Structure and Expression of Human Germline V_H Transcripts

By Jeffrey E. Berman,* Caroline G. Humphries,‡ Jeremy Barth,* Frederick W. Alt,* and Philip W. Tucker‡

*From the *Howard Hughes Medical Institute and the Departments of Biochemistry and Molecular Biophysics, and Microbiology, College of Physicians and Surgeons of Columbia University, New York, New York 10032; and the ‡Department of Microbiology, University of Texas Health Science Center, Dallas, Texas 75235*

Summary

The human V_H5 family consists of two functional genes and one pseudogene. We have found a novel 1.2-kb V_H5 gene transcript in normal fetal liver and cord blood and in transformed B lineage cells. V_H5-positive cDNA clones were isolated from precursor B acute lymphoblastic leukemia, B chronic lymphoblastic leukemia, Epstein-Barr Virus–transformed B cell lines, and cord blood, and were identified as transcripts of unrearranged V_H5 genes (germline transcripts). The cDNA clones were derived from both functional and pseudo-V_H5 genes. Most germline transcripts appear to initiate at the normal V_H promoter and are cleaved and polyadenylated at sites several hundred bases downstream of the V_H5 coding region. Correct splicing of the leader intron was observed in all clones. In functional and pseudo-V_H5 cDNAs, an open translational reading frame extends from the leader to a termination codon in the nonamer. Only limited polymorphisms were observed in the coding as well as flanking regions of the V_H5 transcripts. Functional and pseudo-V_H5 cDNAs are also highly homologous throughout. The many similarities between human germline V_H5 transcripts and previously identified murine germline V_HJ558 transcripts are discussed.

The variable regions of immunoglobulin (Ig) heavy (H) chains are assembled from three groups of germline gene segments, V_H, D, and J_H. During precursor B cell differentiation in fetal liver and adult bone marrow, these segments are joined into a complete V_H-D-J_H variable region by an ordered assembly process. First, D to J_H segments are joined, followed by V_H to D-J_H rearrangement (reviewed in reference 1). In humans, there are 100–200 V_H gene segments, which have been classified into six families (V_H1 to V_H6) on the basis of nucleic acid sequence homology. V_H families range in size from the single-member V_H6 family to the large V_H1 and V_H3 families, each containing >25–30 members (reviewed in reference 2). The small V_H5 family consists of two functional genes and one pseudogene (3–5). One of the functional V_H5 genes is not present in all individuals (4).

In general, most V_H segments are not expressed at detectable steady-state levels in unrearranged, germline configuration. V_H to D-J_H rearrangement brings the V_H segment and its associated promoter elements into proximity of an enhancer located within the J_H-C_μ intron, thus allowing for high-level transcription of the V_H-D-J_H rearrangement initiating at the V_H promoter and continuing into the downstream constant region (reviewed in reference 6). In mice, at least some V_H segments belonging to the large J_H-distal J558 family are transcribed in unrearranged configuration outside the influence of the H chain enhancer (7). Germline V_HJ558 genes are specifically transcribed at the pre-B stage of B cell development. The function of germline V_H transcription is unknown, although the specific expression of germline V_HJ558 transcripts in pre-B cells, the cells in which V_H to D-J_H rearrangement occurs, led to the proposal that germline V_H transcription may be related to accessibility of the V_H locus to recombinase enzymes (7).

The large size of the murine J558 family (up to 500–1,000 members; reference 8) has complicated more detailed analyses of germline V_H transcription. Thus, although it is clear that germline V_H transcripts are derived from multiple V_HJ558 genes (7), it is difficult to ascertain the exact number and nature of the V_H segments that are expressed. Potential germline transcripts derived from the human V_H5 family have been detected in B chronic lymphoblastic leukemia (B-CLL)[1] tumors by Northern analysis (4). In contrast to the murine V_HJ558 family, the small size of the human V_H5 family simplifies studies of germline V_H expression. We have

[1] *Abbreviations used in this paper:* ALL, acute lymphoblastic leukemia; BM, bone marrow; CLL, chronic lymphoblastic leukemia; ORF, open reading frame.

therefore further characterized human germline V_H transcripts in normal and transformed B lineage cells and compared them with murine germline V_HJ558 transcripts.

Materials and Methods

Cells and Tissues. Acute lymphoblastic leukemia (ALL) samples were classified as immature B lineage malignancies (precursor ALL) by standard morphologic and cytochemical criteria (9); three were cell lines (NALL, NALM, and LAZZ) and the remainder were PBMC or bone marrow (BM) samples (>85% lymphoblasts) from ALL patients. B-CLL samples (>98% $CD5^+$) were obtained as described (3, 4). EBV-transformed lines were derived from adult PBMC or from fetal BM as described (10). EBV-21 and EBV-321 were transformants of normal $CD5^+$ cells sorted from adult PBMC (11). Both lines were polyclonal as confirmed by sequencing functionally rearranged V_H5-C_μ transcripts (P.W. Tucker and C. Humphries, unpublished results).

Fetal livers were obtained with permission and Internal Review Board approval from patients undergoing spontaneous or medically indicated therapeutic abortions. Fetal age was estimated by last menstrual period and by fetal foot length. T cell–depleted PBMC were prepared from healthy donors as described (10). Adult spleens were obtained from patients undergoing splenectomy for abdominal trauma. Newborn cord blood was enriched for B cells as described (10).

RNA Blotting and Preparation. Preparation of total and poly$(A)^+$ RNA, fractionation of RNA by formaldehyde/agarose gel electrophoresis, and blotting was done as described (7).

Probes. ^{32}P-labeled nick-translated V_H family–specific probes were prepared and hybridized as described (5). Genomic V_H5 probes were prepared from 5-1R1 (5) or from V_H251 (4).

Preparation of cDNA Libraries. cDNA, made essentially as described previously (7), was blunt-ended, ligated to EcoR1 adaptors, and libraries were constructed in Ch16A (LAZZ) or in λgt10 (CLL-12, CLL-27, EBV-21, and CB-4).

