ-prominent or co-expressed genes, the genes passed SAM test were further analyzed by multiple comparisons using Turkey's honestly significant difference (HSD) tests at an overall confidence level of 95%. Organ-prominent genes are

defined as genes that are expressed significantly higher in one particular organ than any other organs (P<0.05). Similarly, co-expressed genes are the genes that are expressed significantly higher in two organs than any other 4 organs (P<0.05). There were some duplicated genes in single and two organ-prominent groups. The duplicated genes with a lower OSI were filtered. The duplication was due to the HSD-based multiple comparisons. For instance, endothelial cell growth factor protein precursor (VEGF, Genbank ID: NM_031836) was expressed significantly higher in the lung than other organs (p<0.05, OSI for lung=0.975). This gene was also co-expressed significantly higher in the lung and the liver than in other organs (p<0.05, OSI for lung and heart=0.778). In this case, we thus deleted this gene from the lung-liver group; (v) Finally, we further verified the genes identified above by directly comparing the results with spot images in a spreadsheet using the *RealSpot* software [12]. The visually inconsistent genes with spot images were filtered. The final genes were summarized in **Fig.** 4.2 and the hot maps of these genes were shown in **Fig.** 4.3. The liver showed the highest number of prominent genes (306 genes) and spleen the lowest (75 genes). The numbers of other organ-prominent genes were brain (218), kidney (163), lung (147), and heart (95). The lung had a high number of co-expressed genes with other organs: lung-spleen (203), lung-heart (23), lung-liver (37), lung-kidney (98), and lung-brain (10). The kidney also had a high number of co-expressed genes, kidney-liver (151) and kidney-brain (19). A list of all the organ-prominent genes is given in Additional file 1 (Supplementary Table E1, the excel format to show gene expression data) and Additional files 2 and 3 (Supplementary Table E2A and E2B, the PDF format to show spot images).

The prominent genes for one or two organs were further classified into 4 functional categories: function unclear, cellular location, molecular function, and biological process, using ontology annotations from Rat Genome Database (http://rgd.mcw.edu) and Gene Ontology (http://www.geneontology.org). The main functional categories of prominent genes were summarized in Additional file 4 (Supplementary Table E3) for one organ and in Additional file 5 (Supplementary Table E4) for two organs. The functions of the lung-prominent genes include ligand binding, signal transducer, cell communication, development, and metabolism. The cellular location was omitted since only a few genes were documented at the sub-cellular level. It is worthy to note that the functions of 60% or more genes we identified remain unclear in the present time.
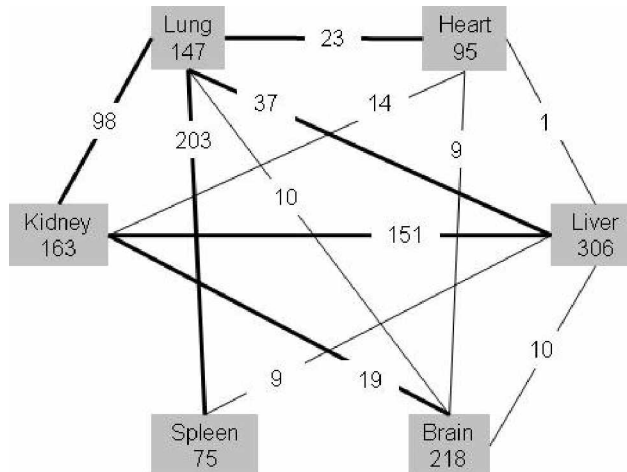
Fig. 3.2: Summary of differentially expressed genes among 6 organs. The number under an organ represents the genes that are expressed significantly higher in the respective organ compared to other organs (p<0.05). Similarly, the number between any two organs represents the genes that are expressed significantly higher in the two organs compared to other organs (p<0.05). Thicker lines highlight a larger number of the genes co-expressed in the respective two organs.

3.4.3 Real-time PCR verification

Based on our research interests, we focused on lung-prominent genes for real-time PCR verification. We selected genes based on both mRNA abundance (signal intensity) and organ specificity index (OSI). OSI was defined as the correlation coefficient of expression levels between an interested gene and a putative gene that had 100% specificity (see *Materials and Methods*). The known lung marker genes have high OSIs, e.g. T1α, 0.996; SP-A, 0.993; SP-D, 0.993; SP-B, 0.933; CCSP, 0.972; and SP-C, 0.912 (Additional file 1). We chose 13 genes, which ranked in the top 30% in signal intensity (high expression level) and the top 10% in OSI (high specificity). In addition, we selected 3 genes that ranked below 30% in signal intensity (low expression level). Real-time PCR verified 13 genes that were expressed significantly higher in the lung than in other organs (**Fig.** 4.4). These genes include BD-2, K19, Calb3, SP-D, ICAM-1, Mapk13, Crhr1, Slc29a1, Ager, Slc34a2, Lcn2, Ddr2, and Mg50. Furthermore, the expression level for most of the genes in the lung was 10 times or more as high as that in other organs. The expression pattern of these genes was consistent with DNA microarray signals (Table 2). Three genes, Nup155, MMP9 and Sp4, did not show a significantly higher mRNA abundance in the lung when compared to other organs under our experiment conditions. This is due to high variations between samples.

3.5   DISCUSSION

In the current study, we developed a parallel hybridization, in which 6 samples can be hybridized onto one single slide. This method provides higher reproducibility and efficiency than the standard co-hybridization, and should be

suitable for experiments investigating multiple biological samples. Using this system, we identified genes prominently expressed in one or two organs of the rat lung, heart, kidney, liver, spleen, and brain. Thirteen out of 16 selected lung-prominent genes were verified by real-time PCR. The genes identified in present study may be useful for further functional investigation in the lung or other organs.

The organ-prominent genes we identified were directly based on statistical comparisons of normalized spot signals. These genes were further ranked by organ specificity index (OSI). The "standard" DNA microarray data process extracts fluorescence intensities of both channels from hybridization images, and calculates and normalizes ratios for further statistical analysis. Our method is different from the "standard" analysis in several ways: (i) we linearly transformed all of the spot signals from each channel of hybridization images into a 0-1,000 scale, which made different channels and slides comparable. Unlike the ratio normalization, we retained relative expression levels in each channel. This is especially useful for multiple sample comparisons; (ii) Gene classification was based on multiple comparisons. Differentially expressed genes among the 6 organs were identified from SAM test, followed by multiple comparison using Tukey's HSD; and (iii) we ranked the genes by organ specificity index (OSI), higher OSI, more specific a gene in one or two organs. In this investigation, we selected lung-prominent genes for verification based on the combination of OSI and normalized spot intensity. We chose the genes ranked in the top 10% in OSI and the top 30% in spot intensity, which ensures both the lung-specificity and the gene expression level.
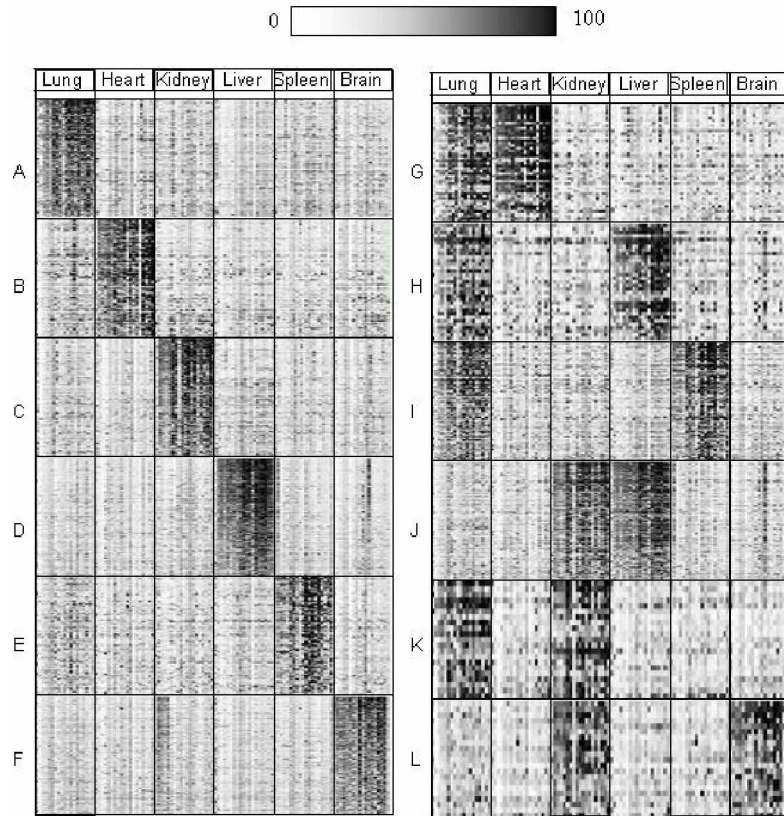
Fig. 3.3: Hot maps of Organ-prominent genes. Left and right panels are the relative expression levels of genes differentially expressed in one and two organs, respectively. Each column represents 19 replicated hybridizations of each organ and each row shows the spot signals of the organ-prominent genes. The scale of normalized spot signals was indicated on the top of the graph. (A): lung: 166 genes; (B) heart: 100 genes; (C) kidney: 186 genes; (D) liver: 324 genes; (E) spleen: 88 genes; (F) brain: 225 genes; (G) lung-heart: 47 genes; (H) lung-liver: 33 genes; (I) lung-spleen: 95 genes; (J) kidney-liver: 174 genes; (K) lung-kidney: 21 genes; (E) kidney-brain: 21 genes.
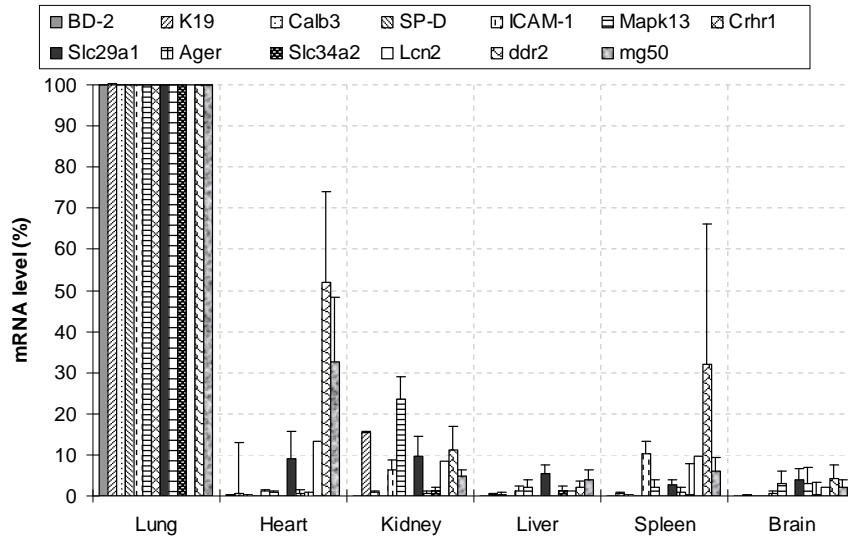
Fig. 3.4 **Relative mRNA abundance of lung-prominent genes determined by relative real-time PCR** The mRNAs from six organs were reverse-transcribed to cDNA and quantified by relative real-time PCR. All of the genes were run on the same plate with 18S rRNA as an endogenous reference. The results were expressed as % of lung. Data shown are means ± S.E. (n=3 biological replications).The mRNA expression level of all the genes in the lung was significantly higher in other organs (P<0.05).

Recently, several studies have compared gene expression profiles in human and mouse [2-6]. Only one report was done on rats with a focus on the brain using commercial Affymetrix chips (7,000 known genes and 1,000 ESTs) [6]. In this data set, the lung and liver were not included and only two replications for the spleen, heart, and kidney were used. In comparison, 2,426 genes out of the 3,576 differential genes (current study, without image-filter) were found to be common with the Walker's study [6]. The correlation coefficient of relative expression between the two data sets was around 0.4 for heart, kidney, or spleen. The low quantitative correlation may be due to the differences between Affymetrix and our in-house microarray platforms such as glass slide/silicon wafer, two/one channel, and 50-mer oligonucleotide/25-mer oligonucleotide set. However, the two data sets showed a consistent gene

expression pattern among heart, kidney, and spleen, when we manually compared differential expression of the genes with top OSI for each pattern. We also compared our dataset with other published datasets from other species that list lung-specific genes. The dataset (http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE2361 and http://www.genome.rcast.u-tokyo.ac.jp/normal/) listed 43 human lung-specific genes. Many known lung-specific genes such as T1α, caveolin and CCSP were not on this list. Among 43 genes on the list, 18 genes were found in our list of lung-prominent genes, including known lung-specific genes, surfactant proteins, ager and a verified gene, Slc34a2. The published dataset was based on a single hybridization of normal lung tissue. Our lung-prominent genes were based on 20 replicated DNA microarray hybridizations (4 biological and 5 technical replications). We believe that our gene lists were statistically confident and had a lower false-positive or false-negative genes.

The 13 lung-prominent genes we verified by real-time PCR have various functions, including pulmonary defenses, ion/solute transport, hormone receptor, differentiation, oxidant response and tumorgenesis. Five of them are defense genes. BD-2 (β-defensin 2) is a cationic peptide with a broad-spectrum antimicrobial activity and contributes to innate immunity in the lung [13]. It is expressed in the airway epithelia [14]. BD-2 was increased in the patients with inflammation and infections [15,16]. SP-D (surfactant protein D) is highly expressed in alveolar epithelial type II cells and plays a pivotal role in cell defense against microbes [17,18]. For instance, it has been reported that SP-D inhibited the proliferation of bacteria by increasing the permeability of the microbial cell membrane. ICAM-1 (intercellular adhesion molecule 1) is a cell

adhesion molecule and a ligand for leukocyte adhesion molecule LFA-1. ICAM-1 also participates in the inflammatory response to lipopolysaccharide-induced lung injury by interacting mainly with neutrophils [19]. Lipcocalin 2 (also known as α_{2u}-globulin-related protein, X13295) is a member of lipocalin protein family composed of small secreted proteins that have the ability to bind to small hydrophobic ligands [20]. Lipocalin 2 expression in the lung is markedly increased in acute lung injury caused by diesel exhaust particles and lipopolysaccharide [21]. Mapk13 (mitogen activated protein kinase 13) plays a role in stress and inflammatory responses via the MAPK cascade signaling pathway. Mapk13 is predominantly expressed in the lung although a small amount of Mapk 13 is also present in kidney [22], which is consistent with our results (**Fig.** 3.4).

Three of the identified lung prominent genes are ion/solute transporters. Calb3 (Calbindin 3), a vitamin D-dependent Ca^{2+} binding protein, was previously studied in the intestine, uterus, placenta, and lung epithelium [23]. It is a Ca^{2+} transporter and regulates Ca^{2+} homeostasis. Slc29a1 (solute carrier family 29 transporter, member 1) is an equilibrative nitrobenzylthioinosine-sensitive nucleoside transporter (ENT1), which transports nucleosides into or out of the cells in a Na^+-independent manner [24]. Northern blot analysis has shown that Slc29a1 is highly expressed in the lung and testes [25]. It plays a role in nucleotide biosynthesis and cellular signaling. Slc34a2 (solute carrier family 34 sodium phosphate, member 2) is a sodium dependent phosphate transporter. It has been shown that Slc34a2 was predominantly expressed in the lung and in situ hybridization revealed that it is localized in alveolar type II cells [26]. Slc34a2 provides inorganic phosphate for

the synthesis of lung surfactant.

Table 3.2   DNA microarray signal intensities and spot images of 13 verified genes*

GeneID	GeneName	Signals	Brain	Heart	Kidney	Liver	Lung	Spleen	OSI
NM_012967	Intercellular adhesion molecule 1 (ICAM-1)								1.00
NM_019231	Mitogen activated protein kinase 13 (Mapk13)								1.00
NM_031764	Discoidin domin receptor family, member 2 (Ddr2)								1.00
NM_012521	Vitamin D-dependent calcium-binding protein (Calb3)								1.00
AF068861	Beta defensin-2 (BD-2)								0.99
NM_031684	Solute carrier family 29 transporters, member 1 (Slc29a1)								0.99
X13295	Lipocalin 2 (alpha-2u protein, Lcn2)								0.99
NM_053380	Solute carrier family 34 sodium phosphate, member 2 (Slc34a2)								0.99
AF346790	Melanoma-associated antigen (Mg50)								0.99
AF089866	Keratin 19 (K19)								0.99
NM_012878	Pulmonary surfactant protein D (SP-D)								0.99
NM_030999	Corticotropin releasing hormone receptor (Crhr1)								0.98
NM_053336	Advanced glycosylation end product-specific receptor (Ager)								0.98

* Gene ID: Genbank accession number; GeneName: name and common symbol of a gene; signals: mean scaled fluorescence intensity of brain, heart, kidney, lung, and spleen; the error bars were based on standard deviation. OSI: Organ specific index (see text for detail)

Crhr1 (corticotropin releasing hormone receptor 1) is a receptor that binds corticotropin-releasing hormone. The mice null for the crhr1 gene died within 48 hours after birth because of a pronounced lung dysplasia [27]. Interestingly, variation of Crhr1 was associated with improved function in the asthma patients who were treated with inhaled corticosteroids [28]. K19 (keratin 19) is expressed in epithelial cells, involved in testicular differentiation and lung cancer [29,30]. Ager (advanced glycosylation end product-specific receptor) is a member of the immunoglobin superfamily and is involved in

oxidant response. It is specifically expressed in alveolar epithelial type I cells [31]. Lung type I cells are squamous, covering >90% of alveolar surface, and, thus, are easily damaged by oxidants. Ager may protect lung type I cells from oxidative injury.

Table 3.3   Gene functions in the lung and 2nd organ

Gene	Function in lung	2nd organ (location)	Function in 2nd organ
Ager	Oxidant response	(AEC I)*	
ICAM-1	AEC-leukocyte adhesion	(AEC)	
K19	Cell differentiation	(AEC)	
SP-D, BD-2	Defense, surfactant	(AEC II)	
slc42a2	*Surfactant synthesis?*	(AEC II)	
Calb3	Ca^{2+} homeostasis		
Mapk13	Inflammatory response		
Slc29a1	*Ion transporter?*		
Lcn2	*Apoptosis?*		
Crhr1	*Hormone receptor?*		
Ddr2	Collagen remodeling		
Mg50	*Tumor pathogenesis?*		
Tnni2, tni3	Lung veins [35]	Heart	Muscle contract [36,37]
Cox6a2, Cox8h	*Energy supply?*	Heart	Muscle energy supply [38] [39]
Anp	*Asthma?* [40]	Heart	Proliferation control [41]
Aqp5	*Edema?*	Liver	Fluid homeostasis [42]
Ces3,gpt	Injury and repair [43] [44]	Liver	Injury [45]
Cyp2615	Oxidantive stress [46]	Liver	*Xenobiotic metabolism?*
Cldn3	Epithelia barrier [47]	Liver	Paracellular permeability [48]
S100a18	Cell migration [49,50]	Spleen	*Cell motility?*
Iga, Igm [51] [52]	Immune response	Spleen	Immune response
Cd37 [53]	*Proliferation?*	Spleen	*Proliferation?*
Cathepsin Y	*Surfactant process?* [54] [55]	Spleen	Endopeptidase [56,57]
Fas, Alp	AEC II injury [58]	Kidney	Renal injury [59] [60]
Tpa66	Inflammatory[61]	Kidney	Anti-arterial thrombosis[62]
Nf2 [63]	Tumor supression	Kidney	Tumor suppression

* AEC I and II: Alveolar epithelial type I and II cells. See the main text for more references.

? Hypothesized function

The two lung-prominent genes with lower mRNA abundance, Ddr2 and Mg50, may be involved in human tumorgenesis (http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=pubmed&do

pt=Abstract&list_uids=15111304&query_hl=1), and the regulation of collagen remodeling in the lung (http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=pubmed&dopt=Abstract&list_uids=11103812&query_hl=11).

The functions of 13 verified genes as well as some highly abundant co-expressed genes in the lung and another organ were summarized in Table 3.3. These co-expressed genes were previously studied in the lung or another organ. The most prominent genes expressed in the lung were relevant to pulmonary protection, including oxidant response, injury and repair, inflammatory, cell defense, and immune response. These genes also contribute to organ construction such as lung veins, energy supply, and epithelial tight junction. Some of these genes may be important for cell proliferation, such as anp and nf2. Two genes, anp and aqp5 may play a role in asthma and edema, respectively. The function of cd37 is currently unclear in any of the organs. Its prominent and specific expression in the lung may imply its important role for lung function. Cd37 may participate in cell proliferation in the lung based on the studies from other members of this gene family. Similarly, cathepsin Y may play a role of surfactant protein processing or apoptosis considering its endopeptidase activity in the spleen and the functions of cathepsin D and H in the lung. These hypothesized functions may serve as a starting point for further functional studies in the respective organs.

The parallel hybridization system has several advantages over the traditional two-color hybridization. First, in this hybridization system 6 paired and dual-color labeled samples were hybridized onto one slide and scanned under identical conditions. The homogenous conditions on one slide improved

the reproducibility and decreased the variation, especially accumulated experimental errors. The latter is problematic in microarray experiments involving a series of samples such as a time course study. Second, any two of the six samples in a parallel hybridization can be directly compared; whereas only two paired samples can be directly compared in the traditional two-color hybridization of a reference or loop design. This increases the experimental efficiency and reduces the number of slides and the amount of RNAs in a whole experiment. In the parallel hybridization system, only one slide is needed for six samples.   In contrast, 6 slides are required for a reference or loop design of six samples in the traditional two-color co-hybridization. The RNA amount is reduced to half that of the traditional hybridization. This is because each sample needs to hybridize twice with neighboring samples in the loop design or hybridize to a common reference consisting of all the samples in the reference design. Multiple-color hybridization on one slide could be developed for three or more samples labeled with distinct fluorescence dyes. However, the potential cross-talk among fluorescence dyes and the need for multiple lasers of a scanner limit its application.

The organ-prominent genes in the current study were identified from 6 organs. Some of them may be expressed higher in other tissues outside the 6 organs we monitored. This limitation may be overcome by further improvement of the parallel hybridization system. One possibility is to include one common control organ (e.g. lung) in all of the parallel hybridization slides. Although it reduces the efficiency, the transcriptomes of more than 6 organs can be directly compared. Another possibility is the potential technical improvement of spot printing and sample arrangement, which may result in more than 6

samples on one parallel slide. In the present study, we printed 10K rat genes in triplicate on three blocks on one slide. Each block contains 16 sub-arrays (4.5 x 4.5 mm) consisting of 625 genes. Therefore, 6 samples can be hybridized to 10K genes in this system. If we print 625 genes onto 48 sub-arrays in replicate, 96 dual-color labeled samples can be hybridized on one slide. Furthermore, if we increase the printing resolution from 160 to 80 microns, we can print 2,500 spots on one sub-array. Consequently, 96 samples can be hybridized to one slide containing 2,500 genes. Another improvement may be the separation of the slide regions. We used thermostatic tapes to divide 3 blocks, which may not be appropriate for more samples. The chambered coverslips of 24 or 48 wells such as CultureWell™ coverslip system or array of arrays glass wafer [32] may be adapted for this purpose.

