Human Molecular Genetics, 1996, Vol. 5, No. 10 ***1649–1655***

Characterization of the human ABC superfamily: isolation and mapping of 21 new genes using the Expressed Sequence Tags database

Rando Allikmets, Bernard Gerrard[1], Amy Hutchinson and Michael Dean*

Laboratory of Viral Carcinogenesis and [1]Intramural Research Support Program, SAIC Frederick, National Cancer Institute, Frederick Cancer Research and Development Center, Frederick, MD 21702, USA

Received May 31, 1996; Revised and Accepted July 22, 1996

GenBank accession nos U66672 and U66692

As an approach to characterizing all human ATP-binding cassette (ABC) superfamily genes, a search of the human expressed sequence tag (EST) database was performed using sequences from known ABC genes. A total of 105 clones, containing sequences of potential ABC genes, were identified, representing 21 distinct genes. This brings the total number of characterized human ABC genes from 12 to 33. The new ABC genes were mapped by PCR on somatic cell and radiation hybrid panels and yeast artificial chromosomes (YACs). The genes are located on human chromosomes 1, 2, 3, 4, 6, 7, 10, 12, 13, 14, 16, 17 and X; at locations distinct from previously mapped members of the superfamily. The characterized genes display extensive diversity in sequence and expression pattern and this information was utilized to determine potential structural, functional and evolutionary relationships to previously characterized members of the ABC superfamily.

INTRODUCTION

The multidrug resistance/ATP-binding cassette (MDR/ABC) superfamily in humans includes genes whose products represent membrane proteins involved in energy-dependent transport of a wide variety of substrates across membranes (for review, see 1,2). The ABC family (also called traffic ATPases) is one of the largest superfamilies of proteins known in both prokaryotic and eukaryotic organisms and can be clearly distinguished from other ATP-binding protein families such as kinases (3). In eukaryotes, ABC genes typically encode four domains consisting of two ATP-binding segments, and two transmembrane (TM) segments; half-molecules, containing one ATP and one TM domain also exist (4). The ATP-binding domains of ABC genes contain motifs of characteristic conserved residues (Walker A and B motifs) spaced by 90–120 amino acids. This conserved spacing and the 'Signature' or 'C' motif just upstream of the Walker B site distinguishes members of the ABC superfamily from other ATP-binding proteins (3,4). These features have allowed new ABC genes to be isolated by hybridization, degenerate PCR and inspection of DNA sequence databases (5–10).

Previously characterized members of the superfamily include those transporting chemotherapeutic drugs (PGP1/3, MRP; 11–14); peptide transporters involved in antigen presentation (TAP1/2; 15) and those involved in a number of inherited human diseases (CFTR, ALD, SUR, PMP70; 16–19). To date only 12 ABC genes have been identified in humans, although over 25 members of this highly conserved superfamily have been described in *Escherichia coli* alone (20,21). 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 characterizing all human genes, several laboratories have been sequencing portions of cDNA clones from different tissues (22,23). During the last year the quantity of random human cDNA sequences has been grown exponentially (24). Such expressed sequence tags (ESTs) are useful for identifying new genes, as tools for genetic mapping, and for the analysis of gene expression (25). We have shown previously the usefulness of using ESTs to characterize a large gene family (7). Here we present an example of an extensive study of the ABC superfamily aimed to characterize a maximum number of genes structurally, and ultimately, functionally.

RESULTS

Identification of ABC sequences and structural characterization of the new ABC genes

Potential ABC genes were initially identified by searching the dbEST database (26) using the N-terminal ATP-binding domain of the human multidrug resistance (MDR) gene PGY1, and the complete amino acid sequence of cystic fibrosis transmembrane conductance regulator (CFTR). The BLAST program (27) was used to search all six reading frames of each EST. Human clones that displayed matches with the highest scores were retrieved and examined for homology to other ABC genes. Analyses using the XREFdb program (30) with representatives from all of the major ABC gene subfamilies identified additional sequences. Table 1 displays the resulting clones and their highest alignment score to a known ABC gene. A total of 105 unique sequences were identified that are not identical to any known mammalian ABC gene. A total of 36 sequences from previously characterized ABC

*To whom correspondence should be addressed at: Laboratory of Viral Carcinogenesis, NCI-FCRDC, Building 560, Room 21-18, Frederick, MD 21702-1201, USA

genes were also identified. All known ABC genes were represented in the dbEST by at least one sequence (clone). This fact and the apparent saturation of the database, i.e. that fewer new ABC were detected in recent searches, leads us to the conclusion, that we have identified most of the human ABC genes. However, some new members of the superfamily may have gone undetected because of their restricted expression in specific tissues or cell types.

Mapping of human ESTs

To map the genes in the human genome, additional sequencing was performed on all available cDNA clones. This data provided additional information on the coding regions of the genes, and the 3′ untranslated regions were used to generate sequence tagged sites (STSs). A monochromosome somatic cell hybrid panel (Coriell Institute) was used to localize the new genes on human chromosomes (Table 1). Additional mapping was carried out on CEPH YAC panels A and B or the GeneBridge 4 Radiation Hybrid Panel (Research Genetics) to define the location more precisely. We were able to map 5 out of 21 genes (25%) on YACs (Table 1), primers to the rest of the clones failed to amplify YAC DNA, most likely due to limited representativity of the YAC pools. Many of these genes were successfully mapped on the radiation hybrid panel.

