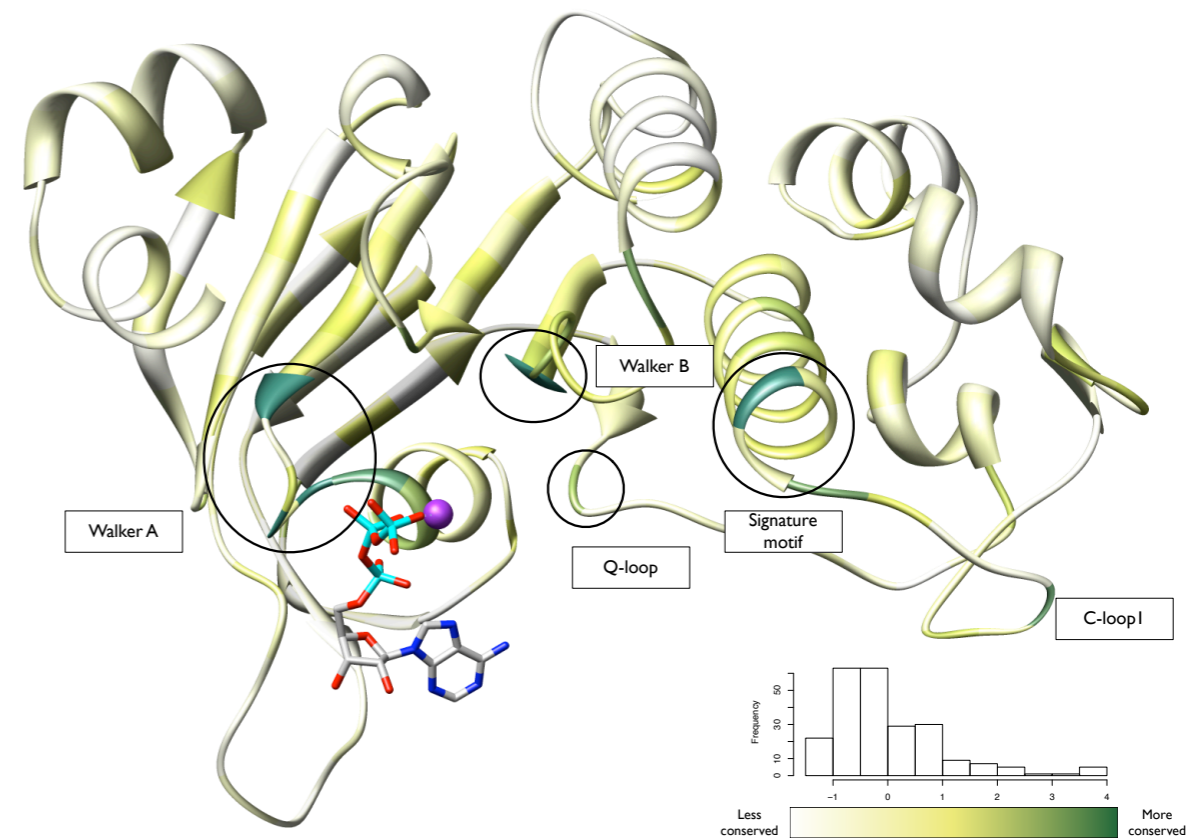


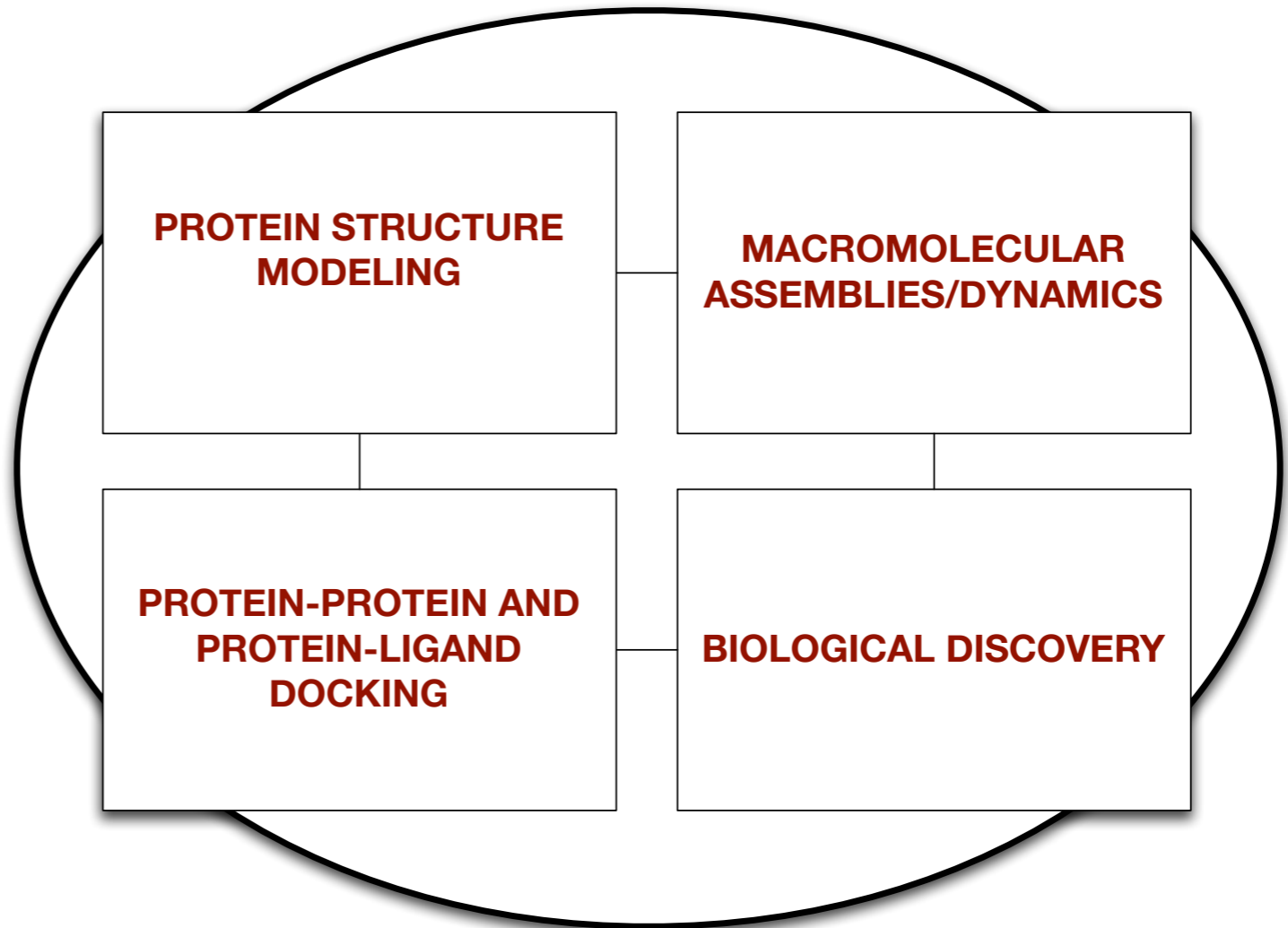
# Functional hotspots revealed by mutational analysis of ABC transporters

Libusha Kelly  
Andrej Sali lab, UCSF  
03.31.08



# Overview of the Sali lab

- predicting the structures of proteins
- determining the structures of macromolecular assemblies
- annotating the functions of proteins using structure



**MODELLER**  
Protein structure modeling

**IMP**  
Integrative Modeling Platform

**MODPIPE**  
Large-scale protein structure modeling

**MODBASE**  
Comprehensive database of protein structure models

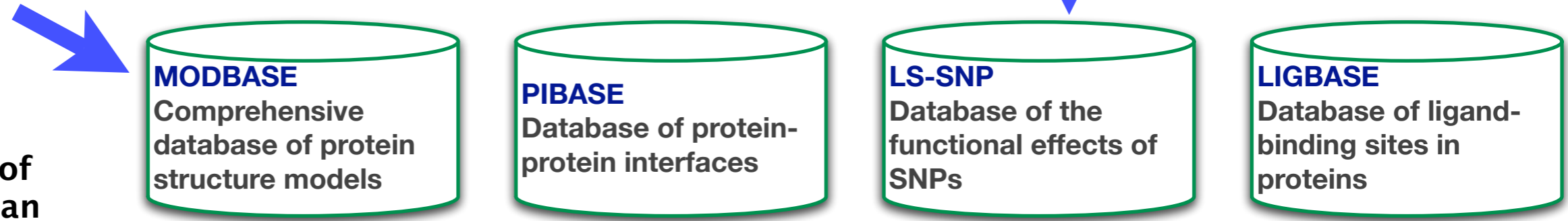
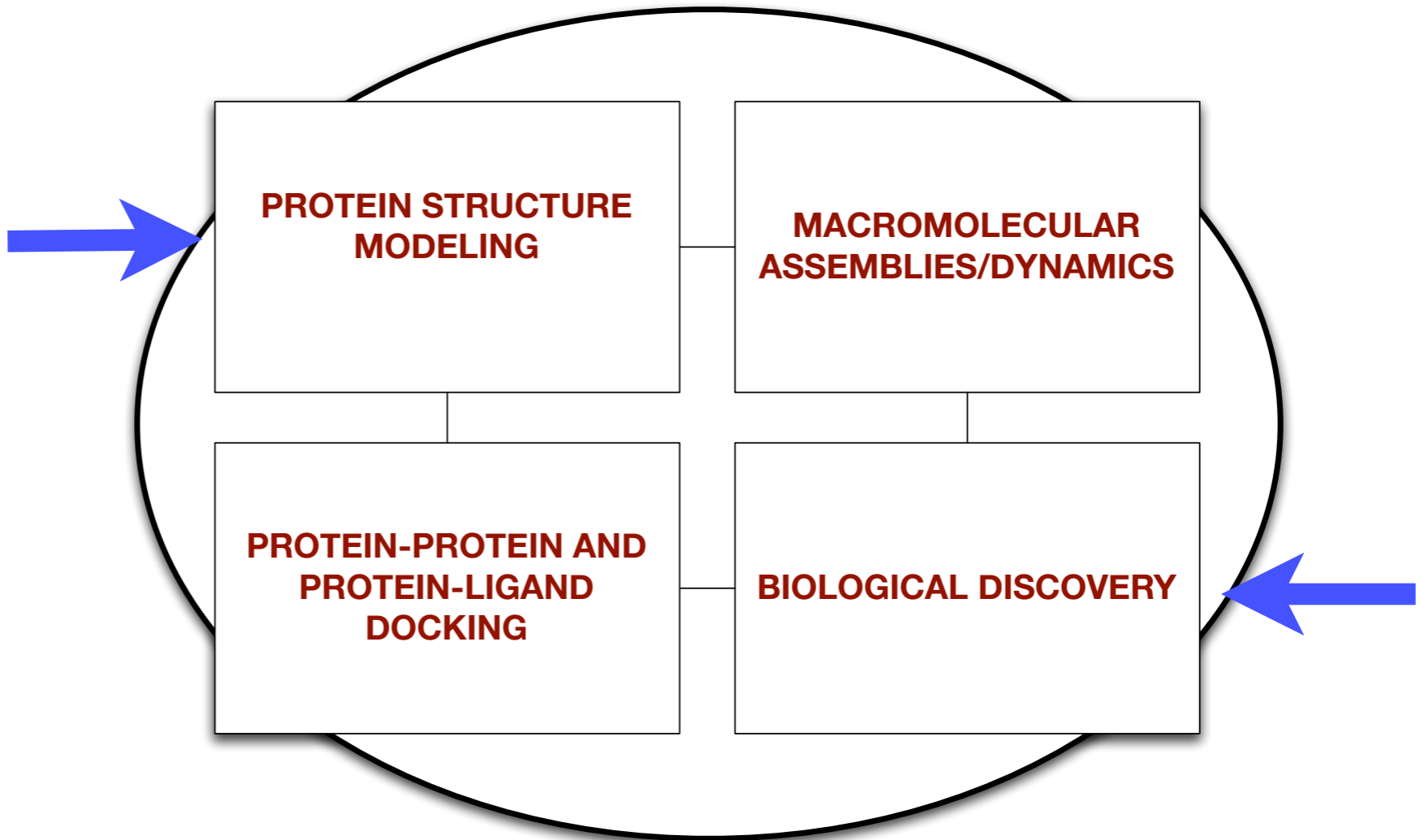
**PIBASE**  
Database of protein-protein interfaces

**LS-SNP**  
Database of the functional effects of SNPs

**LIGBASE**  
Database of ligand-binding sites in proteins

# Overview of the Sali lab

- **predicting the structures of proteins**
- **determining the structures of macromolecular assemblies**
- **annotating the functions of proteins using structure**



Slide courtesy of Eswar Narayanan

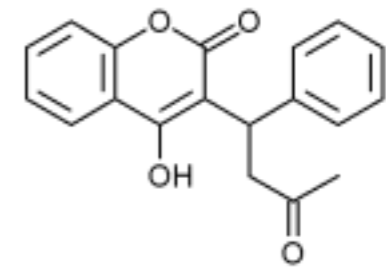
# Genetic variation contributes to drug response

- **Over 10 million common variants in the human genome**
  - ~0.1% difference between two humans
  - ~1.5% difference between a human and a chimp
- **Three widely prescribed drugs now have FDA 'black box' labels advising genetic testing**

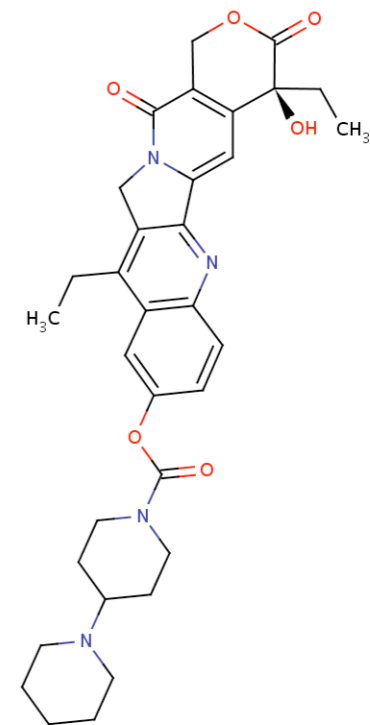


# Commonly prescribed drugs where genetics affect response

- **Warfarin:** anticoagulant prescribed to about two million patients in the US annually.
  - Narrow therapeutic range, dosage needs to be carefully monitored
  - Variants in CYP2C9 linked to differential response
- **Irinotecan:** cancer therapy
  - Membrane-bound UGT1A1 affects drug toxicity



Warfarin



Irinotecan

Higashi MK, Veenstra DL, Kondo LM, Wittkowsky AK, Srinouanprachanh SL, Farin FM, Rettie AE.

Association between CYP2C9 genetic variants and anticoagulation-related outcomes during warfarin therapy.

JAMA. 2002 Apr 3;287(13):1690-8.

Fujita K, Sasaki Y.

Pharmacogenomics in drug-metabolizing enzymes catalyzing anticancer drugs for personalized cancer chemotherapy.

Curr Drug Metab. 2007 Aug;8(6):554-62. Review.

# **How genetic variation in membrane transporters contributes to variable drug response**

# How genetic variation in membrane transporters contributes to variable drug response

- What is the **genetic variation** in the genes encoding membrane transporters in ethnically diverse human populations?

# How genetic variation in membrane transporters contributes to variable drug response

- What is the **genetic variation** in the genes encoding membrane transporters in ethnically diverse human populations?
- What is the **functional significance** of the variant transporters in heterologous expression systems or model organisms?



# How genetic variation in membrane transporters contributes to variable drug response

- What is the **genetic variation** in the genes encoding membrane transporters in ethnically diverse human populations?
- What is the **functional significance** of the variant transporters in heterologous expression systems or model organisms?
- What does the **structure** and **evolutionary history** of membrane transporters tell us about **function**?

# How genetic variation in membrane transporters contributes to variable drug response

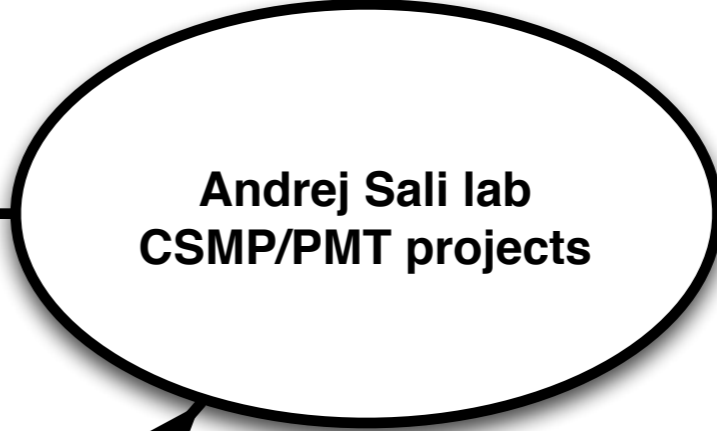
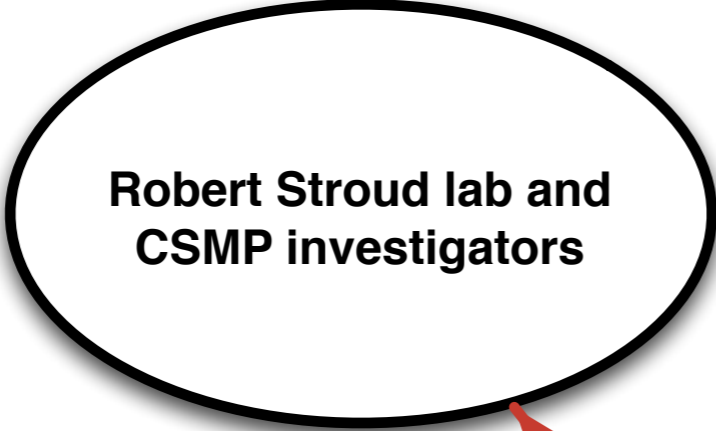
- What is the **genetic variation** in the genes encoding membrane transporters in ethnically diverse human populations?
- What is the **functional significance** of the variant transporters in heterologous expression systems or model organisms?
- What does the **structure** and **evolutionary history** of membrane transporters tell us about **function**?
- How does genetic variation in membrane transporters affect **clinical drug response**?

# Our collaborators let us tackle this problem from multiple perspectives

Structure determination

Modeling structure and predicting function

Franklin Hayes  
Franz Gruswitz

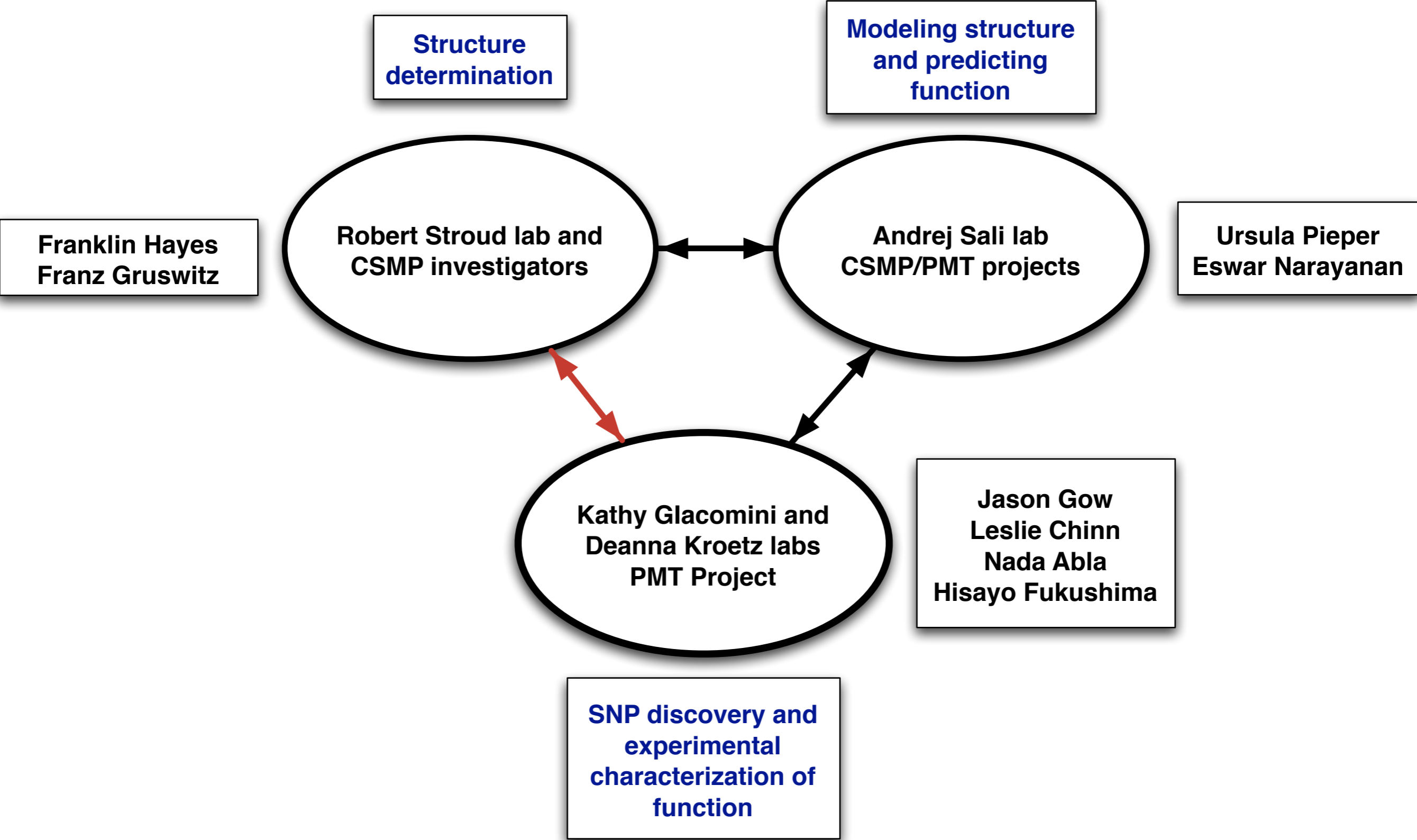


Ursula Pieper  
Eswar Narayanan



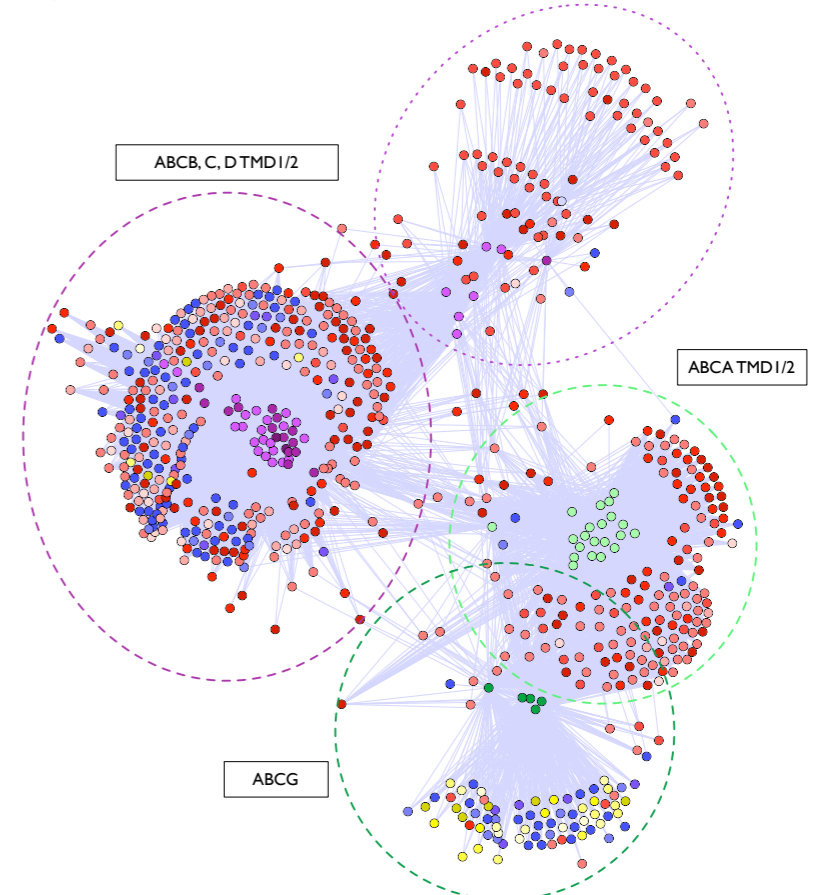
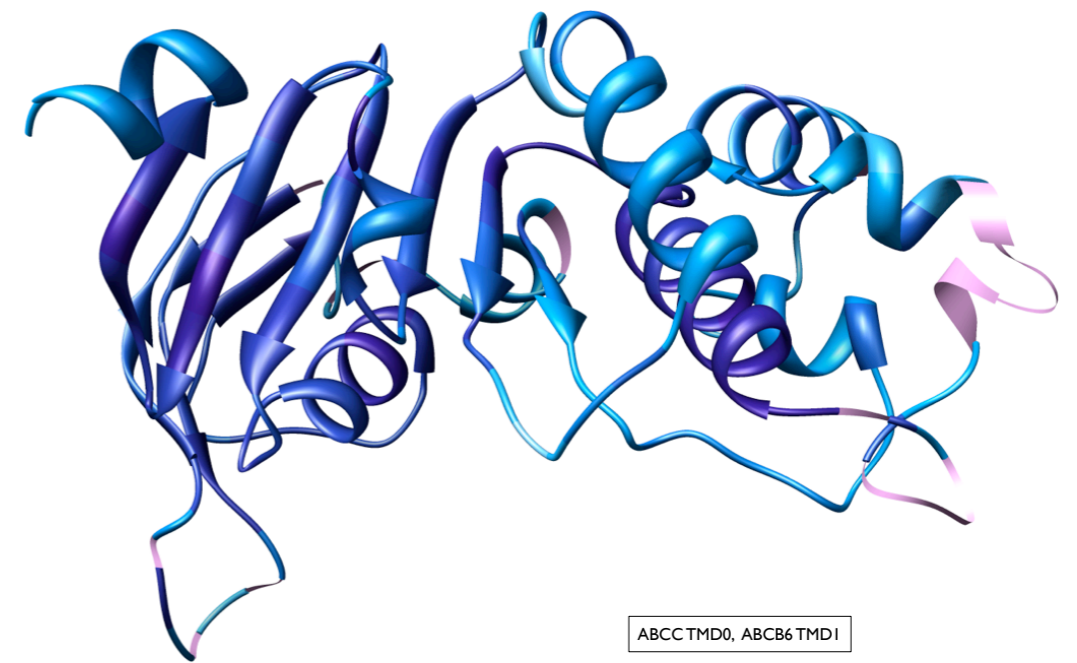
Jason Gow  
Leslie Chinn  
Nada Aba  
Hisayo Fukushima

