BAALC, the human member of a novel mammalian neuroectoderm gene lineage, is implicated in hematopoiesis and acute leukemia

Stephan M. Tanner*, Jamie L. Austin*, Gustavo Leone*†, Laura J. Rush*‡, Christoph Plass*, Kristiina Heinonen§, Krzysztof Mrózek†, Heinz Sill¶, Sakari Knuutila‖, Jonathan E. Kolitz, Kellie J. Archer†, Michael A. Caligiuri*†, Clara D. Bloomfield†, and Albert de la Chapelle*†,††**

*Human Cancer Genetics Program, Ohio State University, 646 Medical Research Facility, 420 West 12th Avenue, Columbus, OH 43210; ‡Department of Veterinary Biosciences, Ohio State University, Columbus, OH 43210; §Department of Clinical Genetics, Kuopio University Hospital, 70211 Kuopio, Finland; †Comprehensive Cancer Center, Ohio State University, Columbus, OH 43210; ¶Division of Hematology, Department of Medicine, Karl-Franzens University, 8010 Graz, Austria; ‖Department of Medical Genetics, Haartman Institute and Helsinki University Central Hospital, 00014 Helsinki, Finland; and **Don Monti Division of Medical Oncology/Division of Hematology, North Shore University Hospital, New York University School of Medicine, Manhasset, NY 11030

Contributed by Albert de la Chapelle, October 3, 2001

The molecular basis of human leukemia is heterogeneous. Cytogenetic findings are increasingly associated with molecular abnormalities, some of which are being understood at the functional level. Specific therapies can be developed based on such knowledge. To search for new genes in the acute leukemias, we performed a representational difference analysis. We describe a human gene in chromosome 8q22.3, *BAALC* (*b*rain *a*nd *a*cute *l*eukemia, *c*ytoplasmic), that is highly conserved among mammals but evidently absent from lower organisms. We characterized *BAALC* on the genomic level and investigated its expression pattern in human and mouse, as well as its complex splicing behavior. *In vitro* studies of the protein showing its subcellular localization suggest a function in the cytoskeleton network. Two isoforms are specifically expressed in neuroectoderm-derived tissues, but not in tumors or cancer cell lines of nonneural tissue origin. We show that blasts from a subset of patients with acute leukemia greatly overexpress eight different *BAALC* transcripts, resulting in five protein isoforms. Among patients with acute myeloid leukemia, those overexpressing *BAALC* show distinctly poor prognosis, pointing to a key role of the *BAALC* products in leukemia. Our data suggest that *BAALC* is a gene implicated in both neuroectodermal and hematopoietic cell functions.

Great progress has occurred in our understanding of the molecular biology of the human leukemias (1). Unlike the situation in most solid tumors, many leukemias are characterized by balanced translocations that give rise to gene rearrangements. In acute myeloid leukemia (AML), about 55% of adult *de novo* cases have clonal cytogenetic abnormalities (2), many of which are specific translocations in which the associated molecular mechanisms are beginning to be understood (1, 3). However, in the remaining cases, no cytogenetic abnormality is found and few clues exist to the molecular events involved. In adult acute lymphoblastic leukemia (ALL), the proportion of patients with no cytogenetic abnormality is ≈31% (4). It is vitally important to gather information regarding the basic events that underlie these molecularly unclassified acute leukemias.

In designing this study to search for novel genes in the acute leukemias, we reasoned that changes in novel genes might account for further cases of acute leukemias. Moreover, cases with trisomy for chromosome 8 (AML+8) might have molecular abnormalities in genes on chromosome 8, analogous to a tandem duplication of the *MLL* gene on chromosome 11 that was found to be associated with AML with trisomy for chromosome 11 (5, 6). Importantly, this also accounts for up to 11% of all cytogenetically normal AMLs (7). In search of putative novel leukemia-associated genes, we therefore studied gene expression in AML+8 cases and in cytogenetically normal AML. We chose cDNA-based representational difference analysis (cDNA-RDA), a method that compares relative expression levels and allows previously unknown genes to be detected and cloned (8, 9).

Here we report our finding of an expressed sequence tag (EST) that was overexpressed in two patients with AML+8 and belongs to the gene *BAALC* (*b*rain *a*nd *a*cute *l*eukemia, *c*ytoplasmic), which is located in human chromosome 8q22.3 and highly conserved in mammals. Normally, *BAALC* is almost exclusively expressed in neuroectoderm-derived tissues. In addition, CD34-positive progenitor cells from bone marrow (BM) express *BAALC*, identifying it as another gene whose expression is shared by neural and hematopoietic cells. Overexpression of *BAALC* was seen in patients with AML and ALL and strikingly, in a subset of AML it marked poor prognosis, suggesting a role for *BAALC* overexpression in acute leukemia.

MEDICAL SCIENCES

Materials and Methods

Human Samples. Pretreatment BM aspirate or peripheral blood samples from patients with different subtypes of leukemia were collected after prior consent. CD34-positive progenitor cells were enriched twice from normal BM aspirates by immunomagnetic separation, using MiniMACS columns (Miltenyi Biotec, Auburn, CA). All cell lines were obtained from American Type Culture Collection. Genomic DNA was extracted according to a standard proteinase K-phenol-chloroform protocol (10). Total RNA was isolated by using Trizol (Invitrogen) or purchased (CLONTECH).

cDNA-RDA. cDNA-based representational difference analysis was performed according to protocols kindly provided by D. G. Schatz (Yale Univ., New Haven, CT) and M. J. O'Neill (Princeton Univ., Princeton) with some modifications. Briefly, 20 μg of total RNA (pooled from three cases with AML+8 for the tester and from four cytogenetically normal cases with AML for the driver) was poly(A)-selected and transcribed into cDNA by using Dynabeads Oligo dT_{25} (Dynal, Great Neck, NY) and Superscript II reverse transcriptase (Invitrogen). The cDNA was *Dpn*II-

Abbreviations: AML, acute myeloid leukemia; BM, bone marrow; PBL, peripheral blood leukocytes; ALL, acute lymphoblastic leukemia; RDA, representational difference analysis; EST, expressed sequence tag; CNS, central nervous system; RT, reverse transcription.

