American Journal of Pathology, Vol. 163, No. 1, July 2003
Copyright © American Society for Investigative Pathology

Two Newly Characterized Germinal Center B-Cell-Associated Genes, *GCET1* and *GCET2*, Have Differential Expression in Normal and Neoplastic B Cells

Zenggang Pan,* Yulei Shen,* Cheng Du,* Guimei Zhou,* Andreas Rosenwald,† Louis M. Staudt,† Timothy C. Greiner,* Timothy W. McKeithan,* and Wing C. Chan*

From the Department of Pathology and Microbiology, University of Nebraska Medical Center, Omaha, Nebraska; and the Metabolism Branch,† National Cancer Institute, Bethesda, Maryland*

A group of genes are highly expressed in normal germinal center (GC) B cells and GC B-cell-derived malignancies based on cDNA microarray analysis. Two new genes, *GCET1* (germinal center B-cell expressed transcript 1) and *GCET2*, were cloned from selected expressed sequence tags (IMAGE clone 1334260 and 814622, respectively). *GCET1* is located on chromosome 14q32 and has four splicing isoforms, of which the longest one is 1787 bp and encodes a 435-amino acid protein. *GCET2* is located on 3q13.13, and the cloned fragment is 3270 bp, which encodes a protein of 178 amino acids. Blast search showed that *GCET1* has a highly conserved serine proteinase inhibitor (SERPIN) domain and is located on a chromosomal locus containing seven other SERPIN family members. *GCET2* is a likely homologue of the mouse gene M17, a GC-expressed transcript. Analysis of the *GCET2* protein sequence indicated that it may be involved in signal transduction in the cytoplasm. Northern blot and real-time polymerase chain reaction analyses confirmed that *GCET1* is highly restricted to normal GC B cells and GCB-cell-derived cell lines. Although *GCET2* is also a useful marker for normal and neoplastic GC B cells, it has a wider range of expression including immature B and T cells. Real-time polymerase chain reaction assay showed that both *GCET1* and *GCET2* are preferentially expressed in follicular lymphoma and diffuse large B-cell lymphoma with GC B-cell differentiation, confirming previous microarray gene expression analysis, but neither one is entirely specific. Multiple markers are necessary to differentiate the GCB from the activated B-cell type of diffuse large B-cell lymphoma with a high degree of accuracy. *(Am J Pathol 2003, 163:135–144)*

In the T-cell regions of secondary lymphoid organs, naïve B lymphocytes have two alternative fates after activation. Some of them will differentiate into primary plasma cells whereas others will migrate to form germinal centers (GC) and undergo further development.[1,2] GC B cells proliferate rapidly and at the same time acquire the capacity to undergo somatic hypermutation and isotype switching. A stringent process of selection also occurs with survival only of cells expressing high-affinity antibody to the selecting antigen.[3–5] This complex maturation process is associated with the acquisition of a unique gene expression profile that distinguishes GC B cells from B cells at other stages of maturation. A number of known genes and uncharacterized genes that are represented as expressed sequence tags (ESTs) are highly expressed in normal and malignant GC B cells, but not in resting or activated peripheral blood B cells.[6,7] Gene expression profiling studies on a large number of diffuse large B-cell lymphomas (DLBCLs) have shown the presence of a unique subset with overexpression of genes in the GC B-cell signature.[6] This type of lymphoma has a better prognosis than other types that do not exhibit this expression signature. The gene expression profile is very heterogeneous even within individual subsets of DLBCL, but a limited number of genes in the GC B-cell signature have been identified to be most diagnostic of the GCB subtype of DLBCL.[7] Two of the ESTs, which are important predictors of prognosis on DLBCL and have the highest *P* values in differentiating GCB-DLBCL from other subtypes,[7] were selected for cloning. The characteristics of these two genes are described below.

Materials and Methods

EST Selection and cDNA Assignment

Two ESTs were selected from the NCI_CGAP_GCB1 library based on the expression profiling data of our pre-

Supported by the National Institutes of Health (grant U01-CA84967-04).

Accepted for publication April 1, 2003.

Address reprint requests to Wing C. (John) Chan, M.D., Department of Pathology, University of Nebraska Medical Center, 983135 Nebraska Medical Center, Omaha, NE 68198. E-mail: jchan@unmc.edu.

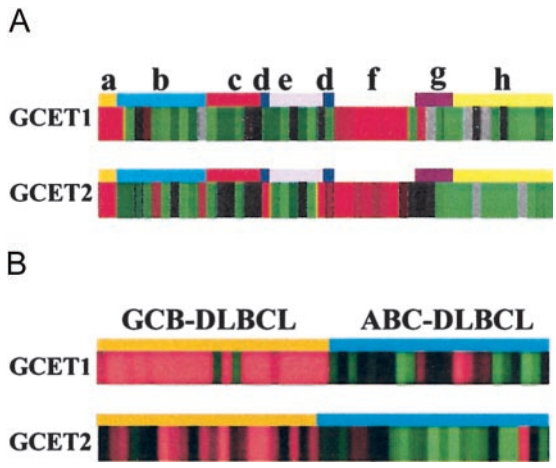


Figure 1. Microarray gene expression profiles of *GCET1* and *GCET2*. **A:** Both *GCET1* and *GCET2* are highly expressed in normal GC B cells (**a**) and FL (**f**); down-regulated in activated and resting peripheral B cells (**b**, **g**), resting and activated T cells (**c**, **e**), and CLL (**h**). **B:** They have high expression levels in GCB-DLBCL but low expression in the ABC group.

vious cDNA microarray study of normal and neoplastic B cells.[6,7] These ESTs are highly expressed in normal GC B cells, GCB-derived DLBCLs, and follicular lymphomas (FLs) (Figure 1).

Using NCBI Genomic BLAST, the chromosomal locations of the selected ESTs were determined. The 50-kb chromosomal regions flanking each of the ESTs were searched to find other human ESTs close to or overlapping with the selected ESTs. Polymerase chain reaction (PCR) primers were designed to determine the presence and orientation of these ESTs on the cDNA (from the lymphoma cell line DHL16; see below) containing the original EST. Also, primers were designed according to known EST sequences for rapid amplification of cDNA ends (RACE)[8,9] (Table 1).

mRNA Isolation and cDNA Synthesis

To clone the corresponding genes from selected ESTs, cytoplasmic total RNA was isolated from a cultured GCB-derived lymphoma cell line, DHL16, using RNeasy Midi Kit (Qiagen, Valencia, CA). Cytoplasmic total RNA was further treated with DNase I (Promega, Madison, WI), and then mRNA was isolated using the Poly(A) Pure mRNA purification kit (Ambion, Austin, TX).

cDNA was synthesized using SMART RACE cDNA amplification kit (Clontech, Palo Alto, CA) following the manufacturer's protocol. A universal adaptor, SMART IIA oligonucleotide, hybridized to the three cytosine residues added to the 3′ end of the cDNA during reverse transcription (RT) and acted as a template for further extension of the cDNA strand, with incorporation of a universal SMART primer sequence.

