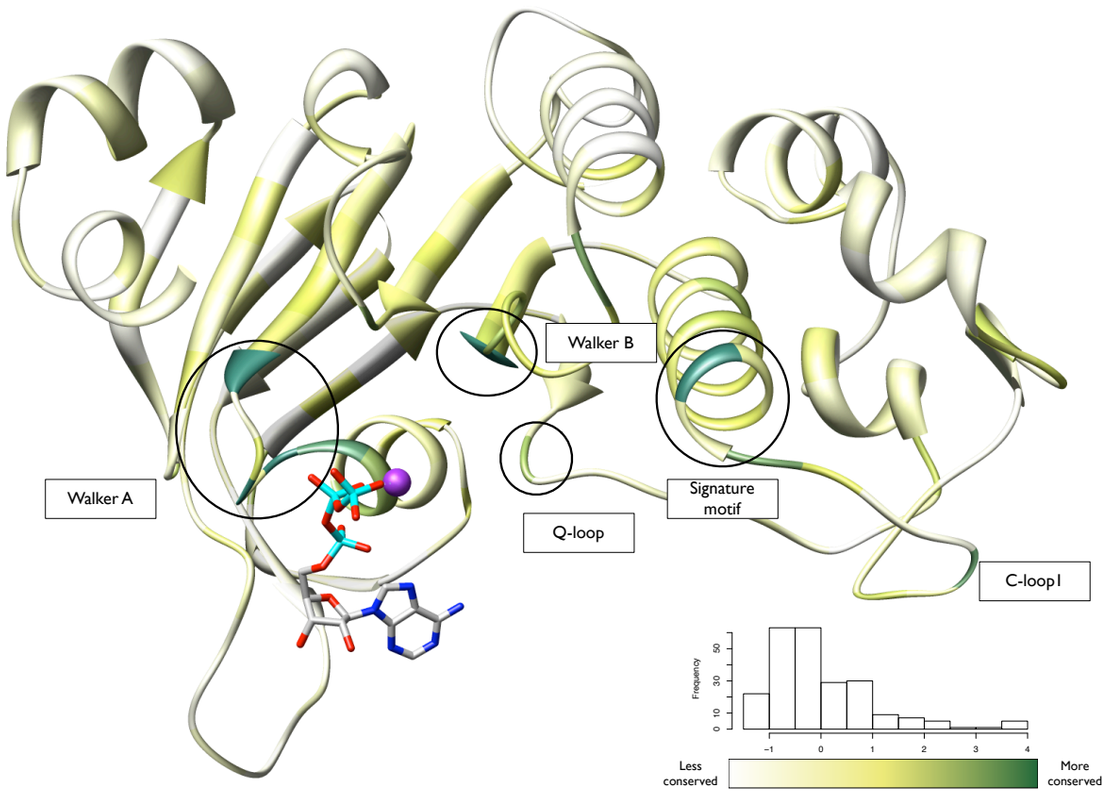
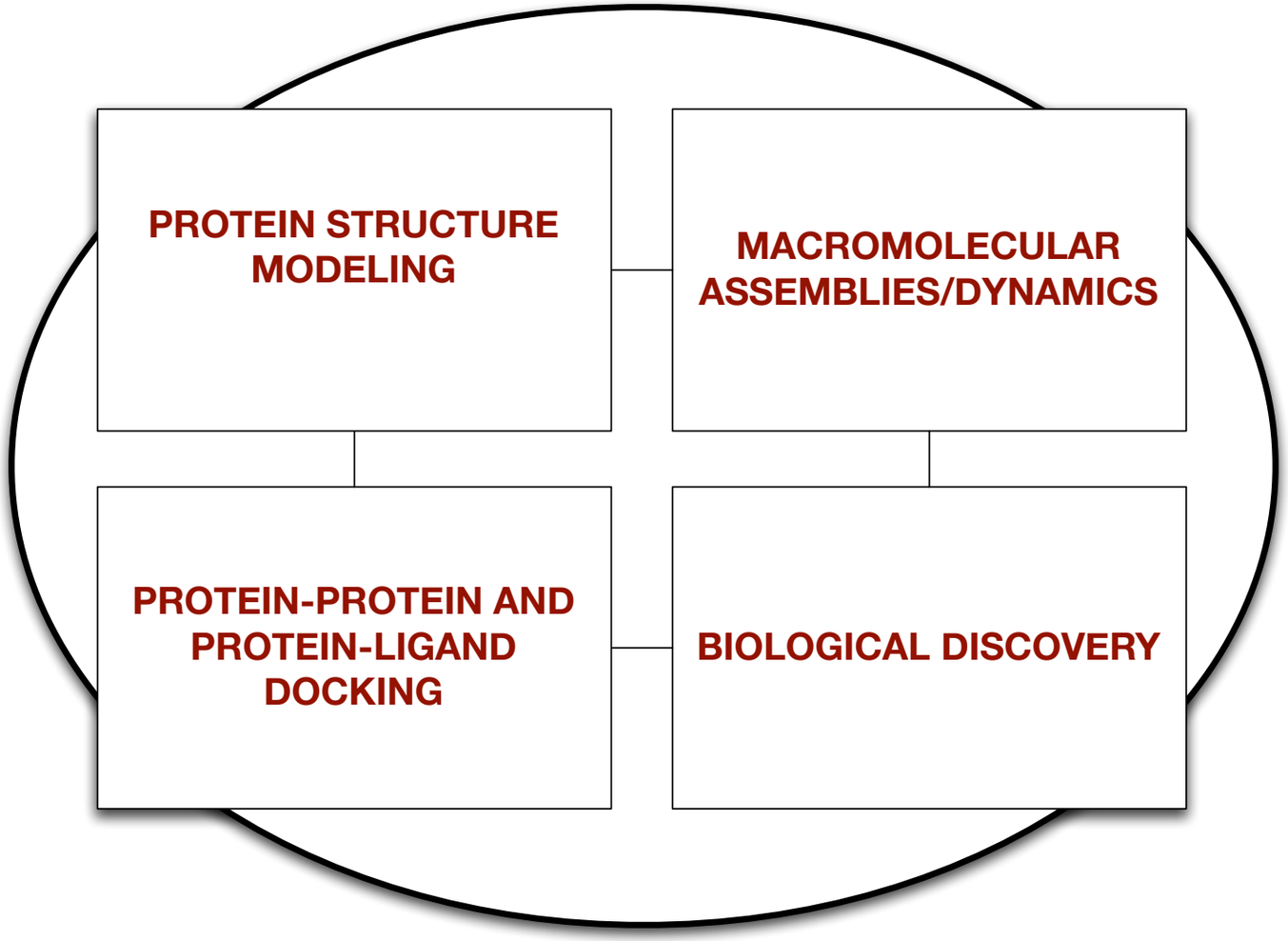
Functional hotspots revealed by mutational analysis of ABC transporters

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

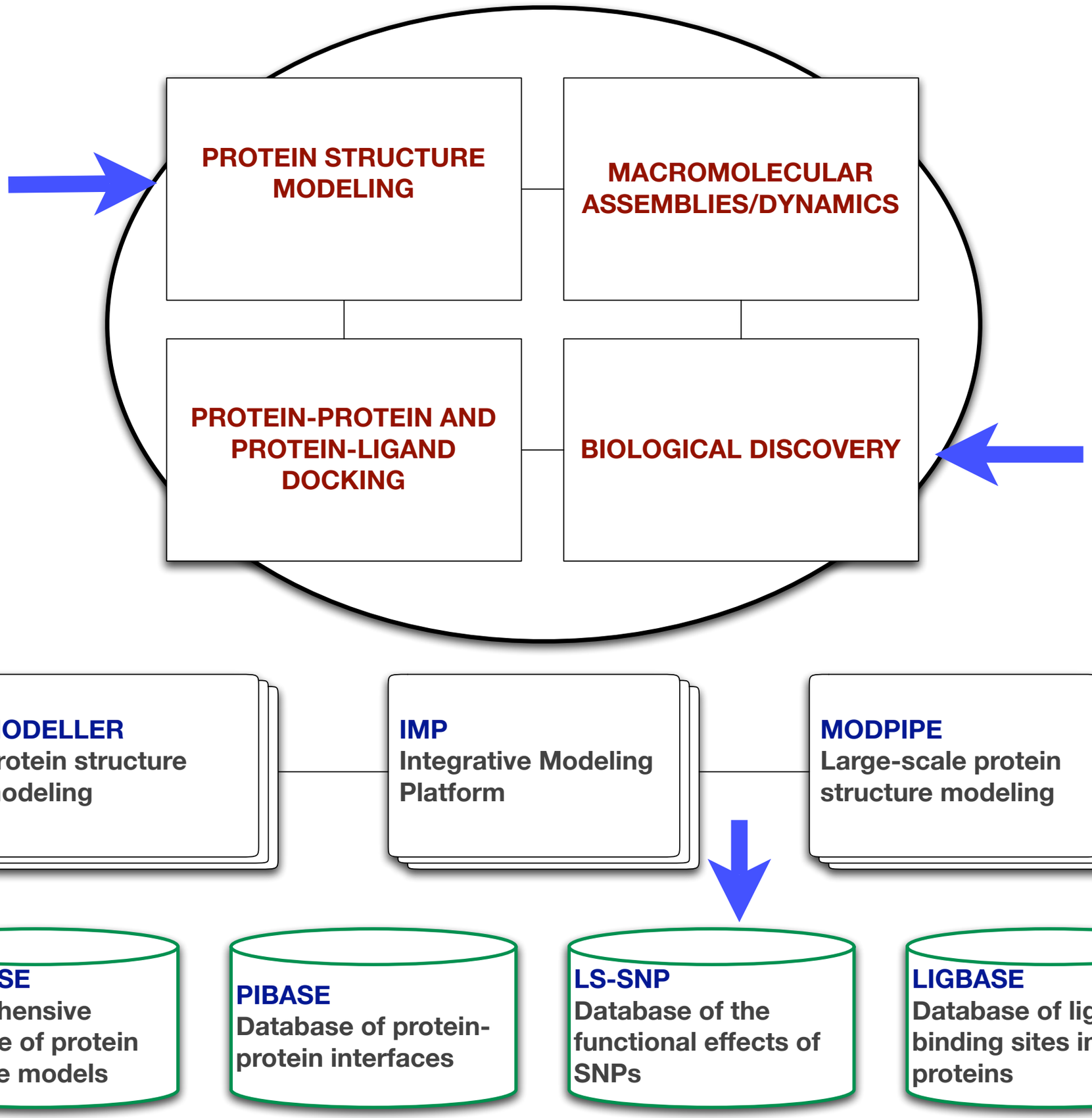
LS-SNP
Database of the functional effects of SNPs

LIGBASE
Database of ligand-binding sites in proteins

Slide courtesy of Eswar Narayanan

Overview of the Sali lab

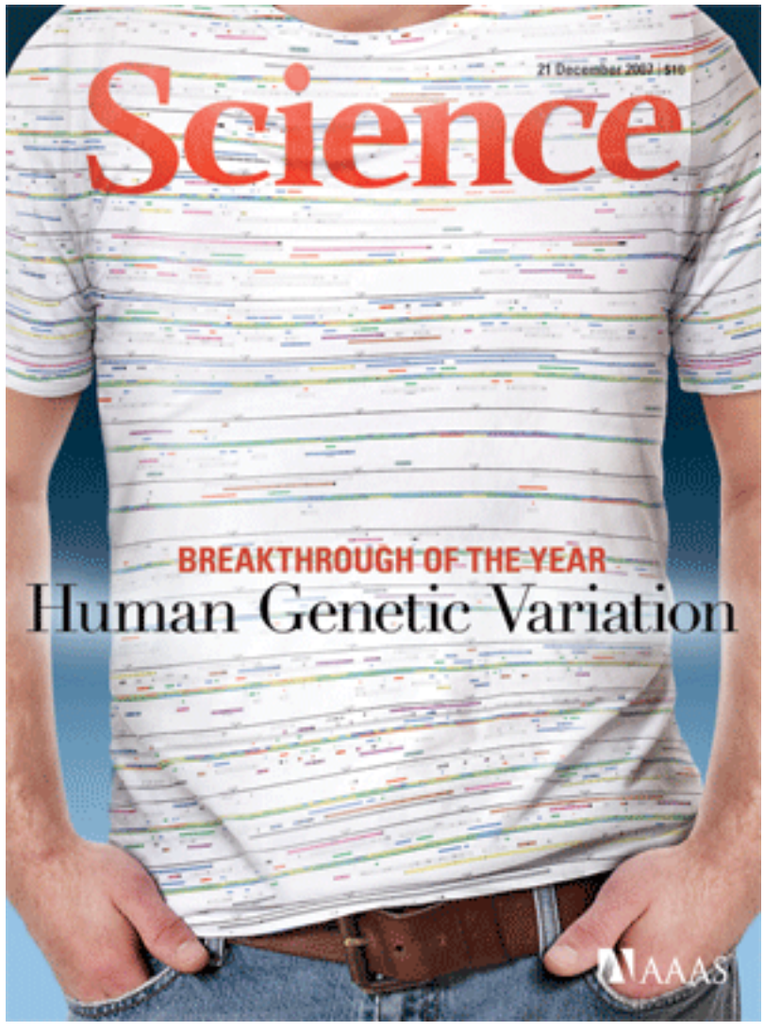
- predicting the structures of proteins
- determining the structures of macromolecular assemblies
- annotating the functions of proteins using structure



Slide courtesy of Eswar Narayanan

Genetic variation contributes to drug response

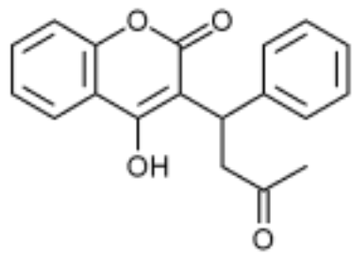
- Over 10 million common variants in the human genome
  - ~0.1% difference between two humans
  - ~1.5% difference between a human and a chimp
- Three widely prescribed drugs now have FDA ‘black box’ labels advising genetic testing



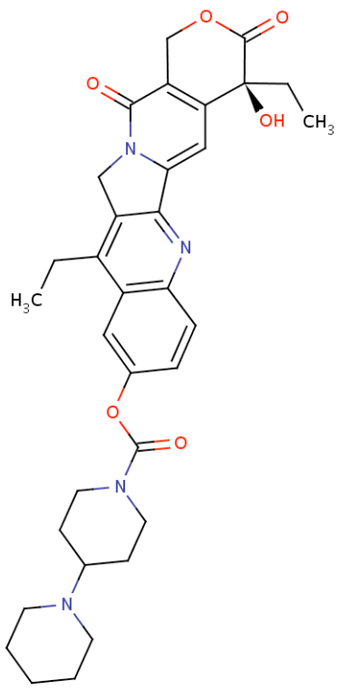
http://pharmacogenetics.ucsf.edu
http://www.hapmap.org/
http://www.sciencemag.org/cgi/content/full/308/5720/362

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

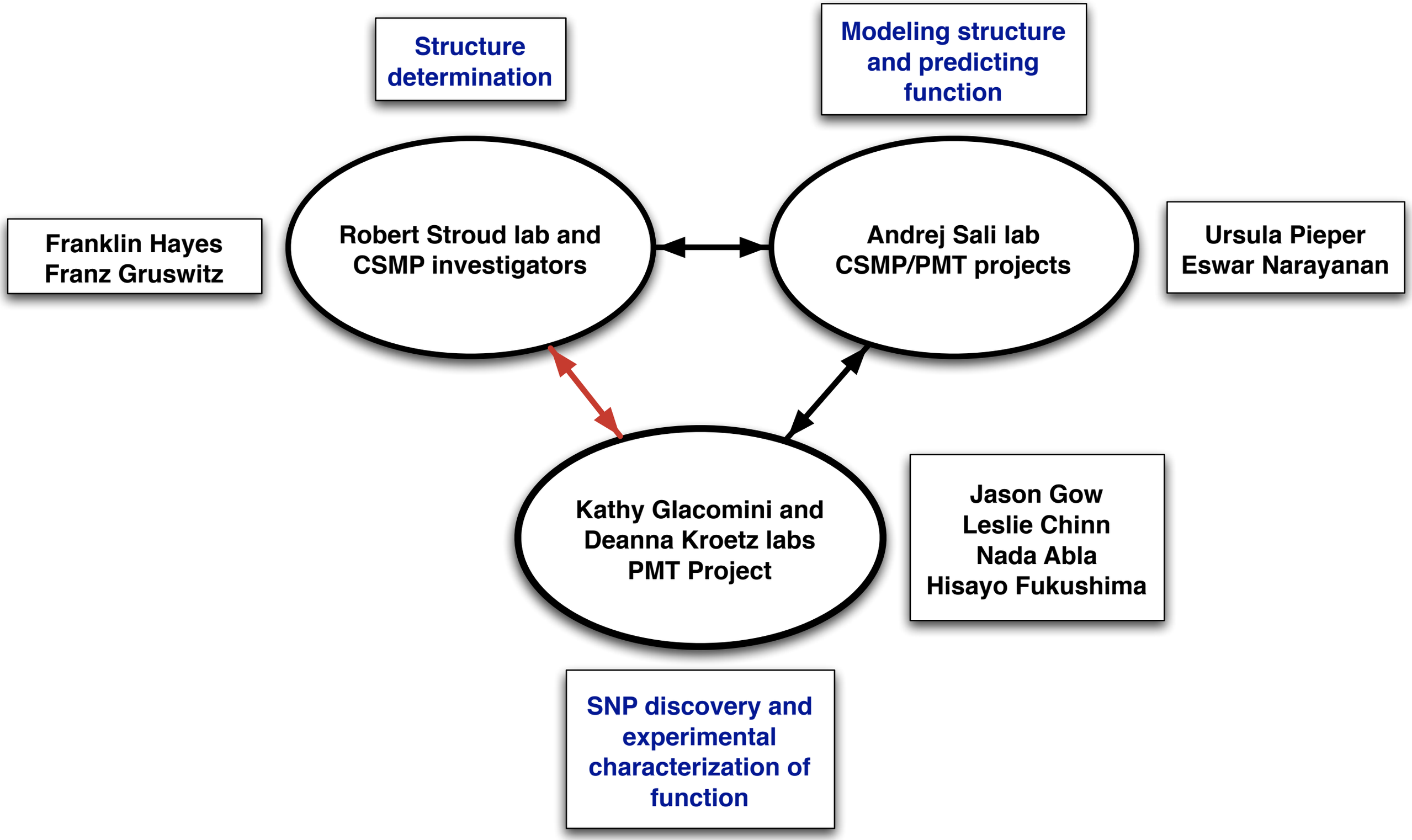
- **What is the genetic variation in the genes encoding membrane transporters in ethnically diverse human populations?**

- **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?**

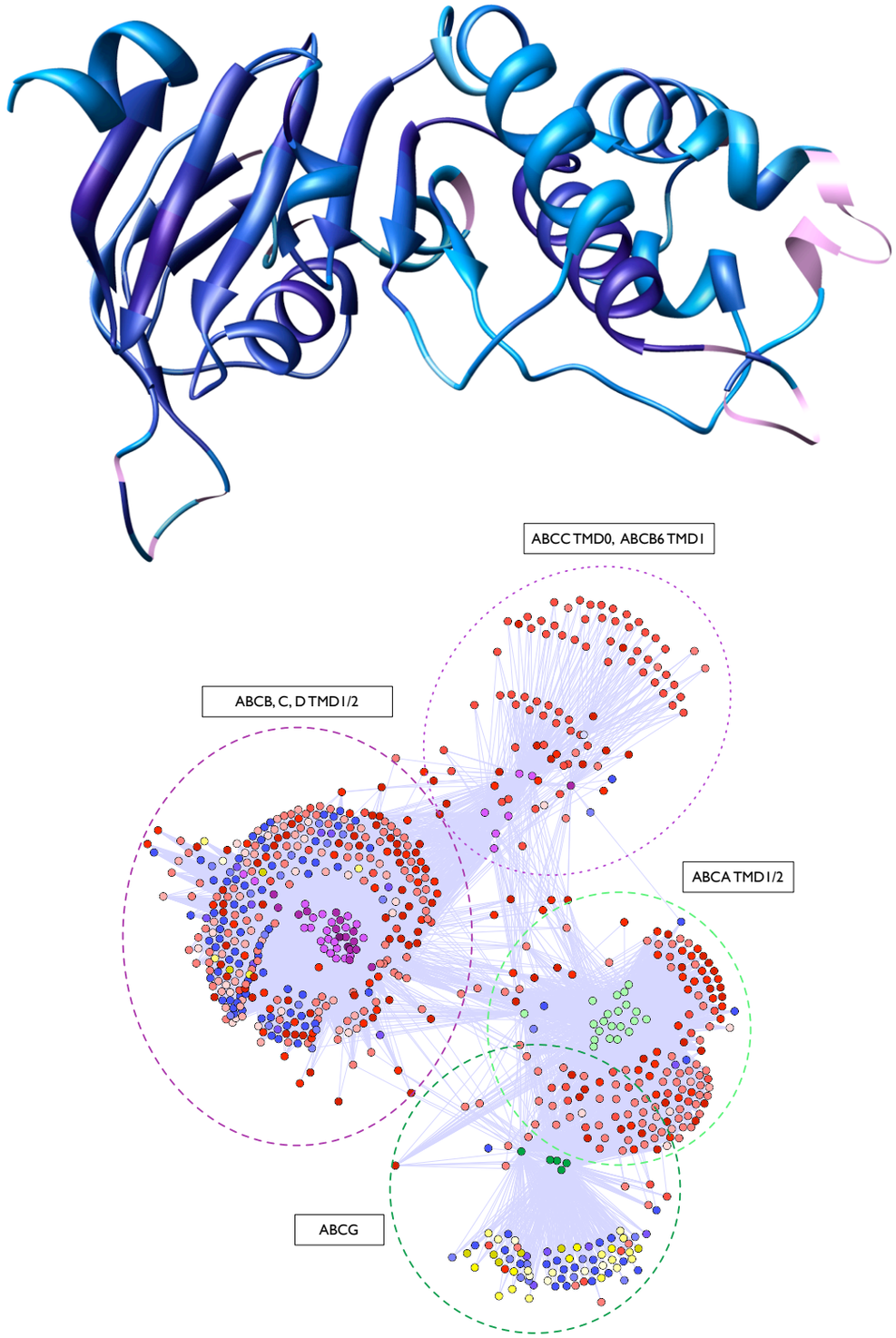
- **How does genetic variation in membrane transporters affect clinical drug response?**

Our collaborators let us tackle this problem from multiple perspectives

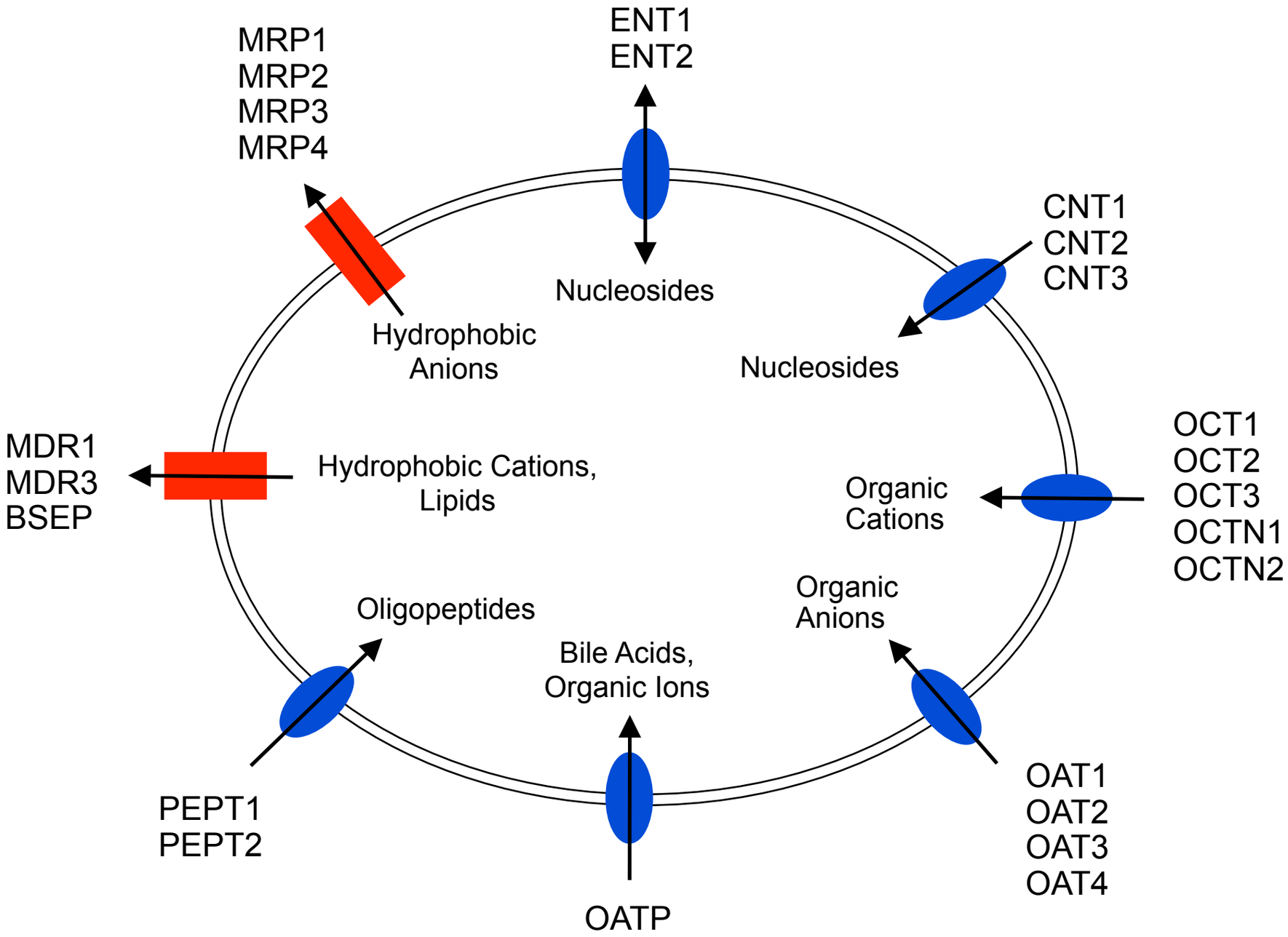


Using structure and evolutionary analysis to functionally characterize membrane proteins