SNP discovery and experimental characterization of function

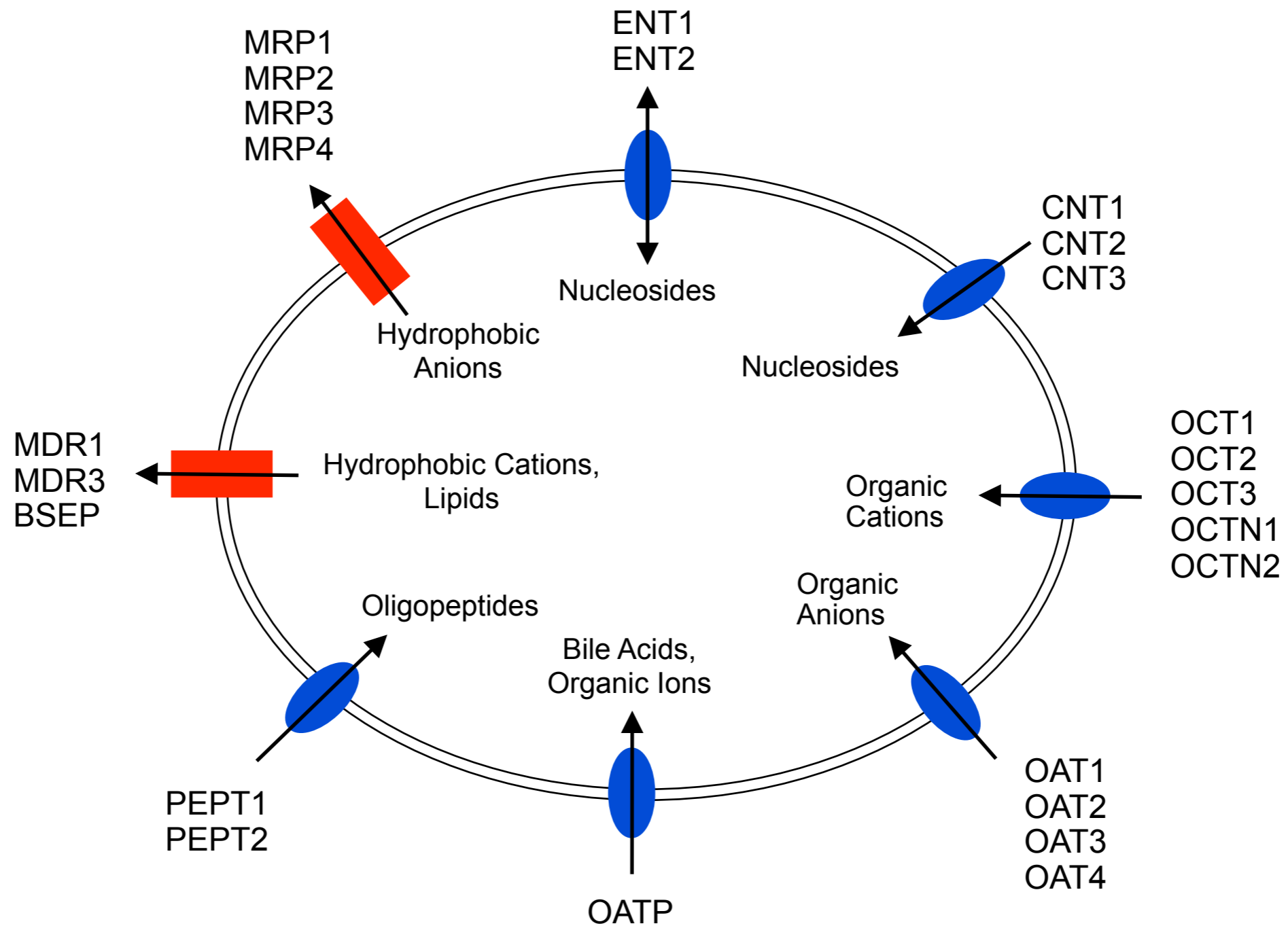


# Using structure and evolutionary analysis to functionally characterize membrane proteins

- **The impact of human genetic variation on membrane transporters**
  - Domain interfaces and disease-associated mutants in ABC transporters
  - Predicting the effects of point mutations
- **Membrane protein evolution**
  - Taxonomy of membrane protein families
  - Domain organization and evolution of ABC transporters
  - Target selection for structural characterization of yeast membrane proteins

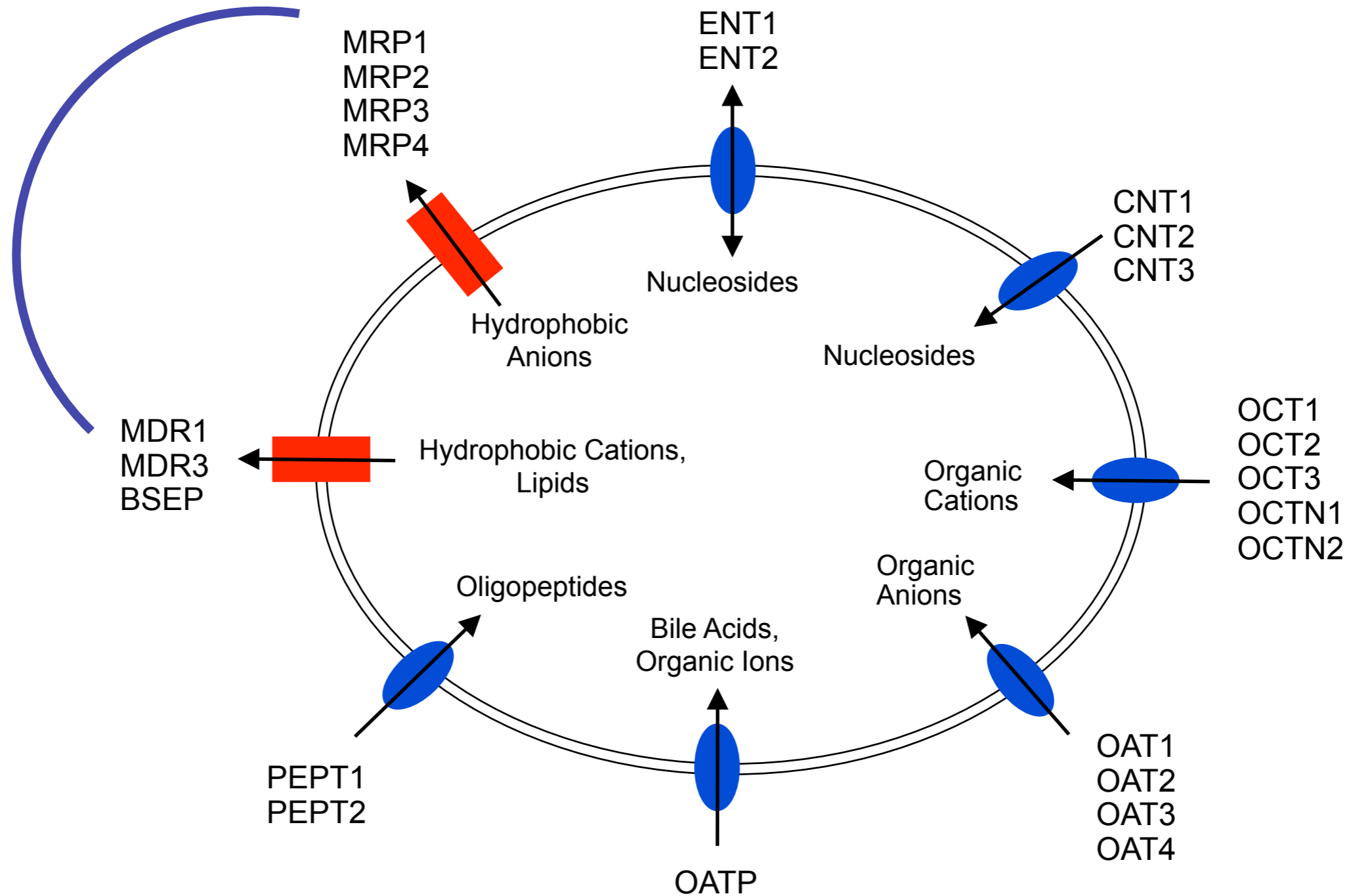


# Cataloguing variation in membrane transporters



# Cataloguing variation in membrane transporters

## ABC transporters



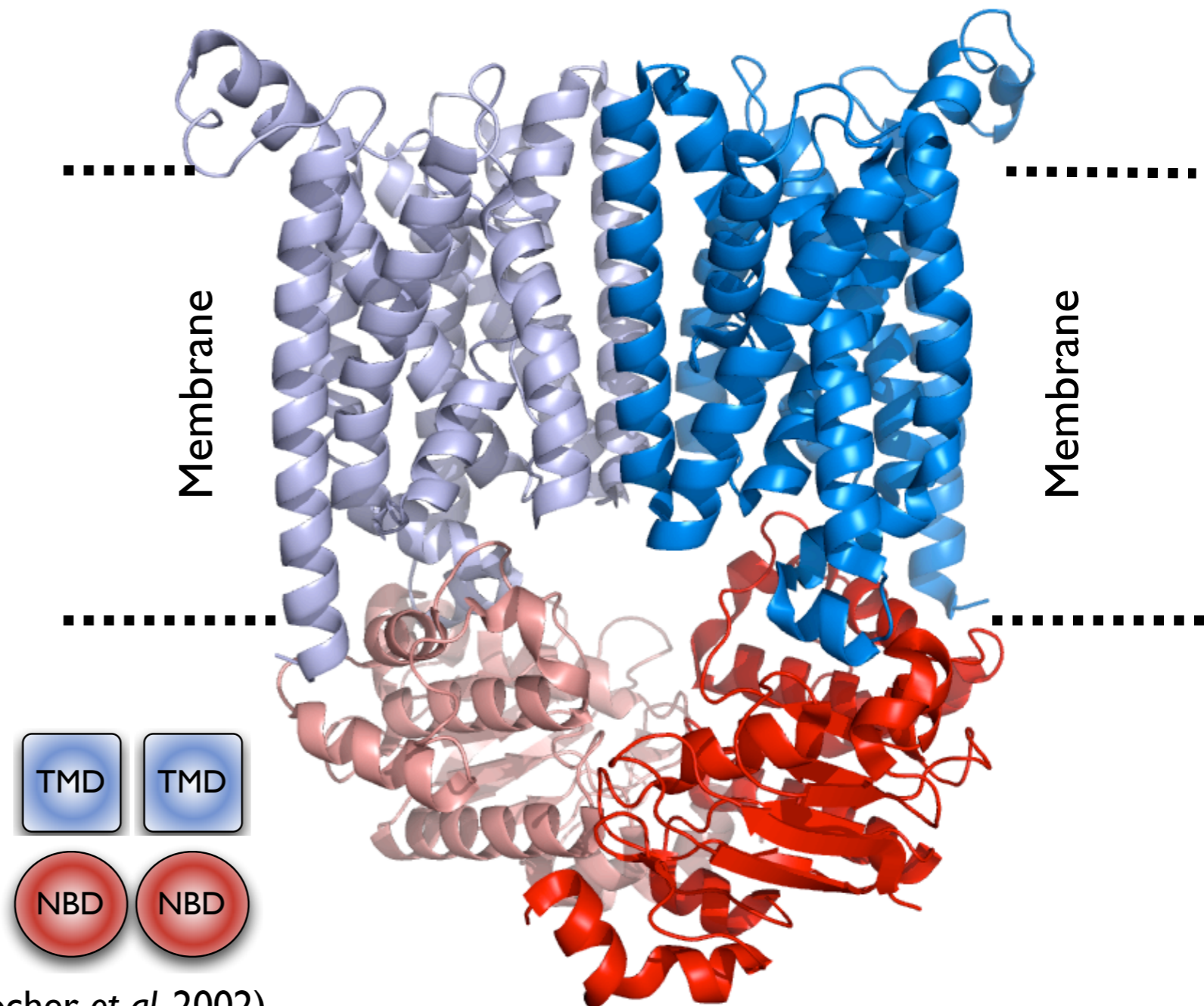
# ABC transporters play roles in disease and drug response

<b>Human ABC transporter</b>	<b>Function</b>	<b>Disease</b>
ABCC7 (CFTR)	Chloride ion transporter	Cystic fibrosis
ABCD1 (ALD)	Likely a very long chain fatty acid transporter	Adrenoleukodystrophy
ABCA4 (ABCR)	Retinoids	Retinitis pigmentosa, AMD STGD
ABCC2	Organic anions, multidrug resistance-associated	Dubin-Johnson syndrome

- **>1,000 clinically characterized disease-associated point mutants in human ABC transporters**
- **Functional analysis of mutations is lacking**

# ABC transporters are membrane proteins that bind a wide range of substrates

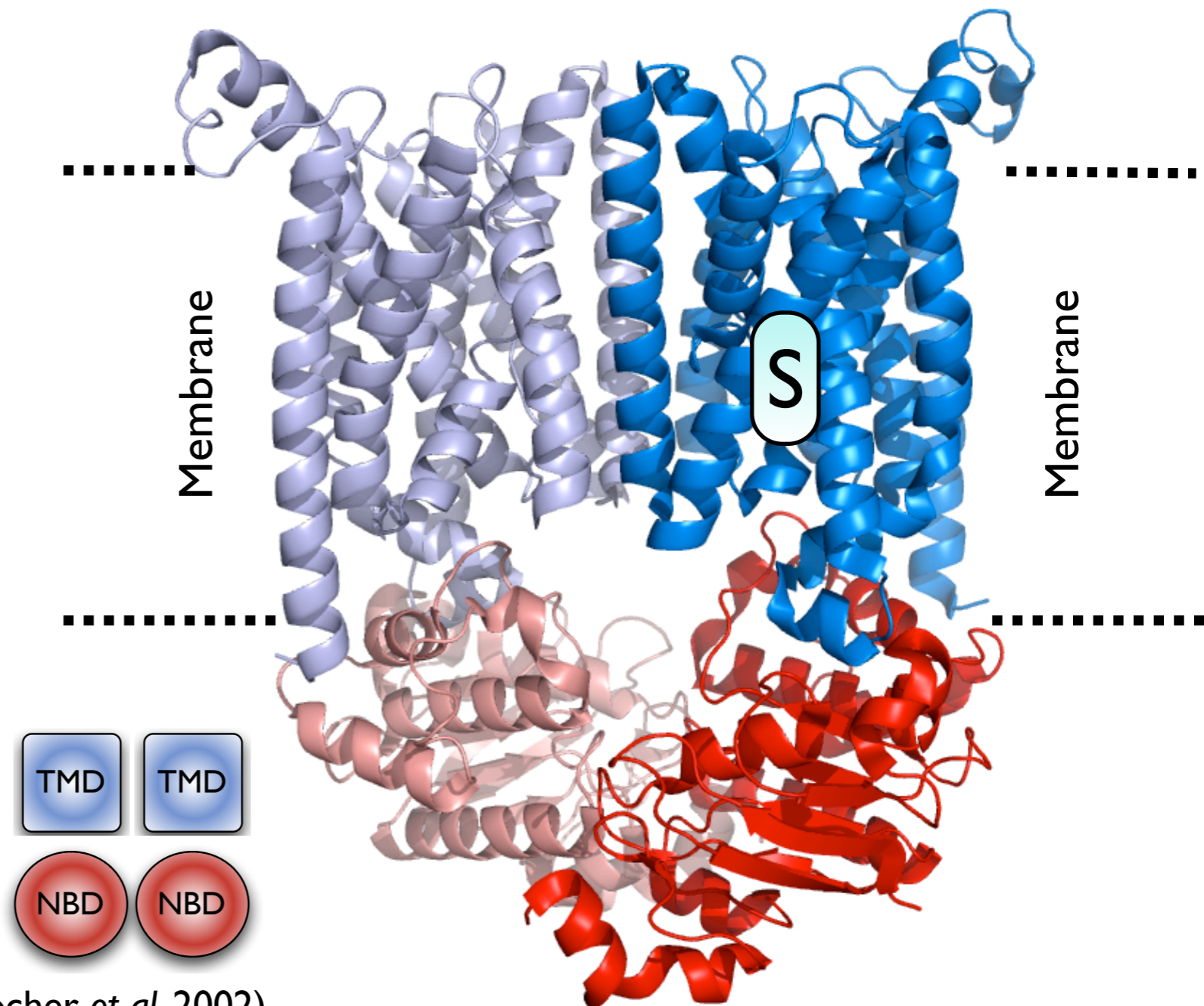
- In humans, ABC transporters are active export pumps
- Substrate binds in transmembrane domains (TMDs), ATP binds in nucleotide binding domains (NBDs)
- ATP binding, hydrolysis and release are coupled to substrate transport





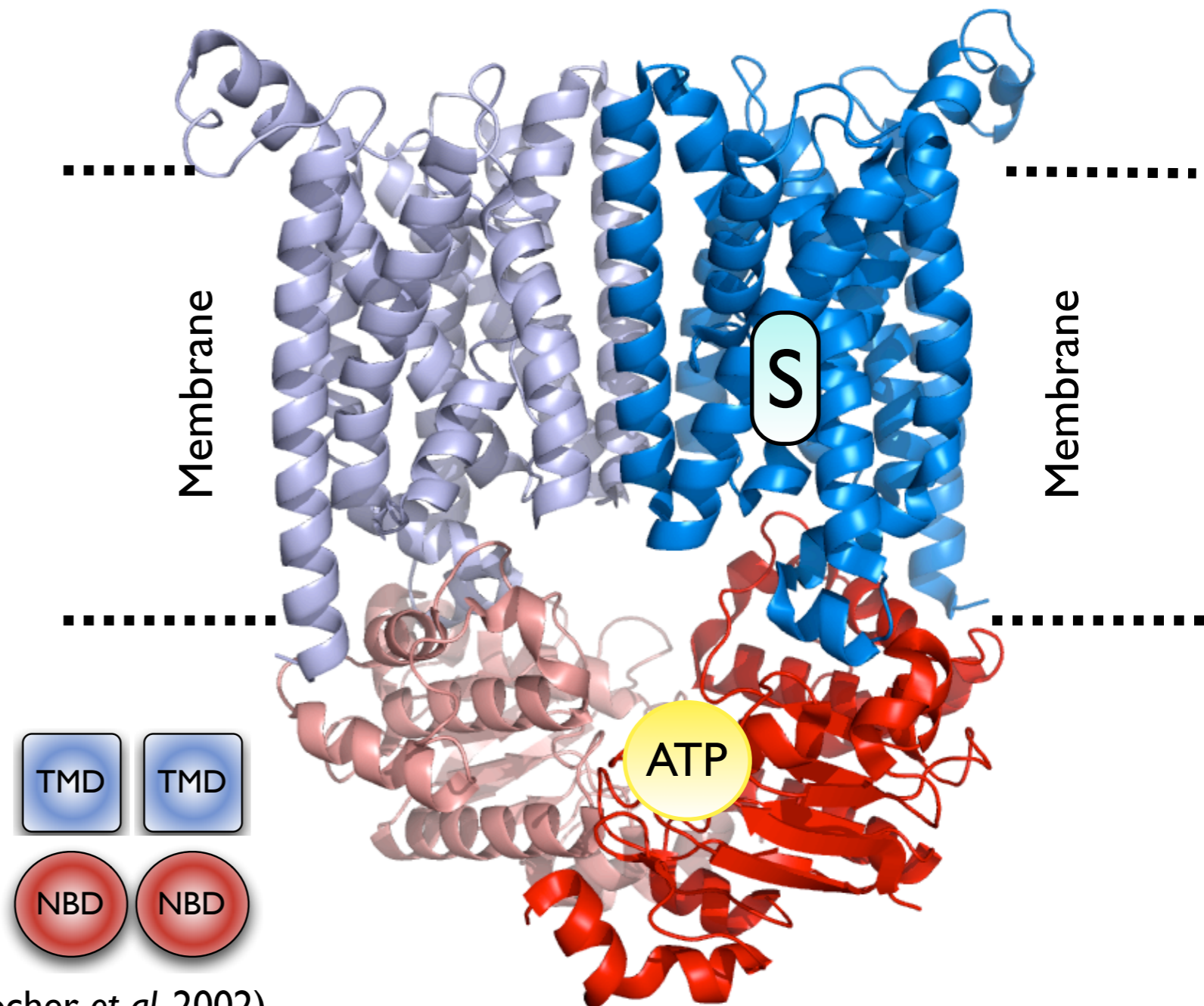
# ABC transporters are membrane proteins that bind a wide range of substrates

- In humans, ABC transporters are active export pumps
- Substrate binds in transmembrane domains (TMDs), ATP binds in nucleotide binding domains (NBDs)
- ATP binding, hydrolysis and release are coupled to substrate transport



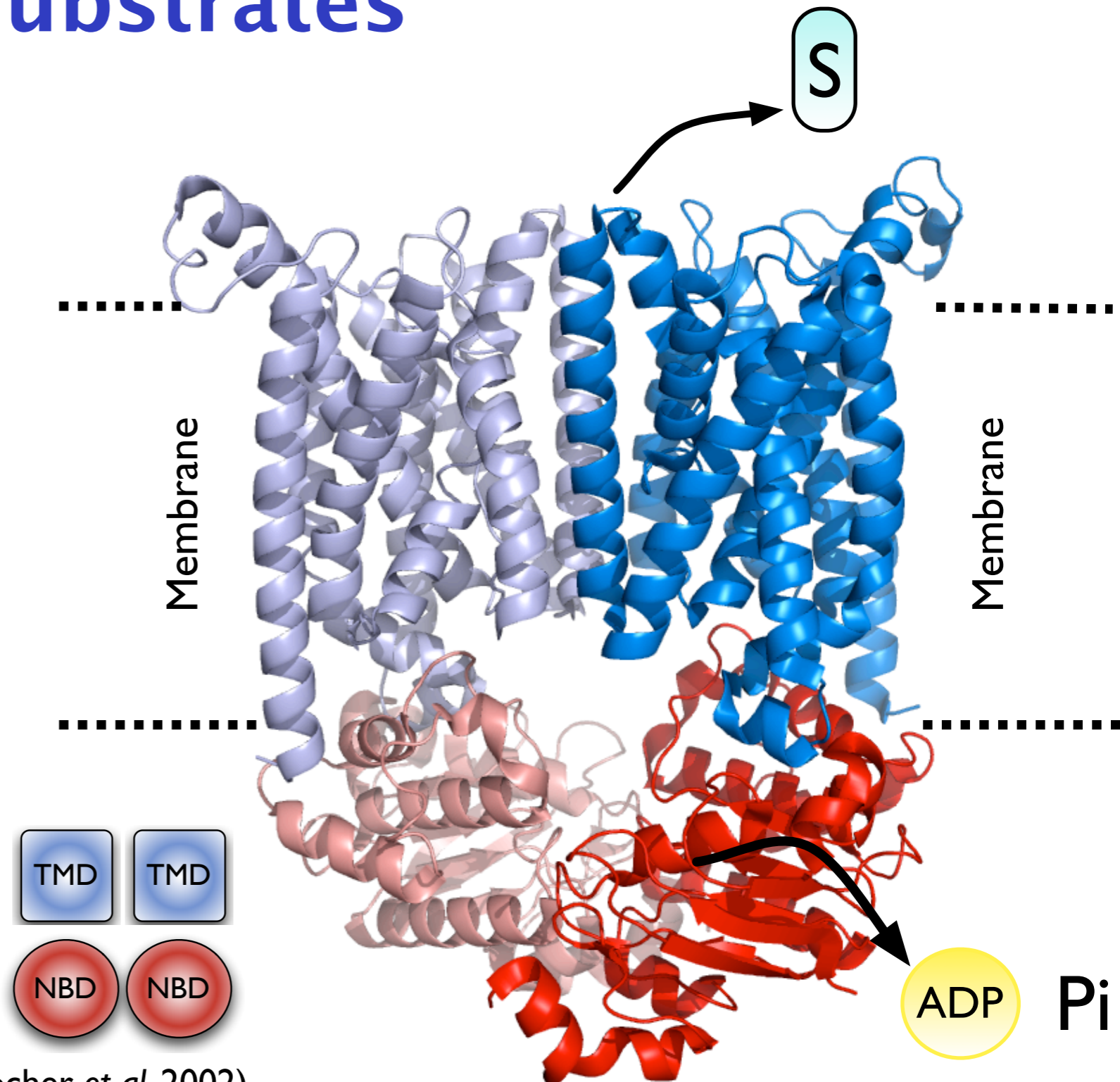
# ABC transporters are membrane proteins that bind a wide range of substrates

- In humans, ABC transporters are active export pumps
- Substrate binds in transmembrane domains (TMDs), ATP binds in nucleotide binding domains (NBDs)
- ATP binding, hydrolysis and release are coupled to substrate transport

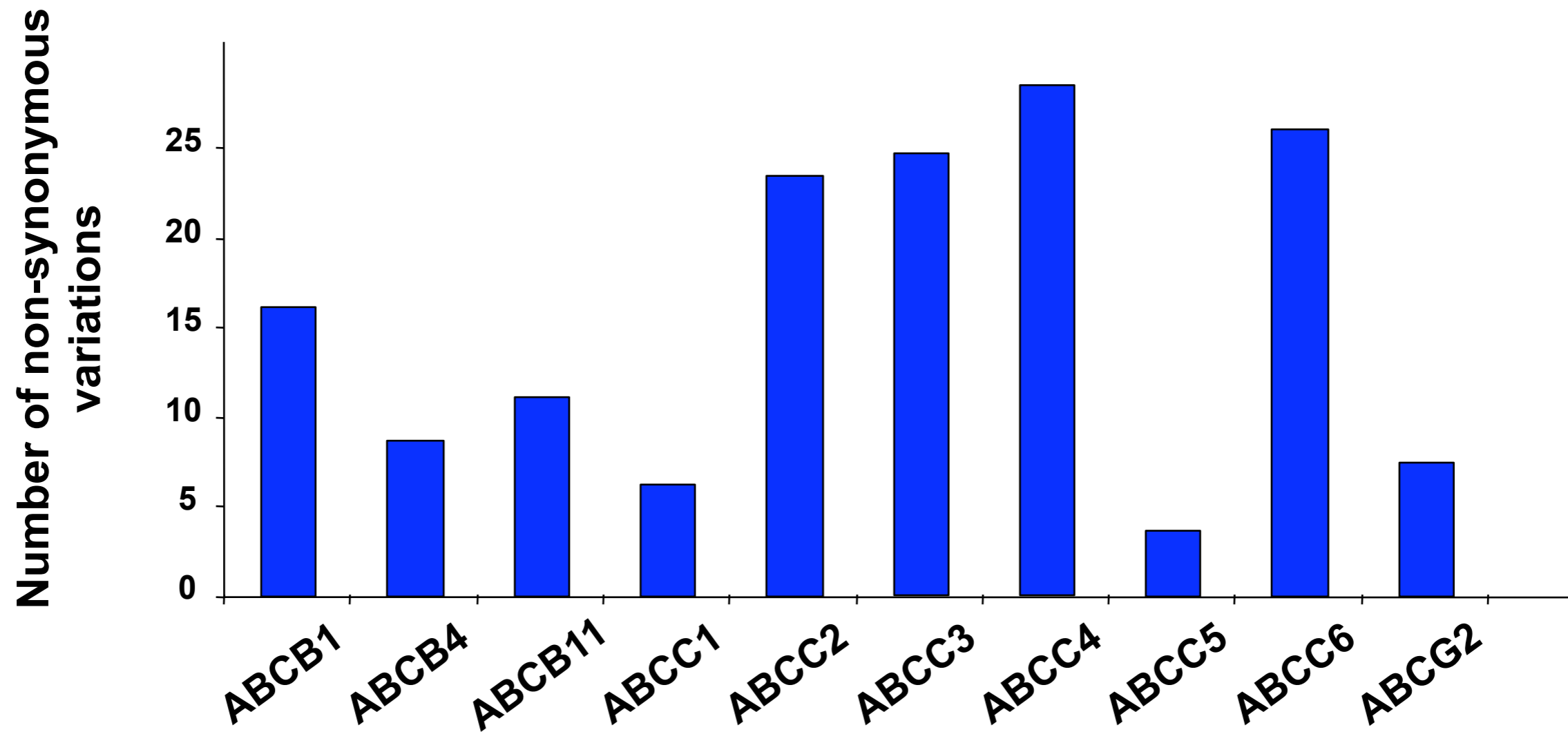


# ABC transporters are membrane proteins that bind a wide range of substrates

- In humans, ABC transporters are active export pumps
- Substrate binds in transmembrane domains (TMDs), ATP binds in nucleotide binding domains (NBDs)
- ATP binding, hydrolysis and release are coupled to substrate transport