All mapping data are summarized in Table 1. Gene EST123147 was mapped to 6p21 and to YACs 924_b_3, 225_b_1 and 368_d_7 within the HLA Class I region, between HLA-E and HLA-A. The hemachromatosis gene has been mapped to this region and is responsible for an iron transport/metabolism disorder (33). However, additional mapping of the EST123147 gene placed it just telomeric to HLA-E, outside of the hemachromotosis critical region (data not shown). The full-length cDNA sequence of EST123147 was obtained by 5′ RACE, and showed that the gene contains two ATP-binding domains and has 42% aa identity to the yeast GCN20 gene (34). GCN20 encodes a cytoplasmic protein that plays an important role in yeast translational control under amino acid starvation conditions. This gene (EST123147), together with the EST133090 and EST201864 genes, that map to 7q35-qter and 3q25 respectively, form a new subfamily of ABC proteins in humans that lack transmembrane segments and consist only of one or two ATP-binding domains.

Expression of the new ABC genes

Expression studies of the new genes were carried out using Northern blots of different human tissue RNA. The genes display considerable variation in transcript sizes and tissue specificity (Table 1, Fig. 1). Most eukaryotic ABC genes encode four domains, full molecule (FM), consisting of two transmembrane (TM) and two ATP-binding segments, although half-molecules (HM), consisting of TM-ATP or ATP-ATP also exist (4). The majority of the newly characterized genes fell into the first category, containing four domains and transcript sizes ranging from 6 to 10 kb (Table 1). Six genes expressed mRNAs in the range of 3.5–4.5 kb and appear to represent TM-ATP half-molecules, whereas three genes, namely EST123147, EST133090 and EST201864, contained only ATP-binding domains in their sequence and are encoded by a 1.3–3.0 kb message. Analysis of the full-length clones will be required to confirm some of these assignments.

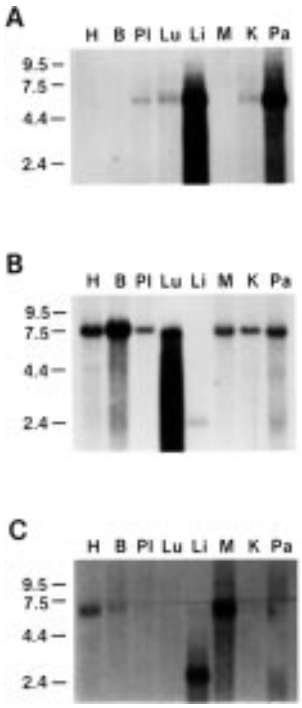

Figure 1. Expression of three EST clones in human tissues. Inserts from the EST clones were isolated and hybridized to blots containing RNA from heart (H), brain (B), placenta (Pl), lung (Lu), liver (Li), muscle (M), kidney (K) and pancreas (Pa). (**A**) EST90757, (**B**) EST111653 and (**C**) EST90625.

Most of the new genes were expressed in all tissues examined (Table 1, Fig. 1). However, in a few cases (Fig. 1B and C) different size messages were detected in different tissues, the nature of these alternate transcripts remains to be investigated. Some genes, i.e. EST90757 and EST172291, showed detectable expression only in selected tissues, mainly liver. In a recent paper (35) a rat liver-specific gene similar to MRP was described. A mutation in this gene, cMOAT (canalicular multispecific organic anion transporter), was found in the TR$^-$ rat. This strain displays congenital jaundice (36) and is a model for the human Dubin-Johnson syndrome (OMIM #237500), a hyperbilirubinemia. Sequence comparison and expression data revealed that the EST172291 gene is most likely the human homologue of the rat cMOAT, the sequence identity being 89% on aa level (Fig. 2). Another liver-specific gene, EST90757, was also shown to be closely related to MRP (Table 1, Fig. 2 and 3). Expression studies of these genes in human tumor cell lines are in progress to investigate the potential role of these MRP homologues in drug resistance and anion transport.

Sequence diversity and phylogenetic analysis

In order to analyze the extent of evolutionary relatedness between the EST sequences and known ABC genes, a phylogenetic analysis was performed. The amino acid sequences were aligned with PILEUP and trimmed to a common overlapping region of 220 residues. The results in Figure 2 show that the EST sequences contain conserved amino acids that are found in all known members of the gene family, but are distinct from all previously characterized genes.

