CHAPTER III
DIFFERETIAL EXPRESSION PROFILES OF PERIPHERAL BLOOD LEUKOCYTES ASSOCIATED WITH CHOLANGIOCARCINOMA AND POSSIBILITY FOR USING AS PROGNOSTC MARKER

3.1 Introduction

Inflammation is a critical component of tumor progression. Many cancers arise from site of infection, chronic irritation and inflammation. It is now becoming clear that the tumor microenvironment, which is largely orchestrated by leukocytes (the so-called inflammatory cells), is as indispensable have co-opted some of signaling molecules of innate immune system such as cytokine, chemokine, their receptor for invasion, migration, and metastasis (Coussens, Werb, 2002).

Circulating leukocytes serve as a guard and comprehensive surveillance system that patrols the body for signs of infection, inflammation, and other threats, including cancer (Whitney et al., 2003). Recent studies have shown that peripheral blood cells share more than 80% of transcriptome with each nine tissue tested namely, brain, colon, heart, kidney, liver, lung, prostate, spleen and stomach (Liew et al., 2006). Moreover, it has been found that the blood transcriptome profile is adjusted according to the physiological or pathological condition of the body. Therefore, a disease –specific gene expression signature in peripheral blood cells could constitute potential disease markers as well as genes involved in pathogenesis of the disease. Peripheral blood cells have been used to identify gene expression signature for autoimmune, cardiovascular, neurological diseases and cancers such as renal cell carcinoma, breast and bladder cancer (Mohr, Liew, 2007). These signature genes have been shown to be useful in identifying pathways relevant to disease and in diagnosis or predicting response to therapy of the diseases (Baranzini et al., 2005).

Cholangiocarcinoma (CCA) is an inflammation associated cancer and possessed high infiltrating leukocytes in tumor microenvironment (Krungmee et al., 2003). The association of blood leukocytes and leukocytes in tumor environment

were demonstrated in our study as mentioned in Chapter II. High level of peripheral blood $CD14^{+}CD16^{+}$ monocytes was associated with high level of tissue infiltrating macrophages (MA387 +ve cells). These inflammatory cells and tumor cells may communicate or interact and influences the expression profile of infiltrating leukocytes. Thus, gene expression profile of infiltrating leukocytes may reflect the presence of the disease (Liew, 2004). Based on these evidences, we hypothesized that there is a unique gene expression profile of peripheral blood leukocytes specifically to CCA and this blood profile can be used as surrogate tissue reflecting the disease for diagnosis or predicting clinical outcome in CCA patients.

The aims of this study were (1) to investigate whether gene expression profile of peripheral blood leukocytes (PBLs) could reflect disease condition- CCA and (2) whether expression patterns of the CCA associated genes in PBLs could predict eventual patient outcome.

3.2 Material s and Method

3.2.1 Subjects

The subjects in this study were composed of healthy subjects (HE), benign biliary tract disease (BBD) and CCA patients. BBD and CCA subjects were patients who admitted at Srinagarind hospital, Faculty of Medicine, Khon Kaen University and had histological proved to be benign biliary tract disease or CCA. Healthy subjects were defined as the persons who had age and sex matched with CCA patients, normal clinically health with normal complete blood count (CBC), normal liver function test (LFT), and no appear chronic inflammatory diseases such as diabetic mellitus or hepatitis. Informed consent was obtained from each subject. The protocol for this study was approved by the Human Research Ethics Committee, Khon Kaen University (HE471214 & HE480312).

For analysis of gene expression profiles in peripheral leukocytes was performed in 9 CCA patients, 5 BBD patients and 8 healthy subjects that were summarized in Table 3.1 and Supplement data:Table B1.

For gene validation, peripheral blood leukocytes from CCA patients were recruited in two phases. The first phases, a "training set" of 35 samples was obtained from the patients with histological proved as CCA and 20 control samples

from healthy subjects according to criteria s described above. In the second phase, a "testing set" of 20 samples was obtained from prospectively enrolled patients with suspected CCA.

The characteristic of the subjects in this study is summarized as shown in Table 3.1. All subjects showed no significant difference in the average age but they had differences in total white blood cells (WBC) count

Table 3.1 Characteristics of CCA, BBD and healthy subjects used in cDNA microarray analysis

	CCA patients (n = 9)	**BBD patients** (n = 5)	**Healthy subjects** (n = 8)	**P value**
Sex (Male/Female)	7/2	5/0	4/4	
Age (yrs)	57.67 ± 12	53.2 ± 7	49.13 ± 3.5	0.086
White blood cells counts (Mean : cells/μl)	12.24×10^3	12.52×10^3	6.08×10^3	0.006**

** CCA or BBD differ from HE but CCA not differ from BBD

3.2.2 RNA extraction

Total RNA was extracted using TRIzol® reagent with Purelink™ Micro-to-Midi system kit (Invitrogen, Carlsbad, CA, USA) including phase separation with TRIzol® reagent and purification. For phase separation step, the heparinized blood (0.5 ml) was added into 4.5 ml TRIzol® reagent. Chloroform (200 μl) was added to 1 ml of TRLzol® reagent and the tube was shaken vigorously by hand for 15 seconds, incubated at room temperature (RT) for 2-3 min and centrifuged at 12,000 xg for 15 minutes. After centrifugation, the mixture was separated into lower, red phenol-chloroform phase, interphase, and colorless upper aqueous phase containing RNA. The aqueous phase was transferred to a fresh tube and an equal volume of 70% ethanol was added to obtain a final ethanol concentration of 35% and mixed well by vortexing. The purification step was performed by transferring up to 700 μl of the sample to the RNA spin cartridge pre-inserted in collection tube. Centrifuged the

cartridge at 12,000 xg for 15 second,discharged the flow-through and re-inset cartridge in the tube. Repeated this step until the entire sample was processed, which then 700 µl of Wash buffer I was added to the spin cartridge centrifuged and discharged the flow-through. The washing step was repeated with 500 µl of Wash buffer II with ethanol, centrifuged and discharged the flow-through. The spin cartridge was then centrifuged at 12,000 x g for 1 min to dry the membrane with attached RNA. The cartridge was then inserted into an RNA Recovery tube and eluted with 20-30 µl of RNase-free water with further incubation at RT for 1 min. The spin cartridge was then centrifuged for 2 min at >12,000 xg. The collected RNA was stored at -80°C until analysis. The quality and quantity of RNA was determined by 2100 Bioanalyzer RNA LabChip (Agilent Technologies, Palo Alto, CA) and Nano drop spectrophotometer (ND-1000 spectrophotometer and ND1000 version 3.2.1 software), respectively. The 28s:18s rRNA ratio of RNA sample showed equal or more than 1.5-1.7 was selected for gene chip experiment and verification by real time RT-PCR.