Data deposition: The sequences reported in this paper have been deposited in the GenBank database [accession nos. AF363578 (human *BAALC* genomic region) and AF371319–AF371326 (human, mouse, rat, and pig *BAALC* coding sequences)].

††To whom reprint requests should be addressed. E-mail: delachapelle-1@medctr.osu.edu.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

digested (New England Biolabs), R-adapters were added with T4 ligase (New England Biolabs), and tester and driver representations were amplified by using Expand High-Fidelity DNA polymerase (Roche Diagnostics). R-adapters were removed by *Dpn*II digest. This step concluded the processing of the driver. For the tester, J-adapters were added by ligation and, the first round of subtractive hybridization was performed at a ratio of 100:1 (driver:tester). Two rounds of PCR with an intermediate mung-bean nuclease digest (New England Biolabs) were performed for difference product 1 (DP1). For a second round of subtractive hybridization, the J-adapters were replaced by N-adapters, and the ratio was increased to 800:1. DP2 was cloned into the *Bam*HI site of pZERO-1 (Invitrogen). Inserts were PCR-amplified by using M13 universal primers and sequenced on an ABI 377 sequencer (Applied Biosystems). Sequences were used in BLAST searches against GenBank. A detailed cDNA-RDA protocol is available on request.

Cloning of Human *BAALC* and Mouse *Baalc*. RDA clone sequences representing partial cDNA clones from *BAALC* were used to design PCR primers for reverse transcription (RT)-PCR, cDNA, and BAC clone sequencing. BAC clones 680F3, 318M2, and 754L23 were from human BAC library RPCI-11 (11). To sequence mouse *Baalc* primers were designed from EST sequences and used for RT-PCR. Coding sequences of all human and mouse transcripts were RT-PCR-amplified from brain or AML blast RNA and cloned into expression vectors pcDNA3 (Invitrogen) and Myc epitope tagging pcDNA3–5xMyc (gift of C. H. Park, Seoul, South Korea). For mapping, the T31 mouse radiation hybrid panel was purchased (Research Genetics, Huntsville, AL) and screened in duplicate by PCR. Results were sent to The Jackson Laboratory for linkage analysis (12).

Northern Blot Analysis. Human *BAALC* transcript 1-6-8 and mouse *Baalc* transcript 1-6-8 were random primer-labeled (Roche Diagnostics) and used to probe species-specific multi-tissue Northern blots (CLONTECH). All blots were probed with human β-actin cDNA as a control.

Expression of BAALC in 293 and NIH/3T3 Cells. All pcDNA3 and pcDNA3–5xMyc-based *BAALC* clones were CsCl-banding-purified twice (10). Equivalent amounts of vector only or expression clone DNA were transfected by the $CaCl_2$-HBS method (10) into human 293 or mouse NIH/3T3 cells grown in DMEM and 10% FBS (both Invitrogen) under 6% CO_2 atmosphere. Cotransfection with vector pcDNA3-GFP expressing green fluorescent protein (GFP) served as a transfection control. Transfections were washed after 14 h twice with PBS and grown for another 24 h for immunoprecipitation and Western blot analysis or fixed with 4% paraformaldehyde for *in situ* immunocytochemistry.

Immunoprecipitation and Western Blot Analysis. After transfection, 293 cells were harvested 40 h after transfection and lysed in RIPA buffer (0.15 mM NaCl/0.05 mM Tris·HCl, pH 7.2/1% Triton X-100/1% sodium deoxycholate/0.1% SDS) (10). Cleared total cell lysate was used for immunoprecipitation by using 9E10 mouse mAb (Santa Cruz Biotechnology) and protein G plus-agarose (Oncogene) against N-terminal Myc epitope tags added to all BAALC isoforms by vector pcDNA3–5xMyc. Proteins were separated by 15% SDS/PAGE and blotted (10) onto poly(vinylidene difluoride) (PVDF) membranes (Millipore). BAALC proteins were detected by either 9E10 (1:1000) or anti-BAALC-specific GN-2214 (1:200) Abs. Antigen-Ab complexes were visualized by using 1:10,000 diluted horseradish peroxidase-conjugated goat anti-mouse (NA 931) or anti-rabbit (NA 934) Ig with an enhanced chemiluminescence system (all Amersham Pharmacia). The polyclonal Ab GN-2214 was raised commercially (Sigma-Genosys) in rabbits against amino acids 9–26 of human isoform 1-6-8 (DAIEPRYYESWTRETEST).

Immunocytochemistry. NIH/3T3 cells were transfected, incubated overnight, and fixed with 4% paraformaldehyde. Primary Abs 9E10 (1:1000) or GN-2214 (1:200) were applied, and species-specific rhodamine-conjugated secondary anti-IgG Abs (Roche Diagnostics no. 605140 for mouse or Chemicon no. AP307R for rabbit) were used to visualize protein–Ab complexes with a fluorescence microscope (Nikon Eclipse E800). Nuclei were stained with 4′,6-diamidino-2-phenylindole (DAPI; Roche Diagnostics).

