# An evolutionary approach to predicting variability in human drug response

**Libusha Kelly**
**Andrej Sali lab**

*In honor of Katie Gettman*

# Reflections on a stupid exit talk title

# Reflections on a stupid exit talk title

- **Mutational analysis of human ABC transporter nucleotide binding domains reveals functional hotspots**

# Reflections on a stupid exit talk title

- **Mutational analysis of human ABC transporter nucleotide binding domains reveals functional hotspots**
  - **“Although I don't understand a word of your project title, I'm sure it makes sense to you.”**

# Reflections on a stupid exit talk title

- **Mutational analysis of human ABC transporter nucleotide binding domains reveals functional hotspots**
  - **“Although I don't understand a word of your project title, I'm sure it makes sense to you.”**
  - **“I have no idea what your talk is about... it sounds very sci-fi to me.”**

# Reflections on a stupid exit talk title

- **Mutational analysis of human ABC transporter nucleotide binding domains reveals functional hotspots**
  - **“Although I don't understand a word of your project title, I'm sure it makes sense to you.”**
  - **“I have no idea what your talk is about... it sounds very sci-fi to me.”**
  - **“Can’t wait for the movie version!”**

# Reflections on a stupid exit talk title

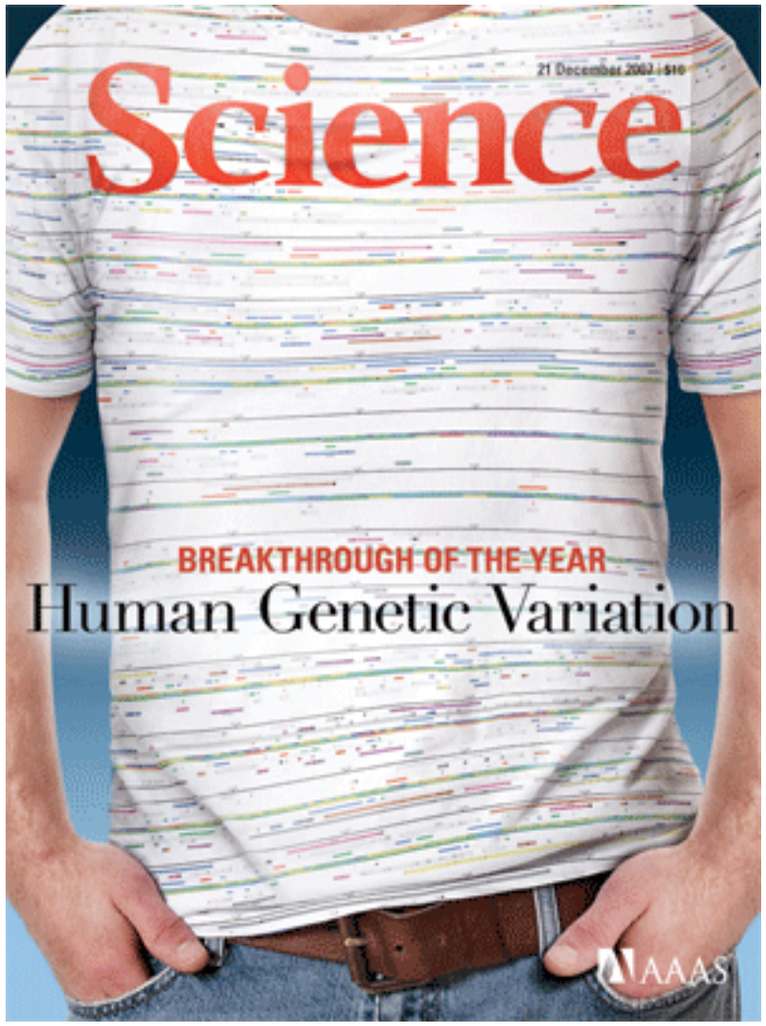
- **Mutational analysis of human ABC transporter nucleotide binding domains reveals functional hotspots**
  - **“Although I don't understand a word of your project title, I'm sure it makes sense to you.”**
  - **“I have no idea what your talk is about... it sounds very sci-fi to me.”**
  - **“Can’t wait for the movie version!”**
  - **“woo-hoo! I'm there...at least for the drinks”**

# An evolutionary approach to predicting variability in human drug response

- **What's known about human genetic variation and drug response?**
- **Examining the role of transport proteins in drug response and disease**
- **Predicting the effects of newly discovered variants on transporter function**
- **Validating these predictions in the wet lab**
- **Towards predicting drug interactions with transport proteins**

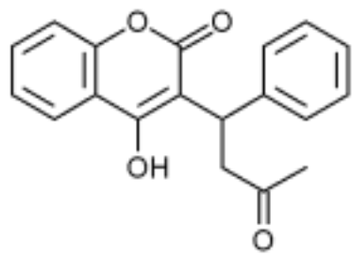
# The human genome has lots of variation

- **Over 10 million common variants in the human genome**
  - **~0.1% difference between two humans**
  - **~1.5% difference between a human and a chimp**
- **Sequences with variants are called alleles : UGT1A1*28**
- **Collections of alleles across some region of the genome is called a genotype**



# Commonly prescribed drugs where genetic variation affects response

- **Warfarin: anticoagulant prescribed to about two million patients in the US annually.**
  - **Narrow therapeutic range, dosage needs to be carefully monitored**



Warfarin

- **Variants in CYP2C9, an enzyme that metabolizes warfarin, are associated with:**
  - **warfarin maintenance dose**
  - **time to stable warfarin dosing**
  - **bleeding events in patients**
- **Variants differ by ethnic group**

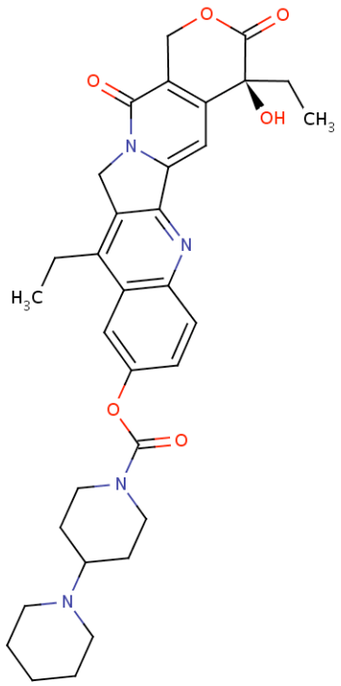
| Allele | *1/*1 | *2/*2 | *3/*3 |
|---|---|---|---|
| Daily dose (mg) | 5.63 | 4.88 | 3.32 |

| Ethnicity | "White", with 3.8% Hispanic |
|---|---|

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.

# Commonly prescribed drugs where genetic variation affects response

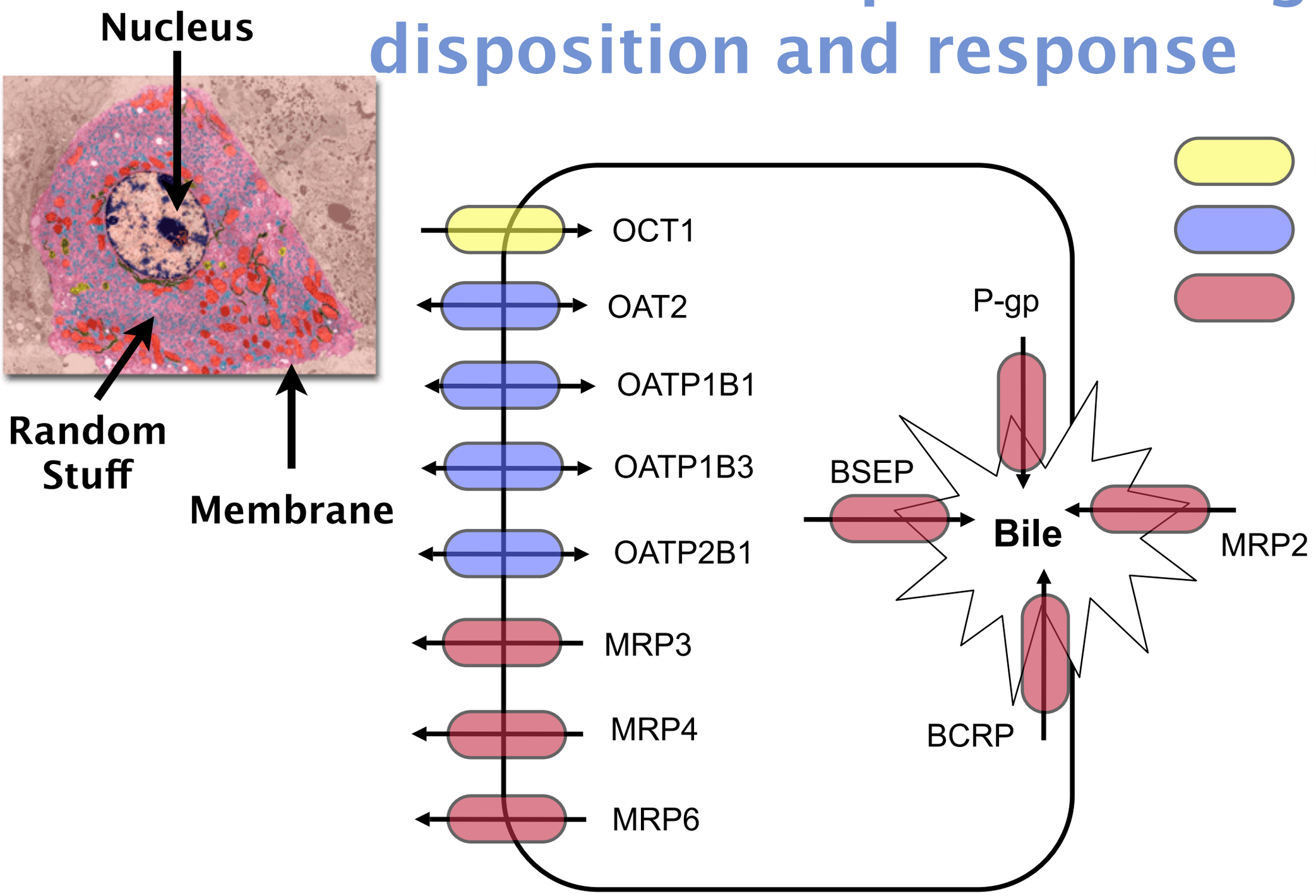
- **Irinotecan**: cancer therapy
  - Variants in membrane-bound protein UGT1A1 affect drug toxicity
  - Possible ability to predict response
  - ‘Personalized’ cancer chemotherapy




**Irinotecan**

Liu CY, Chen PM, Chiou TJ, Liu JH, Lin JK, Lin TC, Chen WS, Jiang JK, Wang HS, Wang WS.
UGT1A1*28 polymorphism predicts irinotecan-induced severe toxicities without affecting treatment outcome and survival in patients with metastatic colorectal carcinoma.
Cancer. 2008 May 1;112(9):1932-40.

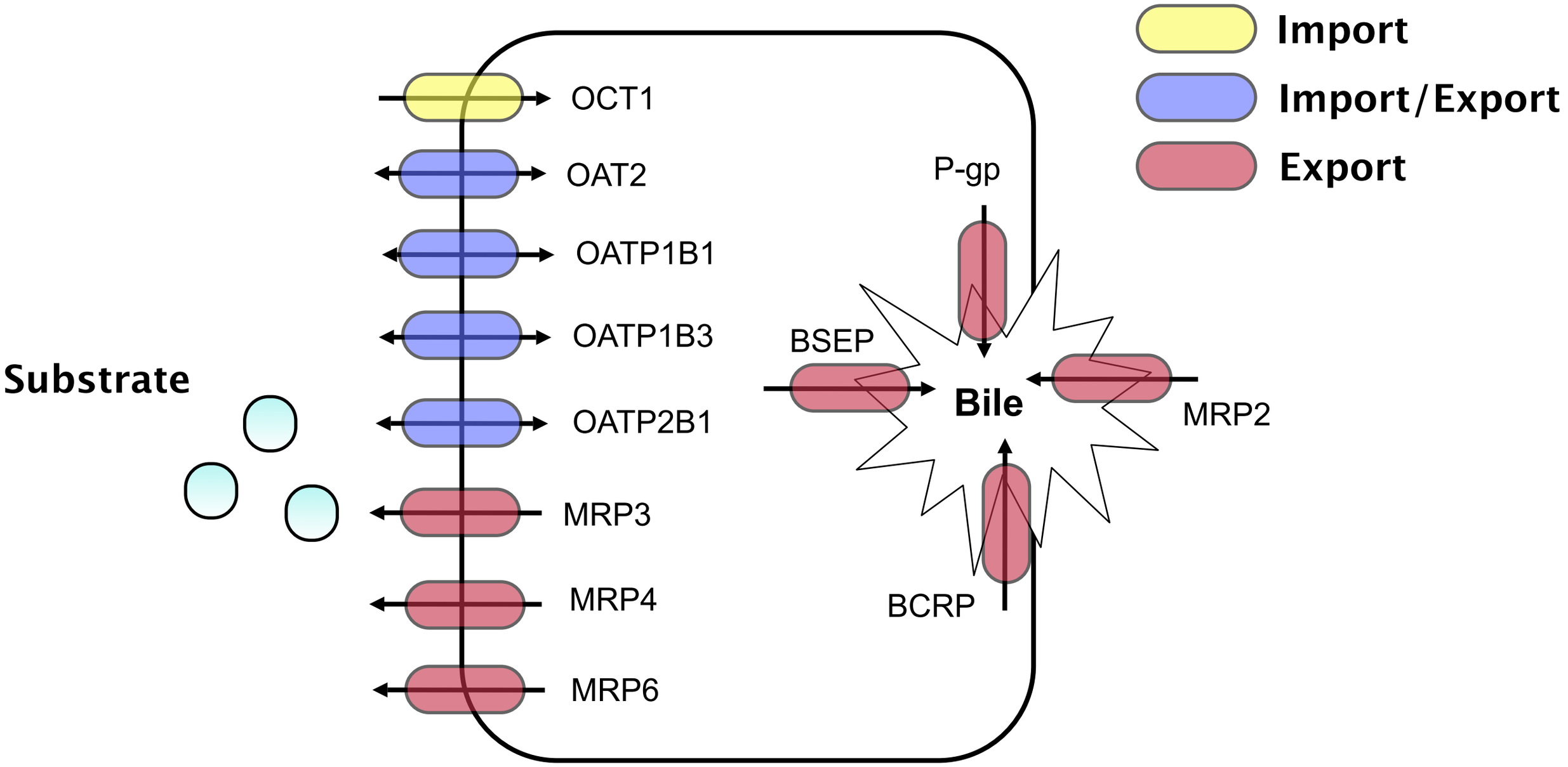
# Membrane transporters: drug disposition and response




- **Transporter proteins get drugs and other molecules in and out of cells**

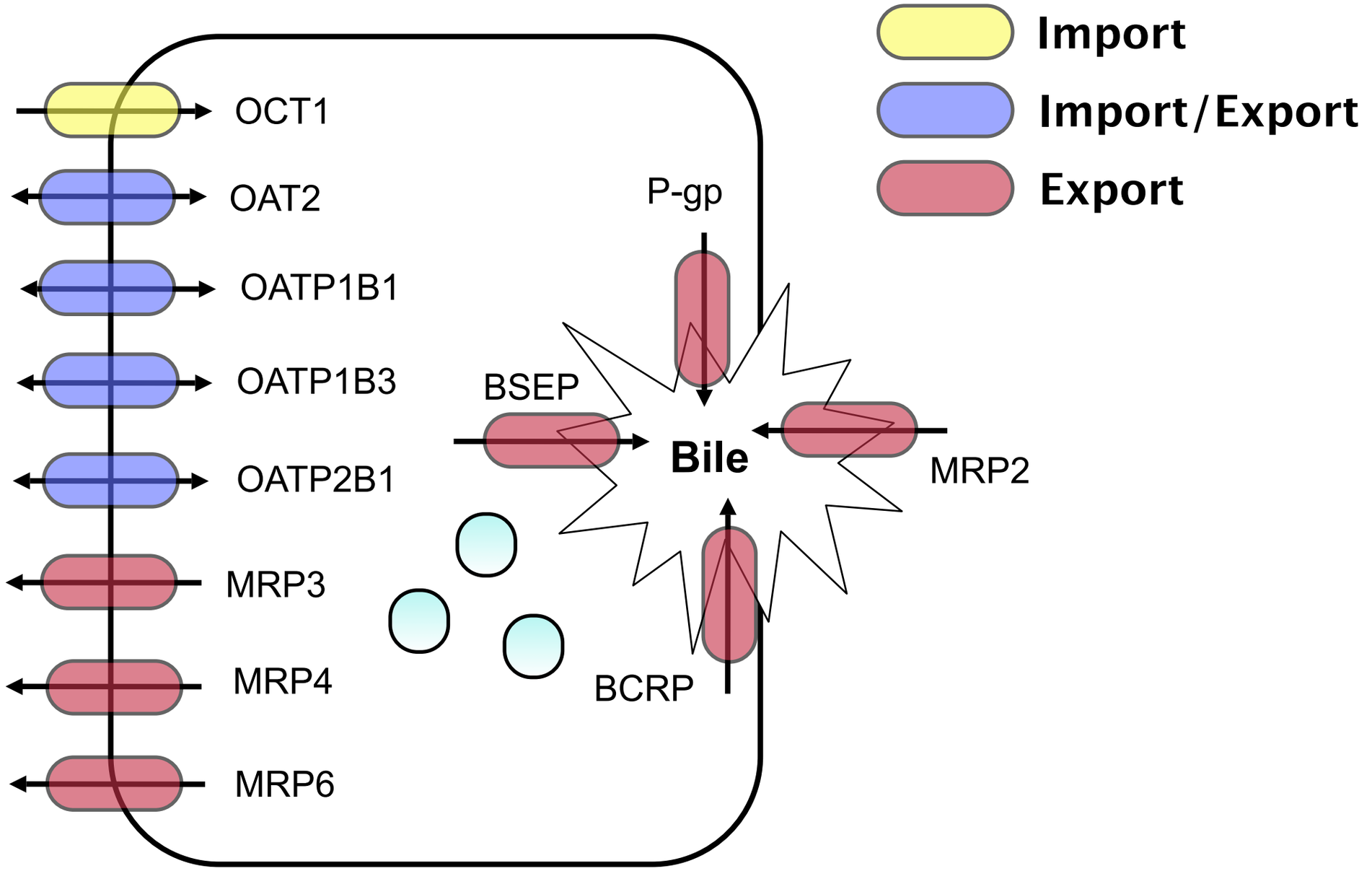
Dennis Kunkel Microscopy, Inc., Kathy Giacomini, UCSF

# Membrane transporters: drug disposition and response



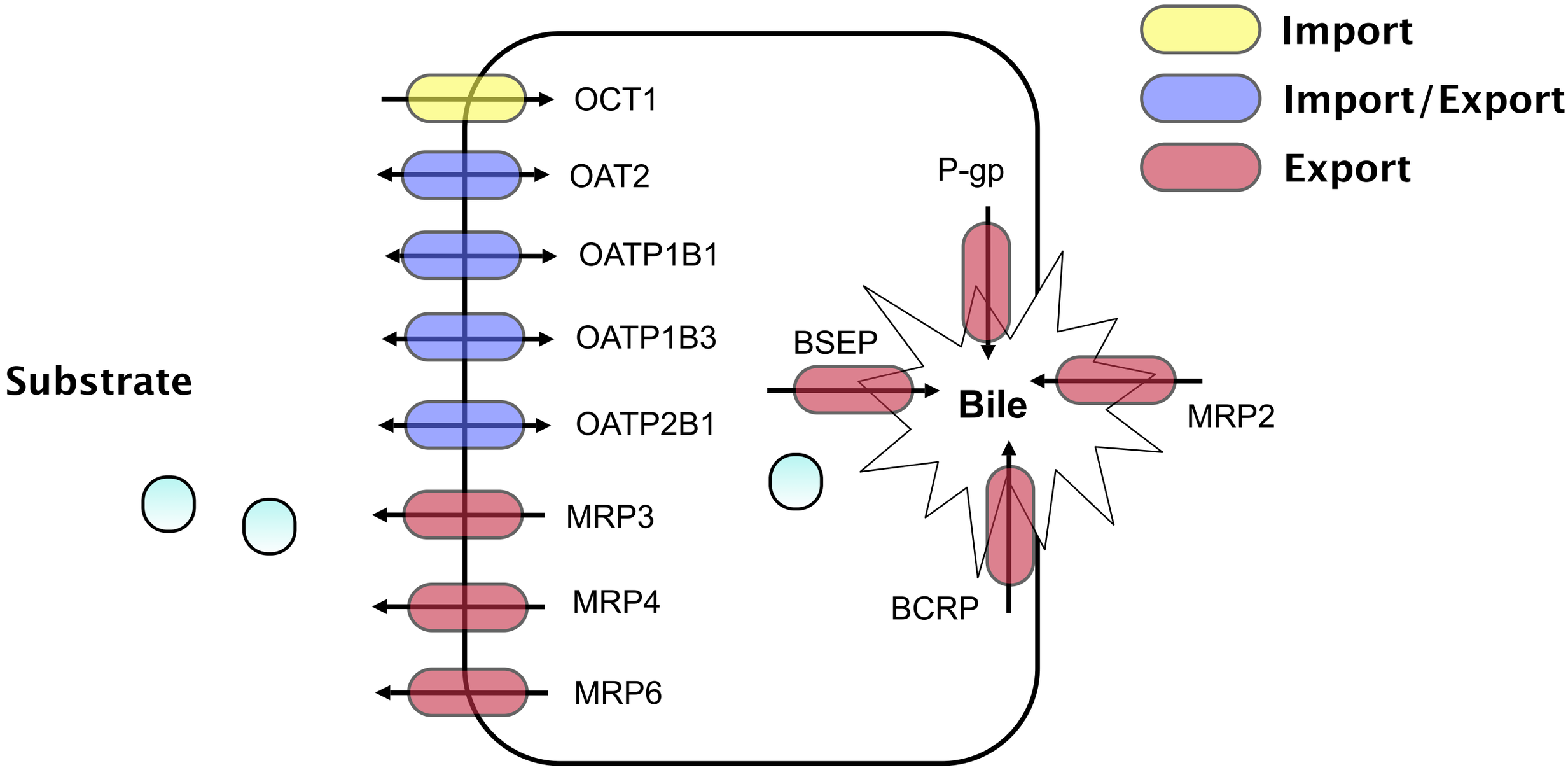

- **Transporter proteins get drugs and other molecules in and out of cells**

# Membrane transporters: drug disposition and response



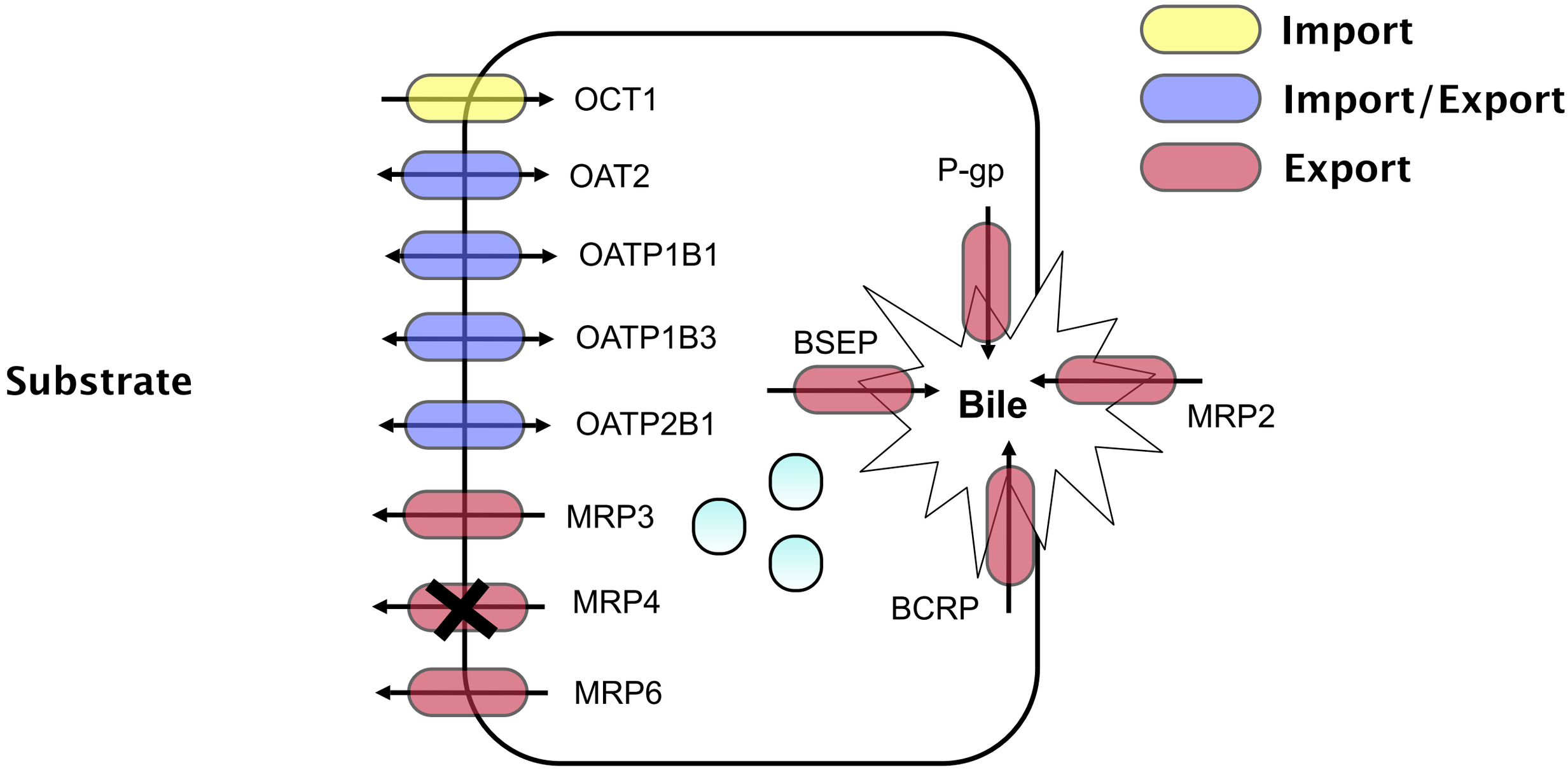

- **Transporter proteins get drugs and other molecules in and out of cells**

# Membrane transporters: drug disposition and response




- **Transporter proteins get drugs and other molecules in and out of cells**

# Membrane transporters: drug disposition and response




- **Transporter proteins get drugs and other molecules in and out of cells**

# Hypothesis: Genetic variation in membrane transporters contributes to drug response

# Hypothesis: Genetic variation in membrane transporters contributes to drug response

- **Catalogue genetic variation in the genes encoding membrane transporters in ethnically diverse human populations**

