Mammalian Genome 6, 114–117 (1995).

Mammalian Genome

Characterization and mapping of three new mammalian ATP-binding transporter genes from an EST database

R. Allikmets,[1] B. Gerrard,[2] D. Glavač,[2] M. Ravnik-Glavač,[2] N.A. Jenkins,[3] D.J. Gilbert,[3] N.G. Copeland,[3] W. Modi,[2] M. Dean[1]

[1]Laboratory of Viral Carcinogenesis, National Cancer Institute, Frederick Cancer Research and Development Center, Building 560, Room 21-19, Frederick, Maryland 21702-1201, USA
[2]Biological Carcinogenesis and Development Program, Program Resources, Inc./DynCorp, Frederick Cancer Research and Development Center, Frederick, Maryland 21702-1201 USA
[3]Mammalian Genetics Laboratory, ABL-Basic Research Program, National Cancer Institute, Frederick Cancer Research and Development Center, Frederick, Maryland 21702-1201, USA

Received: 26 July 1994 / Accepted: 6 October 1994

Abstract. Analysis of the human expressed sequence tag (EST) database identified four clones that contain sequences of previously uncharacterized genes, members of the ATP-binding cassette (ABC) superfamily. Two new ABC genes (EST20237, 31252) are located at Chromosome (Chr) 1q42 and 1q25 respectively in humans, as determined by FISH; at locations distinct from previously mapped genes of this superfamily. Two additional clones, EST 600 and EST 1596, were found to represent different ATP-binding domains of the same gene, ABC2. This gene was localized to 9q34 in humans by FISH and to the proximal region of Chr 2 in mice by linkage analysis. All genes display extensive diversity in sequence and expression pattern. We present several approaches to characterizing EST clones and demonstrate that the analysis of EST clones from different tissues is a powerful approach to identify new members of important gene families. Some drawbacks of using EST databases, including chimerism of cDNA clones, are discussed.

Introduction

The ATP-binding cassette (ABC) superfamily is one of the largest gene families and encodes a functionally diverse group of proteins principally involved in the transport of molecules across membranes. To date at least 25 ABC genes have been identified in *E. coli* alone (Ames and Lecar 1992). These genes are involved in the transport of sugars, amino acids, ions, peptides, and proteins (Higgins 1992). In humans only 10 members of the family have been described. These include a number of genes involved in diseases, that is, cystic fibrosis transmembrane conductance regulator (CFTR; Riordan et al. 1989), adrenoleukodystrophy gene (ALD; Mosser 1993), and/or other important biological processes, such as P-glycoproteins (PGY1, PGY3; Gros et al. 1986) and antigen peptide transporters (TAP1, TAP2; Spies et al. 1990). In humans, ABC genes typically encode four domains consisting of two ATP-binding segments and two transmembrane segments (Higgins 1992). The ABC transporter proteins share significant structural homology within the ATP-binding domains and can be clearly distinguished from other ATP-binding proteins such as kinases (Higgins 1992). The fact that prokaryotes contain a large number of ABC genes suggests that many mammalian members of the superfamily remain uncharacterized. As an approach to characterize all human genes, several laboratories have begun sequencing portions of cDNA clones from different tissues (Adams et al. 1993; Takeda et al. 1993). Such expressed sequence tags (ESTs) are useful for identifying new genes, as tools for genetic mapping, and for the analysis of gene expression (Grausz and Auffray 1993). We have begun to explore the usefulness of using ESTs to characterize a large gene family structurally and, ultimately, functionally.

Materials and methods

Sequence analysis. Searches of the dbEST (Boguski et al. 1993) database were performed with BLAST on the NCBI file service (Altschul et al. 1990). Amino acid alignments were generated with PILEUP (Feng and Doolittle 1987). Sequences were read with an IBI Gel Reader and analyzed with the Genetics Computer Group package of programs (Devereaux et al. 1984) running on a VAX computer. The sequences have been deposited with GENBANK under accession #U18235-7.