3.2.3 Affymetrix one cycle eukaryotic expression sample processing

The first step was one cycle cDNA synthesis which included:

(1) First strand cDNA synthesis required high-quality total RNA 1 µg and Poly A RNA control kit (Positive control). Prepared the Poly A RNA dilution: from first dilution (1:20) by making 1 of 1:50 dilution (2 µl of 1st dilution + 98 µl of H_2O) and 1 of 1:25 dilution (2 µl of 1st dilution + 48 µl of H_2O). Prepared RNA sample by adding 1 µl of 3rd dilution of poly A RNA control, 2 µl T7-oligo (dT) primer (50 µM), 9 µl of total RNA (1 µg) and H_2O to make a final volume of 12 µl. Incubated the RNA at 70°C, 10 min and then placed tube on ice for 2 min and spun down to collect sample at bottom of tube. After that, 7 µl of First strand master mix (4 µl of 5X 1st strand reaction mix, 2 µl of 0.1 M DTT, and 1 µl of 10mM dNTP ; Invitrogen) was added to RNA sample, mixed well and placed the tube at 42°C for 2 min and then added 1 µl of Superscript III (Invitrogen). Mixed the final volume 20 µl, centrifuged and put back to 42°C for 1 hour. (2) Second strand cDNA synthesis. To 20 µl of the first strand synthesis sample, added130 µl of second strand master mix (91 µl of H_2O, 30 µl of 5X 2nd strand reaction mix, 3 µl of 10 mM dNTP, 1 µl E coli DNA ligase,

4 µl of Ecoli DNA polymerase I, and 1 µl of RNaseH ; Invitrogen) for a total volume of 150 µl. Incubated the mixture at 16°C for 2 h,then 2 µl of T4 DNA polymerase was added and incubated at 16°C for 5 min. After that, cleaned the sample using Sample cleanup module (Affymetrix 900371) for cleanup of Double stranded cDNA and determined the cDNA concentration by Nano Drop.

The Second step was Synthesis of Biotin-labeled cRNA. Template cDNA 1 µg in 20 µl was required for adding in IVT labeling master mix (4 µl of 10X labeling buffer, 12 µl of IVT NIT mix, 4 µl of IVT labeling enzyme mix and H_2O to get the final volume of 40 µl), incubated at 37 °C for 16 h. Then, cleanup the sample using sample cleanup module (Affymetrix 900371) for cleanup of Biotin-labeled cRNA. Quantification and quality of Biotin-labeled cRNA were determined using Nano Drop and Agilent 2100 BioAnalyzer, respectively. After calculation to get cRNA equaled to 650 – 690 µg in 36 µl.

The third step was Fragmentation the cRNA for target preparation. This step required 650 µg of Biotin labeled cRNA in 32 µl for adding in 8 µl of fragmentation buffer to get a final volume of 40 µl. Incubated the mixture at 94°C, 35 min and the quality of fragmented biotin-labeled cRNA was measured with Agilent 2100 BioAnalyzer.

Last step was Eukaryotic target hybridization : 49/64 format array. Fragmented biotin labeled cRNA 38 µl was added in Hybridization cocktail (5 µl of control 3nM Oligonucleotide B2, 15 µl of 20X Eukaryotic hybridization control, 3 µl of 10 mg/ml herring sperm DNA, 3 µl of 50 mg/ml acetylated BSA, 150 µl of 2X hybridization buffer, 30 µl DMSO, and H_20 300 µl). Filled array with 200 µl of 1X hybridization buffer, incubated in hybridization oven at 45°C, 10 min with rotation. Heat the hybridization cocktail to 99°C for 5 min and then 45 °C for 5 min. Spun at 13,000 rpm for 5 min, removed the buffer solution and filled with 200 µl of clarified hybridization cocktail. Finally, placed the probe array into a hybridization oven and incubate at 45 °C for 16 hours, rotation at 60 rpm.

3.2.4 Staining and Washing Affymetrix gene chip by Fluidic station.

After 16 hours of Affymetrix gene chip hybridization. Affymetrix gene chip was stained with Biotinylated antibody solution mix and Streptavidin

Phycoerythrin (SAPE) solution mix. After that All of Affymetrix gene chips were washed with washing buffer A and B.

3.2.5 Microarray analysis

Expression data was obtained using an Affymetrix GSC3000 scanner and processed by Gene Chip Operating Software; GCOS (Affymetrix, Santa Clara, CA). Partek software (Agilent Technologies, Palo Alto Ca) was used for downstream analysis of GCOS processed data (Principle component analysis, Hierarchical Clustering analysis, Venn diagram and Gene Ontology). Signals from all probe sets were normalized using Human Genome U133 Plus 2.0 Array Normalization Controls. The criteria for selecting differentially expressed genes between CCA (or BBD patients) and healthy subjects were

(i) Mean fluorescence intensity in each probe set was equaled or more than 8 for all up or down regulated genes

(ii) The significantly expression at P value was less than 0.05 and 1.5 fold difference in up or down regulated genes.

3.2.6 Real-time RT-PCR (*Quantitative Polymerase Chain Reaction)*

Transcript copy number for specific genes of interest was measured using an adaptation of a two-step real-time reverse transcriptase–polymerase chain reaction (Real time RT-PCR) method.

Real time RT-PCR for specific genes of interest and internal control will be performed using a SYBR green assay. The primers were designed using Integrated DNA technologies company (IDT®, http://www.idtdna.com/SciTools/SciTools.aspx). Approximately 200 ng of total RNA from each sample was converted to cDNA using the 1st Strand cDNA Synthesis Kit for RT-PCR [AMV] kit (Roche Applied Diagnostics, Indianapolis IN) according to manufacturer's instructions. After the first-strand synthesis, the reverse transcriptase was denatured by incubation at 99°C for 5 min followed by quick cooling. DNA was stored at –20°C until use. PCR was performed on a LightCycler (Roche Applied Diagnostics, Indianapolis IN) using the LightCycler FastStart DNA Master SYBR Green I kit (Roche Applied Diagnostic, Indianapolis IN) and ~ 5 ng cDNA sample. Amplifications included one cycle of template denaturation at 95 °C for 10 minutes followed by 45 cycles of 95°C for 10 seconds, 68°C for 10 seconds, and 72°C for 16 seconds were performed by

LightCycler 3 Run machine (Roche Applied Diagnostic, Indianapolis, IN). The presence of a single amplified product was confirmed by DNA melting point analysis. Threshold cycles (Ct) for each amplification reaction were determined using LightCycler Software version 3.5 (Roche Applied Diagnostics, Indianapolis IN). All samples were also amplified with the human β-actin LightCycler–Primer Set (Roche Applied Diagnostics, Indianapolis IN). The sequences of gene-specific primers employed are provided in Table 3.2. The gene-specific primers for individual samples (including CCA patients and healthy subjects) were normalized to signals obtained with β-actin from the same sample. Relative change in gene expression of CCA patients were analyzed by using $2^{-\Delta\Delta Ct}$ method where $\Delta\Delta Ct = (Ct_{target} - Ct_{actin})_{patients} - (Ct_{target} - Ct_{actin})_{healthy}$. The relative change in gene expression of healthy subjects was indicated to be 1X expression of each target genes. Thus, relative change value show more than 1X expression represented up-regulation, whereas that showed less than 1X expression represented down-regulation.

