Journal of Bioenergetics and Biomembranes, Vol. 33, No. 6, December 2001 (© 2001)

Complete Characterization of the Human ABC Gene Family

Michael Dean[1,3] and Rando Allikmets[2]

The human ATP-binding cassette (ABC) transporters comprise a large family of membrane transport proteins and play a vital role in many cellular processes. The genes provide functions as diverse as peptide transport, cholesterol and sterol transport, bile acid, retinoid, and iron transport. In addition some ABC genes play a role as regulatory elements. Many ABC genes play a role in human genetic diseases, and several are critical drug transport proteins overexpressed in drug resistant cells. Analysis of the gene products allows the genes to be grouped into seven different subfamilies.

KEY WORDS: ATP-binding cassette; transporter; ABC genes.

INTRODUCTION

The ATP-binding cassette (ABC) genes represent the largest family of transporter genes and many of these genes are implicated in disease processes and/or drug resistance (Allikmets *et al.*, 1996; Decottignies and Goffeau, 1997; Higgins, 1992; Michaelis and Berkower, 1995). The prototype ABC protein binds ATP and uses this energy to transport molecules across cell membranes. While hydrophobic compounds are the most common substrates, ABC genes are capable of transporting metal ions, peptides, and sugars. The functional protein contains two ATP-binding domainsor nucleotide binding folds (NBF) and two transmembrane (TM) domains. The genes are typically organized as full transporters containing two of each domain, or half transporters.

ABC genes are abundant in the genomes of bacteria and archaebacteria where their principal role is in the import of essential molecules (Ames and Lecar, 1992; Higgins *et al.*, 1990). The yeast genome contains 29 ABC genes, and most of these transporters function to move compounds out of the cell or into intracellular organelles (Decottignies and Goffeau, 1997; Michaelis and Berkower, 1995). Analysis of amino acid sequence alignments of the ATP-binding domains has allowed the ABC genes to be classified into subfamilies. There are seven ABC gene subfamilies in the human genome with all except the OABP group containing multiple members. For the most part these subfamilies contain genes that also display considerable identity in the TM domains and have identical gene organization, and similar intron location. Five of these subfamilies are also found in the yeast genome, indicating that these groups were established early in the evolution of eukaryotes, and have been retained. However, the function of ABC genes corresponds poorly to subfamily organization, and often genes in different subfamilies share more similarity in substrate recognition than do genes in the same subfamily.

The *ABC1* gene subfamily is not represented in the yeast genome, but is in the *C. elegans* and Drosophila genomes, suggesting that it arose after eukaryotes achieved multicellularity. The best characterized ABC1-like genes are *ABCA4* (*ABCR*), a rod photospecific transporter of retinoids, and *ABCA1*, a transporter of cholesterol from peripheral tissues onto high density lipoproteins (Allikmets *et al.*, 1997a,b; Langmann *et al.*, 1999; Luciani *et al.*, 1994; Sun *et al.*, 1999). *ABCA1* is also involved in the phagocytic engulfment of cells (Moynault *et al.*, 1998). The *ABCA4* gene is mutated in several retinal degenerative diseases and the *ABCA1* gene is responsible for Tangier disease (a reverse cholesterol transport deficiency) and some forms of hypolipoproteinemia (Allikmets *et al.*, 1997a,b; Bodzioch *et al.*, 1999; Brooks-Wilson *et al.*, 1999; Cremers *et al.*, 1998; Martinez-Mir *et al.*, 1997; Rust *et al.*, 1999). Therefore characterization of new ABCA family genes is likely to yield biologically important transporters that may play a role in human pathology.

[1] Human Genetics Section, Laboratory of Genomic Diversity, NCI-Frederick, Frederick, Maryland.
[2] Departments of Ophthalmology and Pathology, Columbia University, New York.
[3] To whom correspondence should be addressed.