3.6   ACKNOWLEDGMENTS

3.7   References

1. Wheeler DL, Church DM, Edgar R, Federhen S, Helmberg W, Madden TL *et al.*: **Database resources of the National Center for Biotechnology Information: update.** *Nucleic Acids Res* 2004, **32 Database issue:**

D35-D40.

2. Sandberg R, Yasuda R, Pankratz DG, Carter TA, Del Rio JA, Wodicka L *et al.*: **Regional and strain-specific gene expression mapping in the adult mouse brain.** *Proc Natl Acad Sci U S A* 2000, **97:** 11038-11043.

3. Miki R, Kadota K, Bono H, Mizuno Y, Tomaru Y, Carninci P *et al.*: **Delineating developmental and metabolic pathways *in vivo* by expression profiling using the RIKEN set of 18,816 full-length enriched mouse cDNA arrays.** *Proc Natl Acad Sci U S A* 2001, **98:** 2199-2204.

4. Warrington JA, Nair A, Mahadevappa M, Tsyganskaya M: **Comparison of human adult and fetal expression and identification of 535 housekeeping/maintenance genes.** *Physiol Genomics* 2000, **2:** 143-147.

5. Su AI, Wiltshire T, Batalov S, Lapp H, Ching KA, Block D *et al.*: **A gene atlas of the mouse and human protein-encoding transcriptomes.** *Proc Natl Acad Sci U S A* 2004, **101:** 6062-6067.

6. Walker JR, Su AI, Self DW, Hogenesch JB, Lapp H, Maier R *et al.*: **Applications of a rat multiple tissue gene expression data set.** *Genome Res* 2004, **14:** 742-749.

7. Quackenbush J: **Microarray data normalization and transformation.** *Nat Genet* 2002, **32:** 496-501.

8. Kerr MK, Churchill GA: **Statistical design and the analysis of gene**

expression microarray data. *Genet Res* 2001, **77:** 123-128.

9. Pan W: **A comparative review of statistical methods for discovering differentially expressed genes in replicated microarray experiments.** *Bioinformatics* 2002, **18:** 546-554.

10. Konig R, Baldessari D, Pollet N, Niehrs C, Eils R: **Reliability of gene expression ratios for cDNA microarrays in multiconditional experiments with a reference design.** *Nucleic Acids Res* 2004, **32:** e29.

11. Tusher VG, Tibshirani R, Chu G: **Significance analysis of microarrays applied to the ionizing radiation response.** *Proc Natl Acad Sci U S A* 2001, **98:** 5116-5121.

12. Chen Z, Liu L: **RealSpot: Software validating results from DNA microarray data analysis with spot images.** *Physiol Genomics* 2005, **21:** 284-291.

13. Jia HP, Mills JN, Barahmand-Pour F, Nishimura D, Mallampali RK, Wang G *et al.*: **Molecular cloning and characterization of rat genes encoding homologues of human beta-defensins.** *Infect Immun* 1999, **67:** 4827-4833.

14. Wang X, Zhang Z, Louboutin JP, Moser C, Weiner DJ, Wilson JM: **Airway epithelia regulate expression of human beta-defensin 2 through Toll-like receptor 2.** *FASEB J* 2003, **17:** 1727-1729.

15. Schaller-Bals S, Schulze A, Bals R: **Increased levels of antimicrobial peptides in tracheal aspirates of newborn infants during infection.**

Am J Respir Crit Care Med 2002, **165:** 992-995.

16. Ross DJ, Cole AM, Yoshioka D, Park AK, Belperio JA, Laks H *et al.*: **Increased bronchoalveolar lavage human beta-defensin type 2 in bronchiolitis obliterans syndrome after lung transplantation.** *Transplantation* 2004, **78:** 1222-1224.

17. Crouch EC: **Collectins and pulmonary host defense.** *Am J Respir Cell Mol Biol* 1998, **19:** 177-201.

18. Wright JR: **Immunomodulatory functions of surfactant.** *Physiol Rev* 1997, **77:** 931-962.

19. Beck-Schimmer B, Madjdpour C, Kneller S, Ziegler U, Pasch T, Wuthrich RP *et al.*: **Role of alveolar epithelial ICAM-1 in lipopolysaccharide-induced lung inflammation.** *Eur Respir J* 2002, **19:** 1142-1150.

20. Flower DR: **The lipocalin protein family: structure and function.** *Biochem J* 1996, **318:** 1-14.

21. Yanagisawa R, Takano H, Inoue K, Ichinose T, Yoshida S, Sadakane K *et al.*: **Complementary DNA microarray analysis in acute lung injury induced by lipopolysaccharide and diesel exhaust particles.** *Exp Biol Med (Maywood)* 2004, **229:** 1081-1087.

22. Hu MC, Wang YP, Mikhail A, Qiu WR, Tan TH: **Murine p38-delta mitogen-activated protein kinase, a developmentally regulated protein kinase that is activated by stress and proinflammatory**

cytokines. *J Biol Chem* 1999, **274:** 7095-7102.

23. Dupret JM, L'Horset F, Perret C, Bernaudin JF, Thomasset M: **Calbindin-D9K gene expression in the lung of the rat. Absence of regulation by 1,25-dihydroxyvitamin D3 and estrogen.** *Endocrinology* 1992, **131:** 2643-2648.

24. Yao SY, Ng AM, Muzyka WR, Griffiths M, Cass CE, Baldwin SA *et al.*: **Molecular cloning and functional characterization of nitrobenzylthioinosine (NBMPR)-sensitive (es) and NBMPR-insensitive (ei) equilibrative nucleoside transporter proteins (rENT1 and rENT2) from rat tissues.** *J Biol Chem* 1997, **272:** 28423-28430.

25. Lu H, Chen C, Klaassen C: **Tissue distribution of concentrative and equilibrative nucleoside transporters in male and female rats and mice.** *Drug Metab Dispos* 2004, **32:** 1455-1461.

26. Hashimoto M, Wang DY, Kamo T, Zhu Y, Tsujiuchi T, Konishi Y *et al.*: **Isolation and localization of type IIb Na/Pi cotransporter in the developing rat lung.** *Am J Pathol* 2000, **157:** 21-27.

27. Smith GW, Aubry JM, Dellu F, Contarino A, Bilezikjian LM, Gold LH *et al.*: **Corticotropin releasing factor receptor 1-deficient mice display decreased anxiety, impaired stress response, and aberrant neuroendocrine development.** *Neuron* 1998, **20:** 1093-1102.

28. Tantisira KG, Lake S, Silverman ES, Palmer LJ, Lazarus R, Silverman EK *et al.*: **Corticosteroid pharmacogenetics: association of sequence**

variants in CRHR1 with improved lung function in asthmatics treated with inhaled corticosteroids.** *Hum Mol Genet* 2004, **13:** 1353-1359.

29. Bodenmuller H, Ofenloch-Hahnle B, Lane EB, Dessauer A, Bottger V, Donie F: **Lung cancer-associated keratin 19 fragments: development and biochemical characterisation of the new serum assay Enzymun-Test CYFRA 21-1.** *Int J Biol Markers* 1994, **9:** 75-81.

30. Appert A, Fridmacher V, Locquet O, Magre S: **Patterns of keratins 8, 18 and 19 during gonadal differentiation in the mouse: sex- and time-dependent expression of keratin 19.** *Differentiation* 1998, **63:** 273-284.

31. Shirasawa M, Fujiwara N, Hirabayashi S, Ohno H, Iida J, Makita K *et al.*: **Receptor for advanced glycation end-products is a marker of type I lung alveolar cells.** *Genes Cells* 2004, **9:** 165-174.

32. Zarrinkar PP, Mainquist JK, Zamora M, Stern D, Welsh JB, Sapinoso LM *et al.*: **Arrays of arrays for high-throughput gene expression profiling.** *Genome Res* 2001, **11:** 1256-1261.

33. Chen Z, Jin N, Narasaraju T, Chen J, McFarland LR, Scott M *et al.*: **Identification of two novel markers for alveolar epithelial type I and II cells.** *Biochem Biophys Res Commun* 2004, **319:** 774-780.

34. Pfaffl MW: **A new mathematical model for relative quantification in real-time RT-PCR.** *Nucleic Acids Res* 2001, **29:** e45.

35. Millino C, Sarinella F, Tiveron C, Villa A, Sartore S, Ausoni S: **Cardiac and smooth muscle cell contribution to the formation of the murine pulmonary veins.** *Dev Dyn* 2000, **218:** 414-425.

36. Murphy AM, Thompson WR, Peng LF, Jones L: **Regulation of the rat cardiac troponin I gene by the transcription factor GATA-4.** *Biochem J* 1997, **322:** 393-401.

37. Martin AF, Orlowski J: **Molecular cloning and developmental expression of the rat cardiac-specific isoform of troponin I.** *J Mol Cell Cardiol* 1991, **23:** 583-588.

38. Mell OC, Seibel P, Kadenbach B: **Structural organisation of the rat genes encoding liver- and heart-type of cytochrome c oxidase subunit VIa and a pseudogene related to the COXVIa-L cDNA.** *Gene* 1994, **140:** 179-186.

39. Lenka N, Basu A, Mullick J, Avadhani NG: **The role of an E box binding basic helix loop helix protein in the cardiac muscle-specific expression of the rat cytochrome oxidase subunit VIII gene.** *J Biol Chem* 1996, **271:** 30281-30289.

40. Mohapatra SS, Lockey RF, Vesely DL, Gower WR, Jr.: **Natriuretic peptides and genesis of asthma: an emerging paradigm?** *J Allergy Clin Immunol* 2004, **114:** 520-526.

41. Cayli S, Ustunel I, Celik-Ozenci C, Korgun ET, Demir R: **Distribution patterns of PCNA and ANP in perinatal stages of the developing rat heart.** *Acta Histochem* 2002, **104:** 271-277.

42. Matsuzaki T, Tajika Y, Suzuki T, Aoki T, Hagiwara H, Takata K: **Immunolocalization of the water channel, aquaporin-5 (AQP5), in the rat digestive system.** *Arch Histol Cytol* 2003, **66:** 307-315.

43. Wallace TJ, Ghosh S, McLean GW: **Molecular cloning and expression of rat lung carboxylesterase and its potential role in the detoxification of organophosphorus compounds.** *Am J Respir Cell Mol Biol* 1999, **20:** 1201-1208.

44. Lian X, Yan C, Yang L, Xu Y, Du H: **Lysosomal acid lipase deficiency causes respiratory inflammation and destruction in the lung.** *Am J Physiol Lung Cell Mol Physiol* 2004, **286:** L801-L807.

45. Ghosh S, Mallonee DH, Hylemon PB, Grogan WM: **Molecular cloning and expression of rat hepatic neutral cholesteryl ester hydrolase.** *Biochim Biophys Acta* 1995, **1259:** 305-312.

46. Lacroix D, Desrochers M, Lambert M, Anderson A: **Alternative splicing of mRNA encoding rat liver cytochrome P450e (P450IIB2).** *Gene* 1990, **86:** 201-207.

47. Daugherty BL, Mateescu M, Patel AS, Wade K, Kimura S, Gonzales LW *et al.*: **Developmental regulation of claudin localization by fetal alveolar epithelial cells.** *Am J Physiol Lung Cell Mol Physiol* 2004, **287:** L1266-L1273.

48. Rahner C, Mitic LL, Anderson JM: **Heterogeneity in expression and subcellular localization of claudins 2, 3, 4, and 5 in the rat liver, pancreas, and gut.** *Gastroenterology* 2001, **120:** 411-422.

49. Harrison CA, Raftery MJ, Walsh J, Alewood P, Iismaa SE, Thliveris S *et al.*: **Oxidation regulates the inflammatory properties of the murine S100 protein S100A8.** *J Biol Chem* 1999, **274:** 8561-8569.

50. Buhling F, Ittenson A, Kaiser D, Tholert G, Hoffmann B, Reinhold D *et al.*: **MRP8/MRP14, CD11b and HLA-DR expression of alveolar macrophages in pneumonia.** *Immunol Lett* 2000, **71:** 185-190.

51. Bruggemann M, Free J, Diamond A, Howard J, Cobbold S, Waldmann H: **Immunoglobulin heavy chain locus of the rat: striking homology to mouse antibody genes.** *Proc Natl Acad Sci U S A* 1986, **83:** 6075-6079.

52. Petitprez K, Grzych JM, Pierrot C, Godin C, Capron A, Khalife J: **Molecular cloning and expression of an anti-idiotype antibody mimicking a protective oligosaccharide of the parasite Schistosoma mansoni.** *Parasitol Res* 1998, **84:** 38-40.

53. van Spriel AB, Puls KL, Sofi M, Pouniotis D, Hochrein H, Orinska Z *et al.*: **A regulatory role for CD37 in T cell proliferation.** *J Immunol* 2004, **172:** 2953-2961.

54. Li X, Rayford H, Shu R, Zhuang J, Uhal BD: **Essential role for cathepsin D in bleomycin-induced apoptosis of alveolar epithelial cells.** *Am J Physiol Lung Cell Mol Physiol* 2004, **287:** L46-L51.

55. Ueno T, Linder S, Na CL, Rice WR, Johansson J, Weaver TE: **Processing of pulmonary surfactant protein B by napsin and cathepsin H.** *J Biol Chem* 2004, **279:** 16178-16184.

56. Nakazono E, Kamata Y, Yamafuji K: **Determination of the mRNA sequence of cathepsin Y, a cysteine endopeptidase from rat spleen, 1 and confirmation of its ubiquitous expression.** *Biol Chem* 2002, **383:** 1971-1975.

57. Sakamoto E, Sakao Y, Taniguchi Y, Yamafuji K: **Cathepsin Y (a novel thiol enzyme) produces kinin potentiating peptide from the component protein of rat plasma.** *Immunopharmacology* 1999, **45:** 207-214.

58. Kuwano K, Hagimoto N, Nakanishi Y: **The role of apoptosis in pulmonary fibrosis.** *Histol Histopathol* 2004, **19:** 867-881.

59. French LE, Tschopp J: **Constitutive Fas ligand expression in several non-lymphoid mouse tissues: implications for immune-protection and cell turnover.** *Behring Inst Mitt* 1996, **97:**156-160.

60. Crowell JA, Korytko PJ, Morrissey RL, Booth TD, Levine BS: **Resveratrol Associated Renal Toxicity.** *Toxicol Sci* 2004, **82:** 614-619.

61. Nyberg A, Fagerberg A, Ahlqvist M, Jern C, Seeman-Lodding H, Aneman A: **Pulmonary net release of tissue-type plasminogen activator during porcine primary and secondary acute lung injury.** *Acta Anaesthesiol Scand* 2004, **48:** 845-850.

62. Hrafnkelsdottir T, Ottosson P, Gudnason T, Samuelsson O, Jern S: **Impaired endothelial release of tissue-type plasminogen activator in patients with chronic kidney disease and hypertension.** *Hypertension* 2004, **44:** 300-304.

63. Sun CX, Robb VA, Gutmann DH: **Protein 4.1 tumor suppressors: getting a FERM grip on growth regulation.** *J Cell Sci* 2002, **115:** 3991-4000.

64. The following additional files are available upon request from Dr. Lin Liu (liulin@okstate.edu): Additional file 1: Supplementary Table E1, Organ-prominent genes in Excel format. Additional file 2: Supplementary Table E2A, One organ-prominent genes in PDF format with spot images. Additional file 3: Supplementary Table E2B, Two organ-prominent genes in PDF format with spot images. Additional file 4: Supplementary Table E3, Main functional categories of one organ-prominent genes. Additional file 5: Supplementary Table E4, Main functional categories of two organ-prominent genes

Chapter 4

Alveolar Type I Cells Protect Rat Lung Epithelium from Oxidative Injury[7,8]

4.1   Abstract

The lung alveolar surface is covered by two morphologically and functionally distinct cells: alveolar epithelial type I and type II cells (AEC I and II). The functions of AEC II, including surfactant release, cell differentiation, and ion transport, have been extensively studied. However, relatively little is known regarding the physiological functions of AEC I. Using DNA microarray based gene expression profiling, we detected the expression of many type I cell-specific genes with a diversity of functions, including cell defense from freshly isolated AEC I and II. Nine out of 10 selected genes that are differentially expressed in AEC I were verified by quantitative real-time PCR. Two genes, apolipoprotein E (Apo E) and transferrin were further characterized and functionally studied. Immunohistochemistry indicated that both proteins were specifically localized in AEC I. Up-regulations of Apo E and transferrin were observed in hyperoxic lungs. Functionally, Apo E and

---

[7] Contributions of co-authors: Jiwang Chen and Zhongming Chen contributed equally to this chapter. Jiwang Chen prepared AEC I and AEC II, and performed immunohistochemistry; Nili Jin and Jiwang Chen performed real-time PCR, Telugu Narasaraju performed hyperoxia exposure, Narendranath Reddy Chinagari performed animal study, Lin Liu is the principle investigator.

8 Our microarray dataset has been submitted to GEO and can be accessed by following steps: 1. Go to webpage http://www.ncbi.nlm.nih.gov/projects/geo/;

2.   Go to the bottom of this page;

3.   Login to GEO; a. user: **zhongmingchen**; b. password:**123456**; c. Click   button "Login";

4.   Go to our data set; a.   GEO accession: GPL1699; b.   Click button "GO";

5.   View dataset by click hyperlinked items of "Sample id"

transferrin play protective roles against oxidative stress in an animal model. Our studies suggest that AEC I is not just a simple barrier for gas exchange, but a functional cell that protects alveolar epithelium from injury.

4.2 Introduction

AEC II are cuboidal in size and only occupy ~5% of the lung alveolar surface area. They synthesize, secrete and recycle lung surfactant. They also trans-differentiate to AEC I to repair the alveolar epithelium after lung injury or during normal fetal lung development (Fehrenbach, 2001). In contrast, AEC I are squamous in shape and cover ~95% of the surface area of the lung. Traditionally, AEC I are thought to simply function as a barrier for gas exchange. Because of the difficulties in the isolation and culture of AEC I, relatively little was known regarding the functions of AEC I (Williams, 2003). Earlier studies have reported methods to isolate AEC I (Weller & Karnovsky, 1986;Picciano & Rosenbaum, 1978). Recently, several laboratories improved the isolation protocols and obtained relatively pure AEC I (Chen *et al.*, 2004a;Dobbs *et al.*, 1998;Borok *et al.*, 2002). More studies on the identification of AEC I markers (Chen *et al.*, 2004b;Qiao *et al.*, 2003;Gonzalez *et al.*, 2005;Dahlin *et al.*, 2004) and the physiological functions of AEC I (Johnson *et al.*, 2002;Borok *et al.*, 2002;Ridge *et al.*, 2003) were performed. AEC I is highly water-permeable and may be responsible for the high water permeability between the alveolar airspace and the blood vessels (Dobbs *et al.*, 1998). The epithelial Na^+ channel and Na^+-K^+-ATPase have been identified in AEC I, indicating its participation in ion and fluid transport in the lung (Borok *et al.*, 2002;Johnson *et al.*, 2002;Ridge *et al.*, 2003). In this study, we performed DNA microarray analysis of AEC I and AEC II. Further characterization and

functional studies revealed that apolipoprotein (Apo E) and transferrin were specifically synthesized by AEC I in the lung, suggesting that AEC I may function as protective cells to prevent alveolar epithelium from oxidative injury.

4.3   Materials and Methods

4.3.1   Microarray printing and hybridization - The Pan Rat 10K Oligonucleotide Set (MWG Biotech Inc., High Point, NC) contains aminated 50-mer oligonucleotides representing 6,221 known rat genes (4,825 Unigene entries), 3,594 rat ESTs, and 169 Arabidopsis negative controls. The oligonucleotides were suspended in 3x SSC at a final concentration of 25 μM and printed on epoxy coated slides by an OmniGrid 100 arrayer (GeneMachine Inc., San Carlos, CA). Each oligonucleotide was spotted in triplicate in three identical 18 x 18 mm blocks: A, B and C. The spot-spot distance was 160 µm and the space between blocks was 4 mm. The printed slides were air-dried and stored in room temperature until hybridization. Randomly selected 2 slides from each print batch were stained with SYBR green II for quality test and were shown a uniform spot morphology.

Total RNA was extracted from AEC I and AEC II with TRI reagents (Molecular Research Center, Cincinnati, OH). The RNA samples (3 µg each) were reverse-transcribed into cDNA with Alexa 546 (green)- or Alexa 647 (red)-specific primer of the 3DNA 350 expression kit (Genisphere Inc., Hatfield, PA), purified with Microcom YM-30 columns, and mixed with 2x formamide hybridization buffer (50% formamide, 6x SSC, 0.2% SDS, and 10x Denhardt solution) in an equal amount. The green-red color paired cDNA sample was denatured at 80°C for 10 min and added to a DNA microarray slide pre-washed with 0.2% SDS. An additional slide was face-to-face covered on this slide as

cover slip. To eliminate the fluorescence dye-induced bias, dye-flip was applied from different sample preparations. The slides were incubated at 42°C for 48 h, followed by washing and hybridizing with Alexa 546- and Alexa 647-labled dendrimers. Hybridized slides were scanned twice (55% PMT and 90% PMT with 90% laser power) with ScanArray Express scanner (PerkinElmer Life and Analytical Sciences, Boston, MA).

4.3.2   Data analysis - Hybridization images were analyzed by Software package *GenePix pro 4* (Axon Instruments Inc, Union City, CA). The 90% PMT images were used for the alignment of spots with microarray grids and the 50% PMT one for raw data extraction and subsequent analysis. $\log_2$ ratios of AEC I to AEC II were calculated from the background-subtracted mean fluorescence intensities of the respective spots and normalized by local weighted scatter plot smooth (LOWESS) based on print-tip. The quality of each spot was evaluated based on spot morphology and quality index (0~4) assigned by software package *RealSpot* developed in our lab (Chen & Liu, 2005). The quality indices of each gene from replicated hybridizations were averaged. The genes with an average quality index of 1.0 or less were filtered and excluded from further analysis. The differentially expressed genes between AEC I and AEC II were identified by significance analysis of microarray (SAM) (false discovery rate<0.01, q-value<0.01) (http://www-stat.stanford.edu/~tibs/SAM/). The identified differential genes were visually validated with spot images using *RealSpot*. The genes with inconsistent spot images were filtered. Then, the genes were functionally grouped by gene ontology annotation (http://www.geneontology.org and http://www.rgd.mcw.edu). The identified genes with a 2 fold change or higher

were listed in Supplementary information, Tables S1 and S2 online. Additionally, relative gene expression levels were assessed by signal rank in 10K gene list. The gene with the highest signal was ranked 1, and lowest ranked 10,000. The rank average of a gene in AEC I or AEC II was considered as the relative expression level in the respective cell type.