Comparative RT-PCR and 5′ Nuclease Assay (TaqMan). Total RNA from patients with AML was reverse-transcribed with avian myeloblastosis virus reverse transcriptase (Roche Diagnostics) from the same original RNA extraction used for cDNA-RDA. For comparative RT-PCR and TaqMan 5′ nuclease assays, both *BAALC* and the housekeeping gene *GPI* (*g*lucose *p*hosphate *i*somerase; ref. 13) were coamplified in the same tube. The TaqMan assays were carried out for each sample in triplicates. Primer and TaqMan oligo probe sequences and conditions for PCR are available on request. To determine the relative levels of expression of *BAALC* among the tumor and normal tissue samples, the comparative C_T method was used (Applied Biosystems). First, the parameter threshold cycle (C_T) was determined for *BAALC* and *GPI*, and the cycle number difference ($GPI - BAALC = \Delta C_T$) was calculated for each replicate. If *BAALC* failed to reach the software-set threshold, the sample was considered below detection limit. If *GPI* amplification failed, the sample was omitted from the analysis. Finally, the mean ΔC_T from the three replicates was generated [$(\Sigma \Delta C_T)/3 = MC_T$], normalizing *BAALC* expression to *GPI* expression.

Statistical Analysis. Samples obtained from the Cancer and Leukemia Group B (CALGB) Leukemia Tissue Bank for patients included in the clinical analysis were from adult *de novo* patients with AML with evaluable cytogenetics enrolled on a prospective cytogenetic companion study, CALGB 8461 (14), who had peripheral blood samples with >50% blasts. Cases with AML with $MC_T \geq 0.75$ were classified as *BAALC*-positive. Event-free survival was measured from the on-study date until date of treatment failure, relapse, or death, censoring only for patients alive and in continuous complete remission.

Results

cDNA-RDA Comparison of AML+8 with Cytogenetically Normal AML. We performed a cDNA-RDA experiment by using mRNA of AML without cytogenetic abnormalities as the driver and mRNA from AML+8 as the tester. After subtractive hybridization, sequencing of 209 RDA clones detected a total of 27 genes, 5 ESTs, and 5 unknown sequences (data not shown). The most frequently detected sequences (65%) belonged to the well known *DNTT* gene (OMIM 187410). The second most frequent sequence (3.1%) was EST clone AA400649, belonging to UniGene cluster Hs.169395 in chromosome 8q22.3. This putative unique gene, later named *BAALC* for *b*rain *a*nd *a*cute *l*eukemia, *c*ytoplasmic, was chosen for further study.

***BAALC*, a Unique Gene on Chromosome 8, Is Neuroectoderm- and Hematopoiesis-Specific.** A combination of primer walking, cDNA clone sequencing, and RT-PCR allowed us to assemble the gene (GenBank accession no. AF363578), which is fully contained in BAC clone 754L23 (Fig. 1*A*). Most ESTs from *BAALC* (UniGene Hs.169395) were from central nervous system (CNS)-derived cDNA libraries, and we found exon 6 to be alternatively spliced in brain RNA (Fig. 2*A*). Thus, the gene was expressed in two transcripts, transcript 1-6-8 (2827 bp) and transcript 1-8