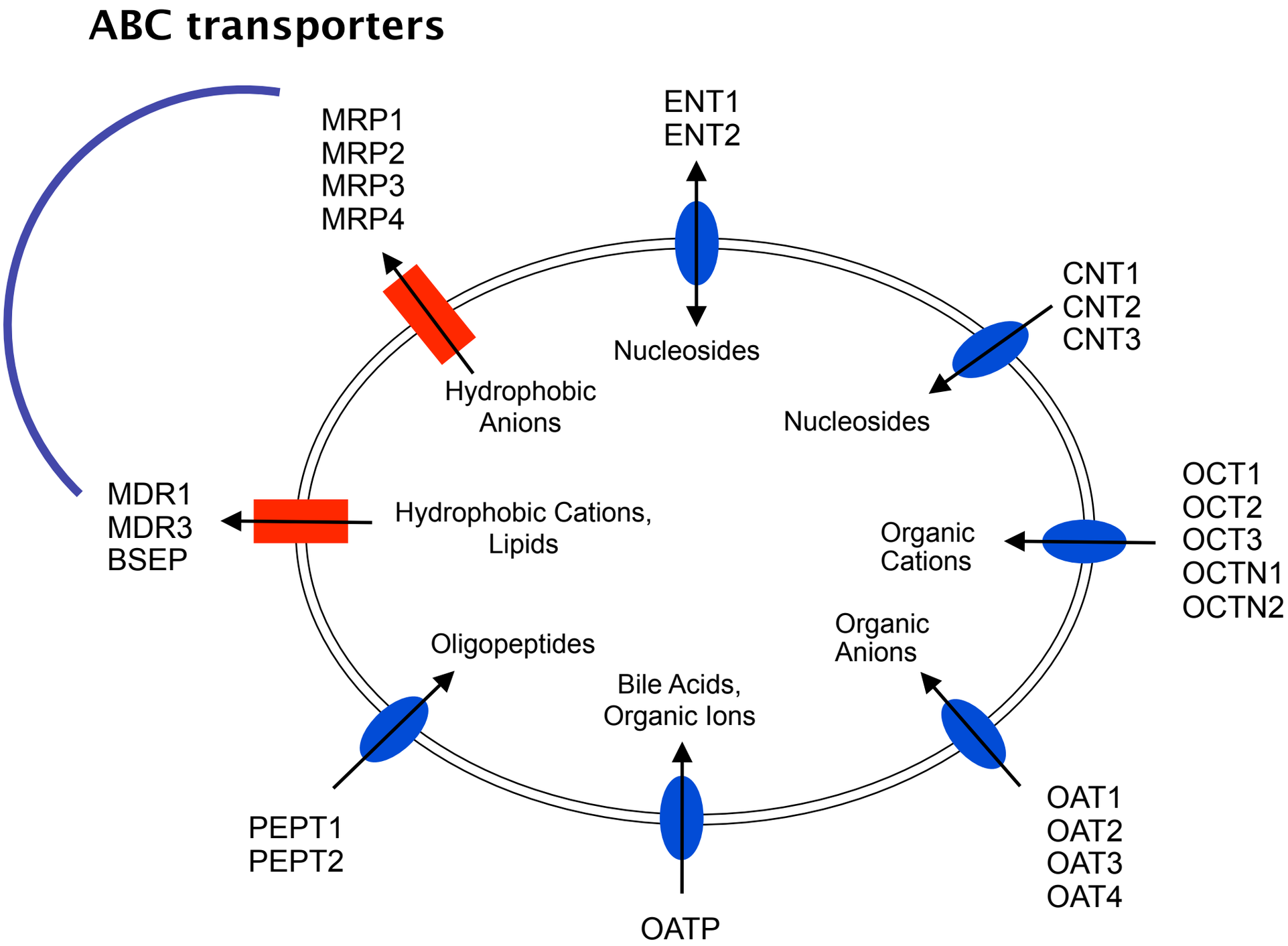
- **The impact of human genetic variation on membrane transporters**
  - **Domain interfaces and disease-associated mutants in ABC transporters**
  - **Predicting the effects of point mutations**
- **Membrane protein evolution**
  - **Taxonomy of membrane protein families**
  - **Domain organization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



Cataloguing variation in membrane transporters



Cataloguing variation in membrane transporters



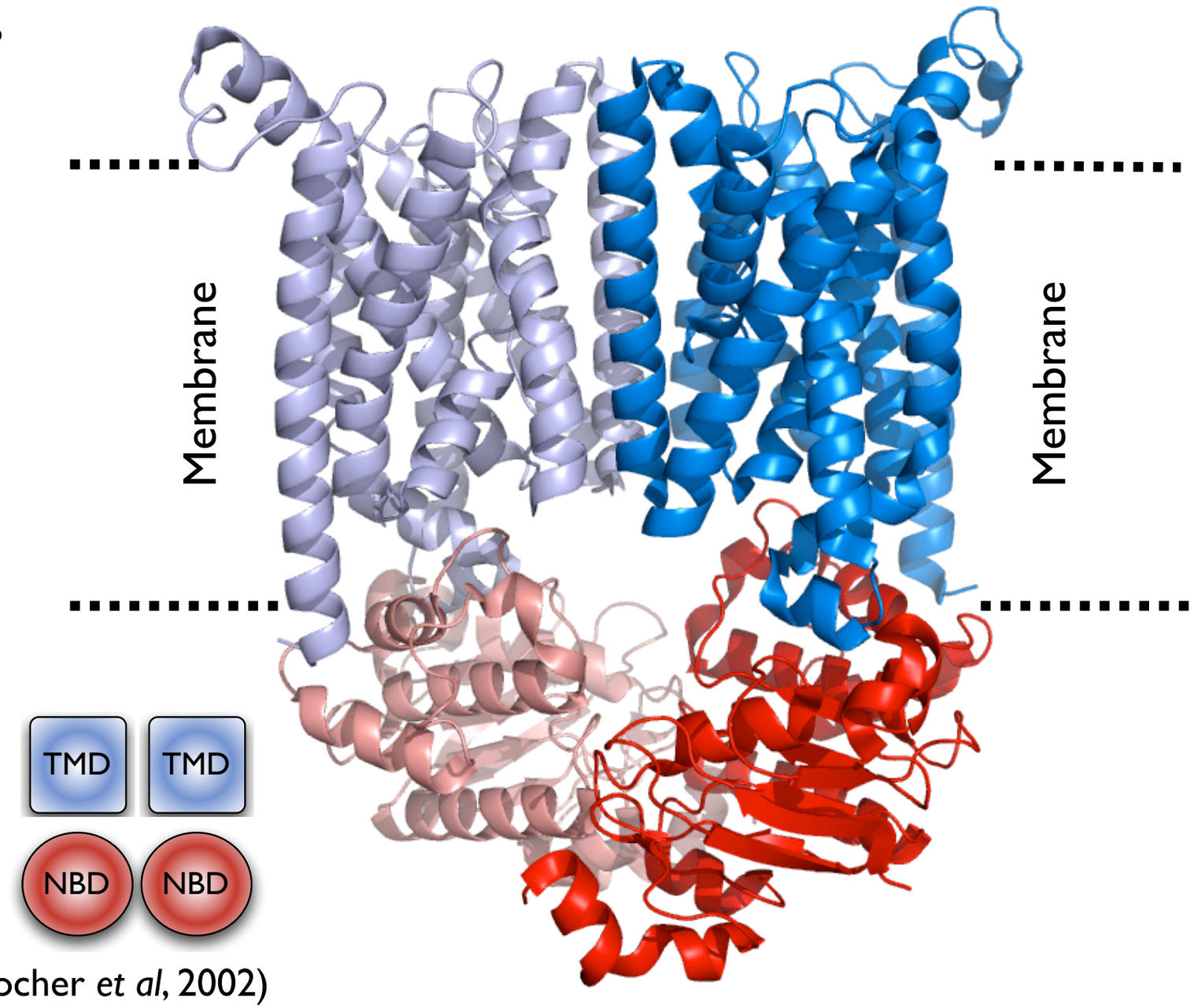
ABC transporters play roles in disease and drug response

Human ABC transporter	Function	Disease
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

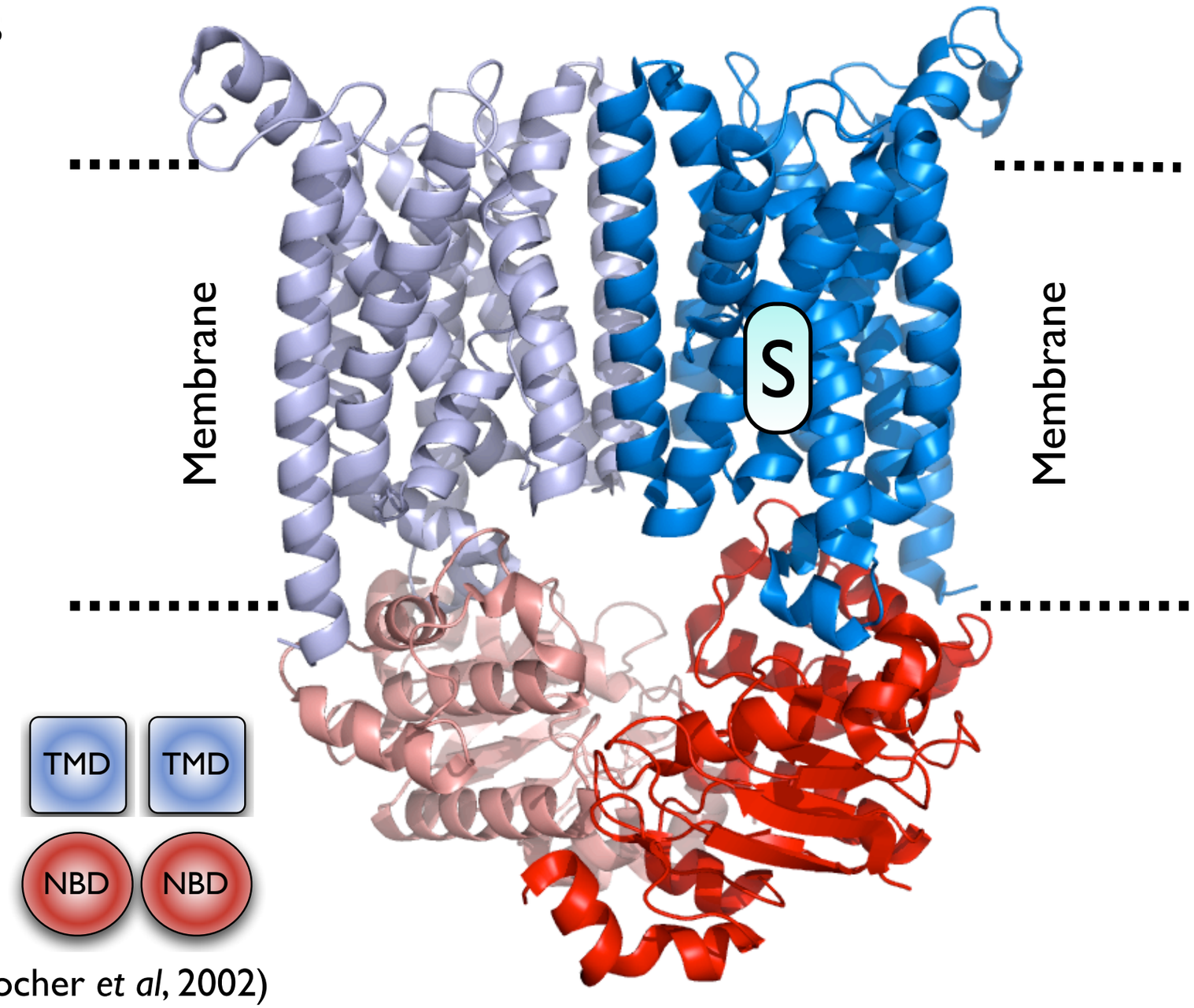
- In humans, ABC transporters are active export pumps
- Substrate binds in transmembrane domains (TMDs), ATP binds in nucleotide binding domains (NBDs)
- ATP binding, hydrolysis and release are coupled to substrate transport



Structure of the *E. coli* BtuCD transporter. (Locher *et al*, 2002)

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

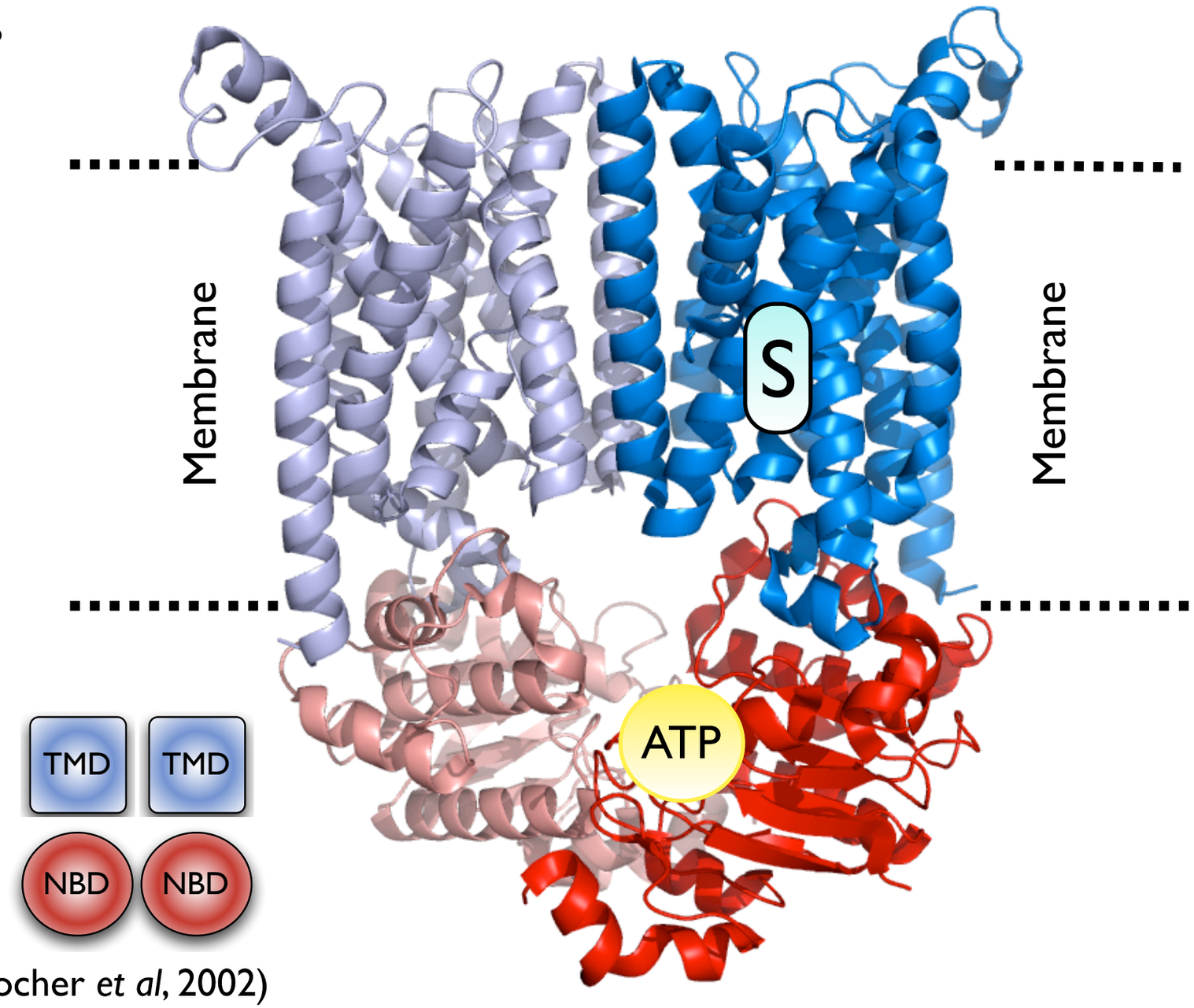
- In humans, ABC transporters are active export pumps
- Substrate binds in transmembrane domains (TMDs), ATP binds in nucleotide binding domains (NBDs)
- ATP binding, hydrolysis and release are coupled to substrate transport



Structure of the *E. coli* BtuCD transporter. (Locher *et al*, 2002)

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

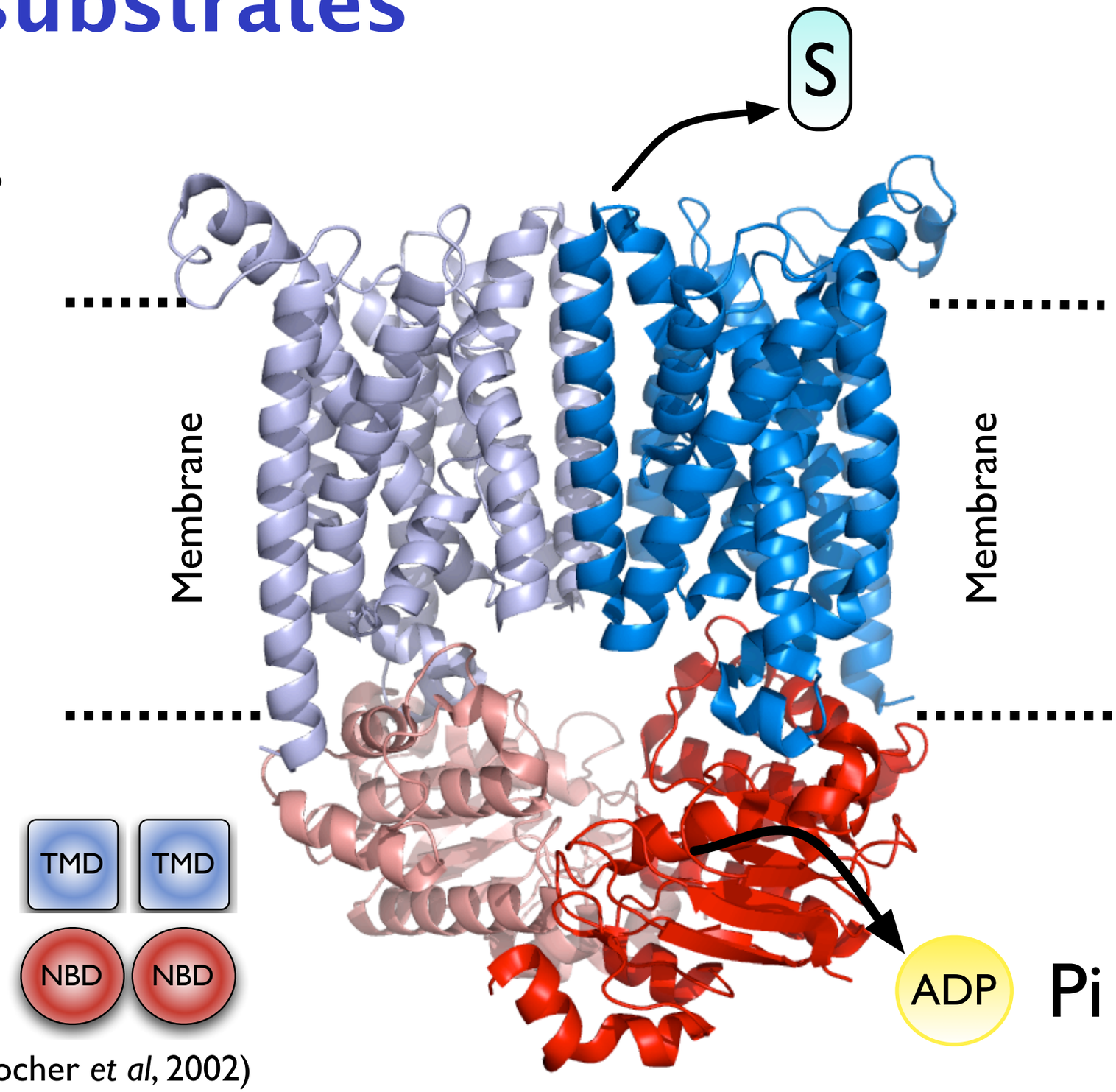
- In humans, ABC transporters are active export pumps
- Substrate binds in transmembrane domains (TMDs), ATP binds in nucleotide binding domains (NBDs)
- ATP binding, hydrolysis and release are coupled to substrate transport



Structure of the *E. coli* BtuCD transporter. (Locher *et al*, 2002)

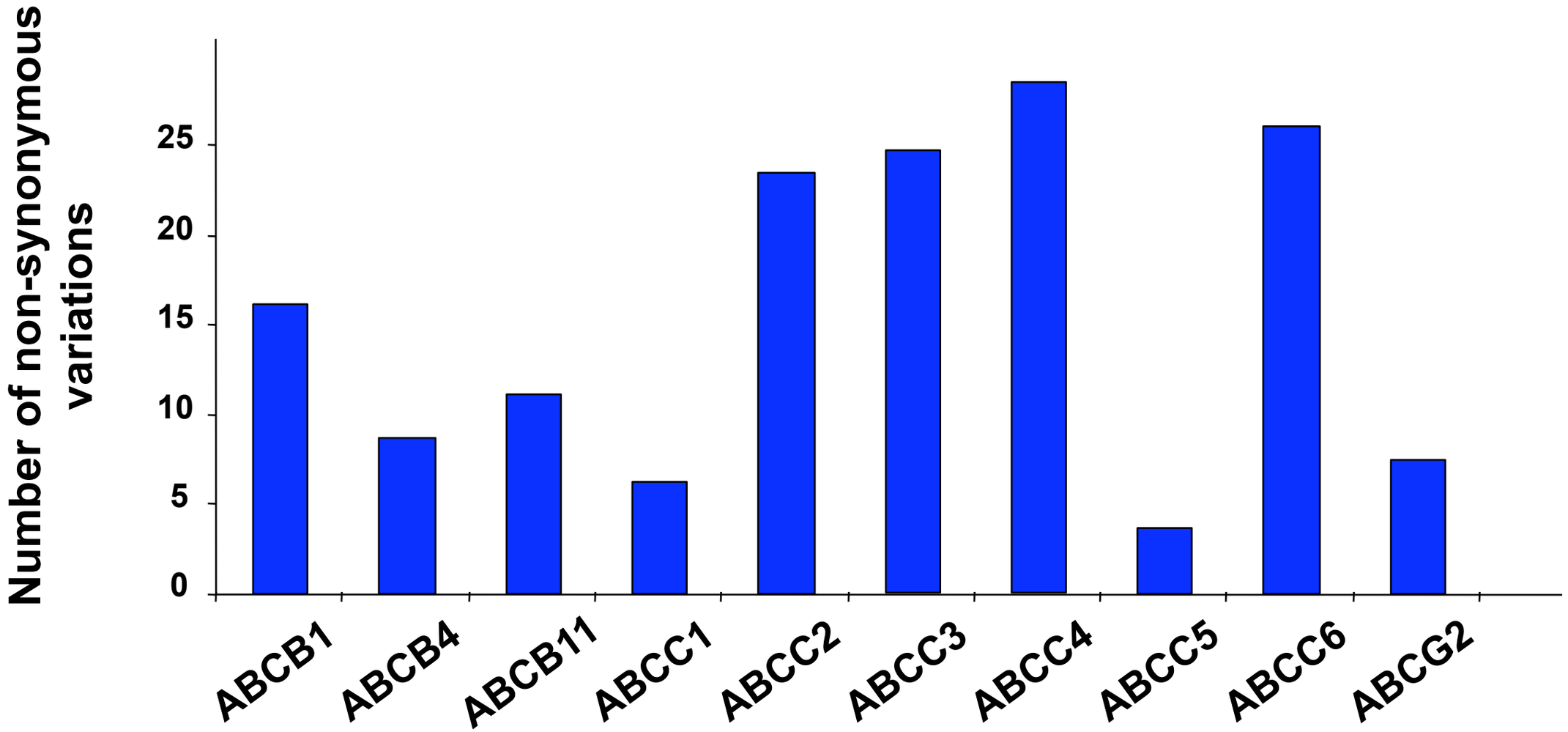
ABC transporters are membrane proteins that bind a wide range of substrates