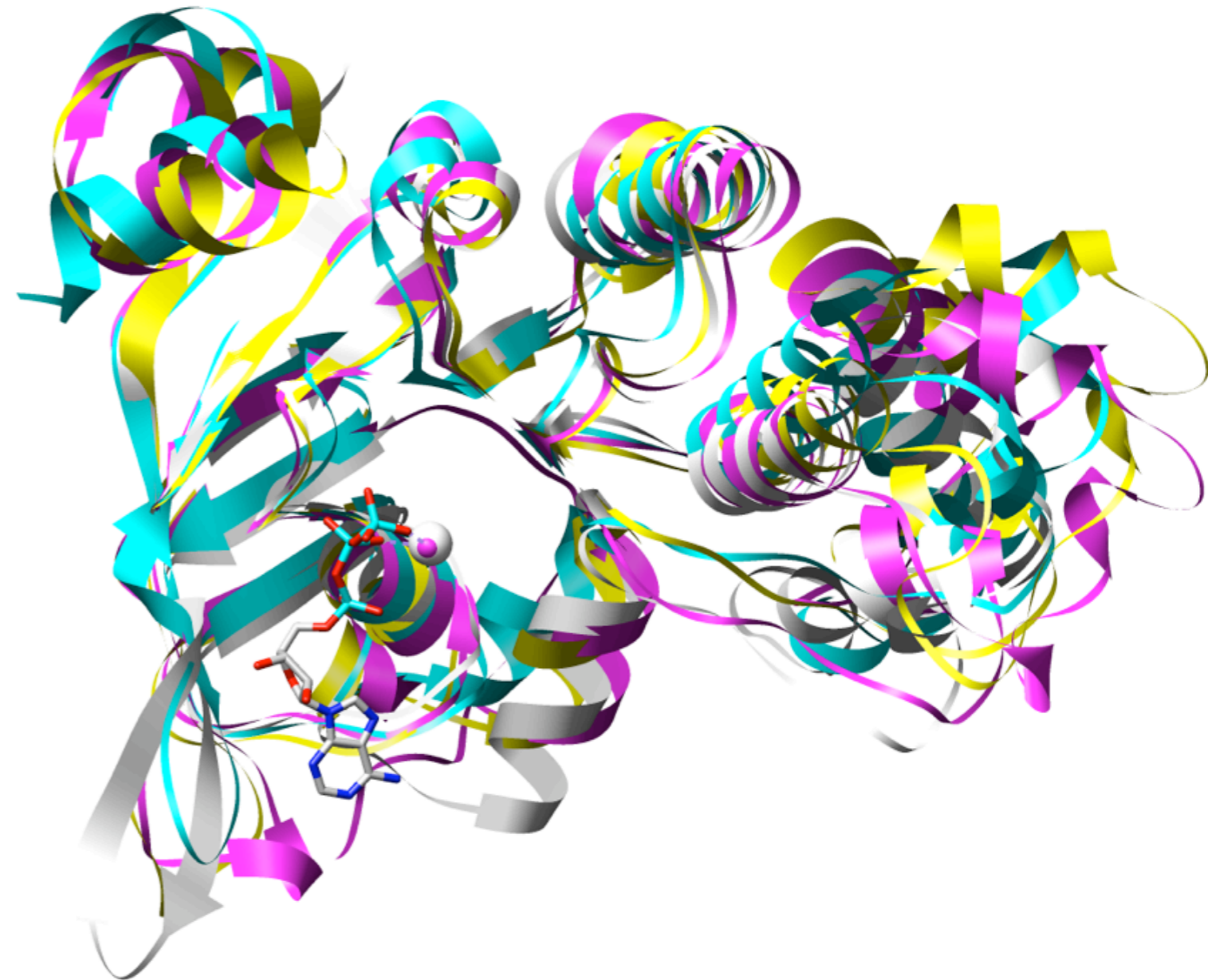
# There is a lot of non-synonymous variation in ABC transporters



# Computational methods allow rapid examination of large numbers of mutations

- **You can't experimentally characterize all SNPs in proteins of interest at all levels you are interested in**
- **Some proteins and systems are difficult to get functional data for; data for 10s of SNPs but not 100s or more**
- **Computational analyses of point mutations can be used to:**
  - **predict the effects of large numbers of clinically relevant mutants**
  - **suggesting a smaller set of experiments**

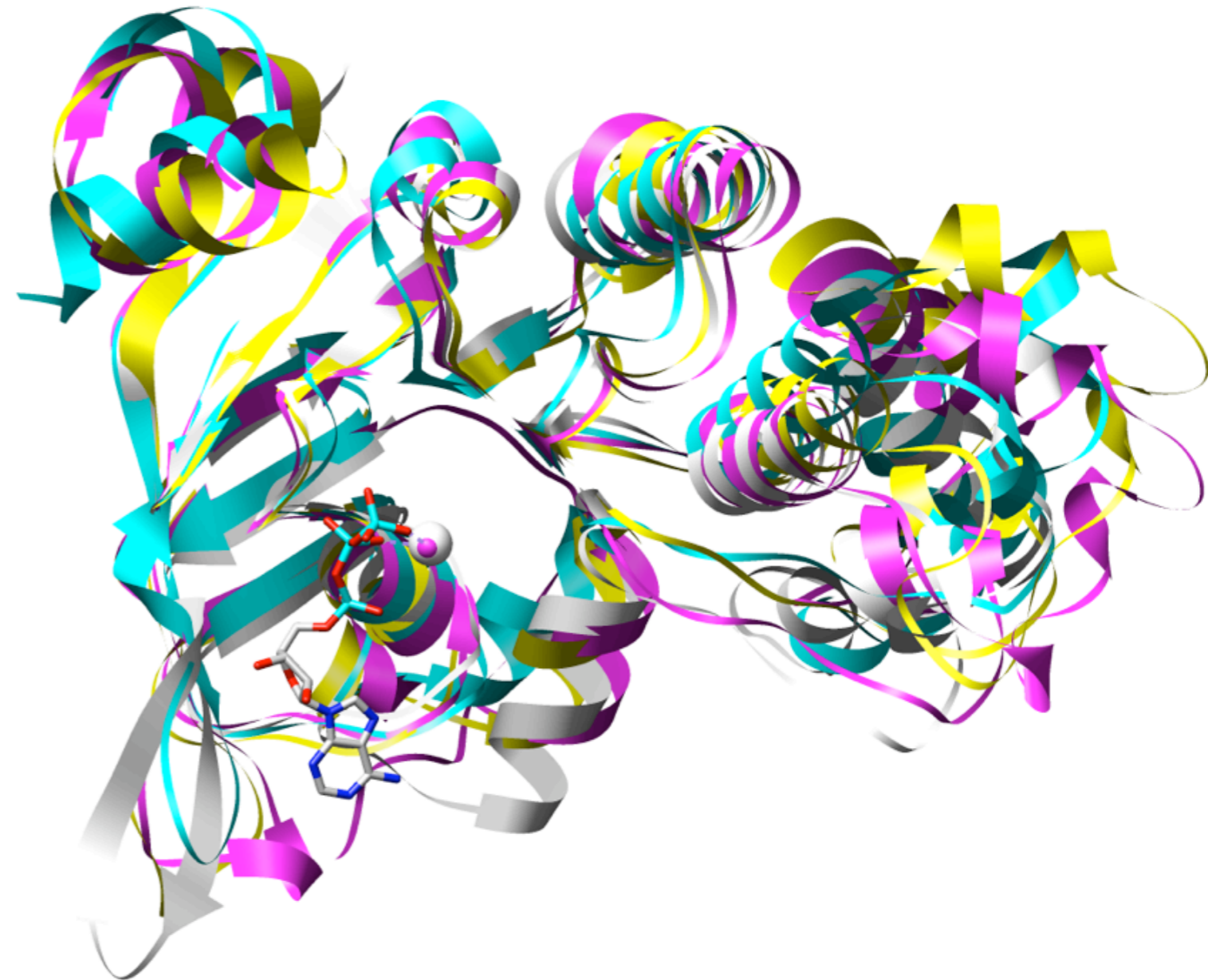
# Using structures and models to examine variation and disease in ABC transporters



structures of four ABC NBDs

# Using structures and models to examine variation and disease in ABC transporters

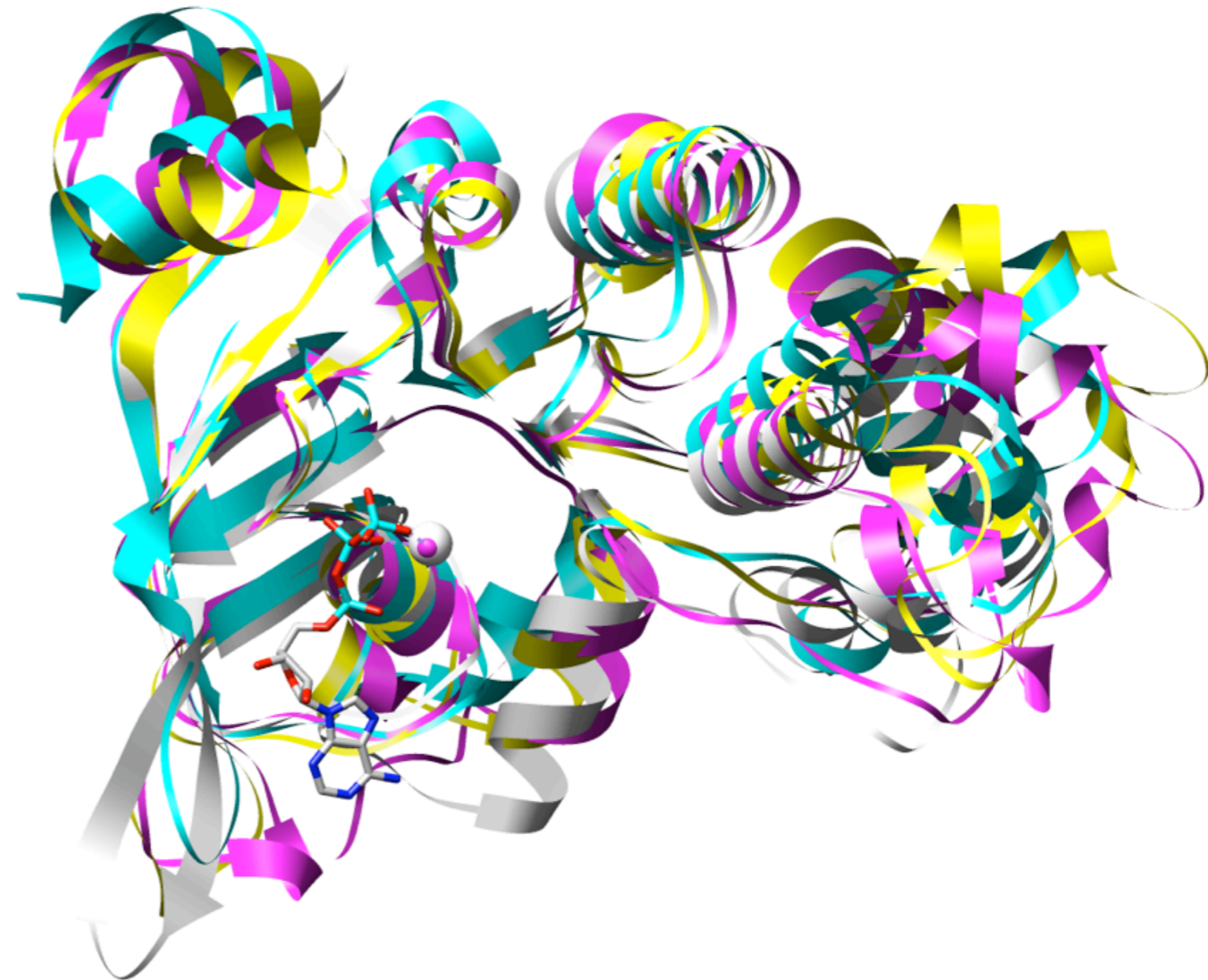
- The overall fold of the NBDs is highly conserved across organisms



structures of four ABC NBDs

# Using structures and models to examine variation and disease in ABC transporters

- The overall fold of the NBDs is highly conserved across organisms
- This enables us to model human NBDs and nsSNPs based on homologs with known structure

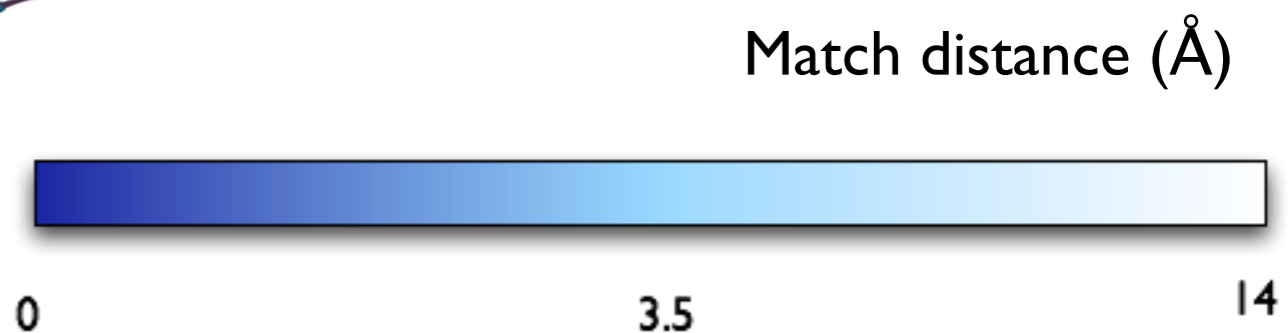
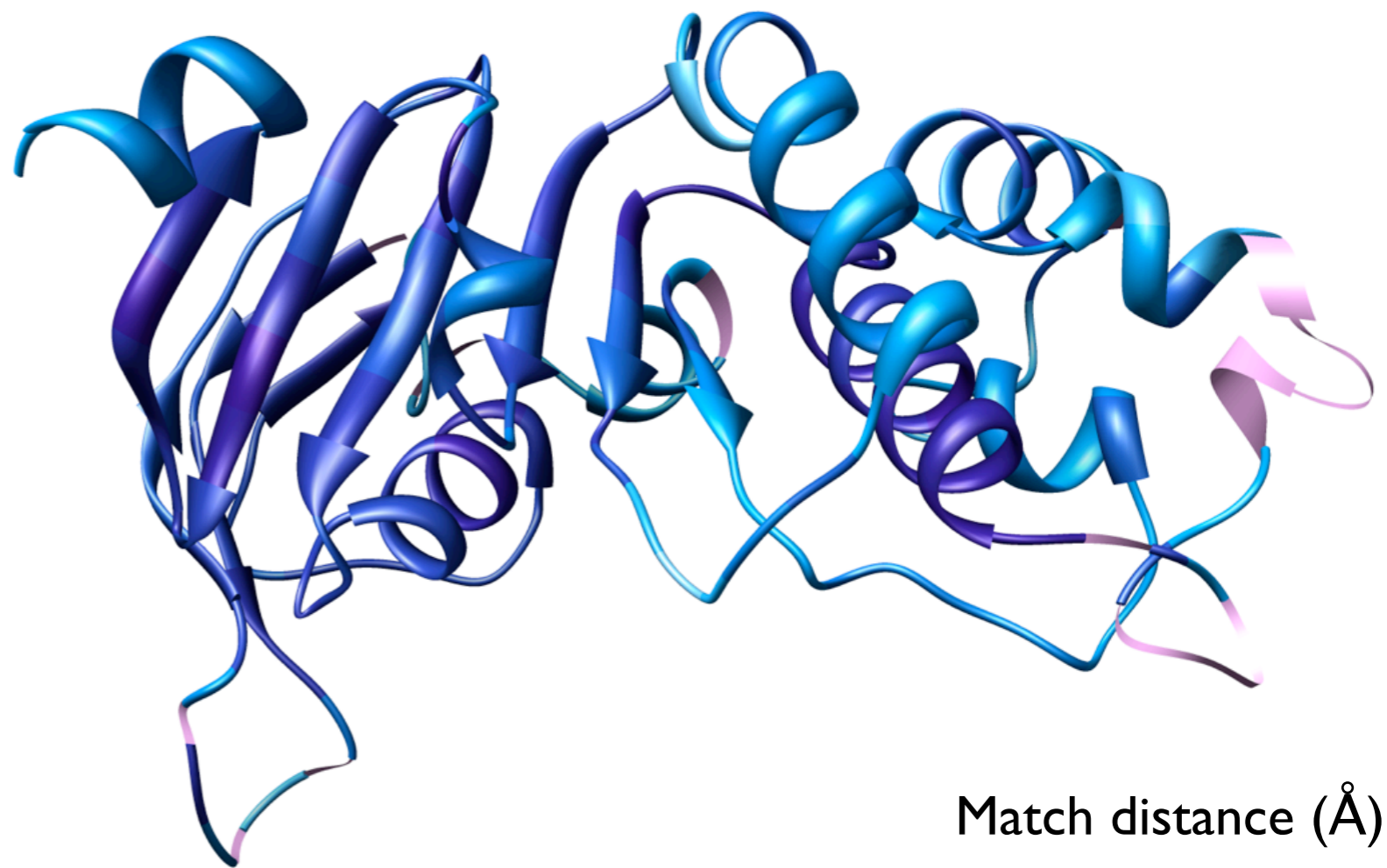


structures of four ABC NBDs



# Using structures and models to examine variation and disease in ABC transporters

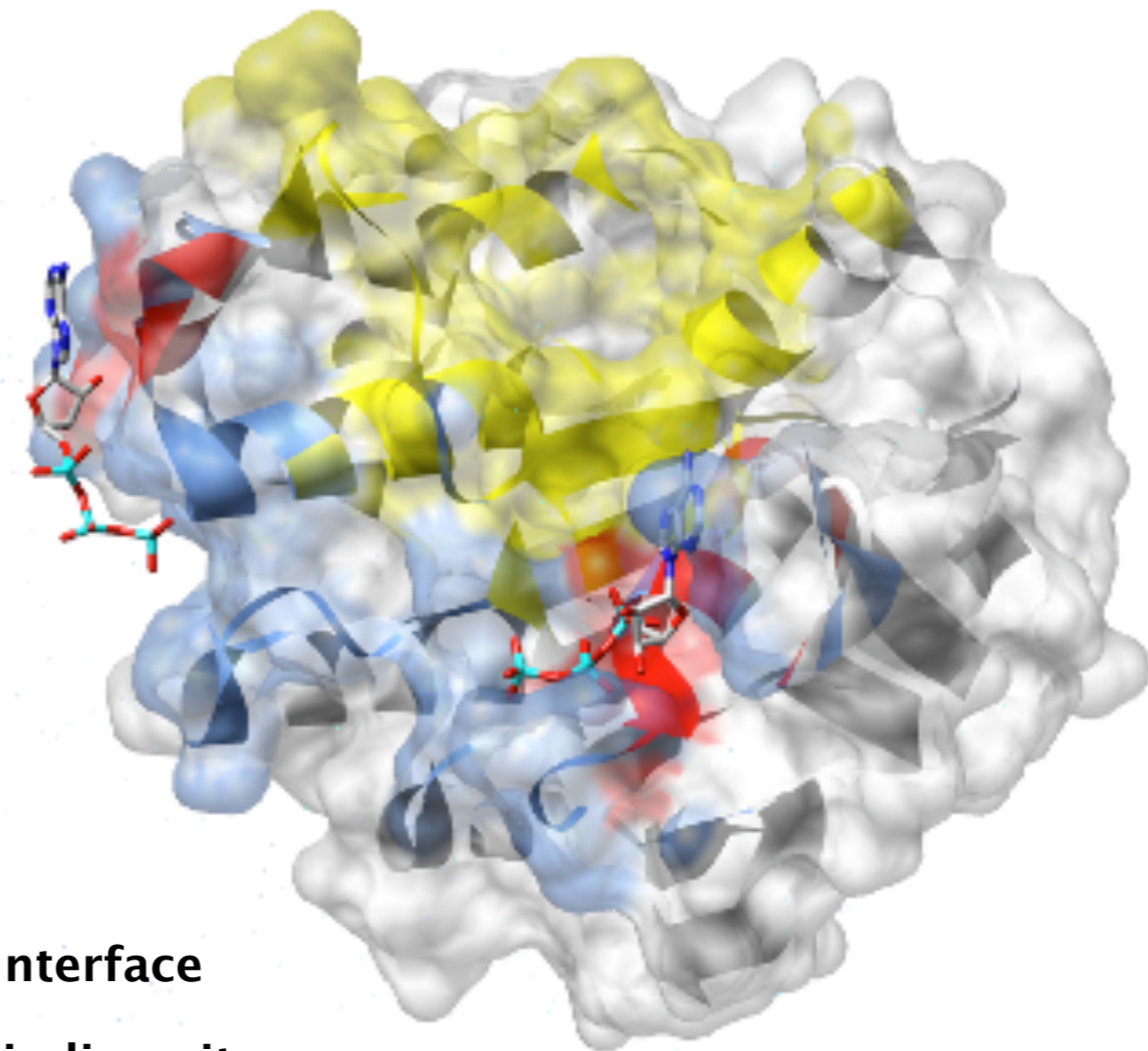
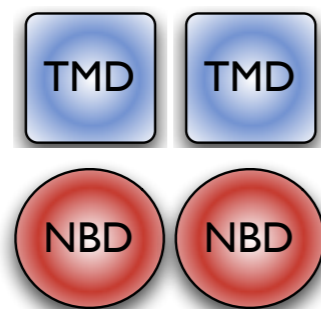
- The overall fold of the NBDs is highly conserved across organisms
- This enables us to model human NBDs and nsSNPs based on homologs with known structure



model of NBD1, human MDR1

# Examining genetic variation at the interfaces in ABC transporters

- Does residue level conservation vary at domain interfaces?
- Might surface conservation suggest unidentified interaction sites?
- Could some disease-associated mutants be affecting domain interactions?



from *M. jannaschii*

# Does sequence conservation vary at ABC transporter domain interfaces?

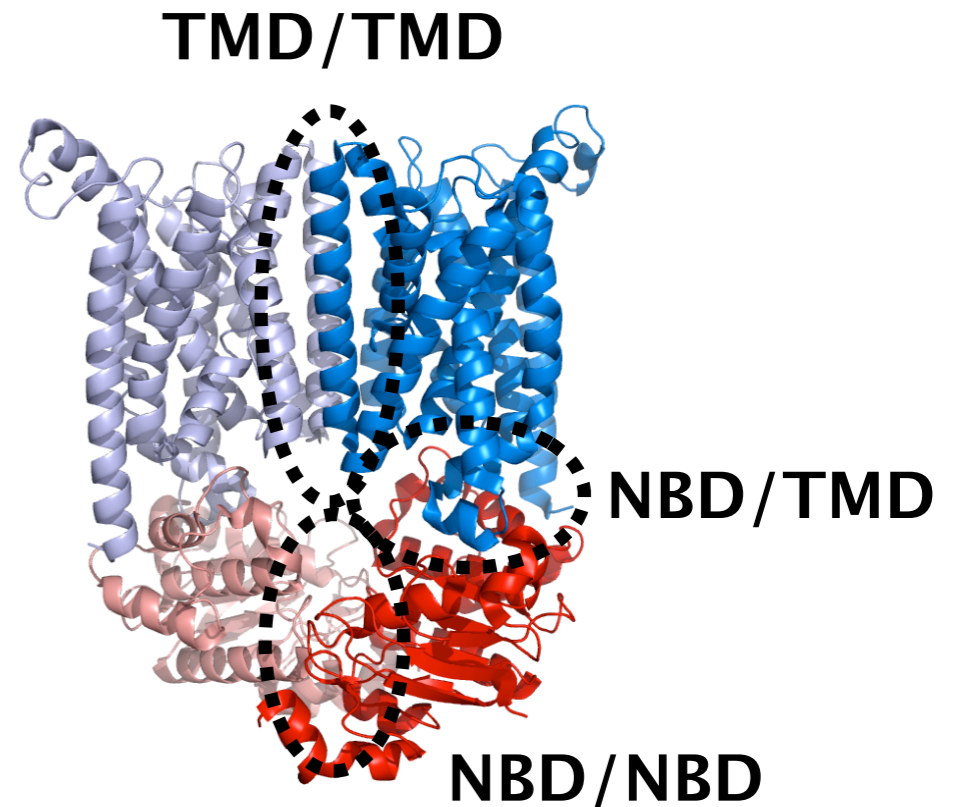
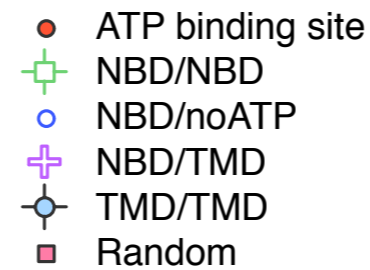
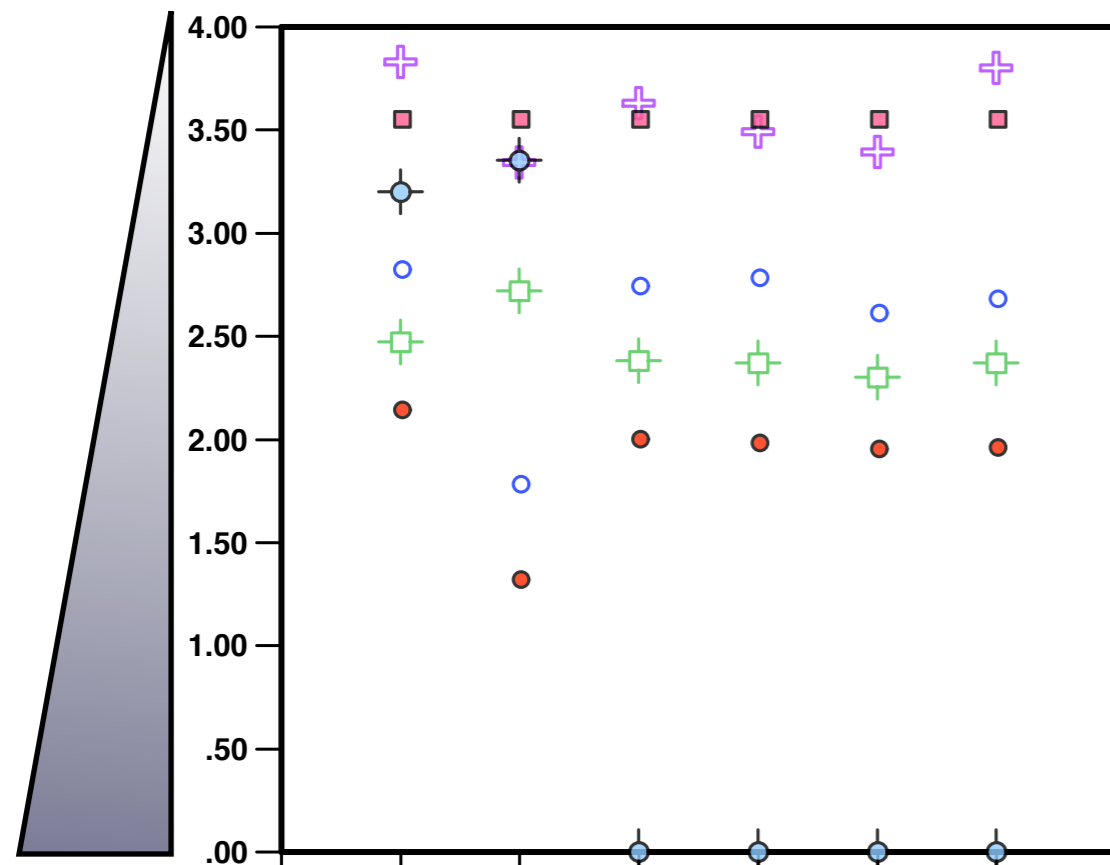
- Automatic multiple sequence alignment profiles for each of six structures and for the NBDs of each human ABC transporter
- We calculate a measure of entropy in the alignment

$$H = - \sum_{aa=1}^{20} P_{aa} \log_2 P_{aa}$$

1L2T:A PDBID CHAIN SEQUENCE/1-235	NVNLNIKÉGEFVS	IN	GP	SG	SGK	STMLNI	IGCLDKPTEGI
ABCG2_HUMAN/1-168	NINGIMKPG-LNATI	U	GPT	GGGKSSLLDVLAARKDPSSGI			
ABCX_CYACA/1-175	NINLQIKTNETHV	IM	GN	SGKSSLLKVIAGHPKVI	EGF		
ABCE1_HUMAN/1-176	IVAGEFTDSEIMVM	L	GENGT	GKTTFIRMLAGRLK	PDEGI		
ADCC_STRPN/1-185	HINYCVDSGEFVTL	T	GENGA	AKTTLIKASLGILQPR	IGF		
ARTP_HAEIN/1-213	DINLEAEEGDTVVL	L	GP	SGAGKSTLIRTLNLL	LEVPKSGI		
ABCX_PORPU/1-178	GVNLSIKPGEIHA	IM	GN	SGKSTLSKVIA--GHPANGI			
ABCBB_HUMAN/1-207	DLNMVIKPGEMTAL	V	GP	SGAGKSTALQLIQRFYDPCEGM			
ABCD1_MOUSE/1-183	--NIRVEEGMHLL	IT	GN	CGKSSLFRILGGLWPTYSGI			
ALSA_ECOLI/1-195	SVNLTVPGEIHAL	L	GENG	AGKSTLMKVLSGIHEPTKGI			