4.3.3 AEC I and AEC II isolation - AEC I and AEC II were isolated from male Sprague-Dawley rats around 250 g according to our recently developed method (Chen *et al.*, 2004a). Briefly, normal rats were anesthetized with intraperitoneal injection of ketamine (40 mg/kg) and xylazine (8 mg/kg). The rats were killed by bleeding through artery after fully numbed. The lungs were perfumed to remove blood cell and removed from rats by surgical procedures (approved by the animal care committee, Oklahoma State University). Sequentially, the lungs were lavaged, digested with elastase, chopped, filtered, plated in rat IgG coated culture dishes to remove macrophages, and negatively selected with rat IgG-, CD4- and T1α-conjugated Dynal-bead to further remove macrophages, lymphocytes and AEC I. The purities of the final AEC I and AEC II preparations were >90% and >96% as assessed by immunocytochemistry using anti-T1α antibodies (E11, a kind gift of Dr. Mary Williams, Boston Univ. and Dr. Antoinette Wetterwald, University of Berne) and the surfactant protein C antibodies (for methodology establishment, and Papanicolaou staining for routine purity assessment), respectively. Cross-contaminations of AEC I and AEC II were less than 0.5%. The viabilities of both cell preparations were >95% as determined by trypan blue dye exclusion.

Table 4.1 Primers used for quantitative real-time PCR

Gene name	GenBank ID	Forward primer	Reverse primer
WAP four-disulfide core domain protein (ps20)	AF037272	TTACAACGGCTGTGCCTATG	TGTAACACTTCCTCCGGACC
Alpha-b-crystallin B	M55534	CAGGAAGATTCCAGCCTCTG	GTTCTCTGGAGCCAGGACAG
secreted acidic cysteine rich glycoprotein (Sparc)	NM_012656	AAACATGGCAAGGTGTGTGA	AAGTGGCAGGAAGAGTCGAA
basigin (Bsg)	NM_012783	GTACACGGTGGATGCAGATG	TCTTTCCCACCTTGATCCTG
transferrin (Tf)	NM_017055	TTCTCATGCTGTTGTGGCTC	CAGCAGGTCTTTCCCAAGAG
glutathione peroxidase precursor (Gpxp)	NM_022525	GAAGATCCATGACATCCGCT	CCGCCTCATATAGGACAGGA
neuronal nicotinic acetylcholine receptor subunit (Chrna10)	NM_022639	AATCTGCACTACTGTGGCCC	CTCTAGTGGCTTGGACTGCC
Advanced glycosylation end product-specific receptor (Ager)	NM_053336	CAGGAAGGACTGAAGCTTGG	GAAAGTCCCCTCATCGACAA
ApoE	S76779	AGGAGCAGACCCAGCAGATA	GGAGTTGGTAGCCACAGAGG
CD9	X76489	GGCTATACCCACAAGGACGA	GCTATGCCACAGCAGTTCAA
18S rRNA		TCCCAGTAAGTGCGGGTCATA	CGAGGGCCTCACTAAACCATC

4.3.4   Quantitative real-time PCR - Data validation of DNA microarray hybridization was performed by quantitative real-time PCR using QuantiTect™ SYBR® Green PCR kit as previously described (Chen *et al.*, 2004b). Primers were listed in Table I. PCR amplification was performed on ABI PRISM 7700 (Applied Biosystems, Foster City, CA). PCR products were purified with GENECLEAN Turbo for constructing standard curves ($10 \sim 10^9$ copies). All gene copy numbers were normalized to 18S rRNA.

4.3.5   **Immunohistochemistry** – This was performed as previously described (Narasaraju *et al.*, 2003). Antigen retrieval was carried out by boiling the sections in microwave for 5 min in 20 mM citrate buffer (pH 6.0) before permeabilization. Primary and secondary antibodies were goat anti-Apo E (1:100, Santa Cruz), rabbit anti-transferrin (1:200, Research Diagnostics), mouse anti-transferrin receptor (1:200, BD Biosciences); mouse anti-LB-180

(1:100, Covance, Richmond, CA); mouse anti-T1α (E11, 1:100); Cy3-conjugated anti-mouse or Alexa 488-conjugated anti-goat antibodies (1:500).

4.3.6 Western blot - This was done according to our previous protocols (Narasaraju *et al.*, 2003). Primary antibodies used include: goat anti-Apo E (1:500), mouse anti-basigin (1:1000, BD Biosciences), mouse anti-CD9 (1:1000, Developmental Studies Hybridoma Bank, Iowa), mouse anti-Sparc (1:1000, Developmental Studies Hybridoma Bank, Iowa), rabbit anti-transferrin (1:1000), rabbit anti-alpha-b-crystallin (1:200, Covance), and anti-β actin (1:4000, Sigma, St. Louis, MO) antibodies. Secondary antibodies were HRP-conjugated anti-goat IgG (1:5000) or HRP-conjugated anti-mouse IgG (1:10000).

4.3.7 Hyperoxia model of rat lung injury - Pathogen free male Sprague-Dawley rats (~ 250 g) were used for animal studies. Oklahoma State University Animal Use and Care Committee approved all animal surgeries and exposure of rats in 95% or higher oxygen for up to 72 hours used in this study. Totally, 20 rats were exposed in the present study. Transferrin (1 mg/rat, Research Diagnostics, Inc), Human recombinant Apo E2 (50 µg/rat, Pan Vera Corporation, Madison, WI) and BSA (1 mg/rat) were intratracheally delivered into rat lungs. The treated rats were placed in a sealed Plexiglas chamber (90 x 45 x 45 cm) and exposed to >95% oxygen for 48, 60 or 72 hours. The flow rate was maintained at 10 L/min. The oxygen concentration was continuously monitored with an oxygen sensor (Vacu-med, Ventura, CA). Rats had free access for food and water. After a 60 hours exposure, rats were anesthetized with

intraperitoneal injection of ketamine (40 mg/kg) and xylazine (8 mg/kg). Pleural fluid was collected with 1~1000µl pipette and 1ml tips before the lungs were perfused with 50 mM phosphate buffered saline (pH 7.4). The volume of pleural fluid from each rat was measured with 1000µl pipette and recorded. Lungs were lavaged with 7 ml of normal saline for three times. Lavage solution (21 ml) was centrifuged to remove the cells and was concentrated to the same volume for western blot analysis.

4.4 Results and Discussion

To explore specific gene expressions and their potential functions in AEC I, we performed DNA microarray analysis using freshly isolated rat AEC I and AEC II (DNA microarray data have been submitted to GEO database: http://www.ncbi.nlm.nih.gov/geo/submission/login/, user name= zhongmingchen, password= 123456, Series entry = GSE2160). The purities of our AEC I and AEC II preparations were >90% and >96%, respectively, and the viabilities were >95%. The cross-contaminations between two cell preparations were <0.5%. The isolated cells have typical characteristics of the *in vivo* lung, remaining functional and suitable for culture (Chen *et al.*, 2004a). We used 5 independent preparations of AEC I and AEC II isolated from adult male rats for microarray analysis and 6 technically replicated hybridizations performed for each preparation, i.e., each labeled AEC I-AEC II cDNA was hybridized on two slides (a slide and its cover slide), on which each gene had 3 replicated spots and totally 6 spots per gene. Dye-flip was applied to dual-color microarray hybridizations of the 5 preparations AEC I vs. AEC II (i.e., Alexa 546-AEC I vs. Alexa 647-ACE II and Alexa 647-AEC I vs. Alexa 546-ACE II were alternatively used), to eliminate the bias resulted from the fluorescence dye during the

image scanning step. After data filter, LOWESS normalization, and statistical SAM test, 2,222 known genes or ESTs in our 10K microarray were significantly and differentially expressed in AEC I or AEC II (p<0.01); 1,080 were expressed higher in AEC I and 1,142 higher in AEC II (**Fig.** 4.1a). The number of differentially expressed genes in AEC I or AEC II was similar due to the ratio centering during LOWESS normalization. Among the 1,265 known genes excluding ESTs, there were 498 for AEC I and 767 for AEC II. To select genes for further study, especially for validation with real-time PCR, the 1,265 known genes were grouped in fold changes and relative gene expression rank (**Fig.** 1b & 1c). More genes were differentially expressed in AEC I than AEC II when the fold change decreased. For instance, 109 genes for AEC I and 87 for AEC II exhibited a fold change of ≥ 2. The differentially expressed genes were distributed in all the gene expression levels (rank meaning: 1 ~ 500 were highly expressed genes, 500 ~ 1500 middle, 1,500 and 4,000 weak, and 4,000~10,000 not visually detectable). Genes differentially (2 or more fold changes) and highly (ranked 1~500) expressed in AEC I were considered to be further investigated.

Gene ontology annotations were another consideration for further targeting novel genes that may essential for AEC I.   The differentially expressed genes were split into functional groups using the information from Rat Genome Database GO association (RGD, www.rgd.mcw.edu) and gene ontology definitions (http://www.geneontology.org). RGD annotated 893 out of the1,265 differentially expressed genes in AEC during the manuscript preparation, 327 for AEC I, and 566 for AEC II. The main gene groups were listed based on GO categories (Table II). In our gene list, each gene was assigned about 5 GO IDs,

which reflected the multiple functions of most of the known genes. Consequently, the total gene number from all the GO categories was larger than 893, the actually annotated genes in our data set. A high percentage of the differentially expressed genes observed for membrane or nucleic proteins (57%, based on GO categories in table II, instead of the actual genes. Other percentages were similar), but relatively less in cytoplasm (31%) or extracellular space (12%). On the other hand, AEC I had more mitochondrial proteins (26 genes) than AEC II (20 genes). Functionally, many of these genes were related to known AEC functions such as phase II xenobiotic metabolism (transferase activity), oxygen metabolism (oxidoreductase activity) and fluid transport (ion channel activity). Other major functions included gene regulation (DNA binding, transcription factor activity, and RNA binding) and signaling pathways (calcium ion binding and G-protein coupling receptor activities).

We next focused on the AEC I-specific genes based on the above information of fold-change, rank and GO annotation. Further characterization and functional studies of these genes are not only able to identify AEC I markers, but also provide clues for potential functions of AEC I. Many known AEC I markers such as caveolin-1β (5.90 ± 1.21), caveolin-1α (4.11 ± 1.22), T1α (3.78 ± 1.08), and aquaporin 5 (4.99 ± 1.08) were shown to be differentially expressed in AEC I in our 10K microarray data set (ratio of AEC I vs. AEC II, means ± SE, for a complete list of AEC I and AEC II genes, see Supplementary information, Table S1 and S2 online).
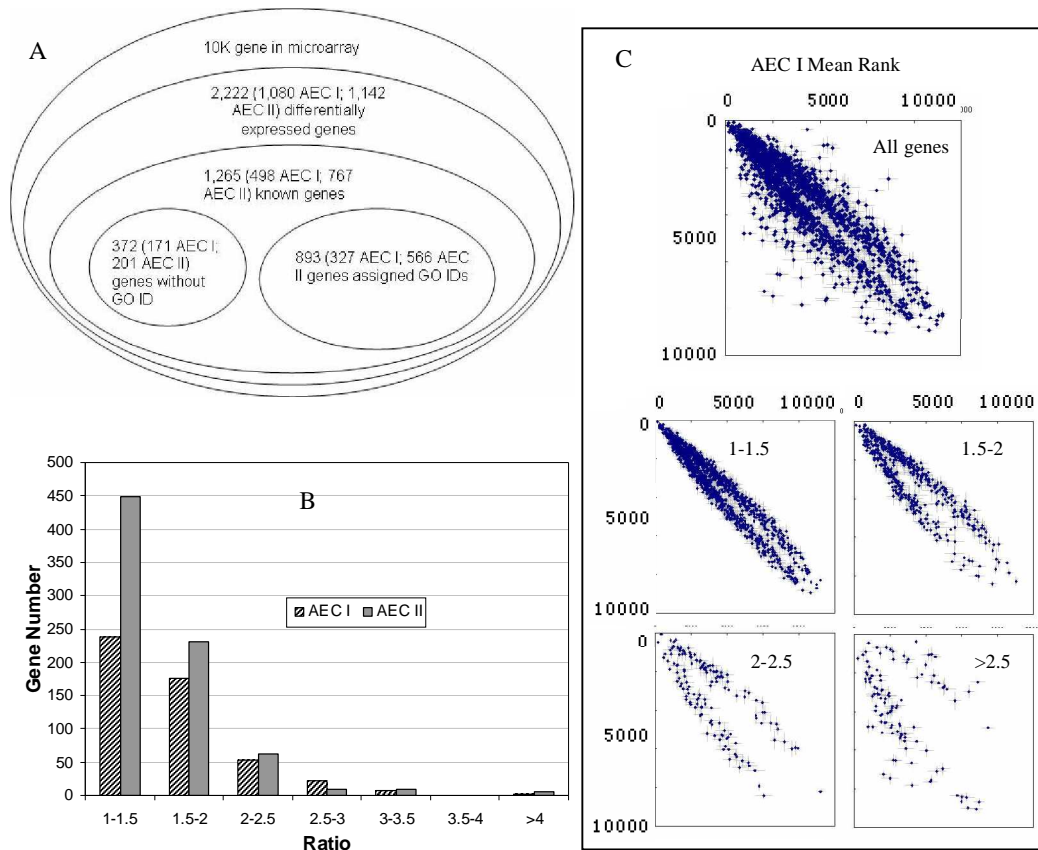
Fig. 4.1 DNA microarray analysis of alveolar epithelial type I and type II cells (AEC I and AEC II). Freshly isolated AEC I and AEC II were subjected to 10 K rat DNA microarray (~6,000 known genes and ~ 4,000 EST) analysis. There were 5 biological replications and 6 technical replications (2 slides, 3 replicates on one slide). Total replications were 30 for each gene. (a) gene numbers during different stages of data analysis. (b) fold change distributions of differentially expressed genes in AEC I and AEC II. (c) Relative expression levels of differentially expressed genes as shown in expression ranks. Data present ranks ± SE (1~10,000) of gene expressions based on 10 K genes, each with 30 replicated spots. Left four panels show rank plots with fold changes of 1-1.5, 1.5-2, 2-2.5 and >2.5, and right panel all of the differentially expressed genes.

94

Table 4.2 Functional groups of differential genes between AEC I and AEC II

GO ID	Meaning	Total	AEC I	AEC II
Cellular location				
GO:0016021	integral to membrane	191	67	124
GO:0016020	Membrane	119	46	73
GO:0005634	Nucleus	108	25	83
GO:0005615	extracellular space	88	31	57
GO:0005622	Intracellular	46	14	32
GO:0005739	mitochondrion	46	26	20
GO:0005737	Cytoplasm	29	12	17
GO:0005576	extracellular	25	5	20
GO:0005783	endoplasmic reticulum	25	14	11
GO:0005840	Ribosome	24	9	15
GO:0005856	cytoskeleton	17	7	10
GO:0005794	Golgi apparatus	14	7	7
Function				
GO:0016787	Hydrolase activity	73	35	38
GO:0016740	transferase activity	69	25	44
GO:0003677	DNA binding	64	13	51
GO:0005524	ATP binding	58	18	40
GO:0005515	protein binding	47	14	33
GO:0004872	receptor activity	45	9	36
GO:0016491	oxidoreductase activity	45	21	24
GO:0005509	calcium ion binding	44	18	26
GO:0003700	transcription factor activity	42	5	37
GO:0003824	catalytic activity	32	15	17
GO:0004672	protein kinase activity	31	8	23
GO:0004930	G-protein coupled receptor activity	29	5	24
GO:0004674	protein serine/threonine kinase activity	28	8	20
GO:0003723	RNA binding	27	9	18
GO:0005216	ion channel activity	27	7	20
GO:0016301	kinase activity	27	9	18
GO:0003735	Structural constituent of ribosome	26	9	17
GO:0000287	magnesium ion binding	22	5	17
GO:0003779	actin binding	21	12	9
GO:0005215	transporter activity	21	4	17
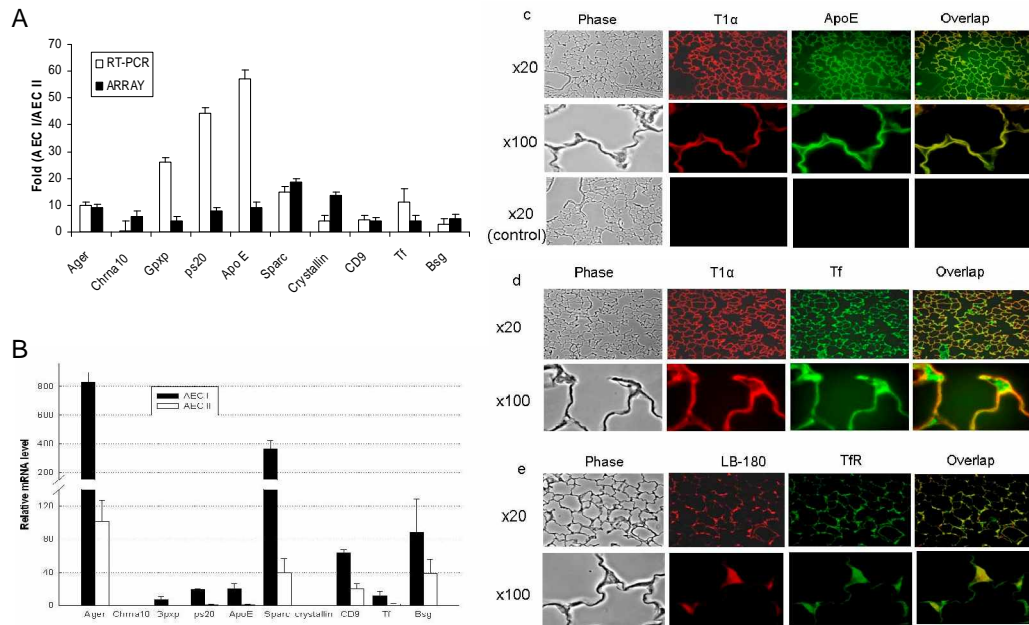
Fig. 4.2 **Verification of selected AEC I-specific genes at mRNA and protein levels**. (a, b) mRNA level: Quantitative real time PCR (qRT-PCR) was performed using an absolute method. Copy number was obtained from a standard curve constructed from PCR products of each gene and normalized to 18S rRNA. Fold changes were calculated from copy numbers of AEC I and AEC II and compared with the DNA microarray data (panel a). Data shown are means ± SE (n=10 for qRT-PCR, 5 independent cell preparations, each run in duplicate, or n=30 for array data, 5 biological and 6 technical replications). The mRNA levels of each gene in AEC I and AEC II were expressed as copy number/10^6 18S rRNA (panel b, means ± SE, n=10). Ager: Advanced glycosylation end product-specific receptor, Chrna 10: neuronal nicotinic acetylcholine receptor subunit alpha 10, Gpxp: plasma glutathione peroxidase precursor, ps20: wap four-disulfide core domain protein, Apo E: apolipoprotein E, Sparc: secreted acidic cysteine-rich glycoprotein, crystalline: alpha-b-crystallin, CD9: platelet-cell surface glycoprotein, Tf: transferrin, and Bsg: basigin. (c-e) Immunolocalization of Apo E, Tf and Tf receptor (TfR) in rat lung. Double labeling were performed using Apo E antibodies and the known AEC I marker, T1 α antibodies (Panel c), Tf and T1 α antibodies (Panel d) or TfR antibodies and the know AEC II marker, LB180 antibodies (Panel e). The control omitting primary antibodies (control) is also shown in Panel c.
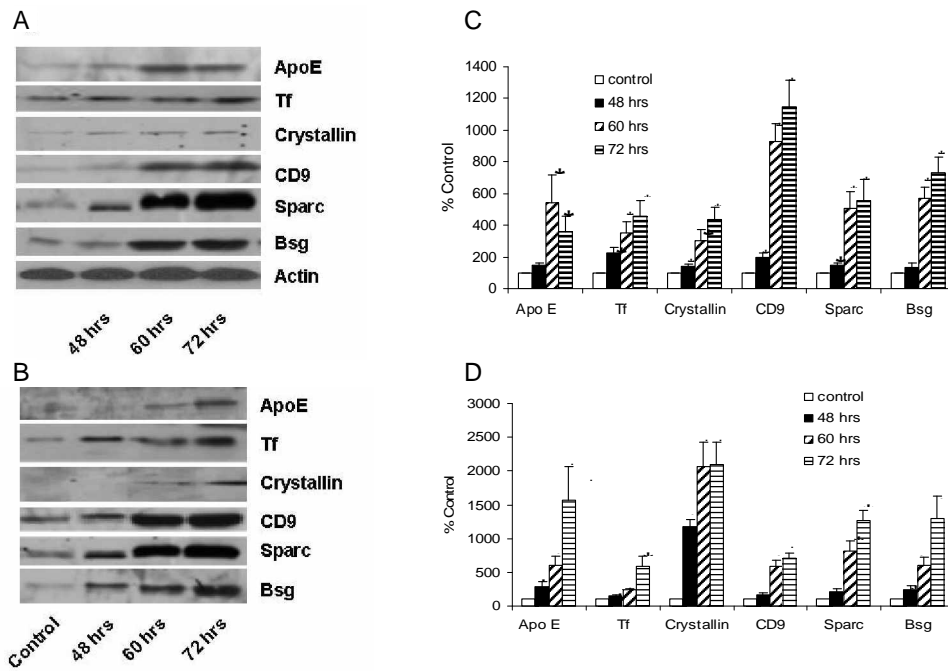
Fig. 4.3 Protein quantification of AEC I genes in lung tissue and lavage fluid after hyperoxia exposure. Rat were exposed to >95% O_2 for 48, 60 and 72 hours. Lung tissue (Panels a & b) and lavage fluid (panels c & d) were collected for measuring protein levels of AEC I genes by western blot. Panels a and b are representative blots; Panels c & d are quantitative data determined by densitometry. Data were normalized to β-actin and expressed as a percentage of control (no hyperoxia exposure) from three different animals (means ± SE). *p<0.05 vs. control; **p<0.01 vs. control.

We selected 10 AEC I-specific genes for further characterization based on the fold change (log 2 ratio >2), rank (1~500) and their cellular location. Real time PCR confirmed 9 genes out of 10 (**Fig.** 4.2a). In general, real time PCR showed a higher fold change than DNA microarray. The large difference in expression measured by real-time PCR may resulted from the large data dynamic range (10~10^{10} copy-number per gene), selections of oligonucleotide primers inherited from PCR, low background-to-signal ratio of DNA microarray (<100, typically 1~10), and different data normalization approach

(housekeeping gene for PCR and LOWESS for microarray). We can not verify neuronal nicotinic acetylcholine receptor subunit alpha 10 (Chrna10), because of its low expression. Relative mRNA levels of these ten genes were shown in **Fig.** 4.2b. Advanced glycosylation end product-specific receptor (Ager, NM_05333) had a highest expression in AEC I, followed by secreted acidic cysteine-rich glycoprotein (Sparc, NM_01265). Platelet-cell surface glycoprotein (CD9, X76489) and basigin (Bsg, NM_01278) had a middle expression in AEC I. Other four genes, plasma glutathione peroxidase precursor (Gpxp, NM_022639), wap four-disulfide core domain protein (ps20, AF037272), apolipoprotein E (Apo E, S76779) and transferrin (Tf, NM_017055) had a significant amount of mRNA expression in AEC I and their expression appeared to be more specific to AEC I. Chrna10 (NM_053336) and alpha-b-crystallin (M55534) mRNA levels were very low in AEC I. Ager and Sparc have previously been shown to be cell markers of AEC I (Shirasawa *et al.*, 2004;Mustafa *et al.*, 2004;Fehrenbach *et al.*, 1998).