# Hypothesis: Genetic variation in membrane transporters contributes to drug response

- **Catalogue genetic variation in the genes encoding membrane transporters in ethnically diverse human populations**

- **Characterize the functional significance of variant transporters**

# Hypothesis: Genetic variation in membrane transporters contributes to drug response

- **Catalogue genetic variation in the genes encoding membrane transporters in ethnically diverse human populations**

- **Characterize the functional significance of variant transporters**

- **Use the structure and evolutionary history of membrane transporters to predict function**

# Hypothesis: Genetic variation in membrane transporters contributes to drug response

- Catalogue genetic variation in the genes encoding membrane transporters in ethnically diverse human populations

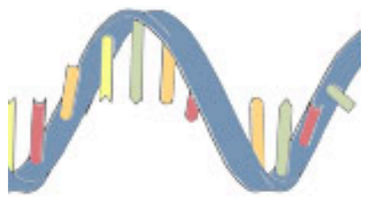
- Characterize the functional significance of variant transporters

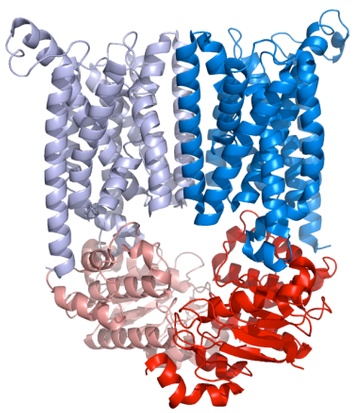
- Use the structure and evolutionary history of membrane transporters to predict function

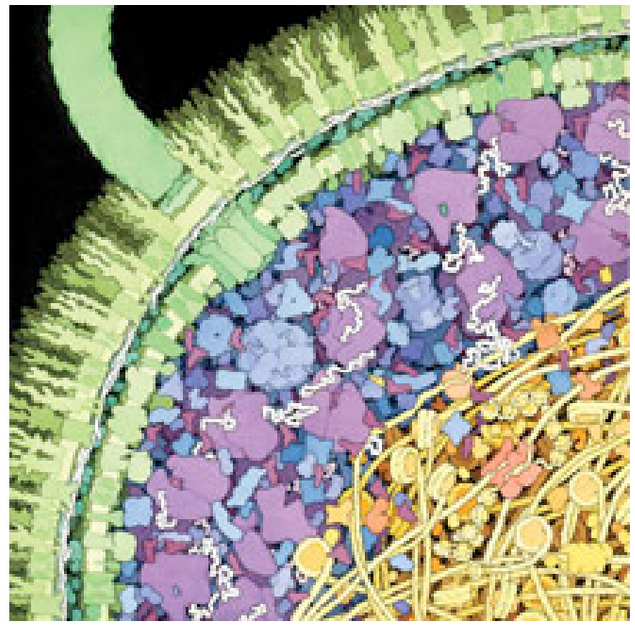
- Assess the role of membrane transporter variants in clinical drug response

# Single-nucleotide polymorphisms and protein function

**DNA** ➤ **RNA** ➤ **Protein** ➤ **Cell**







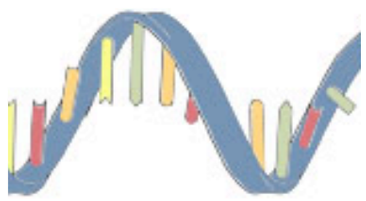
1. **...GTCACTGCGAAG... ...GUC ACU GCG AAG ... ...V T A K...**

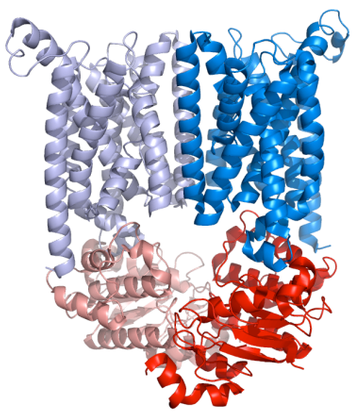
- **Single amino acid residue change**
- **Single protein**

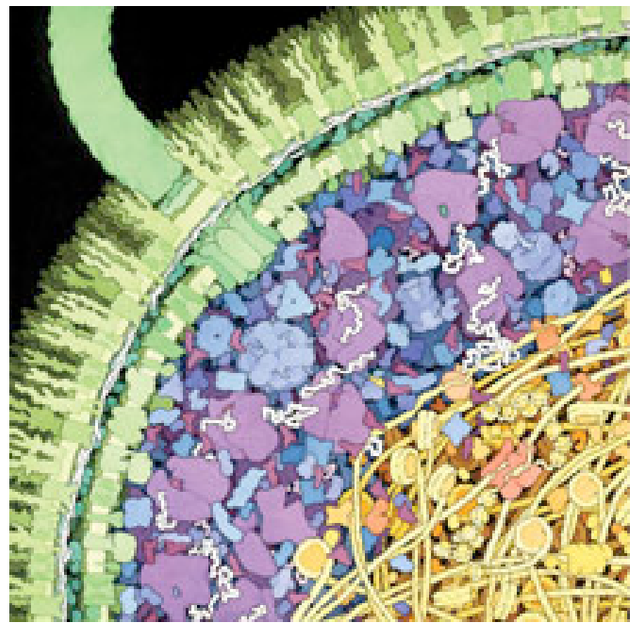
David Goodsell, Scripps and NHGRI

# Single-nucleotide polymorphisms and protein function

**DNA** ➤ **RNA** ➤ **Protein** ➤ **Cell**







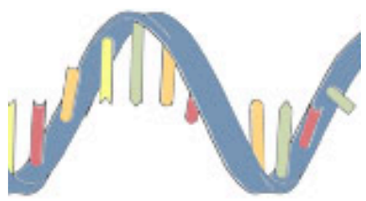
1. ...GTCACTGCGAAG... ...GUC ACU GCG AAG ... ...V T A K...
2. ...GTCGCTGCGAAG... ...GUC GCU GCG AAG ... ...V A A K...

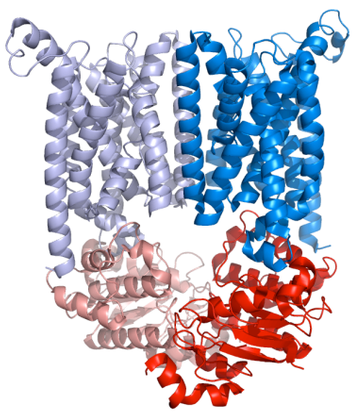
- **Single amino acid residue change**
- **Single protein**

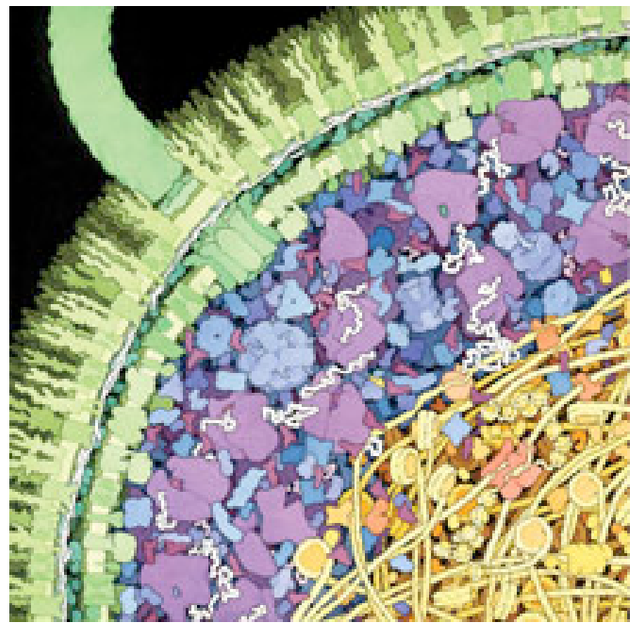
David Goodsell, Scripps and NHGRI

# Single-nucleotide polymorphisms and protein function

**DNA** ➤ **RNA** ➤ **Protein** ➤ **Cell**







1. ...GTCACTGCGAAG... ...GUC ACU GCG AAG ... ...V T A K...
2. ...GTCGCTGCGAAG... ...GUC GCU GCG AAG ... ...V A A K...

- **Single amino acid residue change**
- **Single protein**

David Goodsell, Scripps and NHGRI

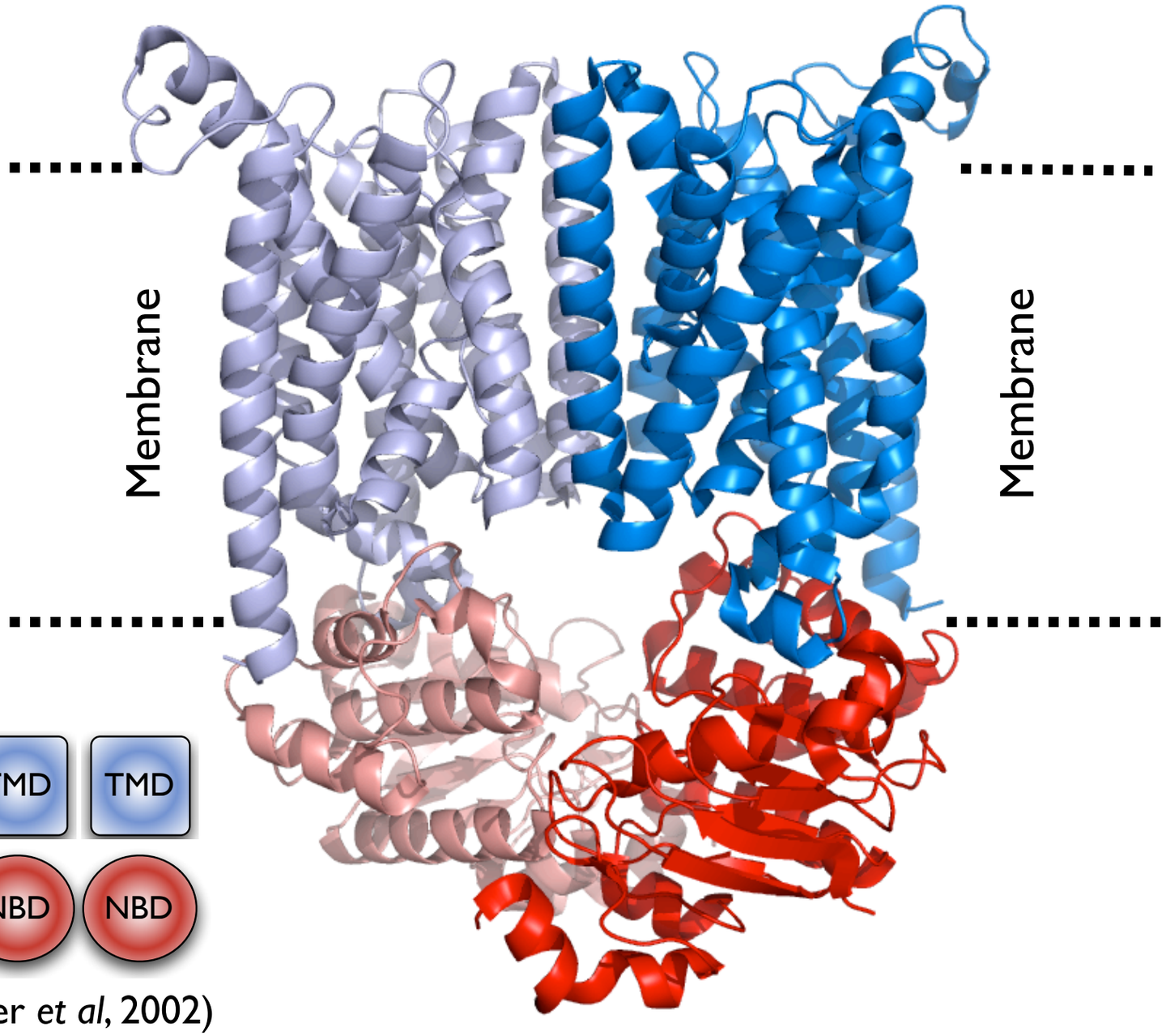
# ABC transporters play roles in disease and drug response

| 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 variants in human ABC transporters**
- **Functional analysis of mutations is lacking**

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

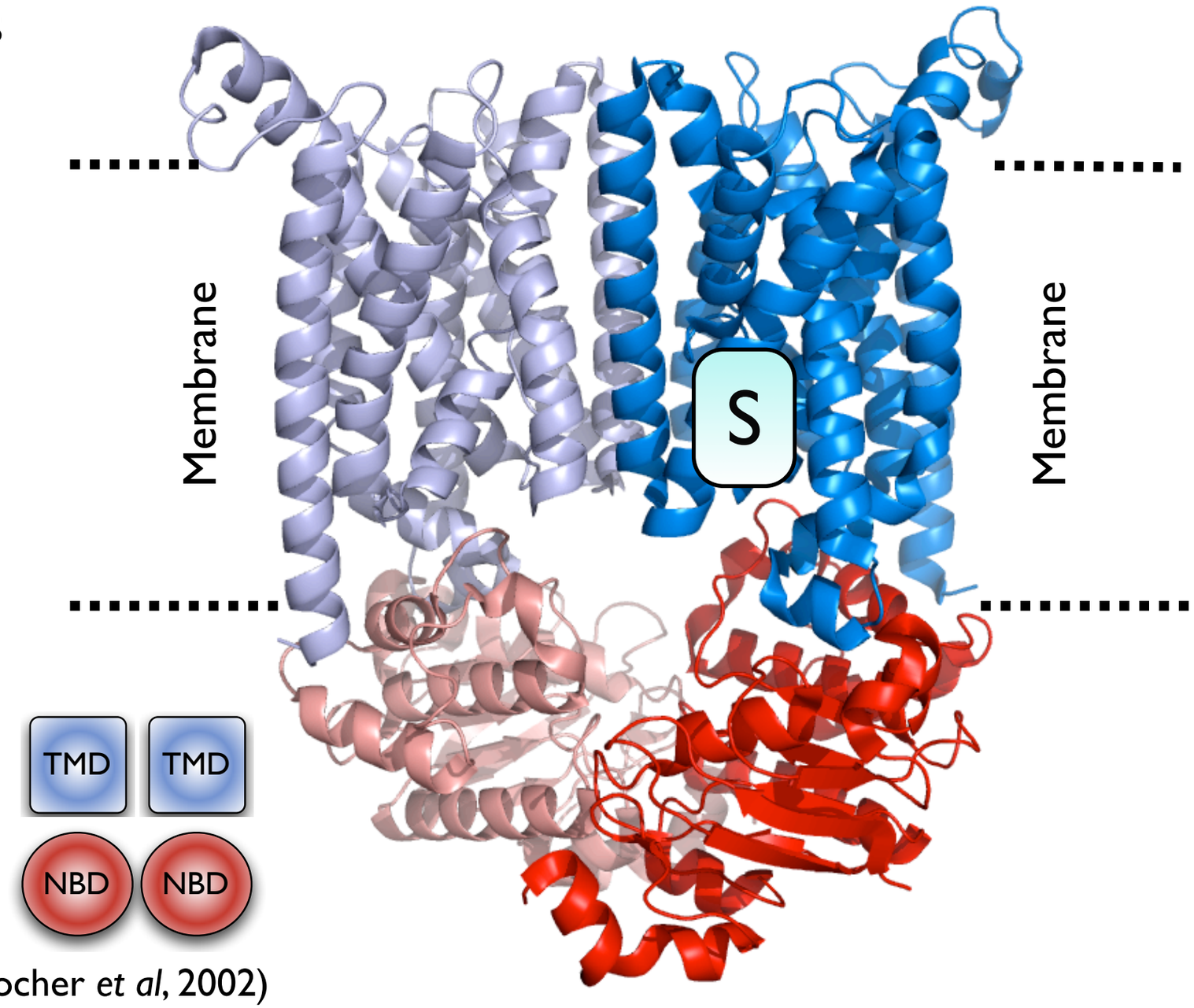
- **In humans, ABC transporters are active export pumps**
- **Substrate binds in transmembrane domains (TMDs), ATP binds in nucleotide binding domains (NBDs)**
- **ATP binding, hydrolysis and release are coupled to substrate transport**




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

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

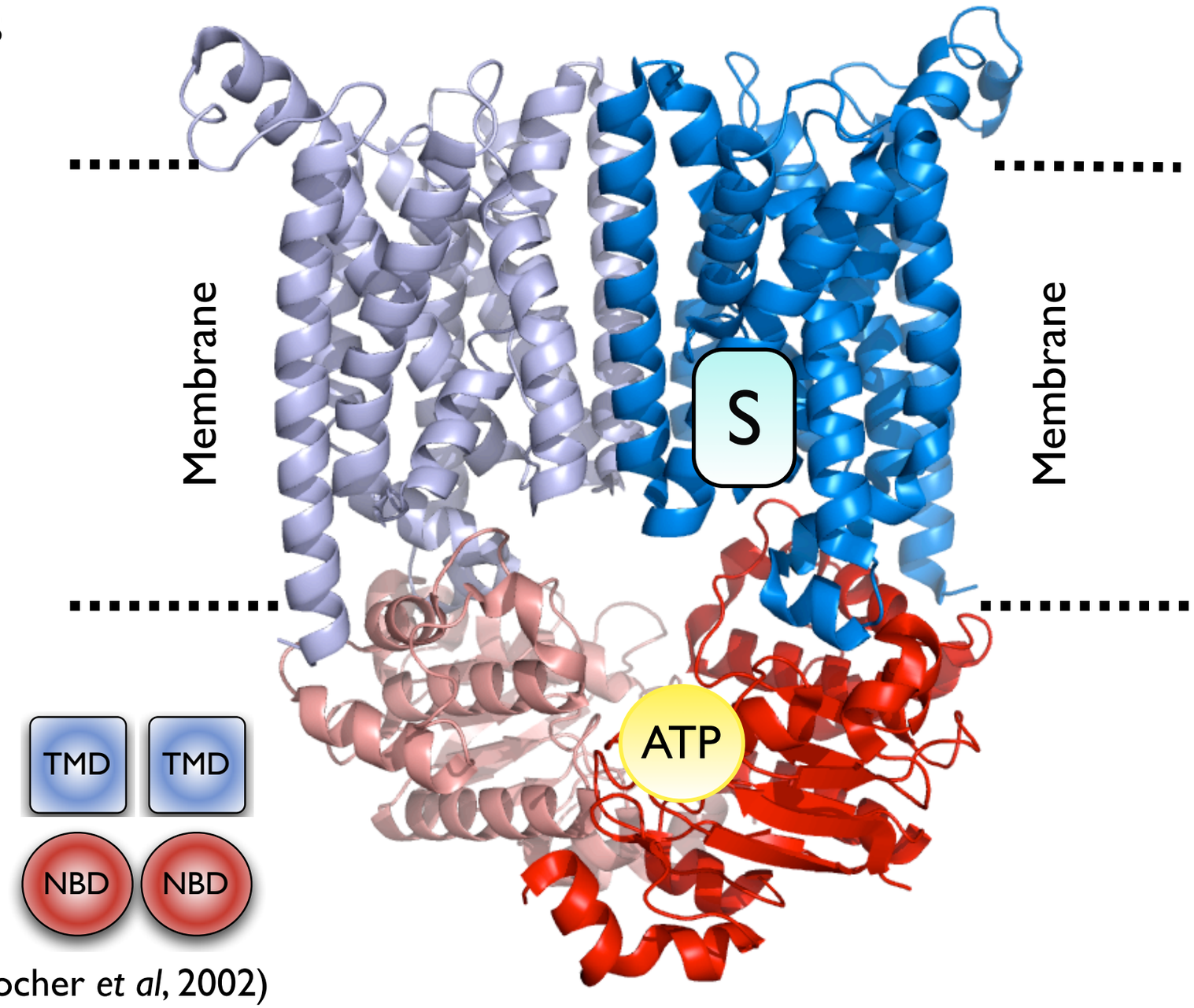
- **In humans, ABC transporters are active export pumps**
- **Substrate binds in transmembrane domains (TMDs), ATP binds in nucleotide binding domains (NBDs)**
- **ATP binding, hydrolysis and release are coupled to substrate transport**




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

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

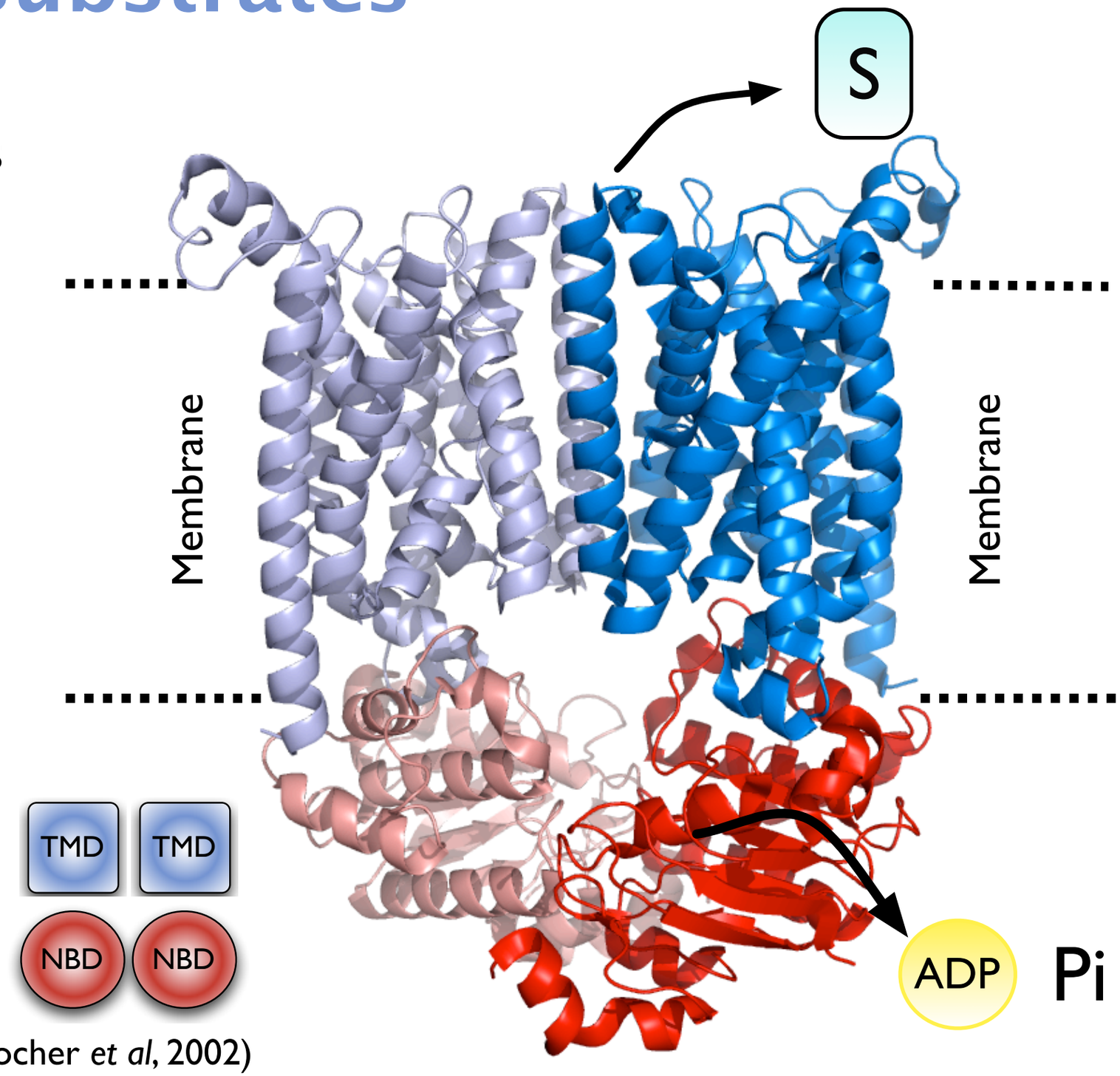
- **In humans, ABC transporters are active export pumps**
- **Substrate binds in transmembrane domains (TMDs), ATP binds in nucleotide binding domains (NBDs)**
- **ATP binding, hydrolysis and release are coupled to substrate transport**