Polymerase chain reactions. Primers to EST sequences were designed with the PRIMER program (Lincoln et al. 1991). Reactions were performed in 1× PCR buffer (Boehringer Mannheim). Samples were heated to 96°C for 5 min, amplified for 35–40 cycles of 96°C, 30 s; 56°C, 30 s; 72°C, 1 min. PCR products were analyzed on 1.5–2% agarose gels after digestion with appropriate restriction enzymes to verify their sequence, if needed.

Northern hybridization. DNA fragments used as probes were purified on a 1% low-melting temperature agarose gel. DNA was labeled directly in agarose with the Random Primed DNA Labeling Kit (Boehringer Mannheim) and hybridized to MTN blots (Clontech), according to the manufacturer's instructions. Each blot contains 2 μg of poly A+ RNA from various human tissues.

Interspecific mouse backcross mapping. Interspecific backcross progeny were generated by mating (C57BL/6J × *M. spretus*) F_1 females and C57BL/6J males as described (Copeland and Jenkins 1991). A total of 205 N_2 mice were used to map the *D2H0S1474E* (Abc2) locus (see text for details). DNA isolation, restriction enzyme digestion, agarose gel electrophoresis, Southern blot transfer, and hybridization were performed essentially as described (Franco del Amo et al. 1993). All blots were prepared with Hybond-N+ nylon membrane (Amersham). The probe, a 1.2 kb *Eco*RI fragment of human cDNA, was labeled with [α^{32}P] dCTP and a random primed labeling kit (Stratagene); washing was done to a final stringency of 1.0 × SSCP, 0.1% SDS, 65°C. Fragments of 3.7 and 1.1 kb were detected in *Sac*I-digested C57BL/6J DNA, and fragments of 4.9 and 1.1 kb were detected in *Sac*I-digested *M. spretus* DNA. The presence or absence of the 4.9-kb *M. spretus*-specific *Sac*I fragment was followed in backcross mice.

A description of the probes and RFLPs for the loci linked to *D2H0S1474E* including vimentin (*Vim*), notch 1 (*Notchl*), and Abelson leukemia oncogene (*Abl*) has been reported previously (Green 1981). Recombination distances were calculated as described (Jenkins et al. 1982)

Correspondence to: M. Dean

with the computer program SPRETUS MADNESS. Gene order was determined by minimizing the number of recombinant events required to explain the allele distribution patterns.

Results

Identification of ABC sequences. Potential ABC genes were identified by searching the dbEST database (Boguski et al. 1993) with one of the ATP-binding domains of the multidrug resistance (MDR) gene PGY1. The BLAST program (Altshul et al. 1990) was used to search all six reading frames of each EST. Human clones that displayed matches with P values <1.0 were retrieved and examined for homology to other ABC genes. Table 1 displays the resulting clones and their alignment score to the MDR sequence as well as the highest score to a known ABC gene. Four unique sequences were identified that are not identical to any known mammalian ABC gene.

Mapping of human ESTs. The EST 20237 cDNA clone served as a probe to isolate a corresponding genomic locus. The sequence of a portion of the genomic clone was used to design PCR primers for mapping of this gene on a somatic cell hybrid panel. The correctly sized product was amplified from a hybrid containing human Chr 1 only. No other hybrid amplified a product except the Chr 17 hybrid, which displayed a larger (800 bp) product (data not shown). Hybrid mapping data were confirmed by FISH, placing this putative ABC gene on 1q42. Northern blot analysis of the cDNA detected a 4.5-kb transcript expressed in all examined tissues (Fig. 1A).