DNA Sequencing. cDNA clones were sequenced on both strands by the dideoxy chain termination method using M13 universal and reverse oligonucleotide primers (U.S. Biochemical Corp., Indianapolis, IN). Sequences were analyzed using the MicroGenie (Beckman Instruments, Inc., Palo Alto, CA) and DNAStar programs.

Results

Transformed B Lineage Cells Express Novel-sized V_H5 Gene Transcripts. Pre-B ALLs have H chain rearrangements but lack sIg and thus appear to represent human pre-B cells (9). 10 pre-B ALLs were examined by Northern analysis for expression of the six human V_H gene families (V_H1–V_H6). The ALLs expressed mature V_H-C_μ H chain mRNA, with individual samples expressing V_H segments from a particular family. In addition, we detected a novel 1.2-kb transcript that hybridized to a V_H5 probe. Novel V_H5 transcripts were found in all precursor B ALLs examined, including three cell lines (Fig. 1 *A*, lanes *1–3*) and seven PBMC or BM samples from ALL patients (Fig. 1 *A*, lanes *4–10*). One ALL line (Fig. 1 *A*, lane *2*) also expressed a novel V_H5 mRNA, slightly

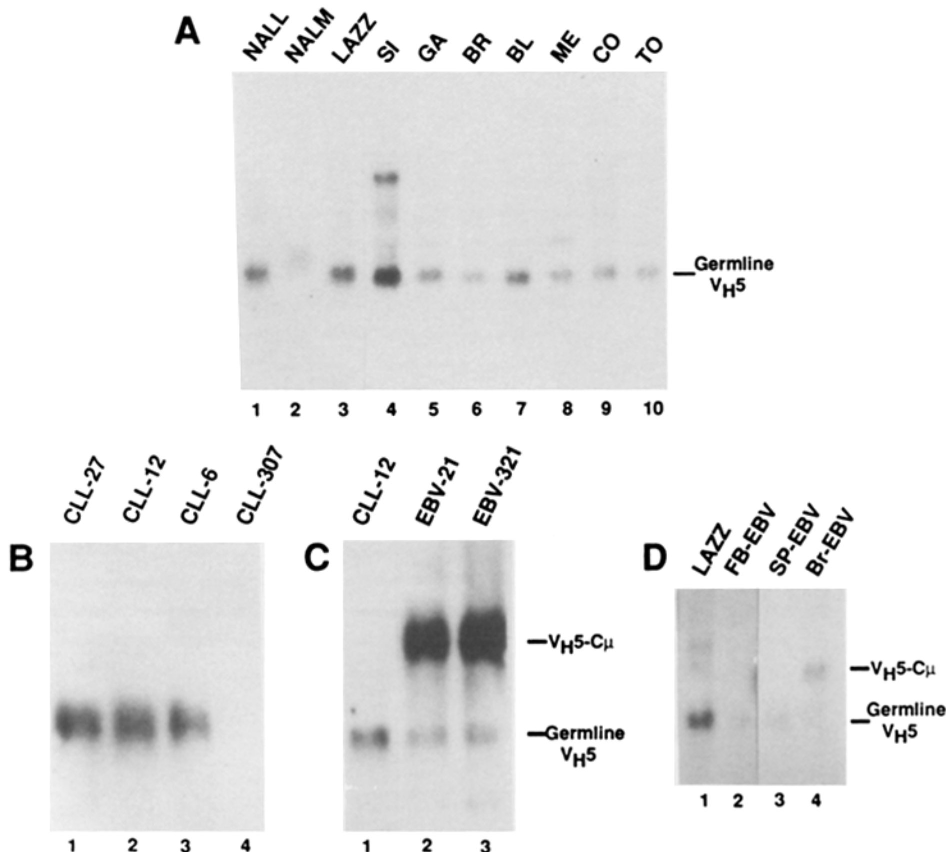

Figure 1. Northern analyses of V_H5 expression in transformed B lineage cells. (*A*) Expression of a 1.2-kb germline V_H5 transcript in pre-B ALL cell lines (lanes *1–3*) and PBMC or BM from ALL patients (lanes *4–10*). (*B*) Germline V_H5 transcription in B-CLLs. (*C* and *D*) Germline V_H5 expression in EBV-transformed B cell lines. B-CLL-12 and LAZZ pre-B ALL RNAs are included for comparison. Mature C_μ H chain transcripts (V_H-D-J_H-C_μ) derived from rearranged V_H genes were detected in some samples; e.g., EBV-21. EBV-21 and EBV-321 are transformants of sorted $CD5^+$ cells. FB-EBV is a fetal BM line, and Sp-EBV and Br-EBV are PBMC lines (10). Amounts of RNA per lane were: A1-10, 10–20 μg total RNA; B1-4, 5 μg poly$(A)^+$; C1-4, 5 μg poly$(A)^+$; D1, 4 μg poly$(A)^+$; D2-4, 5–10 μg total RNA.

greater than 1.2-kb, that is apparently derived from a partially assembled V_H5-D rearrangement (J.E. Berman and F.W. Alt, unpublished results). Occasionally, transcripts larger than mature H chain mRNA were detected with V_H family probes other than V_H5. However, they were not consistently observed in all ALLs and may represent precursor mRNAs (data not shown).

Murine germline V_H transcripts are expressed in pre-B lines and are not observed in more mature B lineage tumors such as B cell lymphomas and myelomas (7). In human sIg^+ CLLs, V_H5 transcripts similar in size to the 1.2-kb species found in pre-B ALLs were detected by Northern blotting (4). Therefore, we examined transformed B lineage cells other than pre-B ALLs for V_H5 expression. The novel 1.2-kb V_H5 transcript was observed in some B-CLLs and EBV-transformed B lymphoblastoid cell lines, but in contrast to the pre-B ALLs, it was not found in all such lines (representative data are shown in Fig. 1, *B*, *C*, and *D*). As with the ALLs, germline transcripts of other V_H families were not observed (data not shown).