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

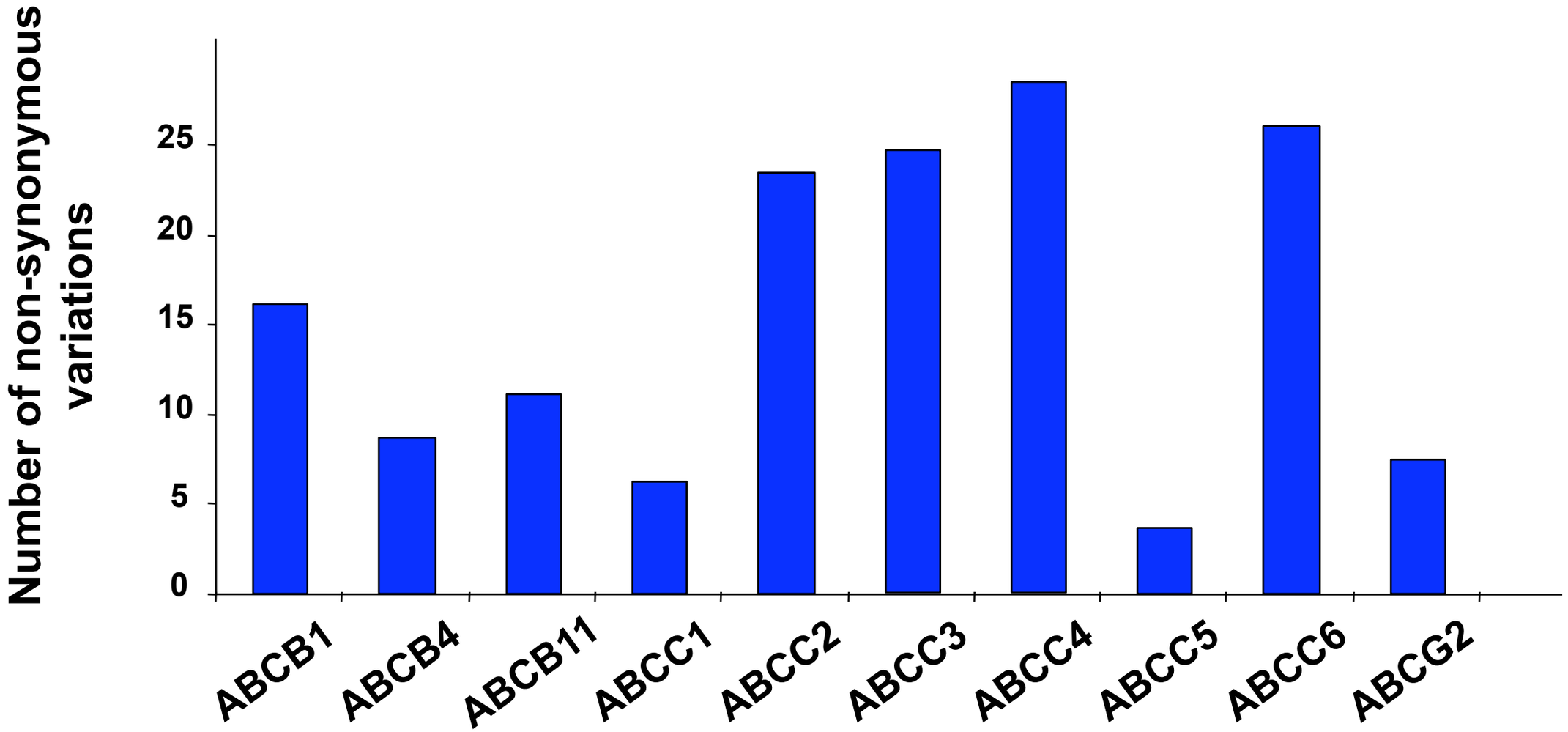
# ABC transporters are membrane proteins that bind a wide range of substrates

- **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**



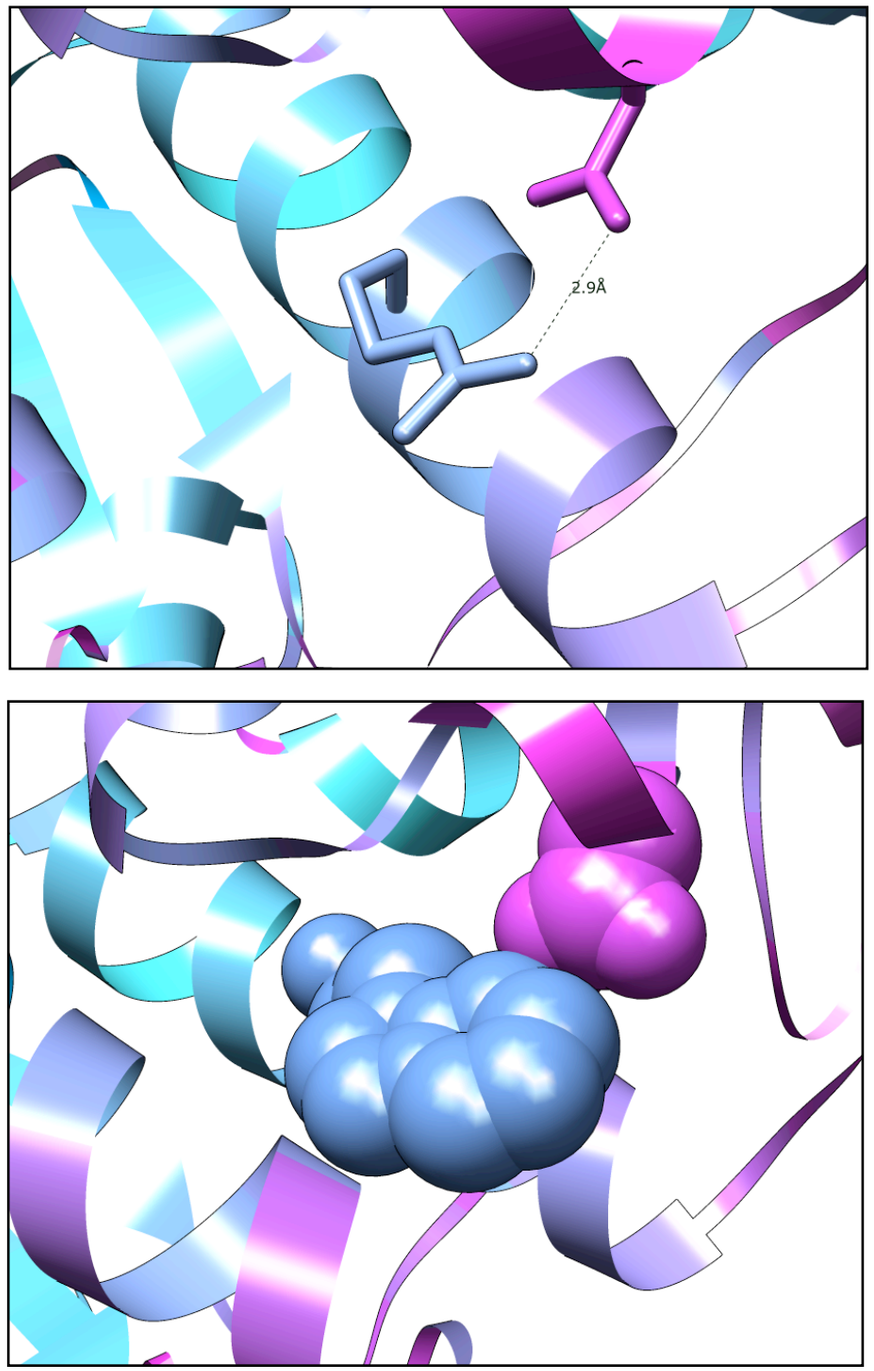

Structure of the *E. coli* BtuCD transporter. (Locher *et al*, 2002)

# Uncharacterized variation in 10 ABC transporters

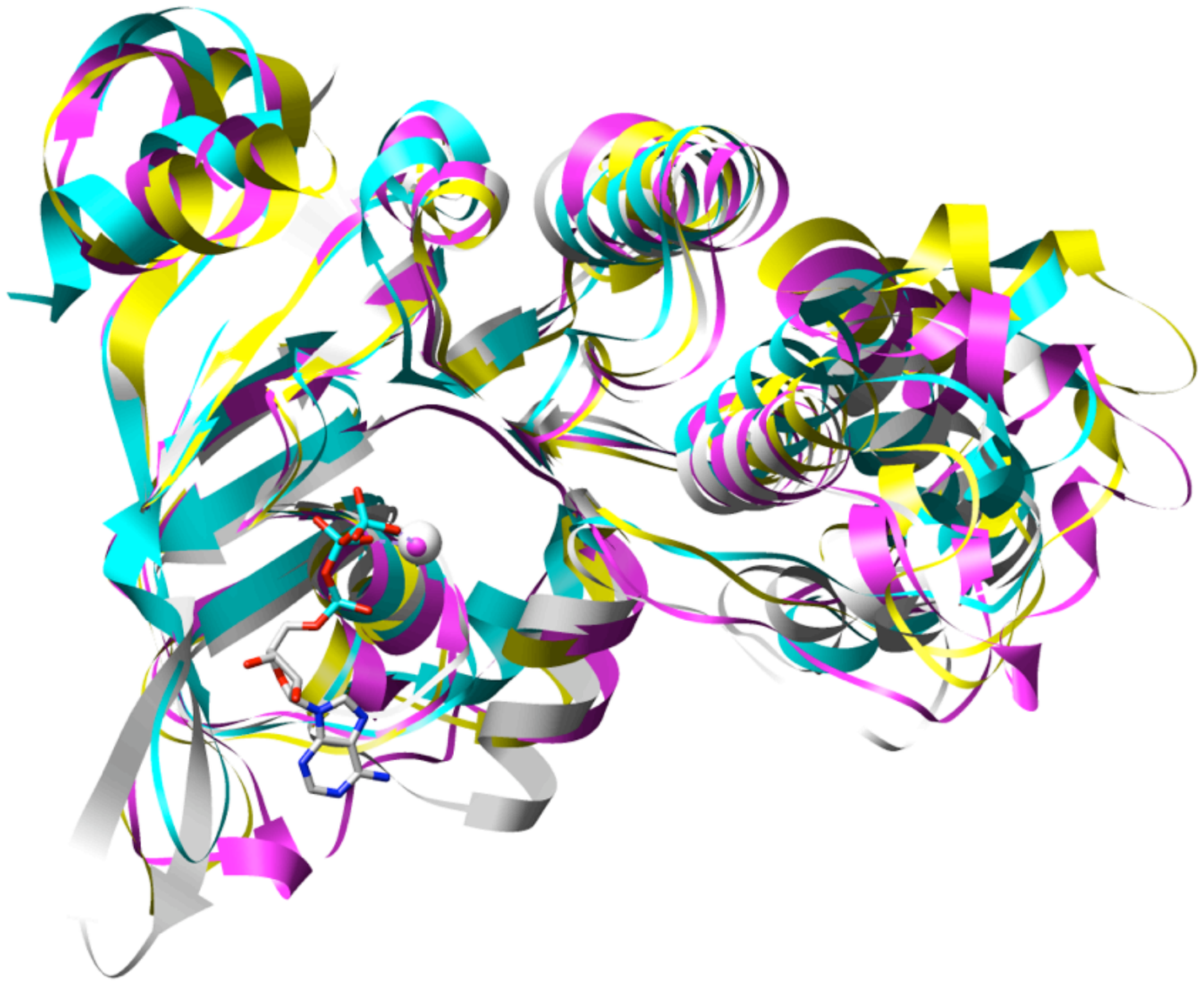


# Single residue changes can have a range of effects

- **Destabilizing:** force a charged residue into the generally hydrophobic interior of the protein
- **Impair domain interactions**
- **Impair residue interactions:** hydrogen bonding, salt bridges
- **Impair interactions with other proteins or ligands:** ATP binding, other partners



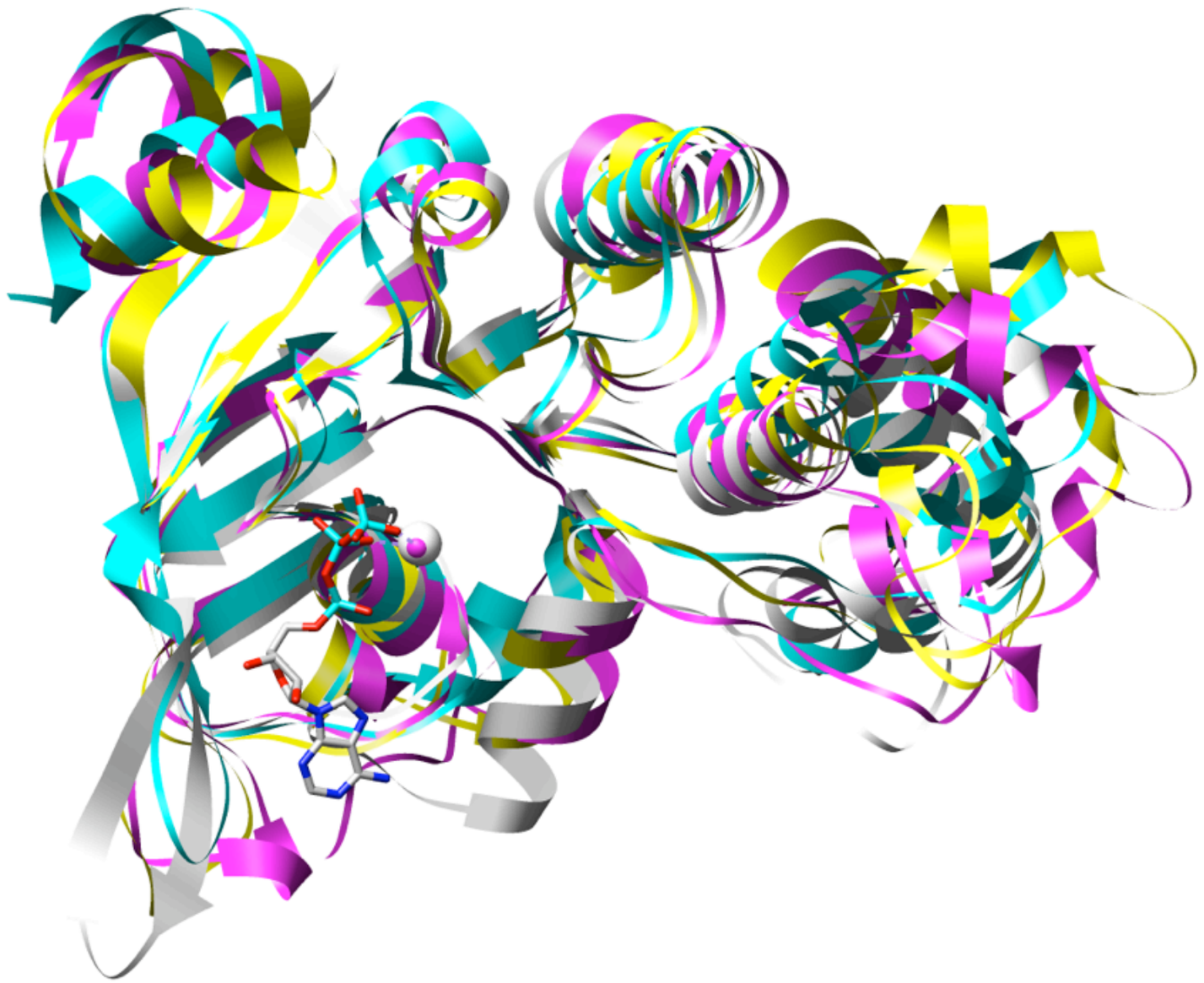
# Evolutionary conservation in ABC transporters



NBD structures of four ABC NBDs

# Evolutionary conservation in ABC transporters

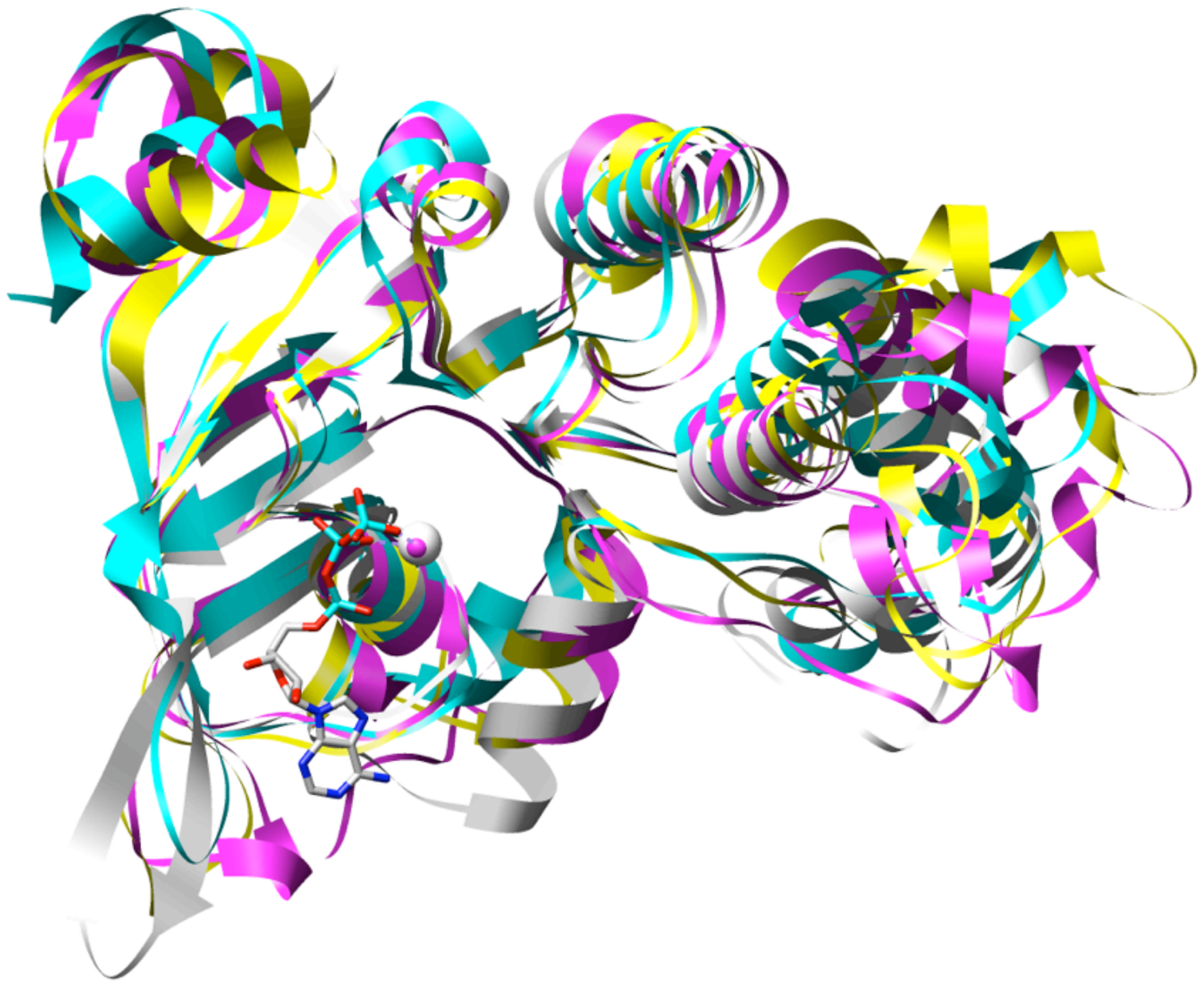
- **The overall fold of the nucleotide-binding domains (NBDs) is highly conserved**



NBD structures of four ABC NBDs

# Evolutionary conservation in ABC transporters

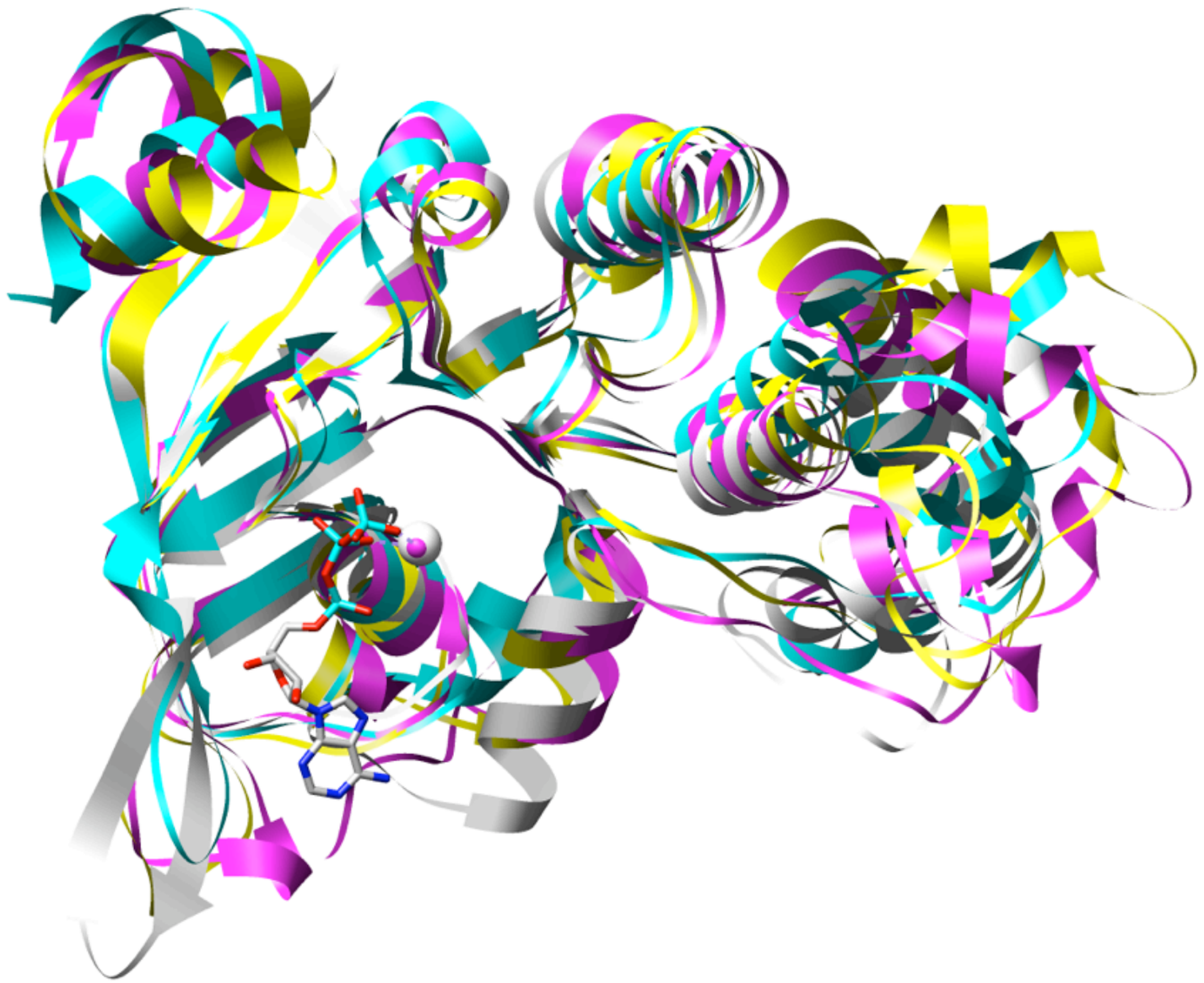
- **The overall fold of the nucleotide-binding domains (NBDs) is highly conserved**
- **Structural features are shared by human proteins**



NBD structures of four ABC NBDs

# Evolutionary conservation in ABC transporters

- **The overall fold of the nucleotide-binding domains (NBDs) is highly conserved**
- **Structural features are shared by human proteins**
- **This enables us to model human NBDs and variants based on homologs with known structure**