Table 1. Primers for 5′ RACE in Cloning *GCET1*

Gene specific primer designed from EST AA805575	CTGCAATAGAGGTGCCTAAC
SMART II A oligonucleotide sequence	AAGCAGTGGTATCAACGCAGAGTACGCGGG
Universal primer A Mix (UPM)	Long CTAATACGACTCACTATAGGGCAAGCAGTGGTATCAACGCAGAGT
	Short CTAATACGACTCACTATAGGGC
Nested universal primer A (NUP)	AAGCAGTGGTATCAACGCAGAGT

Primers for Cloning *GCET2* Utilizing EST Sequences

dbEST ID	GenBank acc. no.		
CS0DG007YD15 (5′)	AL560376	Forward primer	AGTCAGGAGTTGCCCTGTCAG
		Reverse primer	TAGGTCTGGTCAACATTGTCC
814622 (5′)	AA480985	Forward primer	TTAAGAAGAACGATCCTTGGAG
		Reverse primer	ATCAGGACACATAGAGGAGTGC
3064524	AW576133	Forward primer	ATTGGGTATCTTGAAGCACTCC
		Reverse primer	ACTGGACCATTGTGGAAGTAG
684286 (5′)	AA236080	Forward primer	CTGCATCTTGGTATCTCTTCCTC
		Reverse primer	AGAAGCTGAGCCTCCAAGTAG

Primers and Probes Used for Real-Time PCR

GCET1	Forward primer	TGTCAGTGAAGAGGGCACTGA
	Reverse primer	GACCATCCTTCGATCGGACTAT
	Probe	FAM-CCACAGCAGCTACCGCCACCAA-TAMRA
GCET2	Forward primer	AATATGAACTTCTCATGCCTCACAGA
	Reverse primer	AATGCTAGTCCAGCCACTTCACTA
	Probe	FAM-TCTGCAACAGCCACGTCCACTTATGG-TAMRA
GAPDH	Forward primer	GAAGGTGAAGGTCGGAGTC
	Reverse primer	GAAGATGGTGATGGGATTTC
	Probe	JOE-CAAGCTTCCCGTTCTCAGCC-TAMRA
ACTB	Forward primer	TGCCGACAGGATGCAGAAG
	Reverse primer	GCCGATCCACACGGAGTACT
	Probe	FAM-TCAAGATCATTGCTCCTCCTGAGCGC-TAMRA
B2M	Forward primer	TGTGCTCGCGCTACTCTCTCT
	Reverse primer	TCTGCTGGATGACGTGAGTAAAC
	Probe	FAM-CCTGGAGGCTATCCAGCGTACTCCAAA-TAMRA

RACE, Cloning, and Sequencing

5′ RACE was performed using the SMART RACE cDNA amplification kit. In this reaction, a universal primer mix containing a long and a short primer was applied (Table 1). The long primer, combining with the gene-specific primer, amplified the specific template in the initial cycles. With the accumulation of the gene-specific template, the higher concentration short primer acted as a main universal primer for further amplification. Thus, this PCR-suppression effect ensures that the gene-specific cDNA is preferentially amplified with dramatically reduced background.[10]

5′ RACE was performed for 35 cycles at 94°C for 30 seconds, 62°C for 40 seconds, and 70°C for 5 minutes using a T-Personal thermocycler (Biometra, Horsham, PA). The amplified products were electrophoresed in a 1% agarose gel and the specific bands were purified using the QIAquick gel extraction kit (Qiagen) and cloned into pCR2.1 vector using the TA cloning kit (Invitrogen, Carlsbad, CA). Individual white clones were randomly picked from plates and cultured in 5 ml of Amp+ LB medium, and plasmids were extracted using Qiagen QIAprep Spin Miniprep Kit (Qiagen). DNA sequencing was performed using a Perkin-Elmer Applied Biosystems model 377 DNA sequencer (Perkin-Elmer, Emeryville, CA). The sequencing data were analyzed using the GCG program and BLAST search of the GenBank database.

Reverse Transcriptase (RT)-PCR and Real-Time Quantitative RT-PCR

The expression of *GCET1* and *GCET2*, relative to three housekeeping genes, GAPDH (glyceraldehyde-3-phosphate dehydrogenase), ACTB (β-actin), and B2M (β-2-microglobulin), were measured using an ABI Prism 7700 sequence detector (Applied Biosystems, Foster City, CA). Primers and probes for *GCET1*, *GCET2*, ACTB, and B2M were designed using Primer Express software (Applied Biosystems), as shown in Table 1. Probe and primers for GAPDH were purchased from Applied Biosystems.

The probes for *GCET1*, *GCET2*, ACTB, and B2M were labeled with FAM (carboxyfluorescein) at the 5′ ends as the reporter dye and TAMRA (carboxytetramethylrhodamine) at the 3′ ends as the quencher dye. The GAPDH probe was labeled with JOE (carboxy-4,5-dichloro-2,7-dimethoxyfluorescein) as the reporter dye and TAMRA as the quencher dye.

Total RNA or mRNA was extracted from cell lines, clinical samples, and normal tissues: DHL16, SUDHL 6 (FL cell lines);[11] OCI-LY7, OCI-LY19 (GCB-DLBCL cell lines), OCI-LY3, OCI-LY10 (ABC-DLBCL cell lines);[12] BLCL (B-lymphoblastoid cell line derived in the authors' laboratory); U266 (myeloma cell line);[13] Nalm6 (pre-B-cell line);[14] L428 (Hodgkin lymphoma cell line);[15] NK92 (NK cell line)[16] and Jurkat (T-leukemia cell line);[17] lymphoma samples, 11 DLBCL, 10 FL, 6 chronic lymphocytic leukemia (CLL); nonneoplastic tissues, spleen, tonsil, thymus, and peripheral blood B cells. All these RNA samples were treated with DNase I, and then 500 ng to 3 μg of RNA was transcribed into cDNA using Superscript II-reverse transcriptase (Life Technologies, Inc., Grand Island, NY) following the manufacturer's directions. A corresponding PCR on RNA from each sample (non-RT control) was performed to verify the absence of genomic DNA contamination.

Real-time PCR was performed with the TaqMan Universal PCR Master Mix (Applied Biosystems) using 5 μl of cDNA with or without dilution in a 25-μl reaction mixture with a final concentration of 200 nmol/L of probe and 200 nmol/L of primers. After incubation at 50°C for 2 minutes, AmpliTaq Gold was activated by incubation at 95°C for 10 minutes. Forty PCR cycles were performed with denaturation at 95°C for 15 seconds, and combined annealing and extension at 60°C for 1 minute. Serial dilutions of cDNA from DHL16 were used to construct standard curves for the target genes (*GCET1* and *GCET2*) and the endogenous reference genes (GAPDH, ACTB, and B2M). For each unknown sample, the relative amount of target cDNAs and reference cDNAs applied to the PCR reaction system were calculated using linear regression analysis from the corresponding standard curves. Then the normalized expression level of the target gene in each sample was calculated by dividing the quantity of the target transcript with the quantity of corresponding reference transcript. The normalized values of the target transcript were used to compare its relative expression levels in different samples.

Laser Capture Microdissection

The three B-cell compartments (GC, mantle zone, and marginal zone) in 5-μm-thick frozen section of reactive tonsils or spleens were isolated using laser-captured microdissection with Arcturus PixCell II system (Arcturus Engineering, Mountain View, CA). The sections on plain glass slides were fixed with 70% ethanol for 30 seconds, washed in diethyl pyrocarbonate-treated water, stained with Mayer's hematoxylin for 30 seconds, followed by another water wash. The slide was then dehydrated with 70%, 95%, and 100% ethanol for 10 seconds each. Finally, the slide was passed through xylene twice, each for 30 seconds. A consecutive section was immunostained for CD3 to guide the dissection. Only well-defined GC, mantle zone, and marginal zone were dissected to avoid contamination with other populations. Cells were captured at the 15-μm laser setting with the laser pulse at 60 mW for 200 ms. DNA-free RNA was extracted using the Absolutely RNA MicroRNA isolation kit (Stratagene, La Jolla, CA) according to the manufacturer's instruction. After DNase I digestion, the column-adsorbed RNA was eluted into 30 μl of 10 mmol/L Tris-HCl (pH 7.5) and stored at −80°C.