The clone corresponding to EST 31252 hybridized to a 6.0-kb mRNA expressed in lung and placenta (Fig. 1B). Additional sequencing of this clone revealed that it represents an incompletely spliced transcript. The 5′ end of the clone contains a portion of an Alu repeat with a poly-A tract. The cDNA appears to have been primed from the poly-A sequence, and the domain with homology to ABC genes is bounded by consensus splice signals. In addition, this clone contains an exon of nicein, a previously characterized gene (Vailly et al. 1994). The sequences homologous to the ABC genes appear to be located within an intron of nicein (data not shown). Further analysis of these sequences is needed to determine whether these are overlapping genes or whether the ABC portion represents a pseudogene. Zoo blot analysis, with only the 400-bp ABC-homologous part of the EST clone as a probe, showed conservation of this sequence from human to yeast DNA (data not shown). Analysis of somatic cell hybrids assigned this clone to Chr 1; this was confirmed by FISH, mapping this gene to 1q25 (Table 1).

The cDNAs corresponding to the two brain EST clones (600, 1596) were obtained from the ATCC. Clone 1596 was used for in situ hybridization and displayed signals to both 6p21.1 and 9q34. PCR primers from the ABC domain (5′ end) amplified only a hybrid cell line containing Chr 9, whereas a pair of primers from the 3′ end of the cDNA amplified sequence from Chr 6. Most likely this clone is chimeric and the ABC gene resides on Chr 9. Cloning a cosmid contig from the genomic locus containing this gene on human Chr 9 confirmed this (data not shown).

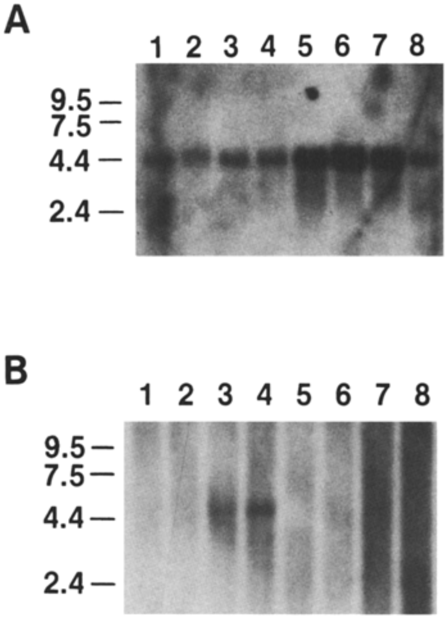

Fig. 1. Expression of EST clones in human tissues. Inserts from the EST clones were isolated and hybridized to blots containing RNA from heart (1), brain (2), placenta (3), lung (4), liver (5), muscle (6), kidney (7), and pancreas (8). **(A)** EST20237; **(B)** EST31252.

The mouse chromosomal location of EST 600 was determined by interspecific backcross analysis. This mapping panel has been typed for over 1500 loci that are well distributed (Copeland and Jenkins 1991). Parent DNAs were digested with several enzymes and analyzed by Southern blot hybridization for informative restriction fragment length polymorphisms (RFLPs) by use of a human cDNA probe. The 4.9-kb *M. spretus SacI* RFLP (see Methods) was used to follow the segregation of the locus in backcross mice. The mapping results indicated that the homolog of EST 600 is located in the proximal region of mouse Chr 2 linked to *Vim, Notchl,* and *Abl.* The gene has been designated *D2H0S1474E.* Although 74 mice were analyzed for every marker (Fig. 2), up to 156 mice were typed for some pairs of markers. Each locus was analyzed in pairwise combinations for recombination frequencies from the additional data. The ratios of the total number of mice exhibiting recombinant chromosomes to the total number of mice analyzed for each pair of loci and the most likely gene order are: centromere-*Vim*–14/156–*Notchl*–0/79–*D2H0S1474E*–6/84–*Abl.* The recombinational frequencies [in centiMorgans (cM)] are: *Vim*–9.0 ± 2.3–[*Notchl, D2H0S1474E*]–7.1 ± 2.8–*Abl.* No recombinants were detected between *Notchl* and *D2H0S1474E* in 79 animals typed in common, suggesting that the two loci are within 3.7 cM of each other (upper 95% confidence limit).