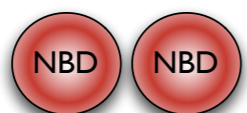
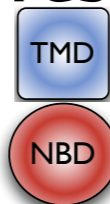

# Sequence conservation varies between the three interfaces

less conserved

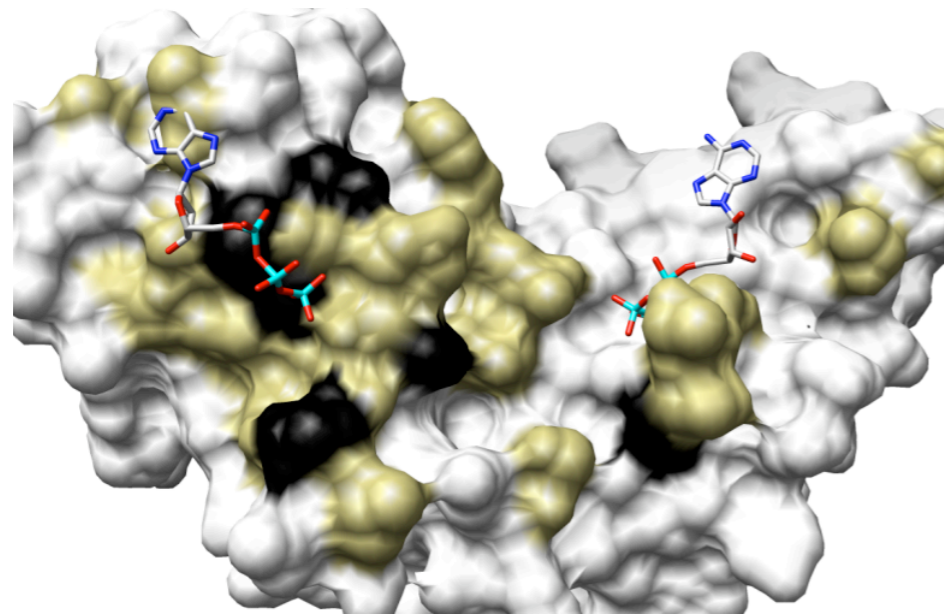




more conserved

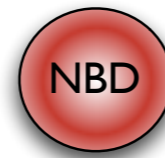
Structure

-  interface was moderately conserved even when ATP binding residues were excluded.
- In contrast,  and  interfaces not conserved at all

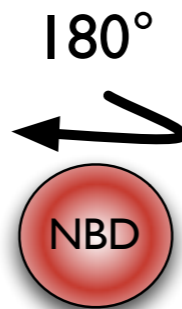
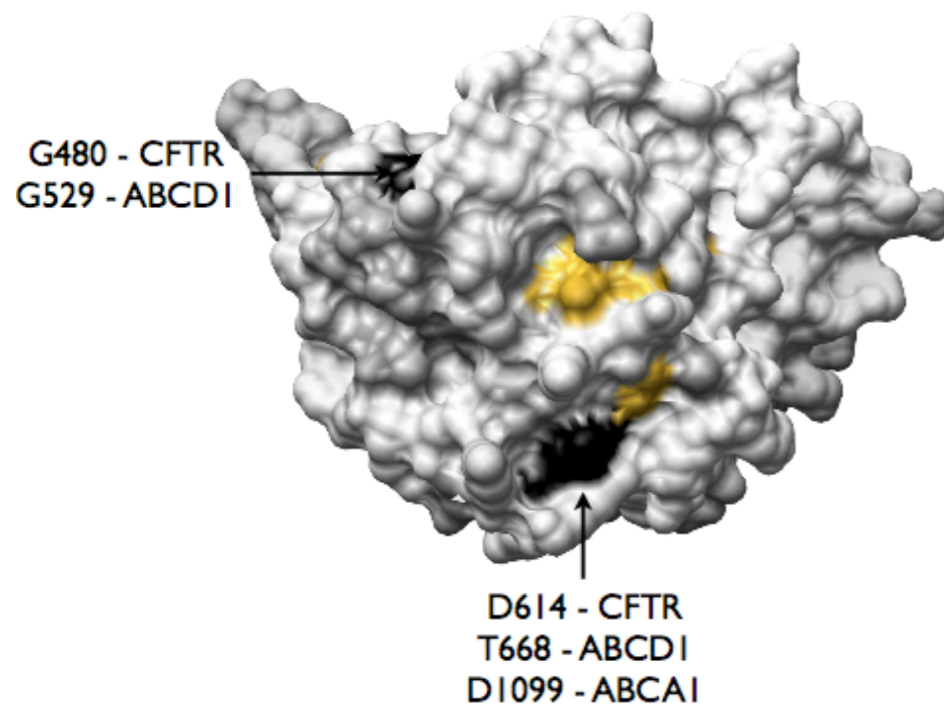
# Some disease-associated mutants affect domain interactions





 conserved  
 disease



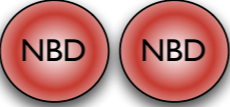
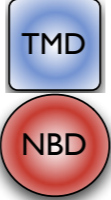
Human cystic fibrosis-associated transproter CFTR NBD1 interface surface



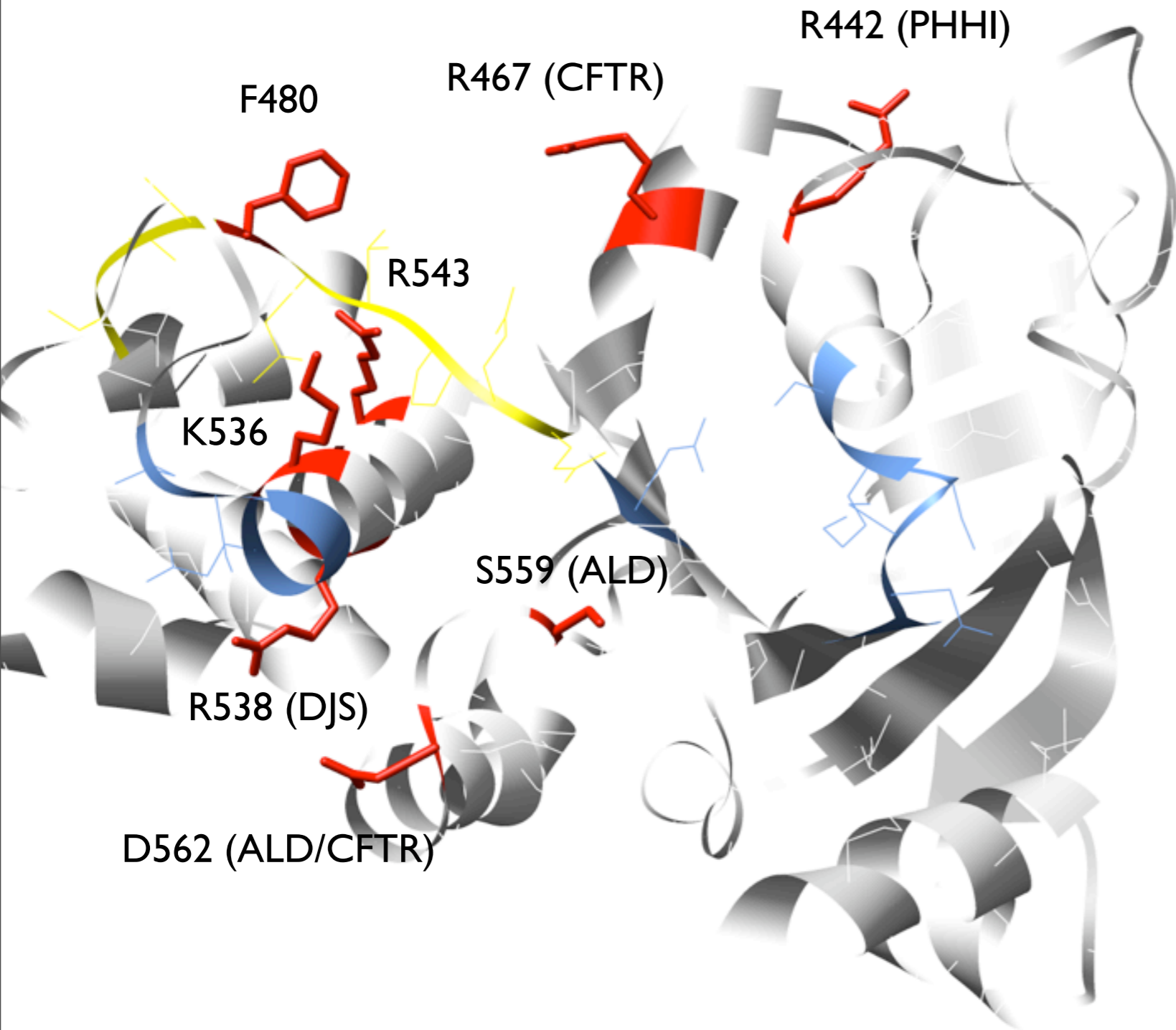
 conserved  
 disease

Exposed surface of bacterial ABC transporter NBD

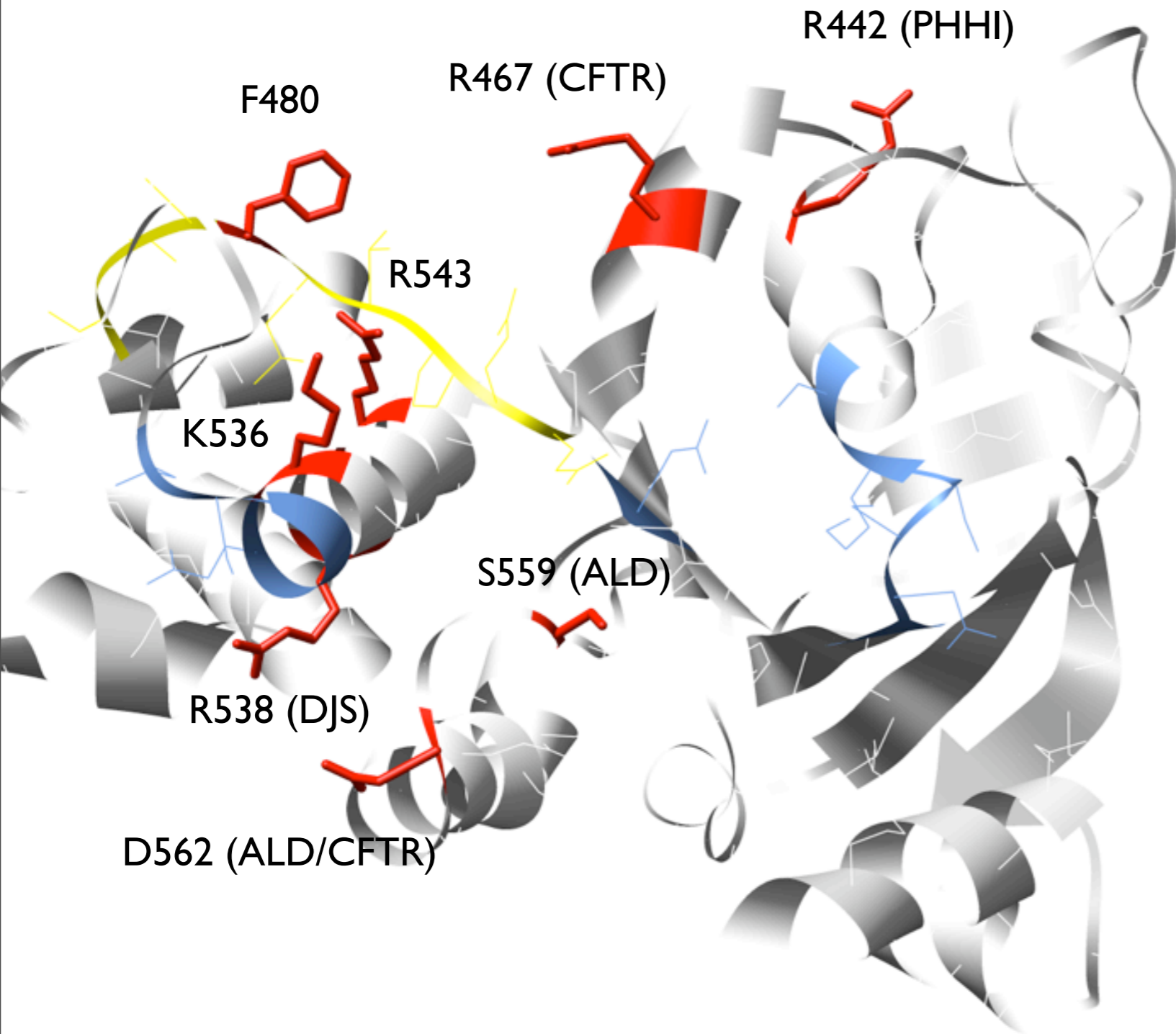
# We found 68 disease-associated positions at putative interfaces

- 10 transporters from four out of seven ABC subfamilies are represented
- 38 were at the  interface
- 30 were at the  interface
- I am working on characterizing experimentally characterizing analogous interface residues in the human ABC transporter MRP4 to examine the functional effects of point mutants at the TMD/NBD interface

# Do disease-associated mutations hint at common mechanisms?



# Do disease-associated mutations hint at common mechanisms?



**Are transporters across the superfamily similar enough that disease-associated mutations in one family are predictive of disease association in other members?**



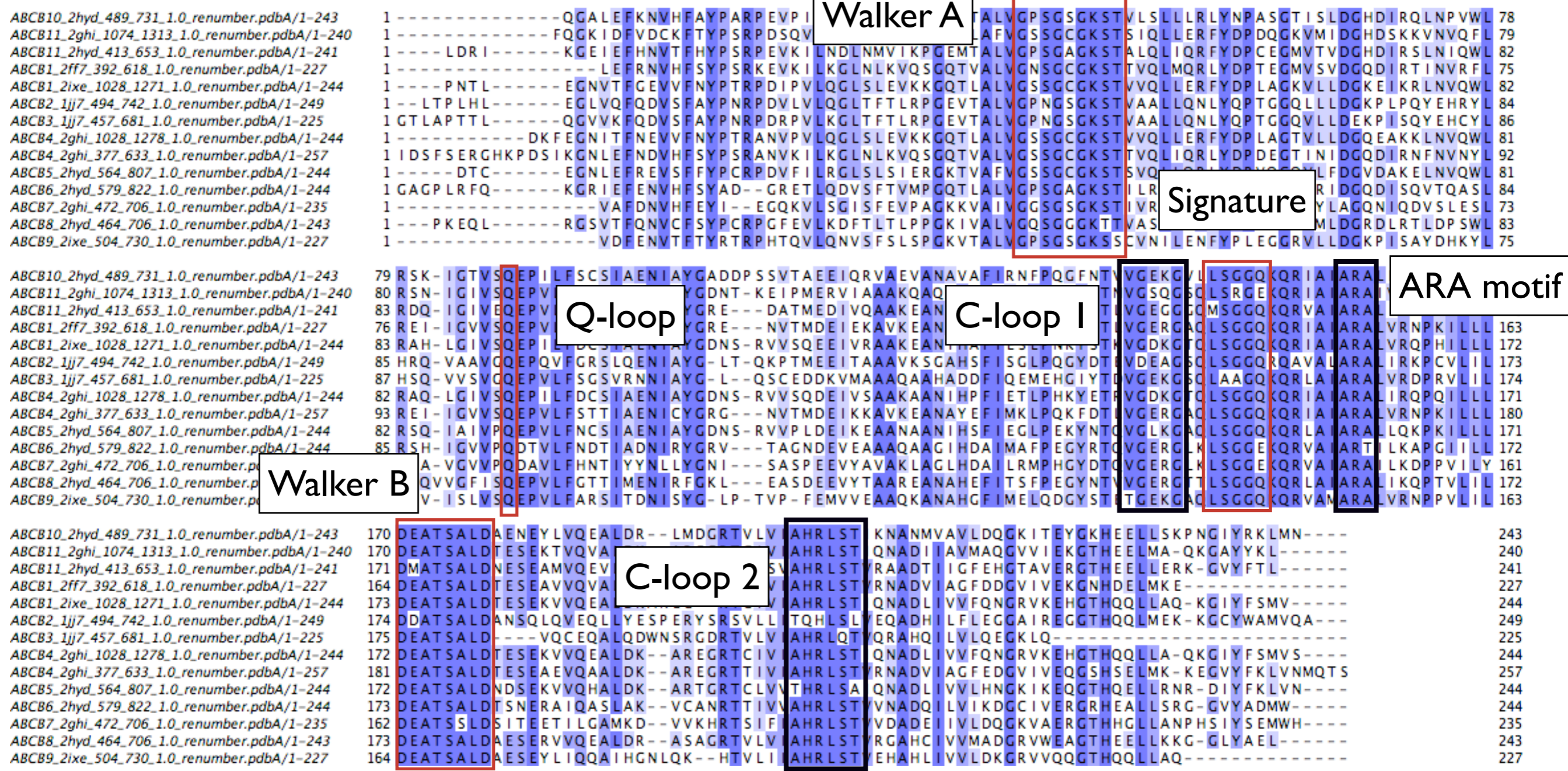
# Common conservation patterns across all human NBDs suggest functional residues

```
ABC B10_2hyd_489_731_1.0_renumber.pdbA/1-243 1-----QGALEFKNVHFAYPARPEVP I FQDFSLSPSGSVTALVGP SGSGKSTVLS LLLRLYNPASGT I SLDGHD I RQLNP VWL 78
ABC B11_2ghi_1074_1313_1.0_renumber.pdbA/1-240 1-----FQGKIDFVDCCKFTYPSRPSQV L NGLSVS I SPGQTLAFV GSSGCGKSTSIQLLERFYDPDQ GKVM I DGHDSKKNVQFL 79
ABC B11_2hyd_413_653_1.0_renumber.pdbA/1-241 1----LDR I-----KGEI EFHNVT FHYP SRPEVK I LNDLNMV I KPGEMTALV GP SGAGKSTALQL IQR FYDPCEGMVTV DGHDIRSLNIQWL 82
ABC B1_2ff7_392_618_1.0_renumber.pdbA/1-227 1-----LEFRNVHF SYPSRKEVK I LKGLNLKVVQSGQTVLVGN SGCGKSTTVQLMQRLYDPT EGMVSV DGDQDIRTINVRFL 75
ABC B1_2ixe_1028_1271_1.0_renumber.pdbA/1-244 1----PNTL-----EGNVT FGEVVF NYPTRPDI PV LQGLSLEVKKGQT LALV GSSGCGKSTVVQLLERFYDP LAGKVL LDGKE I KRLNVQWL 82
ABC B2_1jj7_494_742_1.0_renumber.pdbA/1-249 1--LTP LHL-----EGLVQ FQDVS FAYPNRPDV LV LQGLTFTLRPGEVTALV GPNGSGKSTVAALLQNL YQPTGGQLLDGKPLPQYEHRYL 84
ABC B3_1jj7_457_681_1.0_renumber.pdbA/1-225 1GTLAPTTL-----QGVVK FQDVS FAYPNRPDRPV LKGLTFTLRPGEVTALV GPNGSGKSTVAALLQNL YQPTGGQVLLDEKPI SQYEHCYL 86
ABC B4_2ghi_1028_1278_1.0_renumber.pdbA/1-244 1-----DKFEGNIT FNEVVF NYPTRANVPV LQGLSLEVKKGQT LALV GSSGCGKSTVVQLLERFYDP LAGTVLLDGQEAKKLNQWL 81
ABC B4_2ghi_377_633_1.0_renumber.pdbA/1-257 1IDSFS ERGHK PDS I KGNLEFNDVHF SYPSRANVK I LKGLNLKVVQSGQTVLVGN SGCGKSTTVQLIQR LYDPTDEGT INI DGDQDIRNFVNYL 92
ABC B5_2hyd_564_807_1.0_renumber.pdbA/1-244 1-----DTC-----EGNLEFREVS F FYP CRP DVFI LRGLSLS I ERGKTVA FV GSSGCGKSTSVQLLQR LYDPTVQGVLLFDGVDAKELNVQWL 81
ABC B6_2hyd_579_822_1.0_renumber.pdbA/1-244 1GAGPLRFQ-----KGR I EFENVHFSYAD--GRE T LQDVSFTVMPGQT LALV GP SGAGKST I LRLLFRFYDISSGC I RIDGQDI SQVTQASL 84
ABC B7_2ghi_472_706_1.0_renumber.pdbA/1-235 1-----VAFDNVHF EYI--EGQKVLSG I SFEVPA GKKVA I VGGSGSGKST I VRLLFRFYEPQKGS I YLAGQNIQDVSLES L 73
ABC B8_2hyd_464_706_1.0_renumber.pdbA/1-243 1---PKEQL-----RGSVT FQNVCF SYPCRPGFEV LKDFTLTLP PGK I VALV GQSGGKST I VASLLERFYDPTAGVVM LDGRDLRTLDPSWL 83
ABC B9_2ixe_504_730_1.0_renumber.pdbA/1-227 1-----VDFENVTF TYRTRPHTQV L QNVS FLSLSPGKVTALV GP SGSGKSS CVN I LENFY PLEGG RVL LDGKPI SAYDHKYL 75

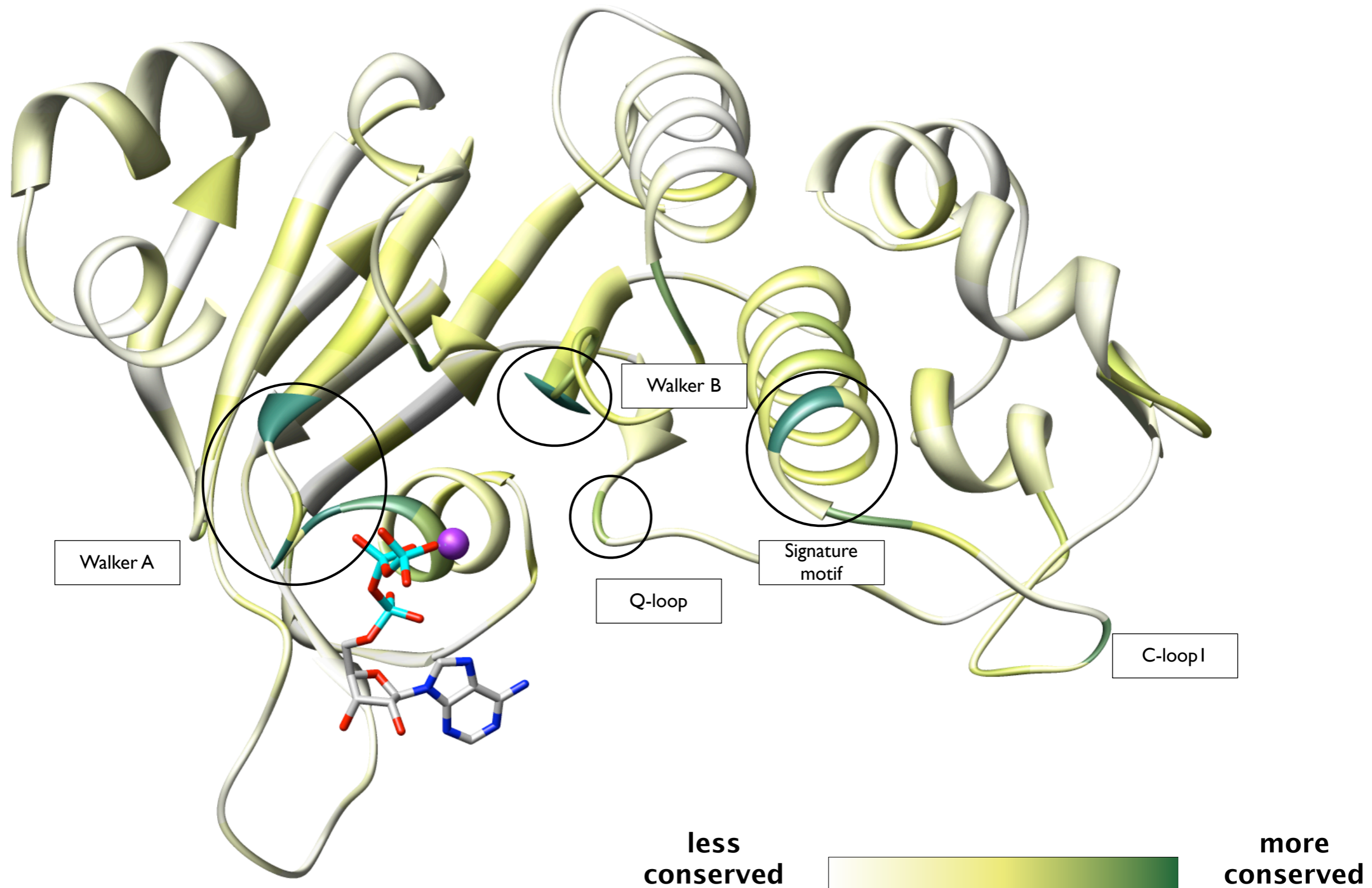
ABC B10_2hyd_489_731_1.0_renumber.pdbA/1-243 79 RSK- IGTV SQEP I L FSCS I AEN IAYGADDP SSVTAE E IQRVAE VANA VAF I RNFPQGFNT V VGEKGV L LSGGQ KQRI A I ARA L LKNPK I LLL 169
ABC B11_2ghi_1074_1313_1.0_renumber.pdbA/1-240 80 RSN- IGIV SQEP V L FACS I MDN I KYGDNT-KE I PMERV I AAKQAQLHDF VMSLPEKYET V VGSQGS LSRGE KQRI A I ARA I VRDPK I LLL 169
ABC B11_2hyd_413_653_1.0_renumber.pdbA/1-241 83 RDQ- IGIV EQEP V L FSTT I AEN IRYGRE--DATMED I VQA AKEANAYNF I MDLPQQFDT V VGEGGGMSGGQ KQRI A I ARA I LRNP K I LLL 170
ABC B1_2ff7_392_618_1.0_renumber.pdbA/1-227 76 REI- IGVV SQEP V L FATT I AEN IRYGRE--NVTMDE I EKAVKEANAYDF I MKLPHKFD T V GERGAC LSGGQ KQRI A I ARA I VRNP K I LLL 163
ABC B1_2ixe_1028_1271_1.0_renumber.pdbA/1-244 83 RAH- LGIV SQEP I L FDCS I AEN IAYG DNS-RVVSQEE I VRAAKEANI HAF I ESLPNKYSTH VGDKGFLSGGQ KQRI A I ARA I VRQPH I LLL 172
ABC B2_1jj7_494_742_1.0_renumber.pdbA/1-249 85 HRQ- VAAV QEPQV FGRS LQEN IAYG-LT-QKPTMEE I TAAAVKSGAHS F I SGLPQGYDT VDEAGS LSGGQ QAV A I ARA I IRKPCV L I L 173
ABC B3_1jj7_457_681_1.0_renumber.pdbA/1-225 87 HSQ- VVSV QEPV L FSGSVRNN IAYG-L--QSC EDDKVM AAAQA AHADDF I QEMEHG IYTD VGEKGS LAAGQ KQRI A I ARA I VRDPRV L I L 174
ABC B4_2ghi_1028_1278_1.0_renumber.pdbA/1-244 82 RAQ- LGIV SQEP I L FDCS I AEN IAYG DNS-RVVSQDE I VSAAKAANIHP F I ETLPHKYET VGDKGFLSGGQ KQRI A I ARA I IRQPQ I LLL 171
ABC B4_2ghi_377_633_1.0_renumber.pdbA/1-257 93 REI- IGVV SQEP V L FSTT I AEN I CYGRG--NVTMDE I KKAVKEANAYEF I MKL PQKFD T V GERGAC LSGGQ KQRI A I ARA I VRNP K I LLL 180
ABC B5_2hyd_564_807_1.0_renumber.pdbA/1-244 82 RSQ- IAI V PQEP V L FNC S I AEN IAYG DNS-RVVP LDE I KEAANAANIHS F I EGLPEKYNTQ VGLKGAQLSGGQ KQRI A I ARA I LQKPK I LLL 171
ABC B6_2hyd_579_822_1.0_renumber.pdbA/1-244 85 RSH- IGVV PQDT V L FNDT IADN IRYGRV--TAGNDEVEAAAQAAG I HDA IMAFPEGYRTQ VGERG L LSGGE KQRI A I ARA I LKAPG I I L 172
ABC B7_2ghi_472_706_1.0_renumber.pdbA/1-235 74 RRA- VGVV PQDAV L FHNT IYNN LLYGNI--SASPEEYAVAK LAGLHDA I LRMPHG YDTQ VGERG L LSGGE KQRI A I ARA I LKDPV I L Y 161
ABC B8_2hyd_464_706_1.0_renumber.pdbA/1-243 84 RGQV VGF I SQEP V L FGTT I MEN I RFGKL--EASDEEYVTAAREANAHEF I TSFPEGYNTV VGERG T LSGGQ KQRI A I ARA I IKQPTV L I L 172
ABC B9_2ixe_504_730_1.0_renumber.pdbA/1-227 76 HRV- I S L V SQEP V L FARS I TDN I SYG-LP-TVP-FEMVVEAAQKANAHGF I MELQDGYST TGEKGAQLSGGQ KQRI A I ARA I VRNP P V L I L 163

ABC B10_2hyd_489_731_1.0_renumber.pdbA/1-243 170 DEATSALDAENEYLVQEALDR--LMDGRTV LV AHR LST KNANMVA VLDQ GK I TEY GKHEELLSKPNGIYRKL MN----- 243
ABC B11_2ghi_1074_1313_1.0_renumber.pdbA/1-240 170 DEATSALDTESEKTVQVALDK--AREGRTCIV AHR LST QNADI IAVMAQGVV I EKGTHEELMA-QKGAYYKL----- 240
ABC B11_2hyd_413_653_1.0_renumber.pdbA/1-241 171 DMATSALDNESEAMVQEVLSK--IQHGHT I ISV AHR LST YRAADTI IGF EHGTAVERGTHEELLERK-GVYFTL----- 241
ABC B1_2ff7_392_618_1.0_renumber.pdbA/1-227 164 DEATSALDTESEAVVQVALDK--ARKGRTTIV AHR LST YRNADV IAGFDDGV IVEKGNHDELMKE----- 227
ABC B1_2ixe_1028_1271_1.0_renumber.pdbA/1-244 173 DEATSALDTESEKVVQEALDKAREG--RTCIV AHR LST QNADL I VV FQNGRVKEHGTHQQLLAQ-KGIYFSMV----- 244
ABC B2_1jj7_494_742_1.0_renumber.pdbA/1-249 174 DDATSALDANSQLQV EQLLYESPERSR SVLL TQHL S LVEQADH I L FLEGGAI R EGGTHQQLMEK-KGCYWAMVQA--- 249
ABC B3_1jj7_457_681_1.0_renumber.pdbA/1-225 175 DEATSALD---VQCEQALQDWN SRGDRTV LV AHR LQT QRAHQ I LVLQEGKLQ----- 225
ABC B4_2ghi_1028_1278_1.0_renumber.pdbA/1-244 172 DEATSALDTESEKVVQEALDK--AREGRTCIV AHR LST QNADL I VV FQNGRVKEHGTHQQLLA-QKGIYFSMVS----- 244
ABC B4_2ghi_377_633_1.0_renumber.pdbA/1-257 181 DEATSALDTESEAEVQAALDK--AREGRTTIV AHR LST YRNADV IAGFEDGV IVEQGSHELMK-KEGVYFKLVNMQTS 257
ABC B5_2hyd_564_807_1.0_renumber.pdbA/1-244 172 DEATSALDNDSEKVVQHALDK--ARTGRTCLV AHR LST YRNADL I VV LKNGK I KEQGT HQELLRN-DIYFKLVN--- 244
ABC B6_2hyd_579_822_1.0_renumber.pdbA/1-244 173 DEATSALDTSNERAIQASLAK--VCANRTTIV AHR LST YVNADQ I LV I KDGC I VERGRHEALLSRG-GVYADMW----- 244
ABC B7_2ghi_472_706_1.0_renumber.pdbA/1-235 162 DEATSALDSITEET I LGAMKD--VVKHRTS I F AHR LST VDADE I I VLDQ GKVA ERGTHHGLLANPHS I YSEMWH--- 235
ABC B8_2hyd_464_706_1.0_renumber.pdbA/1-243 173 DEATSALDAESERVVQEALDR--ASAGRTV LV AHR LST YRG AHC I VVMADGRVWEAGTHEELLKKG-GLYAEL----- 243
ABC B9_2ixe_504_730_1.0_renumber.pdbA/1-227 164 DEATSALDAESEYLIQQA I HGNLQK--HTVLI AHR LST YEHAHL I VVLDKGRVVQQTGTHQQLLAQ----- 227
```