Northern Blot

Total RNAs from cell lines were extracted with Trizol reagent (Invitrogen) according to the manufacturer's instruction. RNAs were separated by agarose gel contain-

ing formaldehyde and transferred to positively charged nylon membrane (Immobilon; Millipore Corp., Billerica, MA) with a vacuum transfer apparatus (Oncor Corp., Gaithersburg, MD). Plasmid clones with full-length open reading frame were labeled by random hexamer with Prime-a-Gene labeling kit (Promega, Madison, WI). Twenty-five ng of the PCR product were heated at 100°C for 3 minutes and then chilled on ice. Reaction mix and ^{32}P-dCTP (3000 Ci/mmol, 10 mCi/ml; ICN Biomedicals, Irvine, CA) were added using the manufacturer's protocol. After a 1-hour incubation at room temperature, the unincorporated ^{32}P-dCTP was removed by a Bio-P6 spin column (Bio-Rad, Hercules, CA). Hybridization was performed in a buffer containing 0.125 mol/L NaPO4, pH 6.8, 0.25 mol/L NaCl, 7% sodium dodecyl sulfate, 1 mmol/L ethylenediaminetetraacetic acid, 10% PEG 8000, and 50% formamide at 50°C overnight. The membrane then was washed three times with a 0.1× standard saline citrate and 0.4% sodium dodecyl sulfate solution at 70°C, and exposed to a Kodak X-OMAT film (Kodak, Rochester, NY) at −70°C in a cassette with intensifying screen.

Results

Selecting Genes that Are Preferentially Expressed in GC B Cells and GCB-Derived Lymphomas

Diffuse large B-cell lymphoma (DLBCL) accounts for ~40% of non-Hodgkin's lymphoma cases, and cDNA microarray assays have identified three gene expression subgroups of DLBCL—GC B cell-like (GCB), activated B cell-like (ABC), and type 3 DLBCL.[7] A group of genes are preferentially expressed in the GCB-DLBCL subtype, including many known markers of GC differentiation, such as BCL-6, CD10, CD38, nuclear factor A-myb, and OGG1, as well as a number of uncharacterized genes represented as ESTs.[6,7,18] Based on the cDNA microarray findings, two uncharacterized genes (IMAGE 1334260 and 814622) were selected from the GCB signature for our study.

According to the cDNA microarray data,[6,7] these two genes have similar expression patterns in the selected normal and malignant lymphoid tissues studied. They are highly expressed in normal GC B cells and GCB-derived lymphomas, including FL and the GCB-DLBCL; and they are poorly expressed in both activated and resting peripheral blood B cells, resting and activated T cells, B-CLL, and the ABC-DLBCL (Figure 1).

BLAST search of the GenBank database did not reveal any known genes matched to our selected EST clones. Searching the 50-kb genomic DNA region flanking either end of the selected ESTs revealed many additional clones in the human EST database, and the majority of these clones are from libraries constructed either from normal GC B cells or GCB-derived lymphoma cell lines.

Sequence Data, Homology, and Chromosomal Localization of *GCET1* and *GCET2*

Two new genes were cloned from the selected EST clones (IMAGE 1334260 and 814622) and named *GCET1* and *GCET2*, respectively. 5′ RACE for *GCET1* revealed two major products of ~1.5 and 1.8 kb (Figure 2A). After cloning and sequencing, we identified four splicing variants of 1787 bp, 1451 bp, 1488 bp, and 1547 bp, which contain an open reading frame of 435 amino acids, 335 amino acids, 337 amino acids, and 286 amino acids, respectively (Figure 2D).

GCET1 contains the serine proteinase inhibitor (SERPIN) motif VSFNRTFLMMI. A Blast search of the human and animal protein databases using *GCET1* protein as a query indicated that *GCET1* contains the conserved SERPIN domain and has the highest homology (50 to 63%) to thyroxine-binding globulin from a variety of species, including human, sheep, rat, and mouse. The longest isoform, *GCET1*A (Figure 2E), was found to be identical to a previously reported gene, centerin,[19] which is a member of the SERPIN family, but no nucleotide sequence and protein sequence had been submitted to GenBank previously.

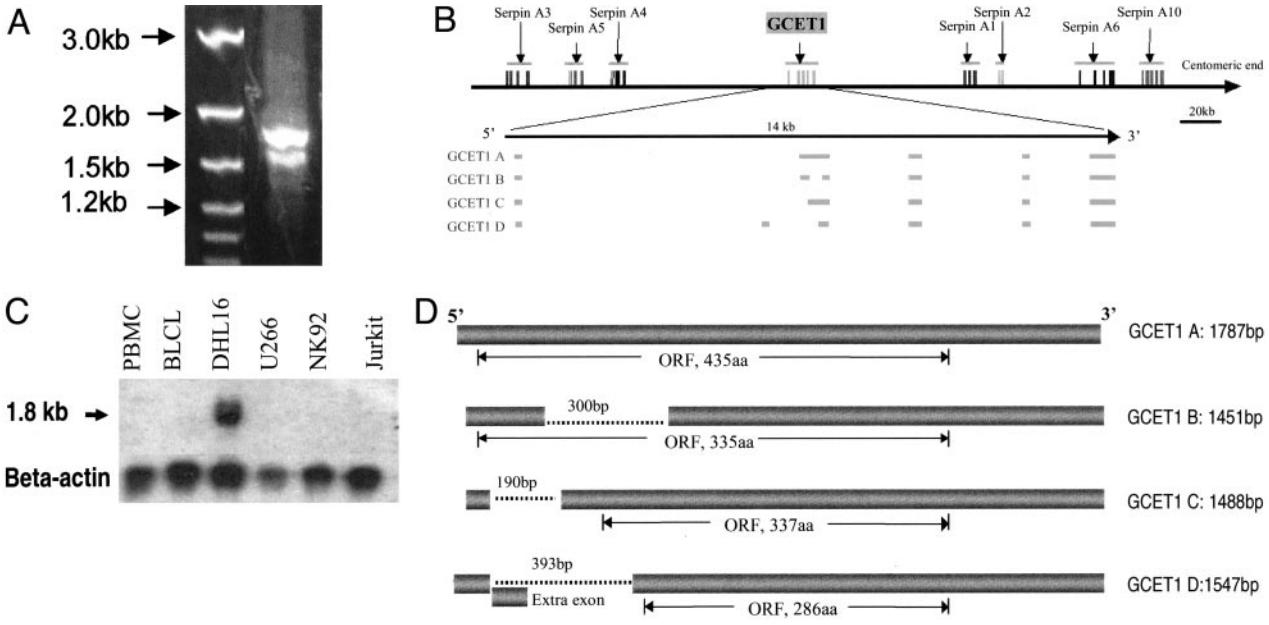
GCET1 is located on chromosome 14q32 and spans ~14 kb of genomic DNA. The longest isoform, *GCET1*A, has 5 exons (Figure 2B). On 14q32, close to *GCET1* are seven other SERPIN members, including SERPIN A1, A2, A3, A4, A5, A6, and A10 (Figure 2B). Taken together, this indicated that *GCET1* belongs to the SERPIN family.

EST clone CS0DG007YD15 and 3064524 were confirmed to be co-localized with clone 814622 and 684286 on the same cDNA by PCR. The PCR products were cloned and sequenced. Multiple attempts at extending the sequence by 5′ or 3′ RACE beyond the most 5′ or 3′ EST clone failed to yield any additional sequence, indicating that we had or were very close to having the full-length sequence of the *GCET2* transcript.