0145-479X/01/1200-0475$19.50/0

CHARACTERIZATION OF THE HUMAN ABC GENE SUPERFAMILY

To characterize the complete complement of ABC genes in the human and murine genomes we identified EST clones that displayed homology with other known ABC genes (Allikmets *et al.*, 1996; Allikmets and Dean, 1998). Each clone was obtained and the complete sequence of the clone determined. An alignment including any additional sequence from the EST database was formed and the gene sequence compared to other known genes. Unique genes were mapped in the mouse and human genomes using radiation hybrid and/or mouse backcross mapping (Schriml and Dean, 2000). Because most ABC genes are dispersed, the genetic location was often sufficient to document that a partially characterized gene indeed represented a new gene family member.

Table I shows the complete list of the 45 ABC genes described to date. The genes are dispersed onto 16 different autosomes and two genes reside on the X chromosome. The amino acid sequence of the ATP-binding domain was used to classify each gene into an existing ABC gene subfamily. Table II displays a list of the subfamilies and the number of genes described in each. The ABC1, MDR, and MRP/CFTR subfamilies are the most abundant containing 10–11 gene each. A phylogenetic tree of all the ABC genes with at least one complete NBF is shown in Fig. 1 and statistical analysis clearly supports the separation of the genes into subfamilies.

Table I. List of Human ABC Genes and Their Chromosomal Location

Symbol	Alias	Location	Subfamily	Function
ABCA1	*ABC1*	9q31	ABC1	Cholesterol transport
ABCA2	*ABC2*	9q34	ABC1	Drug resistance
ABCA3	*ABCC*	16p13.3	ABC1	
ABCA4	*ABCR*	1p13-p21	ABC1	Rod photoreceptor retinoid transport
ABCA5		17q24	ABC1	
ABCA6		17q24	ABC1	
ABCA7		19p13	ABC1	
ABCA8		17q24	ABC1	
ABCA9		17q24	ABC1	
ABCA10		17q24	ABC1	
ABCA12		2q34	ABC1	
ABCA13		7p11-q11	ABC1	
ABCB1	*PGY1*	7q21.1	MDR	Drug resistance
ABCB2	*TAP1*	6p21	MDR	Peptide transport
ABCB3	*TAP2*	6p21	MDR	Peptide transport
ABCB4	*PGY3*	7q21.1	MDR	Bile-acid transport
ABCB5		7p15	MDR	
ABCB6		2q32	MDR	Iron transport
ABCB7	*ABC7*	Xq21-q22	MDR	Iron transport
ABCB8		7q35-q36	MDR	
ABCB9		12q23-q24	MDR	
ABCB10		1q32	MDR	
ABCB11	*SPGP*	2q24	MDR	Bile-acid transport
ABCC1	*MRP1*	16q12	CF/MRP	Drug resistance
ABCC2	*MRP2*	10q23-q24	CF/MRP	Bile-acid transport
ABCC3	*MRP3*	17q21.3	CF/MRP	
ABCC4	*MRP4*	13q31	CF/MRP	Nucleoside transport
ABCC5	*MRP5*	3q25-q26	CF/MRP	
ABCC6	*MRP6*	16p13.1	CF/MRP	
ABCC7	*CFTR*	7q31.2	CF/MRP	Chloride ion channel
ABCC8	*SUR*	11p15.1	CF/MRP	Sulfonylurea receptor
ABCC9	*SUR2*	12p12.1	CF/MRP	
ABCC10	*MRP7*	6p21	CF/MRP	
ABCC11		16p11-q11	CF/MRP	
ABCC12		16p11-q11	CF/MRP	
ABCD1	*ALD*	Xq28	ALD	VLCFA transport
ABCD2	*ALDL1*	12q11	ALD	
ABCD3	*PXMP1*	1p22-p21	ALD	
ABCD4	*PMP69*	14q12	ALD	
ABCE1	*OABP*	4q28	OABP	Oligoadenylate-binding protein
ABCF1		6p21	GCN20	
ABCF2		7q36	GCN20	
ABCF3		3q25	GCN20	
ABCG1	*White*	21q22.3	White	
ABCG2	*ABCP*	4q22	White	Drug resistance
ABCG4	*White2*	11q23	White	
ABCG5	*White3*	2p21	White	

Note. The gene designated *ABCA11* on Chromosome 4p appears to be a pseudogene. An ABCC-like gene is present on Chromosome 21, however, expression of the gene in cells has yet to be documented.