# Common conservation patterns across all human NBDs suggest functional residues

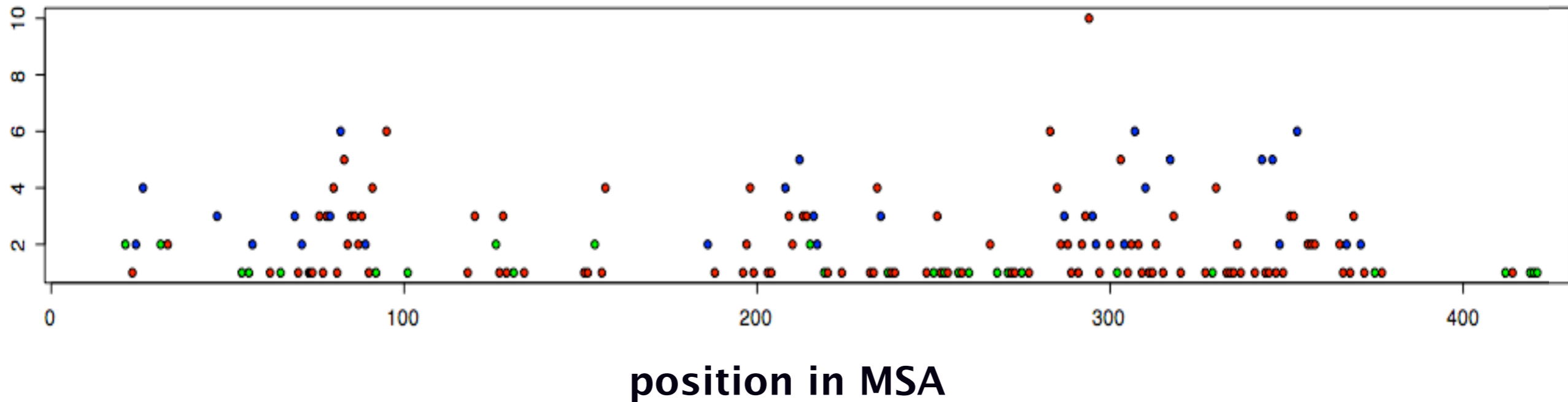


# Common conservation patterns across all human NBDs suggest functional residues

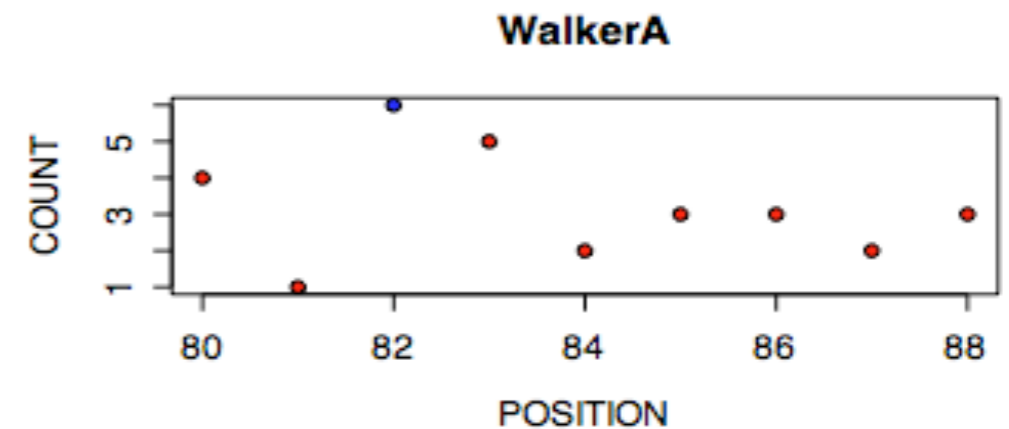


# Clinical data provides evidence for a common mechanism

number of mutants

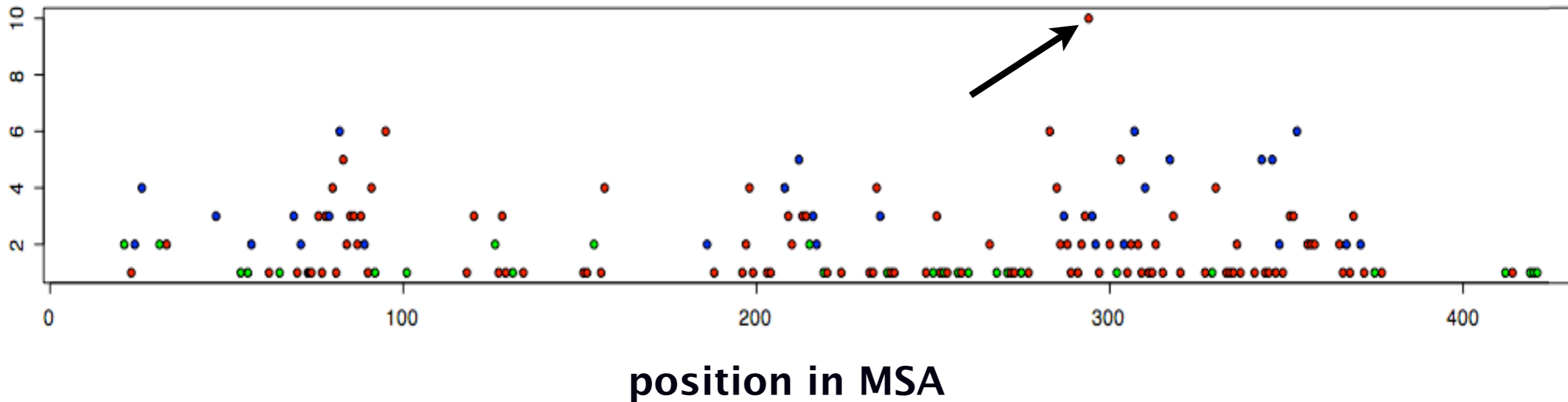


- Mutations mapped to a multiple structure alignment of all human NBD comparative models.
- Known motifs show disease mutations in multiple transporters.

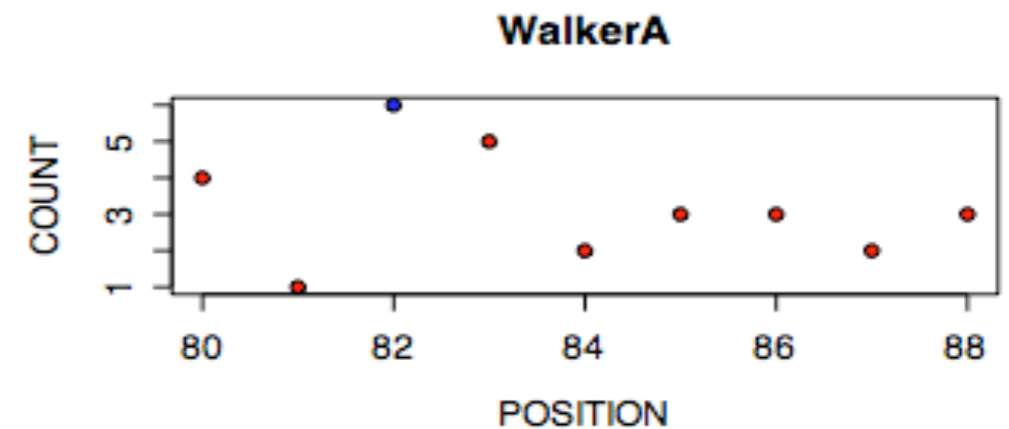


# Clinical data provides evidence for common mechanisms

number of mutants

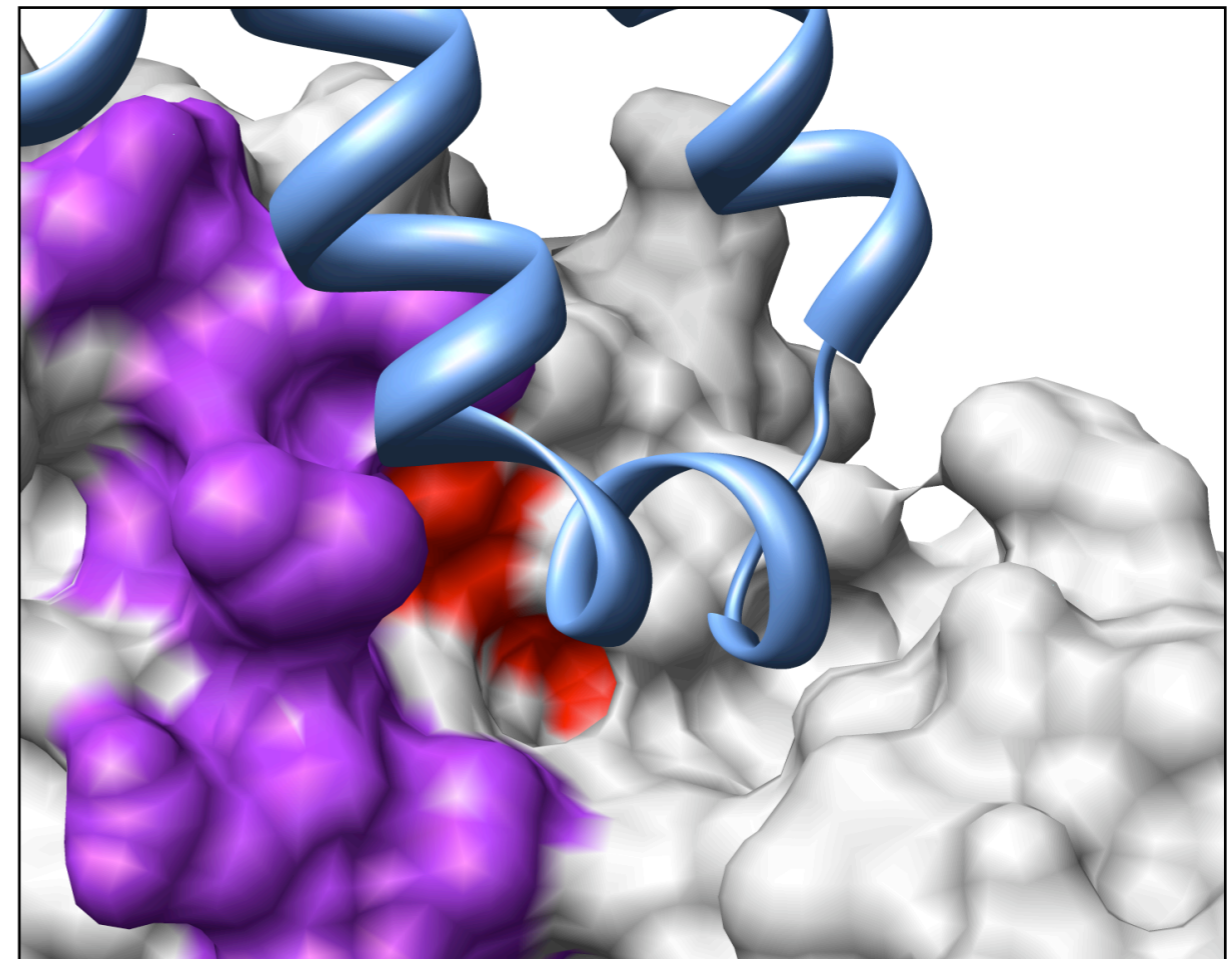


- Mutations mapped to a multiple structure alignment of all human NBD comparative models.
- Known motifs show disease mutations in multiple transporters.



# The context of structure: disease mutants at a putative communication network

NBD1	NBD2
ABCC6_ SLARAVY	ABCB11 AIARA IV
ABCC1_ SLARAVY	ABCC6_ CLARALL
ABCC2_ SLARATY	ABCC10_ CLARALL
ABCC5_ SLARALY	ABCB4_ AIARALI
ABCC7_ SLARAVY	ABCB1_ AIARALV
ABCC4_ NLARAVY	ABCC12_ CVARALL
ABCC9_ CVARALY	ABCC9_ CLARAFV
ABCC8_ SVARALY	ABCC3_ CLARALL
ABCB7_ AIARALI	ABCC11_ CIARAVL
ABCB5_ AIARALL	ABCC5_ CIARALL
ABCB3_ AIARALV	ABCC8_ CLARAFV
ABCB4_ AIARALV	ABCC1_ CLARALL
ABCB2_ ALARALI	ABCC4_ CLARALI
ABCB8_ AIARALI	
ABCB10_ AIARALL	
ABCB11_ AIARALI	
ABCB1_ AIARALV	
ABCC3_ SLARAVY	
ABCF3_ ALARALF	
ABCB9_ AMARALV	
ABCF2_ ALARALF	
ABCF1_ SLARALF	



- Well conserved motif at the TMD/NBD interface
- 15 disease associated mutations

Blue: transmembrane domain  
 Red: ARA  
 Purple: Q-loop

# Developing a general tool to integrate variant data

- Define a general set of features that distinguish between neutral and deleterious point mutants
- Use Random Forests (RF), a supervised learning algorithm, to combine the features for prediction

Karchin R, Diekhans M, Kelly L, Thomas DJ, Pieper U, Eswar N, Haussler D, Sali A.

*LS-SNP: large-scale annotation of coding non-synonymous SNPs based on multiple information sources.*

Bioinformatics. 2005 Jun 15;21(12):2814-20. Epub 2005 Apr 12.

Karchin R, Kelly L, Sali A.

*Improving functional annotation of non-synonymous SNPs with information theory.*

Pac Symp Biocomput. 2005;;:397-408.

# Using supervised learning to predict the effect of point mutants

- Vector of sequence, structure, and evolutionary features representing mutated residues



GENE	MUT	ASA-MUT	RSA-MUT	ASA-WT	RSA-WT	Delt-Res-Chrg	Delt-Res-Vol	Delt-Res-Pol	Pos-cons-score	Rel-Entropy	Granth-score	Buried-charge
CFTR	A1364V	6	0.04	6	0.06	0	-1.71	-1	-8.97	1.14	64	0
CFTR	A455E	0	0	0	0	1	-1.66	9.8	-8.27	0.66	107	1
CFTR	A559T	0	0	0	0	0	-0.92	0.4	-5.4	1.53	58	0



# Using supervised learning to predict the effect of point mutants

- Vector of sequence, structure, and evolutionary features representing mutated residues



GENE	MUT	ASA-MUT	RSA-MUT	ASA-WT	RSA-WT	Delt-Res-Chrg	Delt-Res-Vol	Delt-Res-Pol	Pos-cons-score	Rel-Entropy	Granth-score	Buried-charge
CFTR	A1364V	6	0.04	6	0.06	0	-1.71	-1	-8.97	1.14	64	0
CFTR	A455E	0	0	0	0	1	-1.66	9.8	-8.27	0.66	107	1
CFTR	A559T	0	0	0	0	0	-0.92	0.4	-5.4	1.53	58	0

# Using supervised learning to predict the effect of point mutants

- Vector of sequence, structure, and evolutionary features representing mutated residues



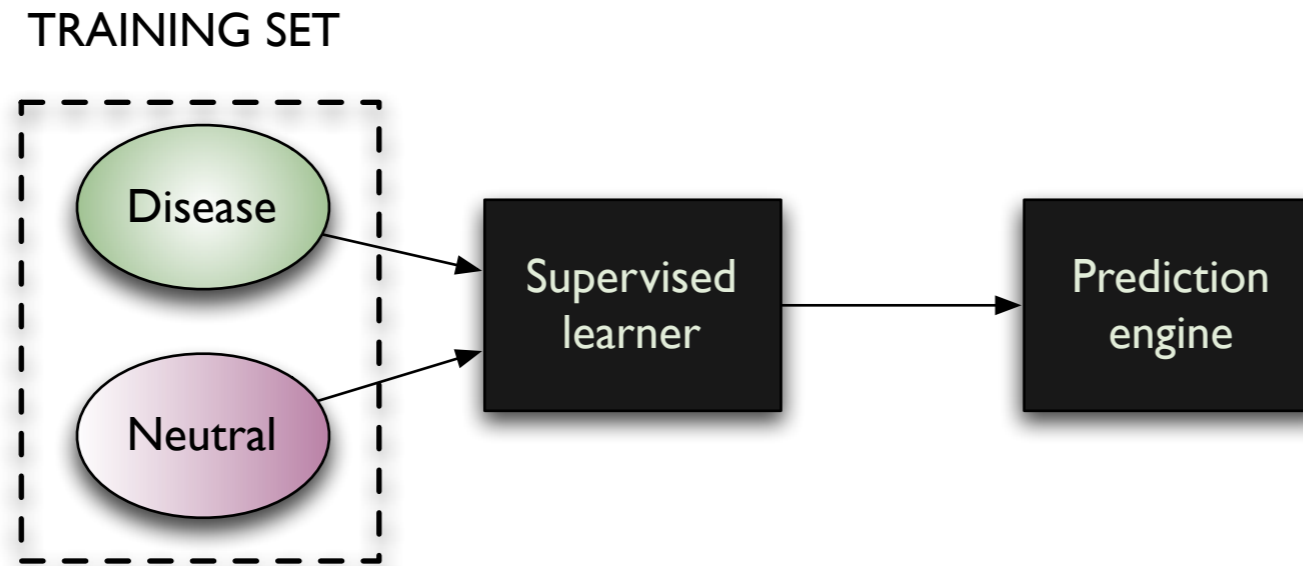
GENE	MUT	ASA-MUT	RSA-MUT	ASA-WT	RSA-WT	Delt-Res-Chrg	Delt-Res-Vol	Delt-Res-Pol	Pos-cons-score	Rel-Entropy	Granth. score	Buried-charge
CFTR	A1364V	6	0.04	6	0.06	0	-1.71	-1	-8.97	1.14	64	0
CFTR	A455E	0	0	0	0	1	-1.66	9.8	-8.27	0.66	107	1
CFTR	A559T	0	0	0	0	0	-0.92	0.4	-5.4	1.53	58	0

# Using supervised learning to predict the effect of point mutants

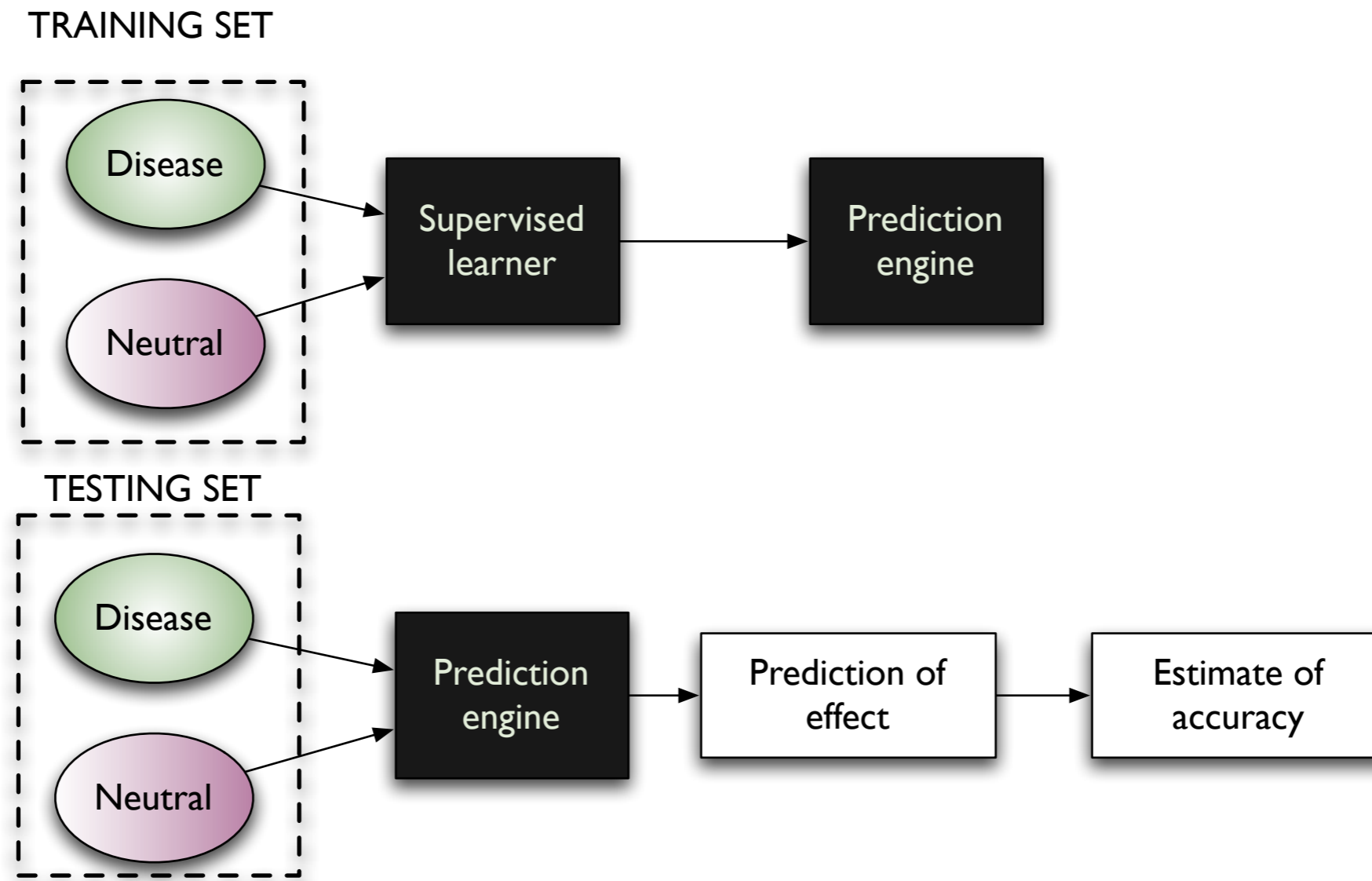
- Vector of sequence, structure, and evolutionary features representing mutated residues
- Binary prediction of effect