The lung epithelium is directly exposed to the environmental oxidants and continuously under stress via the generation of reactive oxygen radicals by inhaled pollutants or during lung inflammation response. AEC II was thought to be a defender cell in the lung (Fehrenbach, 2001). Our current microarray analysis identified many of the AEC I genes that appear to play active roles in lung defense against different stresses. We then directed our attentions to two of the genes, Apo E and transferrin, and examined whether they participate in the defense system against oxidative stress. Apo E is a 34 kDa plasma lipoprotein high-affinity ligand for low density lipoprotein receptors and is involved in the transport of cholesterol and other lipids (Mahley, 1988). It also

participates in the nerve injury and regenerative process and protects against neuronal injury via antioxidant effects (Kitagawa *et al.*, 2002;Lee *et al.*, 2004). It is also involved in Alzheimer's disease (Miyata & Smith, 1996). Low density lipoprotein has been shown to stimulate surfactant synthesis and secretion (Ryan *et al.*, 2002;Pian & Dobbs, 1997).   Transferrin is another plasma major iron transport glycoprotein (80 kDa) that transports iron from the plasma to various cellular localizations (Mateos *et al.*, 1998). The uptake of iron is mediated by transferrin receptors. Transferrin is the major antioxidant in the lower respiratory tract through its binding of iron, which exerts toxic effects by catalyzing highly reactive hydroxyl radicals from superoxide and hydrogen peroxide in the lung (Pacht & Davis, 1988;Lykens *et al.*, 1992;Gutteridge *et al.*, 1996;Gerber *et al.*, 1999). It also contributes to anti-infection defense in the lung. Iron-free transferrin reduces respiratory failure and increases lung surfactant activity (Hallman *et al.*, 1994)

Apo E and transferrin are primarily synthesized in liver, but also in other peripheral tissues including lung (Williams *et al.*, 1985;Driscoll & Getz, 1984;Yang *et al.*, 1997). Transferrin levels in bronchoalveolar lavage (BAL) fluid are higher than that in the plasma (Mateos *et al.*, 1998), indicating its local synthesis, in addition to coming from the plasma through transudation. However, local sources of both proteins in the lung are unknown. Double-labeling of rat lung tissue indicated that Apo E was co-localized with the known AEC I marker, T1α, suggesting its AEC I localization (**Fig.** 4.2c). The control that omits both primary antibodies did not show signals (**Fig.** 4.2c). Other controls that omit one of the primary antibodies only show signals in one "channel" (filter), but not in another (data not shown). In contrast, transferrin

was found to be present in both AEC I and AEC II (**Fig.** 4.2d). The localization of transferrin in AEC II is likely due to the binding of transferrin secreted by AEC I to the transferrin receptor on AEC II because transferrin receptors were only expressed in AEC II, but not AEC I as revealed by double-labeling transferrin receptor with LB-180, a type II cell marker (**Fig.** 4.2e). These data suggest that Apo E and transferrin are synthesized in AEC I, which may partially account for local sources of these proteins in the alveolar air space.

Induction of antioxidant proteins by oxidative stress is an adaptive and protective mechanism for cells responding to the stress. We have previously shown that hyperoxia exposure caused oxidative stress, inflammation, and acute lung injury (Narasaraju *et al.*, 2003). We reasoned that hyperoxia might up-regulate Apo E and transferrin expression in the lung. We therefore exposed rats to >95% oxygen for different times and determined the protein levels of these proteins in lung tissues by Western blot. After different-time point hyperoxia treatment (48 hours, 60 hours, and 72 hours), Apo E and transferrin protein levels in the lung were increased in an exposure time-dependent manner (**Fig.** 4.3 a & b). It is well known that AEC I are more liable to injury than AEC II. As a result, AEC I markers have been detected in alveolar fluids after lung injury and used as biochemical markers for acute lung injury (McElroy *et al.*, 1997). Indeed, Apo E and transferrin proteins were also increased in BAL fluid (**Fig.** 3c & d). These increases correspond well with the severity of lung injury index (Narasaraju *et al.*, 2003). However, the results should be explained with caution. In addition to the AEC I injury, the increase of these proteins in alveolar air space also could be due to the increase of the synthesis and/or the secretion of these proteins by AEC I. In addition to Apo E

and transferrin, other AEC I proteins including α-b-crystallin, CD9, sparc, and basigin were also increased in both lung tissue and BAL fluid in response to hyperoxia exposure. (**Fig.** 4.3 a-d).
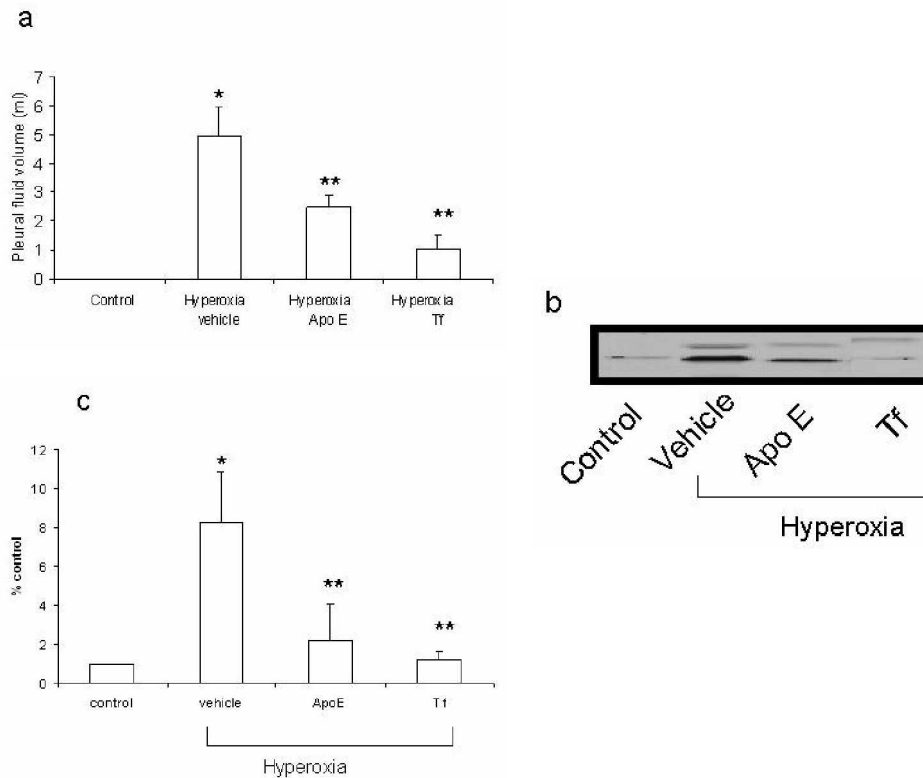
Fig. 4.4 Apolipoprotein E and transferrin protect the lung from hyperoxia injury. Apolipoprotein E (Apo E, 20 µg/100 g BW), or transferrin (Tf, 400 µg/ 100 g BW) were delivered intratracheally into rat lungs. Since there were no significant differences between saline and heat-denatured Tf, the data from both groups were combined as vehicle controls. Rats were exposed to >95% O_2 for 60 hours. The plural fluid volume was measured (panel a). The same volume of lavage fluid was used for the detection of T1 α released into alveolar air space by western blot. The results were quantified by densitometry and expressed as a percentage of control (no hyperoxia exposure). Data shown are means ± SE from six animals. Panel b shows a representative western blot. *P<0.01 vs. control; **P<0.05 vs. vehicle group.

Further functional studies were performed to assess whether Apo E and transferrin can protect alveolar epithelial cells from injury originated from oxidative stress in an animal model. We used an animal model to study

whether the delivery of ApoE, transferrin into rat lungs can prevent hyperoxia-mediated lung injury. Exposure of rats to >95% O_2 for 60 hours resulted in an increase of pleural fluid volume, consistent with previous reports (Narasaraju *et al.*, 2003). No pleural fluid was observed in the control unexposed rats. Delivery of Apo E (20 μg/100 g BW) or transferrin (400 μg/100 g BW) to the lungs before O_2 exposure significantly reduced amounts of pleural fluid (**Fig.** 4.4a). We also used an AEC I marker, T1 α, as a biochemical index of lung injury (McElroy *et al.*, 1997). In the control rats, very little of T1 α protein was detected in BAL fluid. Hyperoxia exposure caused a significant increase of T1 α (**Fig.** 4.4b and 4.4c). However, the increase of T1 α was prevented in the Apo E or transferrin-treated rats. The denatured transferrin or vehicle controls did not provide protection. These data suggest that Apo E and transferrin can prevent hyperoxia-mediated lung injury.

In addition to Apo E and transferrin, several AEC I genes identified in this study may also be involved in lung injury and repair, including Sparc, basigin, alpha-crystallin B, and CD9. These genes were up-regulated in our hyperoxia-mediated lung injury model (**Fig.** 4.3). Sparc (also called osteonectin, BM40) is a Ca^{2+}-binding stress-related protein and may function as an extracellular modulator of Ca^{2+} and other cation-sensitive proteins or proteases that facilitate cellular proliferation in response to injury (Sage, 1992). The accelerated wound closure, a condition contributing to enhanced contractibility, was observed in Sparc-null demise, resulting from its decreased collagen content (Bradshaw *et al.*, 2002). Sparc accumulation was used as a marker for stromal repair (Bradshaw *et al.*, 2003). Basigin (also called CD147 or extracellular matrix metalloproteinase inducer) is a multifunctional

transmembrane protein involved in inflammation and tumor invasion (Muramatsu & Miyauchi, 2003). Basigin induces tumor cell invasion by stimulating the production of matrix metalloproteases from surrounding fibroblasts (Kanekura *et al.*, 2002). Similarly, basigin released from AEC I may also trigger the production or release of metalloproteases in AEC I-surrounding fibroblasts. The released metalloproteases then degrade extracellular proteins in injured cells and facilitate lung repair. Alpha-crystallin B (heat shock protein 20 or hsp20) is highly expressed in mammal eye lens. Together with hsp27, it may respond against cell stresses (Horwitz, 2003). CD9, a member of the tetraspanin family, is involved in cellular activities such as cell migration, proliferation, adhesion (Murayama *et al.*, 2004) and the formation and preservation of various different membrane complexes consisting of several functional proteins (Kijimoto-Ochiai *et al.*, 2004). Anti-CD9 monoclonal antibodies inhibited cell proliferation, reduced cell viability and induced morphological changes specific to apoptosis (Murayama *et al.*, 2004).

This study identified specifically expressed genes in AEC I from the global gene expression of freshly isolated AEC I and AEC II using 10K rat DNA microarray. The data suggest that AEC I play active roles in gas exchange, liquid clearance, lung injury and repair, and lung immune responses. Further studies of two of the AEC I genes, Apo E and transferrin, revealed their functions in the cellular defense against oxidative stress. Therefore, the current studies not only reveal a new function of AEC I cells, cellular defense, but also change the concept that AEC I is only a barrier for gas exchange. AEC I is a fully functional cells. Future studies should discover other important functions of AEC I.

4.5   Acknowledgements

4.6 REFERENCES

Borok, Z., Liebler, J. M., Lubman, R. L., Foster, M. J., Zhou, B., Li, X., Zabski, S. M., Kim, K. J., & Crandall, E. D. (2002). Na transport proteins are expressed by rat alveolar epithelial type I cells. *Am J Physiol Lung Cell Mol Physiol* **282**, L599-L608.

Bradshaw, A. D., Puolakkainen, P., Dasgupta, J., Davidson, J. M., Wight, T. N., & Helene, S. E. (2003). SPARC-null mice display abnormalities in the dermis characterized by decreased collagen fibril diameter and reduced tensile strength. *J Invest Dermatol.* **120**, 949-955.

Bradshaw, A. D., Reed, M. J., & Sage, E. H. (2002). SPARC-null mice exhibit accelerated cutaneous wound closure. *J Histochem.Cytochem.* **50**, 1-10.

Chen, J. W., Chen, Z., Narasaraju, T., Jin, N., & Liu, L. (2004a). Isolation of hightly pure alveolar epithelial type I and type II cells from rat lungs. *Lab Invest* **84**, 727-735.

Chen, Z., Jin, N., Narasaraju, T., Chen, J., McFarland, L. R., Scott, M., & Liu, L. (2004b). Identification of two novel markers for alveolar epithelial type I and II

cells. *Biochem.Biophys.Res.Commun.* **319**, 774-780.

Chen, Z. & Liu, L. (2005). RealSpot: Software validating results from DNA microarray data analysis with spot images. *Physiol Genomics* **21**, 284-291.

Dahlin, K., Mager, E. M., Allen, L., Tigue, Z., Goodglick, L., Wadehra, M., & Dobbs, L. (2004). Identification of genes differentially expressed in rat alveolar type I cells. *Am J Respir Cell Mol Biol.* **31**, 309-316.

Dobbs, L. G., Gonzalez, R., Matthay, M. A., Carter, E. P., Allen, L., & Verkman, A. S. (1998). Highly water-permeable type I alveolar epithelial cells confer high water permeability between the airspace and vasculature in rat lung. *Proc Natl Acad Sci U S A* **95**, 2991-2996.

Driscoll, D. M. & Getz, G. S. (1984). Extrahepatic synthesis of apolipoprotein E. *J.Lipid Res.* **25**, 1368-1379.

Fehrenbach, H. (2001). Alveolar epithelial type II cell: defender of the alveolus revisited. *Respir Res* **2**, 33-46.

Fehrenbach, H., Kasper, M., Tschernig, T., Shearman, M. S., Schuh, D., & Muller, M. (1998). Receptor for advanced glycation endproducts (RAGE) exhibits highly differential cellular and subcellular localisation in rat and human lung. *Cell Mol Biol.(Noisy.-le-grand)* **44**, 1147-1157.

Gerber, C. E., Bruchelt, G., Stegmann, H., Schweinsberg, F., & Speer, C. P.

(1999). Presence of bleomycin-detectable free iron in the alveolar system of preterm infants. *Biochem.Biophys.Res.Commun.* **257**, 218-222.

Gonzalez, R., Yang, Y. H., Griffin, C., Allen, L., Tigue, Z., & Dobbs, L. (2005). Freshly-isolated rat alveolar type I cells, type II cells, and cultured type II cells have distinct molecular phenotypes. *Am J Physiol Lung Cell Mol Physiol* **288**, L179-L189.

Gutteridge, J. M., Mumby, S., Quinlan, G. J., Chung, K. F., & Evans, T. W. (1996). Pro-oxidant iron is present in human pulmonary epithelial lining fluid: implications for oxidative stress in the lung. *Biochem Biophys Res Commun* **220**, 1024-1027.

Hallman, M., Sarnesto, A., & Bry, K. (1994). Interaction of transferrin saturated with iron with lung surfactant in respiratory failure. *J Appl.Physiol* **77**, 757-766.

Horwitz, J. (2003). Alpha-crystallin. *Exp.Eye Res* **76**, 145-153.

Johnson, M. D., Widdicombe, J. H., Allen, L., Barbry, P., & Dobbs, L. G. (2002). Alveolar epithelial type I cells contain transport proteins and transport sodium, supporting an active role for type I cells in regulation of lung liquid homeostasis. *Proceedings of the National Academy of Sciences of the United States of America* **99**, 1966-1971.

Kanekura, T., Chen, X., & Kanzaki, T. (2002). Basigin (CD147) is expressed on melanoma cells and induces tumor cell invasion by stimulating production of

matrix metalloproteases by fibroblasts. *Int.J Cancer* **99**, 520-528.

Kijimoto-Ochiai, S., Noguchi, A., Ohnishi, T., & Araki, Y. (2004). Complex formation of CD23/surface immunoglobulin and CD23/CD81/MHC class II on an EBV-transformed human B cell line and inferable role of tetraspanin. *Microbiol.Immunol.* **48**, 417-426.

Kitagawa, K., Matsumoto, M., Kuwabara, K., Takasawa, K., Tanaka, S., Sasaki, T., Matsushita, K., Ohtsuki, T., Yanagihara, T., & Hori, M. (2002). Protective effect of apolipoprotein E against ischemic neuronal injury is mediated through antioxidant action. *J.Neurosci.Res.* **68**, 226-232.

Lee, Y., Aono, M., Laskowitz, D., Warner, D. S., & Pearlstein, R. D. (2004). Apolipoprotein E protects against oxidative stress in mixed neuronal-glial cell cultures by reducing glutamate toxicity. *Neurochem.Int.* **44**, 107-118.

Lykens, M. G., Davis, W. B., & Pacht, E. R. (1992). Antioxidant activity of bronchoalveolar lavage fluid in the adult respiratory distress syndrome. *Am J Physiol* **262**, L169-L175.

Mahley, R. W. (1988). Apolipoprotein E: cholesterol transport protein with expanding role in cell biology. *Science* **240**, 622-630.

Mateos, F., Brock, J. H., & Perez-Arellano, J. L. (1998). Iron metabolism in the lower respiratory tract. *Thorax* **53**, 594-600.

McElroy, M. C., Pittet, J. F., Allen, L., Wiener-Kronish, J. P., & Dobbs, L. G. (1997). Biochemical detection of type I cell damage after nitrogen dioxide-induced lung injury in rats. *Am J Physiol* **273**, L1228-L1234.

Miyata, M. & Smith, J. D. (1996). Apolipoprotein E allele-specific antioxidant activity and effects on cytotoxicity by oxidative insults and beta-amyloid peptides. *Nat.Genet.* **14**, 55-61.

Muramatsu, T. & Miyauchi, T. (2003). Basigin (CD147): a multifunctional transmembrane protein involved in reproduction, neural function, inflammation and tumor invasion. *Histol.Histopathol.* **18**, 981-987.

Murayama, Y., Miyagawa, J., Oritani, K., Yoshida, H., Yamamoto, K., Kishida, O., Miyazaki, T., Tsutsui, S., Kiyohara, T., Miyazaki, Y., Higashiyama, S., Matsuzawa, Y., & Shinomura, Y. (2004). CD9-mediated activation of the p46 Shc isoform leads to apoptosis in cancer cells. *J Cell Sci* **117**, 3379-3388.

Mustafa, S. B., DiGeronimo, R. J., Petershack, J. A., Alcorn, J. L., & Seidner, S. R. (2004). Postnatal glucocorticoids induce alpha-ENaC formation and regulate glucocorticoid receptors in the preterm rabbit lung. *Am.J.Physiol Lung Cell Mol.Physiol* **286**, L73-L80.

Narasaraju, T. A., Jin, N., Narendranath, C. R., Chen, Z., Gou, D., & Liu, L. (2003). Protein nitration in rat lungs during hyperoxia exposure: a possible role of myeloperoxidase. *Am.J.Physiol Lung Cell Mol.Physiol* **285**, L1037-L1045.

Pacht, E. R. & Davis, W. B. (1988). Role of transferrin and ceruloplasmin in antioxidant activity of lung epithelial lining fluid. *J.Appl.Physiol* **64**, 2092-2099.

Pian, M. S. & Dobbs, L. G. (1997). Lipoprotein-stimulated surfactant secretion in alveolar type II cells: mediation by heterotrimeric G proteins. *Am.J.Physiol* **273**, L634-L639.

Picciano, P. & Rosenbaum, R. M. (1978). The type 1 alveolar lining cells of the mammalian lung. I. Isolation and enrichment from dissociated adult rabbit lung. *Am.J.Pathol.* **90**, 99-122.

Qiao, R., Zhou, B., Liebler, J. M., Li, X., Crandall, E. D., & Borok, Z. (2003). Identification of three genes of known function expressed by alveolar epithelial cells. *Am.J.Respir.Cell Mol.Biol.* **29**, 95-105.

Ridge, K. M., Olivera, W. G., Saldias, F., Azzam, Z., Horowitz, S., Rutschman, D. H., Dumasius, V., Factor, P., & Sznajder, J. I. (2003). Alveolar type 1 cells express the alpha2 Na,K-ATPase, which contributes to lung liquid clearance. *Circ.Res* **92**, 453-460.

Ryan, A. J., Medh, J. D., McCoy, D. M., Salome, R. G., & Mallampalli, R. K. (2002). Maternal loading with very low-density lipoproteins stimulates fetal surfactant synthesis. *Am.J.Physiol Lung Cell Mol.Physiol* **283**, L310-L318.

Sage, E. H. (1992). Secretion of SPARC by endothelial cells transformed by polyoma middle T oncogene inhibits the growth of normal endothelial cells *in*

vitro. Biochem Cell Biol. **70**, 579-592.

Shirasawa, M., Fujiwara, N., Hirabayashi, S., Ohno, H., Iida, J., Makita, K., & Hata, Y. (2004). Receptor for advanced glycation end-products is a marker of type I lung alveolar cells. *Genes Cells* **9**, 165-174.

Weller, N. K. & Karnovsky, M. J. (1986). Isolation of pulmonary alveolar type I cells from adult rats. *Am.J.Pathol.* **124**, 448-456.

Williams, D. L., Dawson, P. A., Newman, T. C., & Rudel, L. L. (1985). Apolipoprotein E synthesis in peripheral tissues of nonhuman primates. *Journal of Biological Chemistry* **260**, 2444-2451.

Williams, M. C. (2003). ALVEOLAR TYPE I CELLS: Molecular Phenotype and Development. *Annu.Rev.Physiol* **65:669-95.**, 669-695.

Yang, F., Friedrichs, W. E., & Coalson, J. J. (1997). Regulation of transferrin gene expression during lung development and injury. *Am J Physiol* **273**, L417-L426.

The following additional files are available upon request from Dr. Lin Liu (liulin@okstate.edu): Additional file 1: Supplementary Table S1, genes predominantly expressed in AEC I in Excel format. Additional file 2: Supplementary Table S2, genes predominantly expressed in AEC II in Excel format.

Chapter 5

Gene Expression Profiles of Rat Type II Pneumocytes During

Hyperoxia Exposure and Early Recovery[9,10]

5.1 ABSTRACT

Rationale: Alveolar epithelial cell (AEC) injury and repair during hyperoxia exposure and recovery stages have been investigated for decades, but the molecular mechanisms of this process are still not completely clear. **Objectives:** Identify the potentially important genes during lung injury and repair. **Methods:** We studied the gene expression profiling of isolated AEC type II (AEC II) from control, 48 hours hyperoxia (>95% O_2)-exposed (Day 0) and 1-7 day recovering (Day 1, 3, 5, and 7) rats using DNA microarray containing 10,000 genes. **Measurements and Main Results:** Fifty genes showed significantly differential expression between two or more time points ($p < 0.05$, fold change>2). These genes can be classified into 8 unique gene expression patterns. Real-time PCR verified 14 selected genes in three patterns related to hyperoxia exposure and early recovery. The change in the

---

protein level for one of the selected genes, Dlk1, paralleled that of the mRNA level. These genes were involved in AEC protection (6), apoptosis (2) and cell proliferation/differentiation (6). In an *in vitro* AEC trans-differentiation culture model, the expression of the cell proliferation/differentiation genes identified above were consistent with their predicted roles in the trans-differentiation of AEC. **Conclusion:** These data indicated that a coordinated mechanism may control the AEC differentiation during *in vivo* hyperoxia exposure and recovery and *in vitro* AEC culture.

