

**Supplementary Figure 1. CFTR protein structure and domain architecture.** (A) Open state CFTR homology model, ribbon representation from Serohijos et al. 2008 PNAS 105:3256. CFTR domains are color coded as shown in the legend. (B) Schematic representation of CFTR domain organization in and around the plasma membrane.



**Supplementary Figure 2. Sequence similarity and domain architecture of CFTR and related ABC transporters.** (A) Statistical significance of BLASTP hits of the human CFTR protein sequence against representative human ABC family members. (B) Dendogram showing the relationship between CFTR and representative ABC family members along with their domain architectures.

	1-1	2-2		3-3
CFTR	MQRSPLEKASVVSKLFFS	NTRPILRKGYRQRLELSDIYQIPSVDSADNI	LSEKLEREWDRELASK	KNPKLINALRRC
ABCC4	MLPVYQEVKPNPLQDANLCSRVFFW	NLNPLFKIGHKRRLEEDDMYSVLPEDRSOHI	LGEELQUEWDKEVLRAE	ENDAQKPSLTRAIIKC
CFTR	FFWR <mark>FMFYGIFLYLG</mark> EVTKAVQPLL	4-4 LGRIIASYDPDNKEERSIAIYLGIGLCI	LLFIVRTLLLHPAIFGI	HHIGMQMRIAMFSLI
ABCC4	YWKS <mark>YLVLGIFTLIE</mark> ESAKVIQPIF.	LGKIINYFENYDPMDSVALNTAYAYATVLTI	FCTLILAILHHLYFYHV	/QCAGMRLRVAMCHMI
CETP	4-4 5-5	- NNINKEDECIALAHEVWITADI OVALIMCI I	6	TATEOACI CRMMMKY
ABCC4	YRKALRLSNMAMGKTTTGQIVNLLS	NDVNKFDQVTVFLHFLWAGPLQAIAVTALLU	WMEIGISCLAGMAVLI	ILLPLQSCFGKLFSSL
	7-7		8-8	
CFTR ABCC4	RÞQRAGKISERLVITSEMIENIQSV RSKTATFTDARIRTMNEVITGIRII	KAYCWEEAMEKMIENLRQTELKLTRKAAYVI KMYAWEKSFSNLITNLRKKEISKILRSSCLI	RYFNSSAFFFSGFFVVI RGMNLASFFSASKIIVI	SLSVLPYALIKGIIL- SVTFTTYVLLGSVITA
	8-8	9-9		10-10
CFTR	RKIFTTISFCIVLRMAVTRQFPWAV	QTWYDSLGAINKIQDFLQKQEYKTLEYNLT	ITEVVMENVTAFWE	EEGFGELFEKAKQNNN
ABCC4	5RVFVAVTLIGAVRLTVTLFFPSAL	<u>ERVSEA</u> IVSIRRIQTFLLLDEISQRNRQLPS	SDGKKMVHVQDFTAFWI	
CFTR	NRKTSNGDDSLFFSNFSLLGTPVLK	DINFKIERGQLLAVAGSTGAGKTSLLMVIM	GELEPSEGKIKHSGRIS	SFCSQFSWIMPGTIKE
ABCC4	ASETPTLQ	GLSFTVRPGELLAVVGPVGAGKSSLLSAVLO	GELAPSHGLVSVHGRIA	AYVSQQPWVFSGTLRS
CFTR	11-11 NITEGVSYDEYRYRSVIKACOLEED	12-12	13-13 RAVYKDADI.YI.LOSPEC	14-14 SYLDVLTEKEIFESCV