The Novel-sized 1.2-kb V_H5 Species Are Germline Transcripts. cDNA libraries were constructed from several ALL, CLL, and EBV-transformed B cell samples tested above, V_H5-hybridizing clones were randomly isolated, and complete nucleotide sequences of two of the longest cDNAs (L3-4, L2-9) from a pre-B ALL library (LAZZ; Fig. 1 *A*) were determined. We also sequenced three V_H5^+ cDNAs randomly isolated from two B CLL samples (CLL-12, CLL-27) and from an EBV-transformed B cell line (EBV-21) that also expressed the 1.2-kb V_H5 transcript (Fig. 1, *B* and *C*). Comparison of

```
   VH251  ATGCAAATGC AAGTGGGGGC CTCCCCACTT AAACCCAGGG CTCCCCTCCA CAGTGAGTCT CCCTCACTGC
   5-1R1  ---------- ---------- ---------- ---------- ---------- ---------- ----------

                                                   Leader>
   VH251  CCAGCTGGGA TCTCAGGGCT TCATTTTCTG TCCTCCACCA TC ATG GGG TCA ACC GCC ATC CTC GCC
   5-1R1  ---------- ---------- ---------- ---------- -- --- --- --- --- --- --- --- ---
 L3-4(1)                         --------- ---------- -- --- --- --- --- --- --- --- ---
 CB-4(2)                                        ----- -- --- --- --- --- --- --- --- ---
EBV-21(3)                 ------- ---------- ---------- -- --- --- --- --- --- --- --- ---
CLL-12(4)  ------ - ---------- ---------- ---------- -- --- --- --- --- --- --- --- ---

                                                        FR1>
   VH251 CTC CTC CTA GCT ATT CTC CAA G/INTRON/GA GTC TGT GCC GAG GTG CAG CTG GTG CAG TCT
   5-1R1 --- --- --G --- G-- --- --- G/INTRON/-- --- --- --- --- --- --- --- --- --- --
 1,2,3,4 --- --- --G --- G-- --- ---     GGA     --- --- --- --- --- --- --- --- --- ---

   VH251 GGA GCA GAG GTG AAA AAG CCC GGG GAG TCT CTG AAG ATC TCC TGT AAG GGT TCT GGA TAC
   5-1R1 --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
 1,2,3,4 --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---

                     CDR1>               FR2>
   VH251 AGC TTT ACC AGC TAC TGG ACC GGC TGG GTG CGC CAG ATG CCC GGG AAA GGC TTG GAG TGG
   5-1R1 --- --- --- --- --- --- -T- --- --- --- --- --- --- --- --- --- --- C-- --- ---
 1,2,3,4 --- --- --- --- --- --- -T- --- --- --- --- --- --- --- --- --- --- C-- --- ---

                 CDR2>                                                           FR3>
   VH251 ATG GGG ATC ATC TAT CCT GGT GAC TCT GAT ACC AGA TAC AGC CCG TCC TTC CAA GGC CAG
   5-1R1 --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
 1,2,3,4 --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---

   VH251 GTC ACC ATC TCA GCC GAC AAG TCC ATC AGC ACC GCC TAC CTG CAG TGG AGC AGC CTG AAG
   5-1R1 --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
 1,2,3,4 --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---

   VH251 GCC TCG GAC ACC GCC ATG TAT TAC TGT GCG AGA CA (CACAGTG) AGAGAAACCAGCCCCGAGCCCG
   5-1R1 --- --- --- --- --- --- --- --- --- --- --- [D-JH
 1,2,3,4 ---  -- --- --- --- --- --- --- --- --- --- -- ------- ----------------------

             *
   VH251 (TCTAAAACC) CTCCACACCGC AGGTGCAGAAT GAG CTGCTAGAGA CTCACTCCCC AGGGGCCTCT CTATTCATCT
     1,2 --------- ---------- ---------- --- ---------- ---------- ---------- ----------
     3,4 --------- ---------- --------G- --- ---------- ---------- ---------- ----------

   VH251 GGGGAGGAAA CACTGGCTGT TTGTGTCCTC AGGAGCAAGA ACCAGAGAAC AATGTGGGAG GGTTCCCAGC
   1,3,4 ---------- ---------- ---------- ---------- ---------- ---------- ----------
       2 ---------- ---------- ---------- ----C----- ---------- ---------- ----------

   VH251 CCCTAAGGCA ACTGTATAGG GGACCTGACC ATGGGAGGTG GATTCTCTGA CGGGGCTCTT GTGTGTTCTA
 1,2,3,4 ---------- ---------- ---------- ---------- ---------- ---------- ----------

   VH251 CAAGGTTGTT CATGGTGTAT ATTAGATGGT TAACATCAAA AGGCTGCCTA ACAGGCACCT CTCCAATATG
   1,2,4 ---------- ---------- ---------- ---------- ---------- ---------- ----------
       3 ---------- ---------- -------A-- ---------- ---------- ---------- ----------

   VH251 ACAGTATTTT AATTAGTGAA AATTTTACAC AGTTCATCAT TGCTTGCTTG CCTTCCTCCC TCCTGTCCAC
   1,2,3 ---------- ---------- ---------- ---------- ---------- ---------- ----------
       4 -T-------- ---------- ---------- ---------- ---------- ---------- --------G-

   VH251 TCTCACTCAC TCCTTCTTTT ATTTTCTACT TAATTTTACA AAATCATTTA ACCCCTTTTT GAACTATTAA
 1,2,3,4 ---------- ---------- ---------- ---------- ---------- ---------- ----------

   VH251 TAGGTTATCT TTGTTTGGTG ATTGTTTTCC TTTCAATAAT ATGTACTGAA TAATTCATCT TTGTGCCAAT
       1 ---------- ---------- ---------- ---------- ---------- ---------- ----------
       2 ----C----- ---------- ---------- ---------- ---------- ---------- --------
       3 ---------- ---------- ---------- ---T------ ---------- ---------- ----------
       4 ---------- ---------- --------T- ---T------ ---------- ---------- ----A-----

   VH251 TCATAAGTAT TCTGGTGTAA TAAAGACTTC TTTCATAAAA ATTGGATAAA TTAAAATAAAGA
       1 ---------- ---------- ---------- ---------- ---------- --
       3 ---------- ---------- ---------- CA
       4 ---------- ---------- ---------- -----AAAAA AAAAAAAAAA AA
```