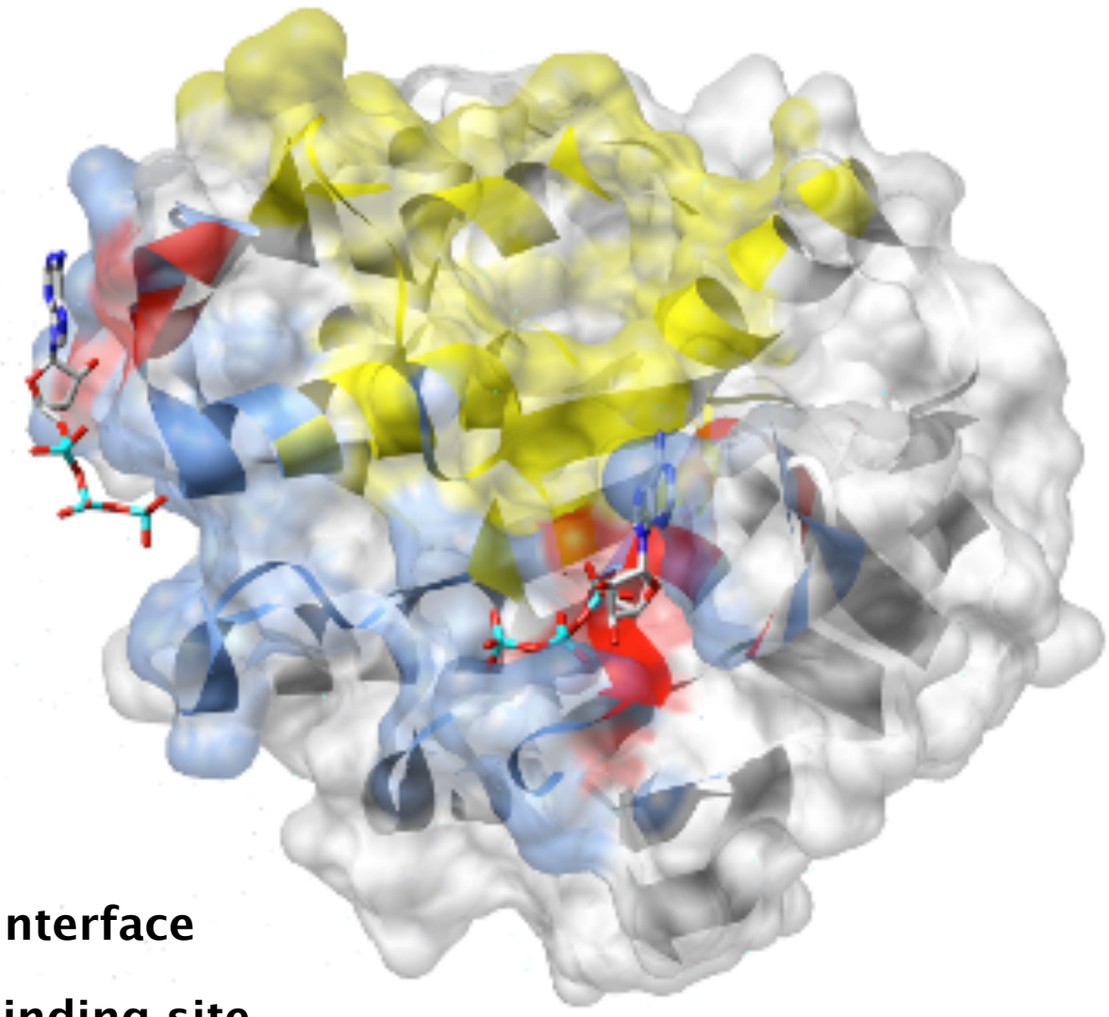


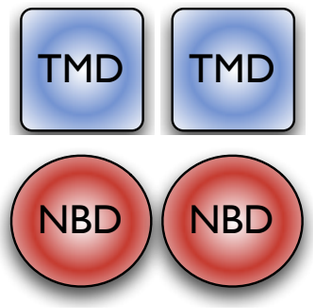
NBD

structures of four ABC NBDs

# Examining genetic variation at interfaces in ABC transporters

- **Does conservation vary at domain interfaces?**
- **Could some disease-associated mutants be affecting domain interactions?**






TMD interface

ATP-binding site

NBD interface

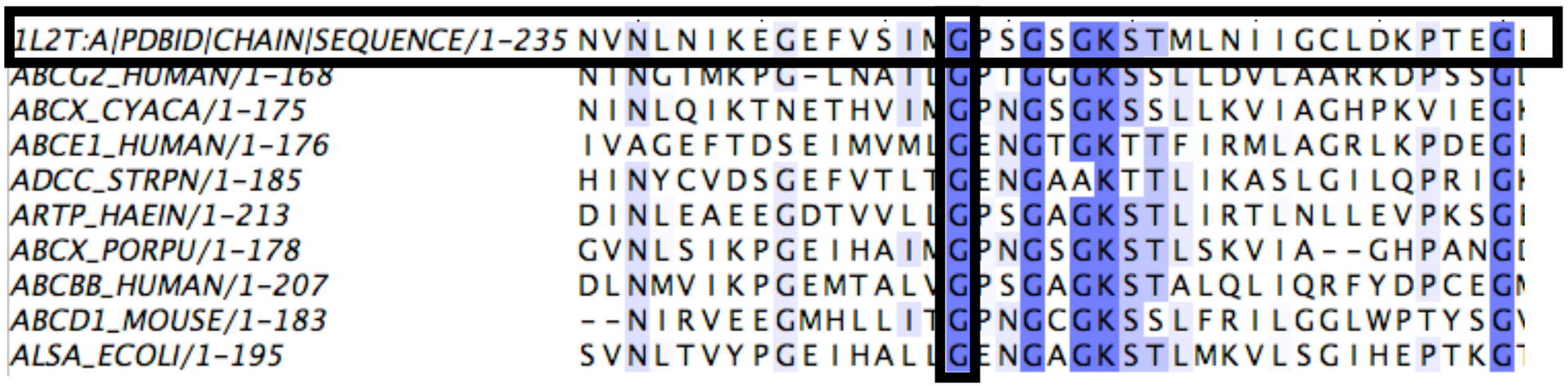

from *M. jannaschii (1L2T)*

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

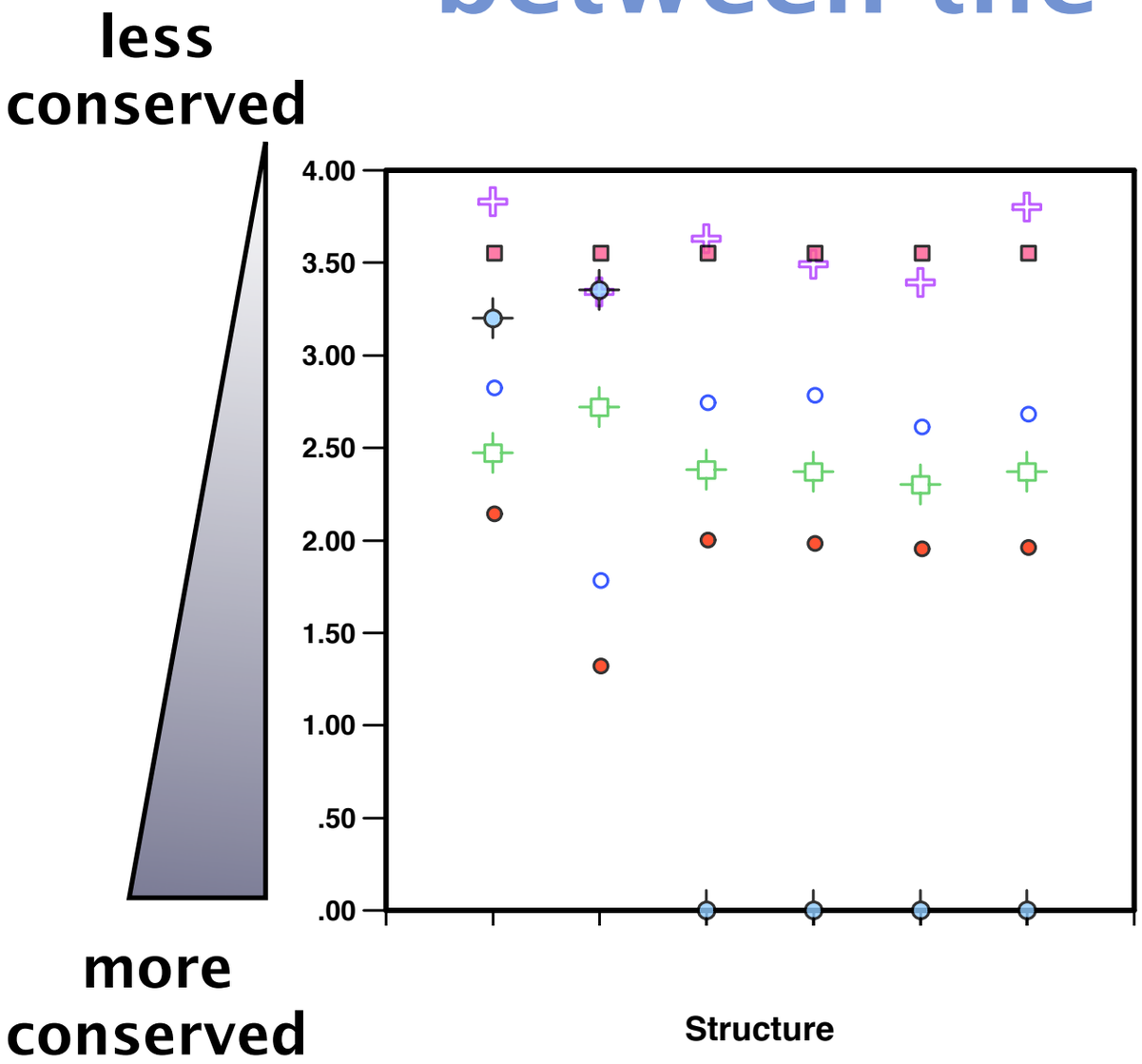
# Calculating sequence conservation at the interfaces

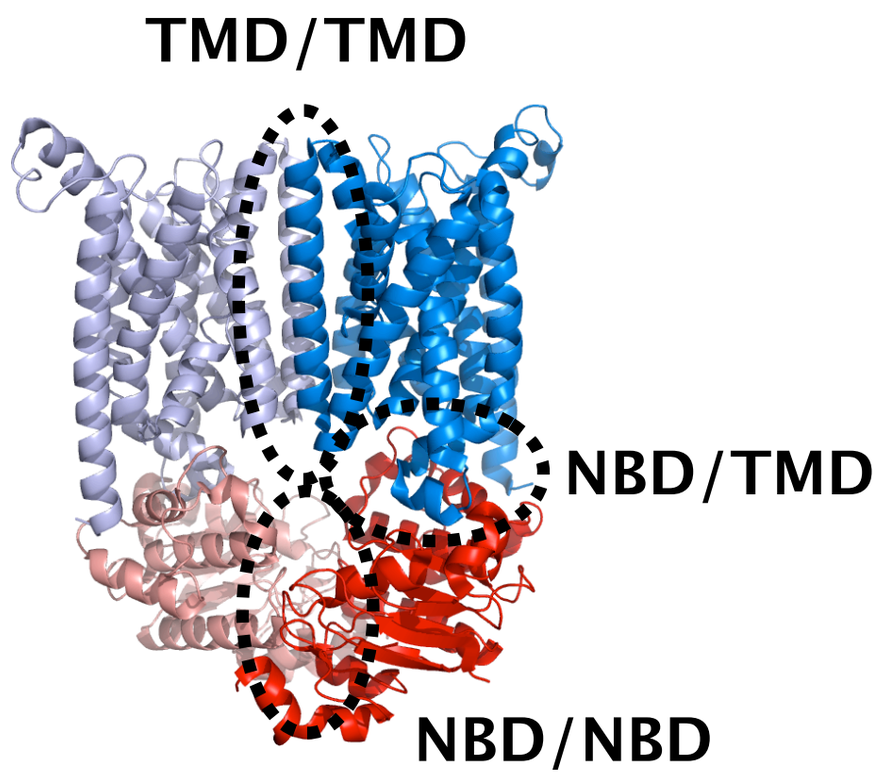
- **Multiple sequence alignments for each structure and for the NBDs of each human ABC transporter**
- **We calculate a measure of entropy in the alignment**

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



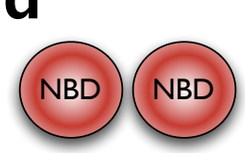
# Sequence conservation varies between the three interfaces

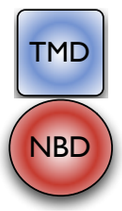



- NBD NBD interface was moderately conserved even when ATP binding residues were excluded.
- In contrast, TMD NBD and TMD TMD interfaces are not conserved

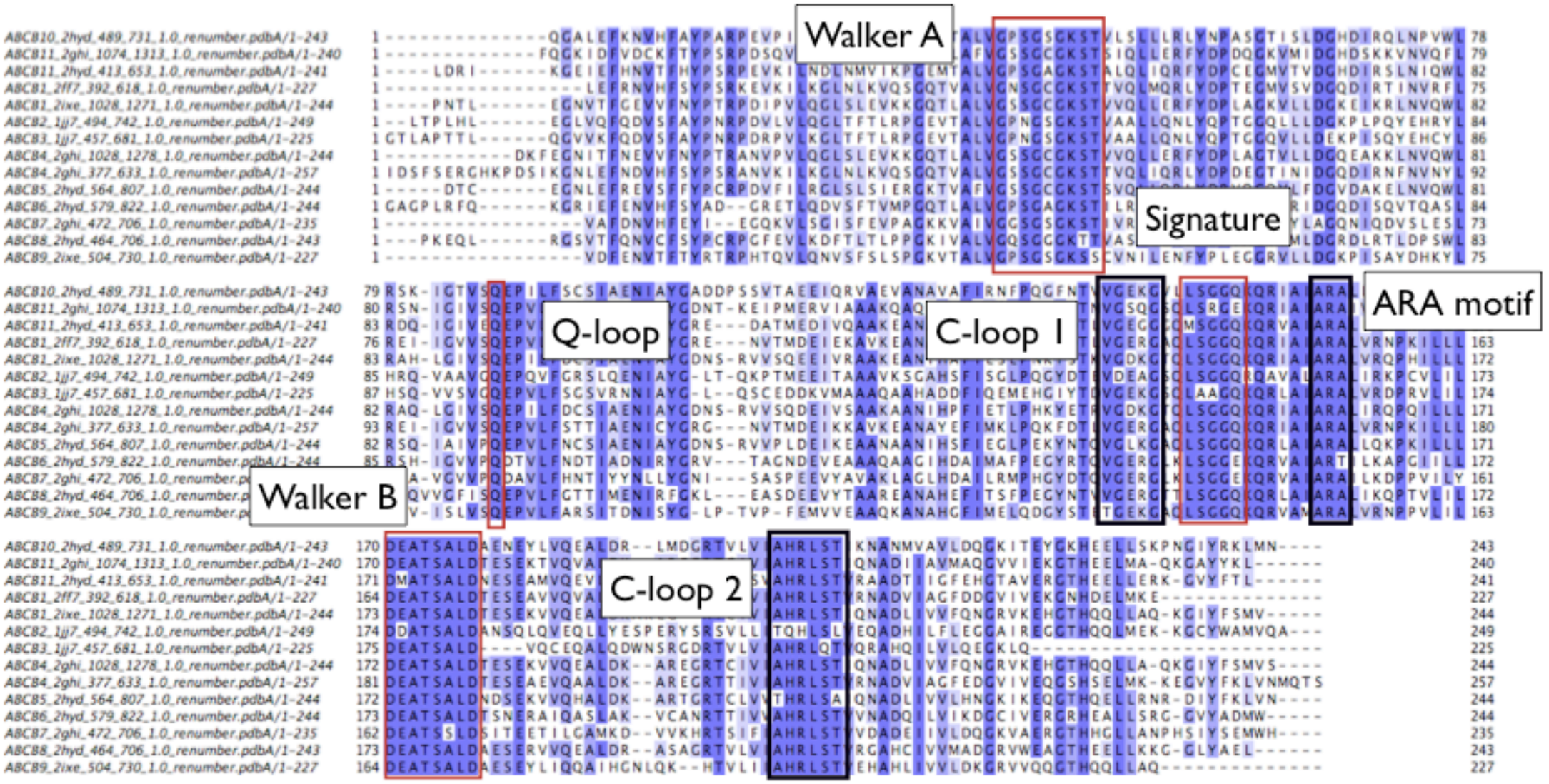
# We found 68 disease-associated positions at putative interfaces

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

interface
- 30 were at the 

interface
- Characterizing analogous interface residues in the human ABC transporter MRP4 to examine the functional effects of point mutants at the TMD/NBD interface

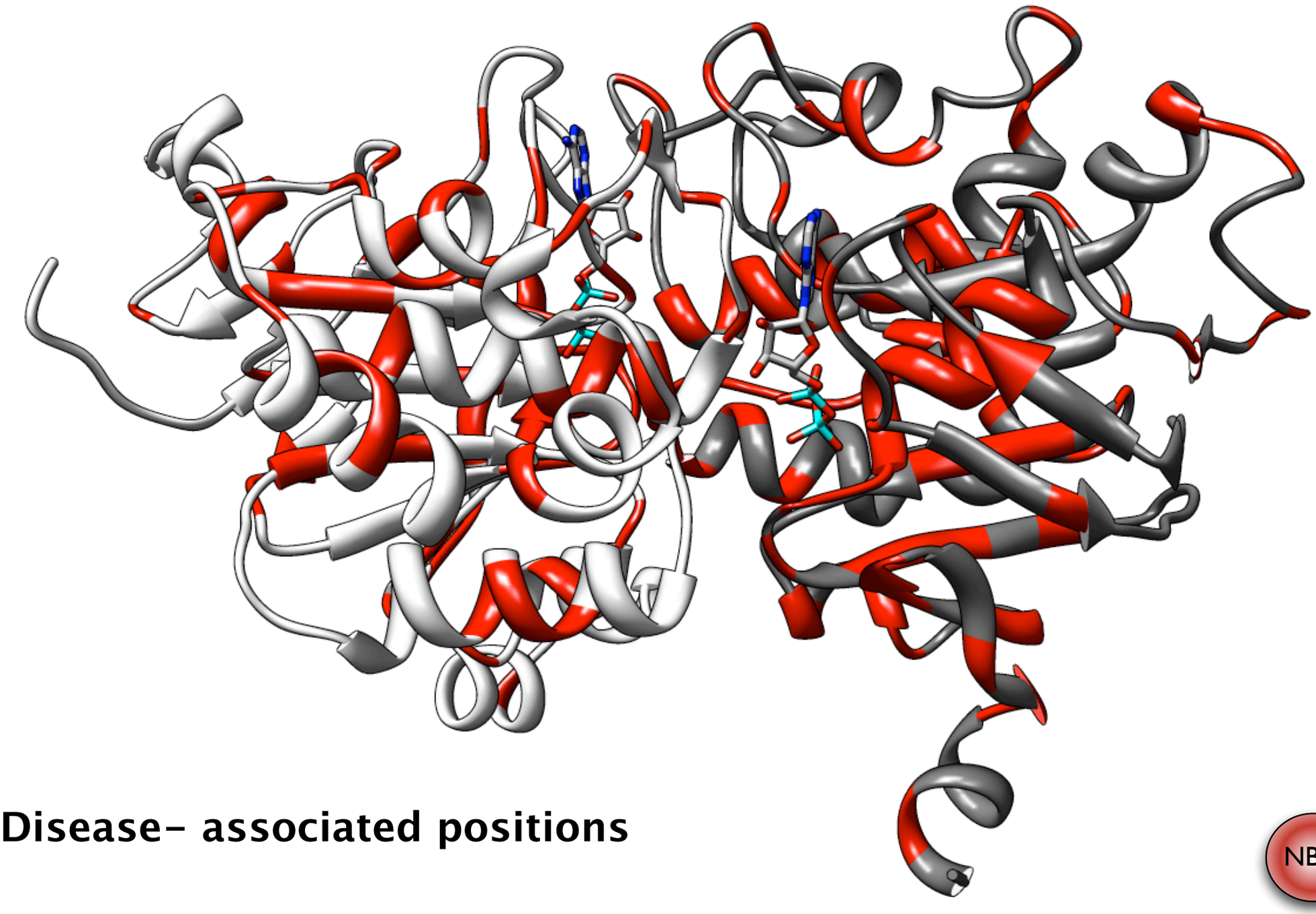
# Sequence alignments suggest functionally important regions




Known functional hotspots

Proposed functional hotspots

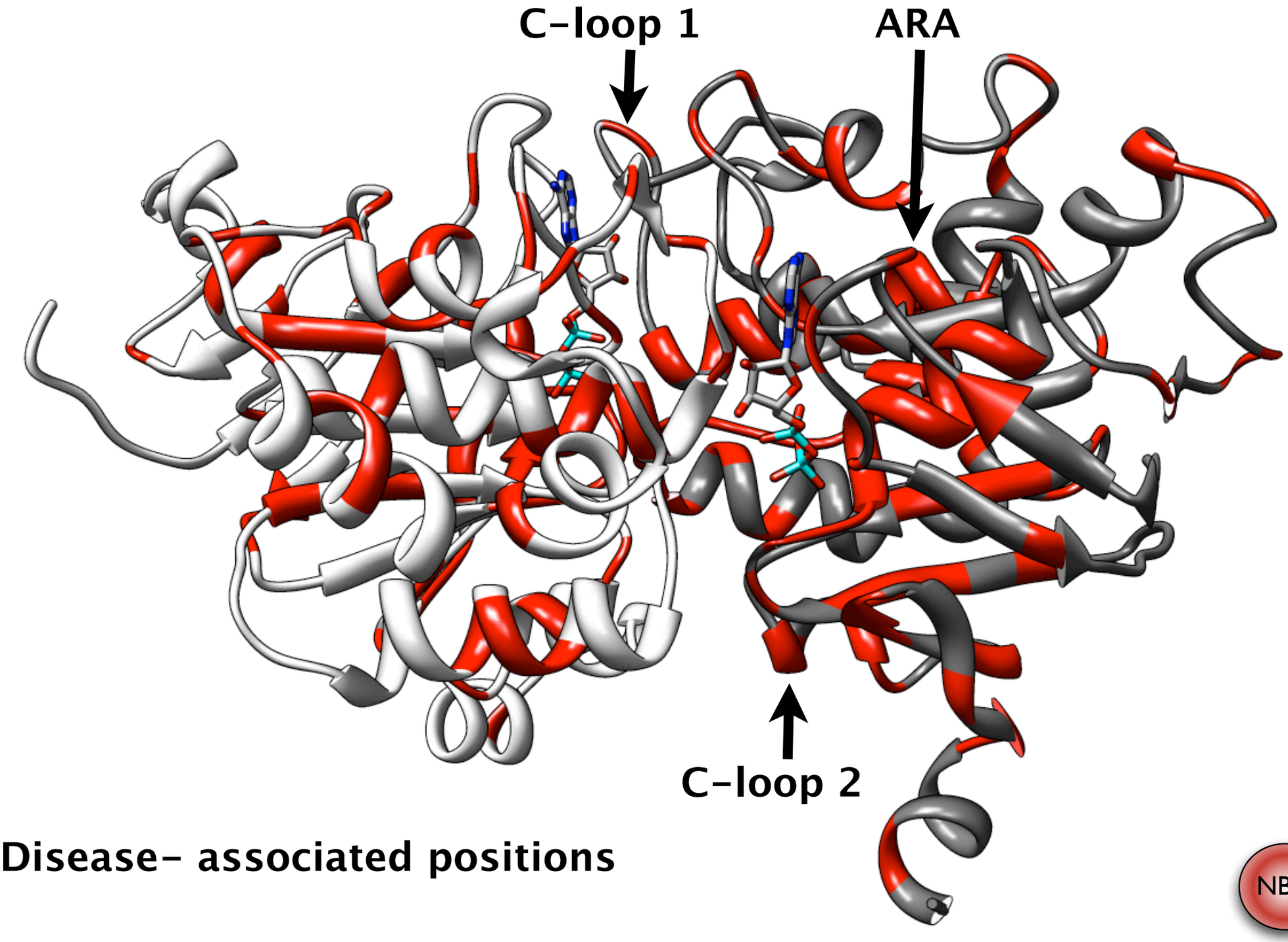
# Disease-associated variants are spread out across the domains