Table 1. New human ABC genes

EST number	Number of clones	Library	RNA expression	Location YAC/RH	Best match/ accession no.	Score	Subfamily
45597	3	Te,IBr,Fli/s	3.5 kb housek.	2	S.po. Hmt1 Q02592	425	MDR/TAP (HM)
90625	2	Fs	7 kb H, Mu 2.5 kb Li only	17q21-q24 845_b_11	C.el. ABC S40724	161	ABC (FM)
90757	6	Li, Fli	6.5 kb Li, Pa only	17q11-q12 WI-4251	HuMRP P33527	806	CF/MRP (FM)
111653	4	Flu,IBr	7.5 kb housek. 2.5 kb Li only	16p13.3 WI-7742	MuABC2 P41234	367	ABC (FM)
122234	3	IBr	3.5 kb H, B	12q24-qter	E.co. HlyB P26760	339	MDR/TAP (HM)
123147	13	IBr,Fs,S,H PBL	3.5 kb housek., Mu highest	6p21.3 225_b_1	S.ce. GCN20 P43535	250	GCN20 (ATP)
131042	2	Lu,IBr		4q31 D4S3137			OligoAbP (ATP)
133090	3	Fli/s,Pla	1.3 kb Lu, Li, Mu	7q35-q36	C.el. GCN20 S56147	482	GCN20 (ATP)
140535	3	Fli/s,Pais	3 kb housek.,	Xq21-q22	MuABC7 U43892	1552	MDR/TAP (HM)
155051	1	Fli/s	~6 kb housek. 10, 4.2 kb H, Mu		MuABC1 P41233	194	CF/MRP (FM)
157481	8	Fli/s,IBr,Pla	3–3.5 kb Pla only	4q22-q23 WI-6336	S.ce. ADP1 P25371	304	MDR/TAP (HM)
170205	1	Fli/s	6.5 kb Pa, Mu	13q31 WI-9265	HuCFTR P13569	111	CF/MRP (FM)
172291	15	Fli/s,Lu,Li	7.0 kb Li only	10q23-q24 WI-4209	HuMRP P33527	418	CF/MRP (FM)
182763	1	IBr	5.5 kb Lu, Li, Mu	6p21 D6S271	cMOAT L49379	611	CF/MRP (FM)
201864	12	Pla,IBr,Lu Fli/s,H,Agl	2.8 kb housek.	3q25.1-q25.2	C.el. ABC U10414	366	GCN20 (ATP)
205858	2	IBr	6.5–7 kb housek. Pancreas high		HuMRP P33527	317	CR/MRP (FM)
277145	6	lBr,Fli/s,B, Hi	7 kb housek. H, B, Mu, Pa	3q25-q26 WI-6365	HuMRP P33527	337	CF/MRP (FM)
328128	3	Br,Fli/s,OlfE	4.5 kb housek.	7q35-q36	S.ma. MDR L26286	475	MDR/TAP (HM)
349056	3	Fli/s	6.5kb Li, Pa	16p13.1 NIB558	HuMRP P33527	129	CF/MRP
352188	6	Fli/s,IBr,Fibr	3 kb housek.	14q12 953_f_11	HuPMP70 P28288	168	ALD/PMP (HM)
394388	1	Ret	N/D	1	MuABC1 P41233	755	ABC (FM?)
422562	2	Pla,Mel	7.5 kb housek.	7p14 D7S673	HuMDR3 P21439	190	MDR/TAP (FM)

Number of clones, how many clones are in an EST contig; Library, tissue/cell line source of a cDNA library; Te, testis; IBr, infant brain; Fli/s, fetal liver/spleen; Li, liver; Flu, fetal lung; S, spleen; H, heart; Lu, lung; Pla, placenta; Pa, pancreas; Pais, pancreatic islands; Agl, adrenal gland; PBL, peripheral blood cells; Br, breast; Hi, hippocampus; Mel, melanocyte; B, brain; OlfE, olfactory epithelium; Fibr, fibroblast; Ret, retina; Mu, skeletal muscle; Location, chromosomal position based on somatic cell hybrids; YAC/RH, YAC address or closest marker on radiation hybrids panel; Hu, human; Mu, murine; S.ma.,*Schistosoma mansoni;* S.po., *Schisosaccharomyces pombe;* S.ce., *Saccharomyces cerevisiae;* C.el., *C. elegans.;* E.co., *E.coli.* Best match/ accession no., ABC gene displaying the highest alignment score /its accession number; **Score**, aa alignment score to the best match as determined by BLAST; Subfamily, subgroup of the ABC genes determined by evolutionary analysis (Fig. 3); FM, full molecule—protein consists of two ATP and two TM domains; HM, half molecule—protein has one ATP and one TM domain; ATP, protein contains only ATP domain(s). N/D, not detected.