We have compared our interspecific map of Chr 2 with a composite mouse linkage map that reports the location of many uncloned mouse mutations (compiled by M.T. Davisson, T.H. Roderick, A.L. Hillyard, and D.P. Doolittle and provided from GBASE, a computerized database maintained at The Jackson Laboratory, Bar Harbor, Maine). *D2H0S1474E* mapped in a region of

Table 1. ABC superfamily genes expressed sequence tags.

NCBI ID	Source ID	Tissue	Species	Location	Size	Homology	Score	MDR
600	EST00632	Brain	Human	9q34	525	NODI RHILV	111	61
1596	EST01643	Brain	Human	9q34	393	NODI RHILV	173	78
20237	HSB06B092	Muscle	Human	1q42	336	MDR1 MOUSE	101	101
31252	hbc1212	Pancreas	Human	1q25	198	INFA ECOLI	56	54

Clone is part of an EST contig; tissue, tissue source of cDNA library; species, confirmed species of origin of the gene; location, chromosomal position based on somatic cell hybrids and/or FISH; homology, ABC gene displaying the highest alignment score; MDR, alignment score to the human MDR N-terminal ATP-binding domain.

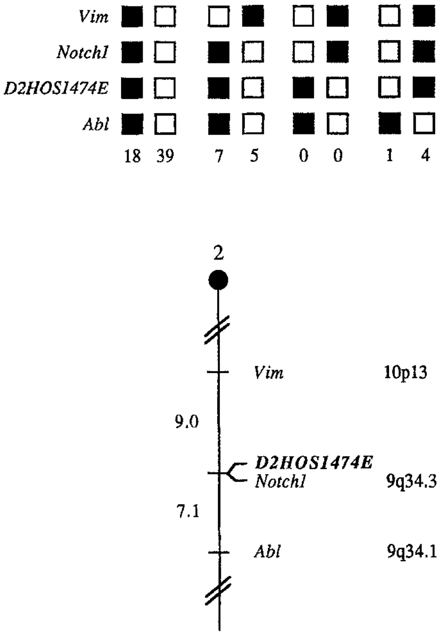

Fig. 2. *D2HOS1474E* (Abc2) maps in the proximal region of mouse Chr 2. The locus was placed on mouse Chr 2 by interspecific backcross analysis. The segregation patterns of *D2HOS1474E* and flanking genes in 74 backcross animals that were typed for all loci are shown at the top of the figure. For individual pairs of loci, more than 74 animals were typed (see text). Each column represents the chromosome identified in the backcross progeny that was inherited from the F_1 parent. The **shaded boxes** represent the presence of a C57BL/6J allele, and the **white boxes** represent *M. spretus* alleles. The number of offspring inheriting each type of chromosome is listed at the bottom of each column. A partial Chr 2 linkage map showing the location of *D2HOS1474E* in relation to linked genes is shown at the bottom of the figure. Recombinational distances between loci in centimorgans are shown to the left of the chromosome, and the positions of loci in human chromosomes, where known, are shown to the right. References for the human map positions of loci cited in this study can be obtained from GDB (Genome Data Base).

the composite map that lacks mouse mutations with a phenotype that might be expected for an alteration in this locus (data not shown). However, in the absence of mouse developmental expression studies, it is difficult to predict what kind of mutation to expect. The proximal region of mouse Chr 2 shares a region of homology with human Chrs 9 and 10 (Fig. 2). The placement of *D2HOS1474E* on human Chr 9q34 confirms and extends this association between mouse Chr 2 and human Chr 9q.

While this work was in progress, the mapping of two new ABC genes was reported by Luciani and coworkers (1994). It was demonstrated that ESTs 600 and 1596 were derived from the same gene, named ABC2. Our data are in agreement with these observations, placing, in addition, the ABC2 gene on mouse linkage map. Both brain-derived ESTs (600 and 1596) were chimeric, the first containing additional sequences from Chr 10, and the second from Chr 6. Frequent chimerism of cDNA clones significantly complicates correct mapping of human EST clones and, respectively, genes in the human genome.