3.2.7 Isolation of neutrophils, lymphocytes and monocytes

Heparinized blood was draw from CCA patients or healthy subjects. Dilute heparin-treated whole blood with an equal volume of 1X PBS (Ca++, Mg++ free, sterile). Gently layer the diluted blood (2 volumes) over Ficoll Hypaque (1volume). After centrifugation at 1,500 RPM at room temperature (Beckman table-top centrifuge, 12 cm radius swing-bucket rotor) for 20 min.at 20°C, brake off, gently removed white colored band of cells at the plasma/Ficoll hypaque interface, and transfered to a new centrifuge tube. Washed cells with 1XPBS and spun at 2,000 RPM for 5 min. utes. Lysed the leftover RBC with 5 ml RBC lysis buffer, incubated for 3 min and spun at 2,000 rpm for 5 min. Peripheral blood mononuclear cells (PBMCs) were suspend in RPMI-1640 with 10% FBS. The PBMCs ($5x10^6$) were incubated at 37°C, 1 hr in plastic tissue culture dishes (diameter 10 cm), to allow selective adherence of monocytes, followed by two successive rinses with warmed medium to remove non-adherent lymphocytes. These preparations were at least 90% monocytes as assessed by immunofluorescent staining with antibodies to the CD14 antigen Figure 3.1A. For granulocyte separation, granulocytes in pack red cells from ficoll separation were washed and resuspended with 1XPBS. Gently layered this diluted (2 volumes) over Percoll gradient 63% and 72% (1 volume) and centrifuged at

the same conditions for ficoll separation. Gently removed white colored band of cells at the 63% percoll/72% percoll interface and transfered to a new centrifuge tube. Fraction of adherent cells (monocytes), non-adherent cells (lymphocytes), and granulocytes were kept in TRIzol® reagent and performed RNA extraction and 1st cDNA synthesis. Each fraction showed at least 90% purity as assessed with cell sedimentation technique and demonstrated with Wright-Giemsa staining (Figure 3.1B).

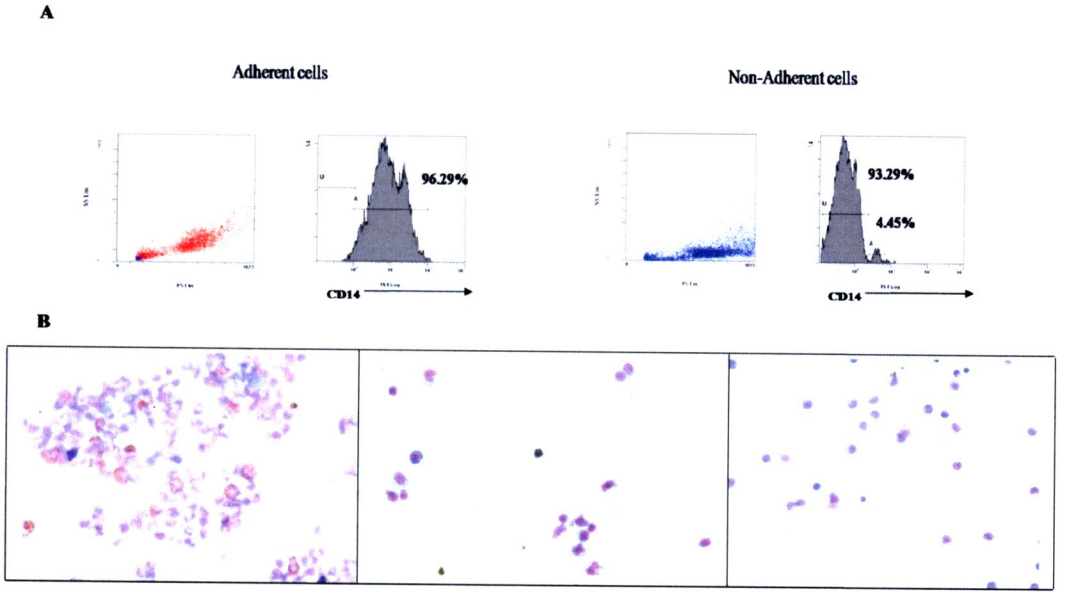

Figure 3.1 Immunofluorescent staining and Wright-Giemsa staining in fractions of polymorphonuclear cells (PMN), monocytes and lymphocytes. Immunofluorescent staining with CD14 antigen of adherent and non-adherent cells after incubation at 37 °C for 1 hr and analyzed by flow cytometer (A). Fractions of adherent cells (monocytes), non-adherent cells (lymphocytes), and PMN were assessed with cell sedimentation technique and demonstrated with Wright-Giemsa staining. Each fraction showed at least 90% purity (B).

3.2.8 Statistical analysis

Statistic analyses were done using SPSS statistical software version16.0.1 (SPSS Inc., Chicago, IL) and STATA version 8. The different expressions of such candidate genes in each group were compared with student-*t*-test. Cox regression was used to establish prognostic index (or risk score) from the expression levels of candidate genes and eventual patient outcome. Kaplan-Meier survival analysis was used to estimate the disease-specific survival and comparison between groups were done with a log-rank test. Cross tabulations were analyzed with χ^2-test for the associations between prognostic index (risk score) with clinico-pathological features of CCA patients.

Table 3.2 Oligonucleotide sequences used

Candidate genes	Sequences of primers and probes (5´to 3´)	Amplicon size (bp)
Beta-Actin	F[a]: CACACTGTGCCCATCTACGA R[b]: CTCCTTAATGTCACGCACGA	162
CTSL	F: AGACATGGATCATGGTGTGCTGGT R: ACAATGGTTTCTCCGGTCTTTGGC	151
CXCL3	F: AGTGGATCACTGTTAGGGTAAGGG R: CGCTGCAGAATGGACATTAAACAAGGC	198
CXCL10	F: TCACCTTTCCCATCTTCCAAGGGT R: GGTAGCCACTGAAAGAATTTGGGC	180
EREG	F: ACCTGTATCTGACCCACTTTGT R: ACGGTCAGAATATATTGGGCATCA	99
IL8	F: GACATACTCCAAACCTTTCCACCC R: TCCAGACAGAGCTCTCTTCCATCA	123
MMP9	F: ATTTCTGCCAGGACCGCTTCTACT R: TTACATGGCACTGCAAAGCAGGAC	141
PTGES	F: CCGTTGGCTTTGGATGTCTTTGCT R: TTTGGAGGGACTCAAACCTTGGGA	109
PLAU	F: TGTGAGATCACTGGCTTTGGA R: AGTCAAAGTCATGCGGCCTT	237
VEGFA	F: TTCCAATCTCTCTCTCCCTGATCGGT R: CCTCTTTCAAAGGAATGTGTGCTGGG	128
SERPINB2	F: TGTAGATGAGCTGTGTGCCTCAGA R: GGGTAGCAGAAGTTGTTCAGAAGAGCAG	100
TLR8	F: GTGAGCTCTGATTGCTTCAGTTGG R: GGTCAACTGCTAAGATGGTCCACA	135

[a] F, Forward primer
[a] R, Reverse prime

3.3 Results

3.3.1 Identification of differentially expressed genes of CCA in peripheral blood leukocytes through cDNA microarray

Using principal component analysis (PCA) for determining expression trend within the dataset of CCA patients and healthy subjects demonstrated the system variance transcriptome indicating 49.2 % difference between transcriptome of peripheral blood leukocytes (PBLs) from 9 CCA patients and those from 8 healthy subjects (Figure3.2). A comparison of the PBLs gene expression profiles of CCA patients and healthy individuals using the Affymetrix GeneChip U133 Plus 2.0 system showed significant differences ($P < 0.05$) in the expression levels of 177 genes. Of the 177 genes, 117 genes were up regulated (Table 3.3) and 60 genes were down regulated in the PBLs from CCA patients (Table 3.4). Cluster analysis on the up or down regulated genes arranged the sample according to the similarities in gene expression patterns. Hierarchical clustering analysis of the up or down regulated genes resulted in a clear separation of the CCA patients from healthy controls (Figure 3.3). Differential genes expressed in the PBLs from CCA patients was used to predict the molecular and cellular functions as determined using ingenuity pathway analysis (IPA) software. The top significant biological functions was predicted to be antigen presentation, cell death, cellular movement, cell to cell signaling and interaction, and cellular growth and proliferation (Supplement data: Figure B2).