- 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



Structure of the *E. coli* BtuCD transporter. (Locher *et al*, 2002)

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

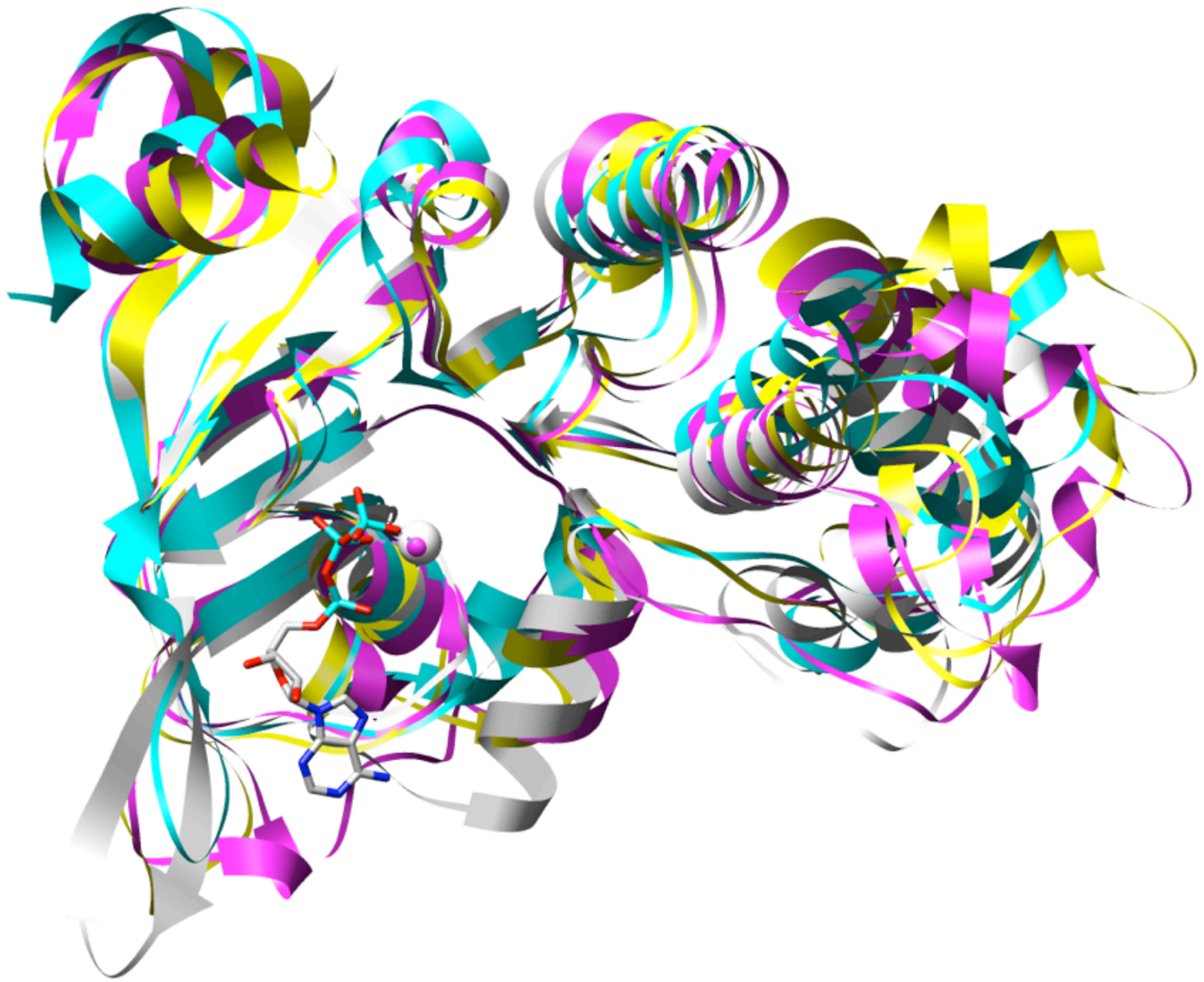
- Some proteins and systems are difficult to get functional data for; data for 10s of SNPs but not 100s or more

- 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

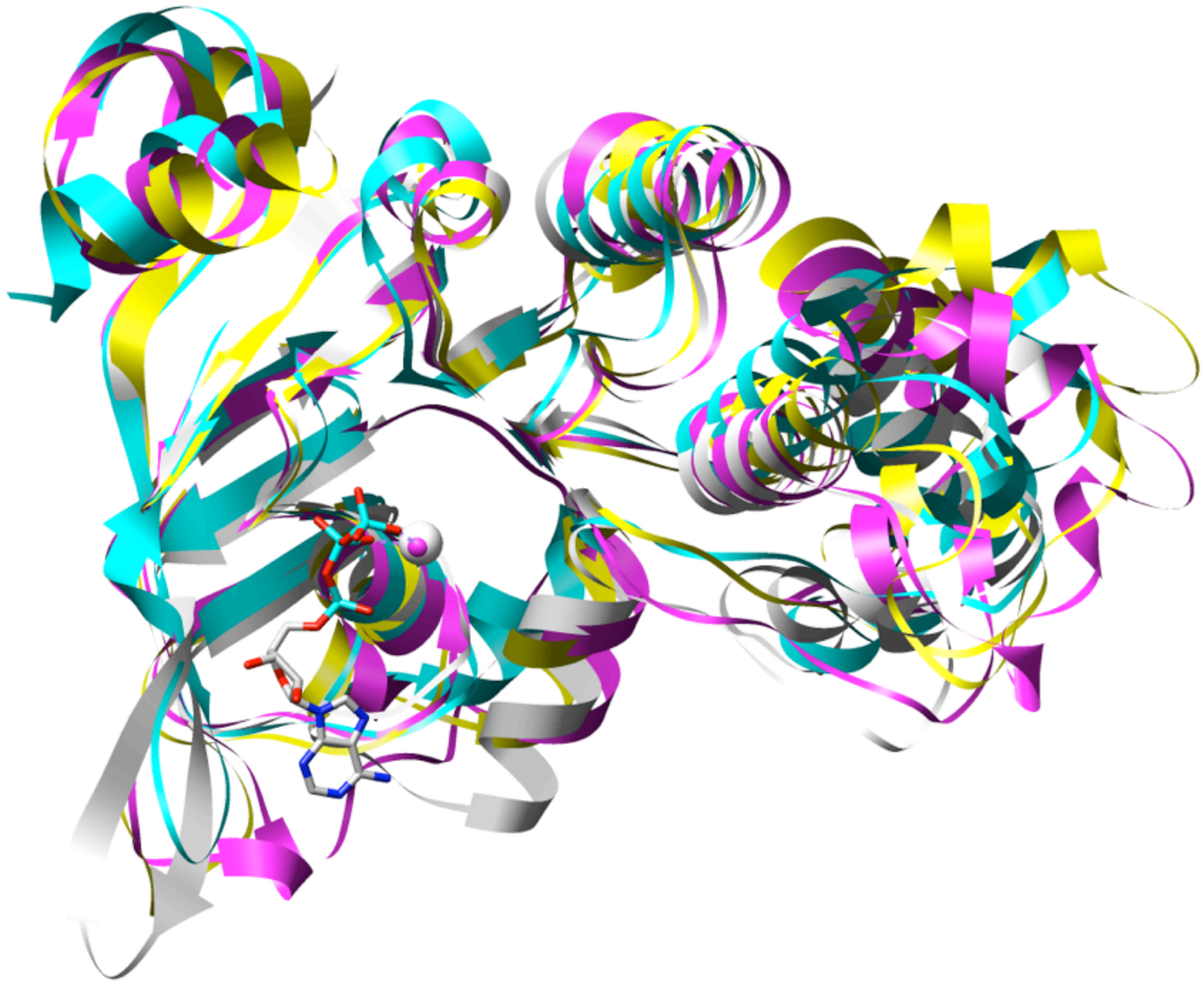


NBD

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**

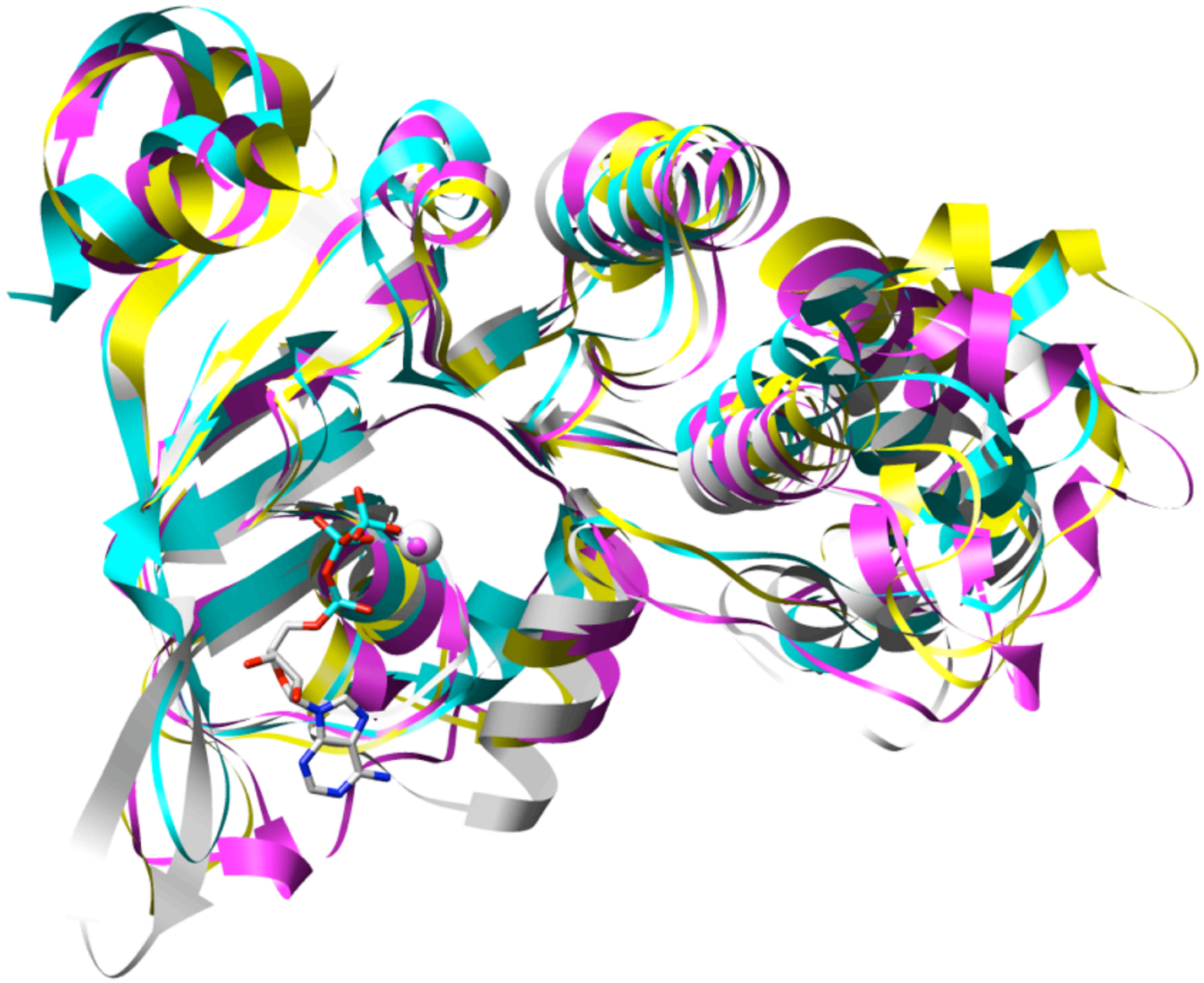


NBD structures of four ABC NBDs

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**

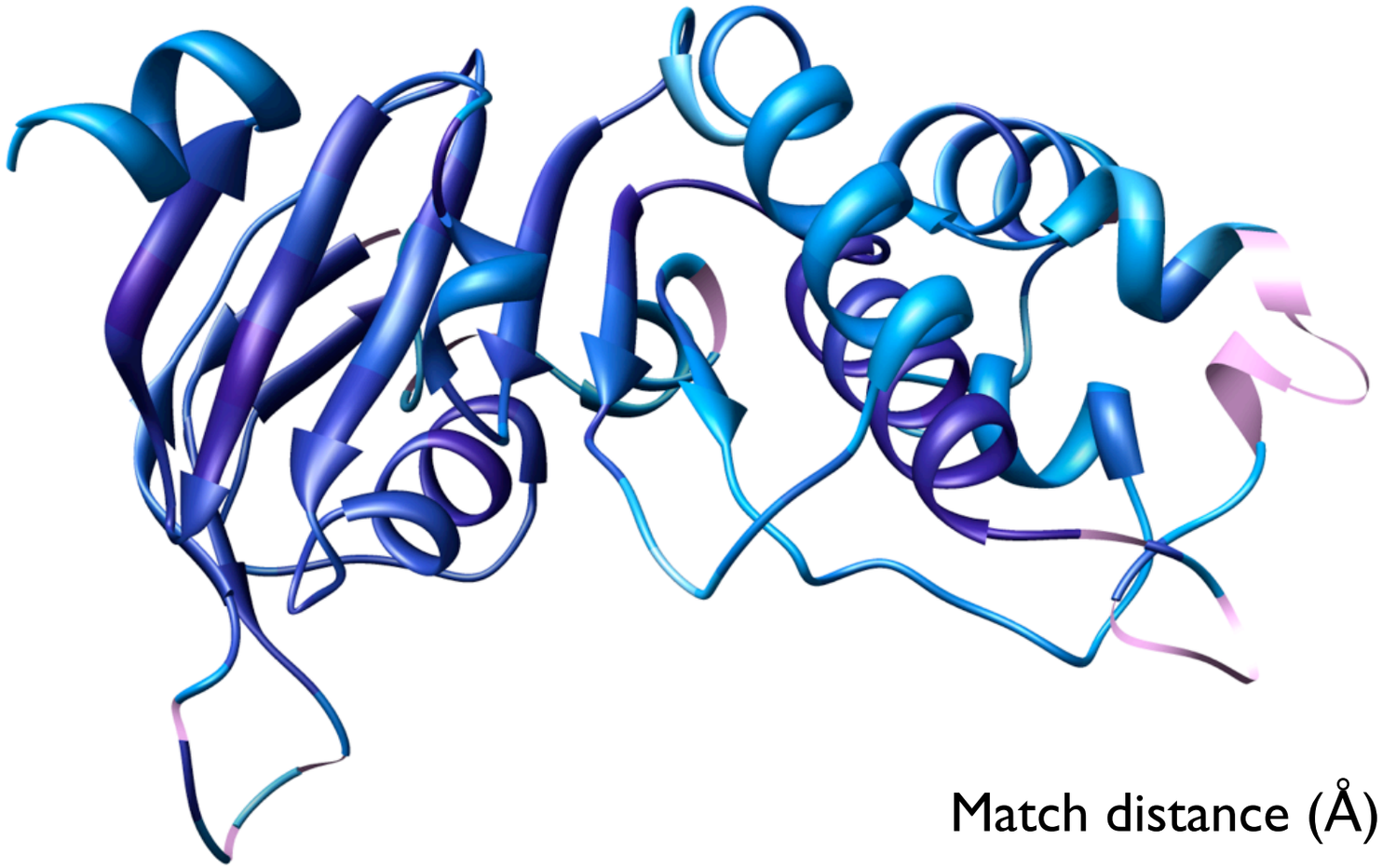


NBD structures of four ABC NBDs

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

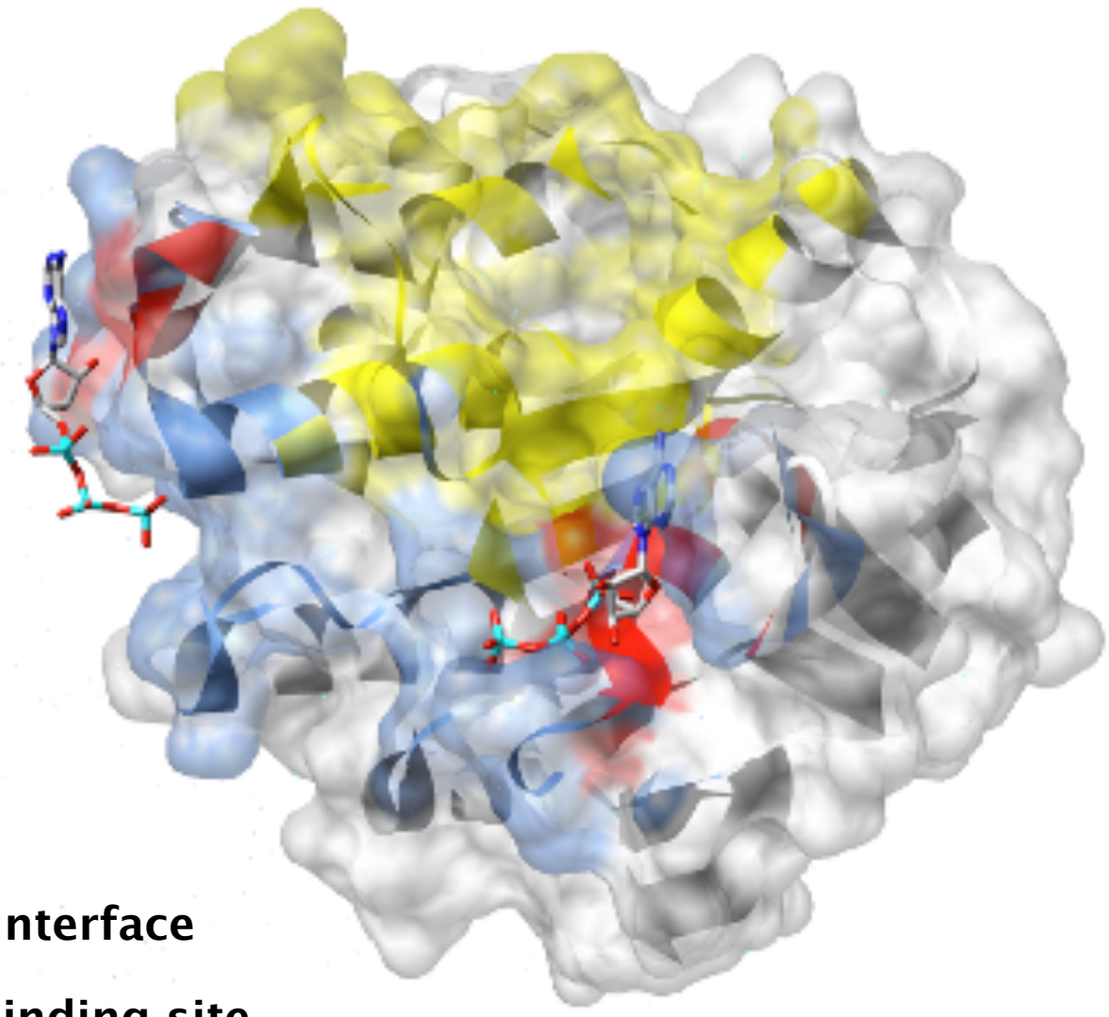
- **The overall fold of the NBDs is highly conserved across organisms**

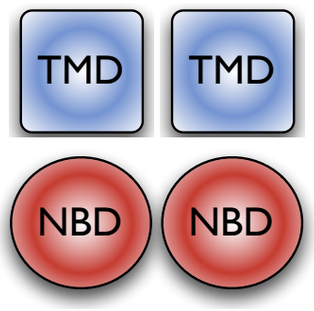
- **This enables us to model human NBDs and nsSNPs based on homologs with known structure**



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?





TMD interface

ATP-binding site

NBD interface

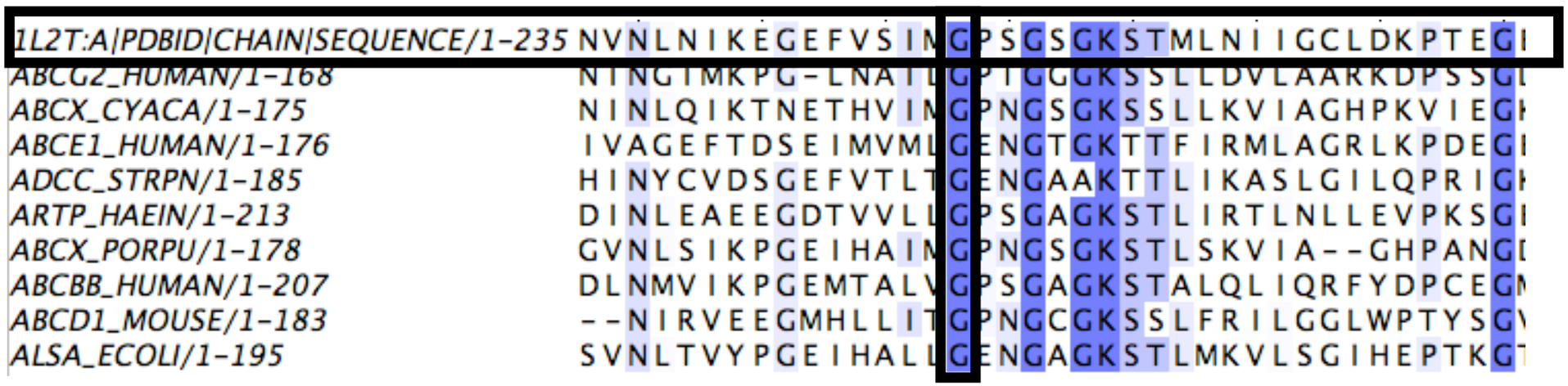
from *M. jannaschii*

Kelly L, Karchin R, Sali A.
Protein interactions and disease phenotypes in the ABC transporter superfamily.
Pac Symp Biocomput. 2007;:51-63.

Does sequence conservation vary at ABC transporter domain interfaces?