Figure 2. Nucleotide sequences of germline cDNAs derived from the functional V_H5 gene. Genomic sequences of unrearranged (V_H251; 4) and rearranged (5-1R1; 5) functional V_H5 genes are provided for comparison. Dashes indicate identity with the V_H251 sequence; differences are indicated. cDNA sequences are condensed into a single line in regions of complete identity (1, L3-4; 2, CB-4; 3, EBV-21; 4, CLL-12). The leader is spliced in the cDNAs so the genomic intron sequences are not shown. Corrections have been made in the originally reported V_H251 sequence (4) immediately 3' of the nonamer. The sequence of V_H251 further downstream of the nonamer is previously unpublished. Heptamer and nonamer recombination sequences are boxed. The internal "V replacement" heptamer at the end of framework 3 is underlined. The asterisk and line above the nonamer indicate the end of the ORF. Polyadenylation signals are underlined in the untranslated region 3' of the nonamer.

these nucleotide sequences to genomic germline V_H5 sequences (4, 5) confirms that they represent germline transcripts of unrearranged V_H5 genes (Figs. 2 and 3). The cDNAs extend from 5 to 39 bp upstream of the leader, through the coding regions, and, in contrast to rearranged V_H gene transcripts, include the normally deleted recombination signals and 3′ flanking regions.

Germline Transcripts Are Expressed from both Functional and Pseudo-V_H5 Genes. As expected, some cDNAs were derived from the functional V_H5 gene (V_H251, Fig. 2), but surprisingly, others were transcripts of the V_H5 pseudogene (V_H15/V_H1-V, Fig. 3). To determine the relative steady-state levels of these two transcripts and to search for other germline V_H5 mRNA species, 10 V_H5^+ clones from the LAZZ pre-B ALL (12; Fig. 1 *A*, lane *3*) cDNA library were randomly characterized. Half of the clones were derived from the functional V_H5 gene and the others were transcripts of the V_H5 pseudogene. 9 of the 10 cDNA clones were from similar regions of the V_H5 genes; however, one pseudogene-derived cDNA, L6-1, extended further upstream to ∼420 bp 5′ of the initiation codon and ended ∼0.1 kb 3′ of the nonamer (data not shown).

Structure of Germline V_H5 Transcripts. Comparison of the V_H5 cDNAs (including L6-1) and germline genomic V_H5 sequences (V_H251 coding region, reference 4; newly determined V_H251 3′ flank, Fig. 2; 1-V, reference 5) indicated that the leader intron was appropriately spliced but no additional splicing occurred in the 3′ flanking regions. The germline V_H5 transcripts are polyadenylated, although there is apparently slight variation in the site of pre-RNA 3′ end cleavage and polyadenylation since some cDNAs such as L3-4 and L2-9 extend beyond the beginning of the poly(A) tails present in CLL-12 and CLL-27 (Figs. 2 and 3). This variation is not surprising because there are multiple consensus polyadenylation/cleavage signals (reviewed in reference 13) at the 3′ ends of both the functional and pseudo-V_H5 genes (underlined in Figs. 2 and 3).

Together, the functional germline V_H5 cDNAs encompass 967 bp of sequence and the pseudo-V_H5 cDNAs span a 1,001-bp region. Assuming an ∼250-base poly(A) tail (13), the cDNA clones correspond to the full-length 1.2-kb transcript. This suggests that transcriptional initiation of the majority of V_H5 germline mRNA sequences occurs within 100 bp upstream of the leader, probably from the normal V_H promoter; although the L6-1 cDNA indicates that there is also some transcriptional initiation upstream of the V_H promoter, at least for the pseudo-V_H5 gene.

Although from different individuals, the coding region sequences of the four functional germline V_H5 cDNAs and the corresponding germline gene (V_H251) are highly conserved

```
      1-V  GCAACTATGC AAATTCAAGT GGGGGCCTCC CCACTTAAAC CCAGGGCTCT CCTCCACAGT GAGTCTCCTT CACCACCCAG
     VH15        ---- ---------- ---------- ----- ---- ---------- ---------- ---------- ----------

                                                  Leader>
      1-V  CTGGGATCTC AGTGCTTTCT TTTCTGTCCT CCTCCAAG ATG GGG TCA ACC GTC ATC CTT TCC CTC GTC CTA
     VH15  ---------- ---------- ---------- ------G- -C- ---G--- --- --- --- --- --- --- --- ---
  L2-9(1)                            ---- ------G- --- --- --- --- --- --- --- --- --- --- --G
CLL-27(2)            - ---------- ---------- ------G- --- --- --- --- --- --- --- --- --- --- --G

                                         FR1>
      1-V  GCT GTT CTC CAA G/INTRON/GT GTC TTT GAC*GAG GTG CAG CTG TTG CAG TCT GCA GCA GAG GTG AAA
     VH15  --- --- --- --- -/INTRON/-- A-- --- -C- --- --- --- --- --- --- --- --- --- --- --- ---
        1  --- --- --- ---    GGT     --- --- -C- --- --- --- --- --- --- --- --- --- --- --- ---
        2  --- --- --- ---    ---     --- --- -C- --- --- --- --- --- --- --- --- --- --- --- ---

                                                                                CDR1>
      1-V  AGA CCC GGG GAG TCT CTG AGG ATC TCC TGT AAG ACT TCT GGA TAC AGC TTT ACC AGC TAC TGG ATC
     VH15  --- --- --- --- --- --- --- -C- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
      1,2  --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---

               FR2>                                                CDR2>
      1-V  CAC TGG GTG CGC CAG ATG CCC GGG AAA GAA CTG GAG TGG ATG GGG AGC ATC TAT CCT GGG AAC TCT
     VH15  --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- -TG --A  -- ---
      1,2  --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---

                                         FR3>
      1-V  GAT ACC AGA TAC AGC CCA TCC TTC CAA GGC CAC GTC ACC ATC TCA GCC GAC AGC TCC AGC AGC ACC
     VH15  --- --- --- --- --- ---  -- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
      1,2  --- --- --- --- --- --- --- -+- --- --- --- --- --- --- --- --- --- --- --- --- --- ---

      1-V  GCC TAC CTG CAG TGG AGC AGC CTG AAG GCC TCG GAC GCC GCC ATG TAT TAT TGT GTG AGA
     VH15  --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
      1,2  --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---

                      *
      1-V  GGGACCATCT AAAACCTTCC GCGGTGCAGG TGCAGAGTGA GCTGCCAGAC ACACCCTCCC TAGGGGCCTC TCTATTCATC
     VH15  ---------- ---------  ---------- ---------- ---------- ---------- A--A- ---- ----------
      1,2  ---------- ---------- ---------- ---------- ---------- ---------- C--------- ----------

      1-V  CGGGGAGGAA ACACTGGCTG TTTGTGTCCT CAGGAGCAAA AACCAGAGAA CAACATGGGA G
        1  ---------- ---------- ---------- ---------- ---------- ---------- -CGTTCCTAAC CCCTAAGGC
        2  ---------- ---------- ---------- ---------- ---------- ---------- ---------G- ---------

        1  AACTGGATGG GAGACCTGAC CCATCCAGTT CTCTGAGGGG GCTCTTGTGT GTTCTACAAG GTTGTTCATG GTGTATATTA
        2  ---------- ---------- ---A------ ---------- ---------- ---------- ---------- ----------

      1,2  CATGGTTAAC ATCAAAAGGC TGCCTAATAG GCACCTCTTC AATATAATAG TCTTTTAATT AGTGAAAATT TTACACAGTT

        1  CATCATTGCT TGCTTGCCTT CCTCCCTTCT GTCCGCTCTT ACTCCCTCCT TCTTTTATTT TCTACTTACT TTTCCAAAAT
        2  ---------- ---------- ---------- ---------- -------T-- ---------- ---------- ----------

        1  CATTTAACCC CTTTTTGTAC CATTAATAAG TTATCTTGTT TTTGTTGTTG TTTTCCTTTT AACAATATGC ACTGAATAAT
        2  ---------- ---------- -----G---- ---------- ---------- ---------- ---------- ----------

      1,2  TCATCTTTGT GCCAATCGTA AATATTTTGA TATAACAAAG ACTTCTGTCA TAAATAGATT TCCTGTGAGT AATCTTGCAA

        1  ATATTTTAGA ACCTGTTTGG TTAAGAATAA ATTAAAATAA ATAGACAAAT TTTTAAAGAC AT
        2  ---------- ---------- ---------- ----------  -----AAAA AAAAAAA
```