GENE	MUT	EFFECT
CFTR	A1364V	DISEASE
CFTR	A455E	DISEASE
CFTR	A559T	DISEASE

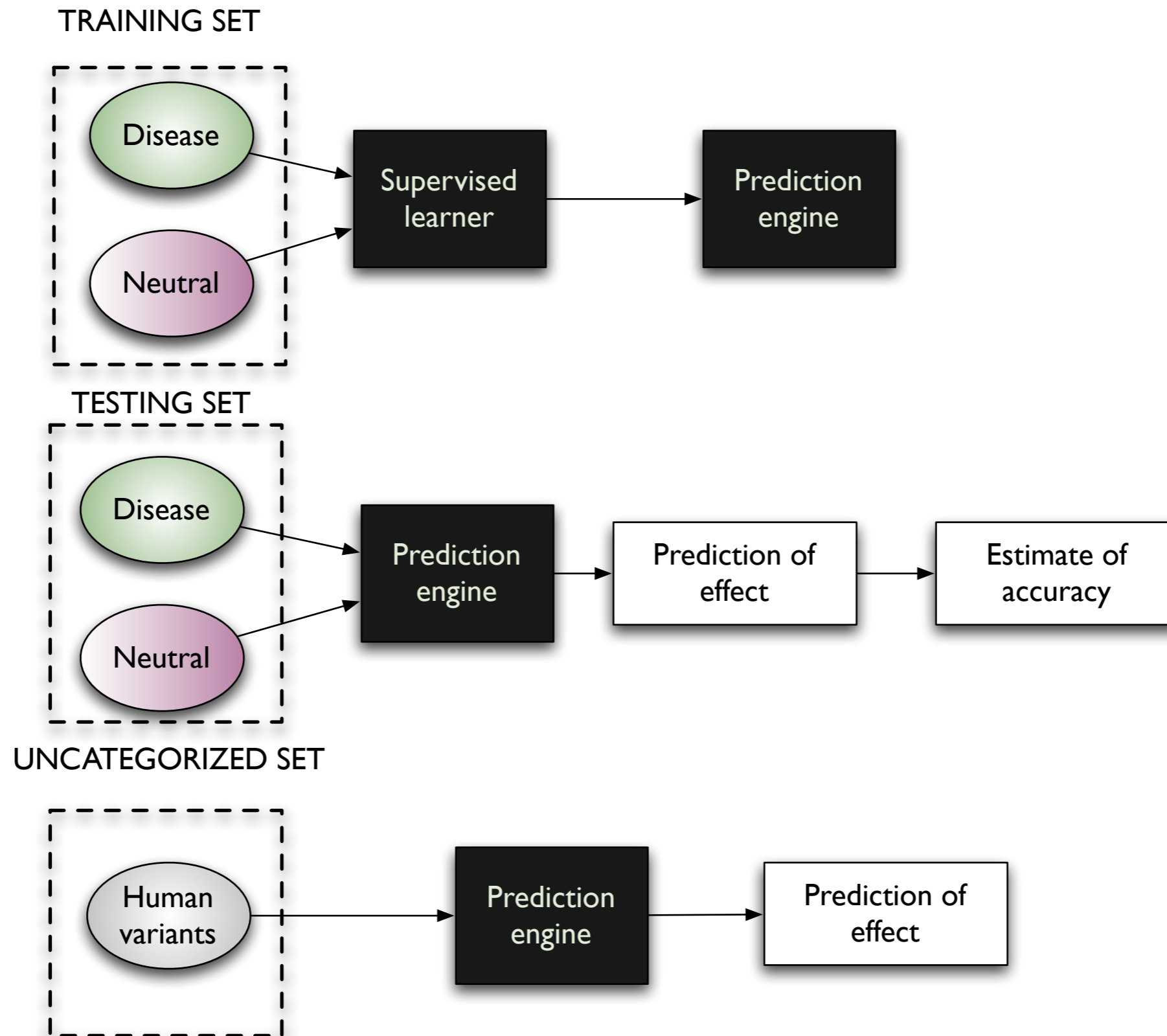
# A supervised learner “learns” classes of data



# A supervised learner “learns” classes of data

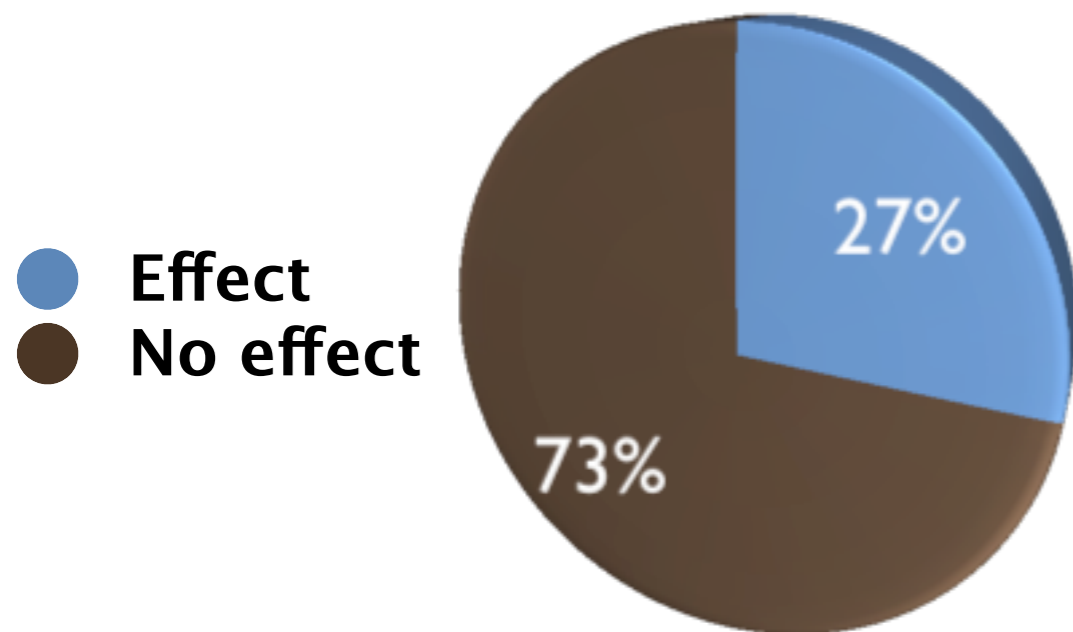


# A supervised learner “learns” classes of data

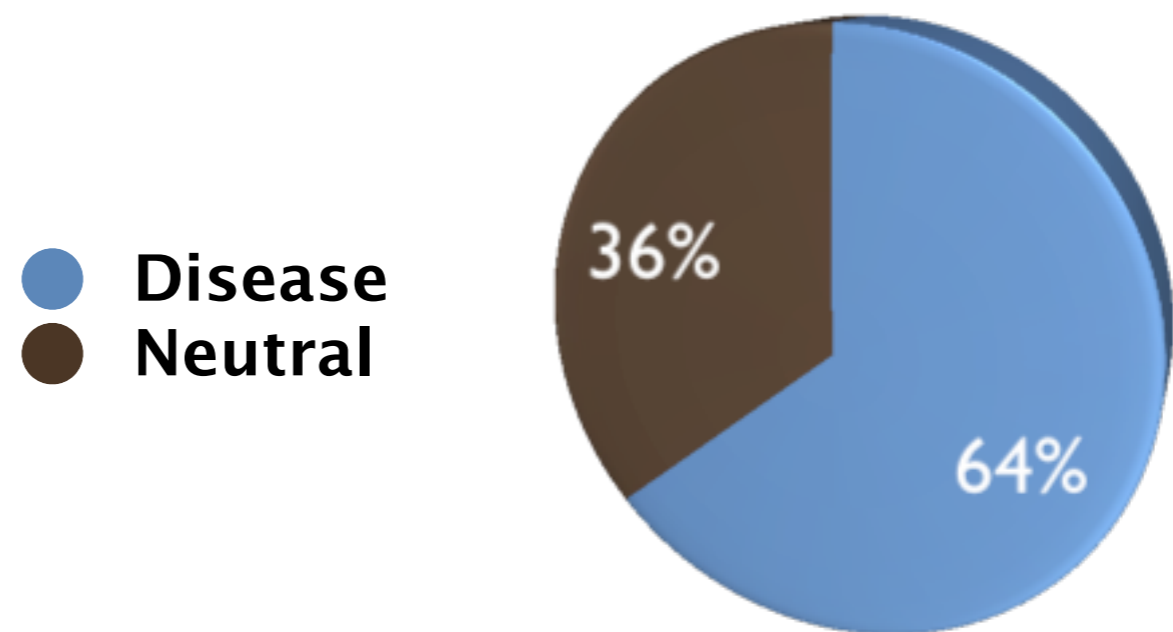


# We use independent training and test sets to validate our predictions

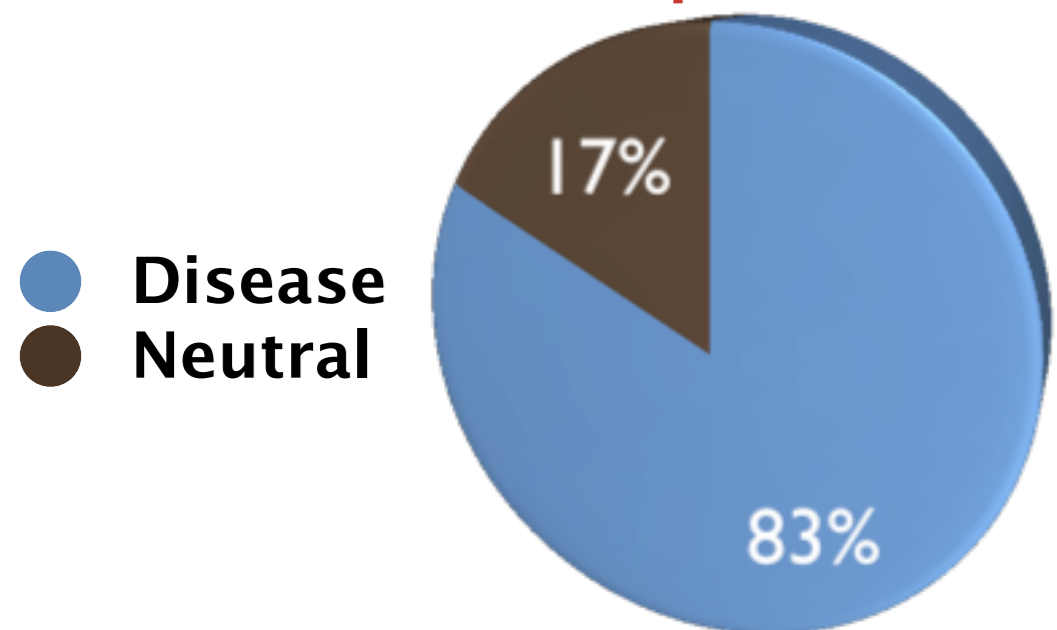
**'Experimental' training set**



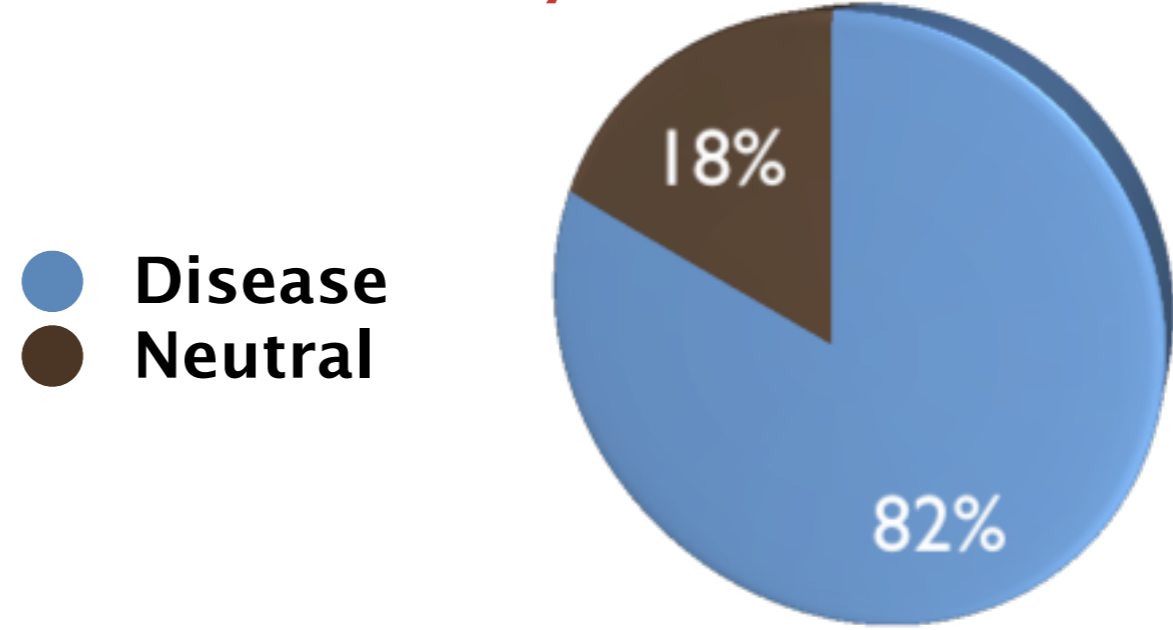
**'Clinical' training set**



**ABC transporter test set**

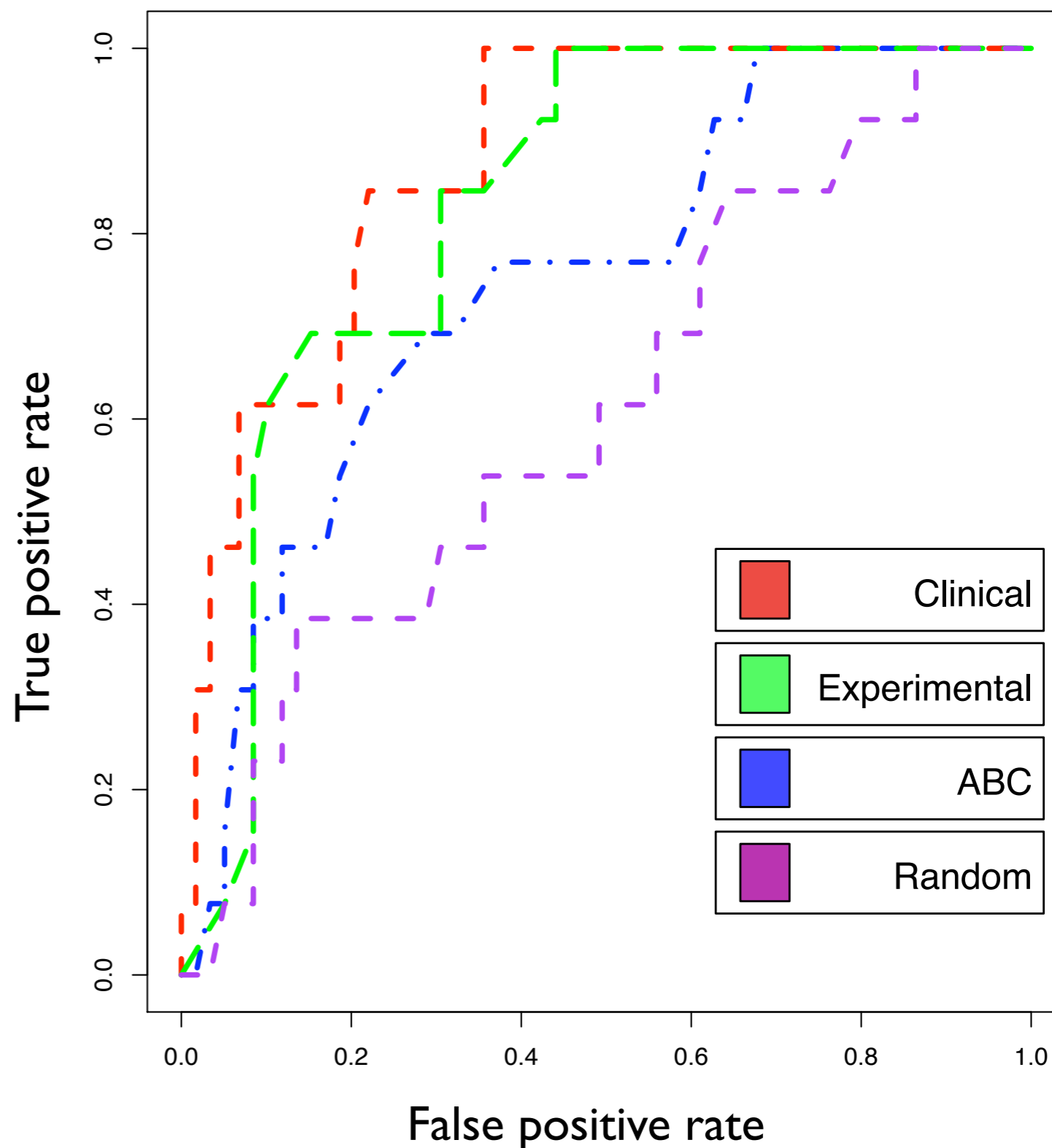


**Cystic fibrosis test set**



# The clinical-trained algorithm is the best performing classifier

- The clinical RF is 86% accurate on our cystic fibrosis test set.
- Now let's try it out on some unknown variants...



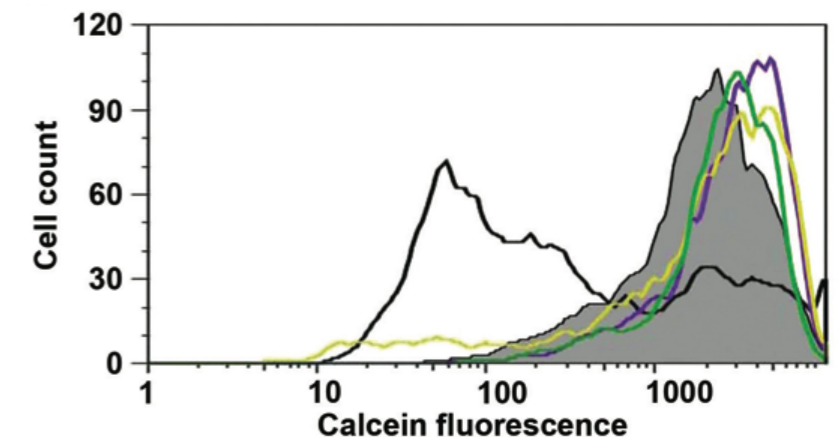


# Experimental functional analysis of ABC transporters

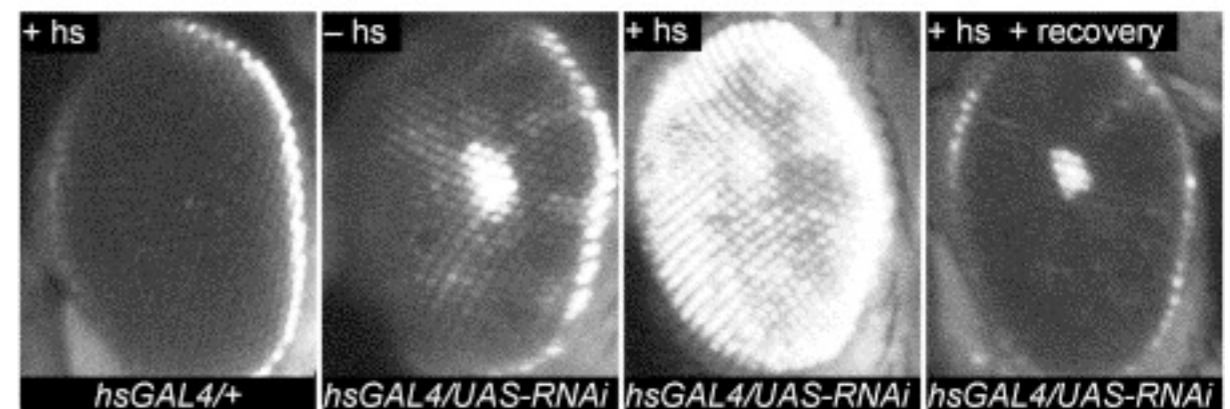
- **Yeast** – transport assays, cytotoxicity



- **Mammalian cells** – transport assays, cytotoxicity, promoter assays, mRNA expression, mRNA and protein expression



- **Drosophila** – live visualization of transport across membranes

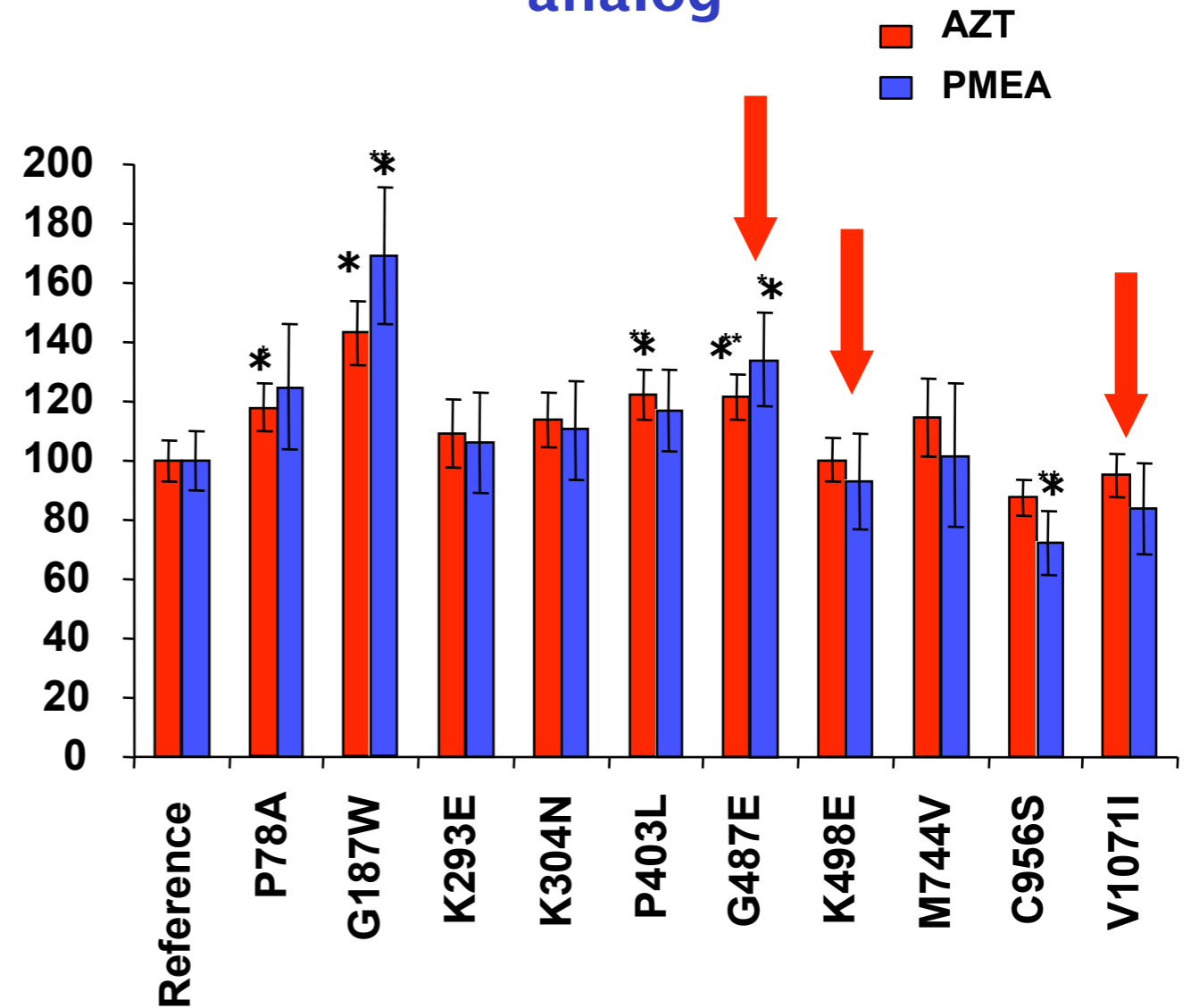


# Validation of MRP4 predictions

Transfected HEK cells,  
radiolabeled  
nucleoside/nucleotide  
analog

Variant	Prediction
G487E	Disease ✓
K498E	Neutral ✓
V1071I	Neutral ✓

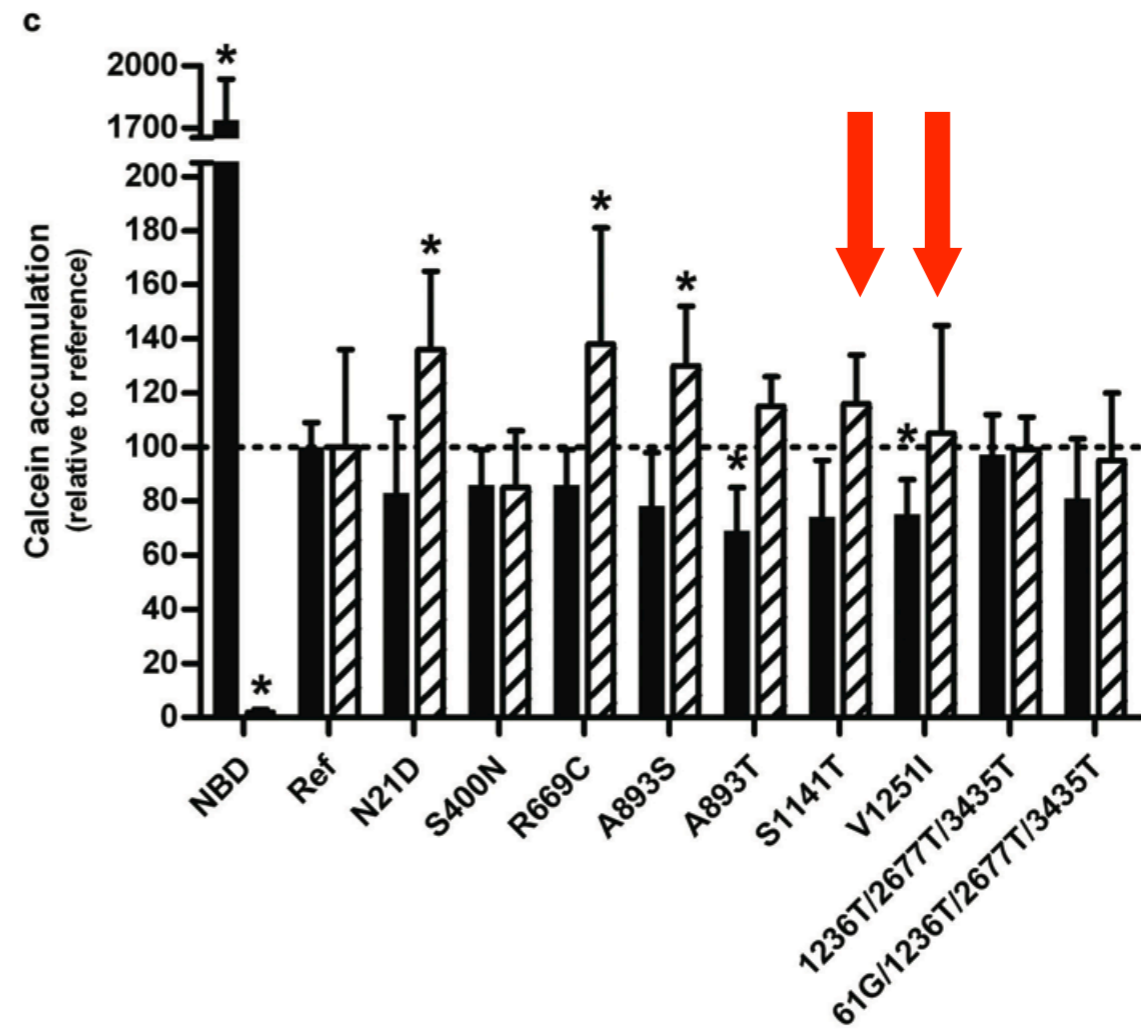
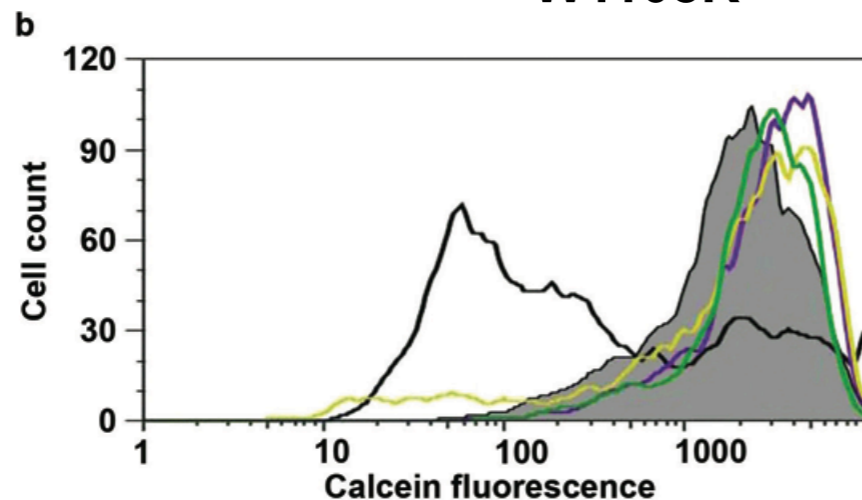
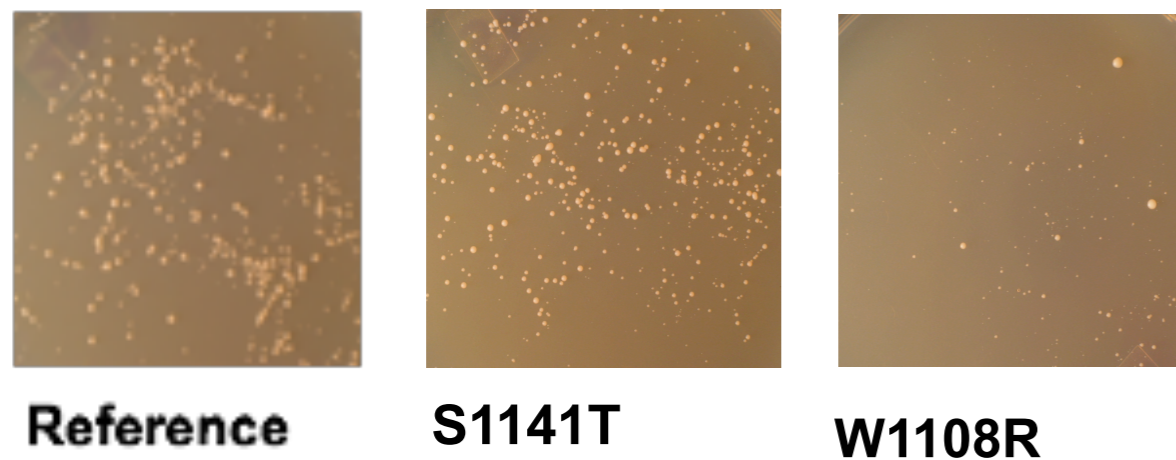
Intracellular Accumulation  
(% of MRP4 reference)