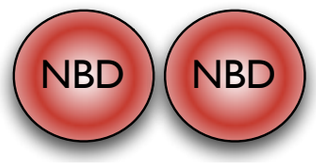
Disease- associated positions

NBD NBD

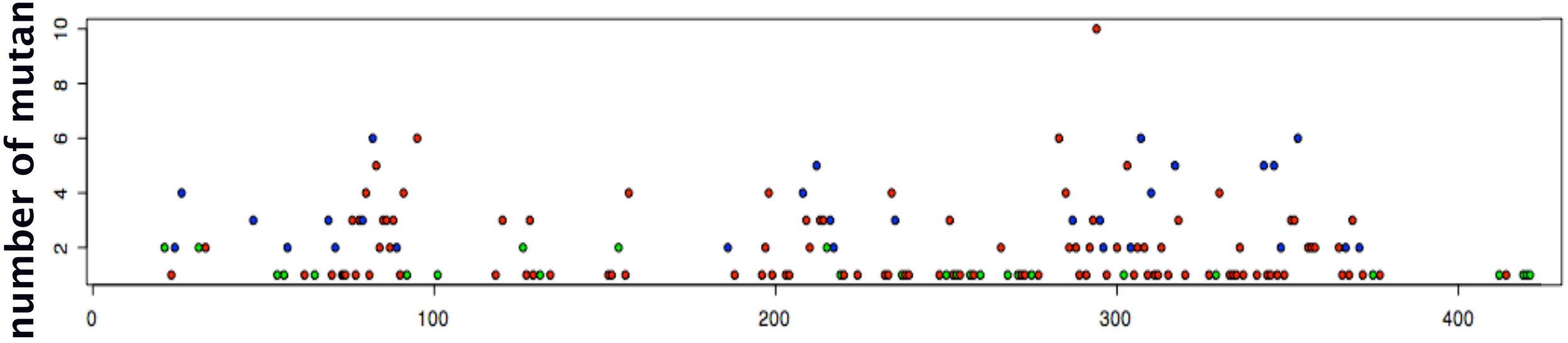
# Disease-associated variants are spread out across the domains



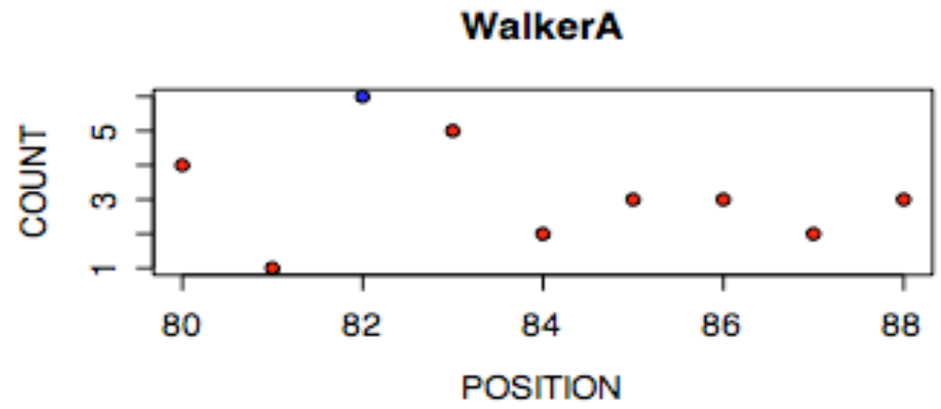

Disease- associated positions



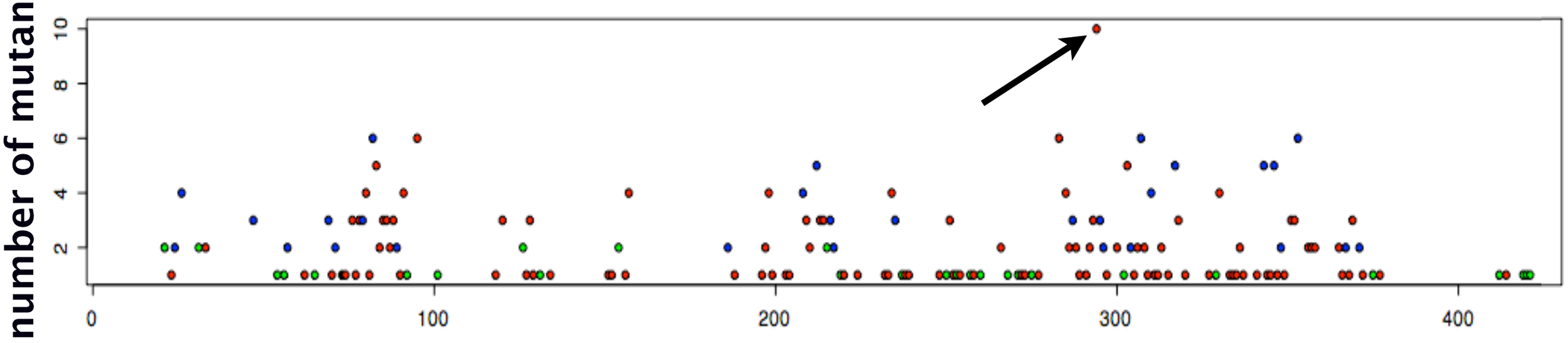
# Mapping mutations to structure suggests mechanisms of effect



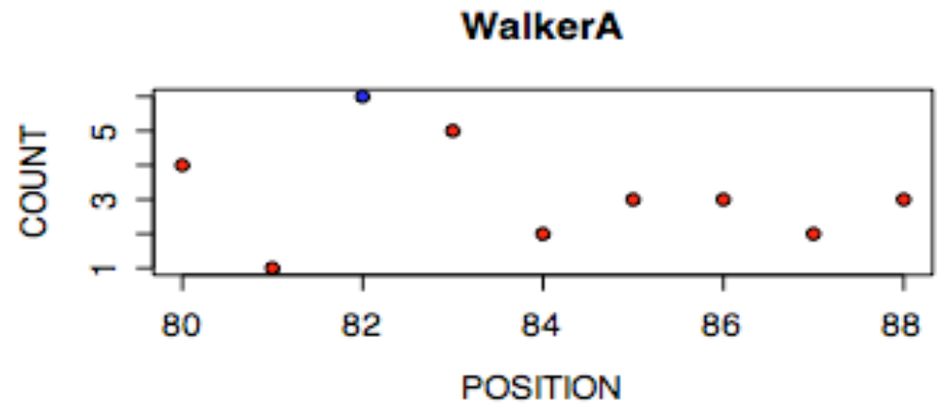

- Mutations mapped to a multiple structure alignment of all human NBD comparative models.
- Known motifs show disease mutations in multiple transporters.



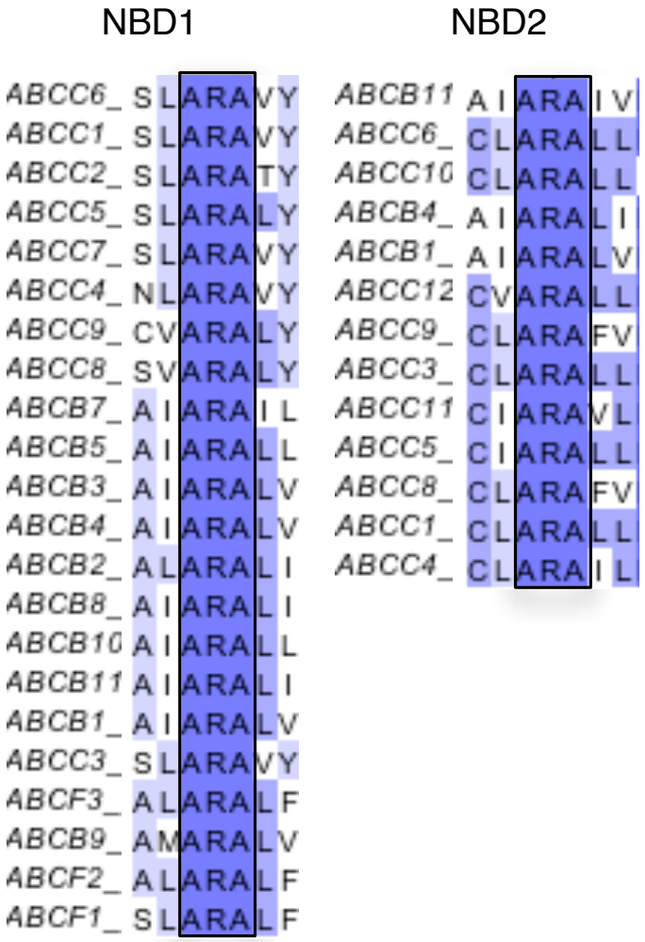
# Mapping mutations to structure suggests mechanisms of effect

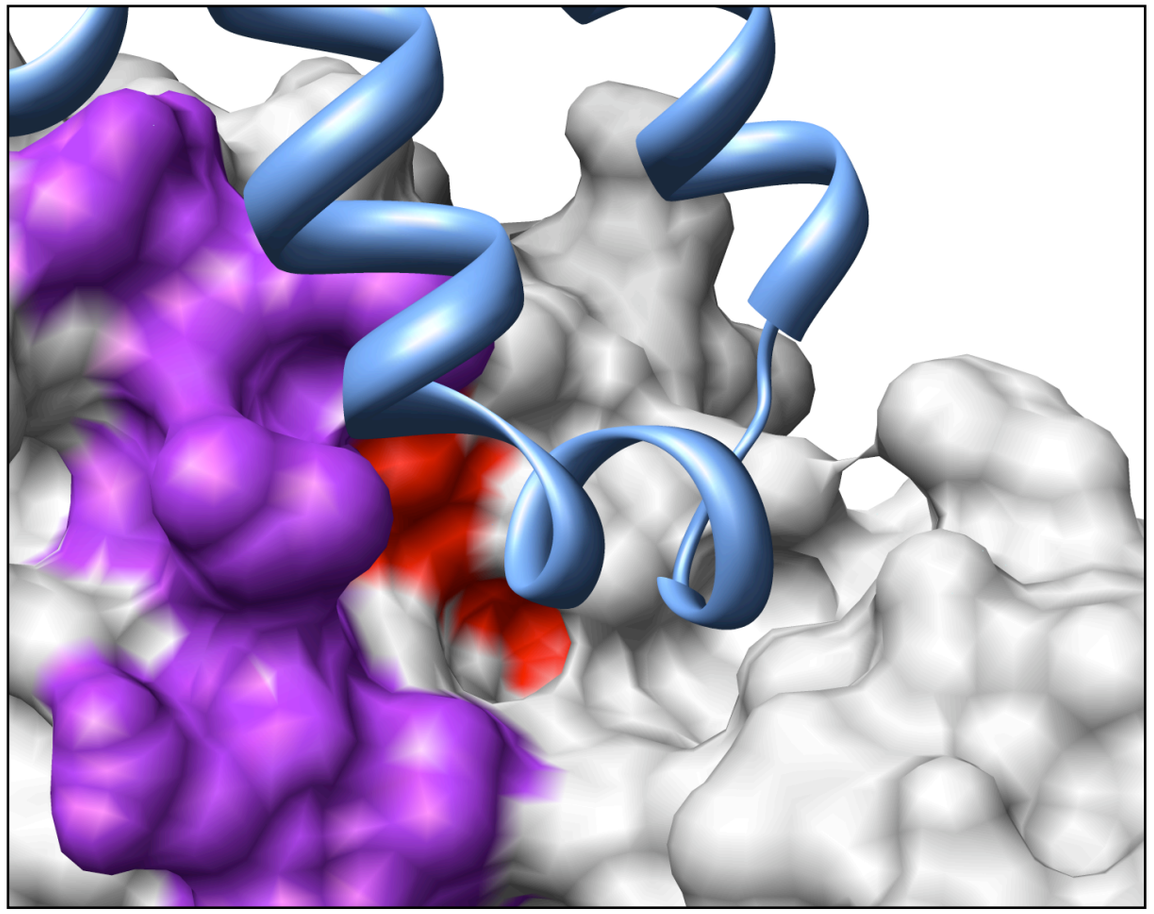



- Mutations mapped to a multiple structure alignment of all human NBD comparative models.
- Known motifs show disease mutations in multiple transporters.



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

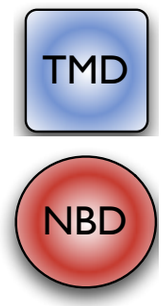


- Well conserved motif at the TMD/NBD interface

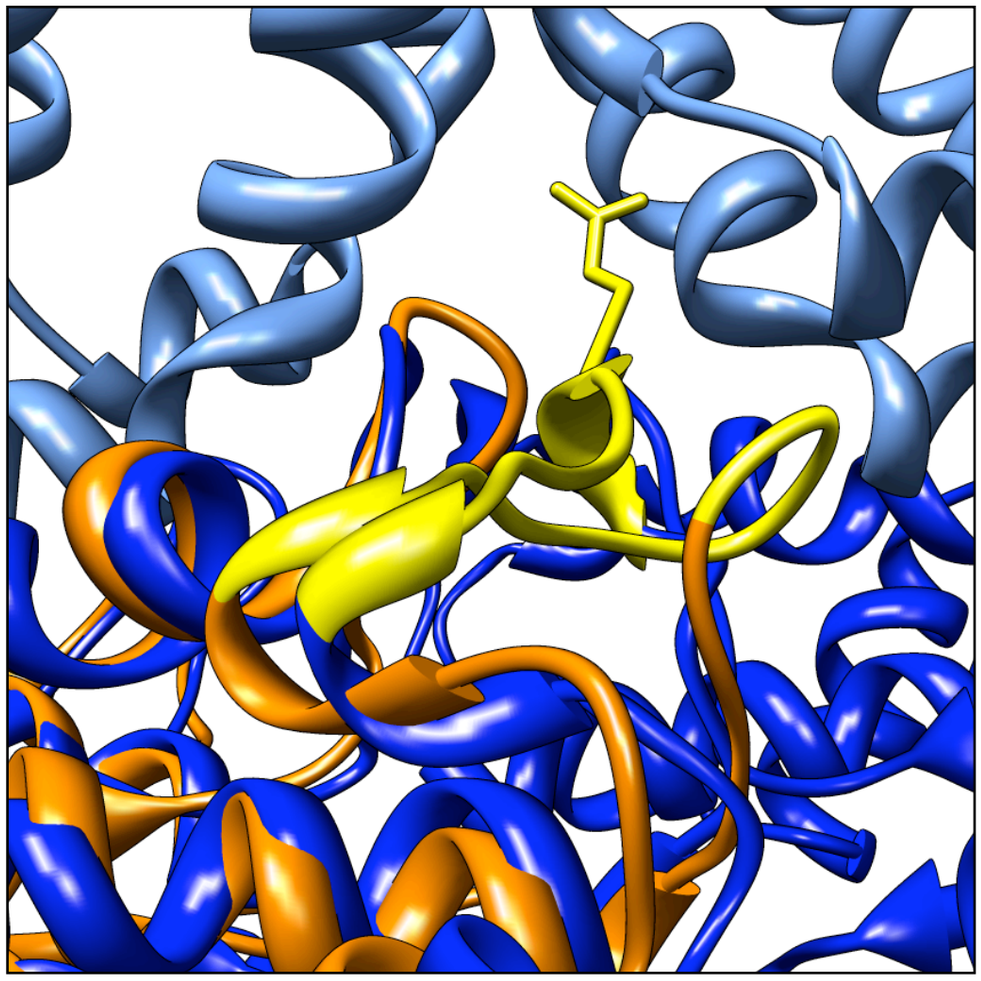
- 10 disease associated mutations

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



# Family-specific domain interactions: C-loop 1

- **Loop oriented toward TMD**
- **Conserved in the ABCB and ABCC subfamilies, absent in five other families**
- **Four mutations in two proteins with disease-association**



**Yellow: C-loop 1**

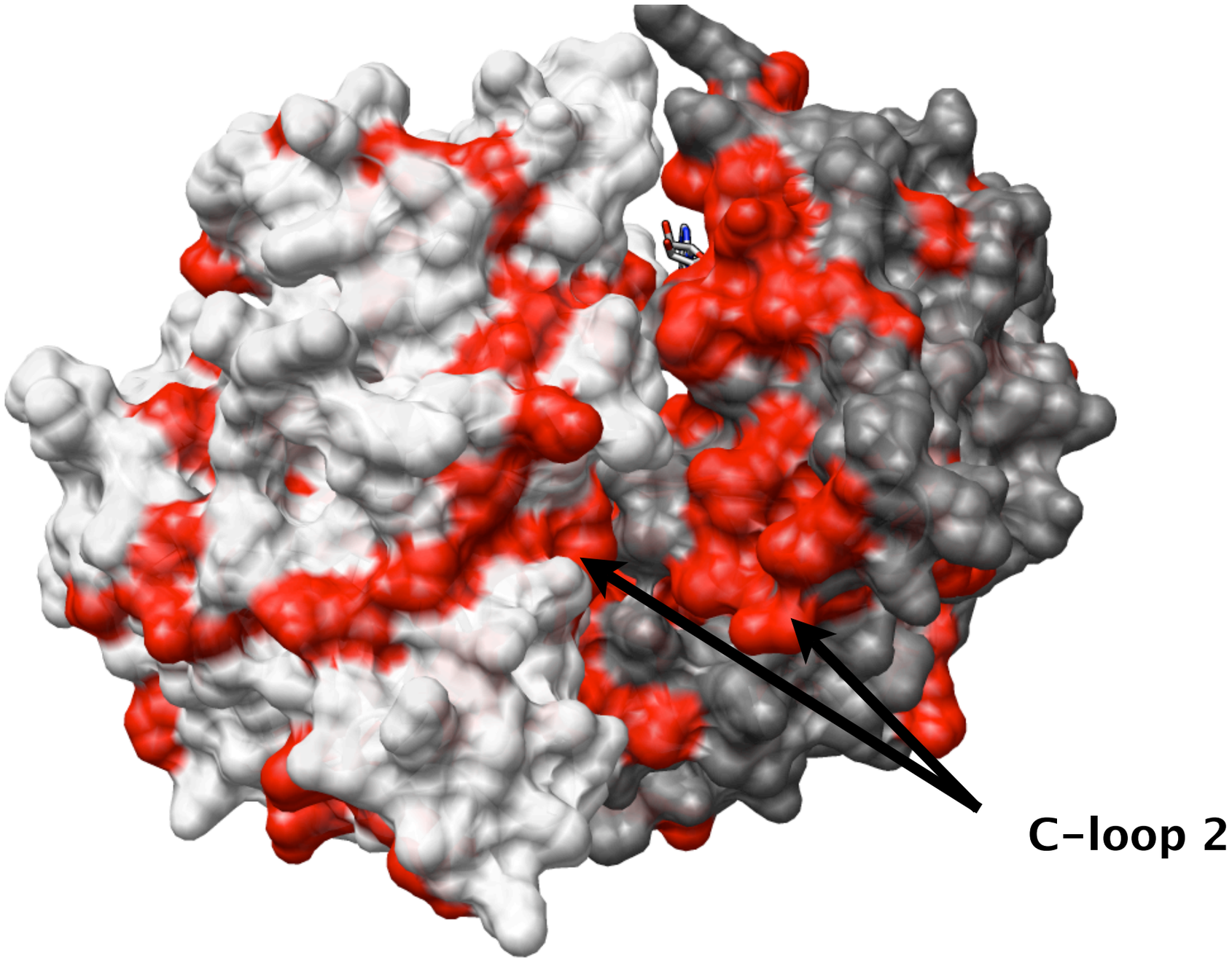
**Light/Dark blue: *S. aureus* 2HYD**

**Orange: Model of MDR1 NBD1**

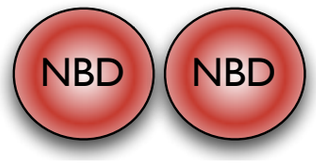
TMD

NBD

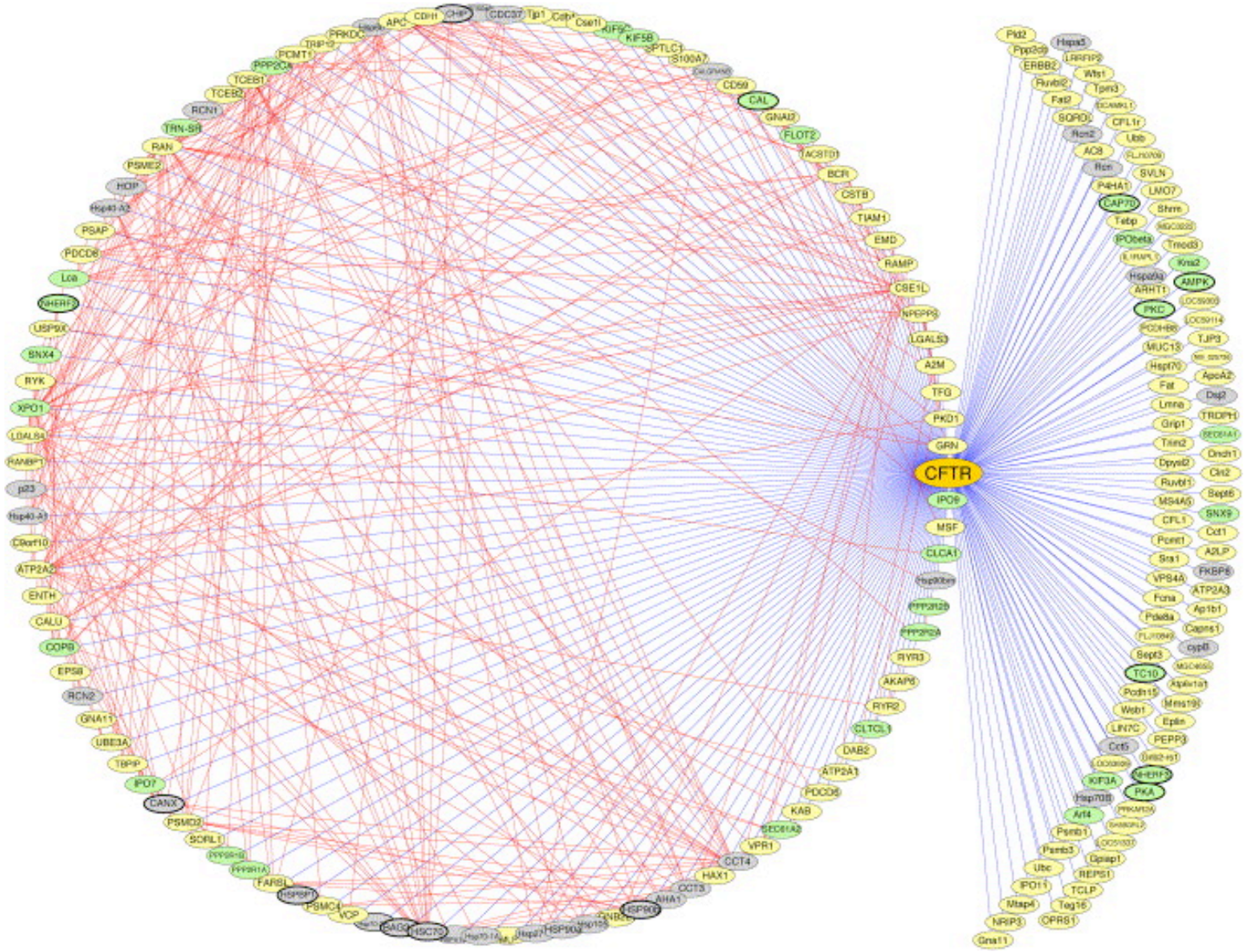
# A disease-associated region at the intracellular NBD surface