Sequence diversity. In order to analyze the extent of evolutionary relatedness between the EST sequences and known ABC genes, a phylogenetic comparison was performed. The amino acid sequences were aligned with PILEUP and trimmed to a common overlapping region. The results (Fig. 3) show that the EST sequences contain most of the conserved residues as the known members of the gene family, but are considerably diverged from the known genes. The complete sequence of the ABC domains of the genes will be needed to perform a more detailed phylogenetic analysis.

Discussion

We have identified several ESTs that represent previously uncharacterized members of the ABC gene superfamily in the human genome. Many ABC genes contain multiple ATP-binding subunits; therefore, some ESTs may derive from the same gene, as

```
EST1596    .......... .......... .......... .......... .......... .......... .......... .......... .......VVK WALEK....L
EST600     .......... .......... .......... .......... .......... .......... .......... ......FYSR LKSMAQEEIR REMDKIIEDL
PMP        IRDL.NFEVR SGANVLICGP NGCGKSSLFR VLGELWPLFG GRLTKPE... ........RR KLFYVPQRPY MTLGTLRDQV IYPDGREDQK RKGISDLVQK EYLDNVQLGH
ALD        VASL.NIRVE EGMHLLITGP NGCGKSSLFR ILGGLWPTYG GVLYKPP... ........PQ RMFYIPQRPY MSVGSLRDQV IYPDSVEDMQ RKGYSEQDLE AILDVVHLHH
CFN        LKDI.NFKIE RGQLLAVAGS TGAGKTSLLM MIMGELEPSE GKIKHSGR.. .......... .ISFCSQFSW IMPGTIKENI IFGVSYDEYR YRSVIKA... CQLEEDISKF
MRPN       LNGI.TFSIP EGALVAVVGQ VGCGKLSLLS ALLAEMDKVE GHVAIKGS.. .......... .VAYVPQQAW IQNDSLRENI LFGCQLEEPY YRSVIQA... CALLPDLEIL
CFC        LENI.SFSIS PGQRVGLLGR TGSGKSTLLS AFLRLLN.TE GEIQIDGVSW DSITLQQWRK AFGVIPQKVF IFSGTFRKNL D...PYEQWS DQEIWKVADE VGLRSVIEQF
MRPC       LRHI.NVTIN GGEKVGIVGR TGAGKSSLTL GLFRINESAE GEIIIDGINI AKAGLHDLRF KITIIPQDPV LFSGSLRMNL D...PFSQYS DEEVWTSLEL AHLKDFVSAL
MDR1C      LQGL.SLEVK KGQTLALVGS SGCGKSTVVQ LLERFYDPLA GKVLLDGKEI KRLNVQWLRA HLGIVSQEPI LFDCSIAENI AYGDNSRVVS QEEIVRAAKE ANIHAFIESL
MDR1N      LKGL.NLKVQ SGQTVALVGN SGCGKSTTVQ LMQRLYDPTE GMVSVDGQDI RTINVRFLRE IIGVVSQEPV LFATTIAENI RYGREN..VT MDEIEKAVKE ANAYDFIMKL
TAP1       LQGL.TFTLR PGEVTALVGP NGSGKSTVAA LLQNLYQPTG GQLLLDGKPL PQYEHRYLHR QVAAVGQEPQ VFGRSLQENI AYGLTQK.PT MEEITAAAVK SGAHSFISGL
TAP2       LKGL.TFTLR PGEVTALVGP NGSGKSTVAA LLQNLYQPTG GQVLLDEKPI SQYEHCYLHS QVVSVGQEPV LFSGSVRNNI AYGL.QS.CE DDKVMAAAQA AHADDFIQEM