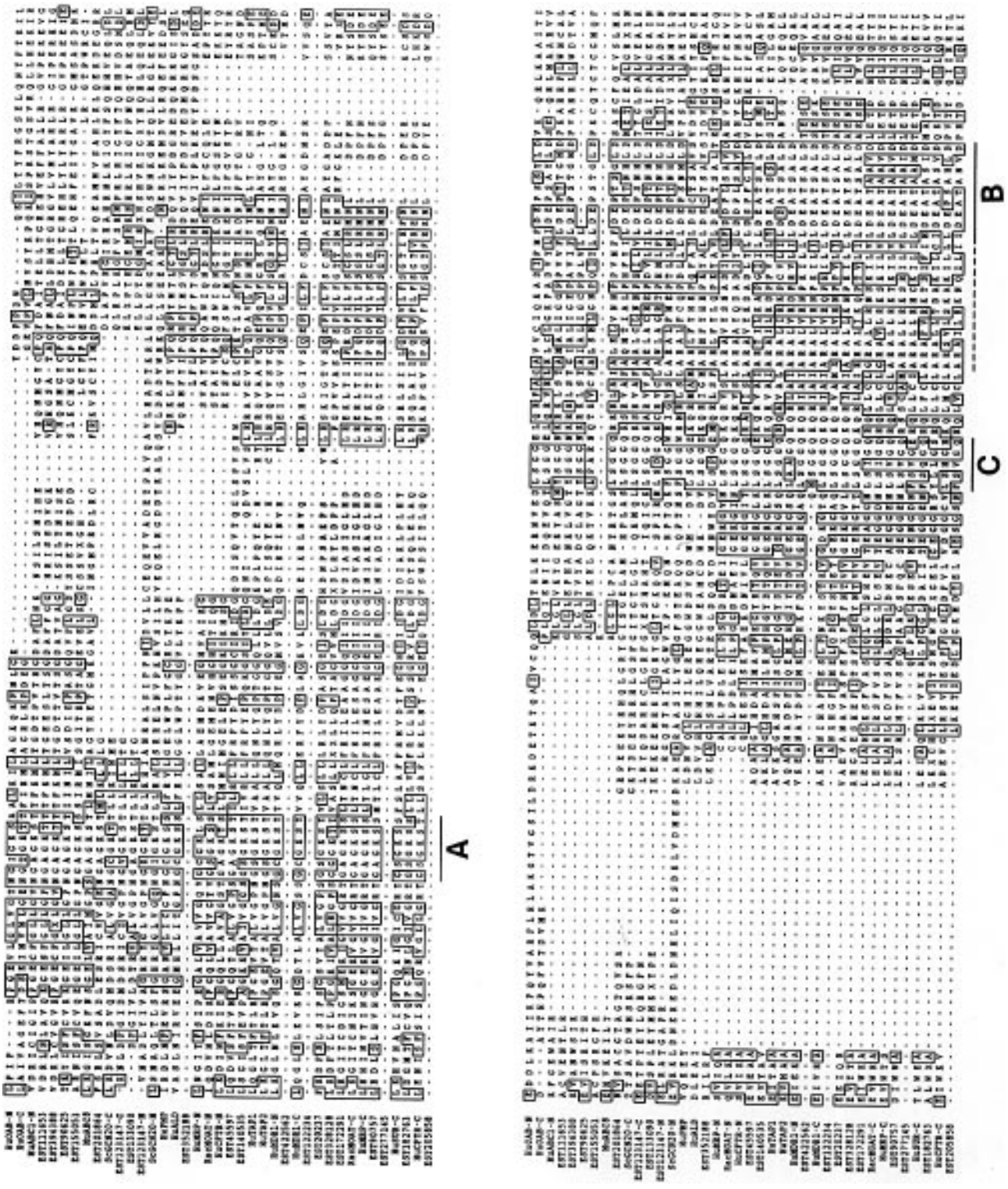

Figure 2. Alignment of EST sequences from the ABC superfamily. Deduced amino acid sequences from 19 EST clones are shown aligned to the known human, rodent and yeast ABC genes. Conserved residues (at least 10 out of 35) are boxed. EST numbers indicate NCBI ID numbers for the new genes; HuOAB, human oligoA binding protein; MuABC1, mouse ABC1 gene; ABC8, mouse ABC8 gene; GCN20, yeast GCN20 gene; HuPMP, human peroxisomal membrane protein; ALD, human adrenoleukodystrophy gene; HuMRP, human multidrug resistance-like protein; cMOAT, rat canalicular membrane protein; HuCFTR, human cystic fibrosis transmembrane conductance regulator; HuSUR, human sulfonylurea receptor; TAP, human antigen presenting protein; HuMDR, human multidrug resistance protein. Walker A and B motifs are designated by solid/dashed lines and letters A and B; C, signature motif for the ABC superfamily; X represents an unknown residue.