ABCC4	NILFGKKYEKERYEKVIKACALKKD	LQLLEDGDLTVIGDRGTTLSGGQKARVNLA	RAVYQDADIYLLDDPLS	SAVDAEVSRHLFELCI
	14	14		
CFTR ABCC4	CKLMANKTRILVTSKMEHLKKADKI CQILHEKITILVTHQLQYLKAASQI	LILHEGSSYFYGTFSELQNLQPDFSSKL <mark>MGG</mark> LILKD	CDSFDQFSAERRNSIL1 	TETLHRFSLEGDAPVS
		14		
CFTR	WTETKKQSFKQTGEFGEKRKNSILN	PINSIRKFSIVQKTPLQMNGIEEDSDEPLE	RRLSLVPDSEQGEAILE	RISVISTGPTLQARR
ABCC4		14		15
CFTR	RQSVLNLMTHSVNQGQNIHRKTTAS	IRKVSLAPQANLTELDIYSRRLSQETGLEIS	SEEINEEDLKECFFDDM	MESIPAVTTWNTYLRY
ABCC4				
CFTR	15 ITVHKSLIFVLIWCLVIFLAE	15		
ABCC4	GKMV	QKGTYTEFLKSGIDFGSLLKKDNEESEQPP	VPGTPTLRNRTFSESV	VSQQSSRPSLKDGALE
CETD	15	1	6-17	17-18
ABCC4	SQDTENVPVTLSEENRSEGKVGFQA	YKNYFRAGAHWIVFIFLILLNTAAQVAYVLQ	QDWWLSYWANKQSMLNV	TVNGGGNVTEKLDLN
	17-18	17-19		18-20
CFTR ABCC4	SYAVIITSTSSYYVFYIYVGVADTL WYLGIYSGLTVAT	LAMGFFRGLPLVHTLITVSKILHHKMLHSVI VLFGIARSLLVFYVLVNSSOTLHNKMFESII	LQAPMSTLNTLKAGGII LKAPVLFFDRNPIGRII	LNRFSKDIAILDDLLP LNRFSKDIGHLDDLLP
	18-20 19-21		20-2	2
CFTR	LTIFDFIQLLLIVIGAIAVVAVLQP	YIFVATVPVIVAFIMLRAYFLQTSQQLKQL	ESEGRSPIFTHLVTSL	KGLWTLRAFGRQPYFE
ABCC4	DIFTDFIQITTQATTTO	3	21-24	22-25
CFTR	TLFHKALNLHTANWFLYLSTLRWFQ	MRIEMIFVIFFIAVTFISILTTG-EGEGRV(	GIILTLAMNIMSTLQWA	AVNSSIDVDSLMRSVSR
ABCC4	ELFDAHQDLHSE <mark>AW</mark> FLFLTTSRWFA	VRLDAICAMFVIIVAFGSLILAK <mark>TLDAGQ</mark> VO	GLALSYALTLMGMFQWC	CVRQSAEVENMM <mark>ISVER</mark>
CFTR	VFKFIDMPTEGKPTKSTKPYKNGOL	22-25 SKVMITENSHVKKDDIWPSGGOMTVKDLTAF	KYTEGGNATLENTSES1	23-26 SpgorVgllgrtgsgk
ABCC4	VIEYTDLEKEAPWEYQKRPPPAWPH	EGVIIFDNVNFMYSPGGPLVLKHLTA-	L]	IKSQEKVGIVGRTGAGK
	23-26	24-27		25-28
ABCC4	STLLSAFLRLLNTEGEIQIDGVSWD SSLISALFRLSEPEGKIWIDKILTT	SITLQQWRKAFGVIPQKVFIFSGTFRKNLD EIGLHDLRKKMSIIPQEPVLFTGTMRKNLDI	PYEQWSDQEIWKVADE PFNEHTDEELWNALQE	/GLRSVIEQFPGKLDFV /QLKETIEDLPGKMDTE
	25-28	26-		30
CFTR ABCC/	LVDGGCVLSHGHKQLMCLARSVLSK.	AKILLLDEPSAHLDPVTYQIIRRTLKQAFAI	DCTVILCEHRIEAMLEC	CQQFL
ADCC4		31		27
CFTR			VIEF	ENKVRQYDSIQKLLNER
ABCC4	PYVLLQNKESLFYKMVQQLGKAEAA	ALTETAKQVYFKRNYPHIGHTDHMVTNTSNO	GQPSTLTIFETAU	
CFTR	-SLFRQAISPSDRVKLFPHRNSSKC	KSKPQIAALKEETEEEVQDTRL		
ABCC4				
	TMD1	NBD1 R TM	D2 NBD2	