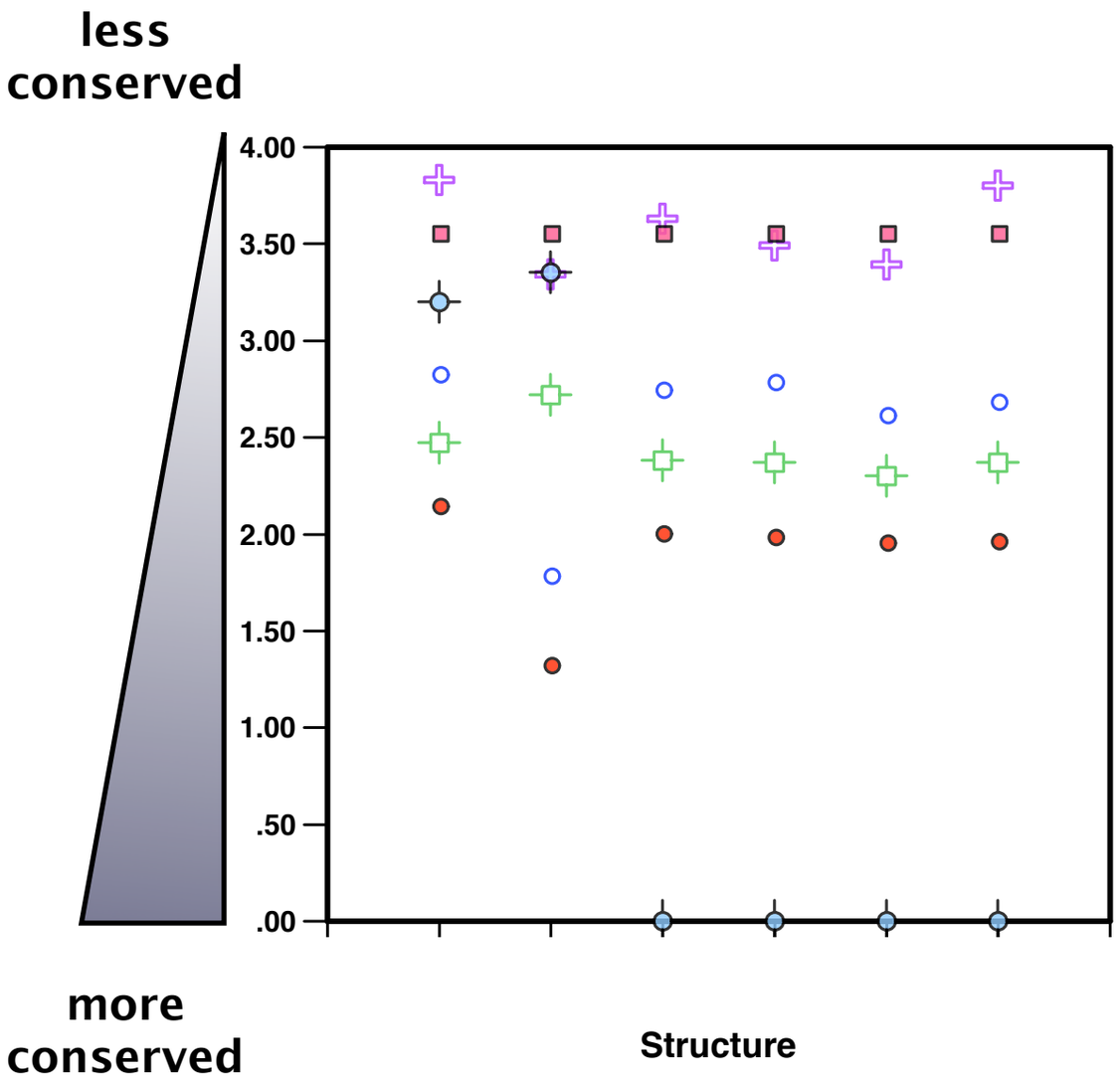
- **Automatic multiple sequence alignment profiles for each of six structures and for the NBDs of each human ABC transporter**
- **We calculate a measure of entropy in the alignment**

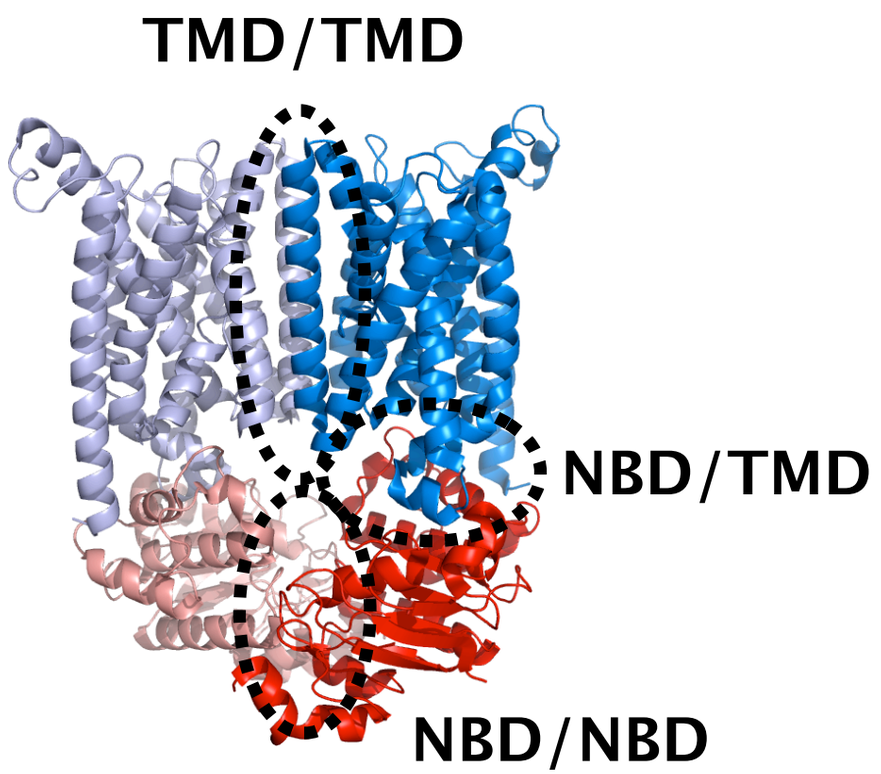
$$H = -\sum_{aa=1}^{20} P_{aa} \log_2 P_{aa}$$



http://salilab.org/modeller

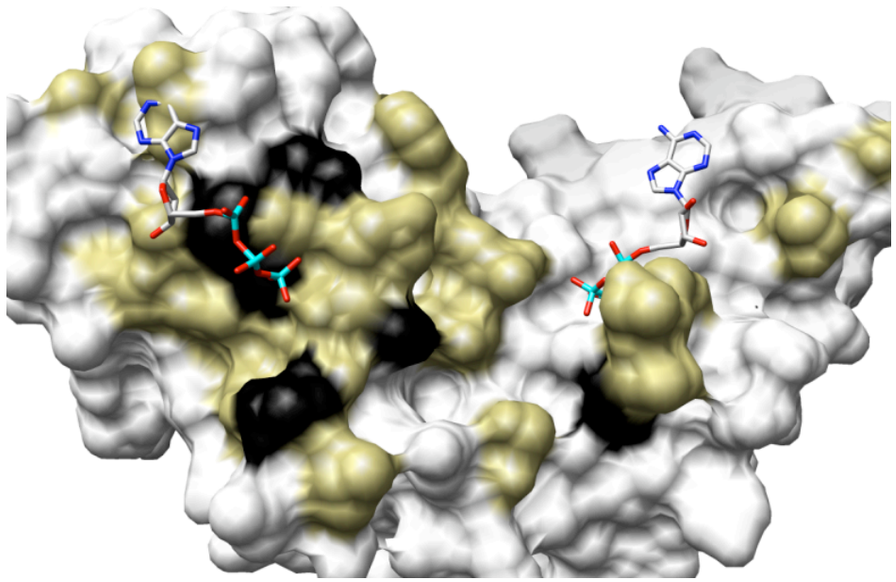
Sequence conservation varies between the three interfaces





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

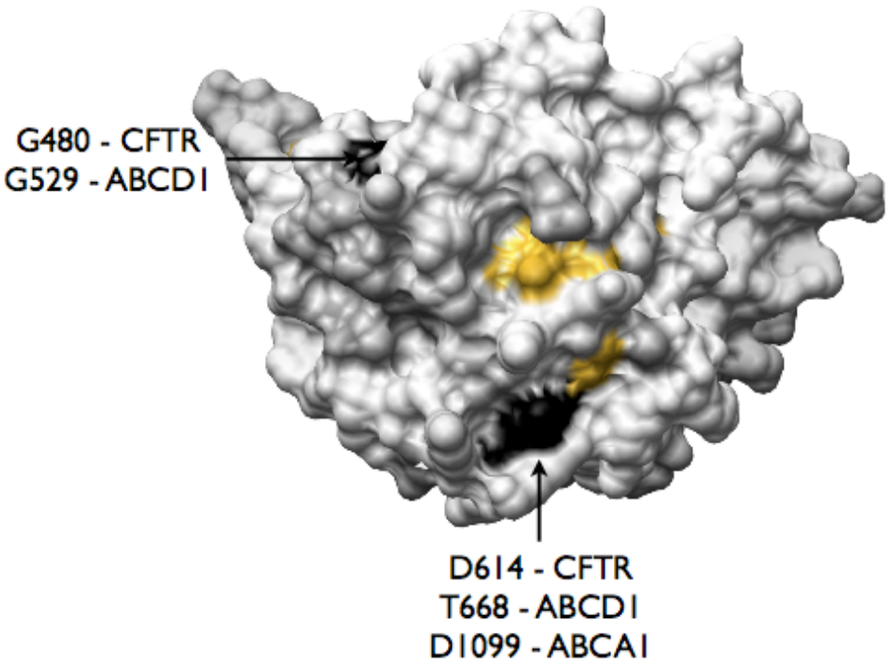
Some disease-associated mutants affect domain interactions



conserved

disease

Human cystic fibrosis-associated transproter CFTR NBD1 interface surface



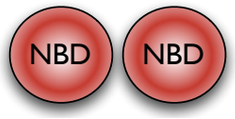
180°

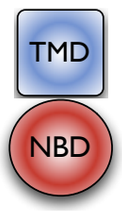
conserved

disease

Exposed surface of bacterial ABC transporter NBD

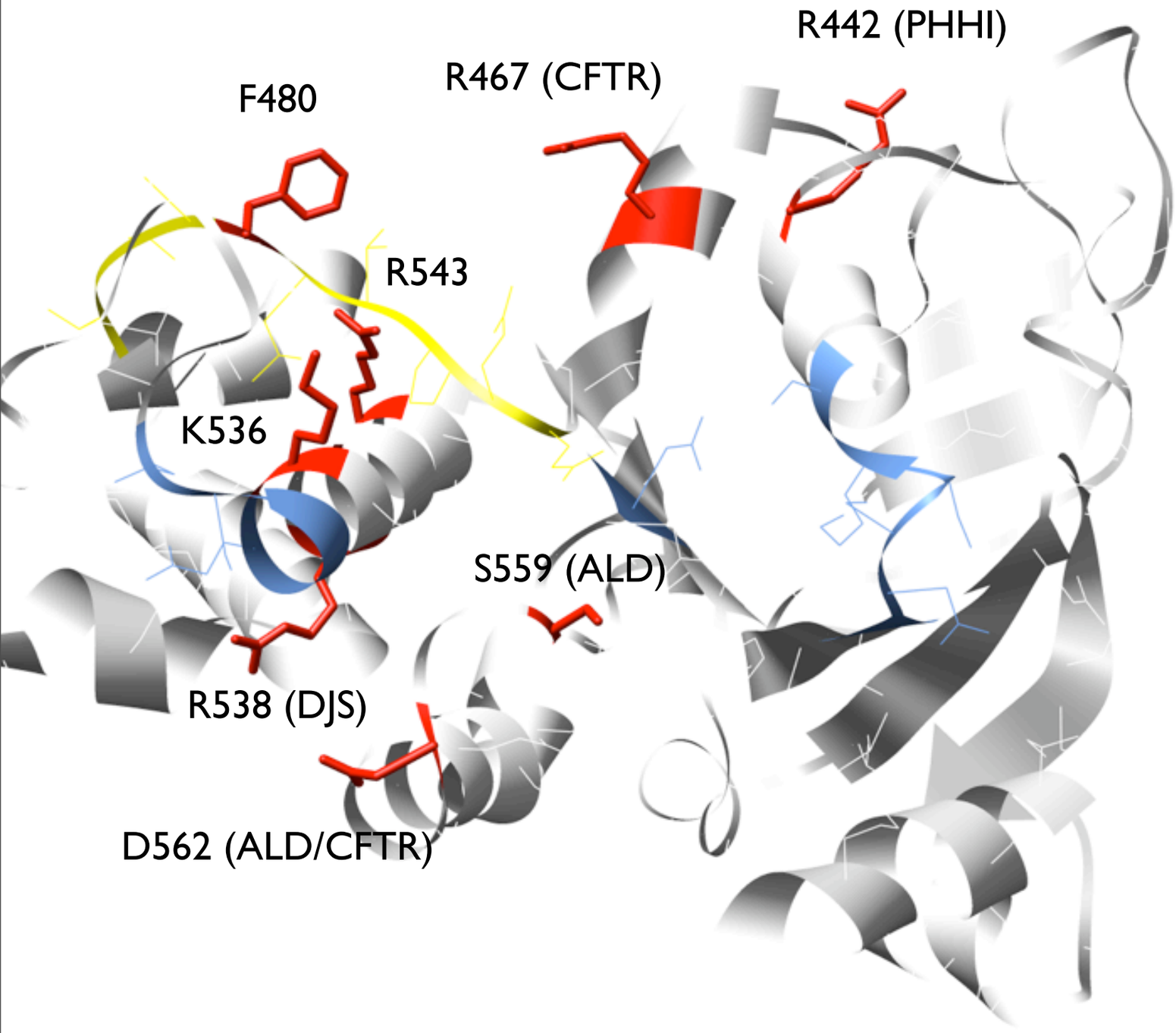
We found 68 disease-associated positions at putative interfaces

- 10 transporters from four out of seven ABC subfamilies are represented
- 38 were at the 

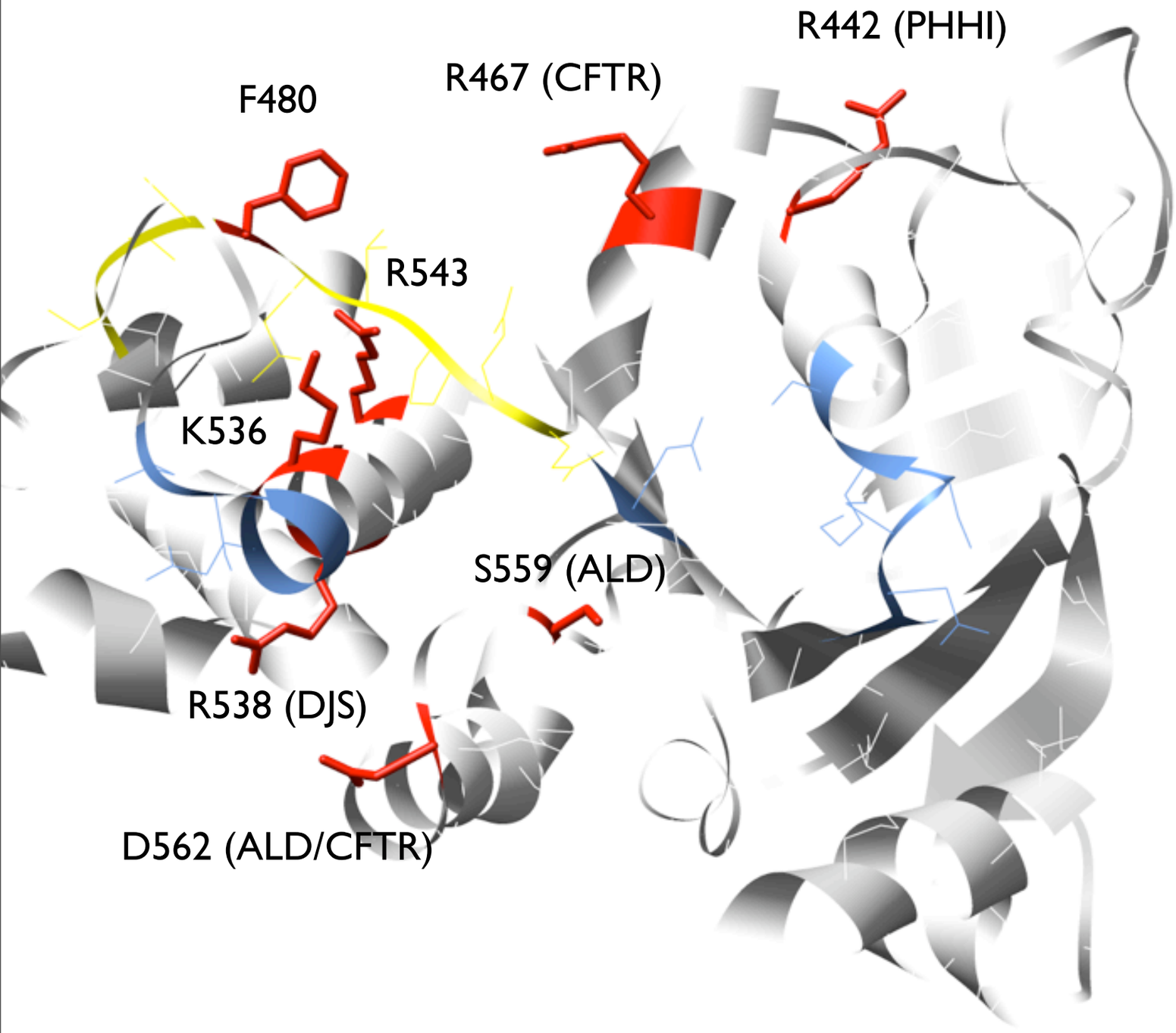
interface
- 30 were at the 

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?

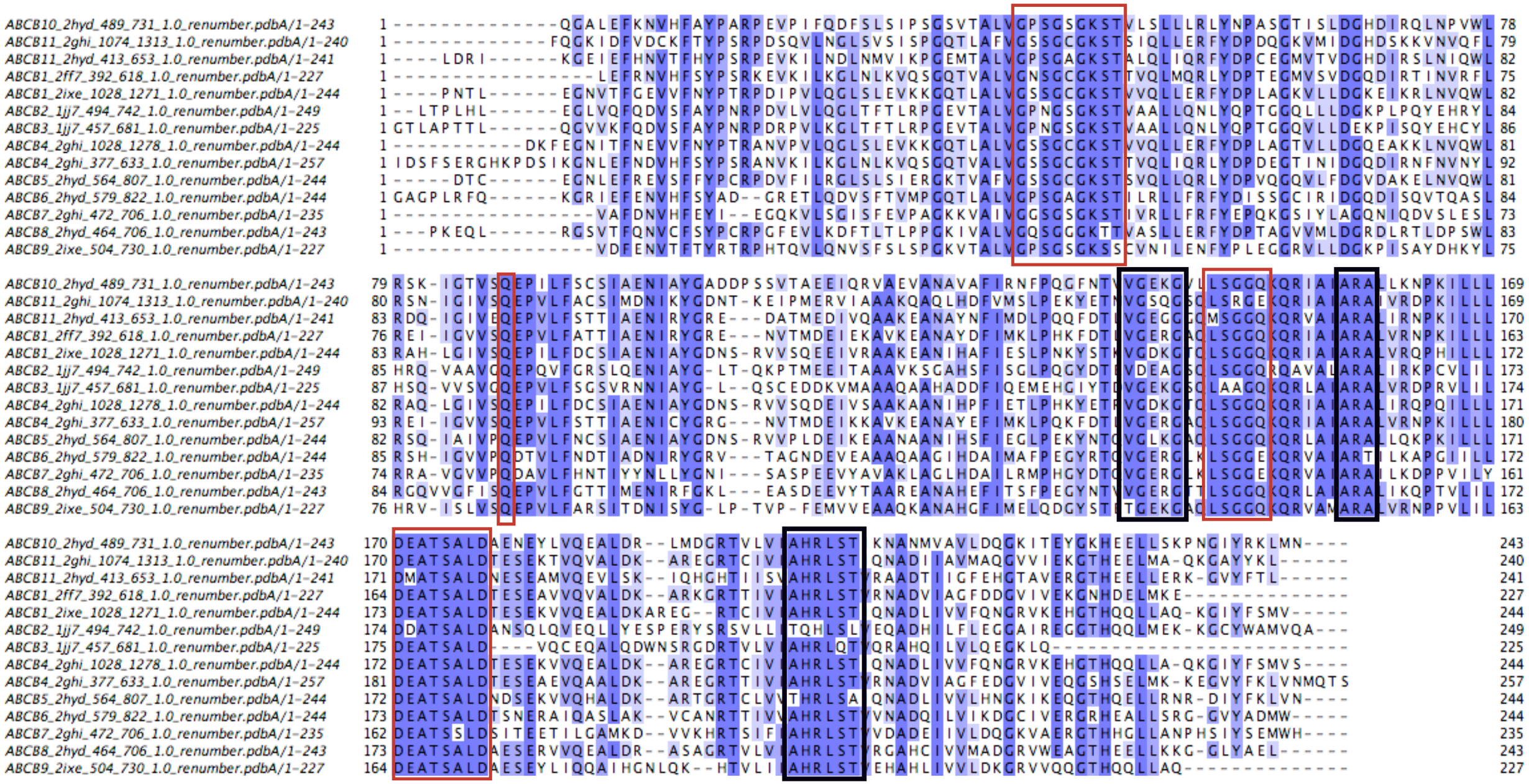


Do disease-associated mutations hint at common mechanisms?

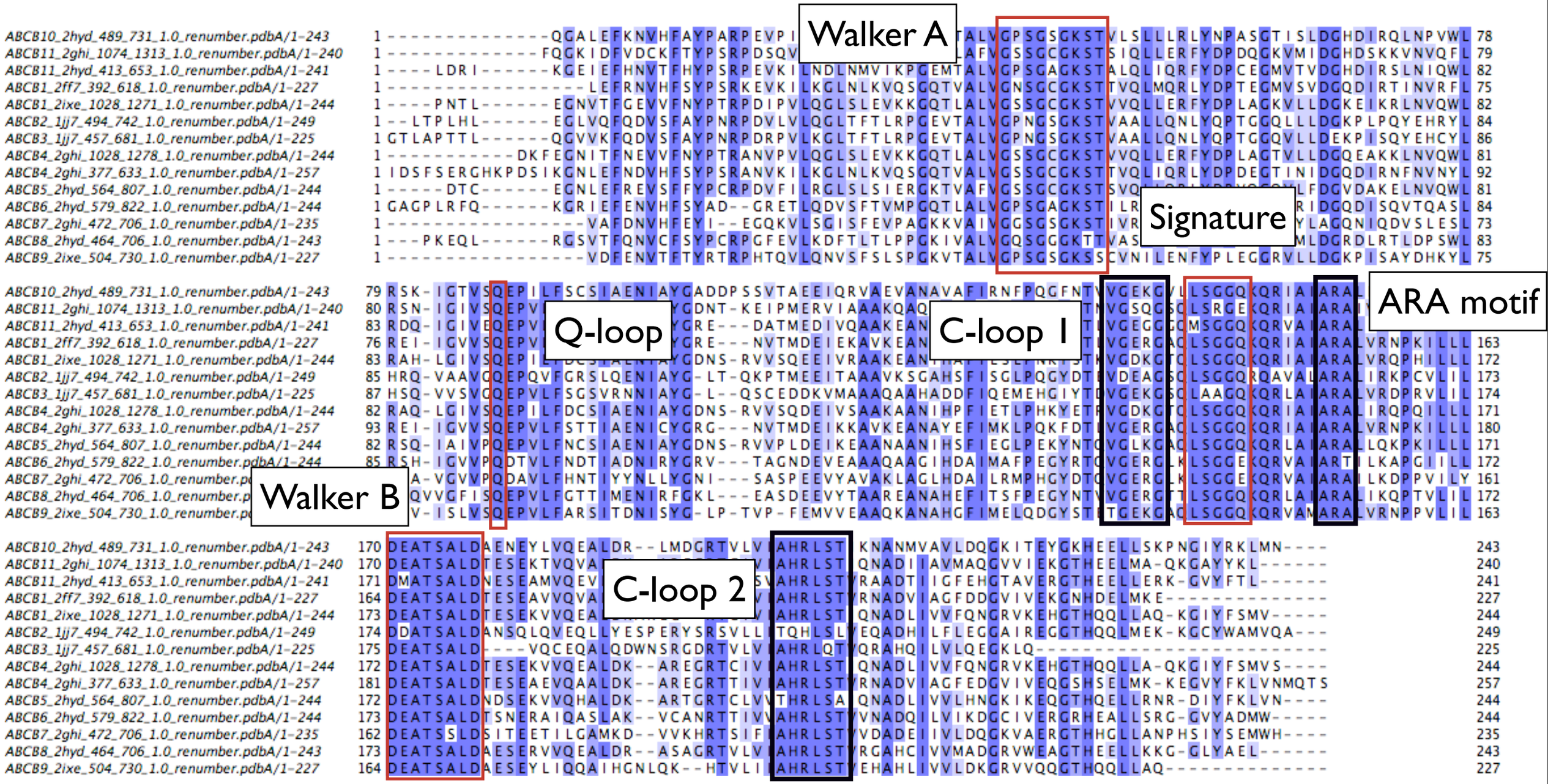


Are transporters across the superfamily similar enough that disease-associated mutations in one family are predictive of disease association in other members?