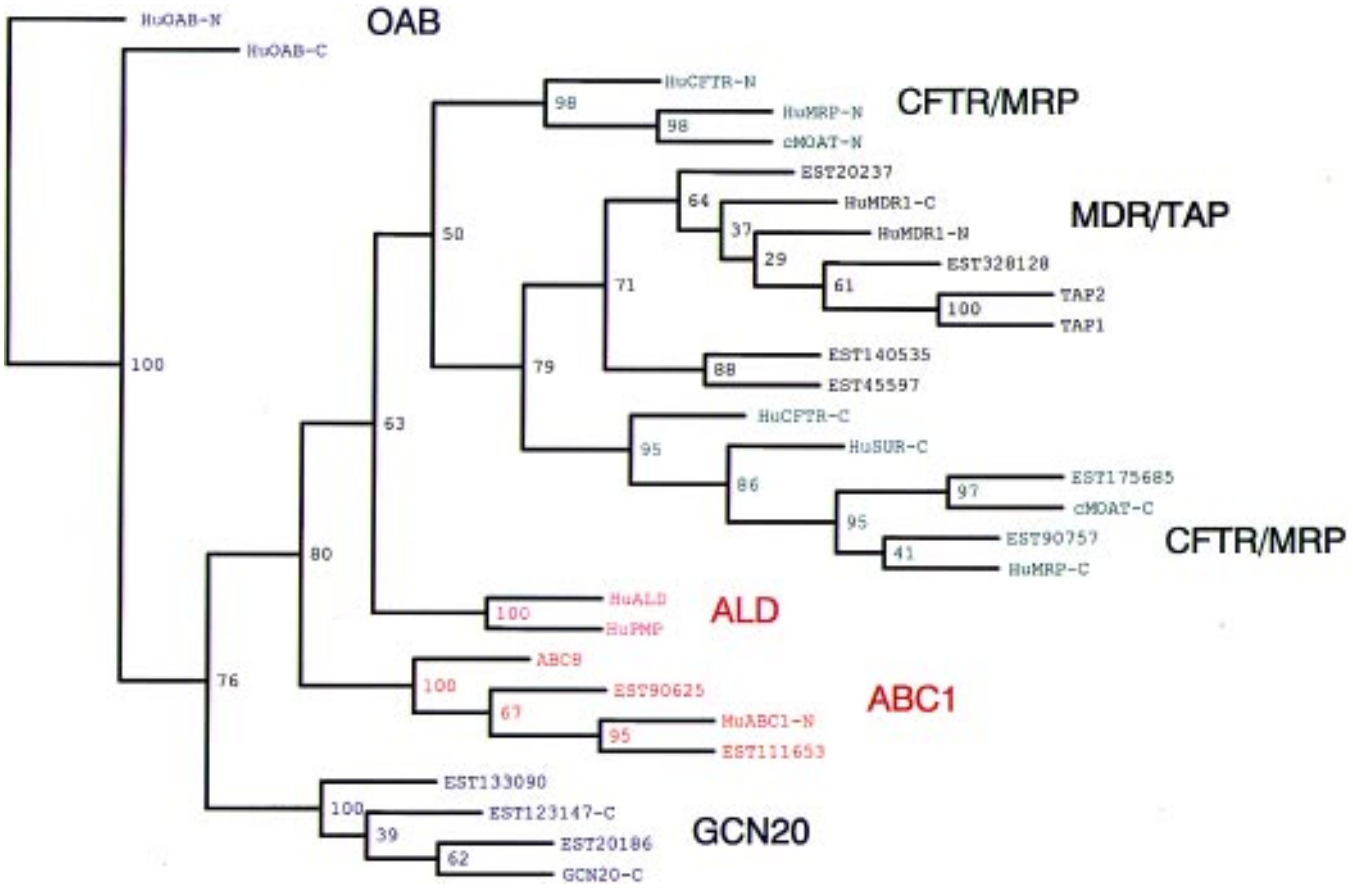

Figure 3. Evolutionary relationship of the new and previously known members of the ABC superfamily. An unrooted maximum parsimony phylogram of members of the superfamily is presented. Designations of the new and known genes are the same as in Figure 2. Numbers on the internal nodes represent bootstrap values out of 100. Subfamilies are indicated in color.

To initially study the phylogenetic relationships of the newly described ABC genes to the rest of the superfamily, two distinct methods of phylogenetic reconstruction were used. Maximum parsimony and distance-based neighbor-joining methods both yielded similar results, supporting our conclusions, shown as the maximum parsimony phylogram on Figure 3. In addition, a bootstrap resampling of 100 iterations in both methods was performed to gauge the reliability of our data to consistently derive the same tree (see Materials and Methods). Groups of genes that were statistically supported by bootstrap values greater than 70 formed six subfamilies (CFTR/MRP, ABC1, GCN20, ALD, OAB, PGP/TAP). Available structure-functional information supported joining genes into these subfamilies.

In the case of the CFTR/MRP subfamily, the N- and C-terminal domains of the different proteins cluster together but are significantly diverged from one another as shown previously (14,37). This is in contrast to the MDR/TAP subgroup of genes, where the two domains are paired together, suggesting a recent duplication event. The GCN20-like genes are clearly more related to one another and to the yeast GCN20 gene than to the other human ABC proteins.

DISCUSSION

We present here an extensive characterization of the human ABC gene superfamily. A total of 21 new genes from this superfamily were mapped and partially sequenced. The phylogenetic relationship of these genes was examined to initially organize the genes into subfamilies. This data provides clues as to the putative function of these new ABC genes.

To date 13 genes of the ABC superfamily have been described (2). The addition of 21 new genes brings the number of the ABC family members in humans to 34, which is close to a number of known bacterial ABC genes. At least one EST hit to each of the known human ABC genes was observed, documenting that the EST database is becoming highly representative. In addition, recent searches of the exponentially growing database have not yielded additional gene family members. Therefore it seems likely that we have identified most of the gene family, and that only a few human members of the superfamily may remain uncharacterized.