ABCA Subfamily

The ABCA subfamily is composed of full transporters that cluster into two groups based on gene structure and homology. The first group contains *ABCA1*, *2*, *3*, *4*, and *7*. The second group is composed of a cluster of at least five genes located on Chromosome 17q24 (*ABCA5*, *6*, *8*, *9*, and *10*) (Schriml *et al.*, 2000). The five genes on 17q24 are closely spaced and arranged in a head-to-tail fashion and all encode full transporters. The two groups of ABC1-like genes are distinguished by having different intron structures with ABCA1 and ABCA4 containing 50 introns and *ABCA6*, *8*, and *9* having 38 introns. The amino terminal portion of the *ABCA1*, *2*, *3*, *4*, and *7* genes are considerably longer than the other group.

The *White/ABC8* Gene Subfamily

The first eukaryotic mutation that was described was the white eye mutant in *Drosophila*. The lack of pigmentation in the eyes of these flies is due to the functional loss of transporters for guanine and tryptophan, two eye pigment precursors. The *white* gene is a half transporter ABC molecule and is one of three related

Table II. Diseases Caused by ABC Genes

Gene	Phenotype
ABCA1	Tangier disease, FHDLD
ABCA4	Stargardt/FFM, RP19, CRD, AMD
ABCB2	Immune deficiency
ABCB3	Immune deficiency
ABCB4	PFIC-3
ABCB7	XLSA/A
ABCB11	PFIC-2
ABCC2	Dubin–Johnson Syndrome
ABCC6	Pseudoxanthoma elasticum
ABCC7	Cystic Fibrosis
ABCC8	FPHHI
ABCD1	ALD

Note. FHDLD – familial hypoapoproteinemia; FFM – fundus flavimaculatis; RP19 – retinitis pigmentosum 19; CRD – cone-rod dystrophy; AMD – age-related macular degeneration; PFIC – progressive familial intrahepatic cholestasis; XLSA/A – X-linked sideroblasosis and anemia; FPHHI – persistent hyperinsulinemic hypoglycemia of infancy due to focal adenomatous hyperplasia; ALD – adrenoleukodystrophy.

genes in flies, the others being *brown* and *scarlet* (Pepling and Mount, 1990). These genes are unique in that they have the NBF at the N-terminal end of the molecule, whereas all other ABC genes are in either a TM-NBF or TM-NBF-TM-NBF arrangement. A mammalian gene with considerable identity to *white* has been described (Chen *et al.*, 1996). This gene, *ABCG1*, (*ABC8/WHITE*) is also a half transporter with the NBF at the N-terminus. *ABCG1* is induced in cells subjected to cholesterol loading and is speculated to play a role in this process (Klucken *et al.*, 2000). At least three other *White/ABC8* subfamily genes are present in the human genome (*ABCG2*, *4*, and *5*) (Shulenin *et al.*, 2001). An additional ABCG-related gene, *Abcg3* (*Abcp2*) has to date only been identified in the mouse genome (Mickley *et al.*, 2001). The yeast genome contains one half transporter that is related to the family (ADP1) and nine full transporter genes arranged in a NBF-TM-NBF-TM fashion. Several of the yeast full transporters are involved in resistance to cytotoxic drugs in both *S. cerevisiae* and other yeast species.