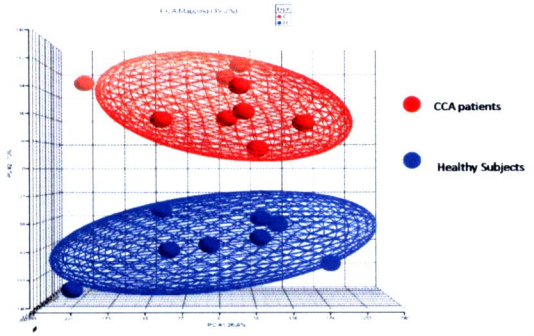

Figure 3.2 Principle component analysis for expression trend between CCA patients and healthy subjects. The Ellipsoid view showed that specimens were grouped by disease. The result showed 49.2 % of the system variance between the expression of peripheral blood leukocytes from 9 CCA patients (red dots) and 8 healthy subjects (blue dots)

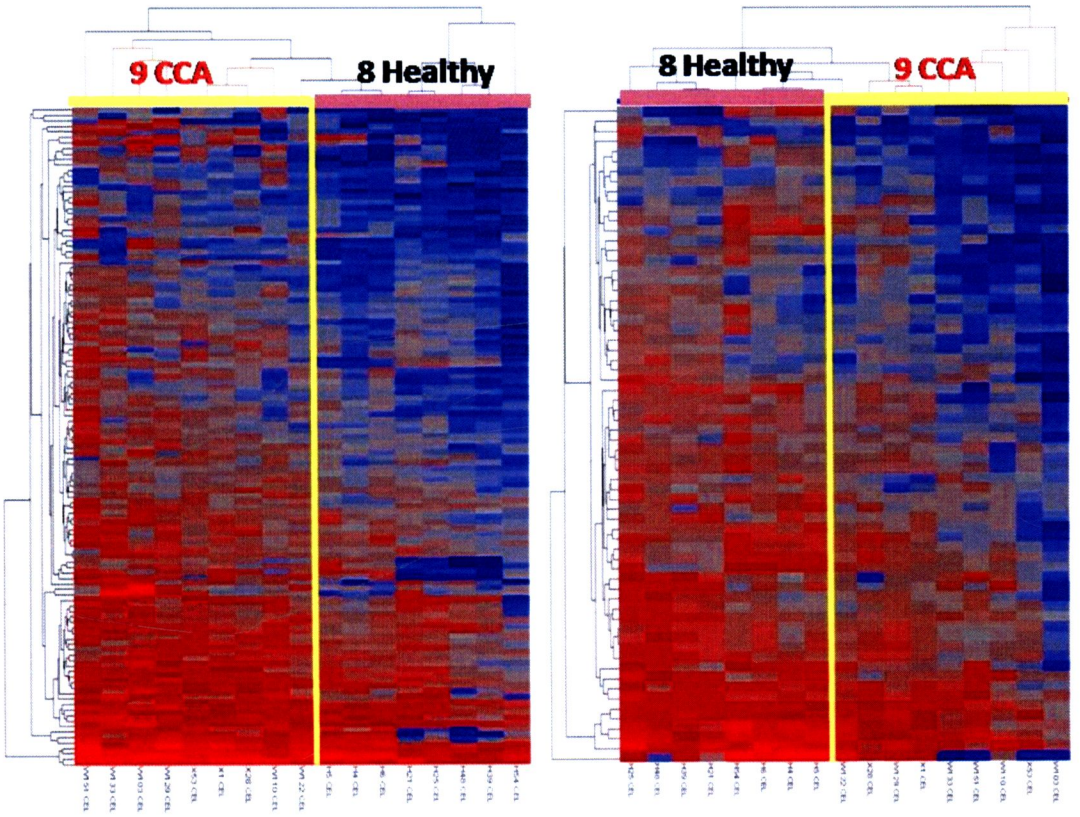

Figure 3.3 Hierarchical clustering analysis of up or down regulated genes in the expression profiles of peripheral blood leukocytes from CCA patients and healthy controls. Left panel showed the hierarchical clustering of 117 up regulated. Right panel showed 60 down regulated genes.

Table 3.3 Lists of up-regulated genes in peripheral blood leukocytes of CCA patients

Probeset ID	Gene Symbol	Gene Name	Fold Change	Biological processes
Protease, Peptidase and Invasion related genes				
211668_s_at	PLAU	plasminogen activator, urokinase	6.3388	proteolysis
204614_at	SERPINB2	serpin peptidase inhibitor, clade B (ovalbumin), member 2	5.60283	anti-apoptosis
203887_s_at	THBD	thrombomodulin	3.87076	blood coagulation
210004_at	OLR1	oxidised low density lipoprotein (lectin-like) receptor 1	2.67693	proteolysis
202087_s_at	CTSL	cathepsin L	2.64564	proteolysis
203936_s_at	MMP9	matrix metallopeptidase 9 (gelatinase B)	2.62559	peptidoglycan metabolism
202381_at	ADAM9	ADAM metallopeptidase domain 9 (meltrin gamma)	1.96734	proteolysis
201666_at	TIMP1	TIMP metallopeptidase inhibitor 1	1.87811	development
201943_s_at	CPD	carboxypeptidase D	1.82203	proteolysis
200742_s_at	TPP1	tripeptidyl peptidase I	1.52183	proteolysis
Growth factor/Angiogenic/ Cell adhesion genes				
205767_at	EREG	epiregulin	4.01995	regulation of progression through cell cycle
210512_s_at	VEGF	vascular endothelial growth factor	3.33421	regulation of progression through cell cycle
211506_s_at	IL8	interleukin 8	2.46387	angiogenesis
202888_s_at	ANPEP	alanyl (membrane) aminopeptidase (aminopeptidase N, aminopeptidase M, microsomal	2.26976	angiogenesis
213746_s_at	FLNA	filamin A, alpha (actin binding protein 280)	2.06052	cell motility
212330_at	TFDP1	transcription factor Dp-1	1.7254	regulation of progression through cell cycle
213524_s_at	G0S2	G0/G1switch 2	1.70691	regulation of progression through cell cycle
203006_at	INPP5A	inositol polyphosphate-5-phosphatase, 40kDa	1.67145	cell communication
212723_at	PTDSR	phosphatidylserine receptor	1.6638	cell differentiation
210184_at	ITGAX	integrin, alpha X (complement component 3 receptor 4 subunit)	1.66255	cell adhesion
203085_s_at	TGFB1	transforming growth factor, beta 1 (Camurati-Engelmann disease)	1.57519	regulation of progression through cell cycle
220000_at	SIGLEC5	sialic acid binding Ig-like lectin 5	1.51549	cell adhesion
Immune response, Cytokine, /Chemokine related genes				
207850_at	CXCL3	chemokine (C-X-C motif) ligand 3	4.90128	chemotaxis
209774_x_at	CXCL2	chemokine (C-X-C motif) ligand 2	2.88502	chemotaxis
202948_at	IL1R1	interleukin 1 receptor, type I	2.77718	inflammatory response
211372_s_at	IL1R2	interleukin 1 receptor, type II	2.08442	immune response
211307_s_at	FCAR	Fc fragment of IgA, receptor for	2.03052	immune response
202878_s_at	CD93	CD93 molecule	1.96563	phagocytosis
221676_s_at	CORO1C	coronin, actin binding protein, 1C	1.78989	phagocytosis
223454_at	CXCL16	chemokine (C-X-C motif) ligand 16	1.56218	receptor mediated endocytosis
218310_at	RABGEF1	RAB guanine nucleotide exchange factor (GEF) 1	1.54423	endocytosis
210140_at	CST7	cystatin F (leukocystatin)	1.51943	immune response