5.2 INTRODUCTION

Alveoli are the basic gas exchange units of the lung. Their epithelia consist of alveolar epithelial cells type I (AEC I) and type II (AEC II). AEC I are squamous and contributes to ~8% of total lung cells. They occupy >95% of alveolar epithelial surface and function in O_2 and CO_2 exchange. AEC II are round and account for ~15% of the lung cells and <5% of the surface space. AEC II secrete surfactant lipid and proteins and also play roles in xenobiotic transformation, fluid clearance, and the repair of injured alveolar epithelium.

Alveoli are constantly exposed to oxygen and environmental stress (e.g. smoke, pollutant gas) and thus are susceptible to injury. AEC I are more susceptible to injury due to their large flat shape and inability to perform mitogenesis. Exposure of animals to hyperoxia, followed by recovery in room air, is a well-established experimental model for AEC injury and repair (1). Hyperoxia exposure results in the death of endothelial cells and AEC I, possibly through necrosis or apoptosis (2). As a result of AEC I injury, these cells detach leaving behind denuded basement membrane. During the recovery stages, AEC II proliferate and differentiate to AEC I to replace the

injured or dead AECs through a cascade of signaling transduction pathways (3).

Recently, genomic approaches have been applied for monitoring gene expression profiles of *in vitro* cultured AEC in response to glucocorticoid, TNFα and LPS, KGF, and disel exhaust particles treatment (4-7). Microarray analysis has shown that the 7-day cultured AEC II are different from the freshly isolated AEC I (8). Using a similar AEC II culture system, we have previously identified two AEC markers, purinergic receptor P2X7 and gamma-aminobutyric acid receptor pi subunit using a commercial DNA microarray (9). However, potential phenotypic changes of AEC during cell culture limit the interpretation of the results.

Several groups have used whole mouse lung tissue for gene expression profiling (10-14). One study has reported changes in gene expression of lung tissue from mice exposed to hyperoxia for up to 48 hours (14). However, no recovery phases have been analyzed. The difference in gene expression among different species has been observed between rats and mice in response to hypoxia (15). The differential gene expression from the whole lung tissue reflects the average effect of over 40 different cell types in the pulmonary system, which makes it difficult to assess changes in a particular cell type. Also, the average effect may significantly reduce the sensitivity of DNA microarray analysis. Furthermore, in whole tissue gene profiling, it is not possible to distinguish between the contributions of lung cells and infiltrating cells accompanied with inflammation induced by hyperoxia exposure.

To eliminate the caveats above, we isolated highly pure AEC II from

hyperoxia-exposed (48 hours) and recovering (day 0, 1, 3, 5, and 7) rats and examined their gene expression profiles immediately after isolation using in-house DNA microarray containing ~6,000 known rat genes and ~4,000 ESTs. Selected genes were verified by real time PCR and western blot. The global gene expression data systematically revealed the genes responding to hyperoxia exposure and recovery. Further studies of these genes should shed light on the mechanisms of alveolar epithelial cell injury and repair, including AEC proliferation and differentiation.

5.3 MATERIALS AND METHODS

5.3.1 Hyperoxia model

Hyperoxia model for the present study was used previously in our laboratory for the investigation of protein nitration (16) and the establishment of the method for isolating a highly pure AEC II preparation from injured lungs (17). The Animal Care and Use Committee at the Oklahoma State University approved the animal protocols used in this study. Male Sprague-Dawley rats (~ 250 gram, Charles River Laboratories, Inc., Wilmington, MA) were housed for 1 week before experiments. Rats were exposed to >95% oxygen in a sealed Plexiglas chamber for 48 hours and then allowed to recover in normal room air for 1-7 days. The control rats were exposed to room air. The oxygen concentration was continuously monitored using an oxygen sensor (Vacu-Med, Ventra, CA) and the flow rate was maintained at 8 L/min using a flow meter. Rats had free access to food and water. The CO_2 in the chamber was removed by sodalime.

5.3.2 AEC II isolation

AEC II were isolated from control, hyperoxia-exposed and recovering

rats according to our recently improved method (Chen et al., 2004b). The purities of AEC II preparations were in the range of 93-97% as determined by the modified Papanicolaou staining. The yields were 20-40 x 10^6 per rat with a peak on Day 1 (40.00 ± 4.08 x 10^6). The viability of these cells was over 92%. Four independent AEC II preparations from each time point [control (A2), 48 hours-exposed (D0), and recovering for 1, 3, 5, and 7 days (D1, D3, D5, and D7)], were used for RNA isolation and microarray hybridization.

5.3.3 RNA isolation

Total RNA was extracted using TRI reagents (Molecular Research Center, Cincinnati, OH). RNA quality and quantity was assessed with agarose gel electrophoresis and A_{280}/A_{260} ratio with spectrophotometer (NanoDrop Technologies, Inc, Rockland, DE). The A_{280}/A_{260} ratios for all RNA preparations were >1.8.

5.3.4 Microarray Printing

The DNA microarray slides used in this study were printed in-house on glass slides with 50-mer aminated oligonucleotides, Pan Rat 10K Oligonucleotide Set (MWG Biotech Inc., High Point, NC). It contained 6,221 known rat genes, 3,594 rat ESTs, and 169 Arabidopsis negative controls. The oliogonucleotides were suspended in 3x SSC at 25 µM and printed on epoxy coated slides (CEL Associates, Pearland, Texas) with an OmniGrid 100 arrayer (GeneMachine, San Carlos, CA) per manufacturer's instruction. Each oligonucleotide was spotted in triplicate on three identical 18 x 18 mm blocks: A, B, and C. The spot-spot distance was 180 µm and the space between blocks was 4 mm. After printing, the slides were incubated in 65% humidity overnight at room temperature. The slides were then dried and stored at room

temperature. Printing spots were uniform as assessed by staining slides with SYTO 61 red fluorescent nucleic acid stain dye.

5.3.5　Microarray hybridization

The 2-step microarray hybridization was carried out with the 3DNA 50 Expression kit (Genisphere Inc., Hatfield, PA). Prior to hybridization, the slides were washed with 0.2% SDS once and with deionized water for 4 times, and then dried by centrifugation. The 3 blocks on a slide were separated by two 2 x 25 x 1 mm thermostatic transparent tape strips to allow the hybridizations of 6 independent dual-color labeled biological samples. The strips were removed after hybridization for washing and scanning slides. Total RNA from each AEC II sample (A2, D0, D1, D3, D5 and D7) was split into two aliquots; 5 µg each for reverse-transcription with Cy3- or Alexa 647-specific primers. cDNAs were purified with Microcom YM-30 columns (Millipore, Billerica, MA), adjusted to 0.5-0.6 µg/µl, and mixed with 2x formamide hybridization buffer (50% formamide, 6x SSC, and 0.2% SDS) in an equal volume. The DNA microarray slides were hybridized with paired cDNAs at 42°C for 48 hours. After being washed, the slides were re-hybridized with Cy3- and Alexa 647-specific capture reagents at 42°C for 2 hours. After the 2-step hybridizations, each of the hybridized slides was scanned twice with a laser confocal scanner, *ScanArray Express* (PerkinElmer Life and Analytical Sciences. Boston, MA), one for the spot quantification with 90% laser power and 70~80% PMT so that about 5% spots were saturated, and another one for the spot alignment with 90% laser power and 95% PMT.

5.3.6　Microarray data analysis

Raw data were extracted from the DNA microarray hybridization

images with *GenePix* Pro 4 (Axon Instruments, Inc. Union City, CA). Spot image visualization, spot quality evaluation, data normalization, and one-way ANOVA for the identification of the differentially expressed genes were performed as previously described using the software package *RealSpot* developed in our laboratory *(18)*. The data set from *RealSpot* contained LOWESS normalized $\log_2$ ratios (D0/A2, D1/D0, D3/D1, D5/D3, D7/D5 and A2/D7), globally normalized fluorescence intensity, quality index (QI), and one-way ANOVA p-value ($p < 0.05$, for all the statistical analyses in the present study). QI was calculated based on signal intensity and signal-to-background ratio (18). A QI of 0, 1, 2, 3, 4, and 5 represents negative, weak, middle, strong, saturated and bad spots, respectively.

The differentially expressed genes among different time points during hyperoxia exposure and recovery were identified based on both fluorescence intensities and normalized $\log_2$ ratios. First, we filtered low quality spots with a mean QI of ≤1.0 (negative spot). Second, a statistical test was performed on fluorescence intensities using one-way ANOVA. The genes with a p-value of >0.05 (no significant changes between any time points) were excluded. Third, the differentially expressed genes between any two adjacent time points were identified by t-test using $\log_2$ ratios. For instance, a gene was differentially expressed between day 0 and 1, if $\log_2$ (D1/D0) was significantly different from 0. Fourth, the remaining genes were further filtered by a fold change of >2 to reduce the false-positive and to facilitate real-time PCR verification. The choice of 2-fold change as a cut-off value was arbitrary.

Finally, the genes from step 4 were classified into different gene expression patterns using a trajectory clustering-like algorithm (19), in which

117

the gene expression directions, up-, down- and flat-regulation, were used to cluster in iterative steps. The mean $\log_2$ ratios of any two time points were replaced, based on the above p-values ($p<0.05$) and $\log_2$ ratios, by a symbol "^" ($\log_2$ ratio>1, up-regulation), "v" ($\log_2$ ratio<-1, down-regulation), or "~" (no significant change). By grouping 6 symbols of $\log_2$ ratios (D0/A2, D1/D0, D3/D1, D5/D3, D7/D5 and A2/D7), the differentially expressed genes were classified into several unique gene expression patterns. For instance, a pattern, "~^~~~~", means D0=A2, D1>D0, D3=D1, D5=D3, D7=D5 and A2=D7. In other words, the genes in this pattern were expressed significantly higher on day1 than on day 0.

The functional categories of each pattern were based on gene ontology annotation from Rat Genome Database gene association file (RGD, http://rgd.mcw.edu), gene ontology definitions (GO, http://www.goontology.org), and GeneRIF (http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=gene& cmd=Retrieve&dopt=Graphics).

5.3.7   Real-time RT-PCR

The validation of selected targeted genes with QuantiTect SYBR Green PCR Kit (Qiagen Inc., Valencia, CA) was performed as previously described (Chen, Jin, Narasaraju, Chen, McFarland, Scott, & Liu, 2004d). Total RNA (1 µg each) were reverse-transcribed into cDNAs with 0.2 µg/µl dT$_{17}$, 0.3 µg/µl random hexamer primer, and 200 U MMLV reverse transcriptase (Invitrogen Inc., Carlsbad, CA). The real-time PCR thermal conditions for all genes were 95°C for 15 min, followed by 40 cycles each at 95°C for 30 sec, 60°C for 30 sec, 72°C for 30 sec and 77°C for 35 sec (signal collection temperature). The primer pairs are listed in Table 5.1. The relative

real-time PCR quantification was based on the delta delta Ct method (20). The endogenous reference gene was beta actin (actb), which was not differentially expressed between any time points during hyperoxia exposure and recovery. One-way ANOVA tests were performed to compare the mRNA levels statistically among the samples from the different time points.

Table 5.1: Real-time PCR primers

Gene name	Sequence (forward and reverse)
Heat shock 10 kd protein 1 (hspe 1)	AAG GAG TGC CGC TGA AAC TGT GCT GAA TCT CTC CAC CCT TTC C
Ccaat binding transcription factor cbf subunit c (nfyc)	TCC AGA CCC TTG CTA CCA ATG GCT GGT TGG CTG ACT GAA TAA A
Mesenchyme homeobox 2 (meox2)	CCC GAG AAA AGG AAC TGG TGA A TAT CGC GAC TGT CGT CAT TTG C
Stearyl-coa desaturase (scd1)	GGA GGA CAC GCT GAA ACT CTC A GGA AAA CCT CTG CCA TGC ATC T
Testis-specific farnesyl pyrophosphate synthetase (fdps)	GCA GAC TCT CGA CCT CAT CAC A CCC ATC AAT TCC AGC CAT G
Resistin-like molecule alpha (retnla)	GCA ACT GCC TGT GCT TAA TCG T TGC GCA TGA GTC AGA TTT CCA
Thiolase aa 1-397 (acaa2)	AGA ATC ACC GCA CAC CTG GTT AAT CCC TTC AGG CTG TGT TCT G
Ferritin light chain subunit (ftl1)	GTC TTC GCG GTT AGC TCC ATA CAC GTC ATC CCG ATC AAA AAA G
Preprocathepsin d (ctsd)	TGA AGA ATG GCA CAT CCT TCG A TGG CTT CCC CAA AGA TCT GTT
Metallothionein 3 (mt3)	TGC ACC TGC TCG GAC AAA T CCC TCT TCG CCT TTG CAA A
Vimentin (vim)	CAG GAT TTC TCT GCC TCT TCC A CCT GTC CGT CTC TGG TTT CAA C
Delta-like homolog drosophila (dlk1)	CCC AGG CCC TTC TAC ATT ATC G GCA GCT TGT CAC ACA GCA ACA C
O-acetyltransferase milk fat glolbule membrane protein (mfge8)	CTC AGA GCA ACA GTG CCA AGG A TCC ACT GCA CAC CAT CAT CAC
Rgc32 protein	CCC TGG CAA AAT CGG CTA CTA TTG AAT ATG GTG CAG GCA GGT A
Urinary plasminogen activator receptor 2 (plaur)	TTC AGA GCT TAC CAC CGA ATG G GCC TCC AAG CAC TGA TTC ATT G
Cytosolic peroxisome proliferators -induced acyl-coa thioesterase (bach)	CAA ACC CCG ATT TCC TTT GAG AGT GTG CTC TGG TGT CCA CGA T
Metallothionein 2a (mt2a)	CGC TTC CTA GGT AAG CGC TCT A TCG CAA CAT ATT CCC CAC TCA T
Bone morphogenetic protein 4 (bmp4)	CCC TGG TCA ACT CCG TTA ATT CT CAC CTT GTC GTA CTC GTC CAG AT
Beta actin (actb)	ACC GTG AAA AGA TGA CCC AGA T CAG TGG TAC GAC CAG AGG CAT A
Surfactant protein A (SP-A)	TAA CTA AGT GCT GCC CTC TGA ACT AGG CTT TCC CAG AAA TCC
Surfactant protein C (SP-C)	AGC CCA CCG GAT TAC TCG AC AGG ACT ACC ACC ACA ACC ACG
T1 alpha (T1a)	GGC AGG TGC TAG AAA CTG GTT GCC GTT GAC GAA CCA AGA AC

5.3.8 Western blot

Total protein was obtained from control, hyperoxia-exposed, and recovering AEC II after the TRI reagent extraction (see RNA isolation) according to the manufacturer's instruction. SDS-PAGE and Western blot were performed as previously described (16). The dilutions of antibodies were: goat polyclonal anti-dlk1 (Abcam, Cambridge, MA), 1:100; mouse monoclonal anti-beta actin (Sigma, St. Louis, MO), 1:1000; horseradish peroxidase conjugated (HRP)- mouse anti-goat (Abcam),1:1000; and HPR-rabbit anti-mouse (Jackson ImmunoResearch Laboratories, Inc., West Grove, PA), 1:1000.

5.3.9 Cell culture

An *in vitro cell* culture system for the trans-differentiation of AEC II to AEC I-like cells was used to assess the expression of the identified genes (9). Briefly, the freshly isolated AEC II were re-suspended in the minimal essential medium with 1% fetal bovine serum. AEC II were cultured in $37^{o}C$ and 5% CO_2 in the plastic dishes for 1, 3, 5 and 7 days. At each time point, total RNA was extracted as described above. T1α (an AEC I marker), surfactant proteins A and C (SP-A and SP-C, AEC II markers), and selected genes from the DNA microarray data were monitored by real-time RT-PCR.

5.4   RESULTS

5.4.1 Hyperoxia model

We have previously observed that the exposure of rats to >95% oxygen leads to lung injury and the severity of injury depends on the duration of the exposure time (16). When rats were exposed to >95% for 48 hours,

alveolar epithelium was disrupted but regained its normal structure during the 7 days of recovery. Coinciding with the alveolar injury, rat body weight significantly decreased during the hyperoxia exposure (control, 252.17 ± 3.98 gram, hyperoxia, 240.08 ± 6.22 gram) and increased during the recovery for 1-7 days (Day 1, 257.00 ± 0.43 gram, Day 3, 253.50 ± 1.92 gram, Day 5, 277.50 ± 1.64 gram and Day 7, 286.00 ± 1.28 gram). To identify the genes in AEC II, which are important for hyperoxia response and early recovery, we performed DNA microarray on the isolated AEC II from hyperoxic and recovering rats. AEC II were isolated from control (A2), 48-hour hyperoxia exposed (D0), and recovering for 1, 3, 5, and 7 days (D1, D3, D5, and D7) rats. Even though there were inflammatory cells flooding into the injured lungs, we were able to obtain >93% pure of AEC II for all of the cell preparations using our recently improved method, which removes inflammatory cells such as macrophages and lymphocytes by negative immunomagnetic selection (17).

5.4.2 Differentially expressed genes

We used 4 independent preparations of AEC II in the current study, each with 6 time points (A2, D0, D1, D3, D5, and D7). The samples at 6 time points for each biological replication were labeled with green or red dyes and simultaneously hybridized to 3 different blocks in a green-red pair on one single slide (**Fig.** 5.1A). cDNA samples were paired in a loop-style, each with dye-flip to compensate the dye-specific bias (**Fig.** 5.1B). After hybridization image quantification and data normalization, we sequentially filtered 2,218 genes with a low quality spot (mean QI <1.0, negative spot), 5,405 genes with no significant changes based on fluorescence intensities (one-way ANOVA), 2,005 genes with no significant changes based on $\log_2$ ratios (t-test), and 322

genes with a fold change of <2. (See Materials and Methods for detail on data analysis). The final 50 genes were listed in supplementary Table S1 (66). Twenty-eight genes changed between A2 and D0, or D0 and D1, and 20 genes between D5 and D7, or D7 and A2 (**Fig.** 5.1C).

Table 5.2 Gene ontology of the significantly changed genes in AEC II

during hyperoxia exposure and subsequent recovery

Pattern	Total	A	B	C	D	E	F	G	O
Symbol		v~~~~^	~^~~~~	~v~~~~	^v~~~~	~~~~v~	~~~~^^	~~~~~^	
Meaning		A2>D0,D7	D1>D0	D0>D1	D0>A2,D1	D5>D7	A2>D7>D5	A2>D7	
Cancer	1	1							
Cell cycle	5	2	1	1		1			
Cell differentiation	2				1			1	
Cell proliferation	1		1						
Chaperone	1					1			
Histone	2	2							
Hormone	1		1						
Intermediate Filament	3	1			1			1	
Ion binding	1							1	
Ion channel	2	1				1			
Lamin	1						1		
Metabolism	6		1	1	1	1	1	1	
mRNA export	1							1	
Peptidase	1				1				
Phagocytosis	1			1					
Phosphorylation	3		1				2		
Telomerase	1						1		
Transcription	6	1	1	1		1	1		1
Unclear	11	1	1	2	2	2	1	1	1
Total	50	9	7	6	6	7	7	6	2

The significantly changed genes were identified by two statistical tests (one-way ANVOA and paired t-test) and using a fold change cut-off value of 2. The gene expression pattern was classified based on the significance of $\log_2$ ratios. Symbol meanings, ^, up-regulation (p<0.05, $\log_2$ ratio>1.0), v, down-regulation (p<0.05, $\log_2$ ratio<-1.0), ~, not significantly changed. The symbols were arranged in order of D0/A2, D1/D0, D3/D1, D5/D3, D7/D5 and A2/D7.
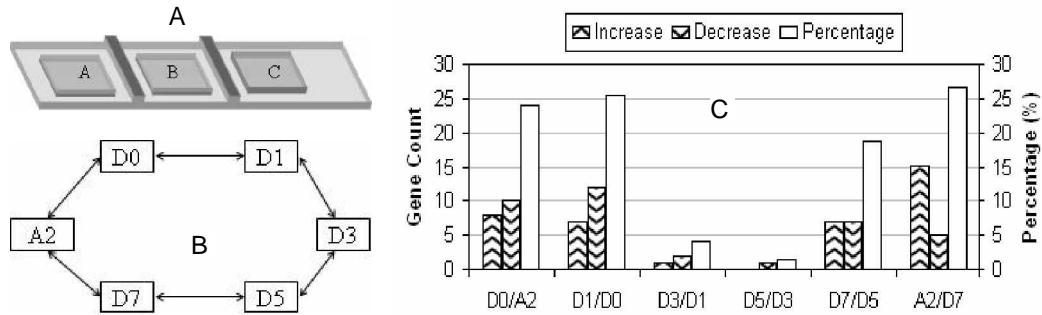
Fig. 5.1 DNA microarray analysis of alveolar epithelial type II cells during hyperoxia exposure and recovery. (A) DNA microarray slide layout. The array has three identical blocks, A, B, and C. Each contains ~ 6,000 known genes and ~ 4,000 ESTs. The three blocks are separated by thermal plastic tapes. Three paired Cy3-(green) and Alexa 647 (red)-labeled cDNA samples were hybridized onto the three blocks. (B) Loop hybridization design. Six AEC II samples were isolated from normal control (A2), 48 hours-hyperoxia exposed (D0), and subsequent recovery in room air for 1, 3, 5, and 7 days (D1, D3, D5, and D7). The samples were paired in a loop style, each with dye-flipping, A2-D0, D0-D1, D1-D3, D3-D5, D5-D7, and D7-A2. There were 4 biological replications (4 independent cell preparations). (C) Differentially expressed genes. Two statistical analyses, a one-way ANVOA and a paired t-test, identified 50 genes that had a fold change of more than 2 for at least two time points. For each ratio, the genes that were >2 (increase) or <2 (decrease) were counted. The percentage of each ratio is the sum of the increase and the decrease count divided by total gene count.