The *GCET2* cDNA we cloned is 3270 bp long, with 6 exons, encoding a 178-amino acid protein. It is located on chromosome 3q13.13 spanning ~15 kb of genomic DNA (Figure 3A). Using *GCET2* protein as a query to various databases showed a 34% homology identity to a human protein (accession no. AAH24174) (Figure 4A) encoded by a predicted gene located on 1q44, and 57% homology to the mouse M17 protein (germinal center expressed transcript) (Figure 4B), whose function has not been defined.[20] (*Note:* GenBank accession numbers: *GCET1A*, AY220118; *GCET1B*, AY220119; *GCET1C*, AY220120; *GCET1D*, AY220121; *GCET2*, AY212246.)

Northern Blot

Initial Northern blot hybridization of *GCET1* showed a dominant band of ~1.8 kb only in the cell line DHL-16 (Figure 2C). *GCET2* has a dominant transcript of ~1.5 kb and another transcript around 3.3 kb. Both transcripts were detected in DHL16 and the pre-B cell line Nalm-6 (Figure 3B). Faint signals were also seen with the NK-cell line, NK92. Analysis of *GCET2* gene sequence showed



E

```
  1 gagctggagggagaggagagattaaagtgaggagagagctacaaccaagtaagcaagtgt
 61 cagggctcaccaaccatgcaaggacagggcaggagaagaggaacctgcaaagacatattt
                M  Q  G  Q  G  R  R  R  G  T  C  K  D  I  F
121 tgttccaaaatggcatcttacctttatggagtactctttgctgttggcctctgtgctcca
    C  S  K  M  A  S  Y  L  Y  G  V  L  F  A  V  G  L  C  A  P
181 atctactgtgtgtccctggccaatgcccccagtgcataccccgcccttcctccacaaag
    I  Y  C  V  S  L  A  N  A  P  S  A  Y  P  R  P  S  S  T  K
241 agcacccctgcctcacaggtgtattccctcaacaccgactttgccttccgcctataccgc
    S  T  P  A  S  Q  V  Y  S  L  N  T  D  F  A  F  R  L  Y  R
301 aggctggttttggagaccccgagtcagaacatcttcttctcccctgtgagtgtctccact
    R  L  V  L  E  T  P  S  Q  N  I  F  F  S  P  V  S  V  S  T
361 tccctggccatgctctcccttggggcccactcagtcaccaagacccagattctccagggc
    S  L  A  M  L  S  L  G  A  H  S  V  T  K  T  Q  I  L  Q  G
421 ctgggcttcaacctcacacacacaccagagtctgccatccaccagggcttccagcacctg
    L  G  F  N  L  T  H  T  P  E  S  A  I  H  Q  G  F  Q  H  L
481 gttcactcactgactgttcccagcaaagacctgaccttgaagatgggaagtgccctcttc
    V  H  S  L  T  V  P  S  K  D  L  T  L  K  M  G  S  A  L  F
541 gtcaagaaggagctgcagctgcaggcaaatttcttgggcaatgtcaagaggctgtatgaa
    V  K  K  E  L  Q  L  Q  A  N  F  L  G  N  V  K  R  L  Y  E
601 gcagaagtcttttctacagatttctccaacccctccattgcccaggcgaggatcaacagc
    A  E  V  F  S  T  D  F  S  N  P  S  I  A  Q  A  R  I  N  S
661 catgtgaaaaagaagacccaagggaaggttgtagacataatccaaggccttgaccttctg
    H  V  K  K  K  T  Q  G  K  V  V  D  I  I  Q  G  L  D  L  L
721 acggccatggttctggtgaaccacattttctttaaagccaagtgggaaaagccctttcac
    T  A  M  V  L  V  N  H  I  F  F  K  A  K  W  E  K  P  F  H
781 cctgaatatacaagaaaggacttcccattcctggtgggcgagcaggtcactgtgcatgtc
     P  E  Y  T  R  K  D  F  P  P  F  L  V  G  E  Q  V  T  V  H  V

841 cccatgatgcaccagaaagagcagttcgcttttggggtggatacagagctgaactgcttt
     P  M  M  H  Q  K  E  Q  F  A  F  G  V  D  T  E  L  N  C  F
901 gtgctgcagatggattacaagggagatgccgtggccttctttgtcctccctagcaagggc
     V  L  Q  M  D  Y  K  G  D  A  V  A  F  F  V  L  P  S  K  G
961 aagatgaggcaactggaacaggccttgtcagccagaacactgagaaagtggagccactca
     K  M  R  Q  L  E  Q  A  L  S  A  R  T  L  R  K  W  S  H  S
1021 ctccagaaaaggtggatagaggtgttcatccccagattttccatttctgcctcctacaat
     L  Q  K  R  W  I  E  V  F  I  P  R  F  S  I  S  A  S  Y  N
1081 ctggaaaccatcctcccgaagatgggcatccaaaatgcctttgacaaaaatgctgatttt
     L  E  T  I  L  P  K  M  G  I  Q  N  A  F  D  K  N  A  D  F
1141 tctggaattgcaaagagagactccctgcaggtttctaaagcaacccacaaggctgtgctg
     S  G  I  A  K  R  D  S  L  Q  V  S  K  A  T  H  K  A  V  L
1201 gatgtcagtgaagagggcactgaggccacagcagctaccgccaccaagttcatagtccga
     D  V  S  E  E  G  T  E  A  T  A  A  T  A  T  K  F  I  V  R
1261 tcgaaggatggtccctcttacttcactgtctccttcaataggaccttcctgatgatgatt
     S  K  D  G  P  S  Y  F  T  V S F N R T F L M M I
1321 acaaataaagccacagacggtattctctttctagggaaagtggaaaatcccactaaatcc
     T  N  K  A  T  D  G  I  L  F  L  G  K  V  E  N  P  T  K  S
1381 taggtgggaaatggcctgttaactgatggcacattgctaatgcacaagaaataacaaacc
1441 acatccctctttctgttctgagggtgcatttgaccccagtggagctggattcgctggcag
1501 ggatgccacttccaaggctcaatcaccaaaccatcaacagggaccccagtcacaagccaa
1561 cacccattaaccccagtcagtgccctttccacaaattctcccaggtaactagcttcatg
1621 ggatgttgctgggttaccatatttccattccttggggctcccaagaatggaaatacgcca
1681 acccaggttaggcacctctattgcagaattacaataacacattcaataaaactaaaatat
1741 gaattcatctgtcaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 1787
```

Figure 2. A: 5′ RACE for *GCET1* reveals two major bands on agarose gel electrophoresis. **B:** Distribution of *GCET1*-splicing isoforms and other known SERPIN family members on chromosome 14q32. **C:** Northern blot of *GCET1* shows a transcript around 1.8 kb in DHL16. **D:** Alignment of *GCET1*-splicing isoforms. *GCET1*A is 1787 bp long encoding a protein of 435 amino acids. *GCET1*B is 300 bp shorter at the 5′ end encoding a protein of 335 amino acids. *GCET1*C is 190 bp shorter at the 5′ end encoding a protein of 337 amino acids. *GCET1*D has 393 bp missing at the 5′ end compared with *GCET1*A and encodes a protein of 286 amino acids, but it has an extra exon as shown. **E:** Nucleotide and protein sequence data of *GCET1*A. **Bolded** sequence is the conserved SERPIN motif and the **underlined** nucleotides represent the possible element for polyadenylation.

five AATAAA polyadenylation signals located after 1540 bp (Figure 3C). According to the Northern blot results and sequence data, one or more of the proximal AAUAAA elements are likely to be functional, yielding the shorter transcript.

Real-Time PCR Analyses of the Expression of GCET1 and *GCET2*

Representative real-time PCR standard curves for ACTB, *GCET1*, and *GCET2* are shown in Figure 5. Although the assays were very accurate and reproducible, the expression of a housekeeping gene was quite variable in different cells and tissues. Normalization using any single housekeeping gene gave occasional spuriously high or low values for the samples studied. We therefore decided to use the median values of the gene expression levels for each sample to avoid these spurious values.