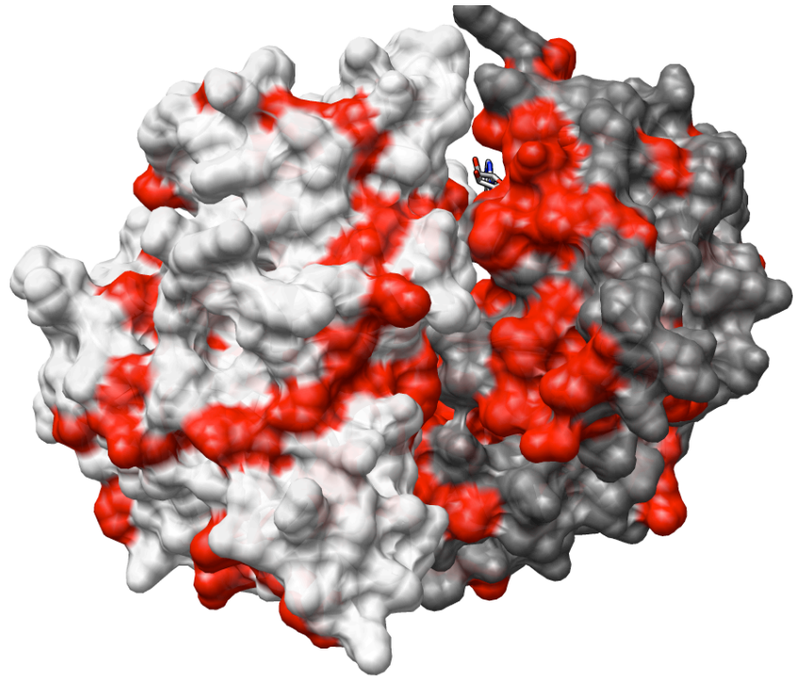



- 11 disease-associated mutations in six different transporters

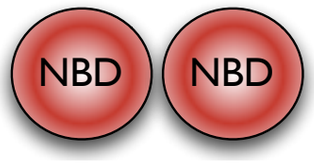


# A disease-associated region at the intracellular surface



- **Intracellular partner interaction surface?**

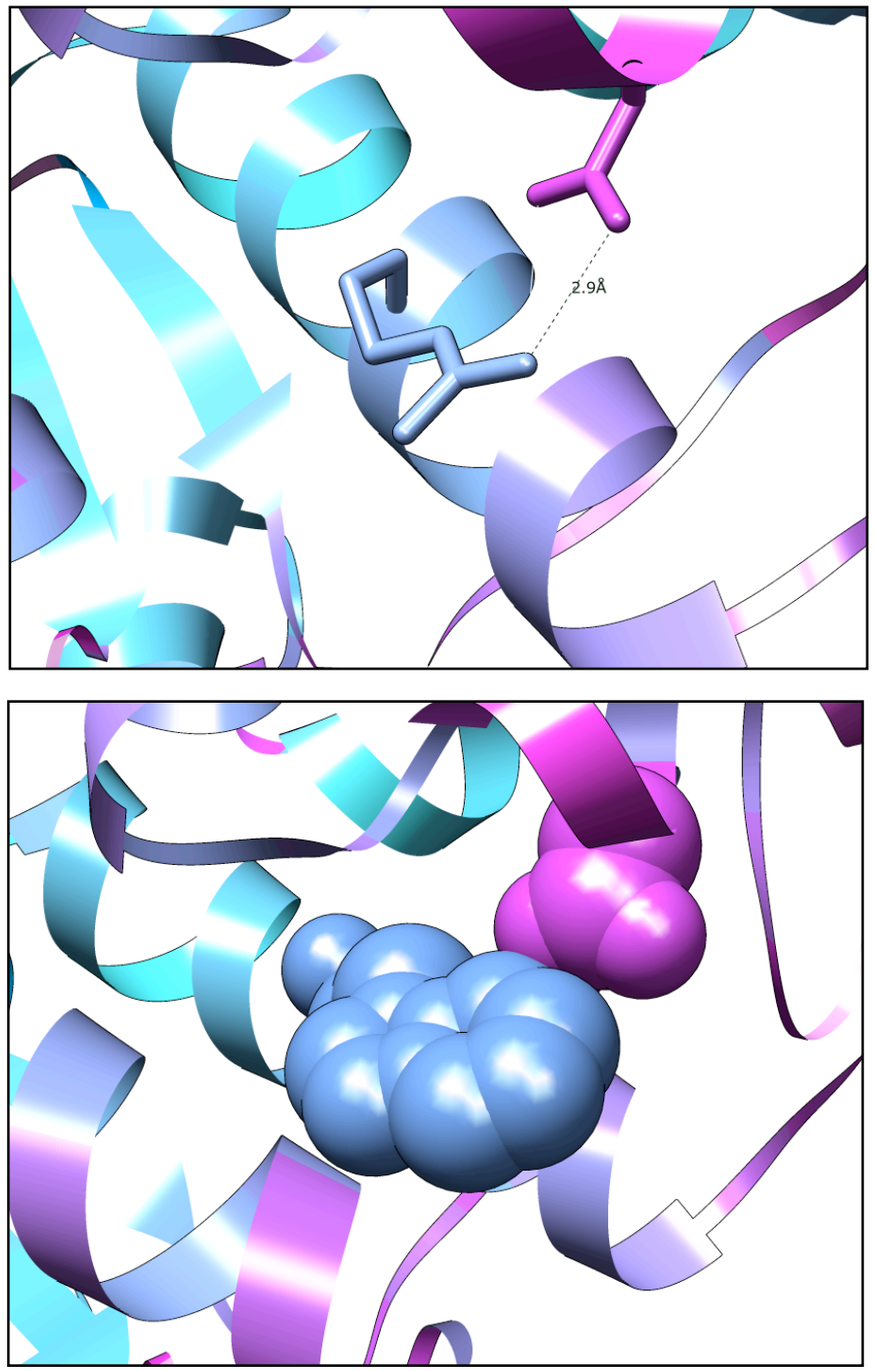




Wang X, et al.
Hsp90 cochaperone Aha1 downregulation rescues misfolding of CFTR in cystic fibrosis.
Cell. 2006 Nov 17;127(4):803-15.

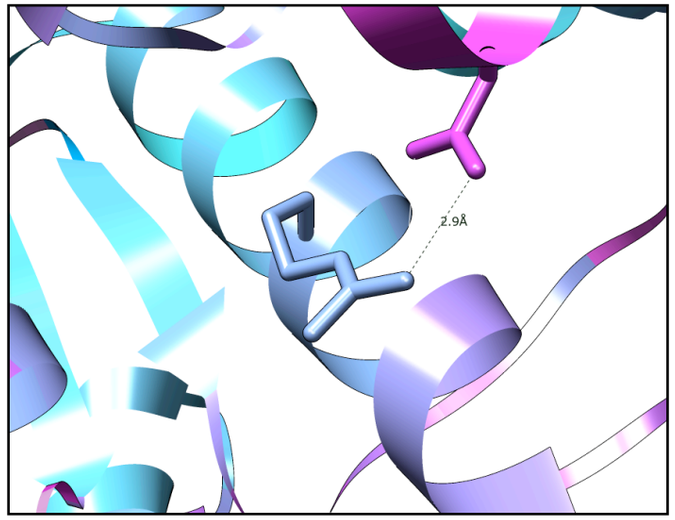
# Representing variation computationally

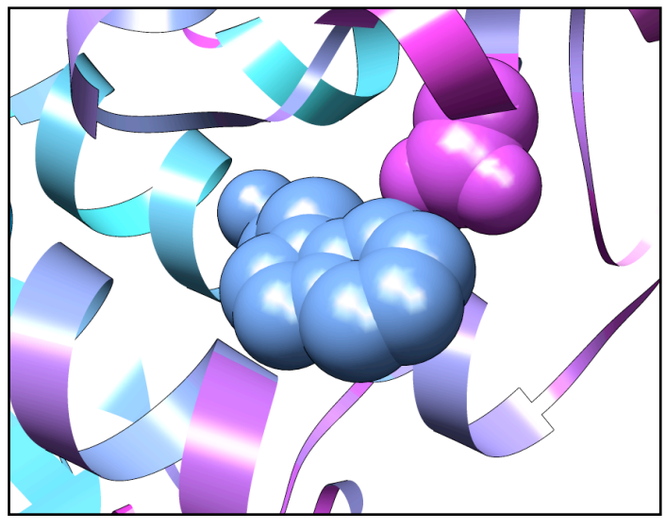
- "Features" of disease-associated mutation R768W in ABCC2
- **Residue**: size change, charge change
- **Evolution**
  - **Sequence**: conserved in an alignment of related sequences
  - **Structure**: buried



# Representing variation computationally

- "Features" of disease-associated mutation R768W in ABCC2
- **Residue**: size change, charge change
- **Evolution**
  - **Sequence**: conserved in an alignment of related sequences
  - **Structure**: buried

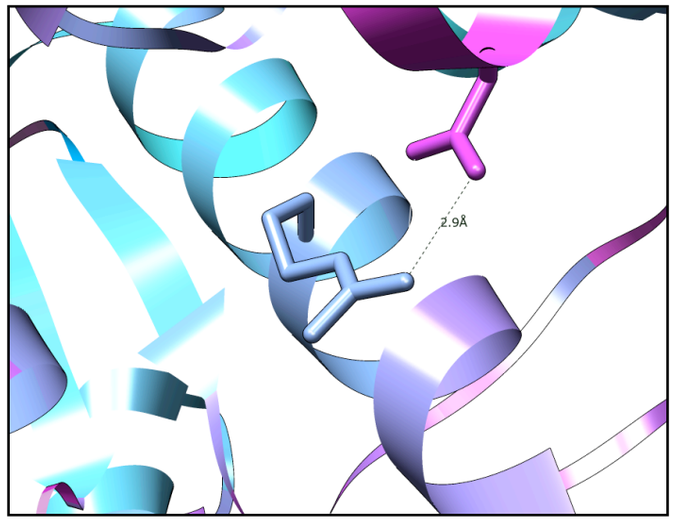


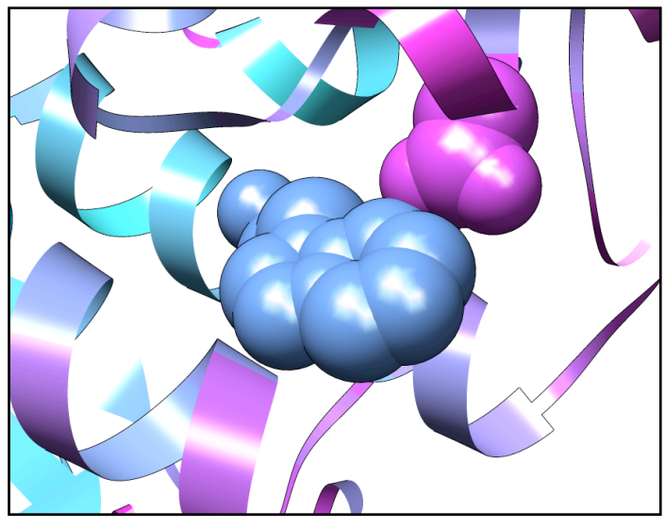



| | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| ABCC2 | R768W | DISEASE | 29 | 0.128 | 30 | 0.12 | 1 | -1.8 | -14.0 | -7.0 | 3.0 | 101 | 0 |

# Representing variation computationally

- "Features" of disease-associated mutation R768W in ABCC2
- **Residue**: size change, charge change
- **Evolution**
  - **Sequence**: conserved in an alignment of related sequences
  - **Structure**: buried

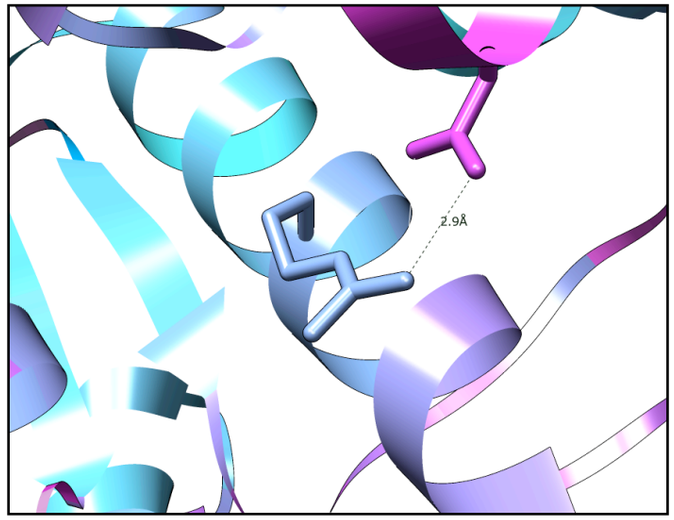


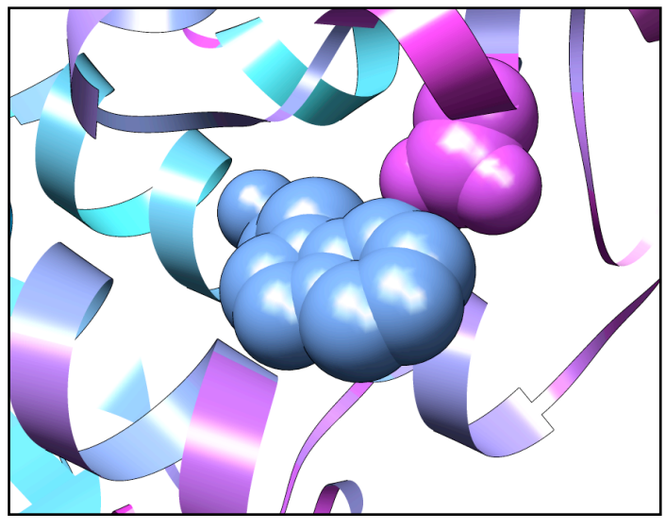



| ABCC2 | R768W | DISEASE | 29 | 0.128 | 30 | 0.12 | 1 | -1.8 | -14.0 | -7.0 | 3.0 | 101 | 0 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| CFTR | A455E | DISEASE | 0 | 0 | 0 | 0 | 1 | -1.7 | 9.8 | -8.3 | 0.66 | 107 | 1 |

# Representing variation computationally

- "Features" of disease-associated mutation R768W in ABCC2
- **Residue**: size change, charge change
- **Evolution**
  - **Sequence**: conserved in an alignment of related sequences
  - **Structure**: buried



| ABCC2 | R768W | DISEASE | 29 | 0.128 | 30 | 0.12 | 1 | -1.8 | -14.0 | -7.0 | 3.0 | 101 | 0 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| CFTR | A455E | DISEASE | 0 | 0 | 0 | 0 | 1 | -1.7 | 9.8 | -8.3 | 0.66 | 107 | 1 |
| MDR1 | S1141T | ? | 57 | 0.4 | 41 | 0.32 | 0 | -0.9 | -0.6 | -7.27 | 0.745 | 58 | 0 |

# Developing a general tool to integrate variant data

- Given a set of features that represents a point mutant

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| MDR1 | S1141T | 57 | 0.4 | 41 | 0.32 | 0 | -0.9 | -0.6 | -7.3 | 0.75 | 58 | 0 |

- Return a binary prediction of effect

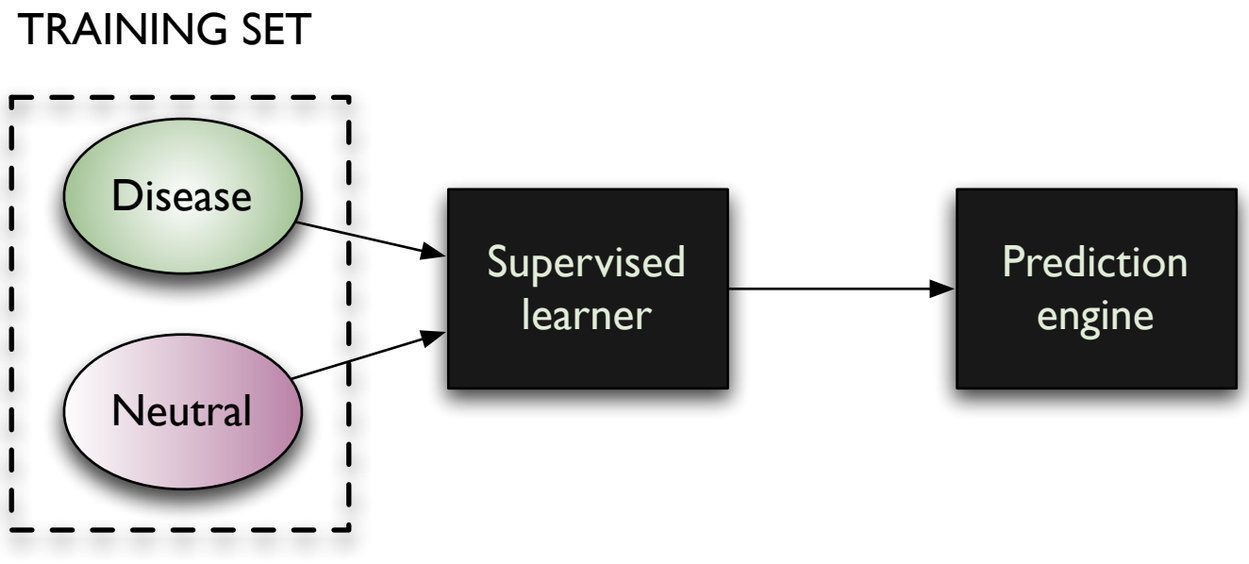
| | | |
|---|---|---|
| MDR1 | S1141T | NEUTRAL |

- Use Random Forests (RF), a supervised learning algorithm, to combine the features for prediction

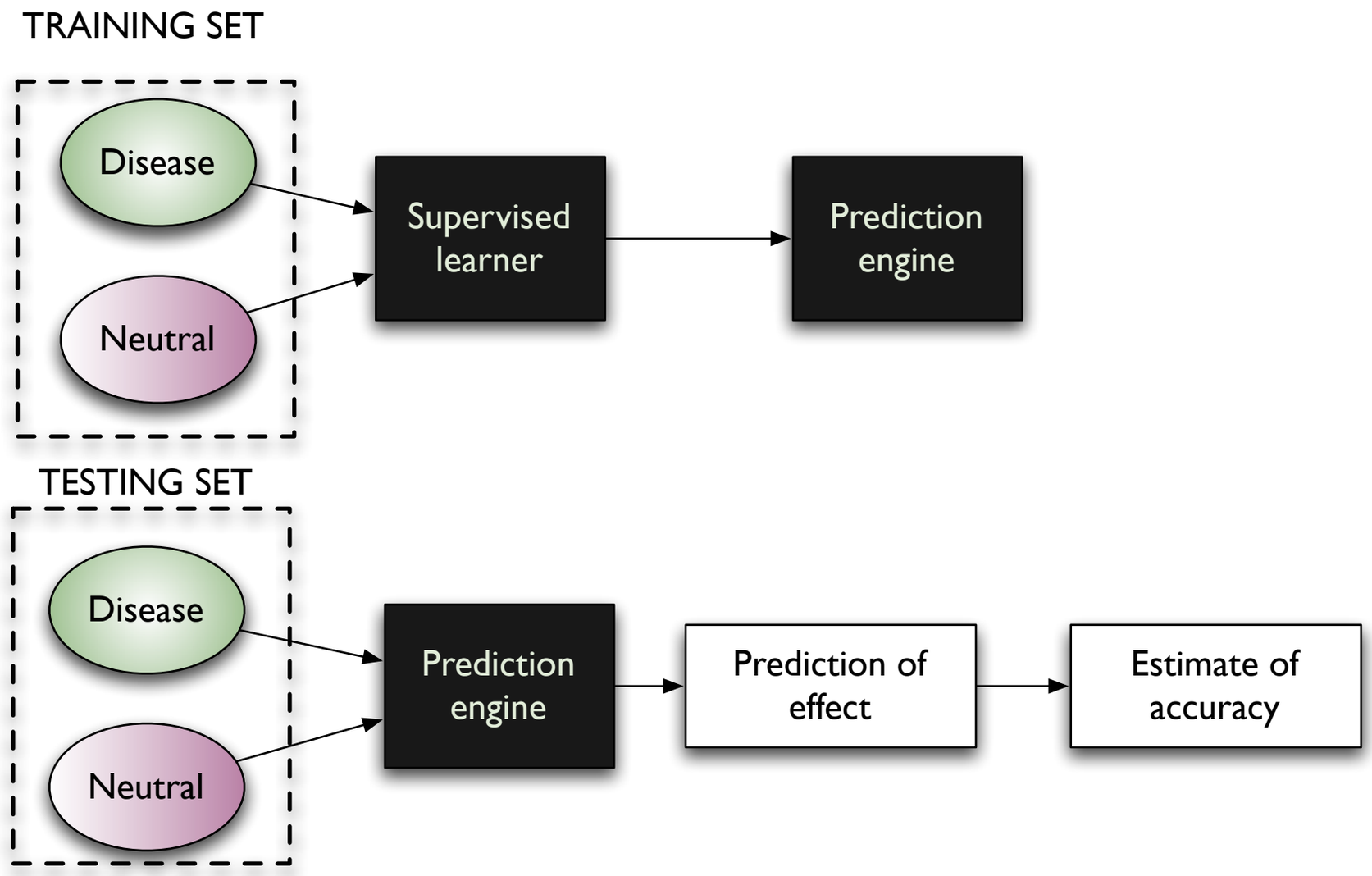

Karchin R, Diekhans M, Kelly L, Thomas DJ, Pieper U, Eswar N, Haussler D, Sali A.
LS-SNP: large-scale annotation of coding non-synonymous SNPs based on multiple information sources.
Bioinformatics. 2005 Jun 15;21(12):2814-20. Epub 2005 Apr 12.

Karchin R, Kelly L, Sali A.
Improving functional annotation of non-synonomous SNPs with information theory.
Pac Symp Biocomput. 2005;:397-408.

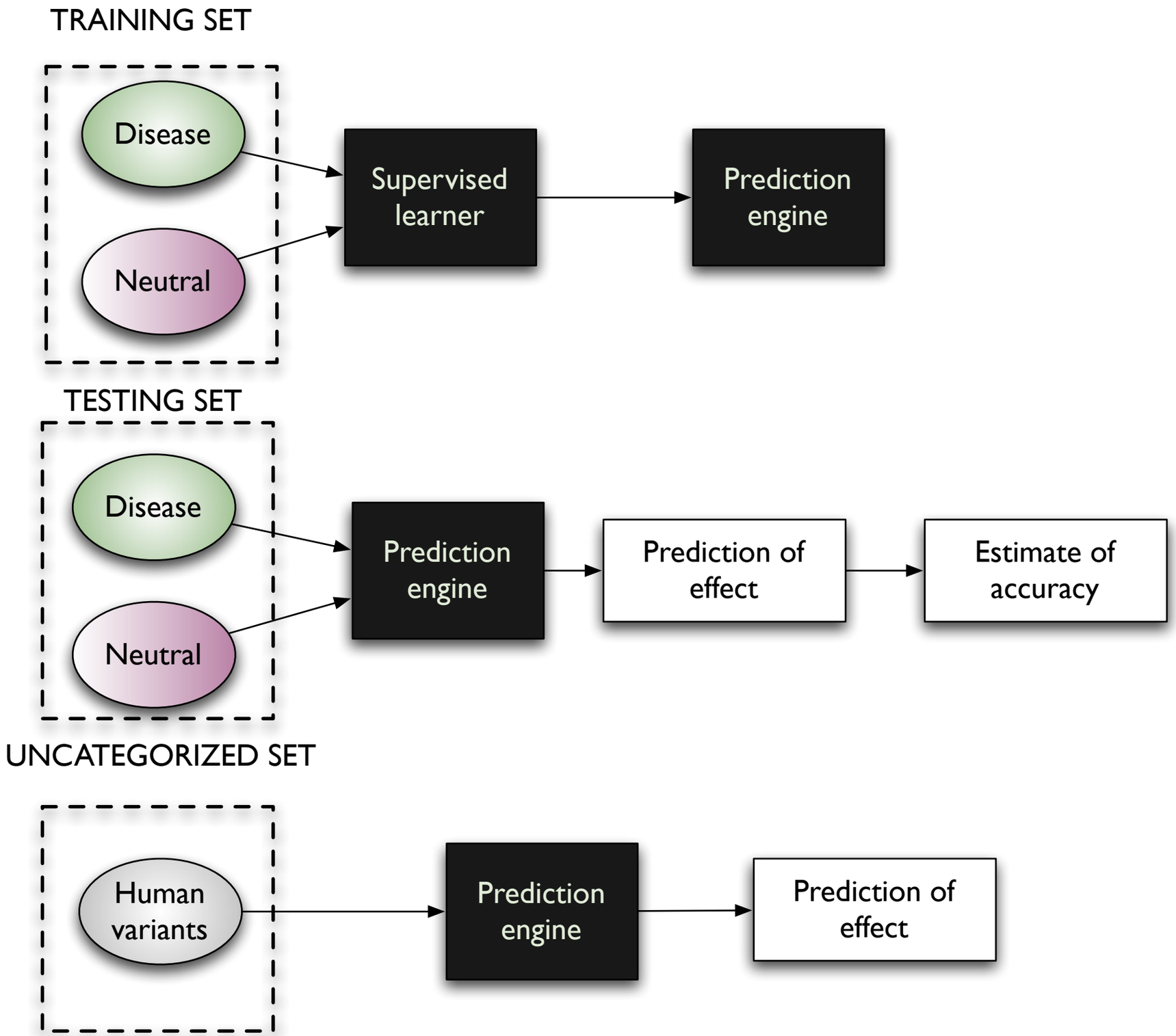
# A supervised learner “learns” classes of data