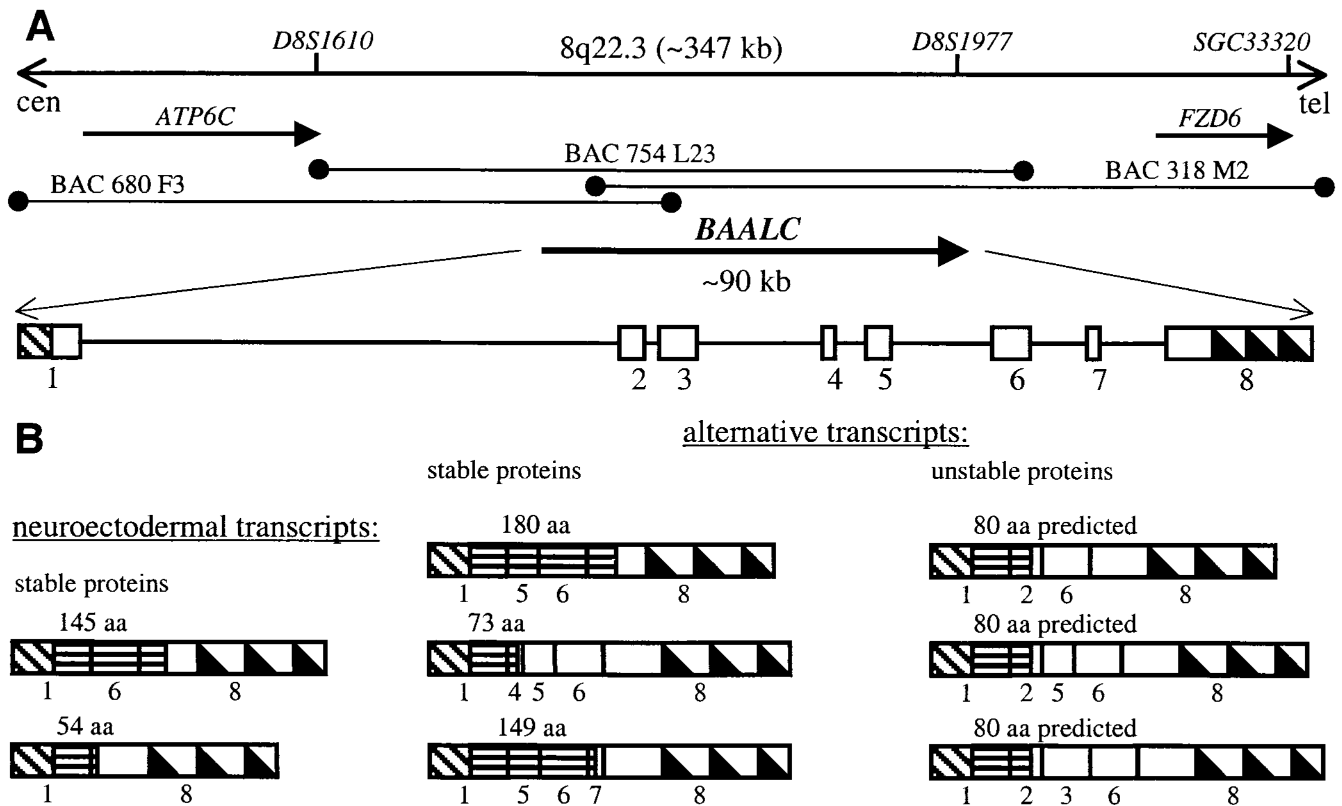

Fig. 1. Genomic organization and transcripts of the human *BAALC* gene. (*A*) *BAALC* is located between *ATP6C* proximal and *FZD6* distal in chromosome 8q22.3 and covers 90 kb of genomic sequence. We assembled the region in the three BAC clones shown. *BAALC* consists of eight exons depicted by boxes, with exon 8 containing three polyadenylation signals (triangles) in the 3′ untranslated region (UTR) leading to three differently sized transcripts (Fig. 3*A*). Exon 1 contains the ATG start codon and a 5′ UTR (diagonally striped). (*B*) The two transcripts 1-6-8 and 1-8 are seen in neuroectoderm tissues, and six more transcripts produced by alternative splicing occur in hematopoietic cells, chiefly in leukemic blasts. The extent of the resulting coding regions are horizontally striped, and the protein sizes are indicated above.

(2660 bp) with predicted ORFs of 145- and 54-aa, respectively (Fig. 1*B*). By comparative RT-PCR, BM showed barely detectable, and peripheral blood leukocytes (PBL) showed no expression of *BAALC* (Fig. 2*A*). As shown below, *BAALC* expression in BM was confined to the CD34-positive progenitor cells.

To explore the tissue specificity of *BAALC*, we assayed multi-tissue Northern (MTN) blots by using transcript 1-6-8 as a probe (Fig. 3*A*). *BAALC* expression was specific for neuroectoderm-derived tissues. In all other tissues, notably BM, PBL, and lymph nodes, no detectable levels of *BAALC* were recorded.

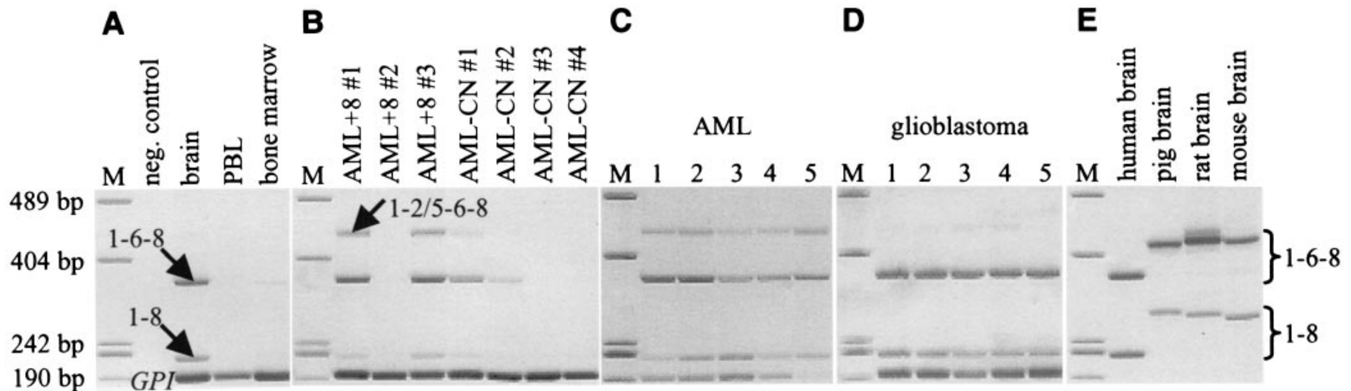

Fig. 2. (*A–D*) Comparative RT-PCR using primers in exons 1 and 8 of human *BAALC* in brain, PBL, BM, 12 cases with AML, and 5 glioblastoma tumors with *GPI* serving as an internal control. M, size marker. (*A*) Exon 6 is alternatively spliced in brain and leads to transcripts 1-6-8 and 1-8. Transcript 1-6-8 is more highly expressed than 1-8 in brain. No *BAALC* expression is detected in PBL, but faint expression occurs in BM. (*B*) The three cases of AML with +8 and four cases of AML with normal karyotype (AML-CN) used in our cDNA-RDA experiment were studied individually by RT-PCR. AML+8 nos. 1 and 3 show very high levels of transcript 1-6-8, whereas AML-CN nos. 1 and 2 show them weakly. Moreover, two alternative transcripts, 1-2-6-8 and 1-5-6-8, are observed that are absent in brain. (*C*) Five cases with AML show the presence of the alternative transcripts when *BAALC* is overexpressed. (*D*) Although the five glioblastoma tumor samples show transcripts 1-6-8 and 1-8 are highly expressed, they are distinguished by their virtual lack of expression of the alternative transcripts. (*E*) Conservation of *BAALC* splicing of transcripts 1-6-8 and 1-8 in brain samples from four mammalian species shown by RT-PCR with primers in exons 1 and 8.