EST31252   .......... .......... .......... .......... .......... .......... .......... .......... .......... ........NR ATFVLVVWHL
EST20237   LRIF.SLSIP SGSVTALVGP SGSGKSTVLS LLLRLYDPAS GTISLDGHDP SAKPSVWLRS KIGTVSQEPI LFSCSIAENI AYGADDPSSV TAEEIQRVAE VANAVAS...
CONSENSUS  L-----F--- -G---A-VG- -G-GKS---- -L--L----- G-----G--- ---------- ------Q-P- --------N- ---------- ---------- --L---I--L

EST1596    ELTKYAD.KP AG....TYSG GNKRKLSTAI ALIGYPAFIF LDEPTTGMDP KARRFLWNLI LD..LIKTGR SVVLTSHSME ECEALCTRLA IMGERSPAVP GQHPAPEEPV
EST600     ELSNKRH.SL VQ....TLSG GMKRKCTVAI AFVGGSRAII LDEPTRAWTP .TRAAIWDLI LK..Y.KPGR TILLSTHHMD EADLLGEPHC IISMGSQVLR .LSALPQGHL
PMP        ILEREGGWDS VQDWMDVLSG GEKQRMAMAR LFYHKPQFAI LDECTSAVSV DVEGYIYSHC .....RKVGI TLFTVSHRKS LWKHHEYYLH MDGRGNYEFK QITEDTVEFG
ALD        ILQREGGWEA MCDWKDVLSG GEKQRIGMAR MFYHRPKYAL LDECTSAVSI DVEGKIFQAA KDAGIALLSI THRPSLWKYH THLLQFDGEG GWKFEKLDSA ARLSLTEEKQ
CFN        ...AEKDNIV LGEGGITLSG GQRARISLAR AVYKDADLYL LDSPFGYLDV LTEKEIFESC VCK..LMANK TRILVTSKME HLKKADKILI LNEGSSYFYG TFSELQNLQP
MRPN       ...PSGDRTE IGEKGVNLSG GQKQRVSLAR AVYSNADIYL FDDPLSAVDA HVGKHIFENV IGPKGMLKNK TRILVTHSMS YLPQVDVIIV MSGGKISEMG SYQELLARDG
CFC        ...PGKLDFV LVDGGCVLSH GHKQLMCLAR SVLSKAKILL LDEPSAHLDP VTYQII.RRT LKQ..AFADC TVILCEHRIE AMLECQQFLV IEENKVRQYD SIQKLLNERS
MRPC       ...PDKLDHE CAEGGENLSV GQRQLVCLAR ALLRKTKILV LDEATAAVDL ETDDLI.QST IRT..QFEDC TVLTIAHRLN TIMDYTRVIV LDKGEIQEYG APSDLLQQRG
MDR1C      ...PNKYSTK VGDKGTQLSG GQKQRIAIAR ALVRQPHILL LDEATSALDT ESEKVV.QEA LDK..AREGR TCIVIAHRLS TIQNADLIVV FQNGRVKEHG THQQLLAQKG
MDR1N      ...PHKFDTL VGERGAQLSG GQKQRIAIAR ALVRNPKILL LDEATSALDT ESEAVV.QVA LDK..ARKGR TTIVIAHRLS TVRNADVIAG FDDGVIVEKG NHDELMKEKG
TAP1       ...PQGYDTE VDEAGSQLSG GQRQAVALAR ALIRKPCVLI LDDATSALDA NSQLQV.EQL LYESPERYSR SVLLITQHLS LVEQADHILF LEGGAIREGG THQQLMEKKG
TAP2       ...EHGIYTD VGEKGSQLAA GQKQRLAIAR ALVRDPRVLI LDEATSALD. ...VQC.EQA LQDWNSRGDR TVLVIAHRLQ TVQRAHQILV LQEGKLQKLA QL*.......
EST31252   SSRPV..... ...LGRKLSG GQMSRMGVGR AMQCCPVLEL LFNGDKDWEV GTRK...... .......... .......... .......... .......... ..........
EST20237   PEFPPRFNTV VGEKGVLLSG GQKQRIAIAR ALLKNPKILL LDEATSALDA ENEYLV.QEA LDR..LMDGR TVLVIAHSLS TIKNANMVAV LDQGKITEYG KHEELLSKPN
CONSENSUS  ---P------ -G--G--LSG GQKQR---AR AL---P--LL LDE-TSALD- -----I---- L--------- T-----H--- ---------- ---G------ ----L-----
```