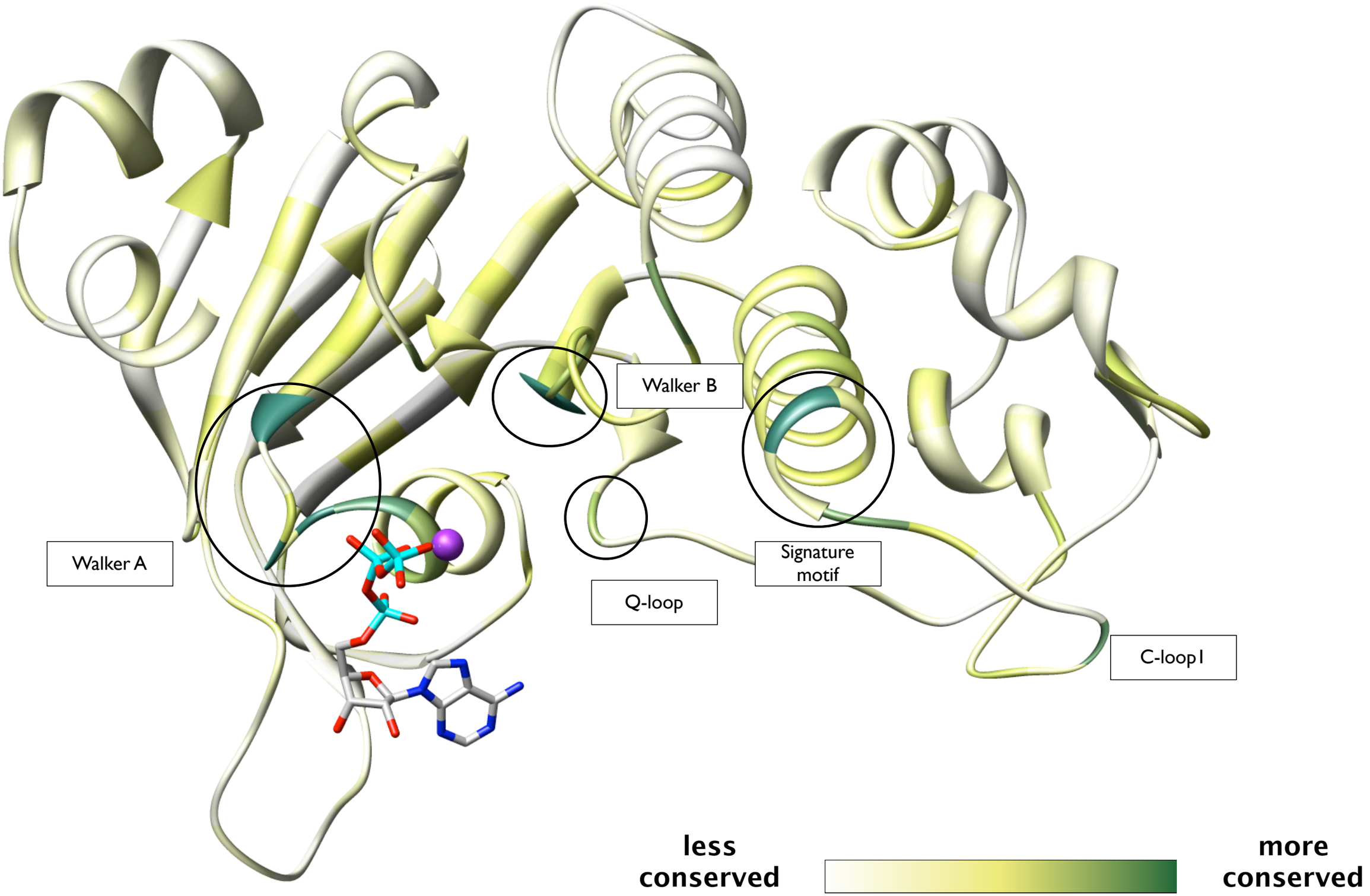
Common conservation patterns across all human NBDs suggest functional residues



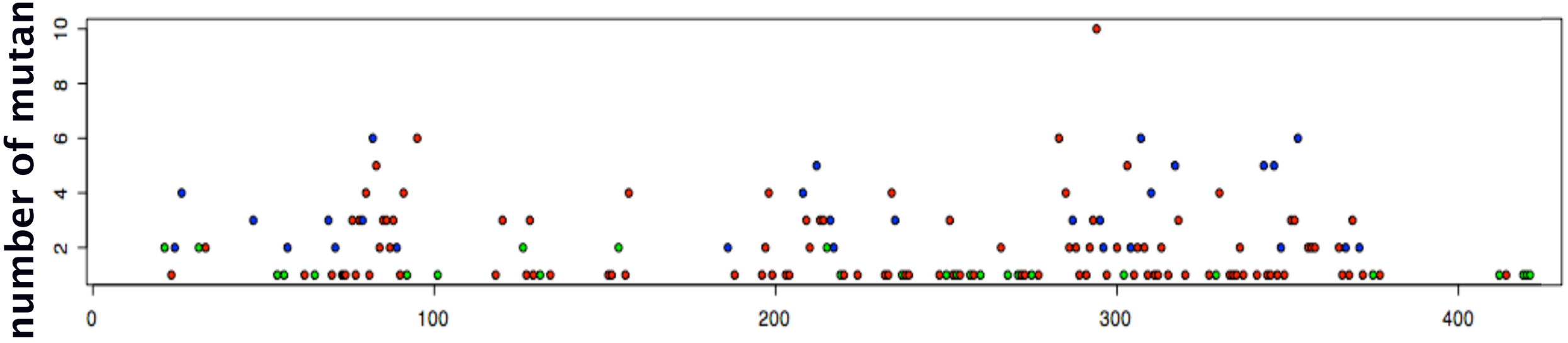
Common conservation patterns across all human NBDs suggest functional residues



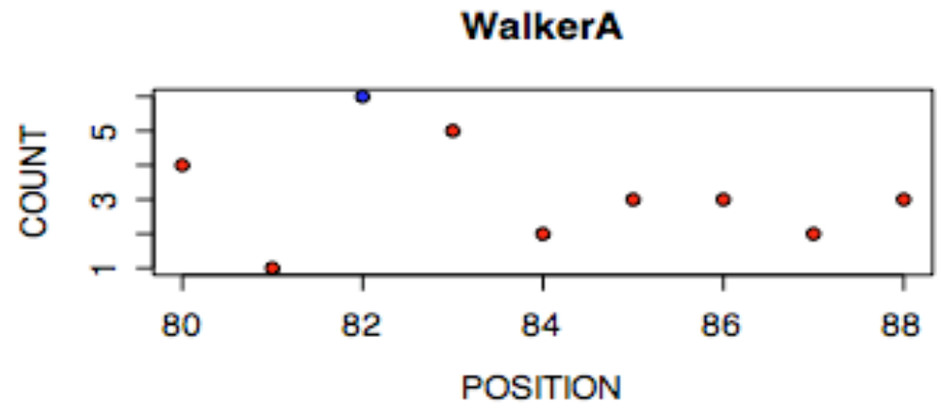
Common conservation patterns across all human NBDs suggest functional residues



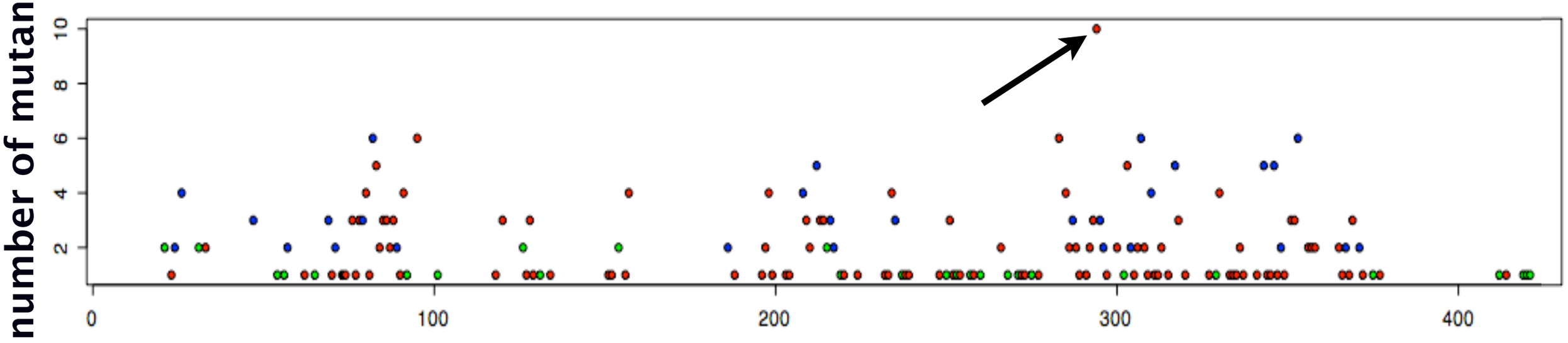
Clinical data provides evidence for a common mechanism



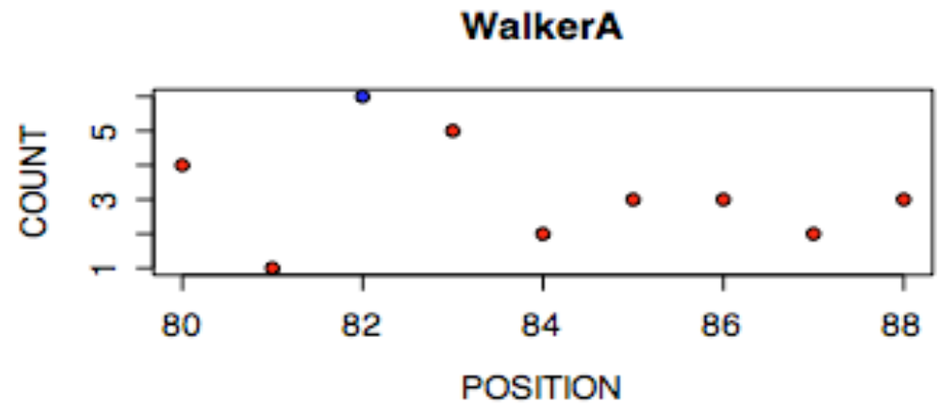
- Mutations mapped to a multiple structure alignment of all human NBD comparative models.
- Known motifs show disease mutations in multiple transporters.



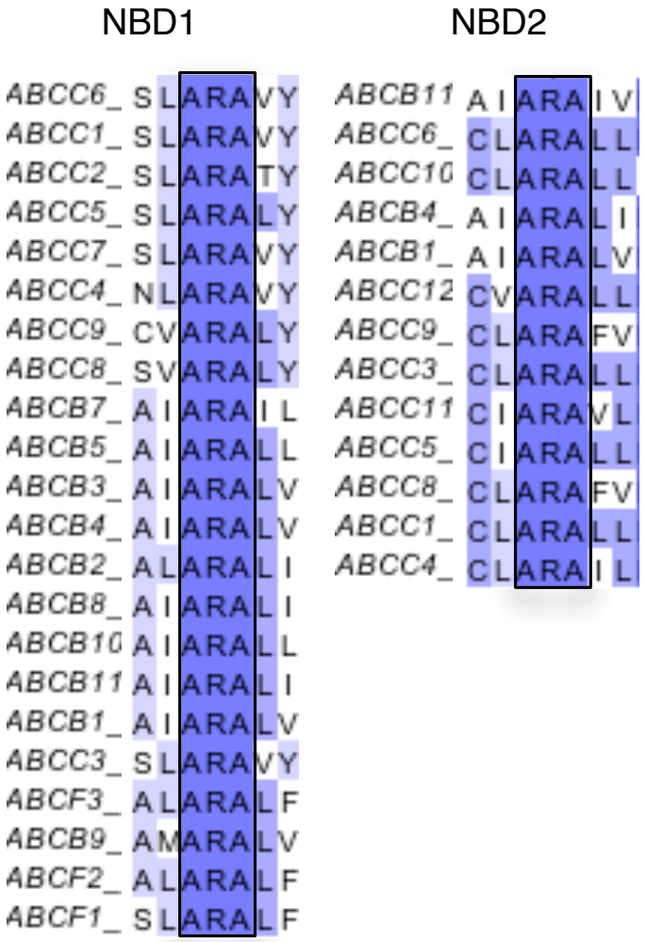
Clinical data provides evidence for common mechanisms

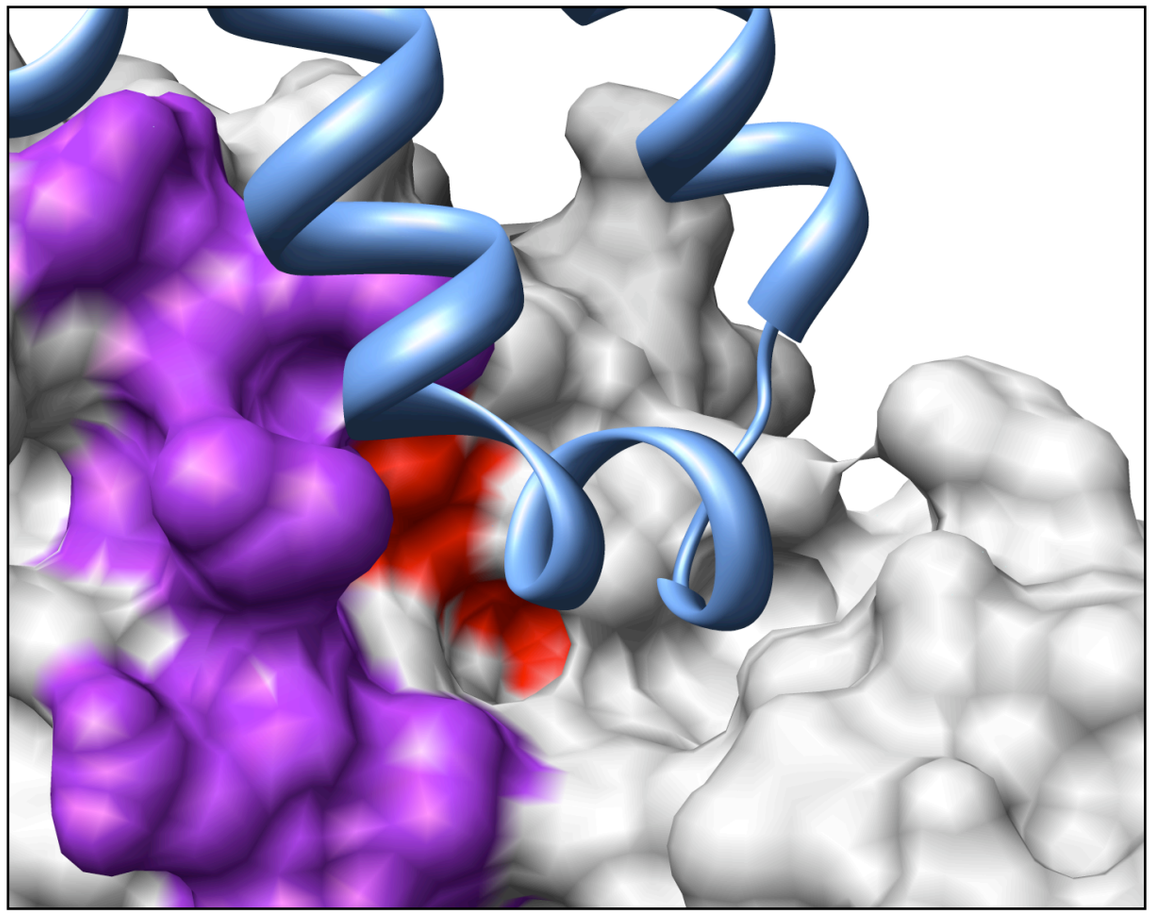


- Mutations mapped to a multiple structure alignment of all human NBD comparative models.
- Known motifs show disease mutations in multiple transporters.



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





- 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-synonomous 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**

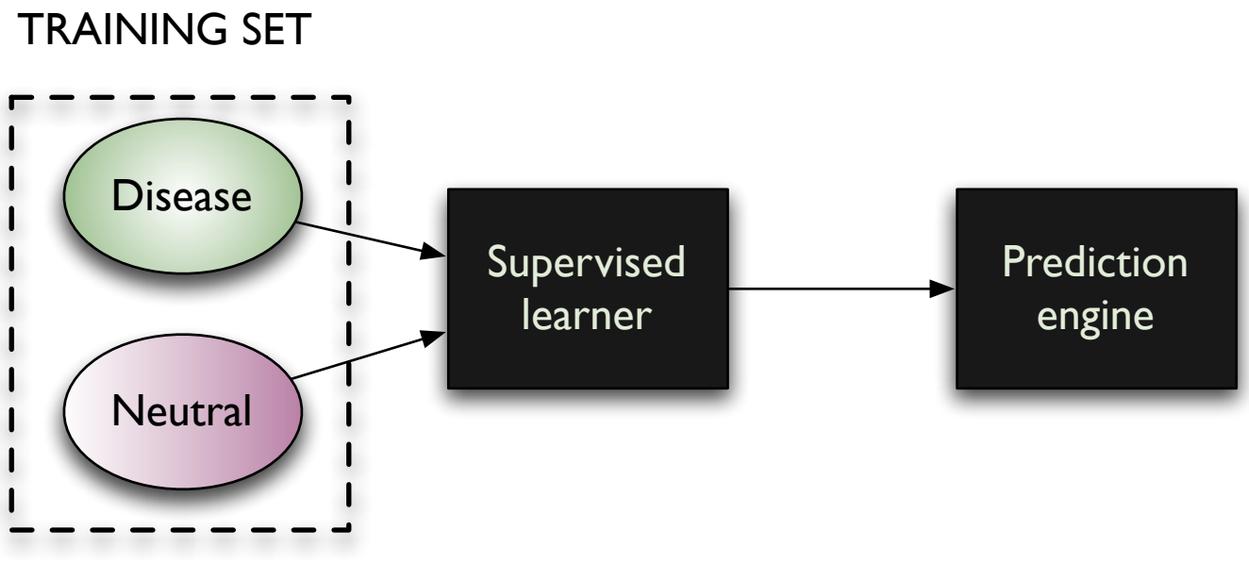
GENE	MUT	ASA-MUT	RSA-MUT	ASA-WT	RSA-WT	Delt-Res-Chrg	Delt-Res-Vol	Delt-Res-Pol	Pos-cons-score	Rel-Entropy	Granth. score	Buried-charge
CFTR	A1364V	6	0.04	6	0.06	0	-1.71	-1	-8.97	1.14	64	0
CFTR	A455E	0	0	0	0	1	-1.66	9.8	-8.27	0.66	107	1
CFTR	A559T	0	0	0	0	0	-0.92	0.4	-5.4	1.53	58	0

Using supervised learning to predict the effect of point mutants

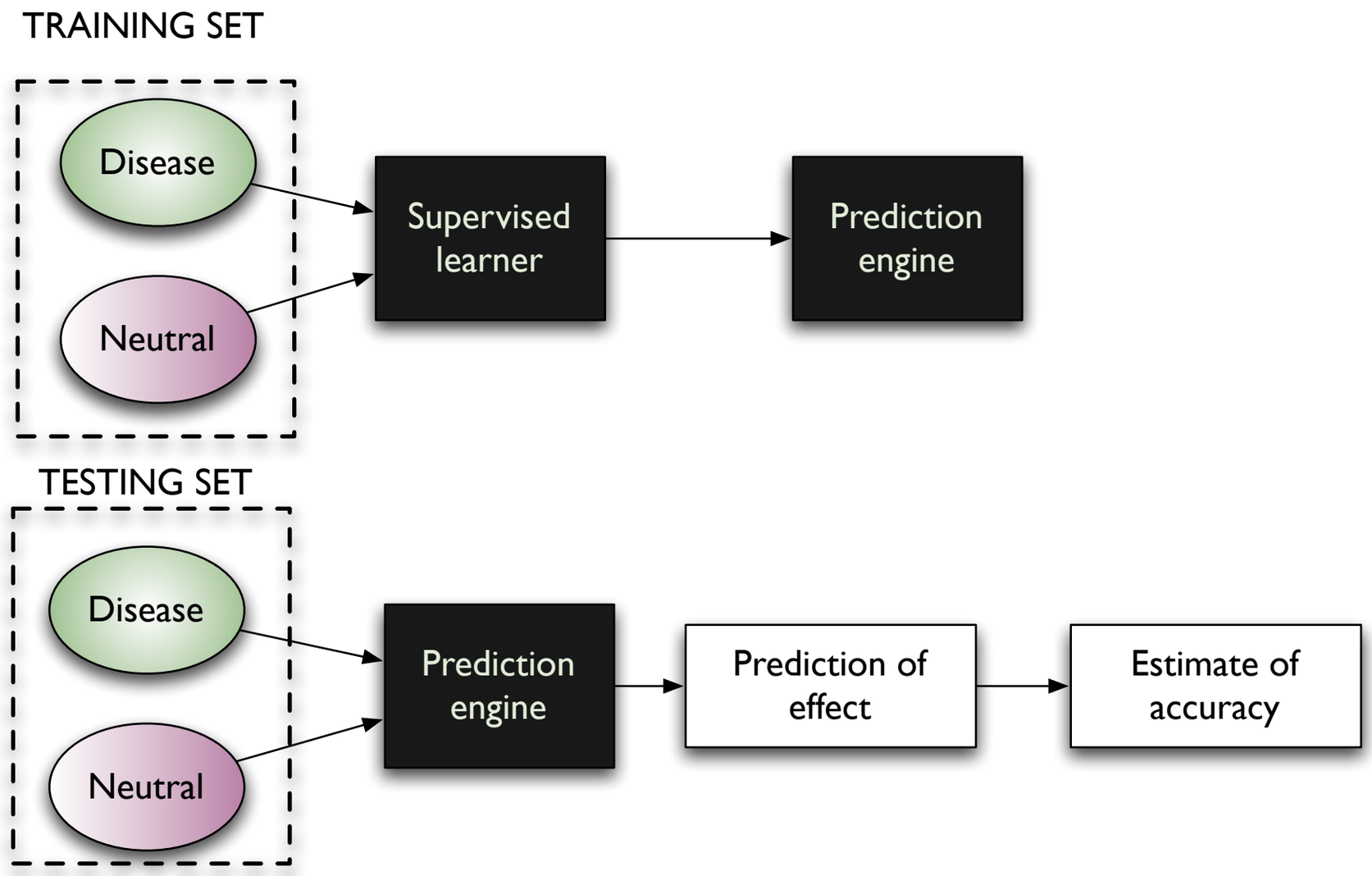
- Vector of sequence, structure, and evolutionary features representing mutated residues
- Binary prediction of effect

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

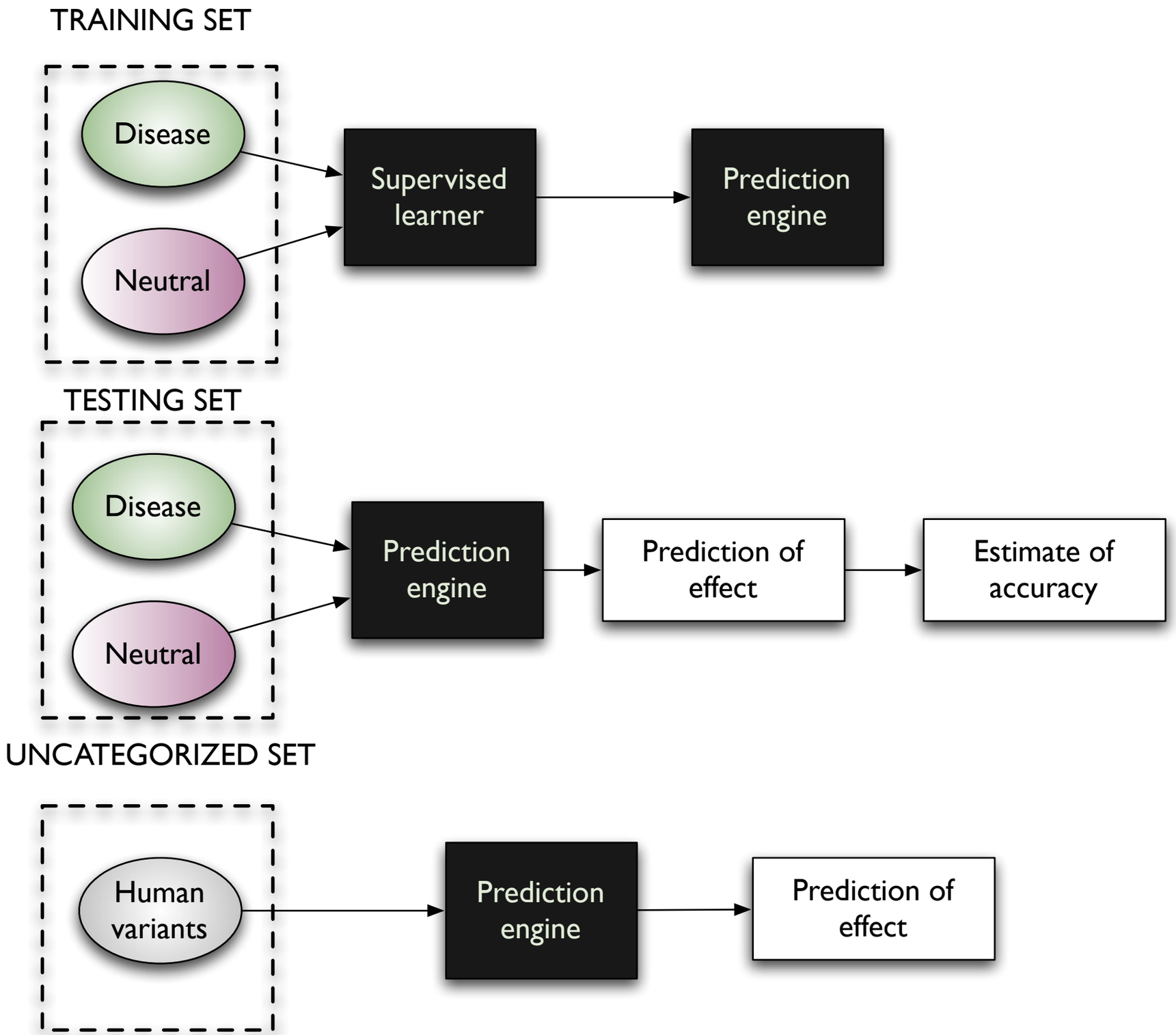
A supervised learner “learns” classes of data