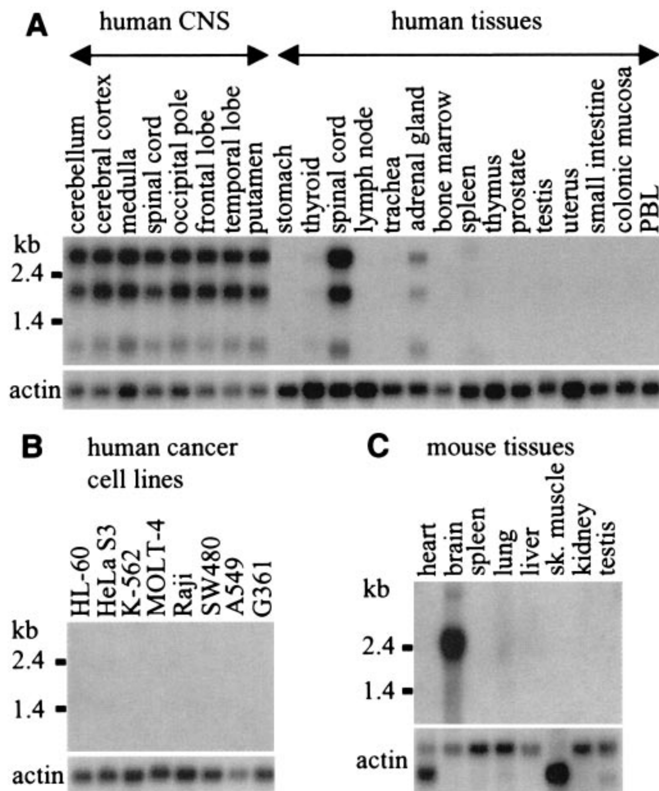

Fig. 3. Tissue-expression pattern of *BAALC*. (*A*) Commercial Northern blots probed with human transcript 1-6-8. High expression of *BAALC* is restricted to neural tissues, and low expression is seen in the neuroectoderm-derived tissues adrenal gland, thyroid, and spleen. Note undetectable levels of *BAALC* in BM, PBL, and lymph nodes, as well as in eight human cancer cell lines (*B*). *BAALC* is expressed as three differently sized transcripts of about 1, 2, and 3 kb, because of the alternative usage of the three poly(A) signals in exon 8. (*C*) A commercial mouse Northern blot probed with *Baalc* transcript 1-6-8 shows a major 2.7-kb transcript exclusively in brain, indicating the same neuroectoderm-specific expression as in human, with a clear preference for the second poly(A) signal.

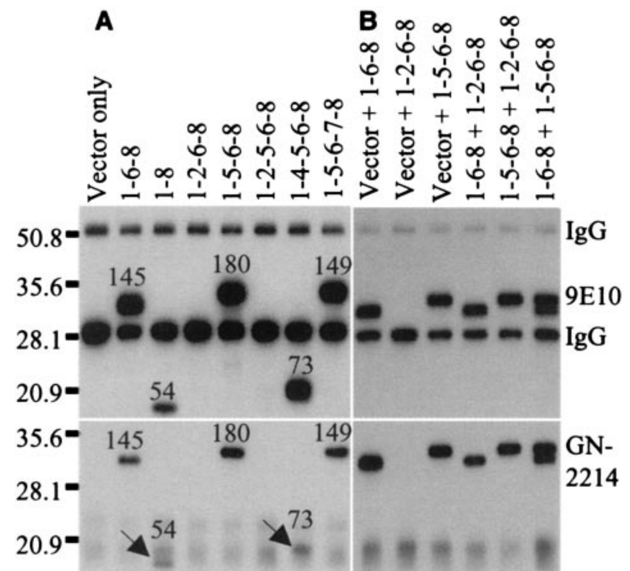

Fig. 4. BAALC protein expression. (*A*) Immunoprecipitation-Western blot analysis of isoforms 1-6-8 and 1-8, and the five alternative isoforms 1-2-6-8, 1-5-6-8, 1-2-5-6-8, 1-4-5-6-8, and 1-5-6-7-8 cloned into pcDNA3–5xMyc and transfected into 293 cells. BAALC protein detection was with anti-Myc mouse Ab 9E10 against the N-terminal Myc tag (*Upper*) or with anti-BAALC exon 1-specific rabbit Ab GN-2214 (*Lower*). Both neuroectodermal transcripts produce the expected proteins of 145- and 54-aa, as do three of the five alternative isoforms (180-, 73-, and 149-aa). Arrows indicate the small 54- and 73-aa isoforms. For the two alternative isoforms 1-2-6-8 and 1-2-5-6-8, no protein expression is detected, indicating potential transcript or protein instability. Note the corresponding signals with GN-2214. All isoforms contain an additional 63 amino acids from the N-terminal Myc tag. The secondary anti-mouse IgG Ab detected the IgG heavy and light chains from 9E10 (used for immunoprecipitation) on the *Upper* blot, which are not seen with the secondary anti-rabbit IgG Ab used on the *Lower* blot. (*B*) Coexpression in 293 cells of neuroectodermal isoform 1-6-8, stable isoform 1-5-6-8, and unstable isoform 1-2-6-8 with vector and with one another as indicated above the lanes. Only isoform 1-2-6-8 fails to express a protein, either alone or together with 1-6-8 or 1-5-6-8, and has no discernible effect on the stable isoforms. The last lane depicts coexpression of both stable isoforms. The results are confirmed with both 9E10 (*Upper*) and GN-2214 (*Lower*).

Furthermore, an MTN blot containing eight human cancer cell lines tested completely negative for *BAALC* (Fig. 3*B*).

Alternative Splicing of *BAALC* in Acute Leukemia. By RT-PCR we detected transcripts in AML that we did not see in brain (Fig. 2*B*). Cloning of these showed the presence of 5 new exons that we numbered 2–5 and 7, which led to 6 additional alternatively spliced transcripts. Including the transcripts 1-6-8 and 1-8, we detected 8 alternatively spliced transcripts (Fig. 1*B*) in the blasts of patients with acute leukemia who were expressing high levels of *BAALC*.

***BAALC* Encodes Several Isoforms of a Protein that Is Cytoplasmic.** Conceptual translation of the 8 different transcripts of *BAALC* observed in acute leukemia blasts suggested 6 different protein isoforms, including 2 (1-6-8 and 1-8) that also occur in brain. All have the same first 53 amino acids encoded by exon 1, but the remaining coding sequences are different. The use of exon 2 in 3 of the transcripts predicted the same 80-aa protein, as the sequence of exon 2 comprises a termination codon. Database similarity searches failed to produce any significant alignments to described proteins.