Fig. 3. Alignment of EST sequences from the ABC superfamily. Deduced amino acid sequences from several EST clones are shown aligned to the known human ABC genes. PMP, peroxisomal membrane protein; ALD, adrenoleukodystrophy gene; CFN and CFC, cystic fibrosis transmembrane regulator, N- and C-terminal domains; MRPN and MRPC, multidrug resistance like protein, N- and C-terminal domains; MDR1N and MDR1C, P-glycoprotein, N- and C-terminal domains; TAP1 and TAP2, antigen peptide transporter genes. A consensus (Consen) sequence is shown at the bottom of the alignment.

shown in the case of two clones, originated from a brain cDNA library. Otherwise each examined sequence maps to a different chromosome and detects an RNA transcript with a unique size and tissue distribution. Despite the large number of sequences examined, none of them corresponds to any of the known ABC genes. Thus, the total number of human ABC genes may be as high as 20–40. Given the large number and diverse function of prokaryotic ABC genes, this is not surprising.

We did, however, encounter several difficulties in using EST clones. Because the sequences are short and multiple clones are present from the same gene, there is considerable redundancy in the database. The transmembrane portions of ABC genes are not conserved among gene family members. Therefore, many additional members of the gene family have probably gone undetected. The BLAST algorithm (Altschul et al. 1990) is very sensitive to identify homologous genes. Several ESTs displayed matches that were below statistical significance to MDR. When these ESTs were compared with other ABC genes, or when the open reading frame was changed to remove putative frame shifts, many of these marginal matches appeared to represent ABC family members. We also encountered numerous clones that are incompletely spliced or chimeric. These clones can confound sequence analysis, expression, and mapping studies. In addition, we found that many of the ESTs from the Clontech CCRF-CEM library (Cat# HL1063g) represent sequences of bacterial origin (R. Allikmets, M. Dean, F. Lorenzo, C. Auffray, manuscript in preparation).

From our data it is clear that the analysis of ESTs is a powerful method to identify new genes. Each of these clones can now be used to generate additional sequence data and perform studies of gene expression. Preliminary data with the *EST20237* suggests that this gene is highly overexpressed in certain myeloid tumor cell lines (data not shown). Further analysis of the expression of these genes in normal, mutant, and tumor cell lines may reveal clues as to their function.

Acknowledgments. We thank J. Takeda and C. Auffray for providing cDNA clones; Stan Cevario for oligonucleotide synthesis; M. Barnstead and D. Barnhardt for technical assistance; Mark Gunnell and Gary Smythers of the Advanced Scientific Computer Facility (Frederick, Maryland) for help with sequence analysis.

This project was supported, in part, by the National Cancer Institute, DHHS, under contract N0-CO-74101 with ABL. The content of this publication does not necessarily reflect the views or policies of the Department of Health and Human Services, nor does mention of trade names, commercial products, or organizations imply endorsement by the U.S. Government. By acceptance of this article, the publisher or recipient acknowledges the right of the U.S. Government to retain a nonexclusive, royalty-free license in and to any copyright covering the article.

References

Adams, M.D., Kerlavage, A.R., Fields, C., Venter, J.C. (1993). 3,400 new expressed sequence tags identify diversity of transcripts in human brain. Nature Genet. 4, 256–267.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., Lipman, D.J. (1990). Basic local alignment search tool. J. Mol. Biol. 215, 403–410.