Human ABC genes have been found on almost all human chromosomes and show no clustering. Known (TAP1 and 2) or predicted (ALD and PMP, 36; EST45597 and EST140535, this study) pairs of half-molecules that form a functional heterodimer are structurally very closely related. At the same time, mapping data indicates that these genes can be located either next to each other on the same chromosome or on different chromosomes. Many of the genes have been placed onto YAC contigs, or mapped with radiation hybrids. Further mapping in both the human and mouse genomes should aid in correlating these genes with genetic diseases/phenotypes.

Phylogenetic studies indicate that at least six well defined subfamilies of proteins exist within the ABC superfamily. This

analysis agrees well with an analysis of the ABC genes of the yeast *Saccharomyces cerevisiae* (37). Yeast have at least five ABC gene subfamilies, and the human genome contains the same groups. The human genome contains an additional subfamily represented by ABC1 and ABC2. The complete sequence of the yeast genome will allow all of the ABC genes from a single eukaryote to be identified.

One subfamily includes three previously undescribed human genes that are homologs of the yeast GCN20 gene. This subfamily is separated from the others not only by the divergence of the ATP-binding domain sequence, but also by the fact that these proteins do not possess transmembrane domain(s). The putative role of these proteins as translation initiation regulators is under investigation by the means of complementation studies. The only other eukaryotic member of the ABC superfamily that does not contain transmembrane domains is the 2′-5′ oligoadenylate binding protein (OAB, 38), which is a representative of a separate subfamily (Fig. 3). Still, the ATP-binding domains of OAB are most closely related to the GCN20 subfamily, suggesting that these putative non-transporters may have a common origin. These proteins are most likely derived by fusion of two different ATP-binding domains.

The MDR/TAP subgroup includes functionally diverse proteins, that consist of either four (FM) or two (HM) domains. The common feature of the proteins within this subfamily is a high similarity of their N- and C-terminal ATP-binding domain sequences, possibly due to a relatively recent duplication event. Proteins of this subfamily, especially those homologous to PGP (i.e. EST20237), could function as drug transporters. Our preliminary data indicates that EST20237 is overexpressed in certain myeloid tumor cell lines.

Similarly, proteins that are members of the CF/MRP subfamily might function as transporters of cytotoxic drugs. The EST90757 gene mRNA is overexpressed in a number of tumor cell lines, whereas its expression is limited to a few normal tissues. All known genes of this subfamily encode for four domain (FM) proteins, but unlike the previous group, the N- and C-terminal ATP-binding domains of the same protein are extremely diverged. This is an interesting case of structurally and functionally similar proteins that have potentially evolved by different means. It has been proposed (37) that CF/MRP subfamily genes may have evolved from a common full-length (FM) progenitor and that the two halves of the protein perform distinct functions. Initial fusion of two half-molecules (HM), containing one TM and one ATP-binding domains, or ancient gene conversion between CF/MRP and MDR/TAP subfamily progenitors are other possible explanations for how this subfamily may have been initiated. Additional analyses of mutation rates may prove fruitful in the future refining our understanding of the evolutionary origin of all subfamilies.

We have shown that characterization of a gene family by analysis of the EST database is an effective strategy. We were aided in this analysis by the fact that ABC genes have a large domain that is highly conserved, and there are a number of characterized genes from this superfamily to use in database searches. In addition, the ATP-binding domain is almost always in the 3′-end of the gene, and EST sequences are biased towards the 3′-end. In fact the only human ABC subfamily that is underrepresented, in comparison to yeast, is the PDR5 subfamily, and ATP-binding domains are located at the N-terminus of PDR5 sufamily genes. There are at least five PDR5-related genes in yeast and only one identified to date in humans. As the EST database grows to better represent the full spectrum of mammalian cell types, this resource will become increasingly more valuable for identifying biologically important genes.

MATERIALS AND METHODS

Sequence analysis

Searches of the dbEST (26) database were performed using BLAST on the NCBI file service (27). Amino acid alignments were generated with PILEUP (28). Sequences were read with an IBI Gel Reader and analyzed with the Genetics Computer Group package of programs (29) running on a VAX computer. Additional database searches were performed by placing diverse members of the ABC gene family in the X-Ref search program. This program provides automatic monthly search of both GenBank and dbEST subsets (30). Additional sequencing of cDNA clones was performed with the Taq Dyedeoxy Terminator Cycle Sequencing kit (Applied Biosystems) according to the manufacturer's instructions. Sequencing reactions were resolved on an ABI 373A automated sequencer. The sequences have been deposited with GenBank under accession nos U66672 and U66692. The gene numbers used in this manuscript, until the assignment of the new gene symbols, represent the appropriate dbEST ID numbers of one of the clones from each contig (http://ncbi.nlm.nih.gov). For some clones 5′ and 3′ RACE reactions were performed using the MARATHON amplification kit (Clontech) according to the manufacturer's protocols.

Phylogenetic trees were generated from the amino acid alignments using PHYLIP (Phylogeny Inference Package) Version 3.5c (31). Two programs were utilized: NEIGHBOR, implementing the Neighbor-Joining distance matrix method, and PROTPARS, Protein Sequence Parsimony Method. Bootstrap resampling of 100 iterations was performed to test the reliability of the associations in both methods. In this analysis resamplings of the original data set is performed to create 100 new data sets. A consensus of the resulting 100 trees measures the consistency of the phylogenetic signal within the original data. Bootstrap proportions >70% were considered strong support for the adjacent node (39).