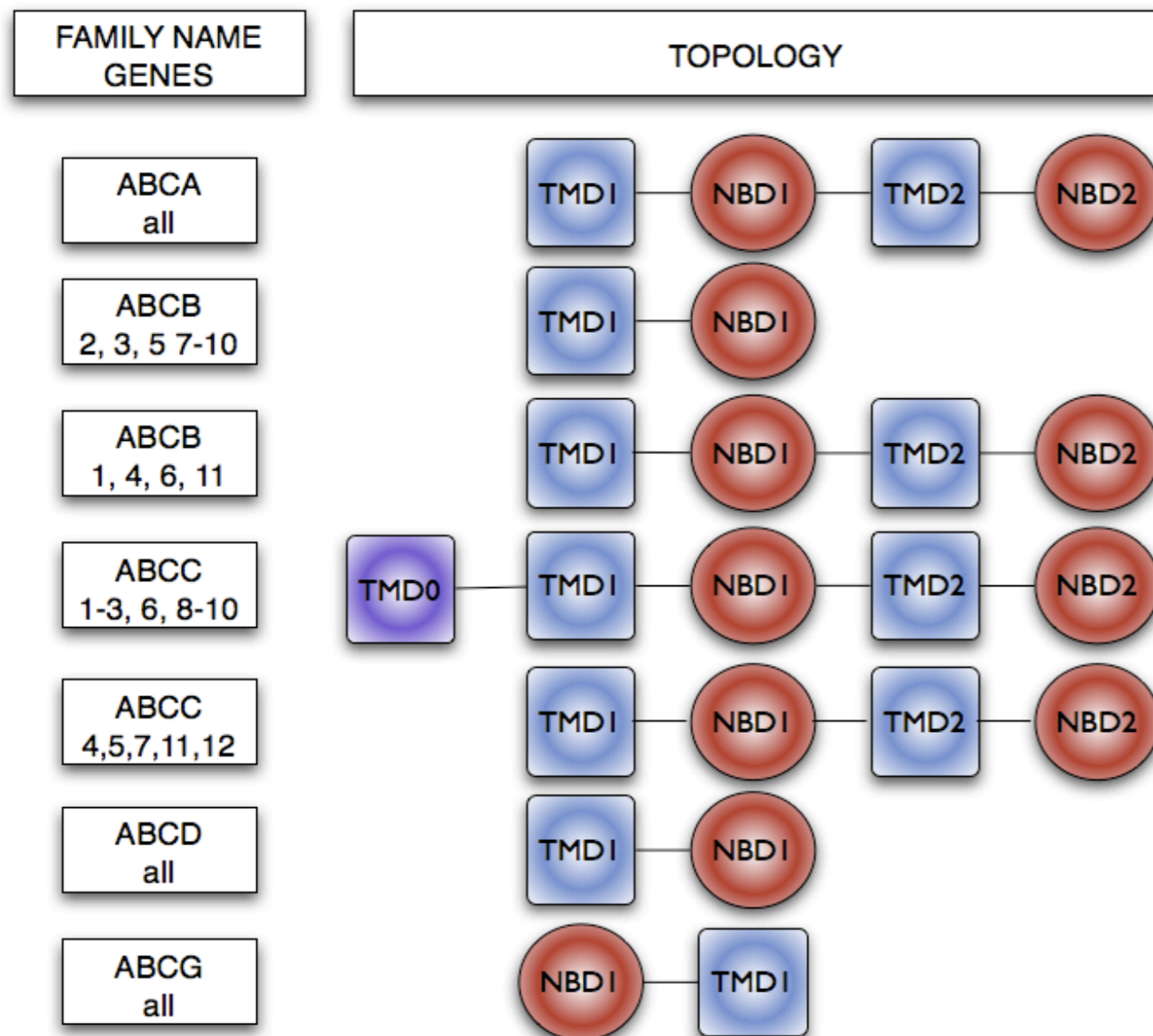
# Validation of P-gp predictions: FACS analysis, yeast assays

Variant	Prediction	
W1108R	Disease	✓
S1141T	Disease	✗
V1251I	Neutral	✓

Bodipy-paclitaxel accumulation in HEK293T cells transiently transfected with P-gp reference and variants

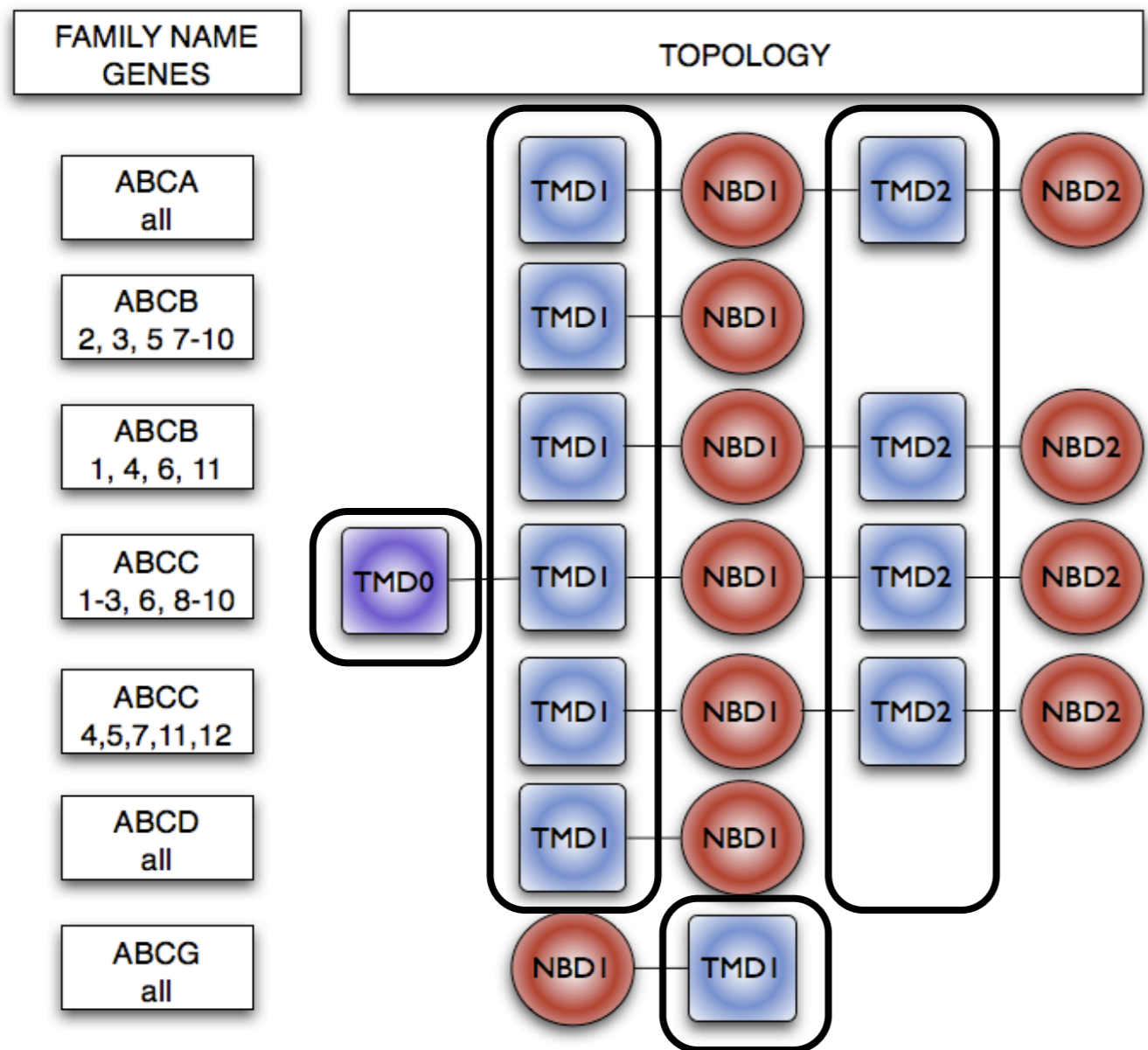


# Towards predicting substrate specificity for membrane transporters



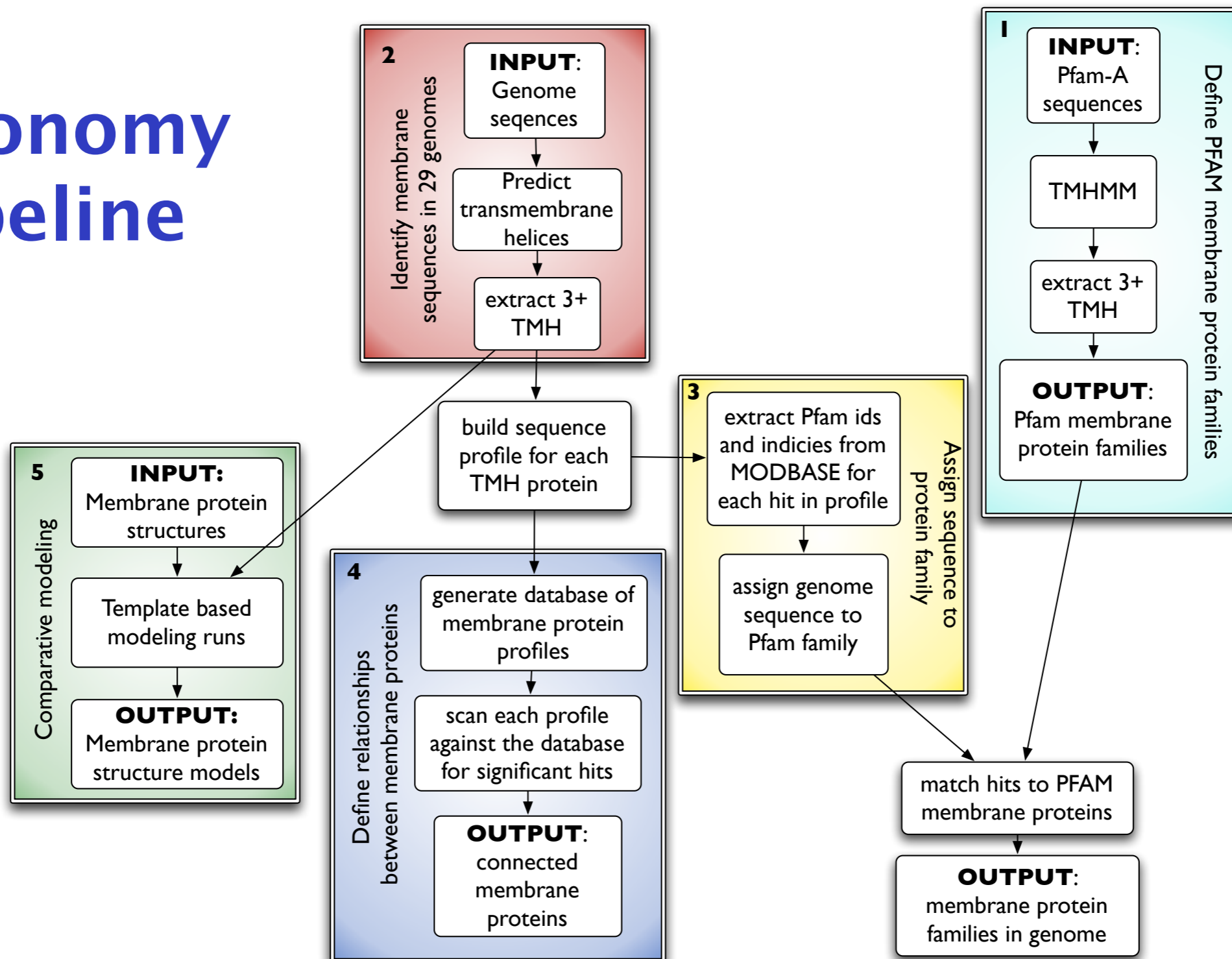
- Substrates bind in the transmembrane domains
- Overlapping substrate specificity
- Multidrug resistance
- Not easily alignable
- Extremely diverse in sequence

# Towards predicting substrate specificity for membrane transporters



- We excise all TMDs and create sequence profiles for each
- Each profile is scanned against a large database of membrane protein profiles
- Profiles that align well with each other are considered “connected”

# Taxonomy pipeline

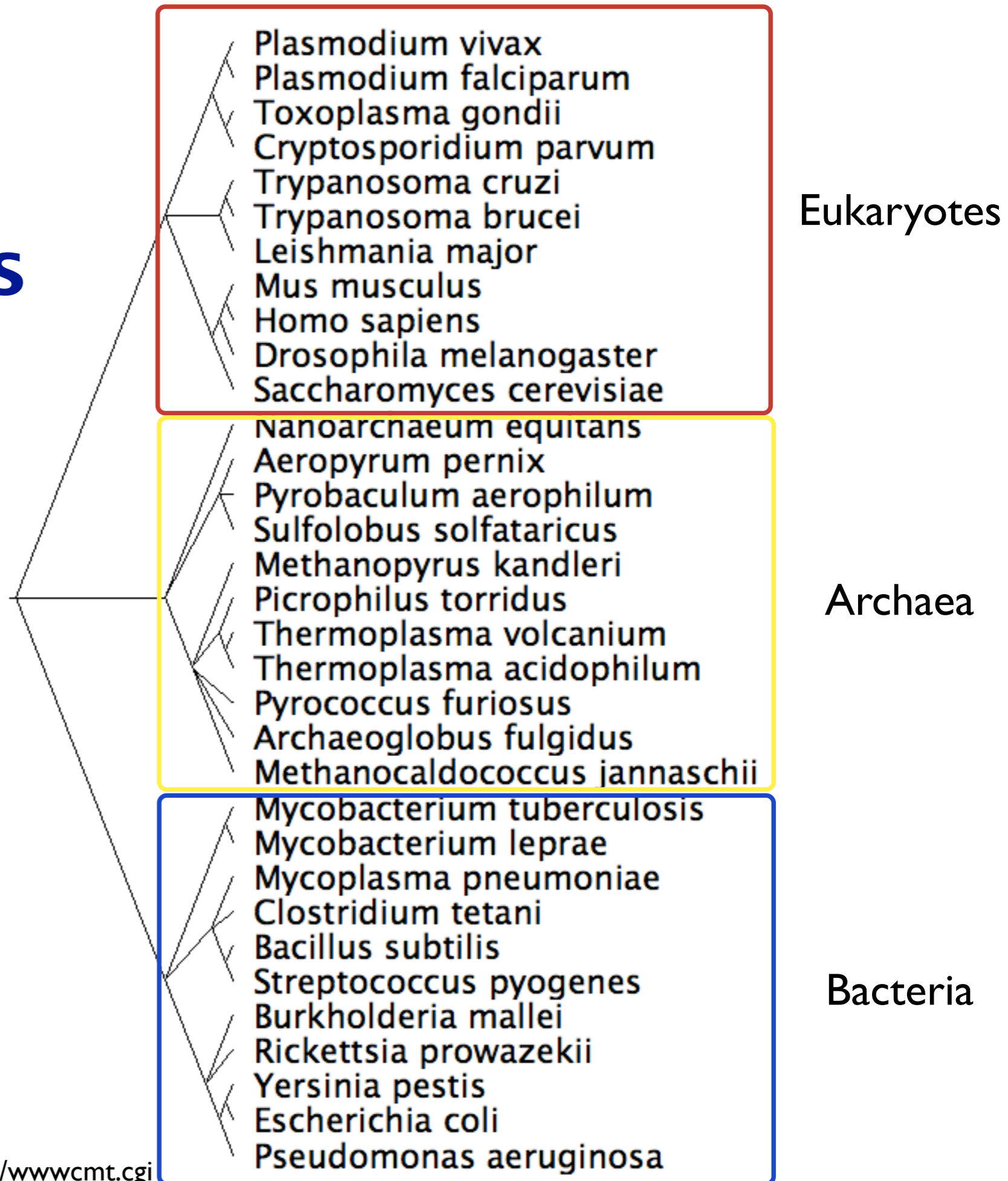


- Create sequence profiles for all alpha-helical membrane proteins with three or more helices in 34 organisms
- Use the sequence profiles to identify Pfam families in each genome and to define new families

# Extending the analysis to whole genomes

We selected:

- model species
- pathogenic species
- complete genomes
- genomic DNA available



# Identifying the membrane proteome of organisms

- 598 membrane protein families in Pfam
- How many times does each appear in a given organism?
- Can we find additional unidentified families?

membrane protein family



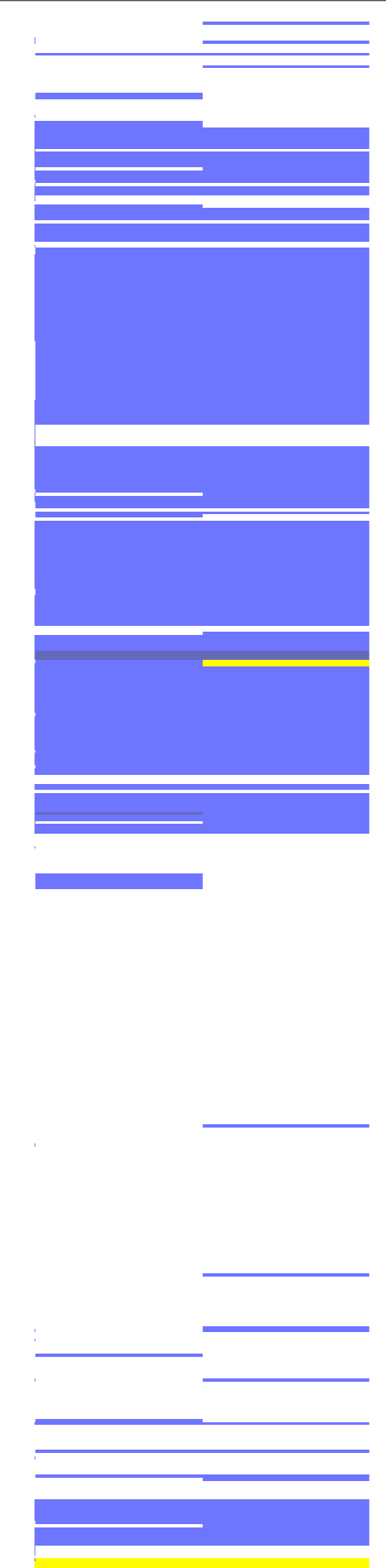
Human



# Identifying the membrane proteome of organisms

- 598 membrane protein families in Pfam
- How many times does each appear in a given organism?
- Can we find additional unidentified families?

membrane protein family



Human Mouse

# Identifying the membrane proteome of organisms

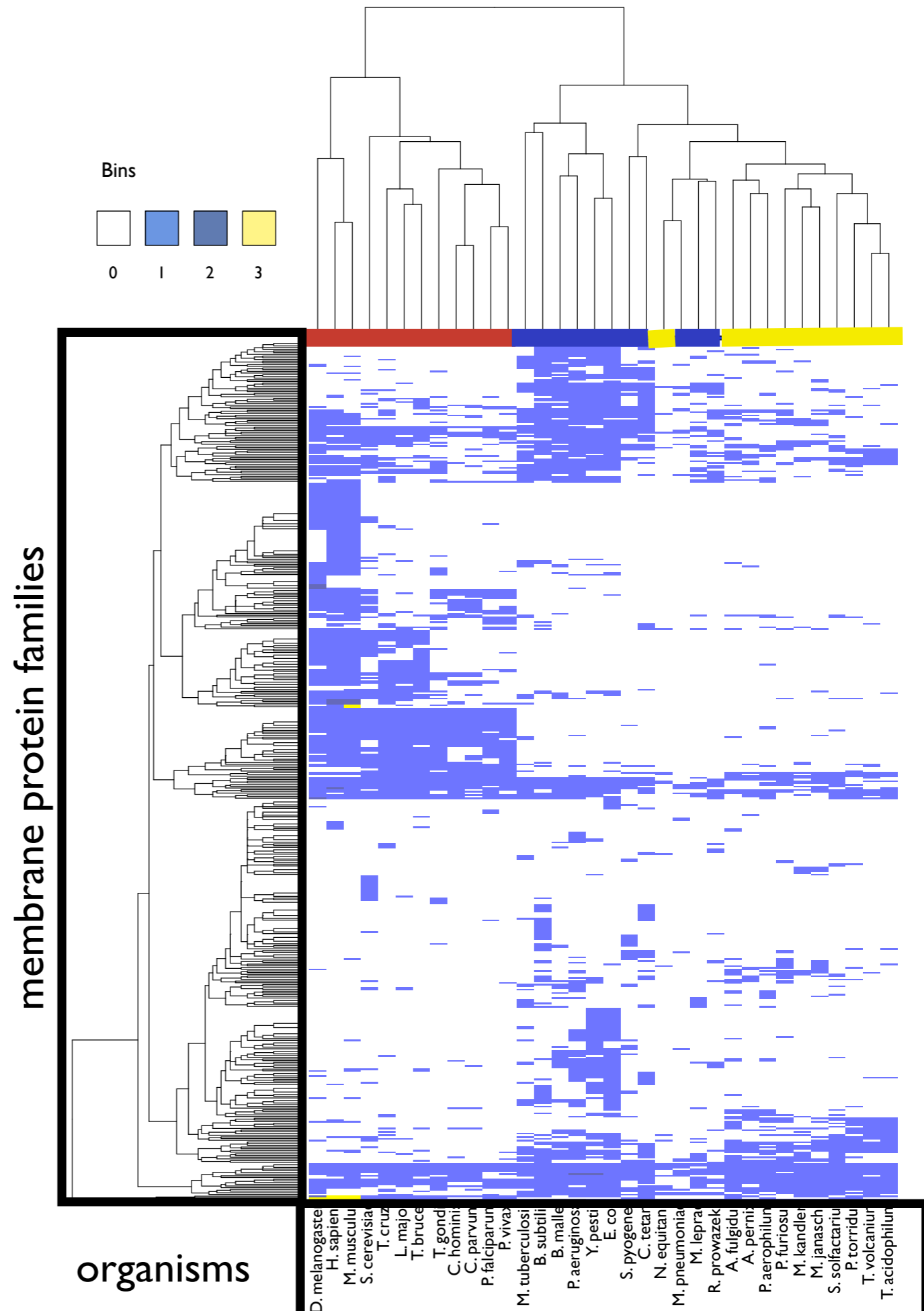
- 598 membrane protein families in Pfam
- How many times does each appear in a given organism?
- Can we find additional unidentified families?

membrane protein family



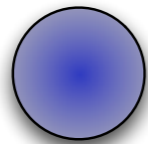
# Family content reveals a clear split between prokaryotes and eukaryotes

- Binning to compensate for large range in the number of family members per genome
- Clusters of families that tend to travel together
- Clusters of families that appear in specific organisms



# We create a large database of membrane protein profiles and use it to scan for similarity

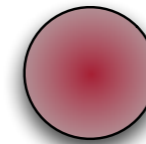
protein 1



Multiple sequence alignment of protein 1

```
4BCB10_2hyd_489_751_1.0_renumber.pdbA/1-245 79 KSK-IGIVSQEPILFSCSIAENIAY
4BCB11_2ghi_1074_1313_1.0_renumber.pdbA/1-240 80 RSN-IGIVSQEPVLFACSIMDNIKY
4BCB11_2hyd_413_653_1.0_renumber.pdbA/1-241 83 RDQ-IGIVEQEPVLFSTTIAENIRY
4BCB1_2ff7_392_618_1.0_renumber.pdbA/1-227 76 REI-IGVVSQEPVLFATTIAENIRY
4BCB1_2ixe_1028_1271_1.0_renumber.pdbA/1-244 83 RAH-LGIVSQEPILFDCSIAENIAY
4BCB2_1jj7_494_742_1.0_renumber.pdbA/1-249 85 HRQ-VAAVQEPQVFGRLQENIAY
4BCB3_1jj7_457_681_1.0_renumber.pdbA/1-225 87 HSQ-VVSVQEPVLFSGSVRNNIAY
4BCB4_2ghi_1028_1278_1.0_renumber.pdbA/1-244 82 RAQ-LGIVSQEPILFDCSIAENIAY
4BCB4_2ghi_377_633_1.0_renumber.pdbA/1-257 93 REI-IGVVSQEPVLFSTTIAENICY
4BCB5_2hyd_564_807_1.0_renumber.pdbA/1-244 82 RSQ-IAIVPQEPVLFNCSIAENIAY
4BCB6_2hyd_579_822_1.0_renumber.pdbA/1-244 85 RSH-IGVVPQDTVLFNDTIADNIRY
4BCB7_2ghi_472_706_1.0_renumber.pdbA/1-235 74 RRA-VGVVPQDAVLFHNTIYYNLLY
4BCB8_2hyd_464_706_1.0_renumber.pdbA/1-243 84 RGQVVGFI SQEPVLFGTTIMENIRF
4BCB9_2ixe_504_730_1.0_renumber.pdbA/1-227 76 HRV-ISLVSQEPVLFARSI TDNISY
```

protein 2



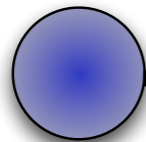
Multiple sequence alignment of protein 2

```
4BCB10_2hyd_489_751_1.0_renumber.pdbA/1-245 79 KSK-IGIVSQEPILFSCSIAENIAY
4BCB11_2ghi_1074_1313_1.0_renumber.pdbA/1-240 80 RSN-IGIVSQEPVLFACSIMDNIKY
4BCB11_2hyd_413_653_1.0_renumber.pdbA/1-241 83 RDQ-IGIVEQEPVLFSTTIAENIRY
4BCB1_2ff7_392_618_1.0_renumber.pdbA/1-227 76 REI-IGVVSQEPVLFATTIAENIRY
4BCB1_2ixe_1028_1271_1.0_renumber.pdbA/1-244 83 RAH-LGIVSQEPILFDCSIAENIAY
4BCB2_1jj7_494_742_1.0_renumber.pdbA/1-249 85 HRQ-VAAVQEPQVFGRLQENIAY
4BCB3_1jj7_457_681_1.0_renumber.pdbA/1-225 87 HSQ-VVSVQEPVLFSGSVRNNIAY
4BCB4_2ghi_1028_1278_1.0_renumber.pdbA/1-244 82 RAQ-LGIVSQEPILFDCSIAENIAY
4BCB4_2ghi_377_633_1.0_renumber.pdbA/1-257 93 REI-IGVVSQEPVLFSTTIAENICY
4BCB5_2hyd_564_807_1.0_renumber.pdbA/1-244 82 RSQ-IAIVPQEPVLFNCSIAENIAY
4BCB6_2hyd_579_822_1.0_renumber.pdbA/1-244 85 RSH-IGVVPQDTVLFNDTIADNIRY
4BCB7_2ghi_472_706_1.0_renumber.pdbA/1-235 74 RRA-VGVVPQDAVLFHNTIYYNLLY
4BCB8_2hyd_464_706_1.0_renumber.pdbA/1-243 84 RGQVVGFI SQEPVLFGTTIMENIRF
4BCB9_2ixe_504_730_1.0_renumber.pdbA/1-227 76 HRV-ISLVSQEPVLFARSI TDNISY
```