Human Diseases and ABC Gene

Several ABC genes have been implicated in both Mendelian and complex genetic disorders including, cystic fibrosis, adrenoleukodystrophy, Stargardt disease, Tangier disease, and several hepatic disorders. Consistent with the wide spectrum of function of the ABC genes, the diseases they are associated with are also diverse (Table II). Cystic fibrosis is characterized by abnormal exocrine secretions causing pancreatic degeneration and lung infections (Tsui, 1995). It is caused by mutations in *CFTR*, the only ABC gene that is an ion channel (Riordan *et al.*, 1989). The incidence of pancreatitis, a complex disorder is increased in *CFTR* mutation carriers (Cohn *et al.*, 1998; Sharer *et al.*, 1998). Adrenoleukodystrophy (ALD) patients suffer from neurological complications and the ALD (*ABCD1*) gene (Mosser *et al.*, 1993). The ALD protein is located in the peroxisome where it is believed to be involved in the transport of very long chain fatty acids. The *ABCA4* gene is expressed exclusively in rod photoreceptors where it transports retinol derivatives from the rod disks into the cytoplasm. A complete loss of ABCA4 function leads to a retinitis pigmentosum phenotype whereas patients with at least one missense allele have Startgardt disease (STGD) (Martinez-Mir *et al.*, 1997). STGD is characterized by juvenile to early adult onset of retinal degeneration and loss of central vision. *ABCA4* mutation carriers are also increased in frequency in age-related macular degeneration (AMD) patients (Allikmets *et al.*, 1997a,b). AMD patients display loss of central vision after the age of 60 and the causes of the disease are poorly understood (Lewis and Lupski, 2000). The abnormal accumulation of retinoids, due to a deficiency in ABCA4, has been postulated to be one mechanism by which this process could be initiated.

Tangier disease is characterized by deficient cholesterol efflux from peripheral cells, such as macrophages, and a very low level of high density lipoproteins (HDL). The disease is caused by alterations in the *ABCA1* gene, implicating this protein in the pathway of removal of cholesterol onto HDL. Patients with hypolipidemia have also been described that are heterozygous for *ABCA1* mutations, suggesting that ABC1 variations may play a role in regulating the level of HDLs in the blood (Marcil *et al.*, 1999).

DISCUSSION

With the increasing availability of human and murine cDNA and genomic sequence, additional members of the ABC gene superfamily continue to be identified. Currently there are 45 ABC genes that have been characterized in either the human or mouse genomes. *ABCA10* and *ABCG4* represent the only ABC genes that we identified exclusively from genomic sequence data. However, they clearly represent expressed genes as there are several ESTs that contain portions of the 3′ untranslated region of *ABCA10*, and PCR expression and Northern blot analysis shows that both genes are expressed in several human tissues.