A supervised learner “learns” classes of data

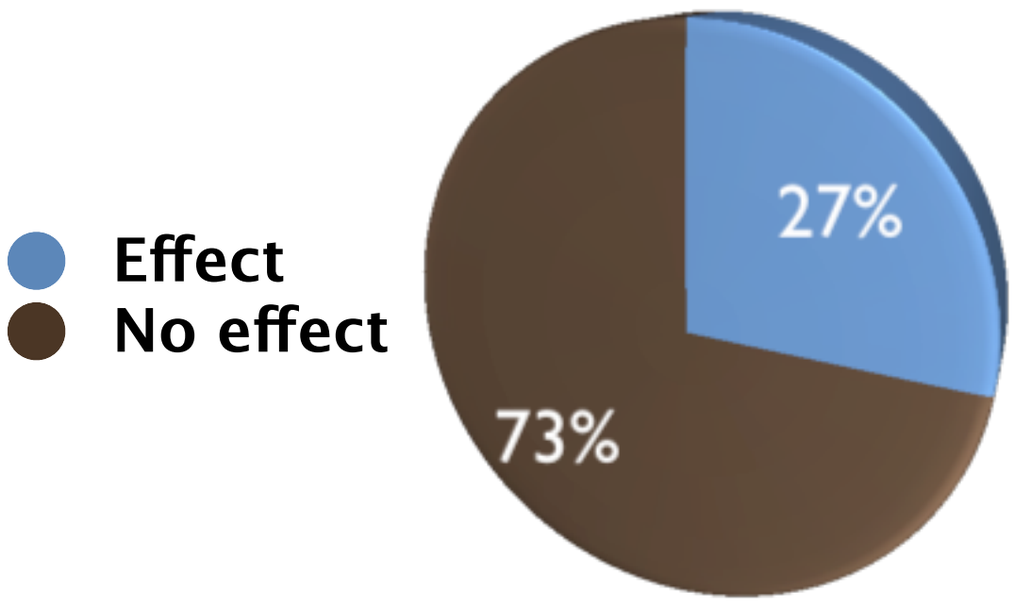


A supervised learner “learns” classes of data

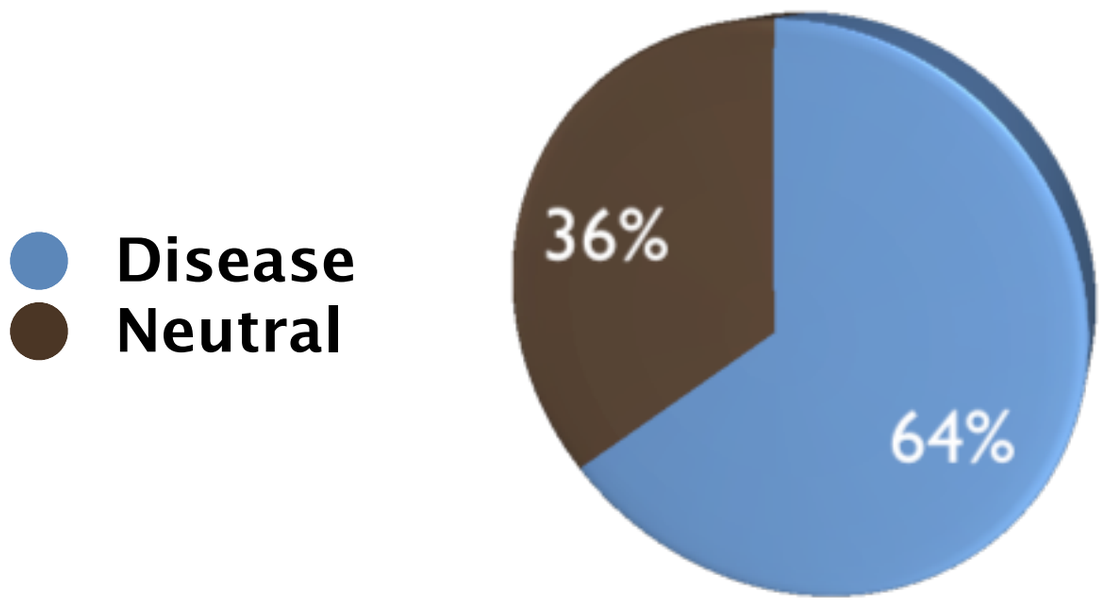


We use independent training and test sets to validate our predictions

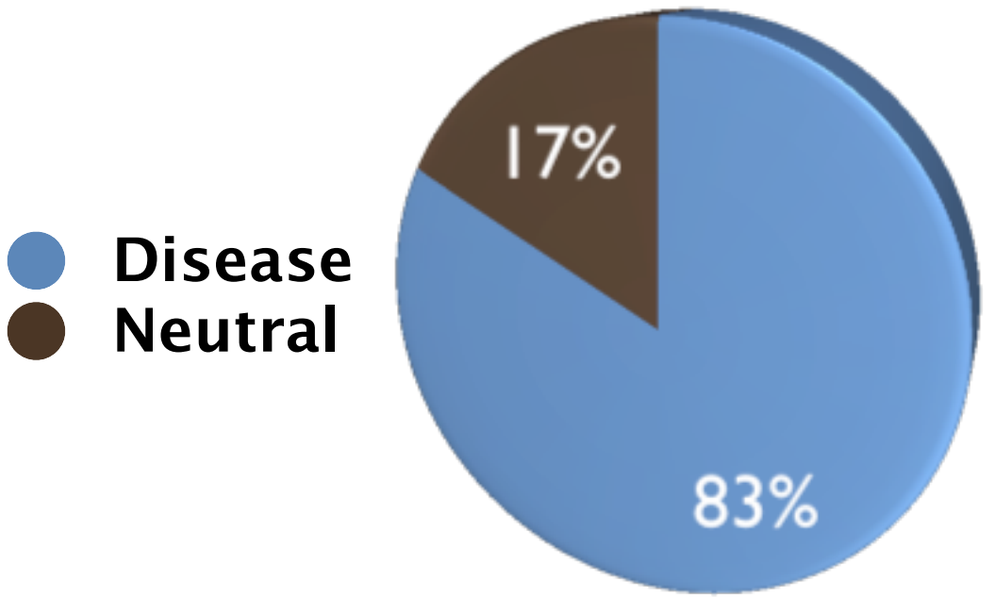
‘Experimental’ training set



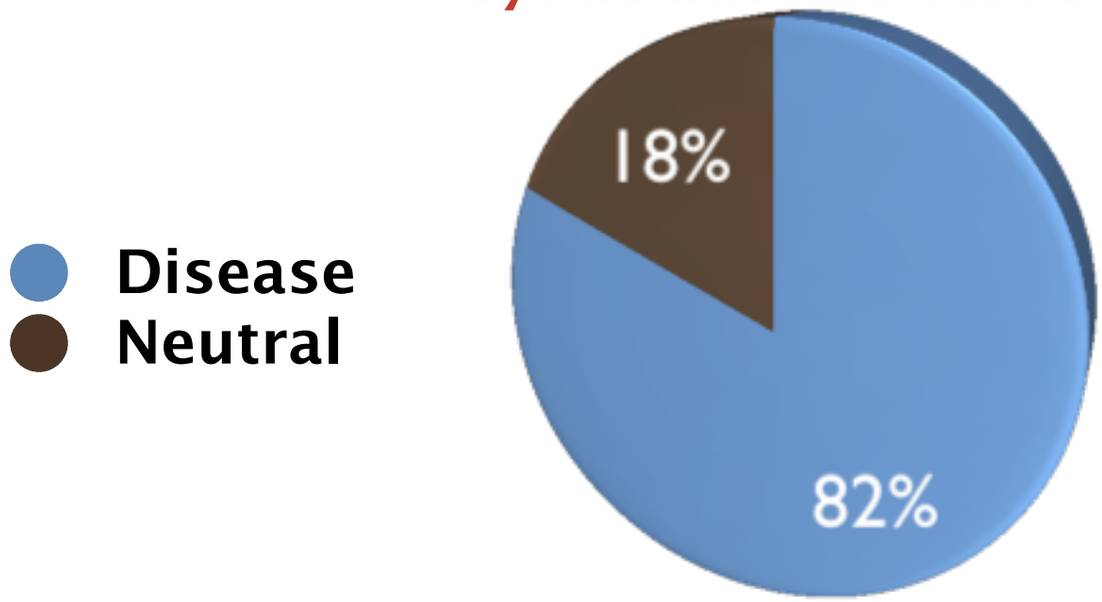
‘Clinical’ training set



ABC transporter test set

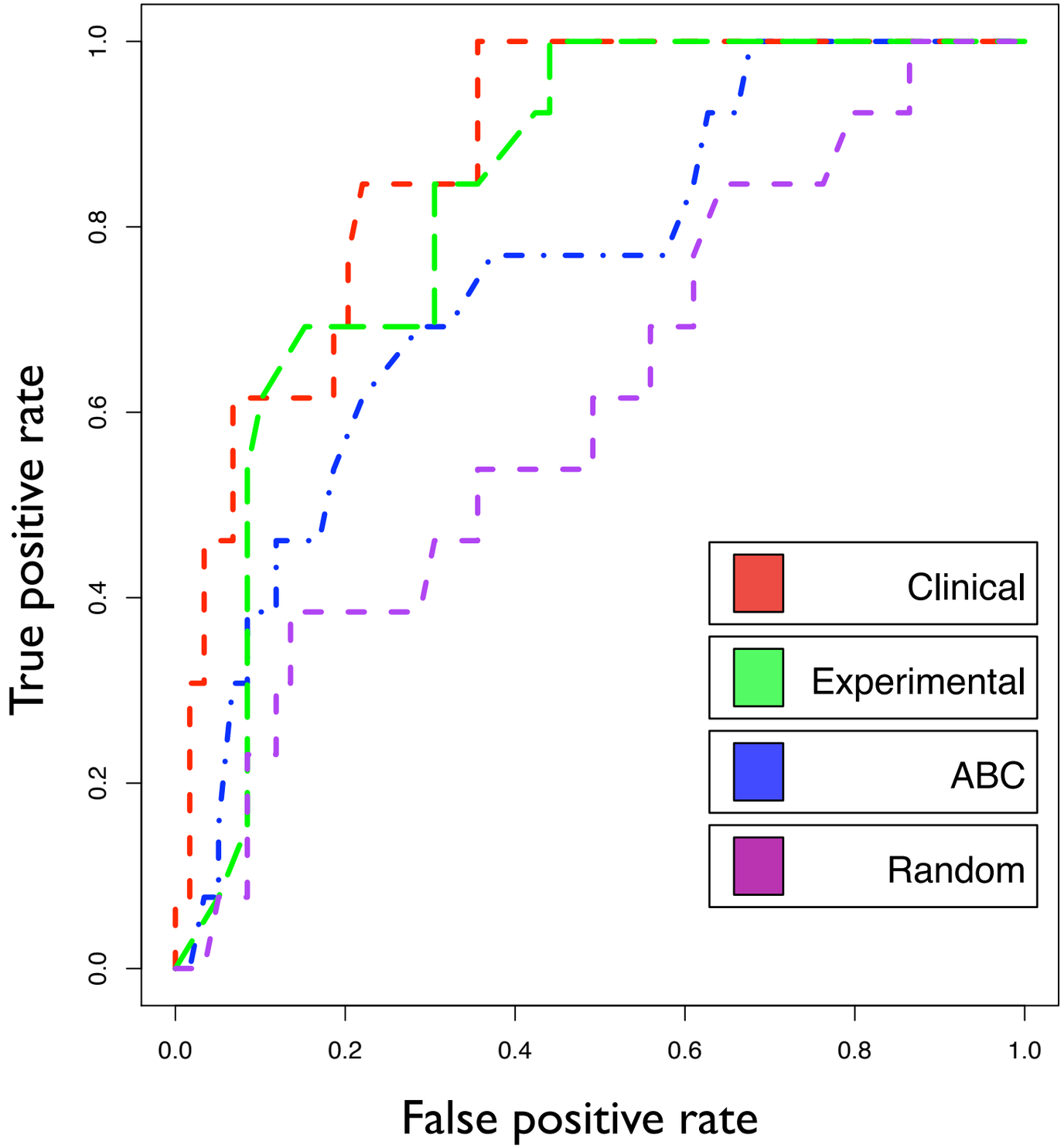


Cystic fibrosis test set



The clinical-trained algorithm is the best performing classifier

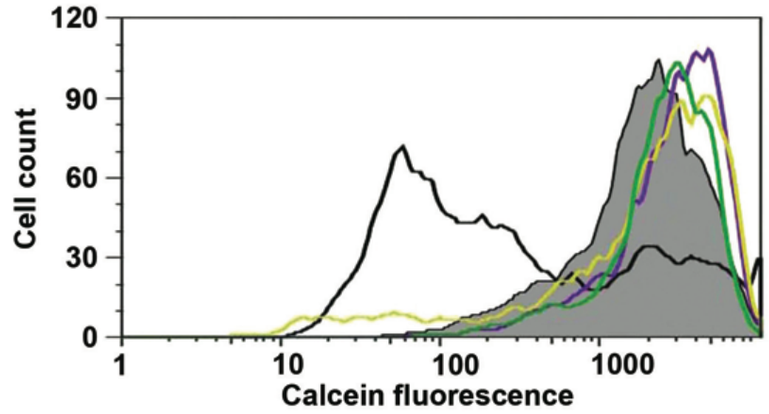
- **The clinical RF is 86% accurate on our cystic fibrosis test set.**
- **Now let's try it out on some unknown variants...**



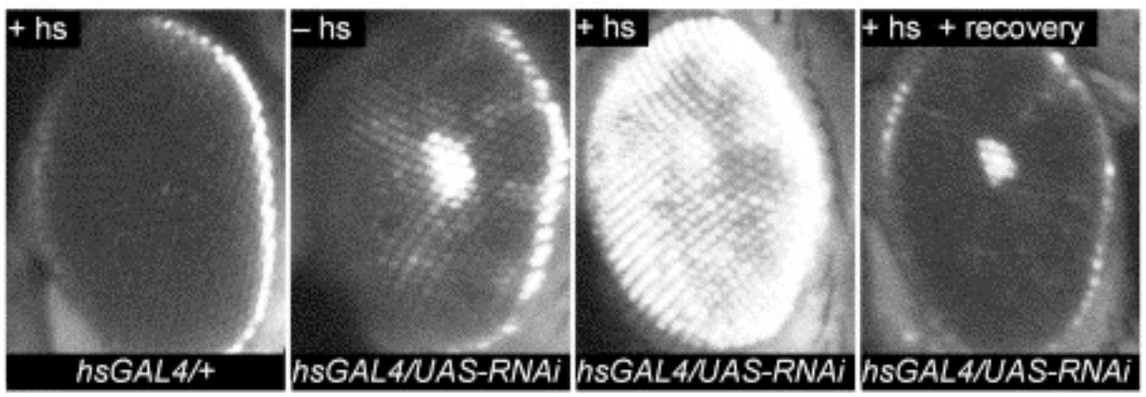
Experimental functional analysis of ABC transporters

- Yeast – **transport assays, cytotoxicity**

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



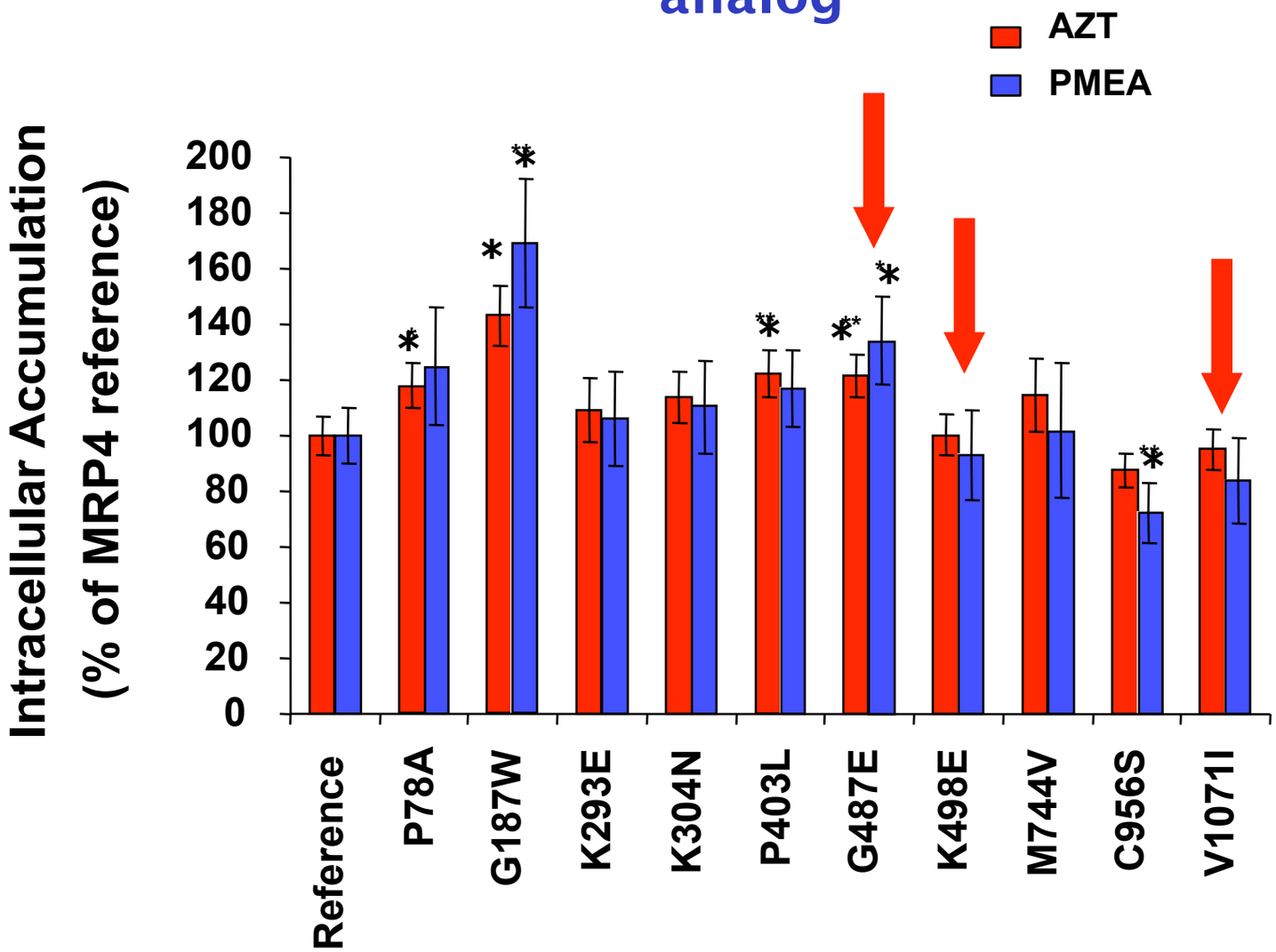
- Drosophila – **live visualization of transport across membranes**



Validation of MRP4 predictions

Variant	Prediction	
G487E	Disease	✔
K498E	Neutral	✔
V1071I	Neutral	✔

Transfected HEK cells, radiolabeled nucleoside/nucleotide analog



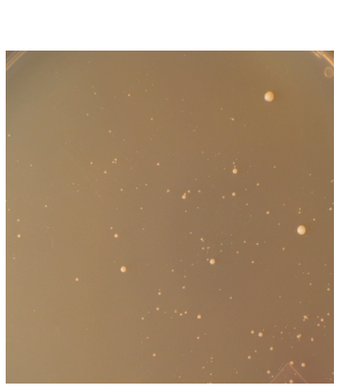
Nada Abla and Deanna Kroetz

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

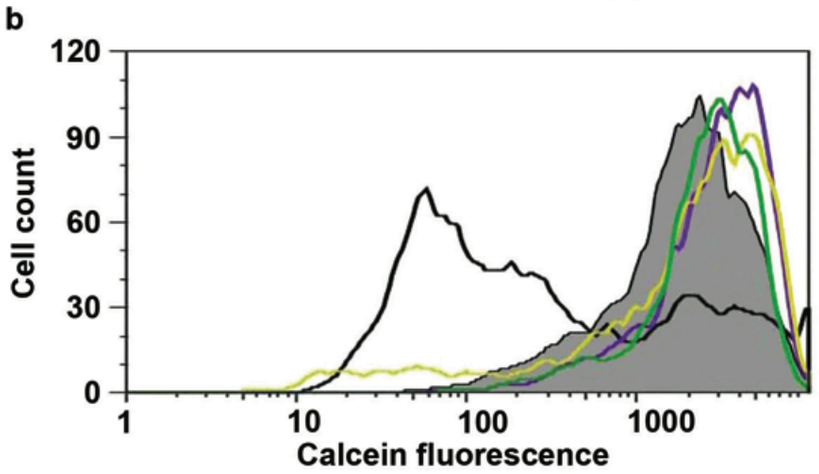
Variant	Prediction	
W1108R	Disease	✔
S1141T	Disease	✗
V1251I	Neutral	✔