Figure 3. Nucleotide sequences of germline cDNAs derived from the V_H5 pseudogene. Two versions of the genomic sequence are provided for comparison (1-V, 5; V_H15, 4). cDNA sequences are condensed into a single line in regions of complete identity (1, L2-9; 2, CLL-27). An error in the published 1-V sequence has been corrected (indicated by the asterisk at the beginning of FR1) and now yields an ORF throughout the coding region. The nonamer recombination sequence is boxed. The internal heptamer is underlined. The asterisk and line above the nonamer indicate the end of the ORF. Polyadenylation signals in the 3′ untranslated region are underlined. The sequence data in Figs. 2 and 3 will be available from EMBL/GenBank/DDBJ under accession numbers X58397–X58402.

```
     FR3>
     GlnValThrIleSerAlaAspLysSerIleSerThrAlaTyrLeuGlnTrpSerSerLeuLysAlaSer
L3-4 CAGGTCACCATCTCAGCCGACAAGTCCATCAGCACCGCCTACCTGCAGTGGAGCAGCCTGAAGGCCTCG
L2-9 CACGTCACCATCTCAGCCGACAGCTCCAGCAGCACCGCCTACCTGCAGTGGAGCAGCCTGAAGGCCTCG
     HisValThrIleSerAlaAspSerSerSerSerThrAlaTyrLeuGlnTrpSerSerLeuLysAlaSer

                                             <------ 23bp ------>
     AspThrAlaMetTyr TyrCysA laArgHi sThrVal ArgGluThrSerProGluProV al *
L3-4 GACACCGCCATGTAT TACTGTG CGAGACA CACAGTG AGAGAAACCAGCCCCGAGCCCG TCTAAAACC
L2-9 GACGCCGCCATGTAT TATTGTG TGAGA                            GGGACCA TCTAAAACC
     AspAlaAlaMetTyr TyrCysV alArg                            GlyThrI le *
                    <------------ 12bp ------------>

L3-4 CTCCACACCGCAGGTGCAGAATGAGCTGCTAGAGACTCACTCCCCAGGGGCCTCTCTATTCATCTGGGGAG
L2-9 TTCCGCGGTGCAGGTGCAGAGTGAGCTGCCAGACACACCCTCCCCAGGGGCCTCTCTATTCATCCGGGGAG

L3-4 GAAACACTGGCTGTTTGTGTCCTCAGGAGCAAGAACCAGAGAACAATGTGGGAGGGTTCCCAGCCCCTAAG
L2-9 GAAACACTGGCTGTTTGTGTCCTCAGGAGCAAAAACCAGAGAACAACATGGGAGCGTTCCTAACCCCTAAG

L3-4 GCAACTGTATAGGGGACCTGACCATGGGAGGTGGATTCTCTGACGGGGCTCTTGTGTGTTCTACAAGGTTG
L2-9 GCAACTGGATGGGAGACCTGACCCATCCAG     TTCTCTGAGGGGGCTCTTGTGTGTTCTACAAGGTTG

L3-4 TTCATGGTGTATATTAGATGGTTAACATCAAAAGGCTGCCTAACAGGCACCTCTCCAATATGACAGTATTT
L2-9 TTCATGGTGTATATTACATGGTTAACATCAAAAGGCTGCCTAATAGGCACCTCTTCAATATAATAGTCTTT

L3-4 TAATTAGTGAAAATTTTACACAGTTCATCATTGCTTGCTTGCCTTCCTCCCTCCTGTCCACTCTCACTCAC
L2-9 TAATTAGTGAAAATTTTACACAGTTCATCATTGCTTGCTTGCCTTCCTCCCTTCTGTCCGCTCTTACTCCC

L3-4 TCCTTCTTTTATTTTCTACTTAATTTTACAAAATCATTTAACCCCTTTTTGAACTATTAATAGGTTATCTT
L2-9 TCCTTCTTTTATTTTCTACTTACTTTTCCAAAATCATTTAACCCCTTTTTGTACCATTAATAAGTTATCTT

L3-4 TGTTTGGTGATTGTTTTCCTTTCAATAATATGTACTGAATAATTCATCTTTGTGCCAATTCATAAGTATTC
L2-9 GTTTTTGTTGTTGTTTTCCTTTTAACAATATGCACTGAATAATTCATCTTTGTGCCAA TCGTAAATATTT

L3-4 TGGTGTAATAAAGACTTCTTTCATAAAAATTGGATAAATT
L2-9 TGATATAACAAAGACTTCTGTCATAAATAGATTTCCTGTGAGTAATCTTGCAAATATTTTAGAACCTGTTT

L2-9 GGTTAAGAATAAATTAAAATAAATAGACAAATTTTTAAAGACAT
```