We cloned the coding sequences of all transcripts except 1-2-3-6-8 into an expression vector that added 63 amino acids containing 5 Myc epitopes to the N terminus of all isoforms. After expressing the clones in 293 cells, we immunoprecipitated the resulting proteins with anti-Myc mAb and assayed for BAALC expression by Western blot analysis. We detected protein expression for neuroectodermal isoforms 1-6-8 and 1-8, and for the 3 isoforms 1-5-6-8, 1-4-5-6-8, and 1-5-6-7-8 (Fig. 4*A*). The two isoforms containing exon 2, 1-2-6-8 and 1-2-5-6-8, did not express protein, indicating that the termination in exon 2 probably conveyed instability to either the resulting transcript or the peptide after translation.

That at least two transcripts were not translated into stable proteins raised the question whether their presence might affect the stability of the other five isoforms. We therefore cotransfected neuroectodermal isoform 1-6-8, stable isoform 1-5-6-8, and unstable isoform 1-2-6-8 with one another and with control plasmid (Fig. 4*B*). There was no obvious negative impact of the unstable isoform on the stability of the other proteins.

To study the intracellular localization, we transfected NIH/3T3 cells with expression constructs from all five stable transcripts. By fluorescence microscopy we localized all isoforms to the cytoplasm, where they appeared as a few large inclusions in the cellular periphery (Fig. 5 *A–C*). We did not observe any differences of localization among the five isoforms or between the native pcDNA3 and Myc-tagged expression constructs (data not shown).

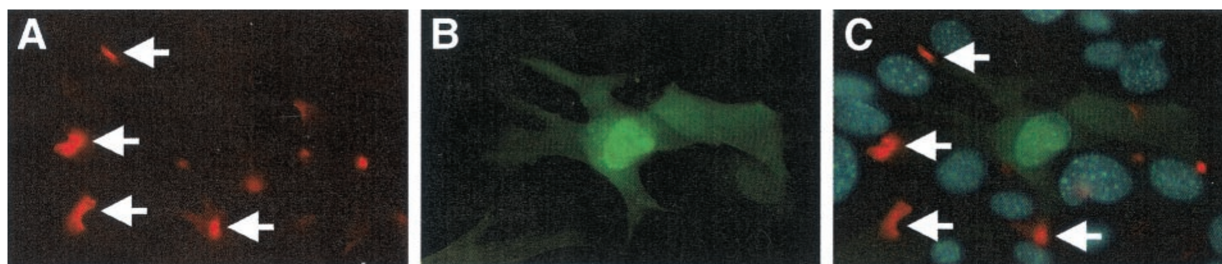

Fig. 5. Subcellular localization of BAALC protein. (*A*) Subcellular localization of human BAALC is shown after transfection of Myc-tagged isoform 1-6-8 into NIH/3T3 cells and staining with 9E10 and rhodamine-conjugated secondary Ab. (*B*) pcDNA3-green fluorescent protein served as a transfection control. (*C*) Triple-filter image. Nuclei were stained with 4′,6-diamidino-2-phenylindole (DAPI). BAALC occurs as a few large inclusions (arrows) in the peripheral parts of the cytoplasm.

***BAALC* Is Conserved in Mammals.** During database searches for *BAALC* orthologs in other species, we detected ESTs from mouse (UniGene Mm.44234 and Mm.85430) and rat (Rn.19969 and Rn.57689). In addition, we found three domestic pig ESTs (BF190130, BF192691, and BF193189) that when assembled contained the complete orthologous ORF for isoform 1-6-8. The putative 145-aa protein was 92% conserved compared with the human. However, we did not find any significant similarities of *BAALC* to sequences from lower organisms, especially *Drosophila melanogaster*, *Caenorhabditis elegans*, or *Saccharomyces cerevisiae*.

From deposited mouse ESTs we predicted that the mouse expresses orthologous transcripts to human *BAALC* 1-6-8 and 1-8. The predicted protein isoform 1-6-8 was 89% conserved compared with the human. We performed RT-PCR on brain RNA from human, domestic pig, rat, and mouse and confirmed the expression of both orthologous transcripts 1-6-8 and 1-8 in all species (Fig. 2*E*). Rat isoform 1-6-8 was highly similar (98%) to the mouse and was also 89% conserved compared with the human. Moreover, pig, mouse, and rat isoforms 1-8 were all 95% conserved compared with human (GenBank accession nos. AF371319–AF371326).

Thus far, mouse cDNA sequences corresponding to the five human alternative exons have not been found, but sequencing revealed that exon 8 of *Baalc* has three potential polyadenylation signals similar to the human. A mouse multitissue Northern blot showed clear preference for the second polyadenylation signal, with high expression of a 2.7-kb transcript in brain (Fig. 3*C*).

Finally, we mapped the mouse *Baalc* gene locus with the T31 radiation hybrid panel to proximal mouse chromosome 15. Our data had the highest anchor logarithm of odds (lod) score of 15.6 to *D15Mit112*, then lod score 14.4 to *D15Mit6*, and finally lod score 11.5 to *D15Mit22*. The best-fit location was between *D15Mit112* proximal and *D15Mit85* distal.

Prevalence of *BAALC* Overexpression in Acute Leukemia. Although *BAALC* overexpression was first detected in two cases with AML+8 (Fig. 2*B*), screening by comparative RT-PCR resulted in 5 of 27 diverse patients with AML being scored *BAALC*-positive by visual comparison to *GPI* (Fig. 2*C*). To more accurately quantify the expression levels of *BAALC* in a larger set of diverse leukemia samples, we switched to real-time PCR.