Quantitative analyses using real-time PCR showed that *GCET1* expression was highly restricted to GC B cells and GCB-derived lymphomas. Microdissected GCs had a 30-fold and 4-fold higher expression of *GCET1* than the mantle zone and marginal zone, respectively (Figure 6A). Its expression was lower in whole tonsil and spleen,

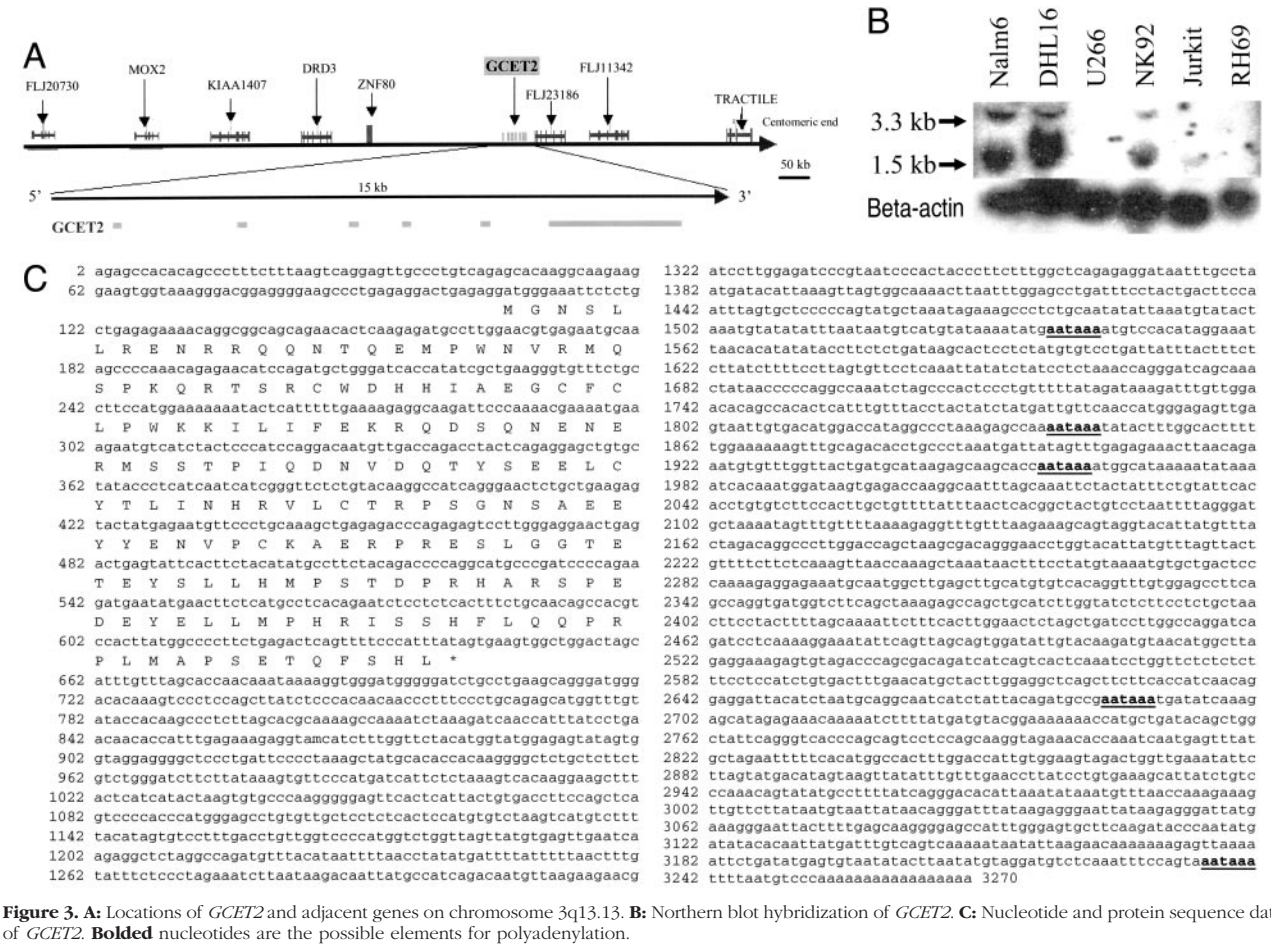


Figure 3. A: Locations of *GCET2* and adjacent genes on chromosome 3q13.13. **B:** Northern blot hybridization of *GCET2*. **C:** Nucleotide and protein sequence data of *GCET2*. **Bolded** nucleotides are the possible elements for polyadenylation.

which was likely because of the dilutional effect of other cells in these two organs. Compared with DHL16, the expression levels were <1% in the thymus, peripheral blood mononuclear cells, and cell lines, including L428, U266, Nalm6, BLCL, and NK92 (Figure 6A). The expression of *GCET1* in 11 of 12 FL samples (DHL16 and SUDHL-6, and 9 clinical FL samples) was at least 1.5 times higher than that in any CLL or ABC-DLBCL samples (median value for all 12 FL samples, 0.49) (Figure 6B). All five CLL samples had low expression of *GCET1* (median value, 0.015). The expression levels varied from moderate to high in eight GCB-DLCBL samples (OCI-LY7 and OCI-LY19, and six clinical samples) (median value, 0.13), and from low to moderate in seven ABC-DLBCL samples (OCI-LY3 and OCI-LY10, and five clinical samples) (median value, 0.025).

Similar to *GCET1*, *GCET2* was highly expressed in the normal GC with approximately 5-fold and 20-fold lower expression in the mantle zone and marginal zone, respectively. In contrast to *GCET1*, *GCET2* had high expression levels in Nalm6 (1.25-fold) and the thymus (0.8-fold) compared with DHL16, suggesting that *GCET2* may be up-regulated in immature B and T cells. *GCET2* also had moderate levels of expression in BLCL (0.5-fold) and L428 (0.3-fold) (Figure 6C). The expression pattern of *GCET2* in neoplastic samples was quite similar to *GCET1*. It was highly expressed in all 12 FLs (median value, 1.9) and 8 GCB-DLBCL samples (median value, 1.5), but 3 of the 7 ABC-DLBCL samples (median value for all 7 samples, 0.31) and 1 of the 5 CLL samples also had moderate to high expression (median value for all 5 samples, 0.28) (Figure 6D).

For both genes, we compared the real-time PCR data with our previous microarray findings in FLs, DLBCL samples, and corresponding cell lines.[6,7] We converted the real-time PCR values into pseudocolors for better visualization. The results of real-time PCR generally parallel the microarray findings with high expression in FL and GCB-DLBCL, and low expression in ABC-DLBCL (Figure 7). However, some overlap in the expression levels in these different entities was found in both systems.

Discussion

We have identified a number of ESTs that show differential expression in normal and neoplastic B cells based on our gene expression profiling data.[6,7,21] Two of the ESTs are preferentially expressed in GC B cells and are among the best classifiers of DLBCL of the GC B cell type.[6,7] We proceeded to clone the genes represented by these two ESTs by a combination of RACE and judicious linking of known ESTs to the cDNA of interest.