Figure 4. Germline transcripts of functional and pseudo-V_H5 genes are highly related. The coding regions of functional (L3-4) and pseudo-V_H5 (L2-9) cDNAs (isolated from LAZZ, see text) are 91% similar (only FR3 is shown), and the 3′ flanking regions are 86% similar. Predicted COOH-terminal amino acid sequences of the putative truncated H chain proteins encoded by the germline V_H5 transcripts are shown. Recombination signal sequences are boxed (from 5′ to 3′ they are: V replacement heptamer, heptamer, nonamer). In L2-9, a 24-bp deletion removes the heptamer and part of the spacer region, thus bringing the internal heptamer 12 bp upstream of the nonamer.

(>98% similarity; Fig. 2). This is consistent with previous findings of limited V_H251 coding region polymorphism (14). Unexpectedly, a high degree of conservation is also found in the 3′ flanking regions of the functional V_H5 sequences, all of which are >98% similar. Even the pseudo-V_H5 sequences are >98% similar throughout both the coding and flanking regions (Fig. 3).

Open Reading Frames (ORF) Are Found in the Germline V_H5 cDNAs. In the germline cDNAs derived from the functional V_H5 gene, an ORF extends throughout the V_H coding region and ends in a termination codon within the nonamer recombination sequence (see asterisk in Fig. 2). Beyond the nonamer there are multiple stop codons in all reading frames. A search of GenBank did not reveal significant homologies between this 3′ untranslated region and other human or murine genes. The germline cDNAs potentially encode a V_H region the same as that of a rearranged gene, except for 11 COOH-terminal residues (His-Thr-Val-Arg-Glu-Thr-Ser-Pro-Glu-Pro-Val) identically encoded within the heptamer/23-bp spacer/nonamer region of all the functional V_H5 cDNAs (Fig. 2). A similar ORF was also found in the pseudo-V_H5 cDNAs, again terminating within the nonamer, but only three additional amino acids (Gly-Thr-Ile) are encoded beyond FR3 due to a deletion within the recombination sequences (Figs. 3 and 4; see Discussion).

Identical Germline V_H5 Transcripts Are Expressed in Normal Tissues. Characteristic 1.2-kb germline V_H5 transcripts were observed on Northern analyses of fetal liver (15 and 17 wk) and cord blood, but were not seen in adult T cell–depleted PBMC or in adult spleen (Fig. 5, *A* and *B*; confirmed by S1 analysis, data not shown). The sequence of a germline V_H5 clone (CB-4) isolated from a cDNA library of normal cord blood (Fig. 5 *B*, lane *2*) was essentially identical to functional V_H5 cDNAs isolated from transformed B lineage cells (Fig. 2). As with the transformed lines, we did not detect germline-length transcripts with V_H family probes other than V_H5 (data not shown).

Discussion

Limited Polymorphism among Germline V_H5 cDNAs. We have demonstrated that unrearranged functional and pseudo-V_H5 gene segments are transcribed in normal and transformed B lineage cells. Although germline transcripts of the poly-

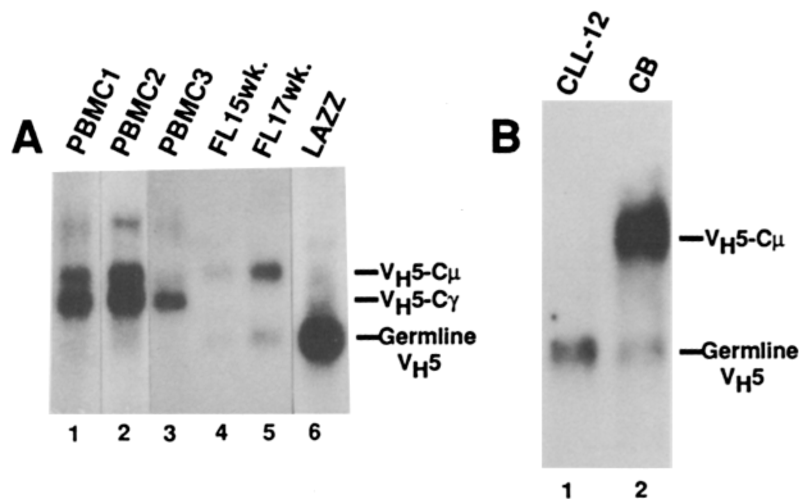

Figure 5. Germline V_H5 expression in normal tissues. (*A*) Germline V_H5 expression in T cell–depleted PBMC (lanes *1–3*) and fetal liver (15 wk, lane *4*; 17 wk, lane *5*). RNA from the LAZZ pre-B ALL line is shown for comparison. (*B*) Germline V_H5 expression in normal cord blood (lane *2*). RNA from B CLL-12 is shown for comparison. Complete V_H5-C_μ H chain mRNA was detected in fetal liver and in cord blood, and both V_H5-C_μ and V_H5-C_γ mRNA were seen in adult PBMC. Amounts of RNA per lane were: A1-3, 10 μg total RNA; A4-5, 6 μg poly(A)$^+$; A6, 4 μg poly(A)$^+$; B1-2, 5 μg poly(A)$^+$.

morphic V_H5 gene (V_H32/V_H-2R1; 4,5) were not isolated, a detailed analysis of multiple clones was only done with the LAZZ cell line, which lacks this gene (J.E. Berman and F.W. Alt, unpublished results). The relative proportion of transcripts derived from the functional and pseudo-V_H5 genes may vary among cell lines; e.g., several germline cDNAs isolated from the polyclonal EBV-321 line (Fig. 1 *C*) were all derived from a functional V_H5 gene (V_H251) (P.W. Tucker and C. Humphries, unpublished observation).