We started by examining *BAALC* expression levels in diverse human tissues and tested normal tissue RNA samples from several organs including BM, PBL, brain and fetal brain, plus different human tumor samples. We analyzed RNA samples from 10 normal human tissues, 10 colorectal cancer–normal colonic mucosa-matched sample pairs, 3 esophageal cancer–normal esophagus-matched sample pairs, 3 lung tumor samples, 5 glioblastoma samples, 2 thyroid carcinoma–normal thyroid-matched sample pairs, 1 thyroid carcinoma sample, and 3 testicular and 3 mammary tumor samples for expression of *BAALC*. Aside from the cases with AML previously scored as *BAALC*-positive, only glioblastoma tumors showed comparable levels of *BAALC* expression. All five cases with glioblastoma strongly expressed *BAALC* in the range of 0.74–3.71 MC_T, but virtually only transcripts 1-6-8 and 1-8 (Fig. 2*D*), distinguishing them from the acute leukemias (Fig. 2*C*). We decided to use an arbitrary cut off of $MC_T \geq 0.75$ to score leukemia samples as positive. The MC_T for normal tissues including brain, fetal brain, spleen, and several normal BM samples were well below this threshold in the range of -12.4–0.14 MC_T. As we detected low expression in five BM samples (-10.57 to -7.2 MC_T), we twice immunomagnetically enriched CD34-positive progenitor cells of two healthy donors. The >95% pure CD34-positive hematopoietic progenitor populations expressed *BAALC* in the range of -4.2 to -2.0 MC_T, whereas the CD34-negative fractions were below detection limit.

As an initial leukemia study, we tested blasts from BM or peripheral blood from 130 diverse AML, 31 ALL, 4 Burkitt's lymphoma (BL), 5 chronic myelogenous leukemia (CML), 5 chronic lymphocytic leukemia (CLL), plus the 7 leukemia cell lines HL-60, KG-1, KG-1a, MC-1010, K-562, D1.1, and RS4;11 for *BAALC* expression. In AML 37/130 (28%) and in ALL 20/31 (65%) were classified as *BAALC*-positive (0.75–8.59 MC_T), but none of the BL, CML, and CLL samples, or leukemia cell lines scored positive (below detection limit to 0.46 MC_T).

Clinical Implications of *BAALC* Overexpression in AML. To investigate whether *BAALC* overexpression had clinical consequences, we statistically analyzed 102 adult *de novo* cases with AML from which we had peripheral blood samples and centrally reviewed clinical data available. Of 102 patients, 29 (28%) were *BAALC*-positive. The distribution of *BAALC* overexpression differed significantly among AML French-American-British subtypes ($P < 0.0001$). It was present in all 5 cases of M0, in 12/28 M1, 7/28 M2, and 3/4 M4Eo cases, but in no M3 (0/6), and only in 1/14 M4 and 1/17 M5 cases. The association with cytogenetic subtype was also nonrandom ($P = 0.0001$). *BAALC* overexpression was seen in 2/2 *inv*(3)/*t*(3;3), 6/8 core binding factor leukemias, and 3/4 isolated trisomy 8, but in 0/6 cases of *t*(15;17) and only 1/9 cases of *t*(11q23). Of 63 patients with normal karotypes, 12 (19%) overexpressed *BAALC*. Overexpression of *BAALC* was an adverse prognostic factor. Among the 29 *BAALC*-positive patients, the median of event-free survival was 0.4 years compared with 1.2 years for the 73 *BAALC*-negative patients ($P = 0.006$).

Discussion

The unique *BAALC* gene is expressed in cells of the CNS and at low levels in other neuroectoderm-derived tissues such as the adrenal gland. In 3T3 cells transfected with *BAALC*, the protein

is cytoplasmic, often with a peculiar inclusion-like localization in the periphery of the cell that suggests a role in cell locomotion or adhesion. That the *BAALC* gene is highly conserved among mammals shows an identical pattern of expression in the CNS in several species (e.g., humans, rats, mice, and pigs), and apparently lacks a conserved ortholog in lower vertebrates such as zebrafish or *Xenopus*, suggests a particular role in the CNS of mammals.

Of major interest are the data on *BAALC* expression in human hematopoietic cells. There is no detectable expression in leukocytes from blood, whereas in BM cells very faint expression is visible by comparative RT-PCR but not by Northern blotting. *BAALC* seems to occur only in the CD34-positive progenitor cells of the BM, at a level comparable to normal brain. In marked contrast, blast cells from 28% of patients with AML and 65% with ALL displayed strong overexpression of *BAALC*. Three facts allow us to propose that overexpression of *BAALC* in these blast cells represents a leukemia-associated change. First, the leukemic blasts express six transcripts that seem to be absent in neuroectodermal tissues such as brain and spleen and virtually absent in glioblastoma tumors, which similarly overexpress *BAALC*. Second, with the exception of glioblastoma, none of the other cancer cells tested showed *BAALC* expression, ruling it out as a general marker for neoplasia. Third, *BAALC* overexpression is clearly nonrandom among AML French-American-British subtypes and AML cytogenetic groups, pointing to a role of *BAALC* in the leukemic phenotype. Its association with an adverse prognosis in AML is a further indication of it having a role in one or more processes that characterize these blasts. In ALL, the situation seems different. A high percentage of patients shows overexpression, but small numbers in the different immunophenotypic and cytogenetic groups preclude a detailed examination. As for a possible mechanism for overexpression, we excluded intragenic mutations by DNA sequencing and by Southern hybridization (data not shown).