A

```
GCET2      1   MGNSLLRENRRQQNTQEMPWNVRMQSPKQRTSRCWDHHIAEGCFCLPWKKILIFEKR-QDS  60
               MGN LLR      + Q  P     +  ++R                  +++  FE++ QD
AAH27174   1   MGNYLLRKLSCLGENQKKPKKGNPDEERKR-----------------QEMTTFERKLQDQ  43

GCET2      61   QNENERMSSTPIQDNVDQTYSEELCYTLINHRVLCTRPSGNSAEEYYENVPCKAERPRES  120
                  +++ +SST  Q+N + + SEE+CYT+INH +   R S +S ++ YEN+     + R+
AAH27174   44   DKKSQEVSSTSNQENENGSGSEEVCYTVINH-IPHQRSSLSSNDDGYENIDSLTRKVRQF  102

GCET2      121 LGGTETEYSLLHMPSTDPRHARSPEDEYELLMPHRISSHFLQQPRPLMAPSETQFSHL  178
                 +ETEY+LL    + P    + E +YE++ PH
AAH27174   103 RERSETEYALLRTSVSRPCSC-THEHDYEVVFPH  135
```

B

```
GCET2  1   MGNSLLRENRRQQNTQEMPWNVRMQSPKQRTSRCWDH----HIAEGCFCLPWKKILIFEKR-QDS  60
           MGN L                      QRT+R +      H  EGC CLPWK I  F+ R Q+S
M17    1   MGNCL----------------------QRTTRYFRGWSCCHSVEGCSCLPWKNIRTFKAR-QES  41

GCET2  61   QNENERMSSTPIQDNVDQTYSEELCYTLINHRVLCTRPSGNSAEEYYENVPCKAERPRES  120
              +NE M+S P+QDN ++TY+EELCY L+ +H  +  RPS N AE +YEN+ KAER +ES
M17    42   PKQNEGMTSAPVQDNANETYTEELCYILVDHEAVRGRPSVNPAEGFYENISNKAERHKES  101

GCET2  121 LGGTETEYSLLHMPSTDPRHARSPEDEYELLMPHRISSHFLQQPRPLMAPSETQFSHL  178
             GTETEYS+L  PS  P+   S +DEYELLMP R SSH  QQPRPL  P ET FS+
M17    102 SRGTETEYSVLRFPSP-PQPLPSTDDEYELLMPSRFSSHAFQQPRPLTTPYETHFSY  157
```

Figure 4. A: Sequence alignment between human *GCET2* and AAH24174. The homology is 34% (46 of 135) (does not include positively charged amino acid residues). **B:** Homology between *GCET2* and mouse M17 is 57% (89 of 157) (does not include positively charged amino acid residues). **Bolded** amino acid sequences are the possible conserved tyrosine phosphorylation sites.

The first gene we cloned, *GCET1*, belongs to the SERPIN family of proteins and is located within a cluster of SERPIN genes on chromosome 14q32. This gene, called centerin,[19] was previously reported but its nucleotide and protein sequence has not been deposited in any of the public databases. Our studies showed that this gene has multiple isoforms with different predicted protein sequences, but the functional significance of these isoforms is not clear. Centerin is induced when B cells are stimulated with CD40L[19] and may play an important role in GC-B cell physiology.

The SERPINs are a large family of proteins, which distribute widely in the body and have various functions. Whereas some are secreted proteins that inhibit extracellular proteases, others seem to function intracellularly. A number of SERPINs have been investigated in their pos-

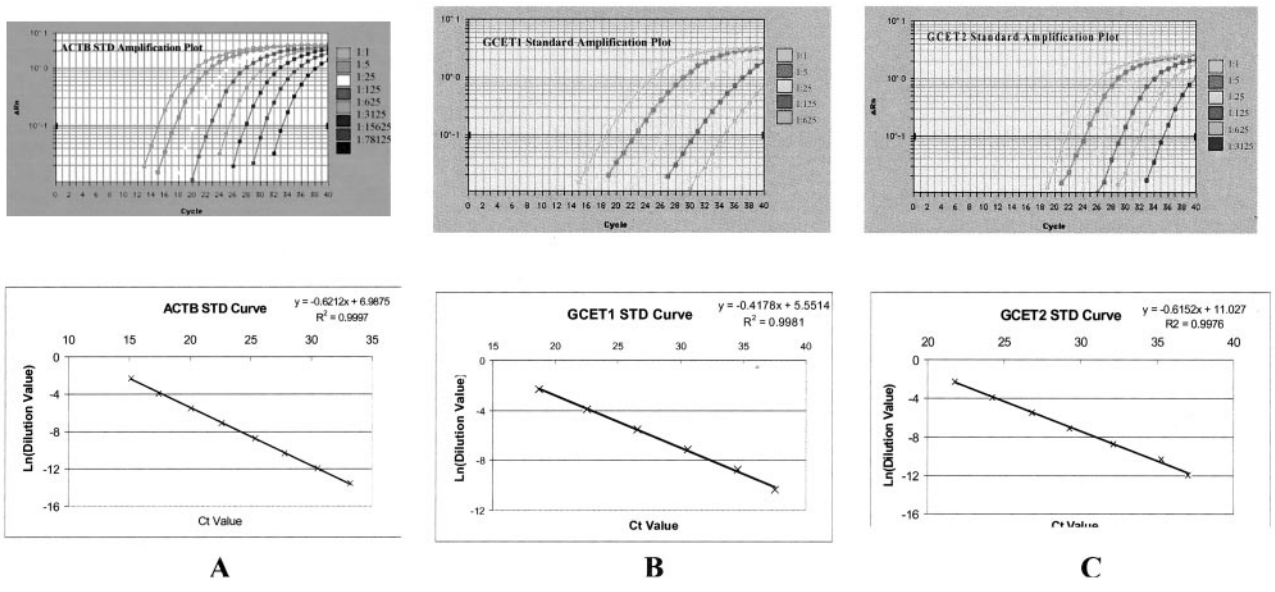


Figure 5. A: Top, Standard amplification plot for reference gene, β actin (ACTB); **bottom**, ACTB standard curve. The serial dilutions of the standard are expressed in nature log values in the *y* axis. **B: Top**, Standard amplification plot for *GCET1*; **bottom**, *GCET1* standard curve. **C: Top**, Standard amplification plot for *GCET2*; **bottom**, *GCET2* standard curve. Standard amplification plots and curves for B2M and GAPDH were similar (not shown).