Ames, G.F.-L., Lecar, H. (1992). ATP-dependent bacterial transporters and cystic fibrosis: analogy between channels and transporters. FASEB J. 6, 2660–2666.

Boguski, M.S., Lowe, T.M.J., Tolstoshev, C.M. (1993). The EST express gathers speed. Nature Genet. 4, 331–332.

Copeland, N.G., Jenkins, N.A. (1991). Development and applications of a molecular genetic linkage map of the mouse genome. Trends Genet. 7, 113–118.

Devereaux, J., Haeberli, P., Smithies, O. (1984). A comprehensive set of sequence analysis programs for the VAX. Nucleic Acids Res. 12, 387–395.

Feng, D-F., Doolittle, R.F. (1987). A multiple sequence alignment using a simplification of the progressive alignment method. J. Mol. Evol. 35, 351–360.

Franco del Amo, F., Gendron-Maguire, M., Swiatek, P.J., Jenkins, N.A., Copeland, N.G., Gridley, T. (1993). Cloning, analysis, and chromosomal localization of *Notch-1,* a mouse homolog of *Drosophila Notch.* Genomics 15, 259–264.

Grausz, J.D., Auffray, C. (1993). Strategies in cDNA programs. Genomics 17, 530–532.

Green, E.L. (1981). Linkage, recombination and mapping. In *Genetics and Probability in Animal Breeding Experiments,* , ed. (New York: Oxford University Press), pp. 77–113.

Gros, P., Croop, J., Housman, D. (1986). Mammalian multidrug resistance gene: complete cDNA sequence indicates strong homology to bacterial transport proteins. Cell 47, 317–380.

Higgins, C.F. (1992). ABC transporters: from micro-organisms to man. Annu. Rev. Cell Biol. 8, 67–113.

Jenkins, N.A., Copeland, N.G., Taylor, B.A., Lee, B.K. (1982). Organization, distribution, and stability of endogenous ecotropic murine leukemia virus DNA sequences in chromosomes of *Mus musculus.* J. Virol. 43, 26–36.

Lincoln, A.L., Daly, M., Lander, E. (1991). PRIMER: a computer program for automatically selecting PCR primers. Whitehead Institute Technical Report, MIT, Cambridge, MA.

Luciani, M.F., Denizot, F., Savary, S., Mattei, M.G., Chimini, G. (1994). Cloning of two novel ABC transporters mapping on human chromosome 9. Genomics 21, 150–159.

Mosser, J. (1993). Putative X-linked adrenoleukodystrophy gene shares unexpected homology with ABC transporters. Nature 361, 726–730.

Riordan, J.R., Rommens, J.M., Kerem, B.-S., Alon, N., Rozmahel, R., Grzelczak, Z., Zielenski, J., Lok, S., Plavsic, N., Chou, J.-L., Drumm, M.L., Iannuzzi, M.C., Collins, F.S., Tsui, L.-C. (1989). Identification of the cystic fibrosis gene: cloning and characterization of the complementary DNA. Science 245, 1066–1073.

Spies, T., Bresnahan, M., Bahram, S., Arnold, D., Blanck, G., Mellins, E., Pious, D., DeMars, R. (1990). A gene in the human major histocompatibility complex class II region controlling the class I antigen presentation pathway. Nature 348, 744–747.

Takeda, J., Yano, H., Eng, S., Zeng, Y., Bell, G.I. (1993). A molecular inventory of human pancreatic islets: sequence analysis of 1000 cDNA clones. Hum. Mol. Genet. 2, 1793–1798.

Vailly, J., Verrando, P., Champliaud, M.-F., Gerecke, D., Wagman, D.W., Baudoin, C., Aberdam, D., Burgeson, R., Bauer, E., Ortonne, J.-P. (1994). The 100-kDa chain of nicein/kalinin is a laminin B2 chain variant. Eur. J. Biochem. 219, 209–218.