5.4.3 Gene expression patterns during hyperoxia exposure and recovery

In order to elucidate the relationship between gene expression changes and the time points of alveolar injury and repair, the 50 identified genes were classified into 7 unique gene expression patterns, each with 6 or more genes (**Fig.** 5.2A~F) and 2 orphan genes (**Fig.** 5.2O). Gene expression patterns, symbols, and meanings are listed in Table 5.2. In pattern B (~^~~~~), C (~v~~~~), E (~~~~v~), and G (~~~~~^), genes were up- or down-regulated between two conditions. Patterns B and C may be involved in the early

123

recovery after hyperoxia exposure since each gene in these two patterns was expressed significantly different between day 0 and day 1. Similarly, patterns E and G may play a role in the later stages of recovery. This is evident because of the difference in gene expression between day 5 and 7, or day 7 and normal control. The other patterns, A (v~~~~^), D (^v~~~~), and F (~~~~^^), showed more complex gene regulations among three or more conditions, and thus may participate in multiple stages of lung injury and repair induced by hyperoxia exposure and recovery. The genes in pattern A were down-regulated during hyperoxia exposure and up-regulated in late recovery stage. Pattern D showed an expression peak on day 1, implying a role in the early recovery. In pattern F, the genes were up-regulated in the late recovery (day 5 to day 7).

The gene functional categories (ontology) of 50 identified genes are listed in Table 5.2. Notably 11 genes are not functionally characterized. There were several genes related to cell growth and differentiation (cell cycle, differentiation and proliferation, and telomerase), indicating that there was a active alveolar repair during hyperoxia exposure and recovery.
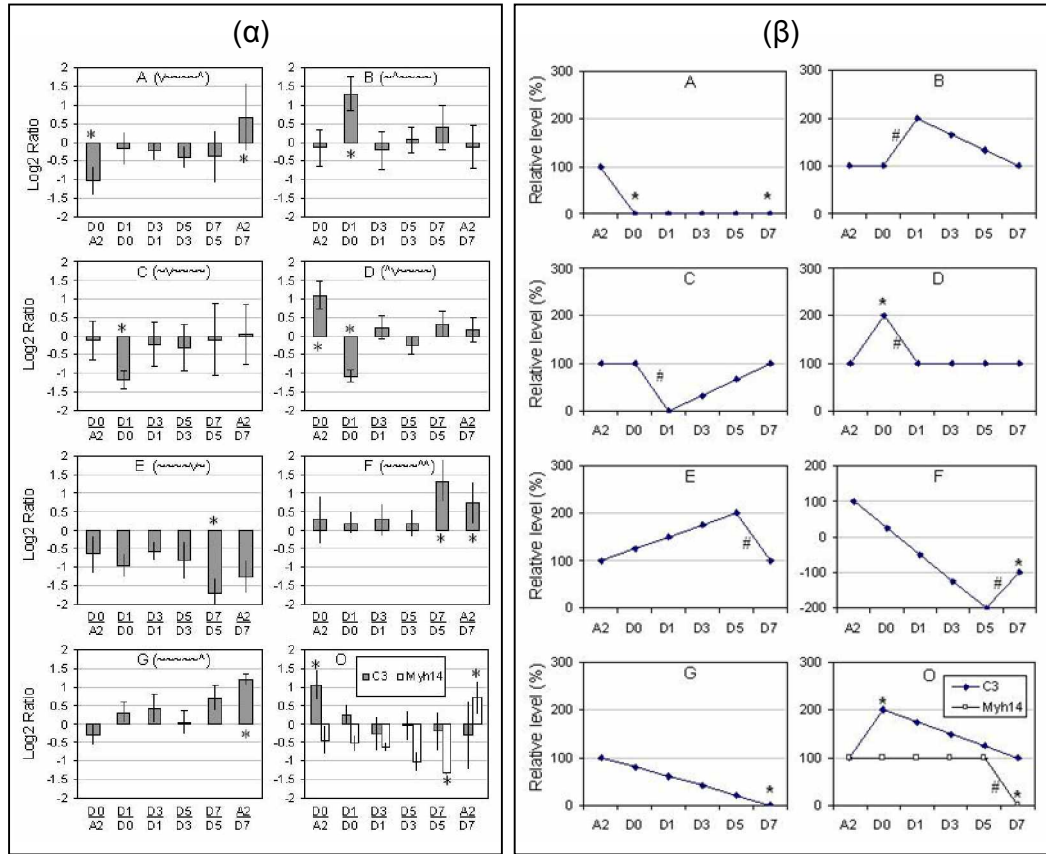
Fig. 5.2 Gene expression patterns in alveolar epithelial type II cells during hyperoxia exposure and recovery. Left panel (α): $\log_2$ ratio changes between time points. Fifty genes, identified by statistical tests and a 2-fold change cut-off value were classified into 7 groups (A-G) and one orphan group (O, 2 genes) based on significant changes at two or more time points. The Y-axis is the $\log_2$ ratio. The values shown are means ± S.D. of all genes in each pattern (gene number: A, 9; B, 7; C, 6; D, 6; E, 7; F, 7; G, 6; and O, 2). A2, normal alveolar type II cells (AEC II), D0-D7: AEC II from 48 hours hyperoxia exposure and subsequent recovery for 0-7 days. *P<0.05, $\log_2$ ratio of each gene in each pattern vs 0. The expression pattern symbol: ~, no significant change; ^, P<0.05, $\log_2$ ratio >2, and v, P<0.05, $\log_2$ ratio <0. **Right panel (β):** diagram of relative mRNA abundance among time points. Y-axes are relative mRNA abundance, and x-axes are time points of hyperoxia exposure and recovery, Meanings of patterns A-O are identical to left panel. * p<0.05: the mRNA D0 or D7 vs A2; # p<0.05, D0 vs D1, or D5 vs D7 (the respective line between them has a -45 or 45-degree slope).

125
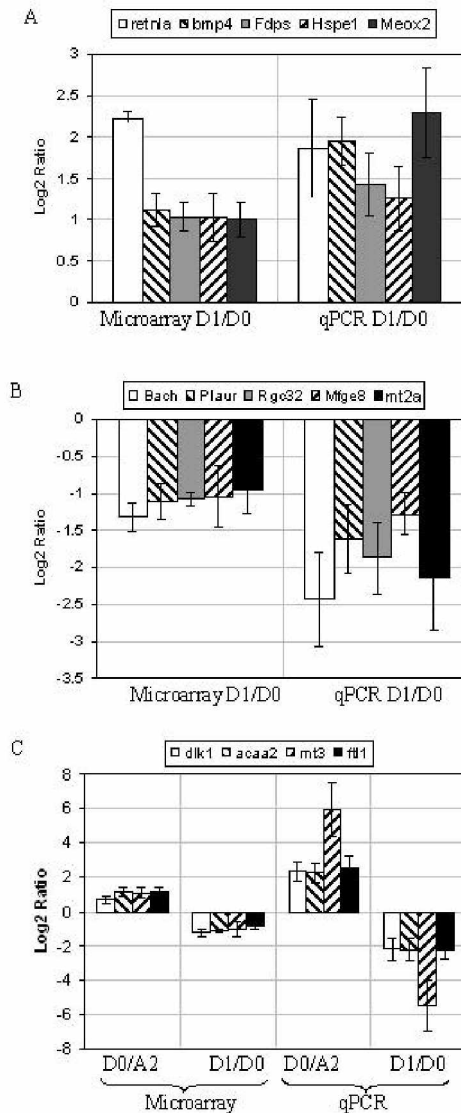
Fig. 5.3 Comparisons of $\log_2$ ratios from DNA microarray and real-time PCR. The genes from pattern B (Panel A), Pattern C (Panel B) and Pattern D (Panel C) were selected for verification by quantitative real time PCR (qPCR). The microarray data was LOWESS normalized and real time PCR data was normalized to β-actin. Error bars represent standard errors. All of the $\log_2$ ratios were significantly different from 0 ($p < 0.05$). There were 4 biological replications for both microarray and real time PCR analyses. A2: normal alveolar type II cells (AEC II); D0 and D1; 48 hours hyperoxia exposure and recovery for 0 and 1 day.

5.4.4 Data validation

Next, we selected some genes for verification with real-time PCR. Because of our research interests in lung injury and repair, in particular, cell proliferation and differentiation, we focused on the genes that change during the early recovery stages between day 0 and 1 (Patterns B, C, and D). Consequently, 19 genes in patterns B, C and D were chosen. Real-time PCR data indicated that 14 genes were consistent with DNA microarray data (**Fig.**

5.3). The known functions of the 14 genes were listed in Table 5.3. In most cases, the real-time PCR resulted in a higher fold change than microarray. This was probably because the real time PCR calculations were based on the delta delta Ct method and also because of the differences between solid-state (microarray) and solution hybridization (PCR) (20). Both microarray and real-time PCR revealed that 5 genes (retnla, bmp4, fdps, hspe1, and meox2) were expressed significantly higher on day 1 than day 0 in pattern B (**Fig.** 5.3A). In patterns C and D, 9 genes (dlk1, acaa2, mt3, ftl1, bach, plaur, rgc32, mfge 8 and mt2a) were consistently expressed higher on day 0 than day 1 in both real-time PCR and microarray assays (**Fig.** 5.3B & C).

The five genes that did not show consistent results may be due to low mRNA abundance, poor amplification (featured by high Ct), and high variation among samples or difference in data normalization (LOWESS vs housekeeping gene). Scd1 was not significantly changed from day 0 to day 1 by real-time RT-PCR, but microarray assay indicated that scd1 was expressed significantly higher on day 1 than on day 0 (log_2 ratio of D1/D0=1.59 ± 0.32). Nfyc and smad5 showed the opposite log_2 ratio of D1/D0 from microarray (1.1±0.28 and -1.6±0.44) and real-time PCR (-3.4±0.95 and 1.66±0.38), respectively. Vim and ctsd showed consistent log_2 ratios of D1/D0 by microarray (-1.2±0.19 and -1.0±0.18) and real-time PCR (-2.1±0.66 and -4.0±1.32), but inconsistent log 2 ratios of A2/D0 by microarray (1.73±0.56 and 0.69±0.12) and real time PCR (-3.7±0.85 and -1.4±0.38).

Table 5.3: The known functions of 14 verified genes

Genbank	Gene name	Function
	Pattern B (~^~~~~): Day1 > Day 0	
NM_053333	Resistin-like molecule alpha (retnla)	May act as a hormone; lung inflammation
NM_012827	Bone morphogenetic protein 4 (bmp4)	Plays a role in induction of cell proliferation
NM_031840	Testis-specific farnesyl pyrophosphate synthetase (fdps)	Isoprene biosynthesis; cell membrane synthesis
NM_012966	Heat shock 10 kd protein 1 (hspe1)	Stress-induced mitochondrial matrix protein and molecular chaperone
NM_017149	Mesenchyme homeobox 2 (meox2)	Homeobox gene; may have a regulatory function in the G0-to-G1 transition of the cell cycle in vascular smooth muscle cells
	Pattern C (~v~~~~): Day 0 > Day 1	
NM_013214	Cytosolic peroxisome proliferator-induced acyl-coa thioesterase (bach)	Involved in fatty acyl-CoA oxidation
NM_017350	Urinary plasminogen activator receptor 2 (plaur)	Involved in plasminogen activation
NM_054008	Rgc32 protein (rgc32)	Increased expression is detected in oligodendrocytes in response to complement activation; may play a role in cell cycle activation
NM_012811	O-acetyltransferase milk fat glolbule membrane protein (mfge8)	Involved in the o-acetylation of gd3 ganglioside sialic acid
M11794	metallothionein 2a (mt2a)	Neuroprotective proteins, may provide a cellular defense strategy in response to DNA damage
	Pattern D (^v~~~~): Day 0 > AEC II and Day 1	
NM_053744	Delta-like homolog drosophila (dlk1)	Involved in maintaining the undifferentiated status of the preadipocyte
X05341	Thiolase aa 1-397 (acca2)	Mitochondrial matrix protein with acetyl-coenzyme A acyltransferase 2 activity
NM_053968	Metallothionein 3 (mt3)	May inhibit cell growth in nervous system
NM_022500	Ferritin light chain subunit 1 (ftl1)	Iron storage

* The function descriptions were from GeneRIF

(http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=gene&cmd=Retrieve&dopt=Graphics).

Based on the availability of suitable antibodies, we further confirmed dlk1 at the protein level in hyperoxic and recovering AEC II. Western Blot was performed on A2, D0 and D1. Data were quantified by densitometry and normalized to β actin. The expression of dlk1 at the protein level was significantly higher on day 0 than on A2 or D1, consistent with the change in mRNA expression (**Fig.** 5.4)
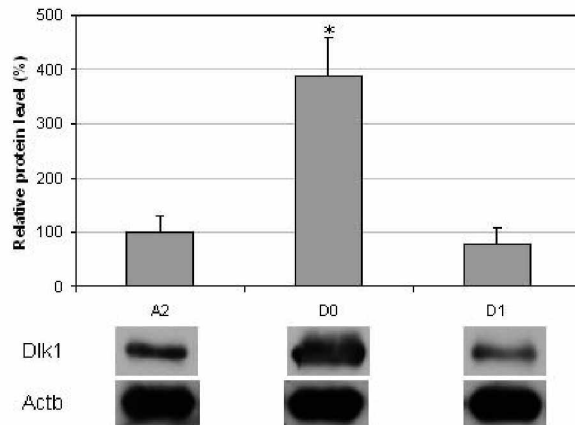
Fig. 5.4 Dlk1 protein expression in hyperoxic and recovering AEC II. The AEC IIs were from control (A2), 48 hours hyperoxia-exposed (D0) and recovering for 1 day (D1) rats. Dlk1 protein level was detected by western blot using anti-Dlk1 antibodies, quantified by densitometry, and normalized to β-actin (Actb). The results were expressed as a percentage of control (A2). Data shown are means ± SE (n=3).
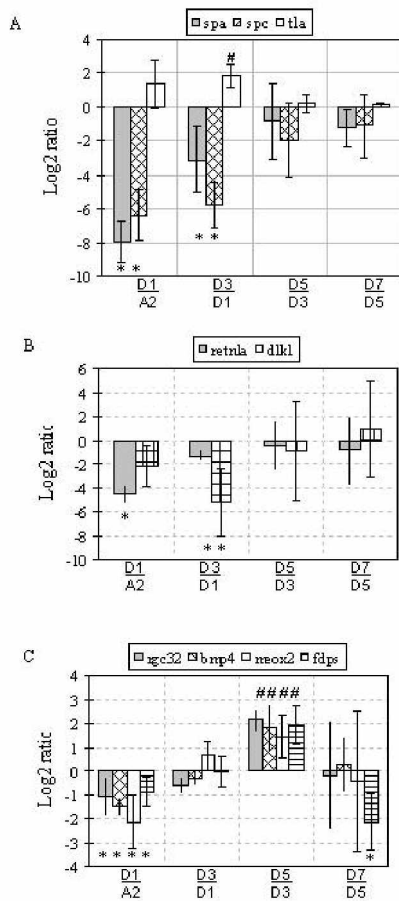
Fig. 5.5 Expression patterns of 6 cell differentiation-related genes in _in vitro_ AEC culture model. Freshly isolated alveolar type II cells (A2) were cultured on plastic dishes for 1, 3, 5, and 7 days (D1, D3, D5, and D7). The mRNA level was measured by quantitative real-time PCR. The results were expressed as $\log_2$ ratios between two time points. The values are mean $\log_2$ ratio, and the error bars are the standard errors. *P<0.05 vs 0; #P<0.05 vs. 0. Data were from 4 Independent cell preparations, each measured in duplicate. (A) SP-A, SP-C and T1α; (B) retnla and dlk1; (C) rgc32, bmp4, mexo2 and fdps.

5.4.5 Gene expression during *in vitro* trans-differentiation of AEC

It is well known that AEC II gradually trans-differentiate to AEC I-like cells when cultured on plastic dishes *in vitro* (21). Using this model, we further studied 6 genes that may be involved in cell proliferation and differentiation. AEC II (A2) isolated from normal rats was cultured with 1% FBS for 1, 3, 5 and 7 days (D1, D3, D5 and D7). The morphology of AEC II (round) gradually changed to AEC I-like cells (flat) (data not shown). The mRNA of T1α, an AEC I marker, was significantly higher on day 3 than on day 1 during *in vitro* culture (**Fig.** 5.5A). In contrast, SP-A and SP-C (AEC II markers) were markedly down-regulated in freshly isolated AEC II compared to day 1 and day 3, indicating the phenotypic change of AEC II to AEC I-like cells at day 3. Dlk1 and retnla decreased significantly in freshly isolated AEC II after 1 or 3 day culture (**Fig.** 5.5B). Rgc-32, bmp4, meox2 and fdps were expressed lower on day 1 than freshly isolated AEC II. However, the expression of these genes increased on day 5 in comparison to day 3 (**Fig.** 5.5C). A simplified gene expression scheme of the 6 genes was drawn based on the results from models of hyperoxia and cell culture (**Fig**. 5.6). Five out of the 6 genes were down-regulated during cell culture from AEC II to day 1. Among other genes identified above, mt2a has a change similar to dlk1 and retnla. Plaur, ftl1 and bach did not exhibit any significant changes at any time points. Mfg8 and acaa2 were expressed significantly higher on day 3 than on day 1. Hspe1 and mt3 were expressed lower on day 3 than on day 1 (data not shown).
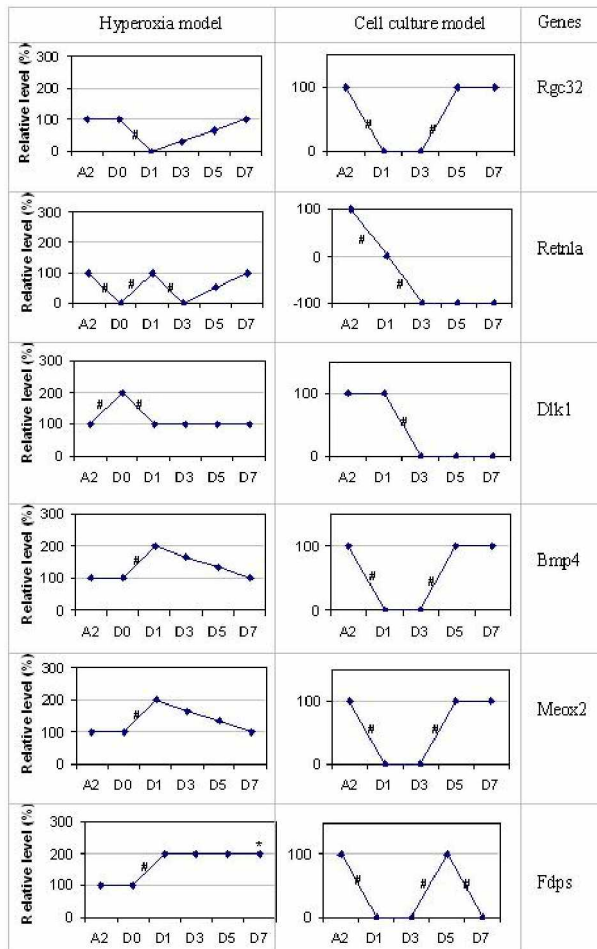
Fig. 5.6 Gene expression diagrams of 6 cell differentiation-related genes in the models of in hyperoxia (left panel, from microarray data) and AEC culture (right panel, from real-time PCR). Y-axes are relative mRNA abundance, and x-axes are time points of hyperoxia exposure and recovery, or cell culture.* $p < 0.05$, A2 vs. D7, # $p < 0.05$, between two adjacent time points (the respective line between them has a -45 or 45-degree slope).

5.5   DISCUSSION

The gene expression profiles of freshly isolated AEC II from normal, hyperoxic, and recovering rats were investigated using DNA microarray. We identified and verified 14 genes, which changed significantly during hyperoxia exposure or early recovery stages. These genes may be involved in AEC protection, apoptosis, and cell proliferation and differentiation. The mRNA expression of the six cell differentiation-related genes was determined in an *in vitro* trans-differentiation culture model of AEC. Common events of AEC II proliferation and differentiation between the models of hyperoxia exposure and cell culture were proposed (**Fig.** 5.7). The information from this study may

serve as a starting point for further investigation of alveolar injury and repair.
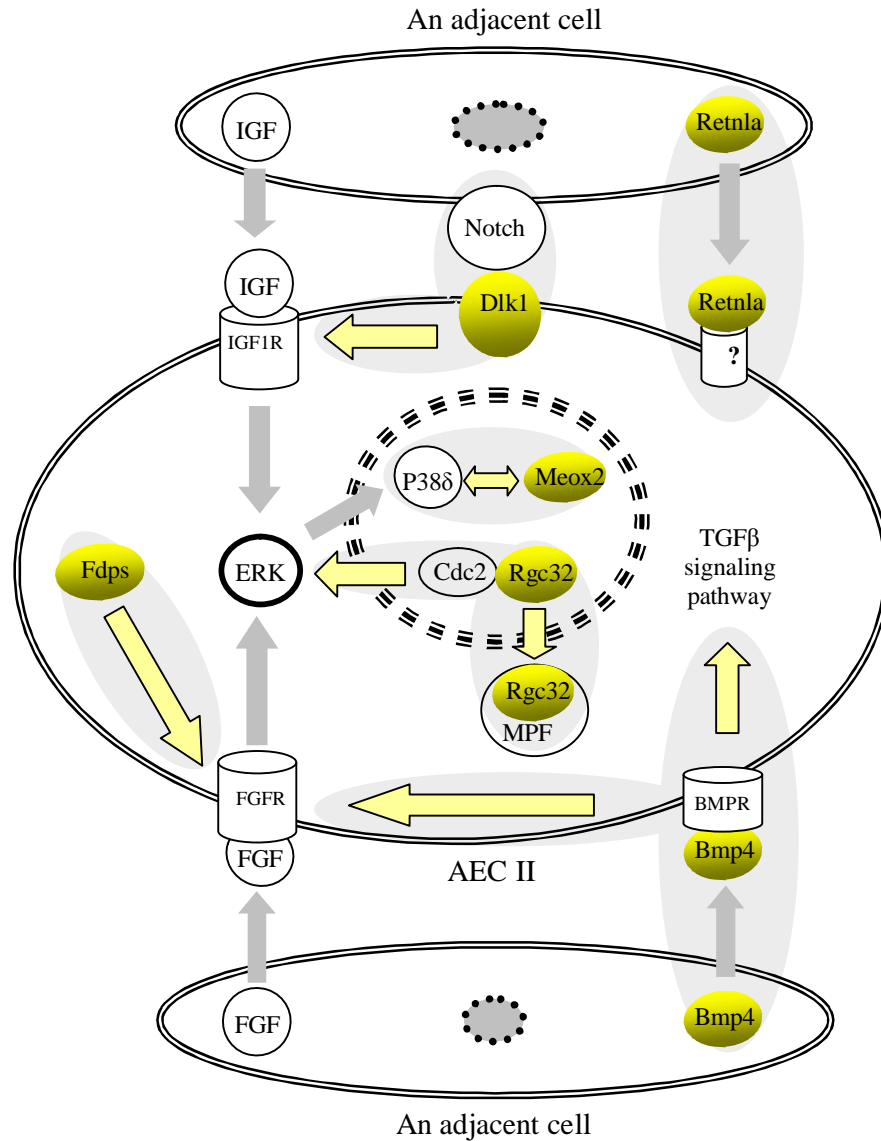
Fig. 5.7 Functional scheme of the 6 genes during AEC II proliferation and differentiation.