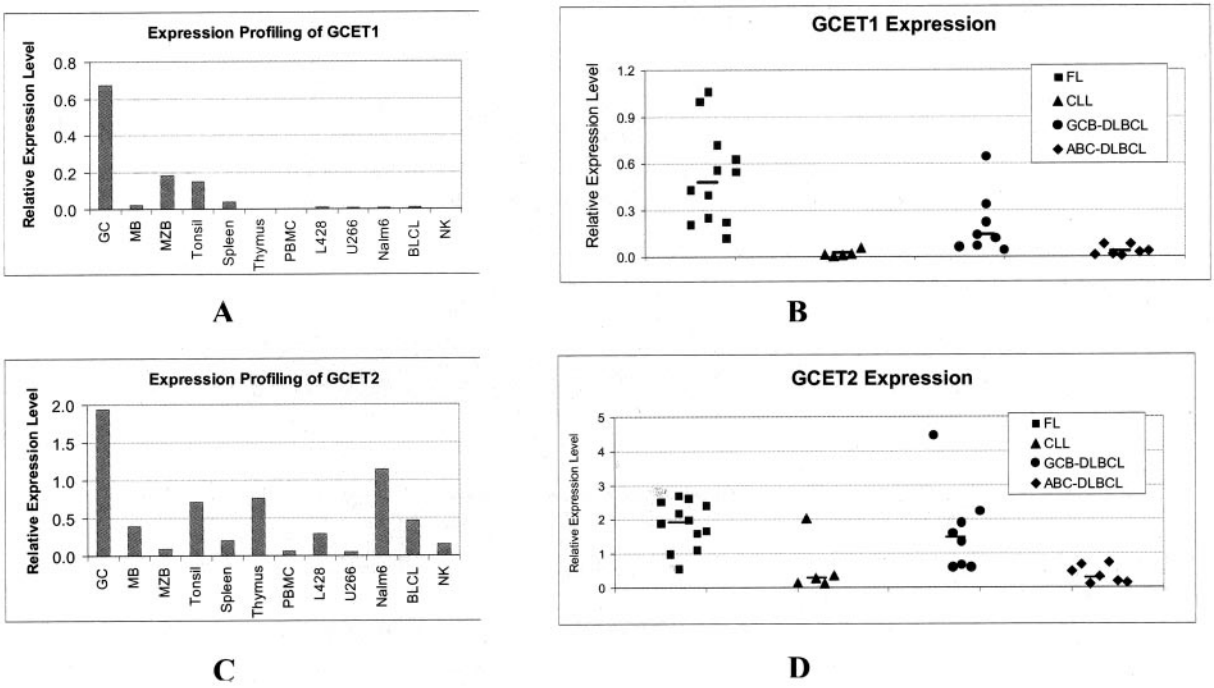


Figure 6. Expression profiling of *GCET1* and *GCET2*. **A:** *GCET1* is highly expressed in microdissected normal GCs, significantly lower in MZB, and very low in MB. In normal tonsil and spleen, it has a moderate expression level. **B:** *GCET1* has a high expression in FL and GCB-DLBCL samples, and its expression is significantly lower in CLL and ABC-DLBCL samples. **C:** Compared with MB and MZB, normal GCs also have a high expression of *GCET2*. Normal tonsil and spleen express a moderate level of *GCET2*. *GCET2* also has moderately high expression in thymus and Nalm6. **D:** In FL, CLL, and DLBCL samples, the expression of *GCET2* is very similar to that of *GCET1* except for one CLL sample. All of the real-time PCR data are expressed as a ratio to DHL16, which is arbitrarily set at 1.

sible role in apoptosis. Serpin squamous cell carcinoma antigen-2 (SCCA2), when overexpressed in squamous cell tumor, can inhibit apoptosis induced by tumor necrosis factor-α.[22] CrmA can inhibit the apoptosis induced by either Fas or tumor necrosis factor.[23,24] It is possible that *GCET1* is involved in the regulation of GC B cell survival. Previous studies and our data indicated that the expression of *GCET1* is primarily restricted to GC B cells and also to lymphoid malignancies with GC B-cell maturation. It is therefore one of the best markers of GC differentiation although some of the GCB type DLBCL have a lower expression level and may thus overlap with occasional ABC-DLBCL with higher expression levels.

GCET2 is 57% identical to a mouse GC-expressed transcript, M17.[20] The exon structures of the two genes are also very similar, suggesting that they are true orthologs. Sequence corresponding to exon 2 of *GCET2* is missing from the reported M17 cDNA. However, a homologous sequence can be detected in the corresponding genomic DNA indicating that this exon may be missing in M17 because of alternative splicing. There is a human gene (AAH24174) located on 1q44 with substantial homology to *GCET2*. It is also a small protein with similar exon/intron pattern and thus is an apparent paralog of *GCET2*.

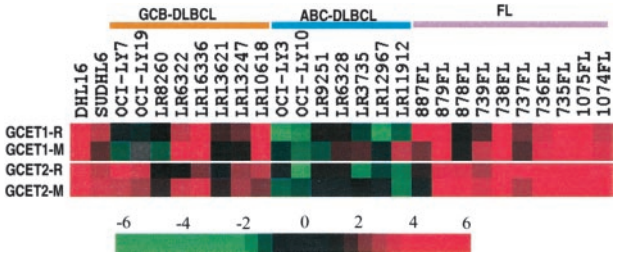


Figure 7. Comparison of real-time PCR with microarray data for *GCET1* and *GCET2*. All of the real-time PCR data were multiplied by 16 and converted into log2 values, with the value of DHL16 standard set at 4 to obtain a better spread of the colors. These values were converted into pseudocolors with green representing low expression and red representing high expression according to the predefined scheme shown. R, Real-time PCR data; M, microarray data.

Northern blot data indicated that *GCET2* has two transcripts. PCR on the 5′ untranslated region (UTR) and open reading frame did not show any isoforms. Five AAUAAA elements were found on the 3′ UTR of the mRNA we cloned. These results suggested that the two mRNA species result from alternative polyadenylation. Usually, the poly(A) tail is added 15 to 30 nucleotides downstream to the element but additional signals are required to determine whether a site is used for polyadenylation. Our data suggested that one or more of the proximal and distal polyadenylation signals are functional.

GCET2 encodes a 178-amino acid protein. Sequence analysis showed that there is no transmembrane domain. However the N-terminal region is likely myristoylated at glycine-2 that may provide the molecule with an anchor to associate with cellular membranes. Thus, myristoylation may regulate GCET2 subcellular localization and function. Previous studies indicated that N-myristoylation is essential for the activities of G protein α (Gα).[25] Of the six tyrosine residues in GCET2, four are conserved in AAH24174, despite the modest overall sequence iden-

tity. All four conserved tyrosines are in the context YXXΦ (Y = tyrosine, X = any amino acid, Φ = hydrophobic), as typical for the binding sites to many SH2 domains. In particular, Scansite software[26] predicted two putative SH2 binding sites at Y-107 (EEYYENV) and Y-148 (EDEYELL), which may interact with SH2 domains from IL2-inducible T-cell kinase (ITK)/Shc/Grb2 and Shc, respectively. Sequence alignment showed that sequences surrounding Y-148 are more conserved among human GCET2, mouse M17 and AAH24174 (Figure 4). There is also a putative PDZ-interacting sequence at the C-terminal end of GCET2. PDZ domains are involved in protein-protein interactions, and bind to the C-terminal four to five residues of their target proteins, frequently transmembrane receptors or ion channels.[27] Similar to mouse M17,[20] the GCET2 protein (amino acids 119 to 151) also contains an ITAM-like motif.[28] ITAM-mediated signaling is critical to a variety of receptor-mediated cell activation, and phosphorylation of the two ITAM tyrosines is essential for function,[29] through binding to the two SH2 domains of ZAP70 or Syk. It should be noted, however, that the spacing between the two tyrosines within the putative motif in GCET2 is considered greater than that found in known functional ITAMs, casting doubt that this motif can function as an ITAM. These structural features suggested that GCET2 might function as an adaptor molecule in signal transduction pathways. Myristoylation may target the protein to membranes, where it recruits signaling molecules by protein-protein interaction through the C-terminal PDZ-binding motif. Conserved tyrosine residues may become phosphorylated and bind to SH2 domains of other signal transduction molecules, leading to activation of downstream signals.

The expression profile of *GCET2* is in general quite similar to *GCET1* but slightly more variable. There is a moderate level of expression in the thymus and the pre-B cell line Nalm6, suggesting that *GCET2* may also be important in the biology of immature T and B cells. Microdissection studies showed that both *GCET1* and *GCET2* are expressed at a high level in GC cells but at low levels in the mantle and marginal zone. A recent paper reported that *GCET2* was induced by IL-4[30] raising the possibility that it may be involved in isotype switch and immunoglobulin production.

Using real-time PCR, we examined the expression levels of *GCET1* and *GCET2* in DLBCL, FL, and CLL and compared the findings with microarray data. In general, the expression of these two genes correlates well with microarray results. Cell lines and clinical samples of GCB-DLBCL and FL had high expression levels whereas CLL and ABC-DLBCL had low levels of expression. The distinction is, however, imperfect, and both genes are necessary to provide a better predictor of GC B-cell differentiation in lymphoid malignancy. Interestingly, Lossos and co-worker[30] have recently reported that using *GCET2* level alone, they were able to identify DLBCL cases with distinctly different prognosis. Further studies will be necessary to determine whether *GCET2* is predictive of survival purely based on its separation of DLBCL into GCB and non-GCB groups or it also has independent biological functions that affect survival. Aside from their diagnostic importance, these two genes are very likely to have important roles in B cell biology and their functions in normal and neoplastic B cells are being actively investigated.