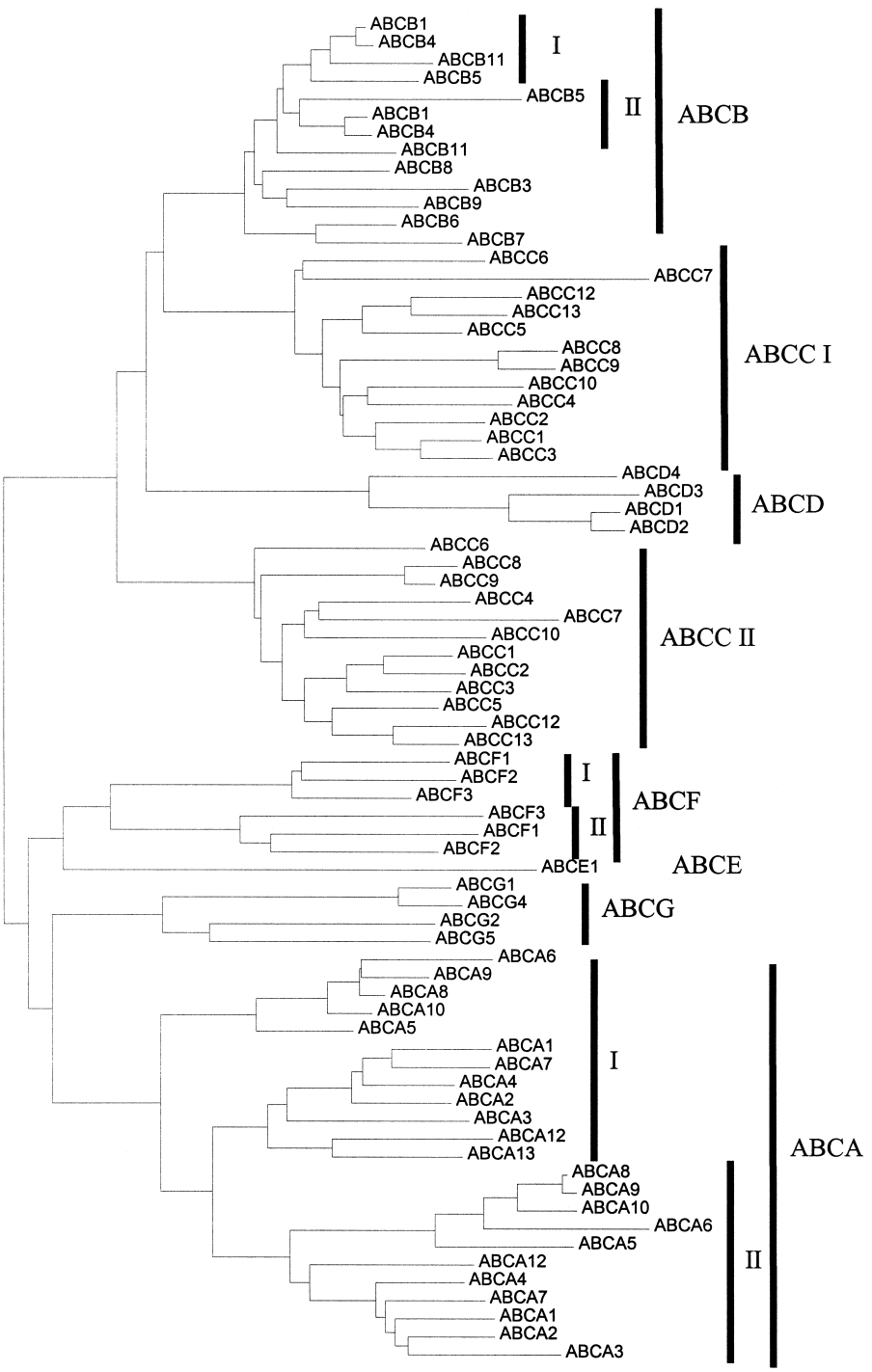

Fig. 1. Phylogenetic analysis of human ABC genes. An alignment of the NBF domain of each human ABC gene was produced and used to derive a phylogenetic tree. The subfamily groups are shown as thick bars on the right and I and II represent the N and C-terminal NBFs respectively.

The ABC1-like genes and OABP represent the only subfamily of ABC genes that do not have identifiable paralogs in the yeast genome. There is however at least one ABC1-related gene in *C. elegans* (*ced-7*) and several in *Drosophila*. Thus the *ABC1* genes appear to have arisen rather recently, as eukaryotes became multicellular and developed more sophisticated transport requirements. To date there are 11 known *ABC1* subfamily genes, making it the largest such group. However, we know relatively little about the function of these genes. In mammals the *ABCA1* gene is highly expressed in macrophages and monocytes and the *ced-7* gene plays a role in the recognition and engulfement of apoptotic cells in *C. elegans*. An important function of macrophages is to process the lipids that were derived from engulfed cells, and *ABCA1* appears to play a critical role in this process.