The middle double line oval represents an AEC II during proliferation and differentiation, and the other two smaller double line ovals are adjacent cells *in vivo* or *in vitro*. The gray circles inside cells are nuclei. Round text boxes represent the proteins. The 6 genes are highlighted. Genes in cylinders are plasma membrane receptors. Highlighted arrows indicate protein-protein interactions related to 6 genes, and gray arrows indicate other protein-protein interactions. Signaling pathways or interactions are showed in gray ovals. Main abbreviations: MPF, M-phase promoting factor; ERK, extracellular receptor kinase.

Hyperoxia injury and repair is a complex model composed of various biological processes, including hyperoxia-mediated lung injury during the exposure and subsequent lung tissue repair (cell proliferation, differentiation and apoptosis) during the recovery. Since these events are sequential, we used a loop design to compare adjacent time points in order to better probe gene expression in a particular event such as cell proliferation and differentiation. This method has several advantages over a reference design. In a reference design, each time point needs to hybridize to a common reference (i.e. control, or D0, or pooled samples). Since any two time points must be indirectly compared through a common reference, the reference design results in accumulated errors from two hybridizations (i.e. each time point vs reference). Also the reference design has a lower dynamic range. In the case of pooled samples as a reference, assuming that the expression level of a gene in A2 is 100 and that of any of other 5 samples (D0, D1, D3, D5, and D7) is 1, the ratio of A2 vs reference is 100/(105/6) = 5.7. This is much lower than the ratio of A2 vs. another sample (100/1 = 100) in the loop design.

In the current study, there were two statistical tests based on fluorescence intensities and $\log_2$ ratios, which eliminated type I (false positive) and type II (false negative) errors resulting from the multiple comparisons of thousands of genes. For instance, there were about 389 false positive genes (5% of 7,782) after one-way ANOVA of fluorescence intensities. The false positive genes were reduced to about 19 (5% of 389) after t-test of $\log_2$ ratios. Thus, the two-step tests greatly improved the reliability of the genes identified from DNA microarray data. With an alternative one-step approach, Bonferroni correction, no significantly differential gene expression was detected because

of the p-value adjustment to 6.43×10^{-6} (=0.05/7782). This correction insured that the global type I error was 0.05, but increased the type II error and excluded many truly differentially expressed genes.

AEC protection   Functionally, 6 genes may be involved in AEC protection during hyperoxia exposure and recovery, including thiolase aa 1-397 (acaa2), cytosolic peroxisome proliferator-induced acyl-coa thioesterase (bach), metallothionein 2a and 3 (mt2a and mt3), plasminogen activator receptor 2 (plaur), and resistin-like molecule alpha (retnla). These genes participate in fatty acid oxidation, growth inhibition, host defense, and lung inflammation. Acaa2 is a mitochondrial acetyl-CoA acyltransferase (22). Bach is a cytosolic acyl-CoA thioesterase and its activity is induced by peroxisome proliferators (23). In context of the rat hyperoxia model, these two genes in the fatty acid beta oxidation pathway may respond to a high concentration of oxygen by consuming oxygen, since both genes showed a higher expression during the hyperoxia exposure. They may have a protective effect on AEC during the exposure stage. Plaur regulates cell-surface plasminogen activation in normal and pathological processes and is critical for pericellular proteolysis. It plays a role in lymphocyte recruitment to the lung (24).   It also has an anti-bacterial effect in the lungs of plaur -/- mice (25). Its augmented expression during hyperoxia exposure indicates its potential role in the protection of AEC against alveolar injury. TNF-α up-regulates the plaur expression in the AEC II cell line, A549. These results support that plaur may be involved in inflammatory response by recruiting the immune cells to injured AEC and activating the subsequent cellular immune response (4). Two highly homological metallothionein genes, mt2a and mt3 show changes during hyperoxia

exposure and early recovery. This is consistent with a previous study, in which metallothionein increases 7 folds in the whole mice lung tissue after 48 hr hyperoxia exposure (14). Metallothioneins are 6-7 kDa metal-binding antioxidant proteins. They are induced by oxidative stress (14; 26; 27) and protect the lung from acute injury (28). Mt3 also inhibits the neuron growth in the neonatal rats (29). Retnla, also known as found in inflammatory zone 1, FIZZ1, is a ~ 9 kDa cysteine-rich secreted protein and has been identified in the bronchoalveolar lavage fluid of the allergen-induced inflamed lungs (30). Its expression is markedly increased in the bronchial epithelial cells and alveolar epithelial type II cells in inflamed lungs, likely by cytokines, IL-4 and IL-13 via STAT6 and c/EBP (31).

AEC apoptosis O-acetyltransferase milk fat globule membrane protein (mfge8) and heat shock 10 kd protein 1 (hspe1) may regulate AEC apoptosis during hyperoxia exposure and subsequent recovery phases. A prolonged hyperoxia exposure is known to damage mitochondria, resulting in AEC apoptosis. Mfge8 is a secreted glycoprotein that bridges apoptotic cells to phagocytes for their removal. It has two EGF-like domains, a mucin-like domain, two C-domains for phospholipid binding and a RGD motif for $\alpha_v\beta_3$ integrin binding (32). Mfge8 binds to apoptotic cells via phosphatidylserine and increases the engulfment of apoptotic cells by phagocytes through $\alpha_v\beta_3$ integrin (33; 34). Hspe1 (also known as hsp10) is a molecular chaperone and plays an anti-apoptotic role by the formation of a complex composed of pro-caspase-3, hsp10 and hsp60 (35). It also inhibits the ubiquitination of the Bcl-xl protein (36). Our current DNA microarray data shows that only a few apoptosis-related genes are significantly changed during hyperoxia exposure

and recovery stages. For instance, a medium to high mRNA abundance of bcl-xl and bax in normal, hyperoxic and recovery AEC II are observed, but no significant changes are evident (data not shown). On the other hand, fas is expressed significantly higher in normal AEC II than in hyperoxic or recovering AEC II (data not shown). A similar result from the microarray analysis of whole mouse lung tissue has been obtained except that Bcl-xl is increased at 24 hr of hyperoxia exposure (14). Clearly, apoptosis is required to resolve the increased number of AEC II during the recovery phase in order to repair alveolar epithelial injury. In addition to mRNA changes, apoptosis-related proteins in AEC may also be regulated at protein translational or post-translational modification levels.

Cell proliferation and differentiation Six genes, delta-like homolog drosophila (dlk1), bone morphogenetic protein 4 (bmp4), mesenchyme homeobox 2 (meox2), testis-specific farnesyl pyrophosphate synthetase (fdps), response gene to complement (rgc-32), and resistin like kinase alpha (retnla) were relevant to cell proliferation and differentiation. Their functional roles were summarized in **Fig**. 5.7. The 6 genes except retnla may modulate MAPK signaling pathway during AEC II proliferation and differentiation. Bmp4 protein may regulate AEC II proliferation and differentiation through TGFβ signaling pathway.

Dlk1 (also known as preadipocyte factor 1, Pref-1) was a member of the notch/delta/serrate family and contains 6 EGF-like repeats and a transmembrane domain. It functioned in cell fate determination and cell differentiation in various tissues and cells. For instance, Dlk1 inhibited the differentiation of adipocytes (38), bone marrow stromal cells (39), human

hematopoietic cells (40), adrenal gland cells (41), and neuroendocrine cells (42). Dlk1 modulated IGF-1 receptor and regulated ERK/MAPK signaling pathway during cell differentiation (43). Dlk1 may play dual roles during AEC II proliferation and differentiation. First, it may interact with notch receptors on the cyctoplasma membrane of other cells (e.g. AEC I, AEC II or macrophage in alveoli, or AEC II cultured in plastic dishes), and regulate cell differentiation through the notch signaling pathway (44). Second, Dlk1 or the notch-dlk1 complex may modulate IGF1 receptor, which in turn affects cell proliferation induced by IGF-ERK-MAPK pathway. Dlk1 was up-regulated during hyperoxia exposure (A2 to D0, **Fig**. 5.6), which may activate a portion of AEC II to cell proliferation and differentiation status. No up-regulation of dlk1 expression during AEC II culture was observed. The Dlk1 showed down-regulation during hyperoxia early recovery stage (day 0 to day 1) and cell culture (day 1 to day 3). Dlk1 repression may decrease the proliferation effect through IGF1R modulation, and shift AEC II to differentiation. The adjacent cells around AEC II may induce cell differentiation *in vivo* or trans-differentiation *in vitro*. In hyperoxia model, AEC, macrophage and endothelial cells may affect AEC II differentiation through Notch-Dlk1 interaction to regenerate normal AEC I and AEC II for alveoli repair. On the other hand, the relatively uniform AEC II in plastic dishes induced AEC I-like cells. These results indicated that cell-cell interactions may play a role in the formation of normal AEC I and AEC II during cell differentiation.

Bmp4, a member of the transforming growth factor-β superfamily, was extensively studied during fetal lung development (45; 46). It was expressed high in the distal tips of the fetal epithelium, but low in the adjacent

mesenchyme. Over-expression of bmp4 in the lung distal epithelium led to the inhibition of cell proliferation and epithelial hypoplasis (47). In the embryonic epithelium culture, the addition of bmp4 reduced epithelial growth and differentiation (48). However, in the intact lung explant organ culture, the inclusion of bmp4 in culture medium was required for morphogenesis branching and cell proliferation. A recent study demonstrated that Bmp4 up-regulated FGF7/10 receptor, and thus acted upstream of FGF signaling pathway (49). However, another study showed that FGF7 (KGF)-MAPK and TGFβ signaling pathways were antagonized during AEC II proliferation and differentiation *in vivo* and *in vitro* (50). In **Fig**. 5.7, Bmp4 plays a role for the crosstalk between the two pathways. Up-regulations of Bmp4 expression were detected in hyperoxia model (day 0 to day 1) and in cell culture model (day 3 to day 5, coming back from decreased level). During the up-regulation phases, KGF-MAPK signaling may be weakened by TGFβ signaling, which should promote AEC II switching to differentiation from proliferation. There was a Bmp4 repression in cell culture model (AEC II to day 1), which was absent in hyperoxia model. The meaning of Bmp4 repression is not clear from current data.

Meox2 is a transcription factor of the homeobox family with a helix-turn-helix motif, regulating cell proliferation, differentiation and migration. Meox2 was also known as growth arrest-specific homeobox (Gax) and played a role in the G0/G1 transition when quiescent cells reentered the cell cycle (51). Meox2 was down-regulated in vascular smooth muscle cells following mitogen activation *in vitro* and *in vivo* (52; 53). It inhibited cell proliferation in a p21$^{WAF/CIP}$-dependent manner (54). Over-expression of meox2 also inhibited

cardiomyocyte proliferation (55). The deletion of meox2 led to defective differentiation and morphogenesis of limb muscles (56). In the placenta, meox2 may regulate the interaction of epithelial and mesenchymal cells (57). Meox2 may have an inhibitory effect on the G0/G1 transition of the cell cycle, and was down-regulated by ERK1/2-MAPK-p38δ activation in vascular smooth muscle cells (51). The ERK1/2-MAPK-p38δ pathway can be activated by growth factors like FGF7 (KGF) and IGF in AEC II. Therefore, the up-regulation of Meox2 expression from day 0 to day 1 after hyperoxia exposure may indicate the deactivation of ERK-MAPK signaling pathway induced by FGF or IGF. With the up-regulation of Meox2, AEC II enter G1 phase of cell cycle during recovery stages. In cell culture model, Meox2 expression was different from hyperoxia model. It was down-regulated from day 0 to day 1, and up-regulated from day 3 to day 5. The meaning of Meox2 expression changes during cell culture is not clear from the present results.

Fdps is an enzyme in the isoprene biosynthetic pathway for the synthesis of the cholesterol and nonsterol metabolites in the peroxisome (58). Fdps played an important role in colon cancer proliferation (59), and prolonged the activation of FGF-ERK-MAPK signaling by binding to FGF receptor (60). It has dual roles in cell proliferation and differentiation by controlling lipid synthesis and stabilizing FGF receptors. In hyperoxia recovery stage (day 0 to day 1), increasing expression of Fdps may enhance AEC II proliferation through FGF induced signaling pathways. Its expression levels were more complicated in cell culture than in hyperoxia model. During cell attachment on plastic dishes, its decreasing expression may repress cellular lipid synthesis. From day 3 to day 5, the increasing Fdps expression may enhance lipid synthesis. The

down-regulation of Fdps from day 5 to day 7 may be resulted from the confluence of cell in plastic dishes, in which cell should not proliferate or synthesize lipid for plasma membrane.

Rgc32 promotes cell proliferation by translocating to nuclei and stimulating quiescent cells to enter into S-phase in aortic smooth muscle cells upon complement activation (61; 62). In cytosol, it is a regulatory subunit of M-phase promoting factor (MPF) for the G2/M phase transition. In nucleus, it directly interacts with cyclin—dependent kinase p34^{cdc2} and is essential for G1/S phase transition. Cdc2 inhibited ERK-MAPK signaling pathway during mitosis through modulation of MEK1 (63). The down-regulation of Rgc32 in the models of hyperoxia (day 0 to day 1) and cell culture (AEC II to day 1) may reduce the inhibitory effect of cdc2 on cell proliferation through ERK-MAPK pathways. Later, Rgc32 may have an inhibitory effect on AEC II proliferation as it was up-regulated in both models (day 7 to normal AEC II level in hyperoxia, and day 3 to day 5 in cell culture), possibly through Rgc32-cdc2 complex.

The above 5 genes appear to modulate MAPK signaling during AEC II proliferation and differentiation. Dlk1 regulates IGF1 receptor, Bmp4 and Fdps affect FGF receptor, and Rgc32 and Meox2 regulate and are regulated by ERK1/2-MAPK signaling pathways, respectively. Since FGF7 (KGF) and IGF have been well studied in AEC II, our results on the 5 genes indicate that there are fine-tuning modulations on MAPK signaling pathway during AEC II proliferation and differentiation.

The last gene, retnla, is a cytokine and may be involved in AEC II proliferation and differentiation, but its cellular receptor or the mechanism is unclear. In adipose tissue, retnla had an inhibitory effect on the differentiation

of 3T3-L1 preadipocytes to adipocytes (64). Retnla in AEC II was also induced in the bleomycin-induced pulmonary fibrosis model. Using co-cultures of AEC II and fibroblasts, retnla was found to promote myofibroblast differentiation independent of TGF-β (37). Additionally, retnla stimulates the proliferation of pulmonary microvascular smooth muscle cells (65). Retnla was repressed in both models of hyperoxia (A2 to day0, and day 1 to day 3) and cell culture (A2 to day 3), in which AEC II were activated for proliferation. The decreasing level of Retnla may help AEC II enter cell proliferation status. An increase of Retnla during the initial recovery of AEC II after hyperoxia exposure may indicate its role in promoting the differentiation of AEC II to AEC I through cell-cell interaction (Fig. 5.7). No up-regulation of Retnla was detected during cell culture. The possible reason is that the up-regulation of Retnla requires the induction of adjacent AEC I as found in the lung but was absent in culture dishes.

The proposed model in Fig. 5.7 has three features. First, the 6 genes play fine-tuning roles for AEC II proliferation and differentiation. These genes modulate the components of MAPK and TGFβ signaling pathways. The modulation changes the effects of the pathway cascades, and indirectly tunes the status of cell proliferation and differentiation. Second, the genes may have promoting or inhibitory effects on AEC II depending on environment (*in vivo* hyperoxia or *in vitro* cell culture) and time courses (hyperoxia exposure and recovery, or cell isolation, attachment and culture). Consequently, the expression levels of the 6 genes were not identical in both models. Cell culture model always showed a down-regulation of gene expression for the first two days, which may be resulted from cell isolation processes. Third, cell-cell

communication or interaction plays a pivotal role in the proposed model. Signaling molecules (IGF, FGF, Bmp4 and Retnla) secreted from adjacent cells or AEC II itself directly stimulate or inhibit AEC II proliferation and differentiation. Dlk1-Notch may induce AEC II differentiation through a direct cell contact to form dlk1-notch signaling pathway. The adjacent cells may determine AEC II differentiation. In hyperoxia model, the adjacent AEC I, endothelial cell or macrophage may induce AEC II differentiation to a normal AEC I. However, the adjacent AEC II in culture dish induce AEC II trans-differentiation to AEC I-like cells.

In summary, we identified 14 genes for AEC protection, apoptosis and cell differentiation from rat model of hyperoxia exposure and recovery using DNA microarray assays. Further study of these genes in cell culture model indicated that 6 genes may be involved in AEC II proliferation and differentiation. We proposed a model for AEC II differentiation, in which the 6 genes modulate MAPK and TGFβ signaling pathways to fine-tune the balance of AEC II proliferation and differentiation. Future studies on the 6 genes will elucidate a high level understanding of AEC II differentiation through the modulation and integration of MAPK, TGFβ and other signaling pathways.

5.6   ACKNOWLEDGMENTS

5.7 Reference List

1. **O'Reilly MA**. DNA damage and cell cycle checkpoints in hyperoxic lung injury: braking to facilitate repair. *Am J Physiol Lung Cell Mol Physiol* 281: L291-L305, 2001.

2. **Buccellato LJ, Tso M, Akinci OI, Chandel NS and Budinger GR**. Reactive oxygen species are required for hyperoxia-induced Bax activation and cell death in alveolar epithelial cells. *J Biol Chem* 279: 6753-6760, 2004.

3. **Adamson IY and Bowden DH**. The type 2 cell as progenitor of alveolar epithelial regeneration. A cytodynamic study in mice after exposure to oxygen. *Lab Invest* 30: 35-42, 1974.

4. **dos Santos CC, Han B, Andrade CF, Bai X, Uhlig S, Hubmayr R, Tsang M, Lodyga M, Keshavjee S, Slutsky AS and Liu M**. DNA microarray analysis of gene expression in alveolar epithelial cells in response to TNFalpha, LPS, and cyclic stretch. *Physiol Genomics* 19: 331-342, 2004.

5. **Gonzales LW, Guttentag SH, Wade KC, Postle AD and Ballard PL**. Differentiation of human pulmonary type II cells *in vitro* by glucocorticoid plus cAMP. *Am J Physiol Lung Cell Mol Physiol* 283: L940-L951, 2002.

6. **Koike E, Hirano S, Furuyama A and Kobayashi T**. cDNA microarray analysis of rat alveolar epithelial cells following exposure to organic

extract of diesel exhaust particles. *Toxicol Appl Pharmacol* 201: 178-185, 2004.

7. **Mason RJ, Pan T, Edeen KE, Nielsen LD, Zhang F, Longphre M, Eckart MR and Neben S**. Keratinocyte growth factor and the transcription factors C/EBP alpha, C/EBP delta, and SREBP-1c regulate fatty acid synthesis in alveolar type II cells. *J Clin Invest* 112: 244-255, 2003.

8. **Gonzalez R, Yang YH, Griffin C, Allen L, Tigue Z and Dobbs L**. Freshly isolated rat alveolar type I cells, type II cells, and cultured type II cells have distinct molecular phenotypes. *Am J Physiol Lung Cell Mol Physiol* 288: L179-L189, 2005.

9. **Chen Z, Jin N, Narasaraju T, Chen J, McFarland LR, Scott M and Liu L**. Identification of two novel markers for alveolar epithelial type I and II cells. *Biochem Biophys Res Commun* 319: 774-780, 2004.

10. **Bonner AE, Lemon WJ and You M**. Gene expression signatures identify novel regulatory pathways during murine lung development: implications for lung tumorigenesis. *J Med Genet* 40: 408-417, 2003.

11. **Clerch LB, Baras AS, Massaro GD, Hoffman EP and Massaro D**. DNA Microarray Analysis of Neonatal Mouse Lung Connects Regulation of KDR with Dexamethasone-Induced Inhibition of Alveolar Formation. *Am J Physiol Lung Cell Mol Physiol* 286: L411-9, 2003.

12. **Kaplan F, Comber J, Sladek R, Hudson TJ, Muglia LJ, Macrae T, Gagnon S, Asada M, Brewer JA and Sweezey NB**. The growth factor midkine is modulated by both glucocorticoid and retinoid in fetal lung development. *Am J Respir Cell Mol Biol* 28: 33-41, 2003.

13. **Mariani TJ, Reed JJ and Shapiro SD**. Expression profiling of the developing mouse lung: insights into the establishment of the extracellular matrix. *Am J Respir Cell Mol Biol* 26: 541-548, 2002.

14. **Perkowski S, Sun J, Singhal S, Santiago J, Leikauf GD and Albelda SM**. Gene expression profiling of the early pulmonary response to hyperoxia in mice. *Am J Respir Cell Mol Biol* 28: 682-696, 2003.

15. **Hoshikawa Y, Nana-Sinkam P, Moore MD, Sotto-Santiago S, Phang T, Keith RL, Morris KG, Kondo T, Tuder RM, Voelkel NF and Geraci MW**. Hypoxia induces different genes in the lungs of rats compared with mice. *Physiol Genomics* 12: 209-219, 2003.

16. **Narasaraju TA, Jin N, Narendranath CR, Chen Z, Gou D and Liu L**. Protein nitration in rat lungs during hyperoxia exposure: a possible role of myeloperoxidase. *Am J Physiol Lung Cell Mol Physiol* 285: L1037-L1045, 2003.

17. **Chen J, Chen Z, Narasaraju T, Jin N and Liu L**. Isolation of highly pure alveolar epithelial type I and type II cells from rat lungs. *Lab Invest* 84: 727-735, 2004.

18. **Chen Z and Liu L**. RealSpot: Software validating results from DNA microarray data analysis with spot images. *Physiol Genomics* 21: 284-291, 2005.

19. **Rudolph MC, McManaman JL, Hunter L, Phang T and Neville MC**. Functional development of the mammary gland: use of expression profiling and trajectory clustering to reveal changes in gene expression during pregnancy, lactation, and involution. *J Mammary Gland Biol Neoplasia* 8: 287-307, 2003.

20. **Pfaffl MW**. A new mathematical model for relative quantification in real-time RT-PCR. *Nucleic Acids Res* 29: e45, 2001.

21. **Paine R and Simon RH**. Expanding the frontiers of lung biology through the creative use of alveolar epithelial cells in culture. *Am J Physiol* 270: L484-L486, 1996.

22. **Abe H, Ohtake A, Yamamoto S, Satoh Y, Takayanagi M, Amaya Y, Takiguchi M, Sakuraba H, Suzuki Y, Mori M and .** Cloning and sequence analysis of a full length cDNA encoding human mitochondrial 3-oxoacyl-CoA thiolase. *Biochim Biophys Acta* 1216: 304-306, 1993.

23. **Engberg ST, Aoyama T, Alexson SE, Hashimoto T and Svensson LT**. Peroxisome proliferator-induced acyl-CoA thioesterase from rat liver cytosol: molecular cloning and functional expression in Chinese hamster ovary cells. *Biochem J* 323: 525-531, 1997.

24. **Gyetko MR, Sud S, Sonstein J, Polak T, Sud A and Curtis JL**. Antigen-driven lymphocyte recruitment to the lung is diminished in the absence of urokinase-type plasminogen activator (uPA) receptor, but is independent of uPA. *J Immunol* 167: 5539-5542, 2001.