Table 3.4 Lists of down-regulated genes in peripheral blood leukocytes of CCA patients.

Probeset ID	Gene Symbol	Gene Name	Fold Change	Biological processes
Immune response, Cytokine, /Chemokine related genes				
204533_at	CXCL10	chemokine (C-X-C motif) ligand 10	-8.25947	cell motility
214329_x_at	TNFSF10	tumor necrosis factor (ligand) superfamily, member 10	-4.19406	apoptosis
211990_at	HLA-DPA1	major histocompatibility complex, class II, DP alpha 1	-2.50858	immune response
228592_at	MS4A1	membrane-spanning 4-domains, subfamily A, member 1	-2.49579	immune response
229560_at	TLR8	toll-like receptor 8	-2.36822	inflammatory response
204502_at	SAMHD1	SAM domain and HD domain 1	-2.36814	immune response
223501_at	unknow	unknow	-2.34527	immune response
223502_s_at	TNFSF13B	tumor necrosis factor (ligand) superfamily, member 13b	-2.30174	immune response
209788_s_at	ARTS-1	type 1 tumor necrosis factor receptor shedding aminopeptidase regulator	-1.90106	angiogenesis
221698_s_at	CLEC7A	C-type lectin domain family 7, member A	-1.89943	phagocytosis, recognition
209823_x_at	HLA-DQB1	major histocompatibility complex, class II, DQ beta 1	-1.88838	immune response
219947_at	CLEC4A	C-type lectin domain family 4, member A	-1.72214	immune response
209374_s_at	IGHM	immunoglobulin heavy constant mu	-1.71396	immune response
213716_s_at	SECTM1	secreted and transmembrane 1	-1.60739	immune response
Protein Biosynthesis				
222858_s_at	DAPP1	dual adaptor of phosphotyrosine and 3-phosphoinositides	-2.44382	protein amino acid dephosphorylation
231866_at	LNPEP	leucyl/cystinyl aminopeptidase	-2.41458	proteolysis
211795_s_at	FYB	FYN binding protein (FYB-120/130)	-1.94799	protein amino acid phosphorylation
224250_s_at	SECISBP2	SECIS binding protein 2	-1.84643	protein biosynthesis
200629_at	WARS	tryptophanyl-tRNA synthetase	-1.71573	protein biosynthesis
225973_at	TAP2	transporter 2, ATP-binding cassette, sub-family B (MDR/TAP)	-1.66772	protein complex assembly
214843_s_at	USP33	ubiquitin specific peptidase 33	-1.63294	ubiquitin-dependent protein catabolism
201998_at	ST6GAL1	ST6 beta-galactosamide alpha-2,6-sialyltranferase 1	-1.55213	protein amino acid glycosylation
218738_s_at	RNF138	ring finger protein 138	-1.50081	protein ubiquitination
Regulatory related genes				
223298_s_at	NT5C3	5'-nucleotidase, cytosolic III	-2.34568	pyrimidine nucleoside metabolism
209199_s_at	MEF2C	MADS box transcription enhancer factor 2, polypeptide C	-2.28759	transcription
206715_at	TFEC	transcription factor EC	-2.2294	regulation of transcription
223980_s_at	SP110	SP110 nuclear body protein	-2.12618	transcription
204838_s_at	MLH3	mutL homolog 3 (E. coli)	-2.01182	mismatch repair
218589_at	P2RY5	purinergic receptor P2Y, G-protein coupled, 5	-1.89457	signal transduction

3.3.2 Verification of candidate genes by RT-PCR

From 177 differentially expressed genes of PBLs from CCA patients, a set of 11 genes were selected as initial candidates to be evaluated by quantitative real time RT-PCR based on two criteria 1) genes which expressed in mononuclear cells (monocytes or lymphocytes) and 2) be noted to promote tumor proliferation, angiogenesis, and metastasis. To prove the robustness of the strategy for identification of genes showing altered expression, eleven genes including 9 up-regulated genes (*PLAU, SERPINB2, VEGFA, EREG, MMP9, IL-8, PTGS, CTSL*, and *CXCL3*) and 2 down-regulated genes (CXCL10 and TLR8) were investigated by quantitative reverse transcription (RT)-PCR analysis using the same set of specimen used for the microarray analysis. As a result, a comparison of the expression ratios of these 11 genes from the microarray analysis and quantitative RT-PCR showed similar result (Figure 3.4).

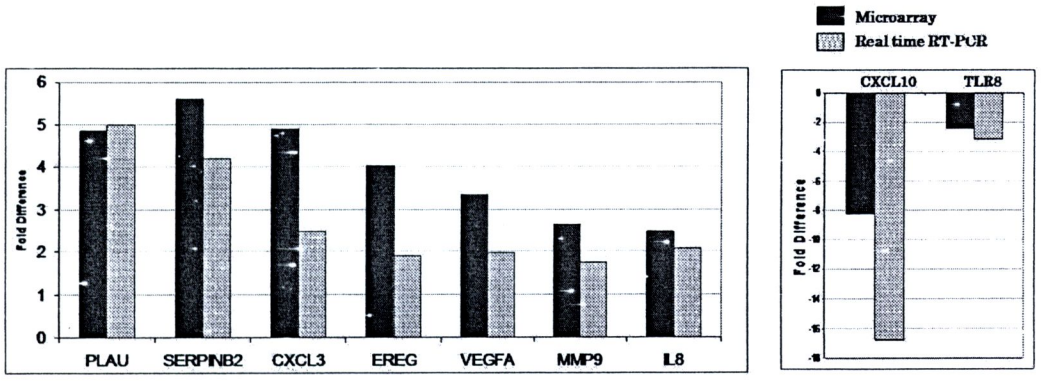

Figure 3.4 Comparison of the expression levels of 9 up- and 2 down-regulated genes of PBLs from CCA patients obtained from cDNA array chip and quantitative real time RT-PCR. Black bar = cDNA array; gray bar = real time PCR.