Nonrandom *BAALC* overexpression among different AML French-American-British subtypes shows a clear association with the more immature blasts. Thus, in the hematopoietic system, *BAALC* expression is found both in normal early progenitor cells and in the most immature types of blasts in acute leukemia, but not in mature hematopoietic cells. It remains unclear what functional aspects are shared by hematopoietic and those neuroectodermal cell types that express *BAALC*, but an implication in cell movement, adhesion, or cell–cell interaction should be considered. Recent research suggests that several genes are common among progenitor cells of the neural and hematopoietic systems (15). *BAALC* might be one of these shared genes postulated to maintain proliferative capacity while inhibiting differentiation. Our observation of increased *BAALC* expression in leukemic blasts and glioblastoma compared with normal BM and brain, supports this hypothesis. We anticipate using animal models to elucidate the role of *BAALC* in hematopoiesis and leukemogenesis, as well as in CNS cells and tumors.

The design of our project was to search for genes that might shed light on the basic mechanisms in molecularly unclassified acute leukemias. We hypothesized that genes in chromosome 8 might play a role, and that patients with AML+8 might be likely to provide evidence thereof. The strategy of using cDNA-RDA to compare AML+8 leukemic cells with cytogenetically normal AML cells did indeed disclose overexpression of a gene on chromosome 8. However, it now seems unlikely that the overexpression of *BAALC* is directly related to trisomy for chromosome 8, because this overexpression is not only seen in AML+8. Instead, *BAALC* overexpression is seen in AML and ALL blasts with different cytogenetic abnormalities, as well as in blasts with normal karyotypes.

In summary, we present the case of a unique, highly conserved neuroectodermal gene that is overexpressed in a subset of patients with acute leukemia and associated with poor prognosis in AML. The mechanistics and functions, in neuroectoderm-derived tissues and tumors, hematopoietic progenitor cells, and leukemic blasts, remain to be explained.

We thank D. Bucci, A. J. Carroll, Z. Dai, C. Edwards, C. Eng, R. Krahe, J. Lockman, G. Marcucci, H. Nakagawa, C. H. Park, N. Pellegata, A. de la Puente, L. Rowe, H. Saavedra, D. Smiraglia, J. W. Vardiman, K. Virtaneva, S. Whitman, F. Wright, L. Wu, Y. Z. Wu, B. Yuan, and X. P. Zhou for advice, reagents, and help with some experiments. We are grateful to D. G. Schatz and M. J. O'Neill for detailed protocols of cDNA-RDA, and to members of the Vogelstein and Kinzler laboratories for helpful discussions. This work was supported by National Institutes of Health Grants CA16058, CA77658 (both to C.D.B.), and CA31946 (to C.D.B. and M.A.C.); the Leukemia Clinical Research Foundation (C.D.B.); and by National Cancer Institute T32 Oncology Training Grant CA09338 (to L.J.R. and M.A.C.). S.M.T. is the recipient of Swiss National Science Foundation Fellowship No. 81BE-53485 and Fellowship No. 965/2336 from the Swiss Foundation for Medical-Biological Grants (SSMBS).

1. Look, A. T. (1997) *Science* **278,** 1059–1064.
2. Mrózek, K., Heinonen, K. & Bloomfield, C. D. (2001) *Baillieres Best Pract. Res. Clin. Haematol.* **14,** 19–47.
3. Caligiuri, M. A., Strout, M. P. & Gilliland, D. G. (1997) *Semin. Oncol.* **24,** 32–44.
4. Wetzler, M., Dodge, R. K., Mrózek, K., Carroll, A. J., Tantravahi, R., Block, A. W., Pettenati, M. J., Le Beau, M. M., Frankel, S. R., Stewart, C. C., *et al*. (1999) *Blood* **93,** 3983–3993.
5. Schichman, S. A., Caligiuri, M. A., Gu, Y., Strout, M. P., Canaani, E., Bloomfield, C. D. & Croce, C. M. (1994) *Proc. Natl. Acad. Sci. USA* **91,** 6236–6239.
6. Caligiuri, M. A., Strout, M. P., Schichman, S. A., Mrózek, K., Arthur, D. C., Herzig, G. P., Baer, M. R., Schiffer, C. A., Heinonen, K., Knuutila, S., *et al*. (1996) *Cancer Res.* **56,** 1418–1425.
7. Caligiuri, M. A., Strout, M. P., Lawrence, D., Arthur, D. C., Baer, M. R., Yu, F., Knuutila, S., Mrózek, K., Oberkircher, A. R., Marcucci, G., *et al*. (1998) *Cancer Res.* **58,** 55–59.
8. Hubank, M. & Schatz, D. G. (1994) *Nucleic Acids Res.* **22,** 5640–5648.
9. O'Neill, M. J. & Sinclair, A. H. (1997) *Nucleic Acids Res.* **25,** 2681–2682.
10. Sambrook, J., Fritsch, E. F. & Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Lab. Press, Plainview, NY), 2nd Ed.
11. Osoegawa, K., Mammoser, A. G., Wu, C., Frengen, E., Zeng, C., Catanese, J. J. & de Jong, P. J. (2001) *Genome Res.* **11,** 483–496.
12. Rowe, L. B., Barter, M. E. & Eppig, J. T. (2000) *Genomics* **69,** 27–36.
13. Krahe, R., Ashizawa, T., Abbruzzese, C., Roeder, E., Carango, P., Giacanelli, M., Funanage, V. L. & Siciliano, M. J. (1995) *Genomics* **28,** 1–14.
14. Byrd, J. C., Lawrence, D., Arthur, D. C., Pettenati, M. J., Tantravahi, R., Qumsiyeh, M., Stamberg, J., Davey, F. R., Schiffer, C. A. & Bloomfield, C. D. (1998) *Clin. Cancer Res.* **4,** 1235–1241.
15. Geschwind, D. H., Ou, J., Easterday, M. C., Dougherty, J. D., Jackson, R. L., Chen, Z., Antoine, H., Terskikh, A., Weissman, I. L., Nelson, S. F. & Kornblum, H. I. (2001) *Neuron* **29,** 325–339.