The sequences of the functional and pseudo-V_H5 transcripts are highly conserved in the 3′ flank as well as in the coding regions. It has been suggested that the limited polymorphism seen in coding regions of the smaller V_H gene families (V_H4, -5, -6) may be related to critical antigen specificities encoded by these genes (14, 15). This explanation does not account for the high degree of conservation among the pseudo-V_H5 coding region sequences or among the 3′ flanking regions of both the functional and pseudo-V_H5 genes. Such conservation could be due to recent emergence and duplication of the V_H5 family, but may also be explained by the fact that these regions are expressed in germline transcripts. Conservation of the 3′ untranslated regions could be due to a functional role such as regulating mRNA stability. The V_H5 pseudogene is likely to be nonfunctional with regard to VDJ rearrangement due to its abnormal recombination sequences; however, conservation of the ORF in the pseudo-V_H5 gene might reflect a function of an encoded germline V_H5 protein.

3′ Flanking Regions of Functional and Pseudo-V_H Germline cDNAs. The functional and pseudo-V_H5 cDNA sequences are highly homologous throughout the coding (91% similar) and 3′ flanking regions (86% similar). A comparison of the 3′ flanking regions of functional (L3-4) and pseudo-V_H5 (L2-9) cDNAs reveals two major regions of divergence: ∼0.16 kb 3′ of the nonamer and immediately 3′ of the coding regions (Fig. 4). A 24-bp deletion in the immediate 3′ flank of the pseudogene results in loss of the heptamer and part of the spacer region, thus placing the nonamer 12 bp 3′ of the internal 'V replacement' heptamer (16, 17). Therefore, on the basis of the 12/23 joining rule (reviewed in reference 1) there could be direct rearrangement between the V_H5 pseudogene (12-bp spacer) and J_H segments (23-bp spacer). As yet there is no evidence for direct pseudo-V_H5 to J_H joining perhaps due to the orientation of the internal heptamer relative to the nonamer. Although it may be a coincidence, it is notable that the 24-bp deletion in the pseudogene does not shift the reading frame; thus, the translational site in the nonamer is conserved.

Comparison of Murine and Human Germline V_H Transcription. Several aspects of human and murine germline V_H transcription are similar. Although mouse and human germline V_H transcripts vary in length, murine germline V_HJ558 cDNAs have an ORF similar in size to that of the human germline V_H5 transcripts, also ending in a termination codon within the recombination signals (but in the 23-bp spacer region rather than in the nonamer; reference 7). Thus, murine and human germline V_H transcripts potentially encode truncated V_H proteins of a similar size. Transcription of only one V_H family has been readily found in humans (V_H5) as well as in mice (J558). Although the V_H5 family has no obvious murine counterpart (i.e., >70% nucleic acid similarity), V_HJ558 members are the most closely related. Of course germline transcripts from other murine and human V_H gene families may exist at undetected, lower steady-state levels. In mice and humans, most germline V_H transcription initiates at the normal V_H promoter, although more 5′ initiation has been seen with V_H pseudogenes that still retain ORFs (see above and reference 7).

Specificity of Germline V_H5 Expression. The consistent expression of germline V_H5 in pre-B ALLs may provide an additional diagnostic marker for disease in tissues such as PBMC, which normally lack detectable germline V_H mRNA. In contrast to the pre-B specificity of murine germline V_H expression, we also detected human germline V_H5 transcripts in more mature B lineage cells. Germline V_H5 expression in some B-CLLs and in some EBV-transformed B cell lines might argue against the proposal that germline V_H transcription reflects processes that regulate V_H to D-J_H rearrangement in pre-B cells. On the other hand, germline V_H expression beyond the pre-B cell stage may be unique to transformed cells and may reflect the difficulty of staging human B lineage tumor cells. Germline V_H expression may also be restricted to certain B cell subpopulations. Germline V_H5 mRNA was observed in fetal liver, but was not detected in PBMC or in adult spleen cells, suggesting that normal mature B cells do not express germline V_H5. Nevertheless, germline V_H5 transcripts may be expressed in some normal $CD5^+$ B cells based on their presence in some $CD5^+$ B-CLLs and in cord blood and fetal liver, tissues that are rich in $CD5^+$ cells (18, 19). We saw no evidence for germline V_H5 transcripts analogous to the unusual species found in human T cells, consisting of a novel 5′ exon spliced to an unrearranged V_H4 gene (20).

Function of Germline V_H5 Transcription. Germline V_H expression may regulate accessibility of the V_H locus during H chain rearrangement; alternatively, truncated V_H proteins encoded by germline transcripts may play a role in B cell development (7). Germline V_H cDNAs presumably possess the same 5′ signals required by mature H chain transcripts for translation since the transcription start sites seem similar. Furthermore, in vitro transcription and translation of the germline V_H5 cDNAs demonstrated that the transcripts are translatable into a protein consistent with the size (13.9 kD) of the predicted ORF (J. E. Berman and F.W. Alt, unpublished results). These putative truncated V_H5 H chain proteins are reminiscent of other members of the Ig gene superfamily, such as Vpre-B (21), and would consist of a single Ig domain possessing the two cysteine residues necessary for formation of the characteristic intra-chain disulfide-bonded loop.

Conservation of germline V_H gene transcription in humans as well as in mice suggests an important functional significance. The complexity of the J558 family has made it difficult to ascertain the heterogeneity of murine germline transcripts and to identify the corresponding genomic germline genes that are transcribed. Detectable germline J558 expression in mice could be the result of low level transcription

of multiple genes and/or high level expression of a limited subset of the large J558 family. The relative simplicity of the human V_H5 family allows us to conclude that human germline transcripts derived from just two unrearranged V_H genes are readily detectable and their genomic counterparts have already been isolated and are available for further analysis. This should facilitate future study of factors such as associated regulatory regions, which may control germline V_H gene transcription.