Since PBLs consist of mixed cell population including poly-morphonuclear cells (PMNs), lymphocytes and monocytes, the cell population which expressed these candidate genes were identified. The expression levels of 9 up-regulated genes (*PLAU, SERPINB2, VEGFA, EREG, MMP9, IL-8, PTGS, CTSL*, and *CXCL3*) were assayed by quantitative RT-PCR in the isolated PMNs, lymphocyte and

monocyte from peripheral blood of 5 CCA patients. The expression levels of these 9 candidate genes in all leukocyte populations were assessed with $2^{-\Delta Ct}$ which ΔCt calculates from Ct value between the candidate gene and a housekeeping gene (ΔCt = Ct of target gene – Ct of housekeeping gene: β-actin). The data demonstrated that the expressions of *PLAU, SERPINB2, CTSL, VEGFA, EREG,* and *CXCL3* were found mainly in the populations of monocytes and lymphocytes *MMP9* expression was mainly observed in PMNs. Whereas *IL-8* and *PTGS* expressions were expressed in all circulating leukocytes (Figure3.5).

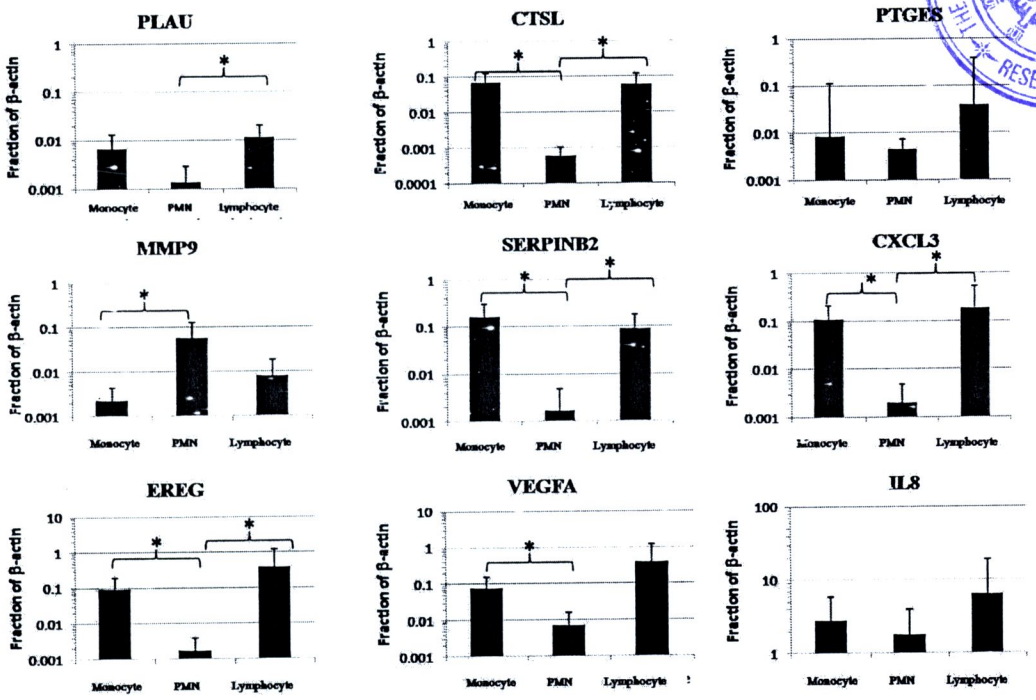

Figure 3.5 Representative expressions of the 9 candidate genes (PLAU, CTSL, PTGES, MMP9, SERPINB2, CXCL3, EREG, VEGFA, and IL8) in fractions of monocytes, leukocytes, and PMNs from CCA patients.

* = $P < 0.05$. Mann Whitney Rank Sum test

3.3.3 Construction and validation of classification models for predicting the outcome of the CCA patients

One of the objectives of this study was to identify a distinct gene signature of PBLs that may allow development of noninvasive screening test to identify individual at risk for CCA or to predict the survival outcome of the patients. The 9 up-regulated genes (*PLAU, SERPINB2, VEGFA, EREG, MMP9, IL-8, PTGS, CTSL*, and *CXCL3*) were evaluated for the potential of being signature in PBLs of CCA patients in the training set of 35 CCA patients and 20 controls. The expression patterns of each gene ($2^{-\Delta\Delta Ct}$) in 35 CCA samples were used to obtain the regression coefficient and hazard ratio of all these genes, which were then used to define "risk score" (so-called prognostic index) by Cox regression model. Our result showed that the best combination equation from three genes related to invasion -PLAU, SERPINB2 and CTSL- for predicting the survival outcome of CCA patients was; risk score = 1.020 PLAU + 1.214 CTSL – 0.882 SERPINB2.

To determine if the risk score was related to patient prognosis, the Kaplan-Meier curve analysis and log rank test based on levels of risk score were tested. Using the cutoff point of risk score = 4.21, (which is the risk score at 50^{th} Percentile of the training set), the CCA patients were stratified into 2 groups: low (<4.21) and high risk score (>4.21). The overall survival of CCA patients was significantly reduced in patient with high risk score ($\geq$ 4.21) (mean survival of 239 days; 95% CI, 167- 311days) when compared with those who had low risk score (< 4.21) (mean survival of 403 days; 95% CI, 320-485 days) (P = 0.021, Figure 3.6). Since, several clinical parameters have been shown to correlate with CCA prognosis; we further determined whether the risk score was a confounding factor underlying the clinical condition by performing univariate and multivariate Cox proportional hazard regression analysis. A univariate of various clinical variables demonstrated that only staging and "risk score" were significant predictors of survival (P < 0.05). The multivariate Cox regression model for survival, which controlled for age, sex, staging and "risk score" (Table 3.5), indicated that both staging and "risk score" are independent predictors of survival for CCA with hazard ratio of 3.85 (95% CI, 0.64 – 23.05; between stage III versus I-II), 7.35 (95% CI, 1.40 – 40.31; between stage IV

versus I-II) and 3.61 (95% CI, 0.95 – 13.72; between high versus low score) respectively.

In addition, to influence clinical decision making, the prognostic index must be shown to be robust when applied in heterogeneous population of the patients. The testing set (20 samples) represented such patients who were speculated to be CCA in clinical practice. The prognosis performance of risk score will be evaluated when the data collection for survival analysis is completed.