# A supervised learner “learns” classes of data

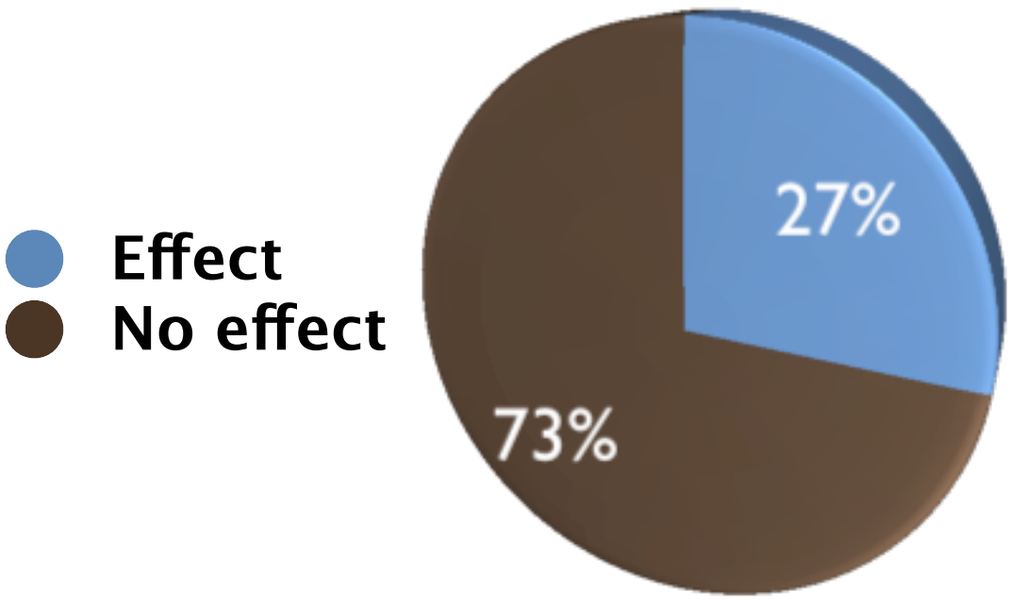


# A supervised learner “learns” classes of data



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

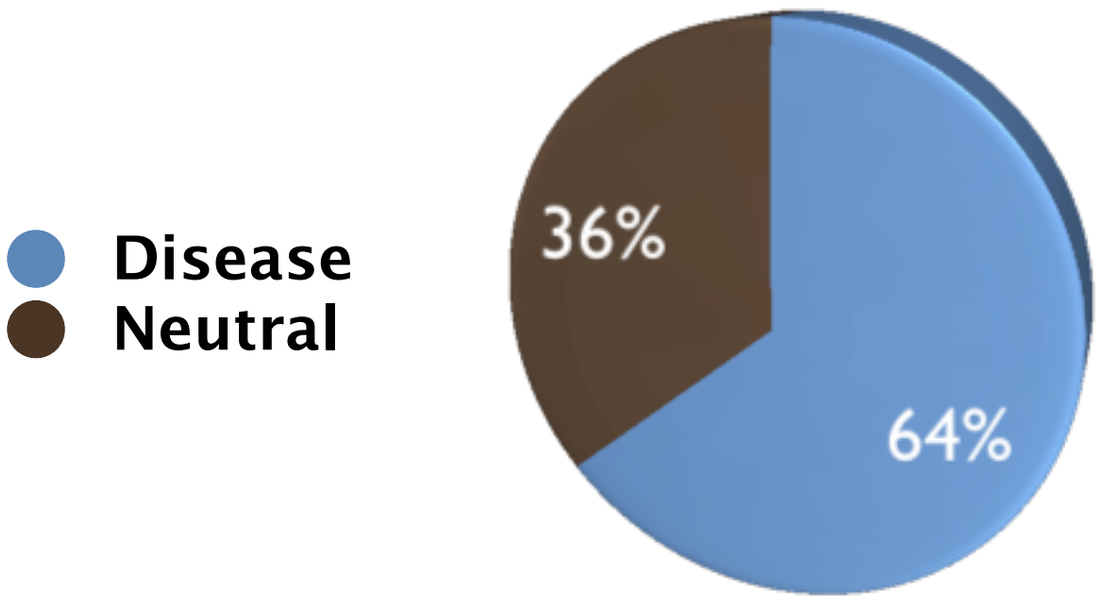
## ‘Experimental’ training set




**Two comprehensively mutated globular proteins**

**4975 mutations total**

## ‘Clinical’ training set




**1. All modelable mutations in human proteins**

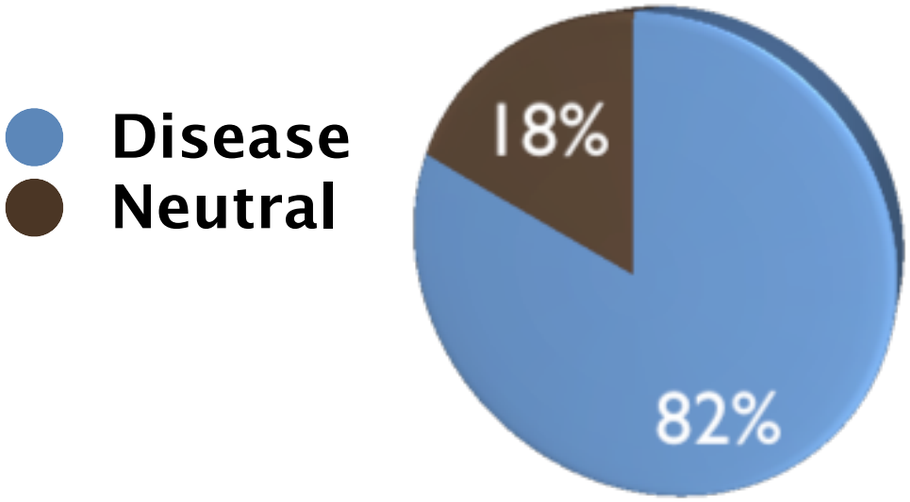
**8396 mutations total**

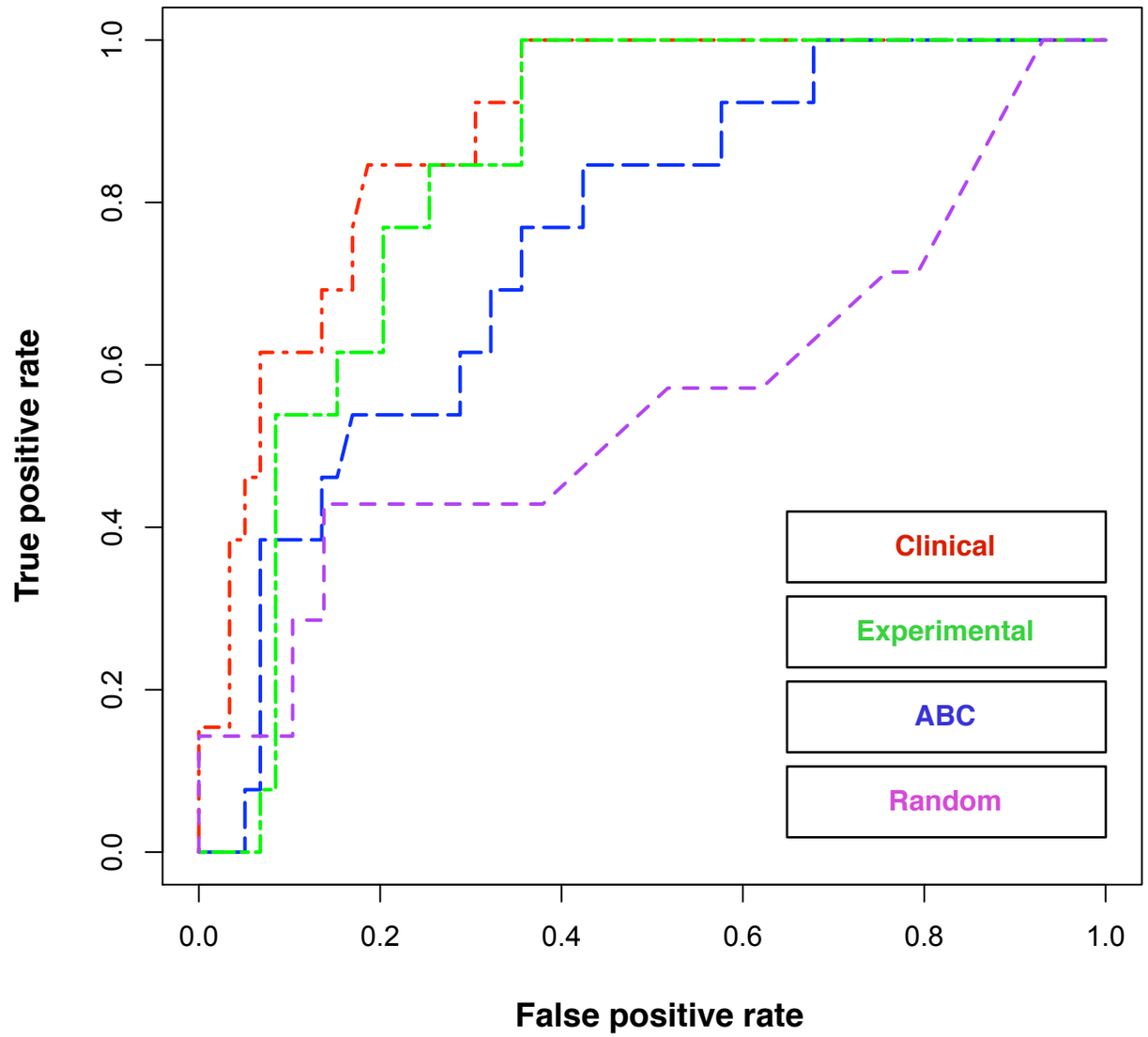
**2. Subset of ABC transporter mutations**

**275 mutations total**

# Test set: 72 mutations from CFTR

- **The classifier trained on clinical data is the best performer**
- **Accuracy: 85%**
- **The ABC-trained classifier suffered from a lack of neutral examples**



# Experimental validation of predictions

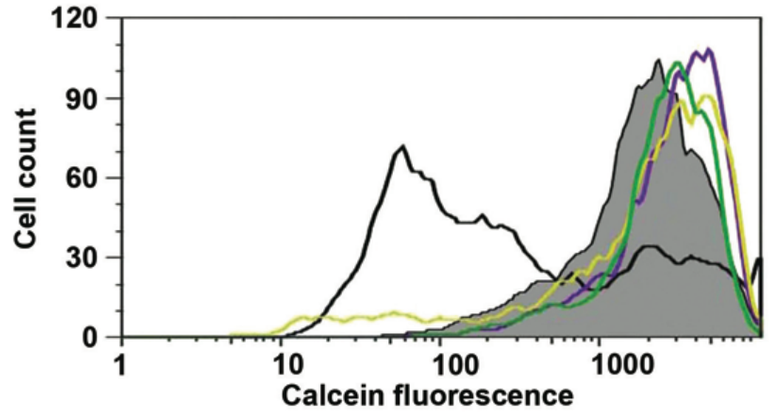
| GENE | HUGO | VARIANT | DOMAIN | PREDICTION |
|---|---|---|---|---|
| MDR1 | ABCB1 | S1141T | NBD2 | NEUTRAL |
| MDR1 | ABCB1 | V1251I | NBD2 | NEUTRAL |
| MDR1 | ABCB1 | W1108R | NBD2 | DISEASE |
| MRP4 | ABCC4 | G487E | NBD1 | DISEASE |
| MRP4 | ABCC4 | K498E | NBD1 | NEUTRAL |
| MRP4 | ABCC4 | V1071I | NBD2 | NEUTRAL |

- **Predicted the effects of 36 point mutants in seven human ABC transporters from three families**
- **Functional assays for two transporters, MDR1 and MRP4**
- **Experimental validation of six predictions**

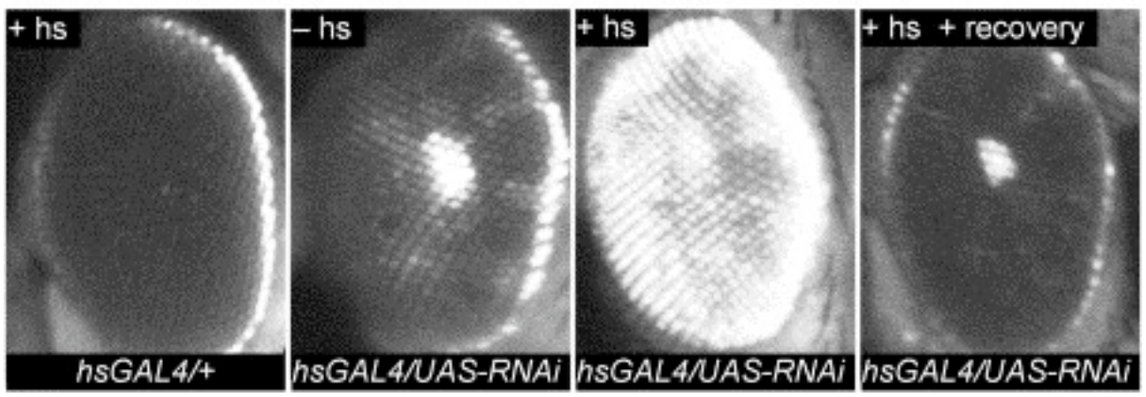
# Experimental functional analysis of ABC transporters

- **Yeast** – **transport assays, cytotoxicity**


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



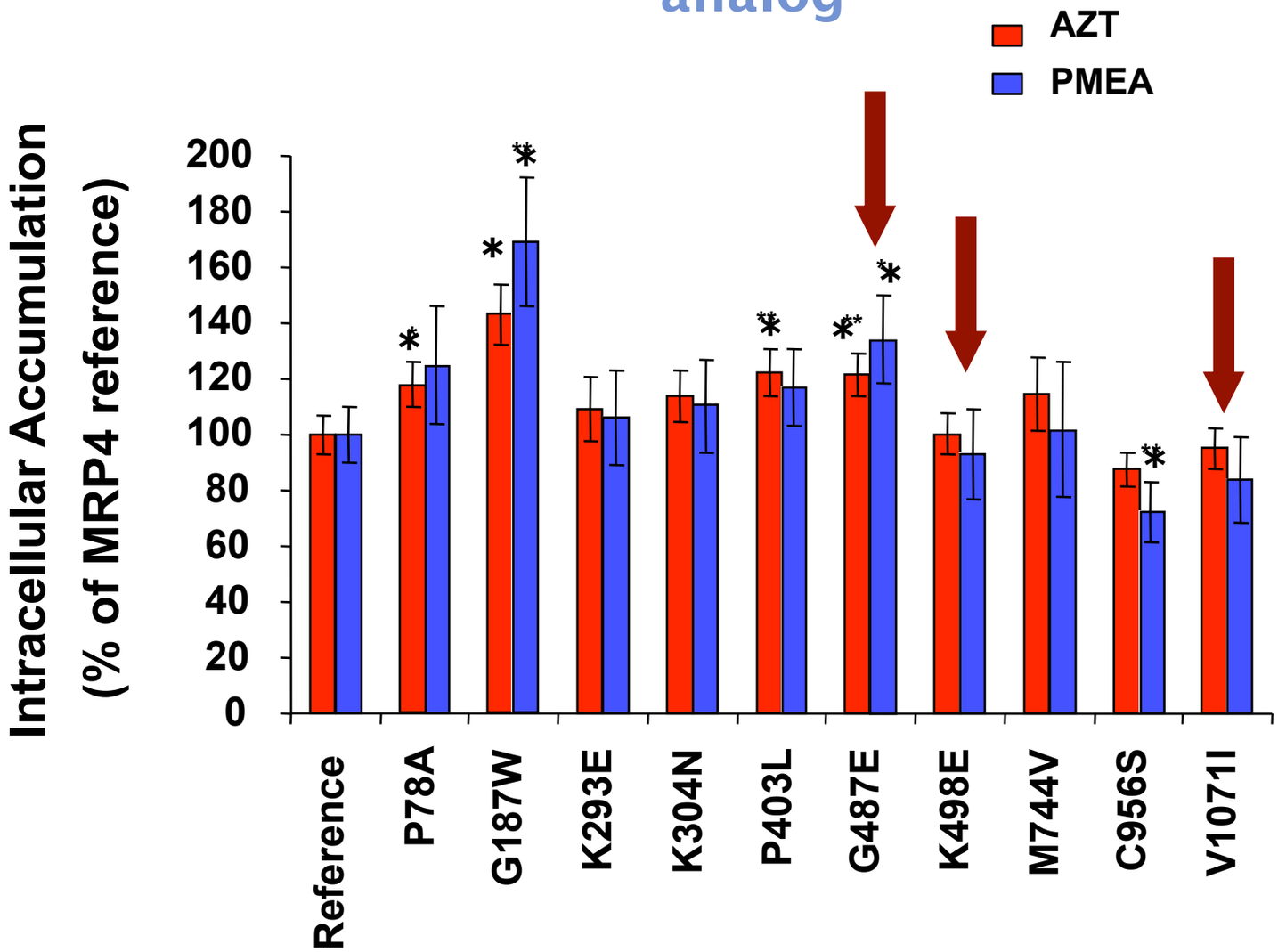

- **Drosophila** – **live visualization of transport across membranes**



# Validation of MRP4 predictions

| Variant | Prediction | |
|---|---|---|
| G487E | Disease | ✔ |
| K498E | Neutral | ✔ |
| V1071I | Neutral | ✔ |

Transfected HEK cells, radiolabeled nucleoside/nucleotide analog

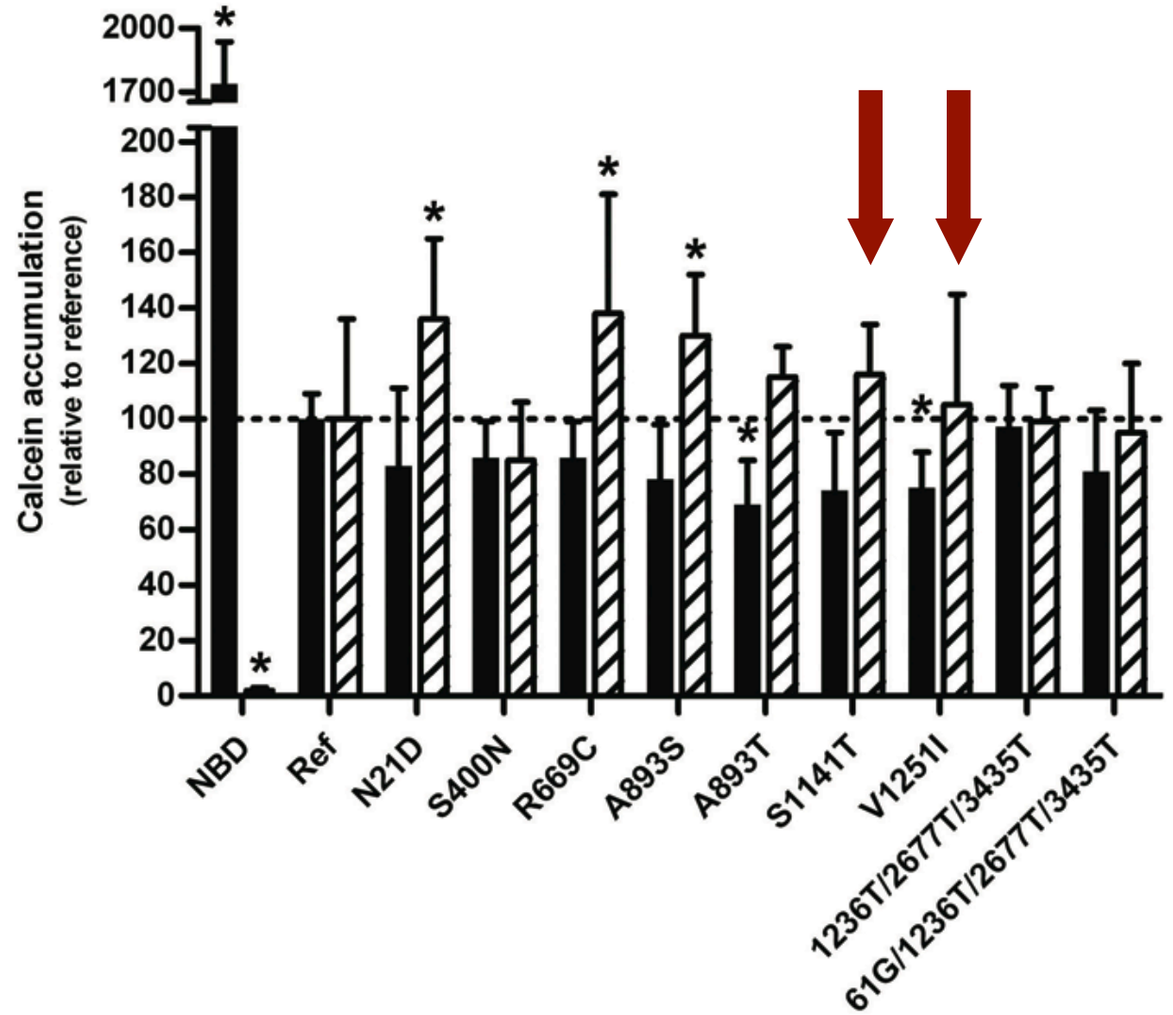



Nada Abla and Deanna Kroetz

# Validation of MDR1 predictions

| Variant | Prediction | |
|---|---|---|
| W1108R | Disease | ✔ |
| S1141T | Neutral | ✔ |
| V1251I | Neutral | ✔ |


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

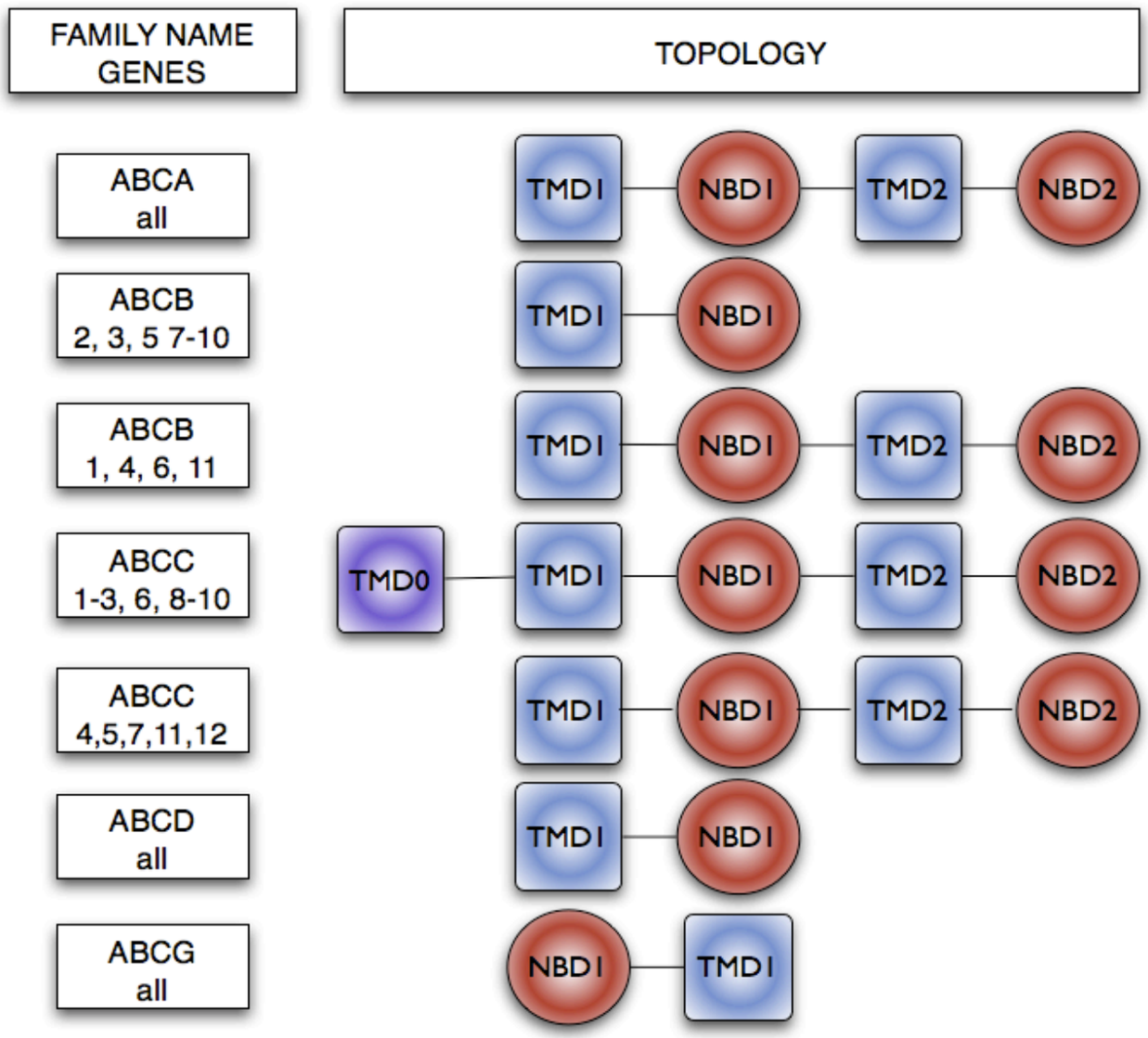



Ho Jeong, Jason Gow and Deanna Kroetz

# Characterizing genetic variation in human transport proteins

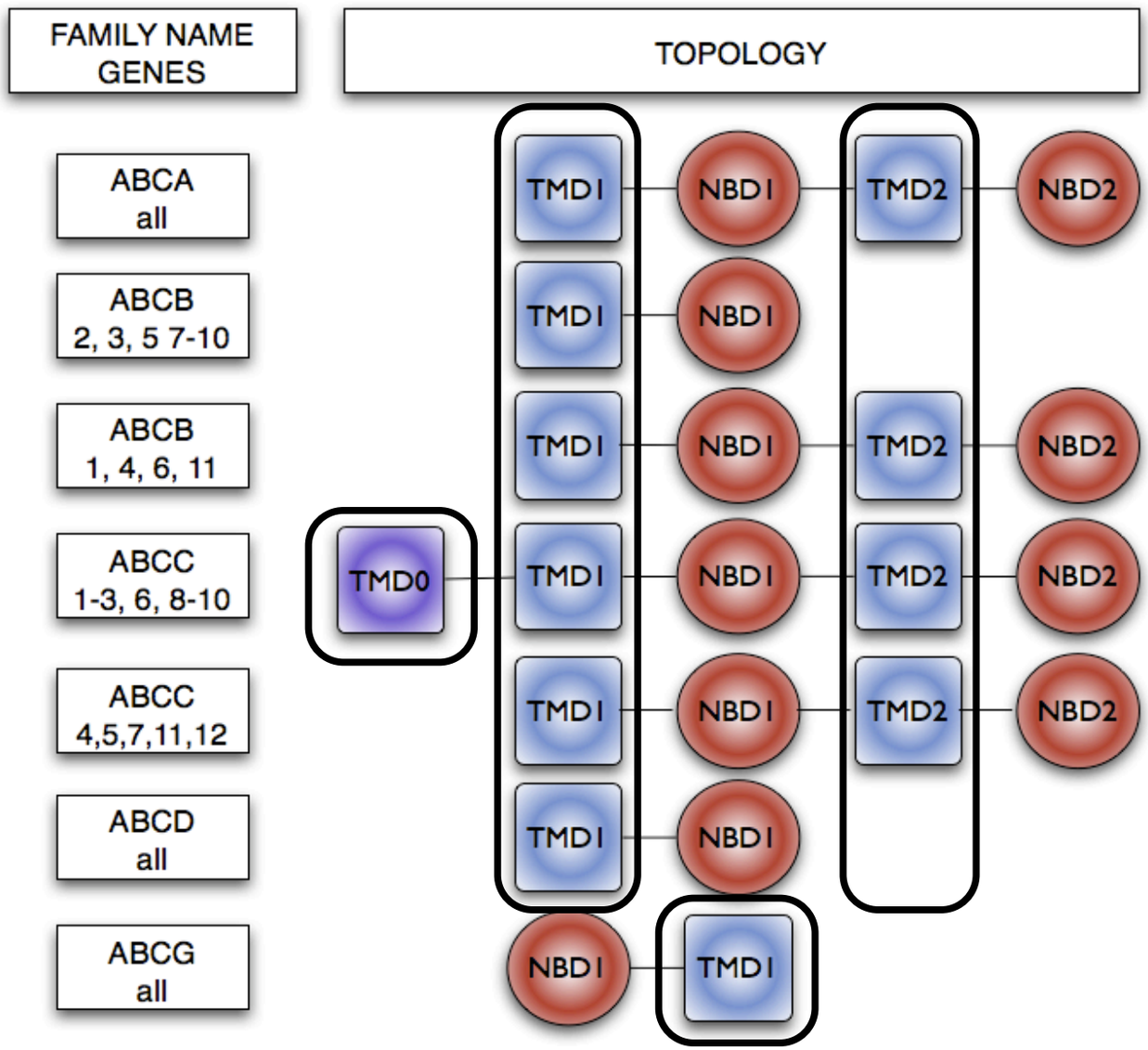
- Comparative modeling of all human ABC transporter NBDs and 300+ point mutants
- Located 68 disease-associated mutations at putative interfaces in 10 human ABC transporters
- Developed a general tool for predicting the impact of point mutations on protein function
    - Correctly predicted the *in vitro*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