Reference

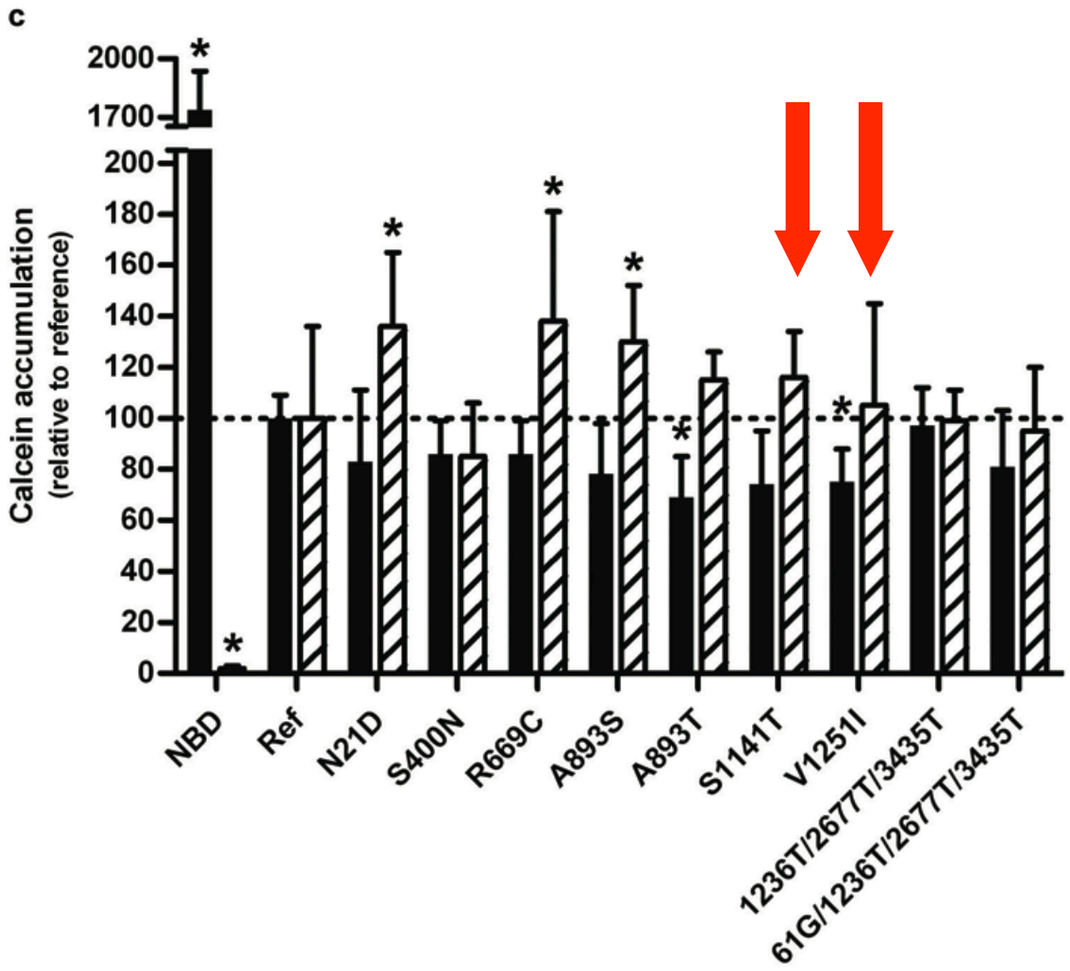
S1141T



W1108R

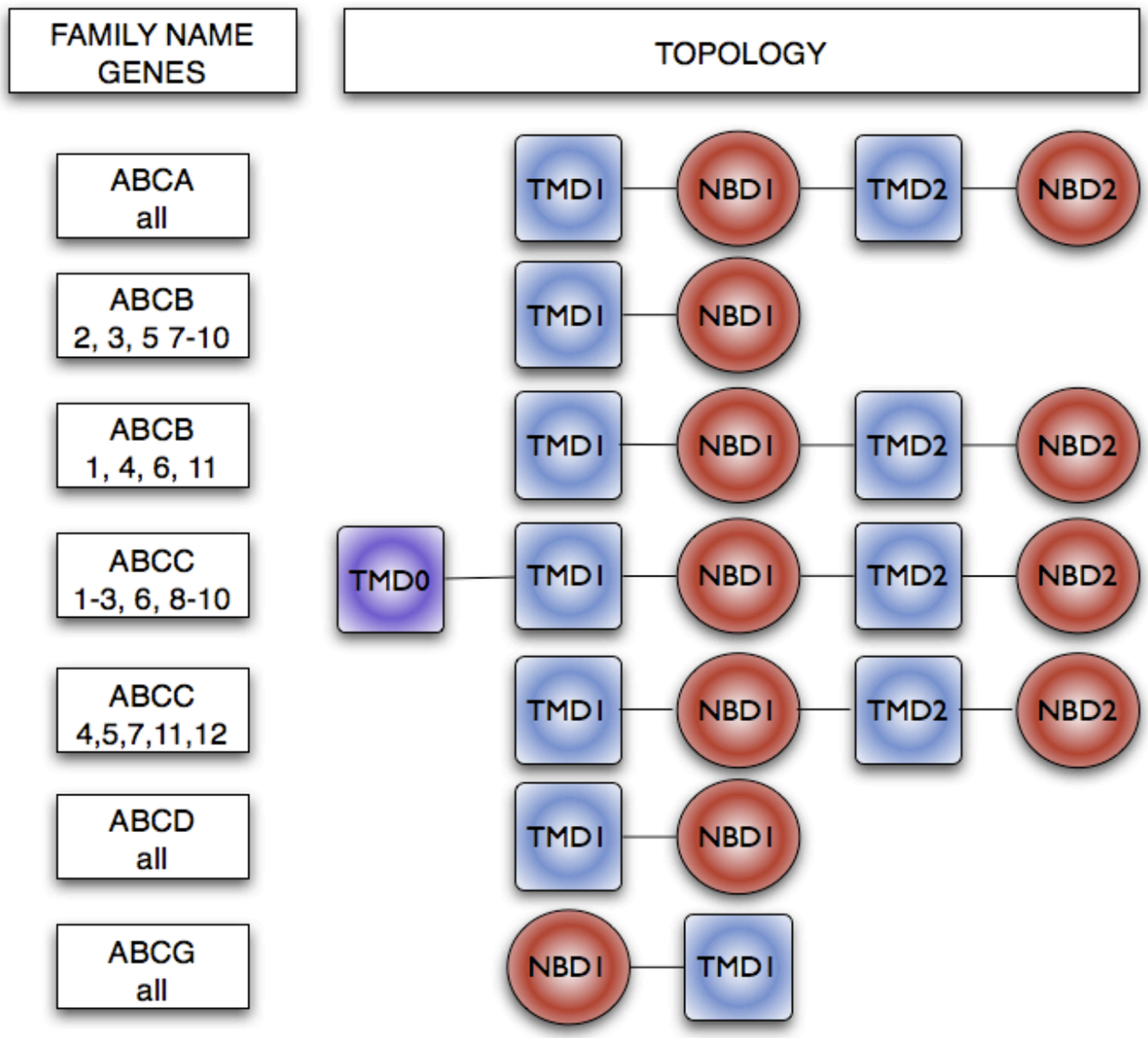


Bodipy-paclitaxel accumulation in HEK293T cells transiently transfected with P-gp reference and variants



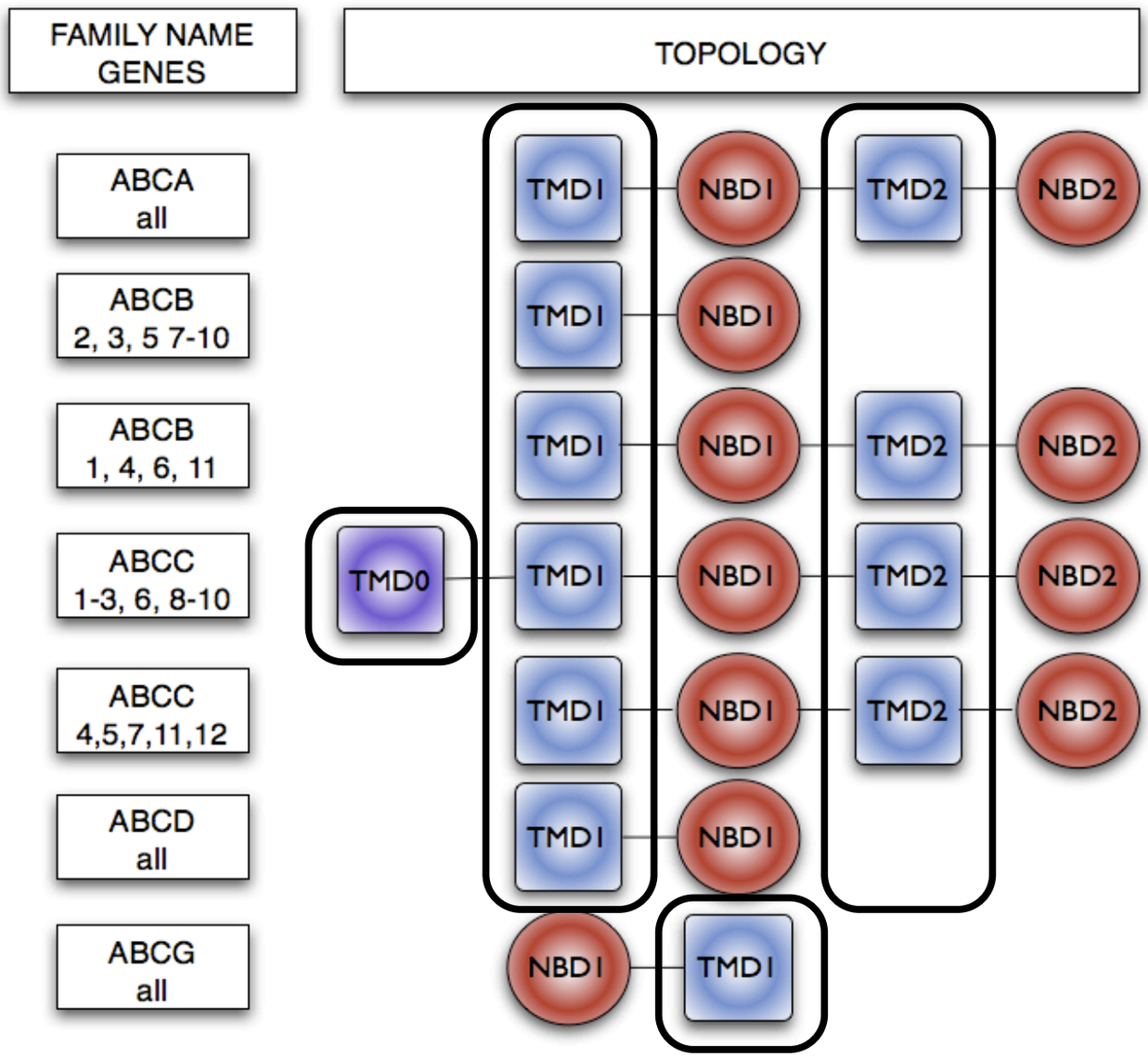
Jason Gow and Deanna Kroetz

Towards predicting substrate specificity for membrane transporters



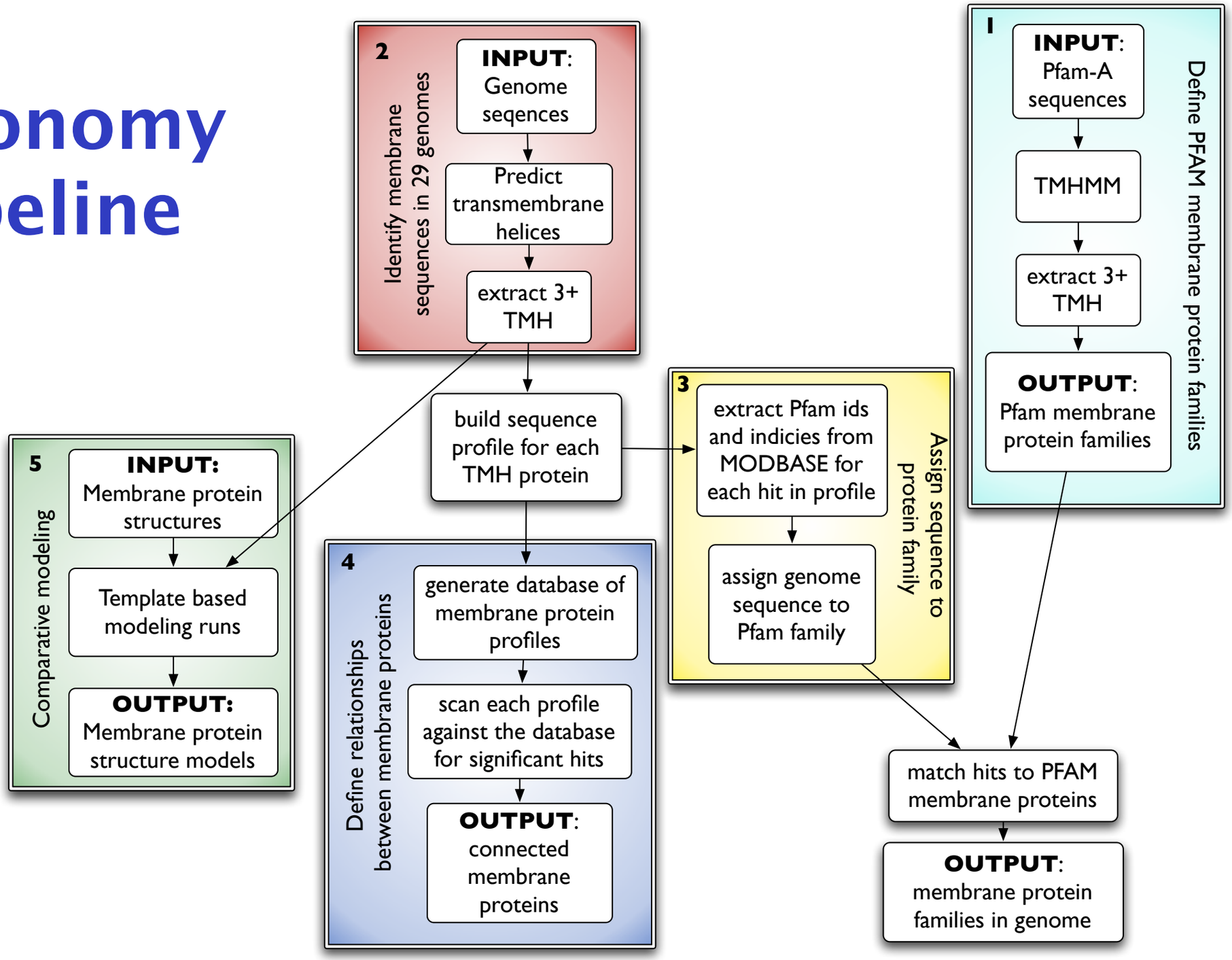
- Substrates bind in the transmembrane domains
- Overlapping substrate specificity
  - Multidrug resistance
- Not easily alignable
  - Extremely diverse in sequence

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

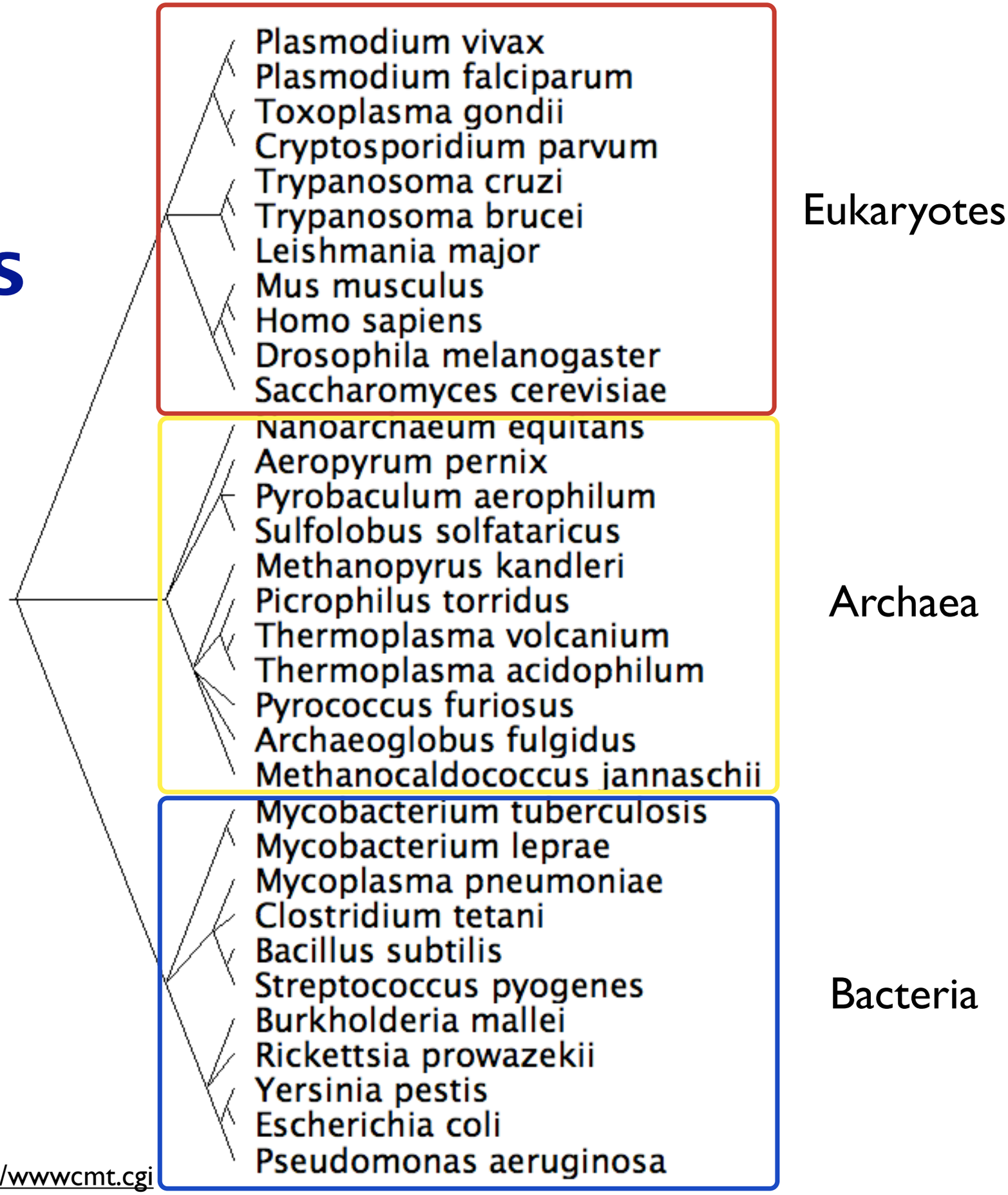


- **Create sequence profiles for all alpha-helical membrane proteins with three or more helices in 34 organisms**
- **Use the sequence profiles to identify Pfam families in each genome and to define new families**

Extending the analysis to whole genomes

We selected:

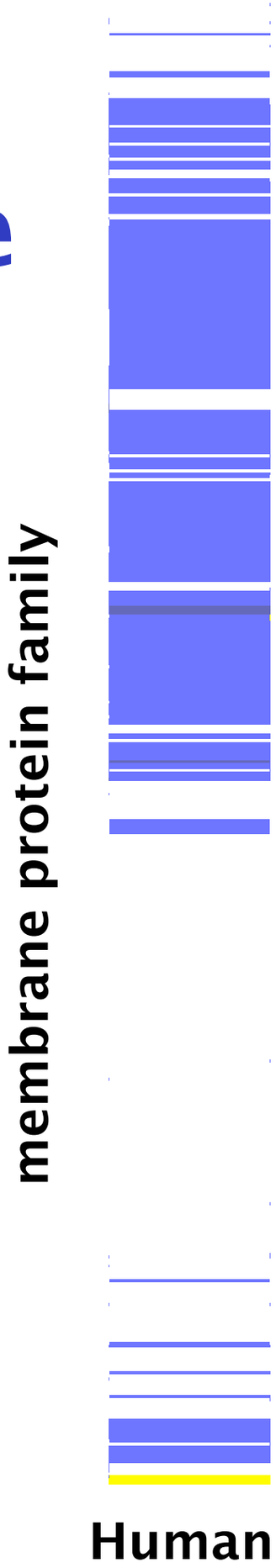
- **model species**
- **pathogenic species**
- **complete genomes**
- **genomic DNA available**



http://www.ncbi.nlm.nih.gov/Taxonomy/CommonTree/wwwcmt.cgi

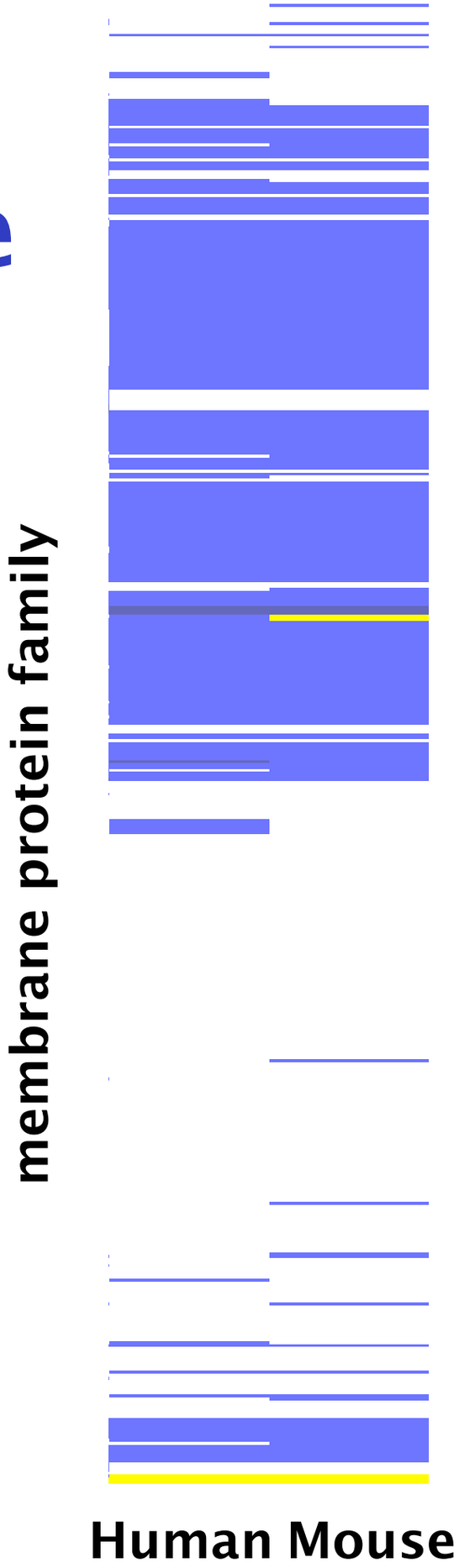
Identifying the membrane proteome of organisms

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



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?

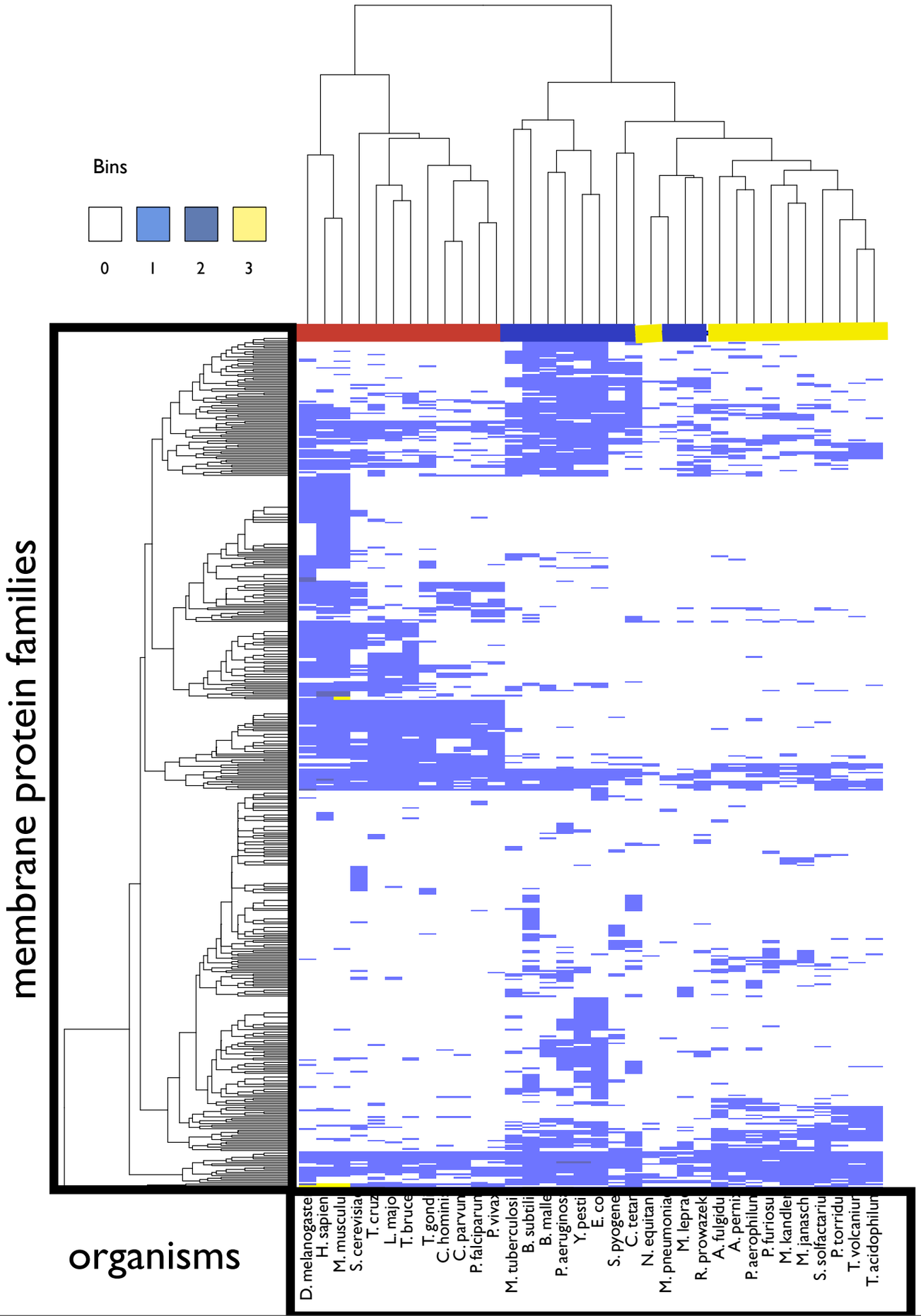


Identifying the membrane proteome of organisms

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

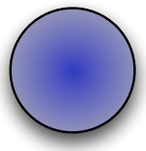
Family content reveals a clear split between prokaryotes and eukaryotes