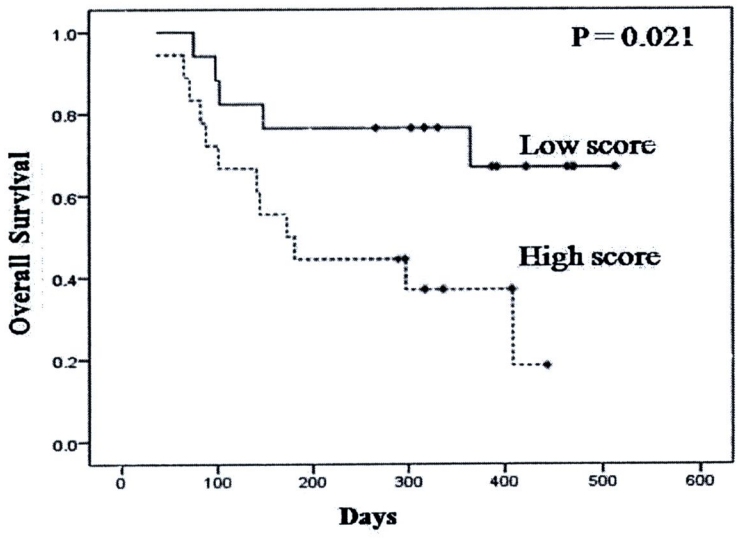

Figure 3.6 Kaplan-Meier survival curves of CCA patients according to high or low risk score. The mean survivals of patients with high and low risk score are 239 days and 403 days, respectively. There was different survival time between both groups of the patients ($P < 0.05$). Risk score = 1.020 PLAU + 1.214 CTSL – 0.882 SERPINB2.

Table 3.5 Univariate and Multivariate analysis of factors associated with survival

Variable	Survival			
	Univariate analysis[a]		Multivariate analysis[b]	
	Hazard ratio (95%CI)[c]	P value	Hazard ratio (95%CI)	P value*
Age (≥ 56 versus < 56)	1.37 (0.52-3.56)	0.518	2.67 (0.90-7.93)	0.074
Sex (male versus female)	1.67 (0.54-5.16)	0.367	1.30 (0.30-5.62)	0.705
Tumor location (ICC versus ECC)[d]	1.12 (0.43-2.90)	0.815	n.a.[1]	
Gross morphology				
(Periductal infiltrating versus intraductal Growth type)	2.4 (0.3-18.83)	0.403	n.a.[1]	
(Mass forming versus versus intraductal Growth type)	2.0 (0.23-17.33)	0.527	n.a.[1]	
Histology type (papillary CCA versus non-papillary CCA)	0.73 (0.27-1.99)	0.545	n.a.[1]	
Tumor Stage				
(III versus I-II)	2.48 (0.48-12.83)	0.278	3.85 (0.64-23.05)	0.023
(VI versus I-II)	8.55 (1.83-39.73)	0.006	7.35 (1.40-40.31)	
Vascular invasion (present versus absent)	0.77 (0.25-2.39)	0.661	n.a.[1]	
Risk score (high versus low)	3.24 (1.13-9.30)	0.029	3.61 (0.95-13.72)	0.046

[a] Univariate analysis, Cox proportional hazards regression.

[b] Multivariate analysis, Cox proportional hazards regression.

[c] 95% CI, 95% confidence interval

[d] ECC, Extrahepatic cholangiocarcinoma: ICC, Intrahepatic cholangiocarcinoma

[1] n.a., not applicable

* Partial likelihood ratio test

3.3.4 Comparison of differentially expressed genes between CCA and BBD patients

To evaluate gene expression profiles of PBLs that are associated to CCA, the expression profile of PBLs from CCA patients were compared to those of in benign biliary tract disease and healthy persons. The principle component analysis was used to show the transcriptome of PBLs from CCA, BBD patients and healthy subjects. As demonstrated in Figure 3.7, there is an overlapping expression between CCA and BBD patients. However, the expression profiles of PBLs from these two groups of the patients can be distinguished from healthy subjects. Differential expressed genes at $P < 0.05$ from CCA and BBD patients were analyzed by Venn diagram after subtracted from those of healthy subjects. The common and unique set of genes between CCA and BBD were identified. There were 116 genes (Supplement data:Table B2) differentially expressed in CCA patients and 54 genes were common between CCA and BBD groups (Figure 3.8). Interestingly, 9 of 11candidate genes selected from XXX (*SERPINB2, VEGFA, EREG, MMP9, PTGS, CTSL, CXCL3*, and CXCL10) were included in the unique set of differential expressed genes in PBLs from CCA patients.

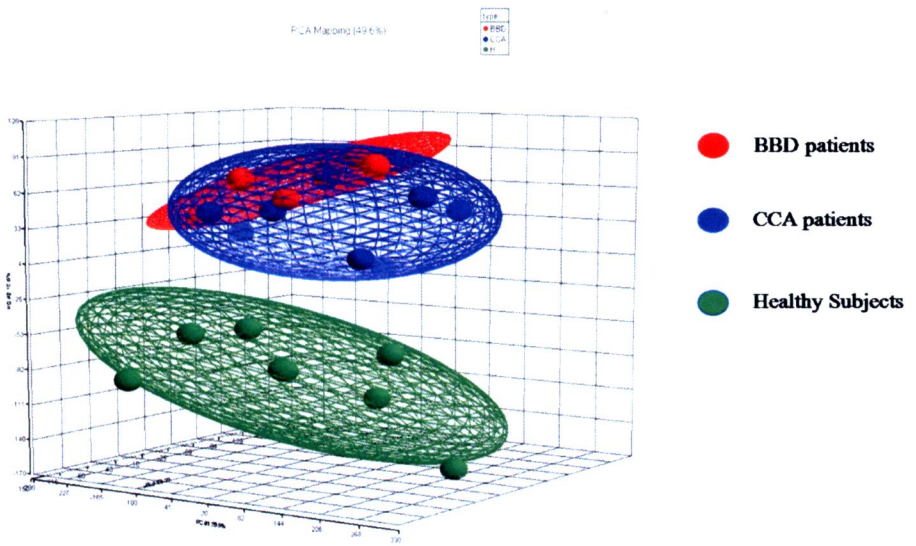

Figure 3.7 Principle component analysis for expression trend between CCA, BBD patients and healthy subjects. The Ellipsoid view showed that specimens were grouped by disease. The result showed 49.2 % of the system variance between the expression of peripheral blood leukocytes from 9 CCA patients (red dots) and 8 healthy subjects (blue dots)

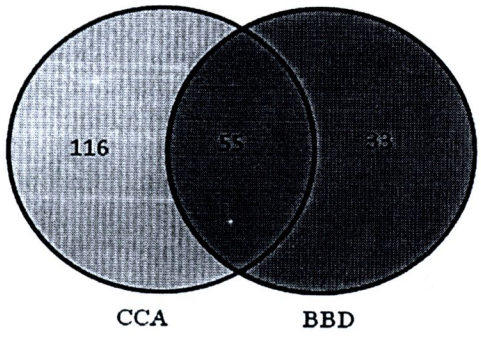

Figure 3.8 Venn diagram analysis for identifying the common and unique set of differentially expressed genes between CCA and BBD patients.

3.4 Discussion

Gene expression profiling in primary tumor has been repeatedly demonstrated to be different from the normal tissues and seems to be the potential diagnostic signature or to correlates with overall survival in cancer patients (Beer et al., 2002; van 't Veer et al., 2003; Wigle et al., 2002). Because of their greater accessibility, expression profiling in surrogate tissue, such as peripheral blood cells, are of interest for determining whether their expression signature may predict clinical outcomes in cancer patients. This study demonstrates that peripheral blood leukocytes (PBLs) from CCA patients did have a unique set of expression profile distinct from those of benign biliary disease patients and healthy persons. Moreover, the gene signature associated with CCA patients can potentially be a prognostic marker for CCA. Our results made several points as discussed below.