25. **Rijneveld AW, Levi M, Florquin S, Speelman P, Carmeliet P and van der PT**. Urokinase receptor is necessary for adequate host defense against pneumococcal pneumonia. *J Immunol* 168: 3507-3511, 2002.

26. **Levy MA, Tsai YH, Reaume A and Bray TM**. Cellular response of antioxidant metalloproteins in Cu/Zn SOD transgenic mice exposed to hyperoxia. *Am J Physiol Lung Cell Mol Physiol* 281: L172-L182, 2001.

27. **Piedboeuf B, Johnston CJ, Watkins RH, Hudak BB, Lazo JS, Cherian MG and Horowitz S**. Increased expression of tissue inhibitor of metalloproteinases (TIMP-I) and metallothionein in murine lungs after hyperoxic exposure. *Am J Respir Cell Mol Biol* 10: 123-132, 1994.

28. **Takano H, Inoue K, Yanagisawa R, Sato M, Shimada A, Morita T, Sawada M, Nakamura K, Sanbongi C and Yoshikawa T**. Protective role of metallothionein in acute lung injury induced by bacterial endotoxin. *Thorax* 59: 1057-1062, 2004.

29. **Tsuji S, Kobayashi H, Uchida Y, Ihara Y and Miyatake T**. Molecular cloning of human growth inhibitory factor cDNA and its down-regulation in Alzheimer's disease. *EMBO J* 11: 4843-4850, 1992.

30. **Holcomb IN, Kabakoff RC, Chan B, Baker TW, Gurney A, Henzel W, Nelson C, Lowman HB, Wright BD, Skelton NJ, Frantz GD, Tumas DB, Peale FV, Jr., Shelton DL and Hebert CC**. FIZZ1, a novel cysteine-rich secreted protein associated with pulmonary inflammation, defines a new gene family. *EMBO J* 19: 4046-4055, 2000.

31. **Stutz AM, Pickart LA, Trifilieff A, Baumruker T, Prieschl-Strassmayr E and Woisetschlager M**. The Th2 cell cytokines IL-4 and IL-13 regulate found in inflammatory zone 1/resistin-like molecule alpha gene expression by a STAT6 and CCAAT/enhancer-binding protein-dependent mechanism. *J Immunol* 170: 1789-1796, 2003.

32. **Stubbs JD, Lekutis C, Singer KL, Bui A, Yuzuki D, Srinivasan U and Parry G**. cDNA cloning of a mouse mammary epithelial cell surface protein reveals the existence of epidermal growth factor-like domains linked to factor VIII-like sequences. *Proc Natl Acad Sci U S A* 87: 8417-8421, 1990.

33. **Hanayama R, Tanaka M, Miwa K, Shinohara A, Iwamatsu A and Nagata S**. Identification of a factor that links apoptotic cells to phagocytes. *Nature* 417: 182-187, 2002.

34. **Hanayama R, Tanaka M, Miyasaka K, Aozasa K, Koike M, Uchiyama Y and Nagata S**. Autoimmune disease and impaired uptake of apoptotic cells in MFG-E8-deficient mice. *Science* 304: 1147-1150, 2004.

35. **Samali A, Cai J, Zhivotovsky B, Jones DP and Orrenius S**. Presence of a pre-apoptotic complex of pro-caspase-3, Hsp60 and Hsp10 in the mitochondrial fraction of jurkat cells. *EMBO J* 18: 2040-2048, 1999.

36. **Shan YX, Liu TJ, Su HF, Samsamshariat A, Mestril R and Wang PH**. Hsp10 and Hsp60 modulate Bcl-2 family and mitochondria apoptosis signaling induced by doxorubicin in cardiac muscle cells. *J Mol Cell Cardiol* 35: 1135-1143, 2003.

37. **Liu T, Dhanasekaran SM, Jin H, Hu B, Tomlins SA, Chinnaiyan AM and Phan SH**. FIZZ1 stimulation of myofibroblast differentiation. *Am J Pathol* 164: 1315-1326, 2004.

38. **Smas CM and Sul HS**. Pref-1, a protein containing EGF-like repeats, inhibits adipocyte differentiation. *Cell* 73: 725-734, 1993.

39. **Abdallah BM, Jensen CH, Gutierrez G, Leslie RG, Jensen TG and Kassem M**. Regulation of human skeletal stem cells differentiation by Dlk1/Pref-1. *J Bone Miner Res* 19: 841-852, 2004.

40. **Li L, Forman SJ and Bhatia R**. Expression of DLK1 in hematopoietic cells results in inhibition of differentiation and proliferation. *Oncogene* 24: 4472-4476, 2005.

41. **Raza FS, Puddefoot JR and Vinson GP**. Pref-1, SF-1 and adrenocortical zonation. *Endocr Res* 24: 977-981, 1998.

42. **Laborda J, Sausville EA, Hoffman T and Notario V**. dlk, a putative mammalian homeotic gene differentially expressed in small cell lung carcinoma and neuroendocrine tumor cell line. *J Biol Chem* 268: 3817-3820, 1993.

43. **Ruiz-Hidalgo MJ, Gubina E, Tull L, Baladron V and Laborda J**. dlk modulates mitogen-activated protein kinase signaling to allow or prevent differentiation. *Exp Cell Res* 274: 178-188, 2002.

44. **Iso T, Hamamori Y and Kedes L**. Notch signaling in vascular development. *Arterioscler Thromb Vasc Biol* 23: 543-553, 2003.

45. **Shannon JM and Hyatt BA**. Epithelial-mesenchymal interactions in the developing lung. *Annu Rev Physiol* 66: 625-645, 2004.

46. **Warburton D and Bellusci S**. The molecular genetics of lung morphogenesis and injury repair. *Paediatr Respir Rev* 5: S283-S287, 2004.

47. **Bellusci S, Henderson R, Winnier G, Oikawa T and Hogan BL**. Evidence from normal expression and targeted misexpression that bone morphogenetic protein (Bmp-4) plays a role in mouse embryonic lung morphogenesis. *Development* 122: 1693-1702, 1996.

48. **Hyatt BA, Shangguan X and Shannon JM**. BMP4 modulates fibroblast growth factor-mediated induction of proximal and distal lung

differentiation in mouse embryonic tracheal epithelium in mesenchyme-free culture. *Dev Dyn* 225: 153-165, 2002.

49. **Tsai PT, Lee RA and Wu H**. BMP4 acts upstream of FGF in modulating thymic stroma and regulating thymopoiesis. *Blood* 102: 3947-3953, 2003.

50. **Zhang F, Nielsen LD, Lucas JJ and Mason RJ**. Transforming growth factor-beta antagonizes alveolar type II cell proliferation induced by keratinocyte growth factor. *Am J Respir Cell Mol Biol* 31: 679-686, 2004.

51. **Saito T, Itoh H, Yamashita J, Doi K, Chun TH, Tanaka T, Inoue M, Masatsugu K, Fukunaga Y, Sawada N, Sakaguchi S, Arai H, Tojo K, Tajima N, Hosoya T and Nakao K**. Angiotensin II suppresses growth arrest specific homeobox (Gax) expression via redox-sensitive mitogen-activated protein kinase (MAPK). *Regul Pept* 127: 159-167, 2005.

52. **Gorski DH, LePage DF, Patel CV, Copeland NG, Jenkins NA and Walsh K**. Molecular cloning of a diverged homeobox gene that is rapidly down-regulated during the G0/G1 transition in vascular smooth muscle cells. *Mol Cell Biol* 13: 3722-3733, 1993.

53. **Weir L, Chen D, Pastore C, Isner JM and Walsh K**. Expression of gax, a growth arrest homeobox gene, is rapidly down-regulated in the rat carotid artery during the proliferative response to balloon injury. *J Biol Chem* 270: 5457-5461, 1995.

54. **Smith RC, Branellec D, Gorski DH, Guo K, Perlman H, Dedieu JF, Pastore C, Mahfoudi A, Denefle P, Isner JM and Walsh K**. p21CIP1-mediated inhibition of cell proliferation by overexpression of the gax homeodomain gene. *Genes Dev* 11: 1674-1689, 1997.

55. **Fisher SA, Siwik E, Branellec D, Walsh K and Watanabe M**. Forced expression of the homeodomain protein Gax inhibits cardiomyocyte proliferation and perturbs heart morphogenesis. *Development* 124: 4405-4413, 1997.

56. **Mankoo BS, Skuntz S, Harrigan I, Grigorieva E, Candia A, Wright CV, Arnheiter H and Pachnis V**. The concerted action of Meox homeobox genes is required upstream of genetic pathways essential for the formation, patterning and differentiation of somites. *Development* 130: 4655-4664, 2003.

57. **Quinn LM, Latham SE and Kalionis B**. The homeobox genes MSX2 and MOX2 are candidates for regulating epithelial-mesenchymal cell interactions in the human placenta. *Placenta* 21: S50-S54, 2000.

58. **Krisans SK, Ericsson J, Edwards PA and Keller GA**. Farnesyl-diphosphate synthase is localized in peroxisomes. *J Biol Chem* 269: 14165-14169, 1994.

59. **Notarnicola M, Messa C, Cavallini A, Bifulco M, Tecce MF, Eletto D, Di LA, Montemurro S, Laezza C and Caruso MG**. Higher farnesyl

diphosphate synthase activity in human colorectal cancer inhibition of cellular apoptosis. *Oncology* 67: 351-358, 2004.

60. **Reilly JF, Martinez SD, Mickey G and Maher PA**. A novel role for farnesyl pyrophosphate synthase in fibroblast growth factor-mediated signal transduction. *Biochem J* 366: 501-510, 2002.

61. **Badea T, Niculescu F, Soane L, Fosbrink M, Sorana H, Rus V, Shin ML and Rus H**. RGC-32 increases p34CDC2 kinase activity and entry of aortic smooth muscle cells into S-phase. *J Biol Chem* 277: 502-508, 2002.

62. **Badea TC, Niculescu FI, Soane L, Shin ML and Rus H**. Molecular cloning and characterization of RGC-32, a novel gene induced by complement activation in oligodendrocytes. *J Biol Chem* 273: 26977-26981, 1998.

63. **Dangi S and Shapiro P**. Cdc2-mediated inhibition of epidermal growth factor activation of the extracellular signal-regulated kinase pathway during mitosis. *J Biol Chem* 280: 24524-24531, 2005.

64. **Blagoev B, Kratchmarova I, Nielsen MM, Fernandez MM, Voldby J, Andersen JS, Kristiansen K, Pandey A and Mann M**. Inhibition of adipocyte differentiation by resistin-like molecule alpha. Biochemical characterization of its oligomeric nature. *J Biol Chem* 277: 42011-42016, 2002.

65. **Teng X, Li D, Champion HC and Johns RA**. FIZZ1/RELMalpha, a novel hypoxia-induced mitogenic factor in lung with vasoconstrictive and angiogenic properties. *Circ Res* 92: 1065-1067, 2003.

66. The following additional file is available upon request from Dr. Lin Liu (liulin@okstate.edu): Additional file 1: Supplementary Table S1: A list of 50 genes identified in alveolar epithelial type II cells during hyperoxia exposure and recovery. in Excel format.

Summary and Conclusion[11]

In this dissertation, a fine-tuning model of AEC differentiation is proposed based on the results from experimental AEC differentiation models using DNA microarray tools. Both experimental models and microarray tools for gene expression profiling have been developed and tested in the five chapters of this dissertation. It is the first time that a common mechanism between AEC II differentiation models of *in vivo* hyperoxia and *in vitro* cell culture is brought forth based on global gene expression profiles. The model integrates the current knowledge of AEC II differentiation, known gene functions, and the experimental data obtained from this dissertation, and thus may serve a basis for future study of AEC II differentiation.

The fine-tuning model is based on two experimental models, and thus may reflect the true mechanism of AEC II differentiation. For instance, there were 14 genes differentially expressed during hyperoxia exposure and recovery, and 6 of them were also changed during *in vitro* cell culture. Therefore, the hyperoxia-specific genes or effects can be removed from the fine-tuning model. The generality of the model can explain the dual roles of many genes in other models of AEC differentiation. For instance, both stimulatory and inhibitory effects of Bmp4 on AEC II differentiation were observed during fetal lung development, and the respective mechanisms were not known. The dual effects of Bmp4 on FGF-MAPK (activates cell proliferation)

---

[11] Only the most relevant references are listed herein. All other references in this section can be found in chapters 1~5.

and TGFβ (inhibits cell proliferation) signaling pathways in the fine-tuning model can explain the dual roles of Bmp4 during fetal lung development. Besides the generality, the fine-tuning model is novel. For instance, an interesting result from the fine-tuning model is the role of Fdps in AEC differentiation, which was not reported previously. The proposed model reveals a novel cell differentiation function of a known check-point enzyme controlling lipid synthesis. Therefore, this model substantially promotes the understanding of AEC differentiation.

This fine-tuning model may be further improved. First, isolated AEC II were used for gene expression profiling in both hyperoxia and cell culture models, and thus cell isolation effects may affect the results. The relative expression levels of the 6 common genes were down-regulated during the first two days of cell culture. Possibly, the repression of some genes may be resulted from cell isolation processes. A possible way to eliminate the cell isolation is to utilize laser captured microdissection, by which AEC II can be dissected from fixed lung tissue from normal or hyperoxia exposed rats (1). These AEC II retain the snapshots of the original mRNA abundance and can be used for global gene expression profiling without enzymatic cell isolation effects. Second, it appears that hyperoxia exposure is a relative gentle treatment that induces AEC proliferation and differentiation, since there are only 50 genes showed detectable differential gene expressions. The small number of changed genes resulted in only 6 common genes. It is very possible that there are more than 6 genes controlling AEC II differentiation. To identify more genes, a stronger treatment may be utilized in the future. For instance, the *in vitro* model, cultured AEC II, may be treated with hydrogen hyperoxide,

which has a hyperoxia-like but stronger effect on cells. An *in vivo* model, lung tissues for neonatal rats may be cultured and treated with hydrogen hyperoxide or growth factors, in which strong AEC proliferation may be stimulated. With stronger treatments, more AEC II will be activated to enter cell cycle, and thus more differentially expressed gene expression during cell differentiation will be identified. Third, this fine-tuning model largely was based on mRNA expression levels, and it should be verified at protein levels. In the future study, high quality antibodies against the 6 genes should be generated (dlk1 antibody has been tested), and thus the consistency of mRNA and protein expression will be tested. Fourth, AEC II isolated from normal or hyperoxia-exposed rats were used for gene expression profiling, and thus no cells undergoing transition from "AEC I-like" to matured AEC I were used. An improvement is to perform gene expression profiling of cultured AEC II from day 1 to day 7, in which AEC II gradually trans-differentiate into AEC I-like cells. Fifth, the proposed model may be functionally tested by gene knockdown. Currently, no transgenic mouse is available for any of the 6 common genes. Alternatively, one or more of the 6 genes may be silenced by RNA interference technique, which selectively degrades the mRNA of target genes and lowers the respective protein expression level (2). Having silenced one or more of the 6 genes, AEC II differentiation should be promoted or slowed down in some stages, and thus the model can be tested. For instance, if Rgc32 is knocked down by RNA interference, AEC II proliferation should stop or slow down at the G1/S transition.

Besides the fine-tuning model for AEC II differentiation, this dissertation also provides new knowledge about lung and experimental tools,

including the identified lung-prominent genes, AEC-specific genes and functions, as well as improved AEC isolation methods and DNA microarray tools. First, the identified genes in the lung or AEC shed light to the novel functions of the target genes or cells. The 147 lung-prominent genes provide clues of lung functions. For instance, MAPK13, a component of MAPK signaling pathway, was not studied in the lung previously. Its predominant expression in lung may indicate its role in stress and inflammatory responses. A further localization of MAPK13 in specific cells of the lung will reveal the physiological functions of the cells. At cellular level, the new AEC-markers may reveal new functions of AEC I or AEC II. For instance, GABRP is a novel AEC II marker (Chapter 1), which plays a role in the fluid clearance through the regulation of chloride channels (unpublished data). Another example is ApoE and transferrin (Chapter 4), which are expressed higher in AEC I and may protect AEC I during hyperoxia exposure. Second, the tools of AEC isolation and DNA microarray can be used for more detailed investigation of AEC differentiation or gene function study. For instance, DNA microarray can be used to evaluate the specificity of RNA interference against a target gene (e.g. Fdps). Further microarray hybridization also can identify down-stream genes affected by the target gene, Fdps, and thus reveal the genes regulated by Fdps or Fdps-MAPK pathway.

In summary, genes and pathways involved in type II cell proliferation and differentiation were identified in this dissertation, including:

1. At organ level, 147 genes are expressed predominantly in lung, rather than in heart, kidney, liver, spleen or brain by directly comparing transcriptomes of the 6 organs.

2. At cellular level, P2X7 and GABRP are potential AEC I and AEC II markers. Among the 1080 AEC I- and 1142 AEC II-specific genes, two AEC I-specific genes, Apo E and transferrin have protective effect against oxidative stress during hyperoxia exposure.

3. During AEC II differentiation, 6 verified genes may be involved in a regulatory network, which fine-tunes AEC II differentiation through MAPK and TGFβ pathways during *in vivo* hyperoxia exposure and recovery and *in vitro* AEC culture.

Reference List

1. **Player A, Barrett JC and Kawasaki ES**. Laser capture microdissection, microarrays and the precise definition of a cancer cell. *Expert Rev Mol Diagn* 4: 831-840, 2004.

2. **Bosher JM and Labouesse M**. RNA interference: genetic wand and genetic watchdog. *Nat Cell Biol* 2: E31-E36, 2000.

Appendices

Appendix A: permission to reprint chapter 1 from Elsevier

Mr Zhongming Chen

zhongmi@okstate.edu

Dear Mr Chen

BIOCHEMICAL AND BIOPHYSICAL RESEARCH COMMUNICATIONS, Vol 319, No 3, 2004, Pages 774-780, Chen et al, 'Identification of …'

As per your letter dated 24 August 2005, we hereby grant you permission to reprint the aforementioned material at no charge **in your thesis** subject to the following conditions:

1.   If any part of the material to be used (for example, figures) has appeared in our publication with credit or acknowledgement to another source, permission must also be sought from that source. If such permission is not obtained then that material may not be included in your publication/copies.

2.   Suitable acknowledgment to the source must be made, either as a footnote or in a reference list at the end of your publication, as follows:

"Reprinted from Publication title, Vol number, Author(s), Title of article, Pages No., Copyright (Year), with permission from Elsevier".

3.   Reproduction of this material is confined to the purpose for which permission is hereby given.

4.   This permission is granted for non-exclusive world **English** rights only.   For other languages please reapply separately for each one required.   Permission excludes use in an electronic form.   Should you have a specific electronic project in mind please reapply for permission.

5.   This includes permission for UMI to supply single copies, on demand, of the complete thesis.   Should your thesis be published commercially, please reapply for permission.

Yours sincerely

Helen Gainford

Rights Manager

Appendix B: permission to reprint chapter 2 from American Physiological Society

From: Margaret Reich
To: Editor, PG
Date: 8/8/05 6:30PM
Subject: Re: FW: Permission request

You may use the manuscript version of your paper in your thesis. Please site where it is published. Thank you.
Margaret

Margaret Reich
Director of Publications and Executive Editor
American Physiological Society
9650 Rockville Pike
Bethesda, MD 20814-3991
301-634-7071
301-634-7245 (fax)
mreich@the-aps.org

APPROVED
By PRipka at 8:43 am, 8/9/05

THE AMERICAN PHYSIOLOGICAL SOCIETY
9650 Rockville Pike-Bethesda, MD 20814-3991

Permission is granted for use of the material specified above provided the publication is credited as the source, including the words "used with permission."

Margaret Reich

Publications Manager & Executive Editor

VITA

Zhongming Chen

Candidate for the Degree of

Doctor of Philosophy

Dissertation: IDENTIFICATION OF GENES AND PATHWAYS INVOLVED IN ALVEOLAR EPITHELIAL CELL DIFFERENTIATION USING DNA MICROARRAY

Major Field: Physiology

Biographical:

   Education:
      B.S. Sichuan University, P. R. China, 1988- 1992, Microbiology
      M.S. Shanghai Second Medical University (a.k.a., School of Medicine, Shanghai Jiaotong University, P. R. China), 1995- 1998, Microbiology (Advisor: Dr. Jingxing Liu).   Thesis: Roles   Of Coxsackievirus Group B Persistent Infection On Murine Model Of Chronic Myocarditis
      Ph.D.: Completed the requirement for the Doctor of Philosophy degree with a major in Physiology at Oklahoma state University in December 2005.

   Experience:
      Assistant Engineer, 1992-1995, Yichang Yeast Base, Chinese Biotechnology Center, P. R. China.
      Assistant Scientist, 1998-2000, Shanghai Institute of Biochemistry, Chinese Academy of Sciences, P.R. China.
   PROFESSIONAL MEMBERSHIPS:
      1. American Physiology Society, 2004-present
      2. Shanghai Society of Biochemistry and Molecular Biology: 1998-2001
      3. Shanghai Society of Microbiology, 1995-2001

Name: Zhongming Chen   Date of Degree: December, 2005

Institution: Oklahoma State University   Location: Stillwater, Oklahoma

Title of Study: IDENTIFICATION OF GENES AND PATHWAYS INVOLVED IN
        ALVEOLAR EPITHELIAL CELL DIFFERENTIATION USING DNA
        MICROARRAY

Pages in Study:161   Candidate for the Degree of Doctor of Philosophy

Major Field: Physiology

Scope and Method of Study:
        In this project, genes and pathways involved in lung alveolar epithelial
        cell (AEC) differentiation were identified. In-house printed DNA
        microarray, data analysis software (*RealSpot)* and multiple sample
        hybridization system were developed. These tools were tested at organ
        and cell level with biological samples from rats. Differentially expressed
        genes during AEC differentiation were identified using the tested tools
        and a rat model of hyperoxia exposure and recovery. Data from DNA
        microarray were verified at mRNA level with real-time PCR, at protein
        level with Western blot, and at cell level with *in vitro* AEC culture.
        Pathways based on the verified genes were proposed to elucidate the
        mechanisms of AEC differentiation.

Findings and Conclusions:
    1. At organ level, 147 genes are expressed predominantly in lung by
       directly comparing transcriptomes of lung, heart, kidney, liver, spleen
       and brain.
    2. At cell level, P2X7 and GABRP are potential type I AEC (AEC I) and
       type II AEC (AEC II) markers. Among the 1080 AEC I- and 1142
       II-specific genes, two AEC I-specific genes, Apo E and transferrin may
       have protective effect against oxidative stress during hyperoxia
       exposure.
    3. During AEC II differentiation, 6 verified genes may be involved in a
       regulatory network, which fine-tunes AEC II differentiation through
       MAPK and TGFβ pathways during *in vivo* hyperoxia exposure and
       recovery and *in vitro* AEC culture.

ADVISER's APPROVAL: Dr. Lin Liu