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

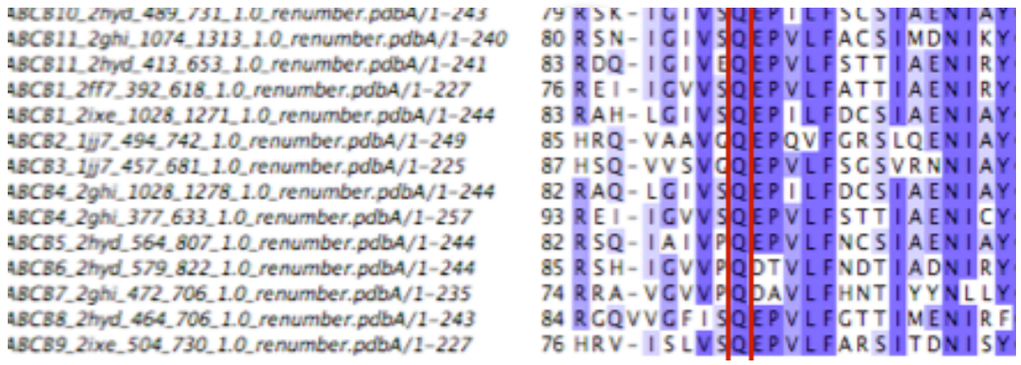


We create a large database of membrane protein profiles and use it to scan for similarity

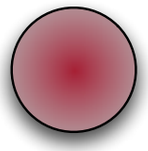
protein 1



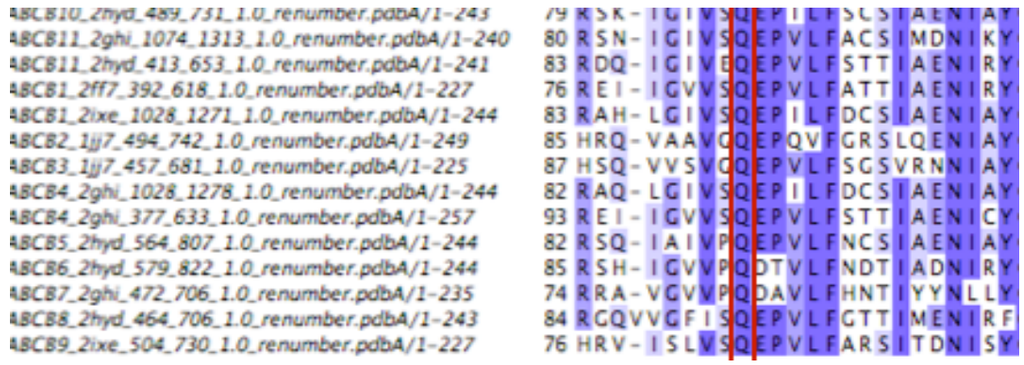
Multiple sequence alignment of protein 1



protein 2

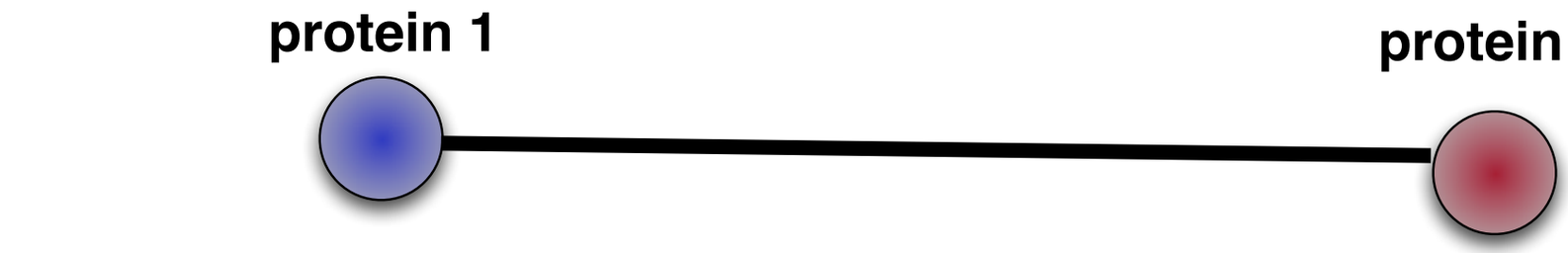


Multiple sequence alignment of protein 2

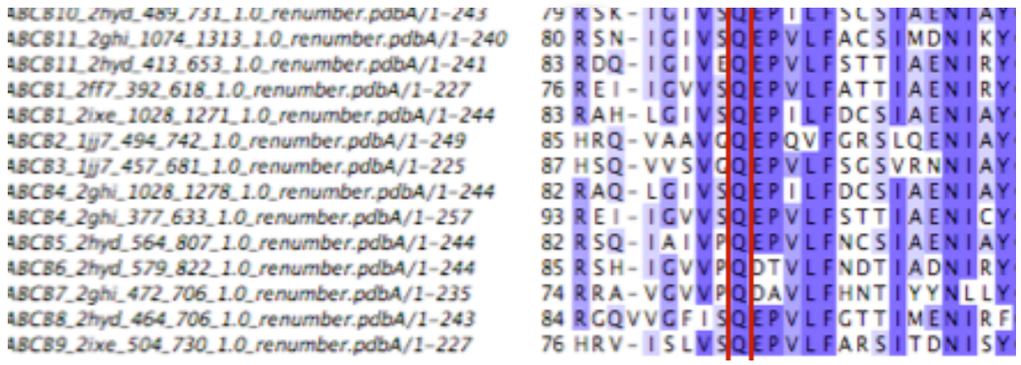


- **Each membrane protein profile can be represented as a node**

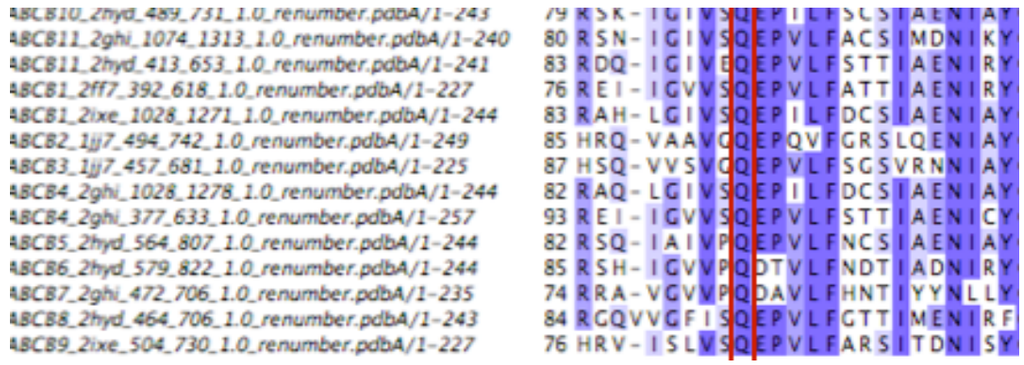
We create a large database of membrane protein profiles and use it to scan for similarity



Multiple sequence alignment of protein 1



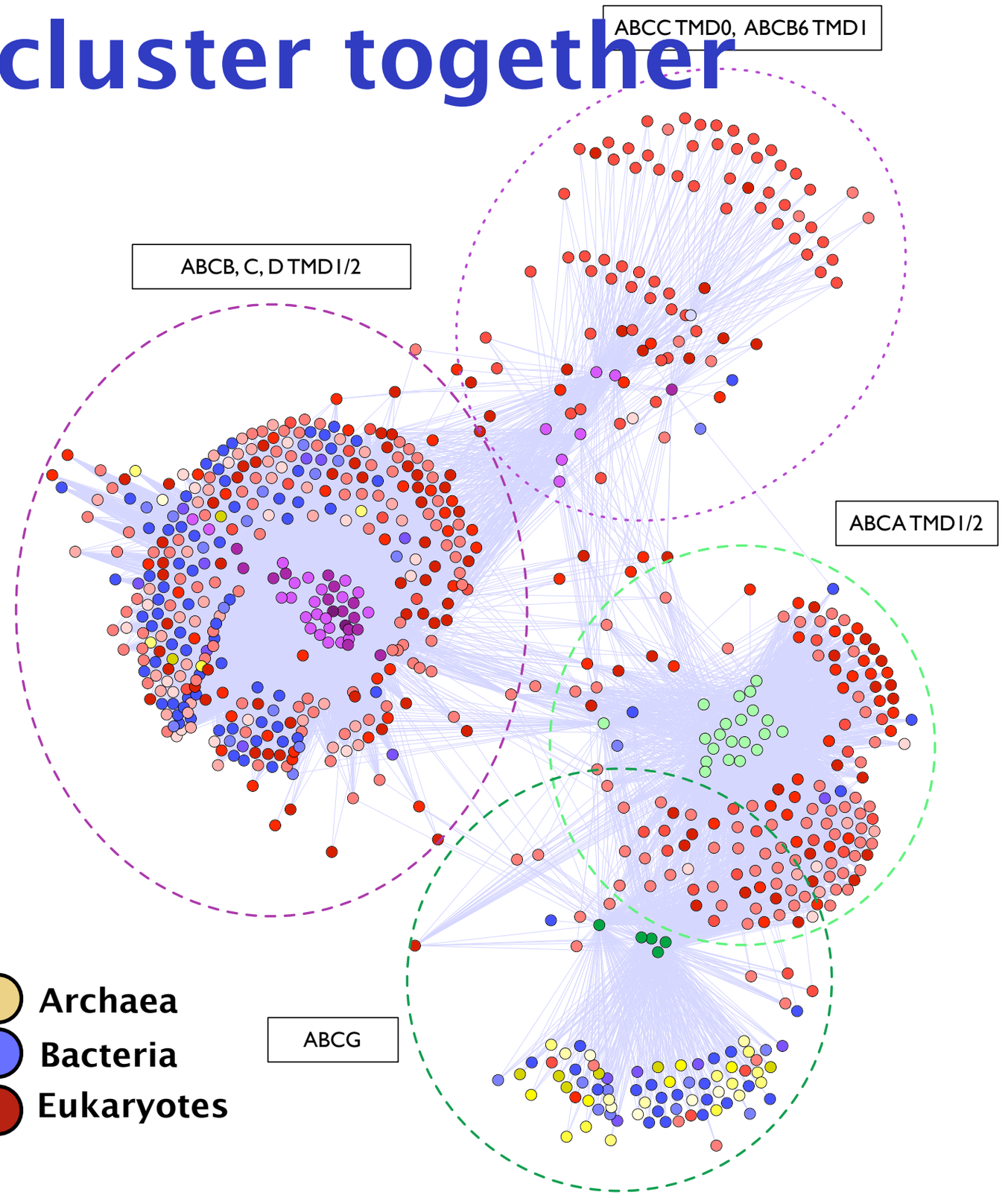
Multiple sequence alignment of protein 2



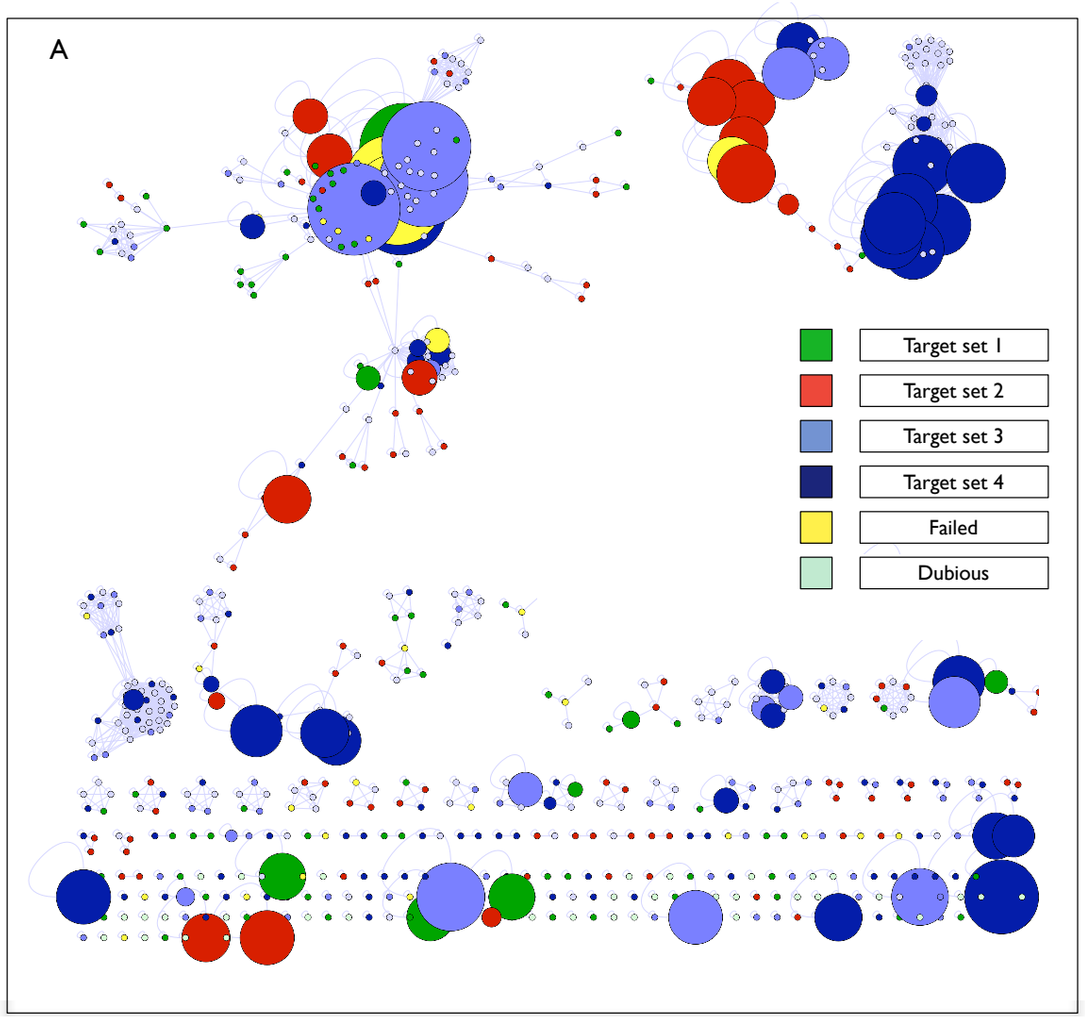
- **Each membrane protein profile can be represented as a node**
- **If two profiles are significantly similar they are “connected” with a line**

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 α-helical membrane proteins in the yeast genome
  - represent all families and clusters
  - two members per family

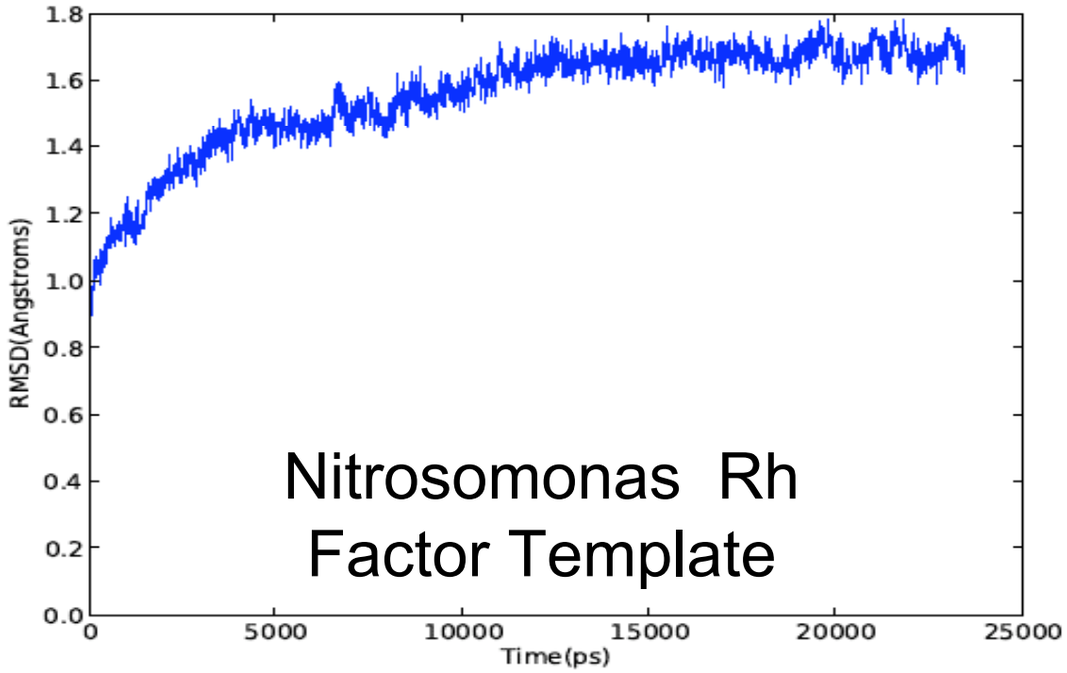
Comparative modeling of a human Rh factor protein based on a bacterial template

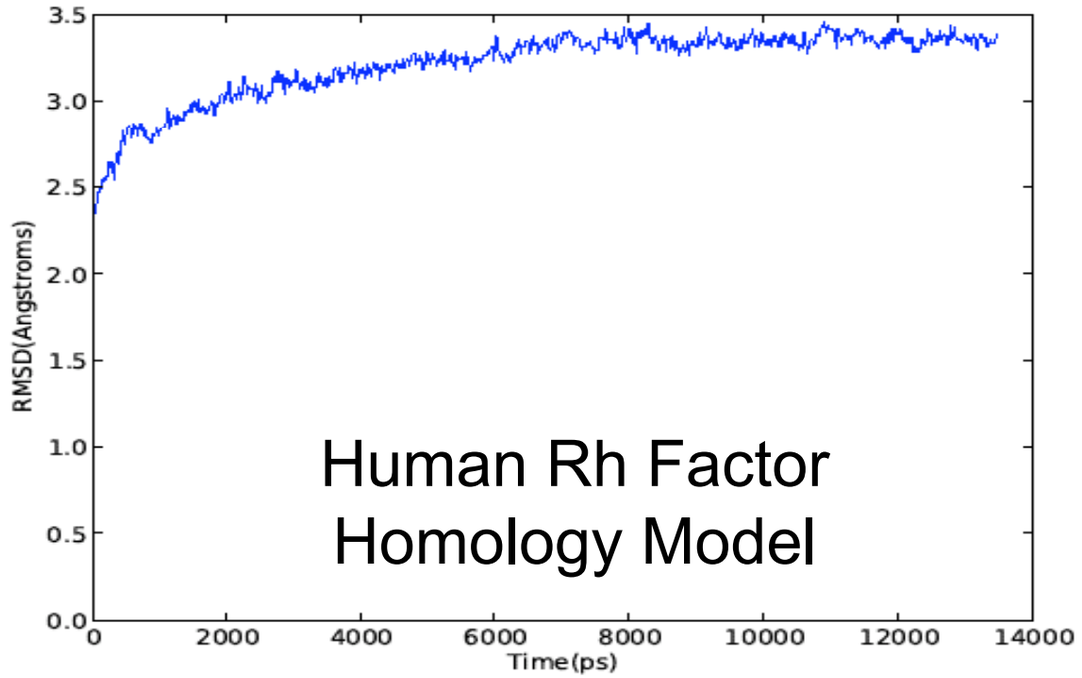
- NeRH is an ammonia transporter from the bacterium *Nitrosomonas Europaea*. The structure was recently solved by Franz Gruswitz in Robert Stroud's lab.
- 40% sequence identical to the human Rh gene hRhBG, alignment is good outside of one long loop in the human protein (~25 residues).

```
hRhBG/1-352     1 -ADNEFYFRYPSFQDVHAMVFVGFGFLMV-FLQ-RYGFSSVGFTFLLAAFALQWSTLVQGFLH-/-------------- 61
NeRH/1-351      1 INEARLVAQYNYSINILAMLLVGFGFLMV-FVR-RYGFSATTGTYLVVATGLPLYILLR-ANGI/-------------- 62

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

hRhBG/1-352   123 DAGGSMTIHTFGAYFGLVLSRVLYRPQLEKSKHRQGSVYHSDLFAMIGTIFLWIFWPSFNAALTA-LGAGQHRTALNT 199
NeRH/1-351    132 DSAGSIAIHAFGAYFGLGVSIALTTAAQRAQP--IESDATSDRFSMLGSMVLWLFWPSFATAIVP--FEQMPQTIVNT 205

hRhBG/1-352   200 YYSLAASTLGTFALSALVGEDGRLDMVHIQNAALAGGVVVGTSSEMMLTPFGALAAGFLAGTVSTLGYKFFTPILESK 277
NeRH/1-351    206 LLALCGATLATYFLSALFH-KGKASIVDMANAALAGGVAIGSVCN-IVGPVGAFVIGLLGGAISVVGFVFIQPMLESK 281

hRhBG/1-352   278 FKVQDTCGVHNLHGMPGVLGALLGVLVAGLAQAMHQLFGLFVTLMFASVGGGLGGLLLKLP--FLDSPPDSQHYEDQ 352
NeRH/1-351    282 AKTIDTCGVHNLHGLPGLLGGFSAILIVPGIA-VAQLTGIGITLALALIGGVIAGALIKLT------GTTKQAYEDS 351
```

With Franz Gruswitz and Ilya Chorny, Robert Stroud lab

The homology model equilibrates!





Equilibration is monitored by the protein RMSD(t)

Figures from Ilya Chorny

In a molecular dyanmics simulation, the model is stable and the pore is recruiting NH4 ions

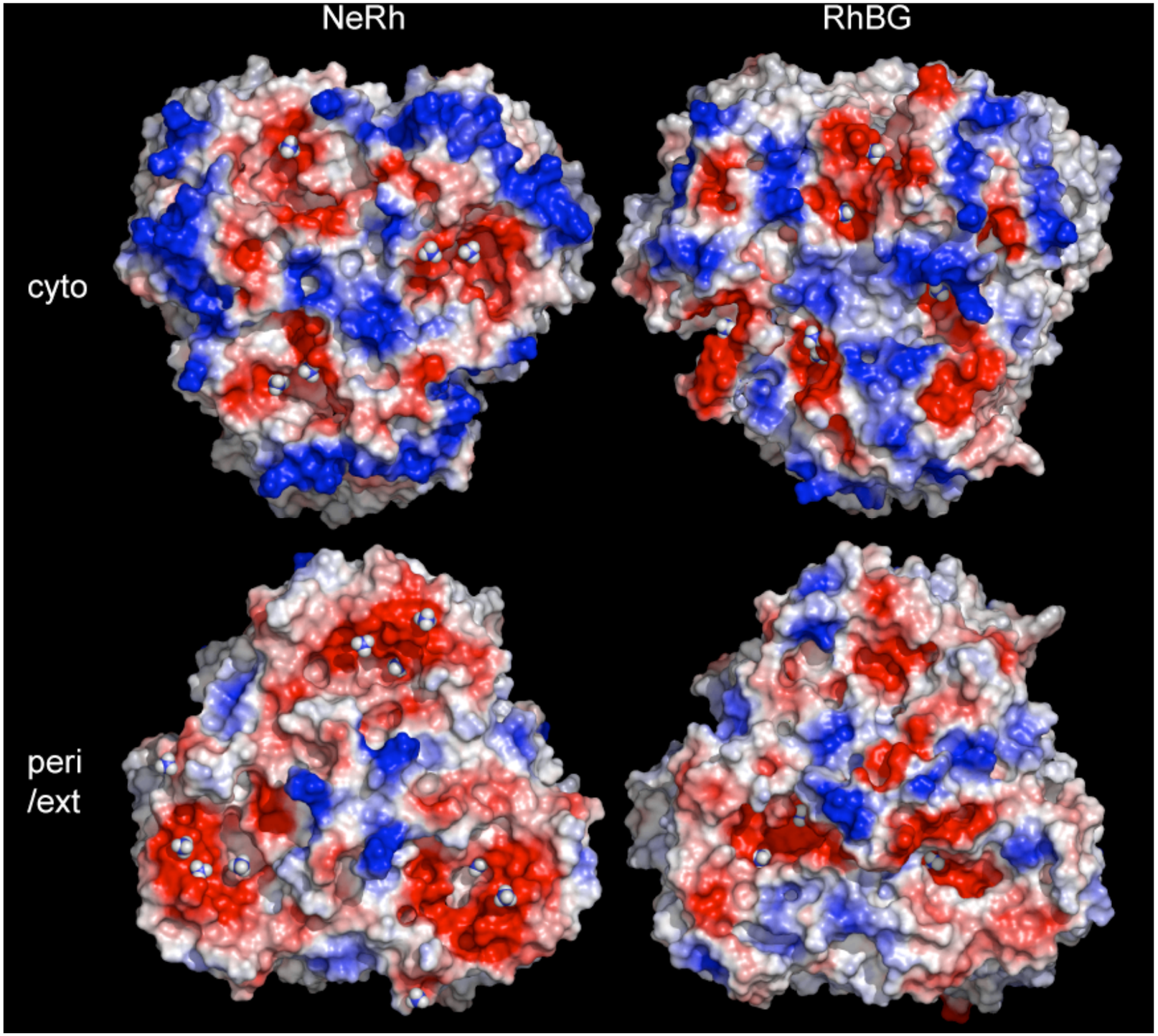


Figure from Ilya Chorny

Characterizing genetic variation in human transporters

- Comparative modeling of all human ABC tranpsorter 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