A comparison with those of healthy subjects, 177 differentially expressed genes were identified, of which 117 were up and 60 were down-regulated in CCA-PBLs. This result indicated that gene expression profiles of PBLs could reliably distinguish CCA patients from healthy subjects. In addition, the biological functions of these 177 genes were shown to involve in the host immune response and associate with tumor progression mediated by CCA peripheral leukocytes. For example, expression of angiogenic chemokines (CXCL2, CXCL3 and CXCL8/IL8) that are potent promoters of angiogenesis, and mediated angiogenic activity were up-regulated whereas CXCL10 (an interferon-inducible chemokines), a potent inhibitor of angiogenesis was suppressed in CCA-PBLs. In melanoma, the persistent expression of CXCL1, 2, or 3 in immortalized murine melanocytes transformed their phenotypes to one with anchorage-independent growth *in vitro* and ability to form tumor *in vivo* in immune-incompetent mice with highly vascularization. On the other hand, when tumors were depleted of CXCL1, 2, or 3 there was a marked reduction of tumor-derived angiogenesis directly related to inhibit tumor growth (Luan et al., 1997; Owen et al., 1997). CXCL8/IL8 is markedly elevated in tumor area that was observed to correlate with neovascularization and inversely correlated with survival of the patients with ovarian carcinoma and non-small-lung cell carcinoma (Smith et al., 1994; Yoneda et al., 1998). CXCL10, mediate its angiostatic activity via CXCR3 on endothelium (Strieter et al., 2006). These evidences suggest that up-regulation of

angiogenic CXC chemokine (CXCL2, 3 and 8) and down-regulation of angiostatic CXC chemokine (CXCL10) in CCA-PBLs may reflect of the disease for promoting tumor angiogenesis and/or inducing immune cell trafficking into tumor area. Chemokines are already known to be differentially expressed in polarized macrophage (e.g., CXCL10 for classical activated macrophages or M1; CCL17 for alternative macrophage or M2) (Martinez et al., 2006). In the present study, the down regulation of CXCL10 in CCA-PBLs may associate with Th 2 response which may promote monocytes/macrophage into M2. These M2 types of macrophages are more related to suppression of immune response and are well known function to support tumor progression.

There was another group of differentially expressed genes in CCA-PBLs that were involved in proteolytic process of extracellular matrix degradation and supported tumor invasion including plasminogen activator urokinase (PLAU), matrix metalloproteinase 9 (MMP9), serpin peptidase inhibitor clade B member 2, and catepsin L. This data is corresponded with the previous data in chapter VI that tumor associated macrophages (TAMs) in CCA tissues especially at leading edge of the tumor expressed PLAU and MMP9 proteins. In addition, the patients with high density of MMP9 and PLAU expressing TAMs in CCA tissues were associated with reducing overall survival after surgical resection. Not only chemokines and proteases but also growth factor related genes (epiregulin, EREG; transforming growth factor Beta1, TGFβ1) that were regulated in PBLs from CCA patients to promote tumor progression. A unique expression profile of PBLs found in CCA patients indicates that PBLs react to the microenvironment by capturing in a complex transcriptional response measured through profiling (Burczynski et al., 2005).

In general, tumor microenvironment which is largely orchestrated by variety of stromal cells, especially leukocytes co-opted some of signaling molecule such as cytokine, chemokine, growth factors, their receptor for invasion, migration, and metastasis of tumor cells (Coussens, Werb, 2002). This assumption can be supported by the current data that tumor promoting related genes - namely *PLAU, SERPINB2, CTSL, VEGFA, EREG,* and *CXCL3* in the transcriptome of CCA-PBLs were mainly from monocytes and lymphocytes. These verifications indicate that monocytes and

lymphocytes may serve as a sensitive monitor which may response or reflect the progression of CCA.

The expression profile, which is typically composed of several genes can be used as a platform to select a small set of candidate genes for verification based on their relative discriminatory power between disease and controls. Recently, a small set of genes from differential expression profiles of several cancers was used as molecular signature for tumor diagnosis and prognosis, including urinary bladder cancer, breast cancer, and adrenocortical tumor and intrahepatic cholangiocarcinoma (de Reynies et al., 2009; Ma et al., 2006; Ma et al., 2008; Nishino et al., 2008; Osman et al., 2006). In the present study, a set of proteolytic related genes (*PLAU, SERPINB2, CTSL)* were identified as signature for prognostic predictor of CCA. By computing into "risk score" or "prognostic index", it had significant power to predict the prognosis of CCA patients. In addition, the multivariate Cox regression model for survival indicated that the risk score was an independent predictor of survival for CCA.

Comparison of gene expression profiling of PBLs clearly identified differentially expressed genes, the expression levels of which were altered in PBLs from CCA patients relative to PBLs of healthy subjects. The small set of proteolytic related genes was also shown to relate with prognosis of the patients. In addition, comparison of gene expression profile of PBLs between CCA patient and BBD patients after subtracting with gene set from healthy subjects gave different data set. A set of 54 genes was found commonly in both patient groups. These overlapping genes may be an immunologic response associated with biliary tract disease such as chronic inflammation and may represent the generalized phenomenon or early pathological process of the lesion at biliary tract. On the other hand, the 116 genes which was differentially expressed only in PBLs from CCA patients, represents the molecular signature related to CCA. In summary, this study supports the idea of using blood transcriptome as an accessible surrogate monitor of tissue and system that are not easily obtained by standard approach.

In the technical point of view, Since several studies can demonstrate the presence of tumor cells in peripheral blood using high sensitivity array or RT-PCR (Lacroix, 2006; Paterlini-Brechot, Benali, 2007). It is possible that the PBLs from

CCA patients may contaminate by circulating tumor cells and this may affect gene expression pattern in the present study. To prove this, three database of gene expression profile data from primary tissues of CCA (Jinawath et al., 2006; Obama et al., 2005; Wang et al., 2006) were checked for similarity with those from blood transcriptome. No highly expressed transcripts of primary tumor CCA transcriptome were found in our database and *vice versa.* In addition, differentially expressed gene in CCA-PBLs doesn't appear any epithelial-related genes. These evidences support that the CCA-associated genes found in peripheral leukocytes did not contribute from circulating tumor cells and can reflect the disease.

The data set obtained from the blood transcriptome of these three subject groups can initiate several valuable data. It is possible in further study that many independent set of the differentially expressed genes may develop a better discrimination power for diagnostic, prognostic or predictive parameters in clinical assay of CCA and other pathological disorders at biliary system. In conclusion, the informative gene expression profiles of PBLs that could reliably distinguish CCA patients from healthy subjects were indentified. On basis of this data, a small set of candidate genes differentially expressed in CCA-PBLs had potential of being a predictor for survival of CCA patients. Thus, this study showed that readily accessible PBLs can be used as a surrogate tissue to assess cancer patients with greater chances for long or short time survival. However, the potential of this approach is needed to be evaluated more extensively in a larger sample size for a better discrimination power before applying in clinical practice.