The ABC genes are highly dispersed in the human genome with no cluster larger than two related genes (TAP1 and TAP2; PGY1 and PGY3; ABCC1 (MRP) and ABCC6) identified to date. The PGY/MDR gene cluster contains three genes in rodents. Thus it is surprising to find a cluster of at least five ABC genes at the 17q24 locus. No obvious disease or phenotypes have been mapped to this region of the genome that would suggest a function for this group. Preliminary analysis of the mouse EST clones indicates that there is also a cluster of related genes in a region syntenic to Chromosome 17, suggesting that this gene cluster is rather ancient. In addition, the high level of expression of some of these genes in the ovary and other tissues is intriguing. Since all steroid hormones are derived from cholesterol, it is tempting to speculate that the gene products may play a role ovarian hormone biogenesis or transport. Knockout technology in the mouse may be needed to begin to understand the function of these transporters.

REFERENCES

Allikmets, R., and Dean, M. (1998). *Methods Enzymol.* **292**, 116–130.

Allikmets, R., Gerrard, B., Hutchinson, A., and Dean, M.(1996). *Hum. Mol. Genet.* **5**, 1649–1655.

Allikmets, R., Shroyer, N. F., Singh, N., Seddon, J. M., Lewis, R. A., Bernstein, P. S., Peiffer, A., Zabriskie, N. A., Li, Y., Hutchinson, A., Dean, M., Lupski, J. R., and Leppert, M. (1997a). *Science* **277**, 1805–1807.

Allikmets, R., Singh, N., Sun, H., Shroyer, N. F., Hutchinson, A., Chidambaram, A., Gerrard, B., Baird, L., Stauffer, D., Peiffer, A., Rattner, A., Smallwood, P., Li, Y., Anderson, K. L., Lewis, R. A., Nathans, J., Leppert, M., Dean, M., and Lupski, J. R. (1997b). *Nat. Genet.* **15**, 236–246.

Ames, G., and Lecar, H. (1992). *FASEB J.* **6**, 2660–2666.

Bodzioch, M., Orso, E., Klucken, J., Langmann, T., Bottcher, A., Diederich, W., Drobnik, W., Barlage, S., Buchler, C., Porsch-Ozcurumez, M., Kaminski, W. E., Hahmann, H. W., Oette, K., Rothe, G., Aslanidis, C., Lackner, K. J., and Schmitz, G. (1999). *Nat. Genet.* **22**, 347–351.

Brooks-Wilson, A., Marcil, M., Clee, S. M., Zhang, L. H., Roomp, K., van Dam, M., Yu, L., Brewer, C., Collins, J. A., Molhuizen, H. O., Loubser, O., Ouelette, B. F., Fichter, K., Ashbourne-Excoffon, K. J., Sensen, C. W., Scherer, S., Mott, S., Denis, M., Martindale, D., Frohlich, J., Morgan, K., Koop, B., Pimstone, S., Kastelein, J. J., and Hayden, M. R. (1999). *Nat. Genet.* **22**, 336–345.

Chen, H., Rossier, C., Lalioti, M. D., Lynn, A., Chakravarti, A.,Perrin, G., and Antonarakis, S. E. (1996). *Am. J. Hum. Genet.* **59**, 66–75.

Cohn, J. A., Friedman, K. J., Noone, P. G., Knowles, M. R., Silverman, L. M., and Jowell, P. S. (1998). *N. Engl. J. Med.* **339**, 653–658.

Cremers, F. P., van de Pol, D. J., van Driel, M., den Hollander, A. I., van Haren, F. J., Knoers, N. V., Tijmes, N., Bergen, A. A., Rohrschneider, K., Blankenagel, A., Pinckers, A. J., Deutman, A. F., and Hoyng, C. B. (1998). *Hum. Mol. Genet.* **7**, 355–362.

Decottignies, A., and Goffeau, A. (1997). *Nat. Genet.* **15**, 137–145.

Higgins, C. F. (1992). *Annu. Rev. Cell. Biol.* **8**, 67–113.