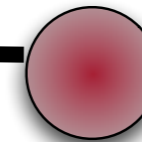
- Each membrane protein profile can be represented as a node

# We create a large database of membrane protein profiles and use it to scan for similarity

protein 1



protein 2



Multiple sequence alignment of protein 1

```

4BCB10_2hyd_489_751_1.0_renumber.pdbA/1-245 79 KSK-IGIVSQEPVLFACSIAMDNIKY
4BCB11_2ghi_1074_1313_1.0_renumber.pdbA/1-240 80 RSN-IGIVSQEPVLFACSIAMDNIKY
4BCB11_2hyd_413_653_1.0_renumber.pdbA/1-241 83 RDQ-IGIVEQEPVLFSTTIAENIRY
4BCB1_2ff7_392_618_1.0_renumber.pdbA/1-227 76 REI-IGVVSQEPVLFATTIAENIRY
4BCB1_2ixe_1028_1271_1.0_renumber.pdbA/1-244 83 RAH-LGIVSQEPILFDCSIAENIAY
4BCB2_1jj7_494_742_1.0_renumber.pdbA/1-249 85 HRQ-VAAVQEPQVFGRLQENIAY
4BCB3_1jj7_457_681_1.0_renumber.pdbA/1-225 87 HSQ-VVSVQEPVLFSGSVRNNIAY
4BCB4_2ghi_1028_1278_1.0_renumber.pdbA/1-244 82 RAQ-LGIVSQEPILFDCSIAENIAY
4BCB4_2ghi_377_633_1.0_renumber.pdbA/1-257 93 REI-IGVVSQEPVLFSTTIAENICY
4BCB5_2hyd_564_807_1.0_renumber.pdbA/1-244 82 RSQ-IAIVPQEPVLFNCSIAENIAY
4BCB6_2hyd_579_822_1.0_renumber.pdbA/1-244 85 RSH-IGVVPQDTVLFNDTIADNIRY
4BCB7_2ghi_472_706_1.0_renumber.pdbA/1-235 74 RRA-VGVVPQDAVLFHNTIYYNLLY
4BCB8_2hyd_464_706_1.0_renumber.pdbA/1-243 84 RQVVGFISQEPVLFGTTIMENIRF
4BCB9_2ixe_504_730_1.0_renumber.pdbA/1-227 76 HRV-ISLVSQEPVLFARSIITDNI SY
    
```

Multiple sequence alignment of protein 2

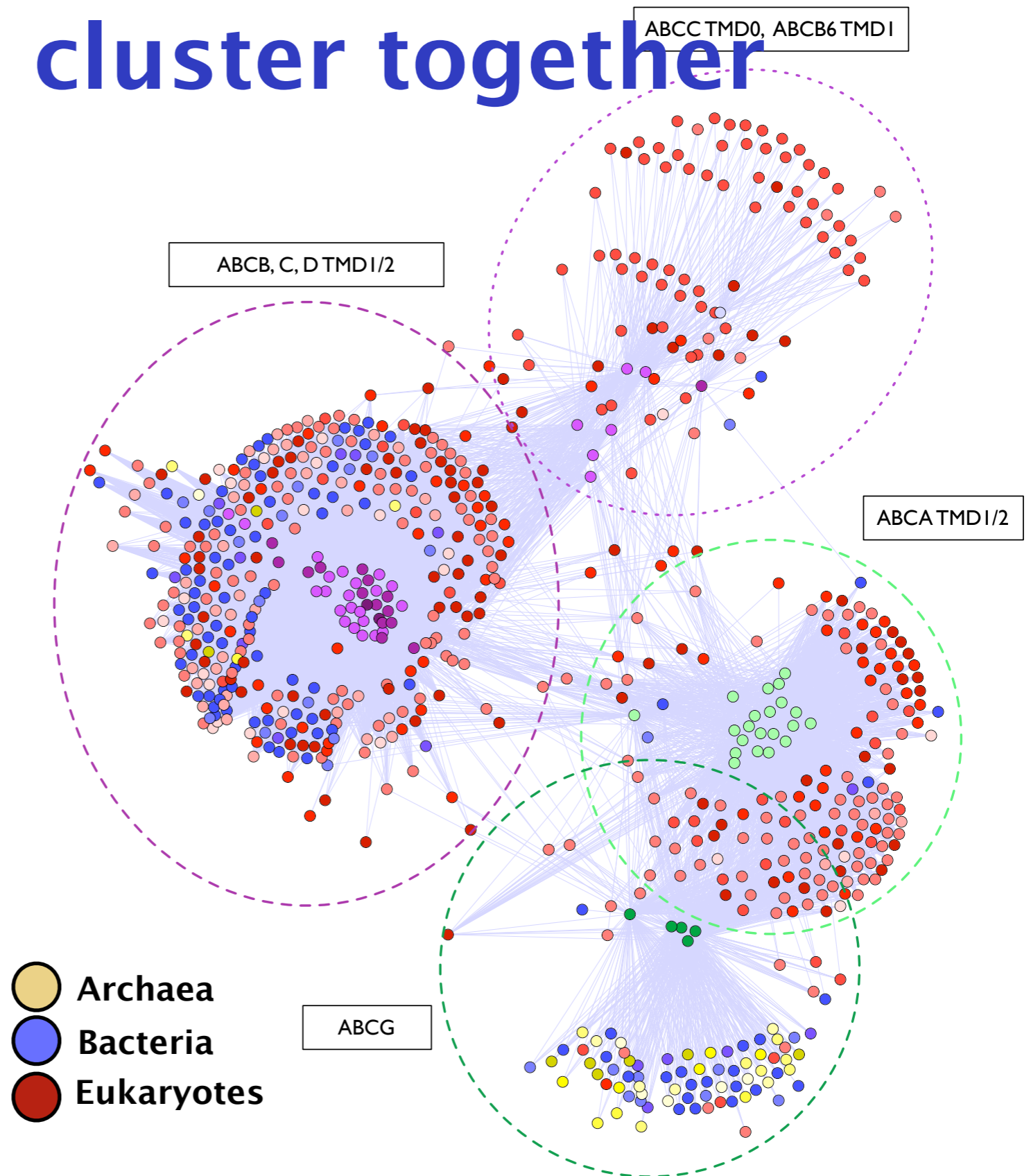
```

4BCB10_2hyd_489_751_1.0_renumber.pdbA/1-245 79 KSK-IGIVSQEPVLFACSIAMDNIKY
4BCB11_2ghi_1074_1313_1.0_renumber.pdbA/1-240 80 RSN-IGIVSQEPVLFACSIAMDNIKY
4BCB11_2hyd_413_653_1.0_renumber.pdbA/1-241 83 RDQ-IGIVEQEPVLFSTTIAENIRY
4BCB1_2ff7_392_618_1.0_renumber.pdbA/1-227 76 REI-IGVVSQEPVLFATTIAENIRY
4BCB1_2ixe_1028_1271_1.0_renumber.pdbA/1-244 83 RAH-LGIVSQEPILFDCSIAENIAY
4BCB2_1jj7_494_742_1.0_renumber.pdbA/1-249 85 HRQ-VAAVQEPQVFGRLQENIAY
4BCB3_1jj7_457_681_1.0_renumber.pdbA/1-225 87 HSQ-VVSVQEPVLFSGSVRNNIAY
4BCB4_2ghi_1028_1278_1.0_renumber.pdbA/1-244 82 RAQ-LGIVSQEPILFDCSIAENIAY
4BCB4_2ghi_377_633_1.0_renumber.pdbA/1-257 93 REI-IGVVSQEPVLFSTTIAENICY
4BCB5_2hyd_564_807_1.0_renumber.pdbA/1-244 82 RSQ-IAIVPQEPVLFNCSIAENIAY
4BCB6_2hyd_579_822_1.0_renumber.pdbA/1-244 85 RSH-IGVVPQDTVLFNDTIADNIRY
4BCB7_2ghi_472_706_1.0_renumber.pdbA/1-235 74 RRA-VGVVPQDAVLFHNTIYYNLLY
4BCB8_2hyd_464_706_1.0_renumber.pdbA/1-243 84 RQVVGFISQEPVLFGTTIMENIRF
4BCB9_2ixe_504_730_1.0_renumber.pdbA/1-227 76 HRV-ISLVSQEPVLFARSIITDNI SY
    
```

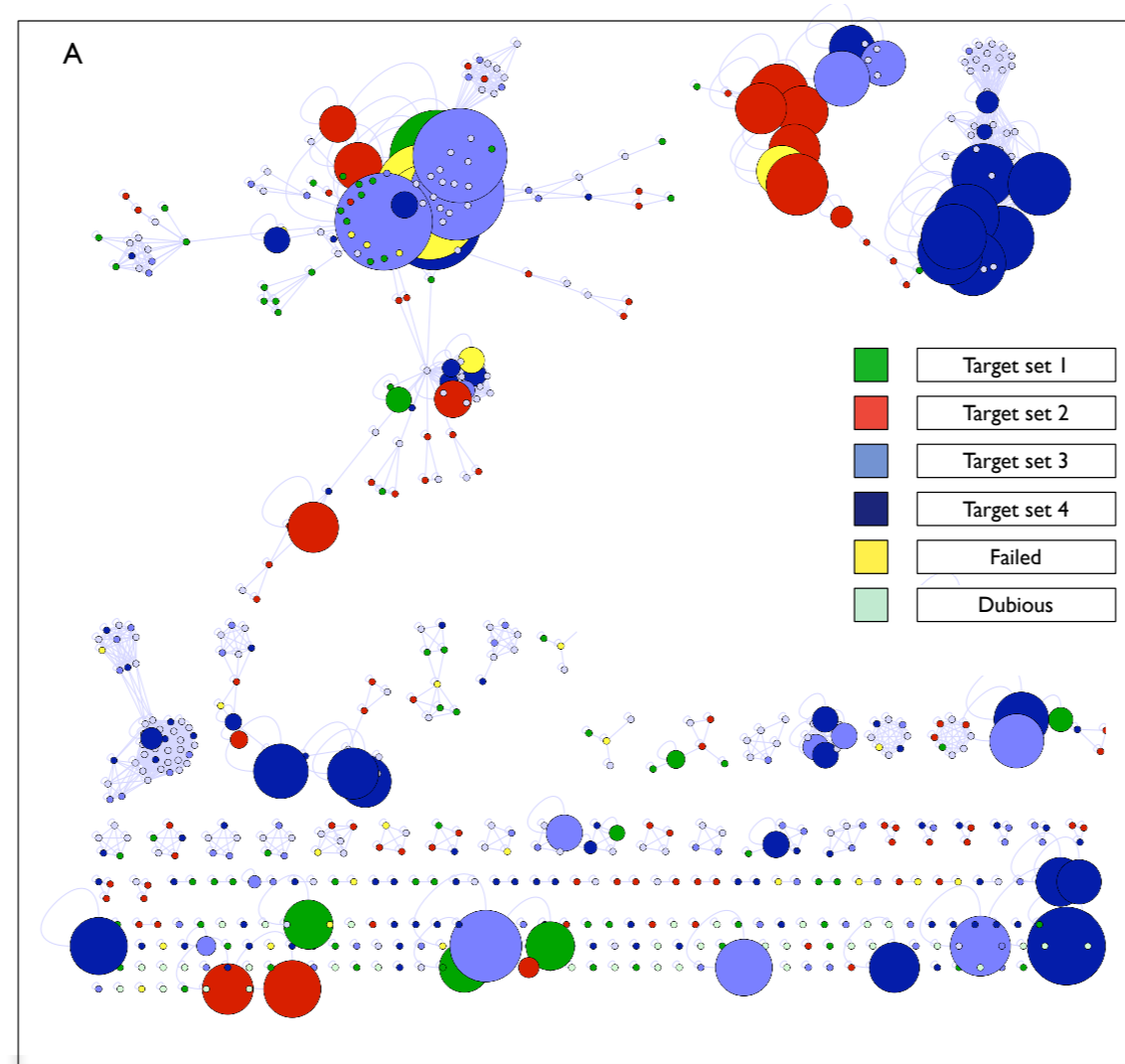
- Each membrane protein profile can be represented as a node
- If two profiles are significantly similar they are “connected” with a line

# ABC transporters with similar substrates cluster together

- The clusters reproduce evolutionary trees based on the NBDs
- NBD/TMD domains evolved together
- ABCA, with no archaeal hits and few bacterial hits, may be the most recently evolved transporter family



# This analysis can also be used to guide target selection



- **Select a set of three or more  $\alpha$ -helical membrane proteins in the yeast genome**
  - **represent all families and clusters**
  - **two members per family**

# Comparative modeling of a human Rh factor protein based on a bacterial template

- NeRH is an ammonia transporter from the bacterium *Nitrosomonas Europaea*. The structure was recently solved by Franz Gruswitz in Robert Stroud's lab.
- 40% sequence identical to the human Rh gene hRhBG, alignment is good outside of one long loop in the human protein (~25 residues).

```

hRhBG/1-352      1 -ADNEFYFRYPSPQDVHAMVFVGFGLMV-FLQ-RYGFSVGFTFLLAAFAALQWSTLVQGFLH- /----- 61
NeRH/1-351      1 INEARLVAQYNYSINILAMLLVGFGLMV-FVR-RYGFSATTGTYLVVATGLPLYILLR-ANGI /----- 62

hRhBG/1-352     62 -----GHIHVGVESMINADFCAGAVLISFGAVLGKTGPTQLLLMALLEVVL-FGINEFVLLHLLG-----VR 122
NeRH/1-351     63 -----FGHALTPHSVDAVIYAEFAVATGLIAMGAVLGRRLRVFQYALLALFIVPV-YLLNEWLVLDNASGLTEGFQ 131

hRhBG/1-352    123 DAGGSMTIHTFGAYFGLVLSRVLYRPQLEKSKHRQGSVYHSDLFAMIGTIFLWIFWPSFNAA LTA-LGAGQHRTALNT 199
NeRH/1-351    132 DSAGSIAIHAFGAYFGLGVSIALTTAAQRAQP--IESDATSDRFSMLGSMVLWLFWPSFATAIVP--FEQMPQTI VNT 205

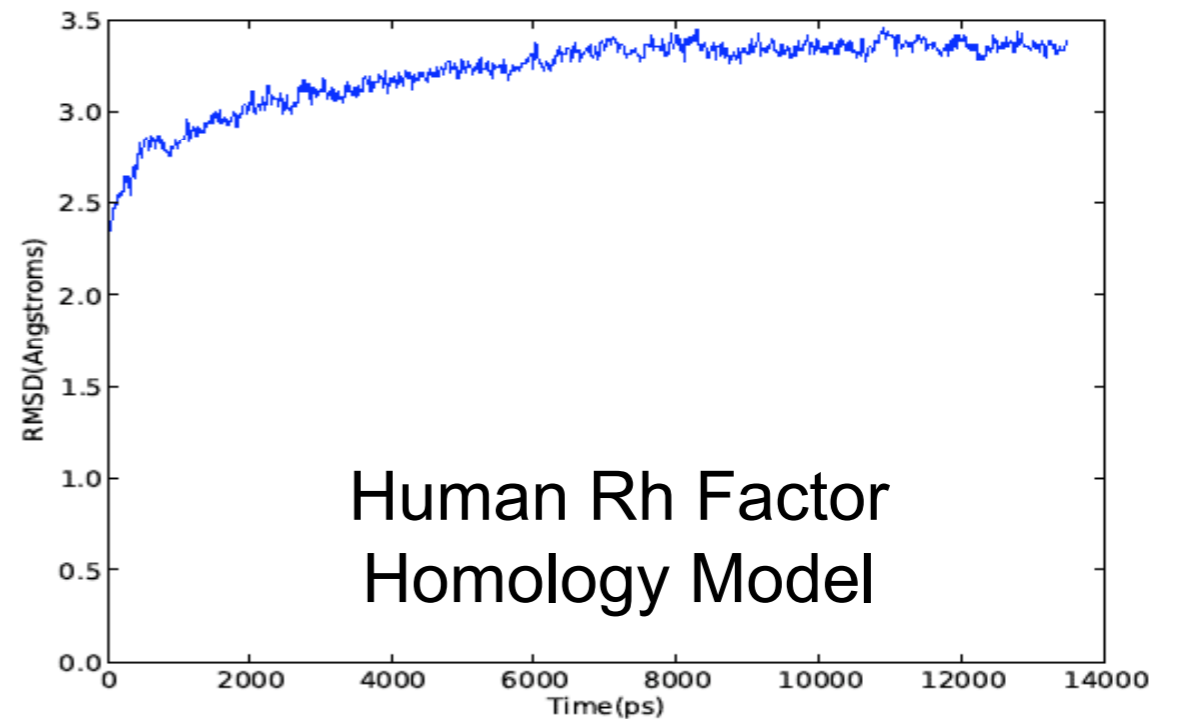
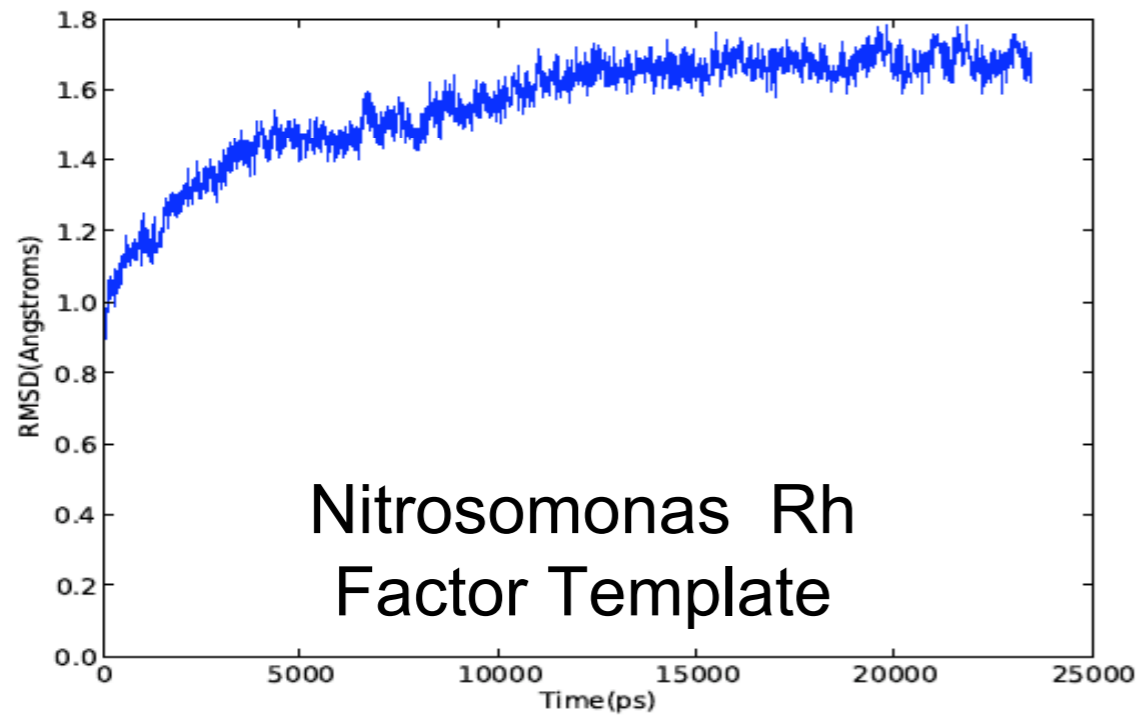
hRhBG/1-352    200 YYS LAASTLGT FALSALVGEDGR LDMVHIQNAALAGGVVVGTSS EMLT PFGALAAGFLAGTVSTLGYKFFTP ILESK 277
NeRH/1-351    206 LLALCGATLATYFLSALFH-KGKASIVDMANAALAGGVAIGSVCN-IVGPVGA FVIGLLGGAISVVG FVFIQPMLESK 281

hRhBG/1-352    278 FKVQDTCGVHNLHGMPGVLGALLGVLVAGLAQAMHQLFGLFVTLMFASVGGGLGGLLLKLP--FLDSPPD SQHYEDQ 352
NeRH/1-351    282 AKTIDTCGVHNLHGLPGLLGGSALIVPGIA-VAQLTGIGITLALALIGGVIAGALIKLT-----GTTKQAYEDS 351
  
```

With Franz Gruswitz and Ilya Chorny, Robert Stroud lab



# The homology model equilibrates!



Equilibration is monitored by the protein RMSD(t)

In a molecular dynamics simulation, the model is stable and the pore is recruiting  $\text{NH}_4$  ions

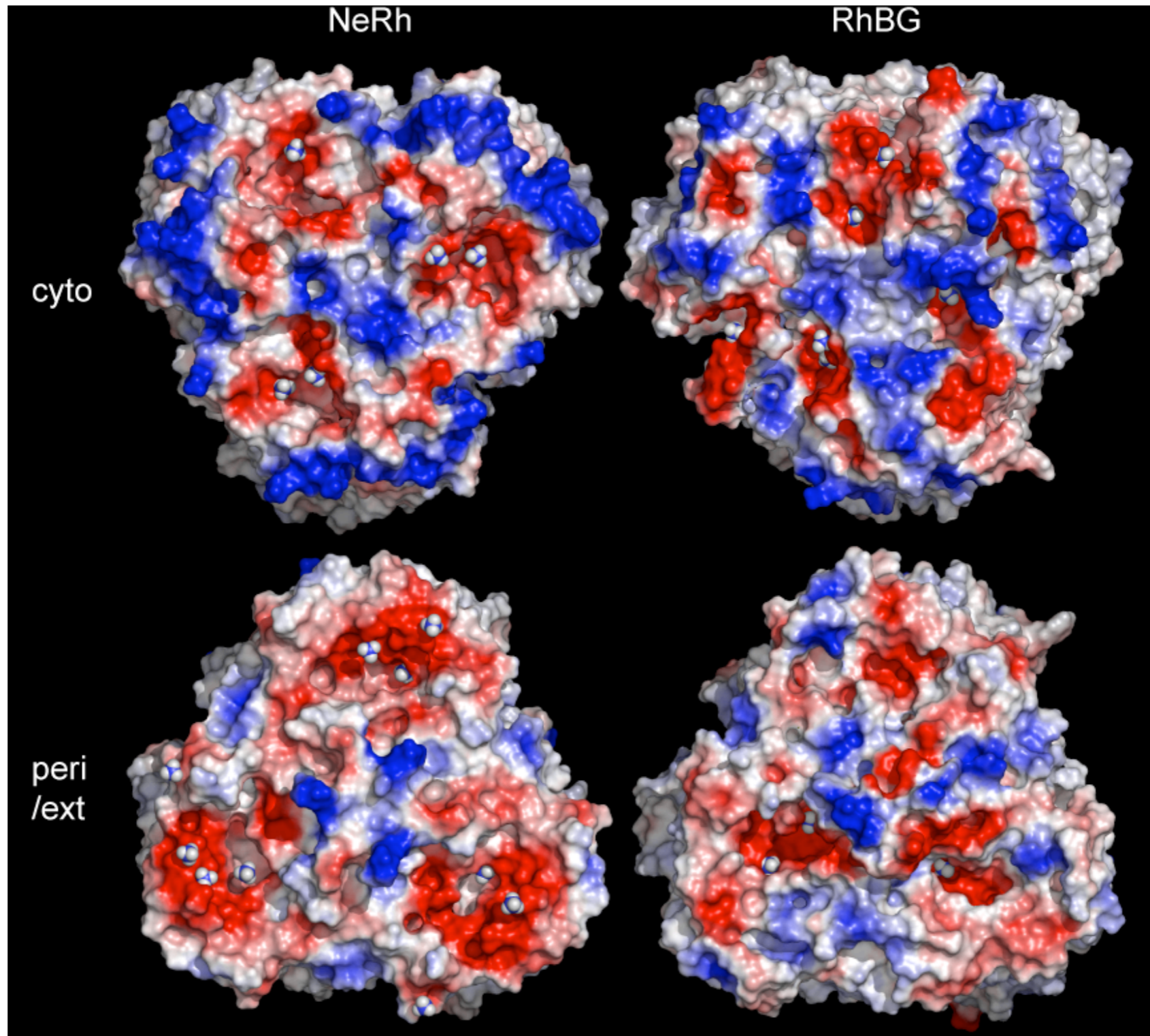


Figure from Ilya Chorny

# Characterizing genetic variation in human transporters

- **Comparative modeling of all human ABC transporter NBDs and 300+ point mutants**
- **Located 68 disease-associated mutations at putative interfaces in 10 human ABC transporters**
- **Developed a general tool for predicting the impact of point mutations on protein function**
- **Correctly predicted the *in vitro* function of five out of six previously uncharacterized ABC transporter variants found in a healthy population**

# **A taxonomic profile of the membrane protein universe**

- **Identified ~20,000 membrane proteins in 34 organisms and created a database of sequence profiles**
- **Human ABC transporter NBDs and TMDs likely evolved together on a single polypeptide chain**
- **Identified ~300 multidrug-resistance family members in pathogenic organisms**
- **Added to current estimates of ~600 membrane protein families with the identification of 51 putative new membrane protein families**
- **Target selection for the structural genomics of integral membrane proteins in yeast**

# Future Directions

- **Atomic level modeling of membrane proteins and substrates**
  - **NeRH and human RH factor proteins**
- **More genomes**
- **Better alignments of membrane proteins**
- **Organismal transport-omes**

# It takes a village

- **Andrej Sali lab, UCSF**
- Ursula Pieper
- Eashwar Narayanan
- Min-Yi Shen
- David Eramian
- **Rachel Karchin, JHU**
- Mark Breidenbach, Carolyn Bertozzi lab, Berkeley
- **Deanna Kroetz lab, UCSF**
- Leslie Chinn
- Hisayo Fukushima
- Jason Gow
- Nada Abla
- **Kathy Giacomini, UCSF**
- **Robert Stroud lab, UCSF**
- Franklin Hayes
- Franz Gruswitz