Acknowledgments

We thank Drs. John Shaughnessy, Georgio Inghirami, and Hans-Georg Klingemann for kindly providing us with the cell lines U266, L426, and NK92.

References

1. Ahmed R, Gray D: Immunological memory and protective immunity: understanding their relation. Science 1996, 272:54–60
2. Cariappa A, Pillai S: Antigen-dependent B-cell development. Curr Opin Immunol 2002, 14:241–249
3. Steinman RM: Dendritic cells and the control of immunity: enhancing the efficiency of antigen presentation. Mt Sinai J Med 2001, 68:106–166
4. Han S, Zheng B, Takahashi Y, Kelsoe G: Distinctive characteristics of germinal center B cells. Semin Immunol 1997, 9:255–260
5. Kelsoe G: Life and death in germinal centers (redux). Immunity 1996, 4:107–111
6. Alizadeh AA, Eisen MB, Davis RE, Ma C, Lossos IS, Rosenwald A, Boldrick JC, Sabet H, Tran T, Yu X, Powell JI, Yang L, Marti GE, Moore T, Hudson J Jr, Lu L, Lewis DB, Tibshirani R, Sherlock G, Chan WC, Greiner TC, Weisenburger DD, Armitage JO, Warnke R, Lavy R, Wilson W, Grevor MR, Byrd JC, Botstein D, Brown PD, Staudt LM: Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. Nature 2000, 403:503–511
7. Rosenwald A, Wright G, Chan WC, Connors JM, Campo E, Fisher RI, Gascoyne RD, Muller-Hermelink HK, Smeland EB, Giltnane JM, Hurt EM, Zhao H, Averett L, Yang L, Wilson WH, Jaffe ES, Simon R, Klausner RD, Powell J, Duffey PL, Longo DL, Greiner TC, Weisenburger DD, Sanger WG, Dave BJ, Lynch JC, Vose J, Armitage JO, Montserrat E, Lopez-Guillermo A, Grogan TM, Miller TP, LeBlanc M, Ott G, Kvaloy S, Delabie J, Holte H, Krajci P, Stokke T, Staudt LM: The use of molecular profiling to predict survival after chemotherapy for diffuse large-B-cell lymphoma. N Engl J Med 2002, 346:1937–1947
8. Schaefer BC: Revolutions in rapid amplification of cDNA ends: new strategies for polymerase chain reaction cloning of full-length cDNA ends. Anal Biochem 1995, 227:255–273
9. Bertioli D: Rapid amplification of cDNA ends. Methods Mol Biol 1997, 67:233–238
10. Matz M, Shagin D, Bogdanova E, Britanova O, Lukyanov S, Diatchenko L, Chenchik A: Amplification of cDNA ends based on template-switching effect and step-out PCR. Nucleic Acids Res 1999, 27: 1558–1560
11. Epstein AL, Levy R, Kim H, Henle W, Henle G, Kaplan HS: Biology of the human malignant lymphomas IV: functional characterization of ten diffuse histiocytic lymphoma cell lines. Cancer 1978, 42:2379–2391
12. Davis RE, Brown KD, Siebenlist U, Staudt LM: Constitutive nuclear factor kappaB activity is required for survival of activated B cell-like diffuse large B cell lymphoma cells. J Exp Med 2001, 194:1861–1874
13. Nilsson K, Bennich H, Johansson SG, Ponten J: Established immunoglobulin producing myeloma (IgE) and lymphoblastoid (IgG) cell lines from an IgE myeloma patient. Clin Exp Immunol 1970, 7:477–489
14. Han T, Dadey B, Minowada J: Unique leukemic non-T/non-B lymphoid cell lines (REH and KM-3): absence of MLR-S and presence of suppressor cell activity for normal T-cell response. J Clin Lab Immunol 1978, 1:237–243
15. Schaadt M, Diehl V, Stein H, Fonatsch C, Kirchner HH: Two neoplastic cell lines with unique features derived from Hodgkin's disease. Int J Cancer 1980, 26:723–731
16. Gong JH, Maki G, Klingemann HG: Characterization of a human cell line (NK-92) with phenotypical and functional characteristics of activated natural killer cells. Leukemia 1994, 8:652–658
17. Schneider U, Schwenk HU, Bornkamm G: Characterization of EBV-genome negative "null" and "T" cell lines derived from children with

acute lymphoblastic leukemia and leukemic transformed non-Hodgkin lymphoma. Int J Cancer 1977, 19:621–626
18. Chan WC, Huang JZ: Gene expression analysis in aggressive NHL. Ann Hematol 2001, 80(Suppl 3):B38–B41
19. Frazer JK, Jackson DG, Gaillard JP, Lutter M, Liu YJ, Banchereau J, Capra JD, Pascual V: Identification of centerin: a novel human germinal center B cell-restricted serpin. Eur J Immunol 2000, 30:3039–3048
20. Christoph T, Rickert R, Rajewsky K: M17: a novel gene expressed in germinal centers. Int Immunol 1994, 6:1203–1211
21. Huang JZ, Sanger WG, Greiner TC, Staudt LM, Weisenburger DD, Pickering DL, Lynch JC, Armitage JO, Warnke RA, Alizadeh AA, Lossos IS, Levy R, Chan WC: The t(14;18) defines a unique subset of diffuse large B-cell lymphoma with a germinal center B-cell gene expression profile. Blood 2002, 99:2285–2290
22. McGettrick AF, Barnes RC, Worrall DM: SCCA2 inhibits TNF-mediated apoptosis in transfected HeLa cells. The reactive centre loop sequence is essential for this function and TNF-induced cathepsin G is a candidate target. Eur J Biochem 2001, 268:5868–5875
23. Tewari M, Telford WG, Miller RA, Dixit VM: CrmA, a poxvirus-encoded serpin, inhibits cytotoxic T-lymphocyte-mediated apoptosis. J Biol Chem 1995, 270:22705–22708
24. Tewari M, Dixit VM: Fas- and tumor necrosis factor-induced apoptosis is inhibited by the poxvirus crmA gene product. J Biol Chem 1995, 270:3255–3260
25. Chen CA, Manning DR: Regulation of G proteins by covalent modification. Oncogene 2001, 20:1643–1652
26. Yaffe MB, Leparc GG, Lai J, Obata T, Volinia S, Cantley LC: A motif-based profile scanning approach for genome-wide prediction of signaling pathways. Nature Biotechnol 2001, 19:348–353
27. Ponting CP, Phillips C, Davies KE, Blake DJ: PDZ domains: targeting signalling molecules to sub-membranous sites. Bioessays 1997, 19: 469–479
28. Yao XR, Flaswinkel H, Reth M, Scott DW: Immunoreceptor tyrosine-based activation motif is required to signal pathways of receptor-mediated growth arrest and apoptosis in murine B lymphoma cells. J Immunol 1995, 155:652–661
29. Johnson SA, Pleiman CM, Pao L, Schneringer J, Hippen K, Cambier JC: Phosphorylated immunoreceptor signaling motifs (ITAMs) exhibit unique abilities to bind and activate Lyn and Syk tyrosine kinases. J Immunol 1995, 155:4596–4603
30. Lossos IS, Alizadeh AA, Rajapaksa R, Tibshirami R, Levy R: HGAL is a novel IL-4-inducible gene that strongly predicts survival in diffuse large B-cell lymphoma. Blood 2003, 101:433–440