Polymerase chain reactions

Primers to EST sequences were designed using the PRIMER program (32). 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; 58°C, 30 s; 72°C, 1 min. PCR products were analyzed on 1.5–2% agarose gels and digested before that with appropriate restriction enzyme to verify their sequence, if needed. Primer sequences and specific reaction conditions have been submitted to GDB.

Northern hybridization

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

ACKNOWLEDGEMENTS

We would like to thank Gary Smythers of the Frederick Biomedical Supercomputing Center for assistance in the sequence analysis, Stan Cevario for primer synthesis, and Jill Slattery and Gavin Huttley for helpful discussions regarding the data. 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 US Government.

REFERENCES

1. Childs,S. and Ling,V. (1994) The MDR superfamily of genes and its biological implications. In DeVita,V.T., Hellman,S. and Rosenberg,S.A. (eds) *Important Advances in Oncology.* Philadelphia, Lippincott Company, pp. 21–36.
2. Dean,M. and Allikmets,R. (1995) Evolution of ATP-binding cassette transporter genes. *Curr. Opin. Genet. Dev.*, **5**, 79–785.
3. Higgins,C.F., Hiles,I.D., Salmond,G.P.C., Gill,D.R., Downie,J.A., Evans,I.J., Holland,B., Gray,L., Buckel,S.D., Bell,A.W. and Hermodson,M.A.(1986) A family of related ATP-binding subunits coupled to many distinct biological processes in bacteria. *Nature*, **323**, 448–450.
4. Hyde,S.C., Emsley,P., Hartshorn,M.J., Mimmack,M.M., Gileadi,U., Pearce,S.R., Gallagher,M.P., Gill,D.R., Hubbard,R.E. and Higgins,C.F. (1990) Structural model of ATP-binding proteins associated with cystic fibrosis, multidrug resistance and bacterial transport. *Nature*, **346**, 362–365.
5. Luciani,M.F., Denizot,F., Savary,S., Mattei,M.G. and Chimini,G. (1994) Cloning of two novel ABC transporters mapping on human chromosome 9. *Genomics*, **21**, 150–159.
6. Dean,M., Allikmets,R., Gerrard,B., Stewart,C., Kistler,A., Shafer,B., Michaelis,S. and Strathern,J. (1994) Mapping and sequencing of two yeast genes belonging to the ATP-binding cassette superfamily. *Yeast*, **10**, 377–383.
7. Allikmets,R., Gerrard,B., Glavac,D., Ravnik-Glavac,M., Jenkins,N.A., Gilbert,D.J., Copeland,N.G., Modi,W. and Dean,M. (1995) Characterization and mapping of three new mammalian ATP-binding transporter genes from an EST database. *Mammalian Genome*, **6**, 114–117.
8. Allikmets,R., Gerrard,B., Court,D. and Dean,M. (1993) Cloning and organization of the *abc* and *mdl* genes of *Escherichia coli*: relationship to eukaryotic multidrug resistance. *Gene*, **136**, 231–236.
9. Leighton,J. and Schatz,G. (1995) An ABC transporter in the mitochondrial inner membrane is required for normal growth of yeast. *EMBO J.*, **14**, 188–195.
10. Gerrard,B., Stewart,C. and Dean,M. (1993) Analysis of Mdr50: a *Drosophila* P-glycoprotein/multidrug resistance gene homolog. *Genomics*, **17**, 83–88.
11. Gerlach,J.H., Kartner,N., Bell,D.R. and Ling, V. (1986) Multidrug resistance. Imperial Cancer Research Fund. *Cancer Surv.* **5**, 25–45.
12. Gros,P., Croop,J. and Housman,D. (1986) Mammalian multidrug resistance gene: complete cDNA sequence indicates strong homology to bacterial transport proteins. *Cell*, **47**, 317–380.
13. Croop,J., Gros,P. and Housman,D. (1988) Genetics of multidrug resistance. *J. Clin. Invest.* **81**, 1303–1309.
14. Cole,S.P.C., Bhardwaj,G., Gerlach,J.H., Mackie,J.E., Grant,C.E., Almquist,K.C., Stewart,A.J., Kurz,E.U., Duncan,A.M.V. and Deeley,R.G. (1992) Overexpression of a transporter gene in a multidrug-resistant human lung cancer cell line. *Science*, **258**, 1650–1654.
15. Spies,T., Bresnahan,M., Bahram,S., Arnold,D., Blanck,G., Mellins,E., Pious,D. and 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.
16. 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., Ianuzzi,M.C., Collins,F.S. and Tsui,L.-C. (1989) Identification of the cystic fibrosis gene: cloning and characterization of the complementary DNA. *Science*, **245**, 1066–1073.