**Supplementary Figure 3. Pairwise alignment of human CFTR and ABCC4 proteins shown along with locations of the protein coding sequence exons.** The CFTR exon number is shown first followed by the corresponding ABCC4 exon. The locations of CFTR domains are indicated with colors as in Supplementary Figure 1.



В

					TMD1									NBD1								R						
CFTR	0	0	1	2	3	4	4	5	6	6	7	7	8	8	9	10	10	11	11	11	12	12	13	14	14	15	0	0
ABCC5	1	2	3	3	4	4	5	5	6	7	7	8	8	9	10	10	11	11	12	13	13	14	14	14	0	0	15	16
TMD2											NBD2																	
	-	_		_	_	_		_	_	_	_	_			_		_		_	_	_	_						
CFTR	16	17	17	17	18	18	19	19	20	20	21	21	22	22	23	23	24	24	25	25	25	26	27	27	27	0		
ABCC5	17	17	18	19	19	20	20	21	21	22	22	23	23	24	24	25	25	26	26	27	28	28	28	29	0	29		
	Conserved CFTR exons 8/27										0.30																	
	C	Conserved ABCC5 exons 5/29									0.17																	
	Average exon conservation									0.23																		

С



Supplementary Figure 4. Comparison of exon-intron structure between CFTR and ABCC subfamily members. Correspondence of exons between CFTR-ABCC4 (A) and CFTR-ABCC5 (B). Corresponding exonic regions are placed in the same column, and exons (or exonic regions) that do not have corresponding sequences are marked with 0. Locations of CFTR domains are indicated above. Fractions of conserved exons are shown below. (C) Average exon conservation (i.e. fractions of conserved exons) are shown between CFTR and members of the ABCC subfamily.



**Supplementary Figure 5. Comparison of gene structures between CFTR and closely related ABCC subfamily members.** The CFTR exon-intron structure is schematically represented in comparison to those of the most closely related ABCC subfamily members: ABCC4 & ABCC5. Exons are represented as black boxes and vertical blue lines between gene models connect corresponding (*i.e.* homologous) exons. For purposes of visual clarity, exons are drawn to scale, whereas introns are arbitrarily truncated.



Supplementary Figure 6. Comparative splice site analysis for CFTR-specific R domain splice sites versus the splice sites of the remaining CFTR exons. (A-B) Rates of sequence conservation (PhyloP scores) are compared for 5' splice acceptor sites and 3' splice donor sites for CFTR R domain encoding exons (red) versus the remaining CFTR exons (blue). (C-D) Average ± SE rates of splice site sequence conservation are shown along with the *P*-values for the differences. (E-F) Splice acceptor and donor sites are shown for CFTR R domain encoding (and proximal) exons 14, 15 & 16.



Supplementary Figure 7. Comparison of cell type-specificity for CFTR R domain encoding exons (red) versus the remaining CFTR exons (blue). Cell type-specifity values are shown for individual CFTR exons (A) along with average ± SE vales and (B) the *P*-values for the differences.



**Supplementary Figure 8. Phylogenies of CFTR and individual CFTR domains.** Phylogenies show two main groups: fish and other vertebrates. For each domain, the ratio B1/C shown in Figure 6 are calculated as the between group branch length (shown in red) divided by the average within group branch lengths (shown in black).

Human CFTR and ABCC subfamily protein & gene sequences								
Sequence Name	<b>RefSeq Accession</b>	Gene ID						
CFTR (ABCC7)	NP_000483	1080						
ABCC1	NP_004987	4363						
ABCC2	NP_000383	1244						
ABCC3	NP_003777	8714						
ABCC4	NP_005836	10257						
ABCC5	NP_005679	10057						
ABCC6	NP_001162	368						
ABCC8	NP_000343	6833						
ABCC9	NP_005682, NP_064693	10060						
ABCC10	NP_001185863	89845						
ABCC11	NP_115972	85320						
ABCC12	NP_150229	94160						
Vertebrate CFTR protein and coding nucleotide sequences								
Species	Protein Accession	Nucleotide Accession						
Human ( <i>Homo Sapiens</i> )	NP_000483	NM_000492						
Chimp (Pan troglodytes)	NP_001073386	NM_001079917						
Orangutan ( <i>Pongo abelii</i> )	NP_001162017	NM_001168545						
Rhesus ( <i>Macaca mulatta</i> )	NP_001028110	NM_001032938						
Rabbit (Oryctolagus cuniculus)	NP_001076185	NM_001082716						
Dog (Canis lupus familiaris)	NP_001007144	NM_001007143						
Cow (Bos taurus)	NP_776443	NM_174018						
Pig (Sus scrofa)	NP_001098420	NM_001104950						
Opossum (Monodelphis domestica)	XP_001368450	XM_001368413						
Rat (Rattus norvegicus)	NP_113694	NM_031506						
Mouse ( <i>Mus musculus</i> )	NP_066388	NM_021050						
Chicken (Gallus gallus)	NP_001099136	NM_001105666						
Frog ( <i>Xenopus laevis</i> )	AAC60023	U60209						
Zebrafish (Danio rerio)	NP_001038348	NM_001044883						
Salmon ( <i>Salmo salar</i> )	NP_001117005	NM_001123533						
Killifish (Fundulus heteroclitus)	AAC41271	AF000271						
Takifugu ( <i>Takifugu rubripes</i> )	NP_001041505	NM_001048040						
Pufferfish (Tetraodon nigroviridis)	AAR16330	AE017192						

Supplementary Table 1. CFTR and ABC transporter sequences used in the analyses reported here.