This work was supported by the Howard Hughes Medical Institute and by grants from the National Institutes of Health (CA-44016, CA-40427, and AI-20047). J. Berman was a fellow of the Howard Hughes Medical Institute.

Address correspondence to Frederick W. Alt, Howard Hughes Medical Institute Research Laboratories, Columbia University College of Physicians and Surgeons, 701 West 168th Street, New York, NY 10032. J.E. Berman's present address is the Medical Biotechnology Center and the Department of Microbiology and Immunology, University of Maryland, 655 West Baltimore Street, Baltimore, MD 21201.

Received for publication 5 November 1990 and in revised form 11 March 1991.

References

1. Alt, F.W., T.K. Blackwell, R.D. DePinho, M.G. Reth, and G.D. Yancopoulos. 1986. Regulation of genome rearrangement events during lymphocyte differentiation. *Immunol. Rev.* 89:5.
2. Berman, J.E., and F.W. Alt. 1990. Human H-chain variable region gene diversity, organization, and expression. *Int. Rev. Immunol.* 5:203.
3. Shen, A., C. Humphries, P.W. Tucker, and F.R. Blattner. 1987. Human H-chain V region gene family non-randomly rearranged in familial chronic lymphocytic leukemia. *Proc. Natl. Acad. Sci. USA.* 84:8563.
4. Humphries, C.G., A. Shen, W.A. Kuziel, J.D. Capra, F.R. Blattner, and P.W. Tucker. 1988. A new human Ig V_H family preferentially rearranged in immature B-cell tumours. *Nature (Lond.).* 331:446.
5. Berman, J.E., S.J. Mellis, R. Pollock, C.L. Smith, H. Suh, B. Heinke, C. Kowal, U. Surti, L. Chess, C.R. Cantor, and F.W. Alt. 1988. Content and organization of the human Ig V_H locus: definition of three new V_H families and linkage to the Ig C_H locus. *EMBO (Eur. Mol. Biol. Organ.) J.* 7:727.
6. Sen, R., and D. Baltimore. 1989. Factors regulating Ig-gene transcription. *In* Immunoglobulin Genes. T. Honjo, F.W. Alt, and T. Rabbitts, editors. Academic Press, San Diego, CA. 327–342.
7. Yancopoulos, G.D., and F.W. Alt. 1985. Developmentally controlled and tissue-specific expression of unrearranged V_H gene segments. *Cell.* 40:271.
8. Livant, D., C. Blatt, and L. Hood. 1986. One H-chain V region gene segment subfamily in the BALB/c mouse contains 500-1000 or more members. *Cell.* 47:461.
9. Foon, K.A., and R.F. Todd. 1986. Immunologic classification of leukemia and lymphoma. *Blood.* 68:1.
10. Nickerson, K.G., J.E. Berman, E. Glickman, L. Chess, and F.W. Alt. 1989. Early human IgH gene assembly in Epstein-Barr virus-transformed fetal B cell lines. Preferential utilization of the most J_H-proximal D segments (DQ52) and two unusual V_H-related rearrangements. *J. Exp. Med.* 169:1391.
11. Nakamura, M., S.E. Burastero, A.L. Notkins, and P. Casali. 1988. Human monoclonal rheumatoid factor-like antibodies from CD5 (Leu-1)$^+$ B cells are polyreactive. *J. Immunol.* 140:4180.
12. Berman, J.E., K.G. Nickerson, R.R. Pollock, J.E. Barth, R.K.B. Schuurman, D.M. Knowles, L. Chess, and F.W. Alt. V_H Gene usage in humans: biased usage of the V_H6 gene in immature B lymphoid cells. *Eur. J. Immunol.* In press.
13. Jackson, R.J., and N. Standart. 1990. Do the poly(A) tail and 3′ untranslated region control mRNA translation? *Cell.* 62:15.
14. Sanz, I., P. Kelly, C. Williams, S. Scholl, P. Tucker, and J.D. Capra. 1989. The smaller human V_H gene families display remarkably little polymorphism. *EMBO (Eur. Mol. Biol. Organ.) J.* 8:3741.
15. Van Dijk, K.W., H.W. Schroeder, Jr., R.M. Perlmutter, and E.C.B. Milner. 1989. Heterogeneity in the human Ig V_H locus. *J. Immunol.* 142:2547.
16. Reth, M., P. Gehrmann, E. Petrac, and P. Wiese. 1986. A novel V_H to V_HDJ_H joining mechanism in H-chain-negative (null) pre-B cells results in H-chain production. *Nature (Lond.).* 322:840.
17. Kleinfield, R., R. Hardy, D. Tarlinton, J. Dangl, L. Herzenberg, and M. Weigert. 1986. Recombination between an expressed Ig H-chain gene and a germline V gene in a Ly-1$^+$ B-cell lymphoma. *Nature (Lond.).* 322:843.
18. Antin, J.H., S.G. Emerson, P. Martin, N. Gadol, and K.H. Ault. 1986. Leu-1$^+$(CD5$^+$) B cells. A major lymphoid subpopulation in human fetal spleen: phenotypic and functional studies. *J. Immunol.* 136:505.
19. Hardy, R.R., K. Hayakawa, M. Shimizu, K. Yamasaki, and T. Kishimoto. 1987. Rheumatoid factor secretion from human Leu1$^+$ B cells. *Science (Wash. DC).* 236:81.
20. Baer, R., A. Forster, I. Lavenir, and T.H. Rabbitts. 1988. Ig V_H genes are transcribed by T cells in association with a new 5′ exon. *J. Exp. Med.* 167:2011.
21. Kudo, A., and F. Melchers. 1987. A second gene, VpreB in the λ5 locus of the mouse, which appears to be selectively expressed in pre-B lymphocytes. *EMBO (Eur. Mol. Biol. Organ.) J.* 6:2267.