17. Mosser,J., Douar,A.-M., Sarde,C.-O., Kioschis,P., Feil,R., Moser,H., Poustka,A.-M., Mandel,J.-L. and Aubourg,P. (1993) Putative X-linked adrenoleukodystrophy gene shares unexpected homology with ABC transporters. *Nature*, **361**, 726–730.
18. Thomas,P.M., Cote,G.J., Wohllk,N., Haddad,B., Mathew,P.M., Rabl,W., Aguilar-Bryan,L., Gagel,R.F. and Bryan,J. (1995) Mutations in the sulfonylurea receptor gene in familial persistent hyperinsulinemic hypoglycemia of infancy. *Science*, **268**, 426–429.
19. Shimozawa,N., Tsukamoto,T., Suzuki,Y., Orii,T., Shirayoshi,Y., Mori,T. and Fujiki,Y. (1992) A human gene responsible for Zellweger syndrome that affects peroxisome assembly. *Science*, **255**, 1132–1134.
20. Ames,G. and Lecar,H. (1992) ATP-dependent bacterial transporters and cystic fibrosis: analogy between channels and transporters. *FASEB J.*, **6**, 2660–2666.
21. Higgins, C.F. (1992) ABC transporters: From micro-organisms to man. *Annu. Rev. Cell. Biol.*, **8**, 67–113.
22. Adams,M.D., Kerlavage,A.R., Fields,C. and Venter,J.C. (1993) 3400 new expressed sequence tags identify diversity of transcripts in human brain. *Nature Genet.*, **4**, 256–267.
23. Boguski,M.S., Lowe,T.M.J. and Tolstoshev,C.M. (1993) The EST express gathers speed. *Nature Genet.*, **4**, 331–332.
24. Lennon,G.G., Auffray,C., Polymeropoulos,M. and Soares,M.B. (1996) The I.M.A.G.E. consortium: An integrated molecular analysis of genomes and their expression. *Genomics*, **33**, 151–152.
25. Grausz,J.D. and Auffray,C. (1993) Strategies in cDNA programs. *Genomics*, **17**, 530–532.
26. Boguski,M.S., Lowe,T.M. and Tolstoshev,C.M. (1993) dbEST-database for 'expressed sequence tags'. *Nature Genet.*, **4**, 332–333.
27. Altschul,S.F., Gish,W., Miller,W., Myers, E.W. and Lipman,D.J. (1990) Basic local alignment search tool. *J. Mol. Biol.*, **215**, 403–410.
28. Feng,D.-F. and Doolittle,R.F. (1987) A multiple sequence alignment using a simplification of the progressive alignment method. *J. Mol. Evol.*, **35**, 351–360.
29. Devereaux,J., Haeberli,P. and Smithies,O. (1984) A comprehensive set of sequence analysis programs for the VAX. *Nucleic Acids Res.*, **12**, 387–395.
30. Bassett,D.E., Boguski,M.S., Spencer,F., Reeves,R., Goebl,M. and Hieter,P. (1995) Comparative genomics, genome cross-referencing and XREFdb. *Trends Genet.*, **11**, 372–373.
31. Felsenstein,J. (1989) PHYLIP—Phylogeny Inference Package Version 3.5c.
32. Lincoln,A.L., Daly,M. and Lander,E. (1991) PRIMER: a computer program for automatically selecting PCR primers. Whitehead Institute Technical Report.
33. Totaro,A., Grifa,A., Roetto,A., Lunardi,C., D'Agruma,L., Sbaiz,L., Zelante,L., De Sandre,G., Camaschella,C. and Gasparini,P. (1995) New polimorphisms and markers in the HLA class I region: relevance to hereditary hemochromatosis (HFE). *Human Genet.* **95**, 429–434.
34. Vazquez de Aldana,C.R., Marton,M.J. and Hinnebusch,A.G. (1995) GCN20, a novel ATP binding cassette protein, and GCN1 reside in a complex that mediates activation of the elF-2a kinase GCN2 in amino acid-starved cells. *EMBO J.* **14**, 3184–3199.
35. Paulusma,C.C., Bosma,P.J., Zaman,G.J.R., Bakker,C.T.M., Otter,M., Scheffer,G.L., Scheper,R.J., Borst,P. and Oude Elferink,R.P.J. (1996) Congenital jaundice in rats with a mutation in a multidrug resistance-associated protein gene. *Science* **271**, 1126–1128.
36. Valle,D. and Gartner,J. (1993) Penetrating the peroxisome. *Nature* **361**, 682–683.
37. Michaelis,S. and Berkower,C. (1995) Sequence comparison of yeast ATP binding cassette (ABC) proteins. In Cold Spring Harbor Symposium on Quantitative Biology, vol.LX: *Protein Kinesis—The Dynamics of Protein Trafficking and Stability.* Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
38. Salehzada,T., Silhol,M., Steff,A.M., Lebleu,B. and Bisbal,C. (1993) 2′,5′-oligoadenylate-dependent RNase L is a dimer of regulatory and catalytic subunits. *J. Biol. Chem.*, **268**, 7733–7740.
39. Hillis,D.M. and Bull,J.J. (1993) An empirical test of bootstrapping as a method for assessing confidence in phylogenetic analysis. *Syst. Biol.*, **42**, 182–192.