Higgins, C. F., Hiles, I. D., Slamond, 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. (1990). *Nature* **323**, 448–450.

Klucken, J., Buchler, C., Orso, E., Kaminski, W. E., Porsch-Ozcurumez, M., Liebisch, G., Kapinsky, M., Diederich, W., Drobnik, W., Dean, M., Allikmets, R., and Schmitz, G. (2000). *Proc. Natl. Acad. Sci. U.S.A.* **97**, 817–822.

Langmann, T., Klucken, J., Reil, M., Liebisch, G., Luciani, M. F., Chimini, G., Kaminski, W. E., and Schmitz, G. (1999). *Biochem. Biophys. Res. Commun.* **257**, 29–33.

Lewis, R. A., and Lupski, J. R. (2000). *Hosp. Pract. (Off. Ed.)* **35**, 41–50, 56–58.

Luciani, M. F., Denizot, F., Savary, S., Mattei, M. G., and Chimini, G. (1994). *Genomics* **21**, 150–159.

Marcil, M., Brooks-Wilson, A., Clee, S. M., Roomp, K., Zhang, L. H., Yu, L., Collins, J. A., van Dam, M., Molhuizen, H. O., Loubster, O., Ouellette, B. F., Sensen, C. W., Fichter, K., Mott, S., Denis, M., Boucher, B., Pimstone, S., Genest, J. Jr., Kastelein, J. J., and Hayden, M. R. (1999). *Lancet* **354**, 1341–1346.

Martinez-Mir, A., Bayes, M., Vilageliu, L., Grinberg, D., Ayuso, C., del Rio, T., Garcia-Sandoval, B., Bussaglia, E., Baiget, M., Gonzalez-Duarte, B., and Balcells, S. (1997). *Genomics* **40**, 142–146.

Michaelis, S., and Berkower, C. (1995). *Sequence comparison of yeast ATP binding cassette (ABC) proteins.* In *Cold Spring Harbor Symposium*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Mickley, L., Jain, P., Miyake, K., Schriml, L. M., Rao, K., Fojo, T., Bates, S., and Dean, M. (2001). *Mamm. Genome.* **12**, 86–88.

Mosser, J., Douar, A. M., Sarde, C. O., Kioschis, P., Feil, R., Moser, H., Poustka, A. M., Mandel, J. L., and Aubourg, P. (1993). *Nature* **361**, 726–730.

Moynault, A., Luciani, M. F., and Chimini, G. (1998). *Biochem. Soc. Trans.* **26**, 629–635.

Pepling, M., and Mount, S. M. (1990). *Nucl. Acids Res.* **18**, 1633.

Riordan, J. R., Rommens, J. M., Kerem, B., Alon, N., Rozmahel, R., Grzelczak, Z., Zielenski, J., Lok, S., Plavsic, N., Chou, J. L., *et al.* (1989). *Science* **245**, 1066–1073.

Rust, S., Rosier, M., Funke, H., Real, J., Amoura, Z., Piette, J. C., Deleuze, J. F., Brewer, H. B., Duverger, N., Denefle, P., and Assmann, G. (1999). *Nat. Genet.* **22**, 352–355.

Schriml, L. M., and Dean, M. (2000). *Genomics* **64**, 24–31.

Sharer, N., Schwarz, M., Malone, G., Howarth, A., Painter, J., Super, M., and Braganza, J. (1998). *N. Engl. J. Med.* **339**, 645–652.

Shulenin, S., Schriml, L. M., Remaley, A. T., Fojo, S., Brewer, B., Allikmets, R., and Dean, M. (2001). *Cytogenet. Cell Genet.* **92**, 204–208.

Sun, H., Molday, R. S., and Nathans, J. (1999). *J. Biol. Chem.* **274**, 8269–8281.

Tsui, L. C. (1995). *Am. J. Respir. Crit. Care. Med.* **151**, S47–S53.