# 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

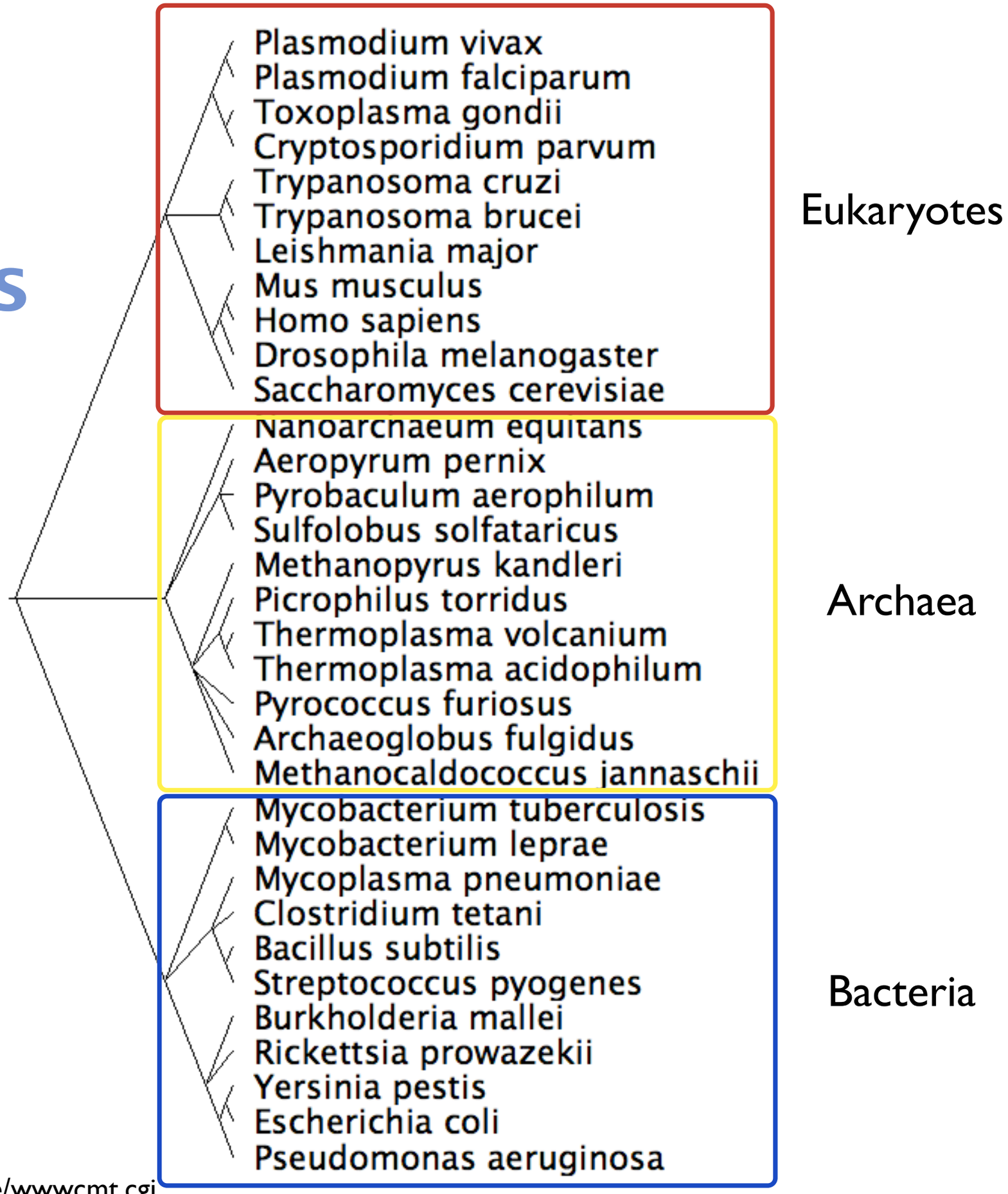



- We excise all TMDs and create sequence profiles for each
- Each profile is scanned against a large database of membrane protein profiles

# Extending the analysis to whole genomes

**We selected:**

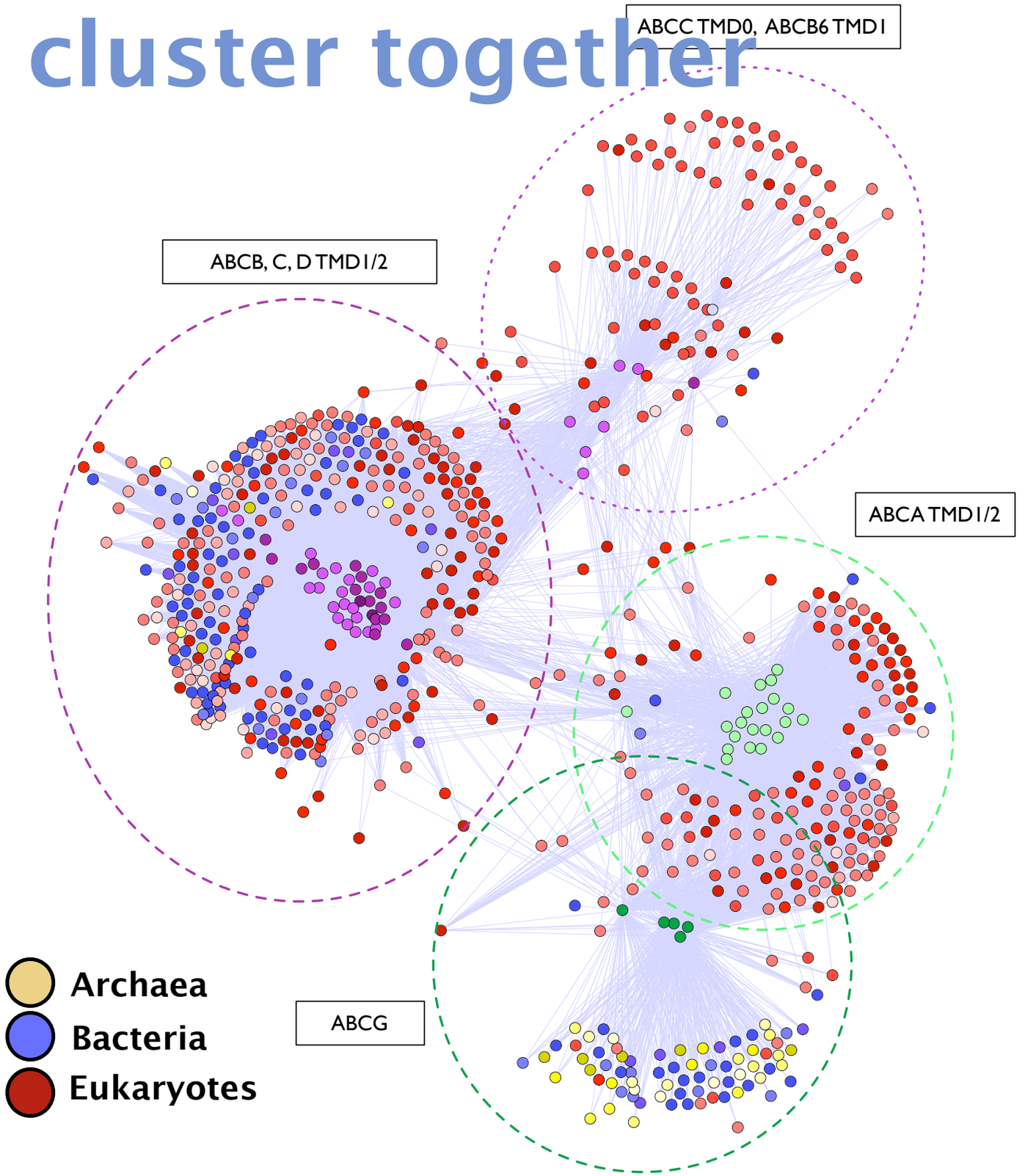
- **model species**
- **pathogenic species**
- **complete genomes**
- **genomic DNA available**




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

# 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

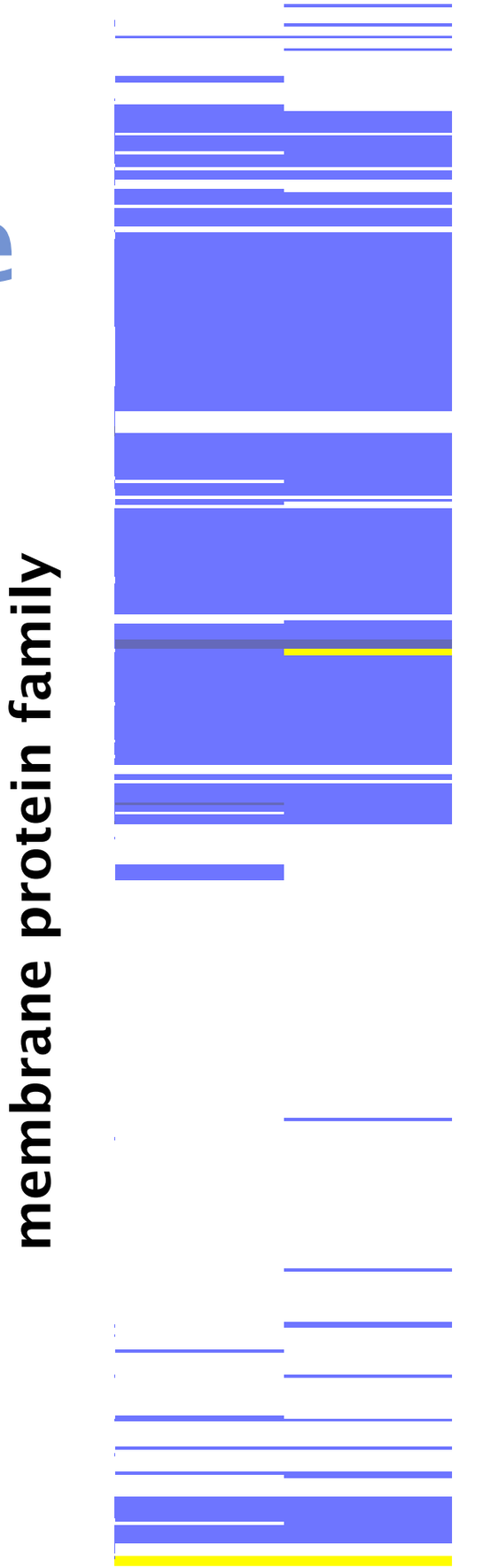


# Identifying the membrane proteome of organisms

- 598 membrane protein families in Pfam
- How many times does each appear in a given organism?

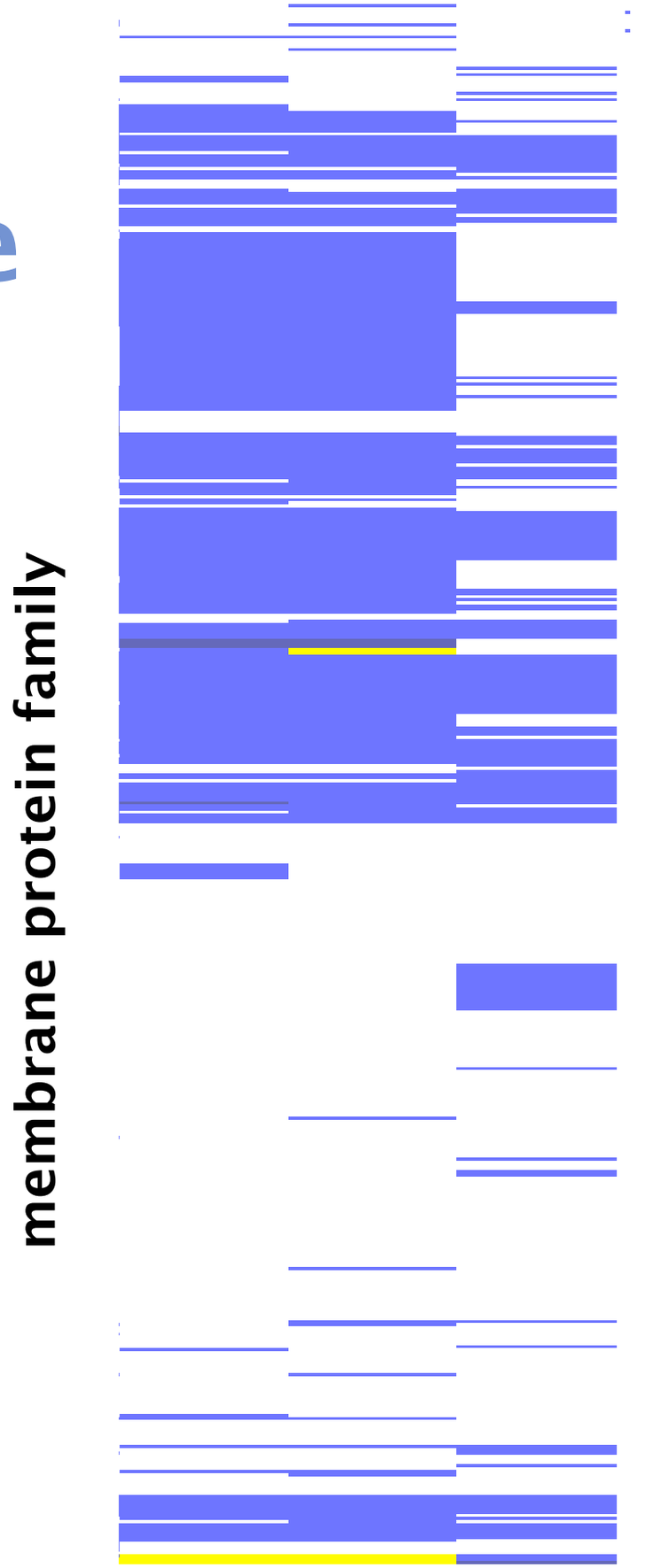
# Identifying the membrane proteome of organisms

- 598 membrane protein families in Pfam
- How many times does each appear in a given organism?



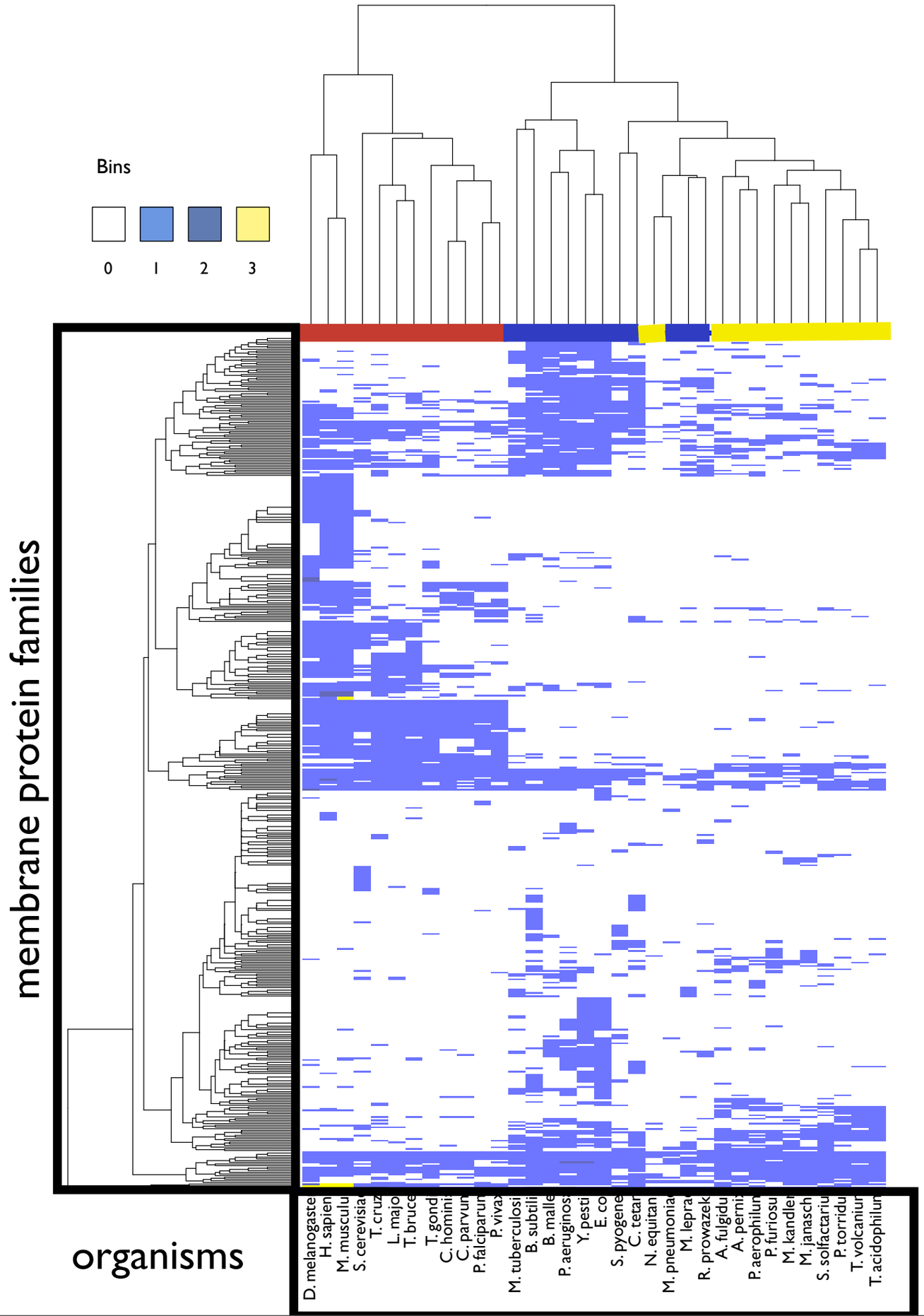
# Identifying the membrane proteome of organisms

- 598 membrane protein families in Pfam
- How many times does each appear in a given organism?



# Family content reveals a clear split between prokaryotes and eukaryotes

- Clusters of families that tend to travel together
- Clusters of families that appear in specific organisms



# 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**

# Thanks!

- Questions?

# Awesome thesis committee

Andrej Sali

Bob Stroud


Deanna Kroetz

# Awesome thesis committee

Andrej Sali

More accurate

Bob Stroud

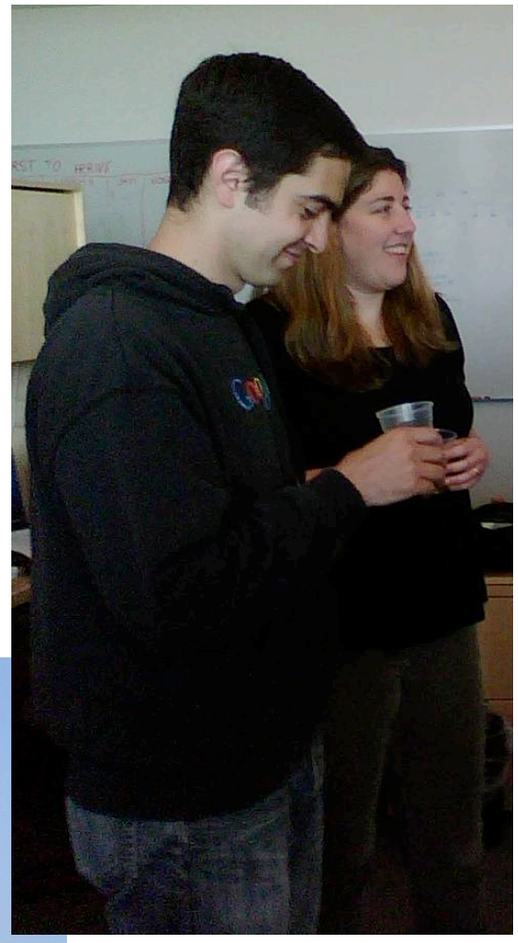
Non-silly picture

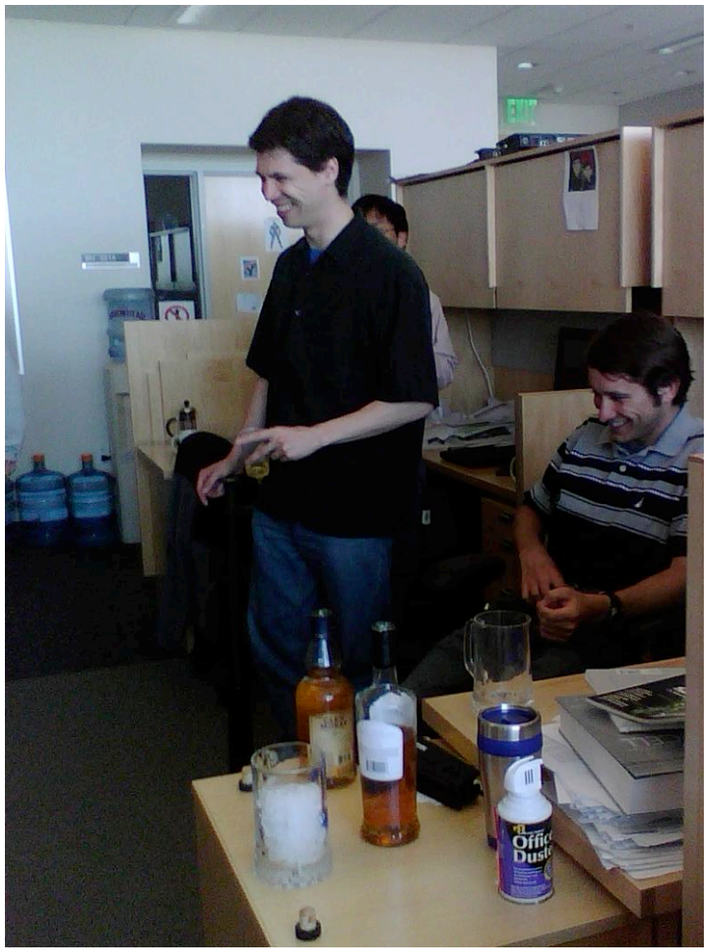
Sweet photo montage

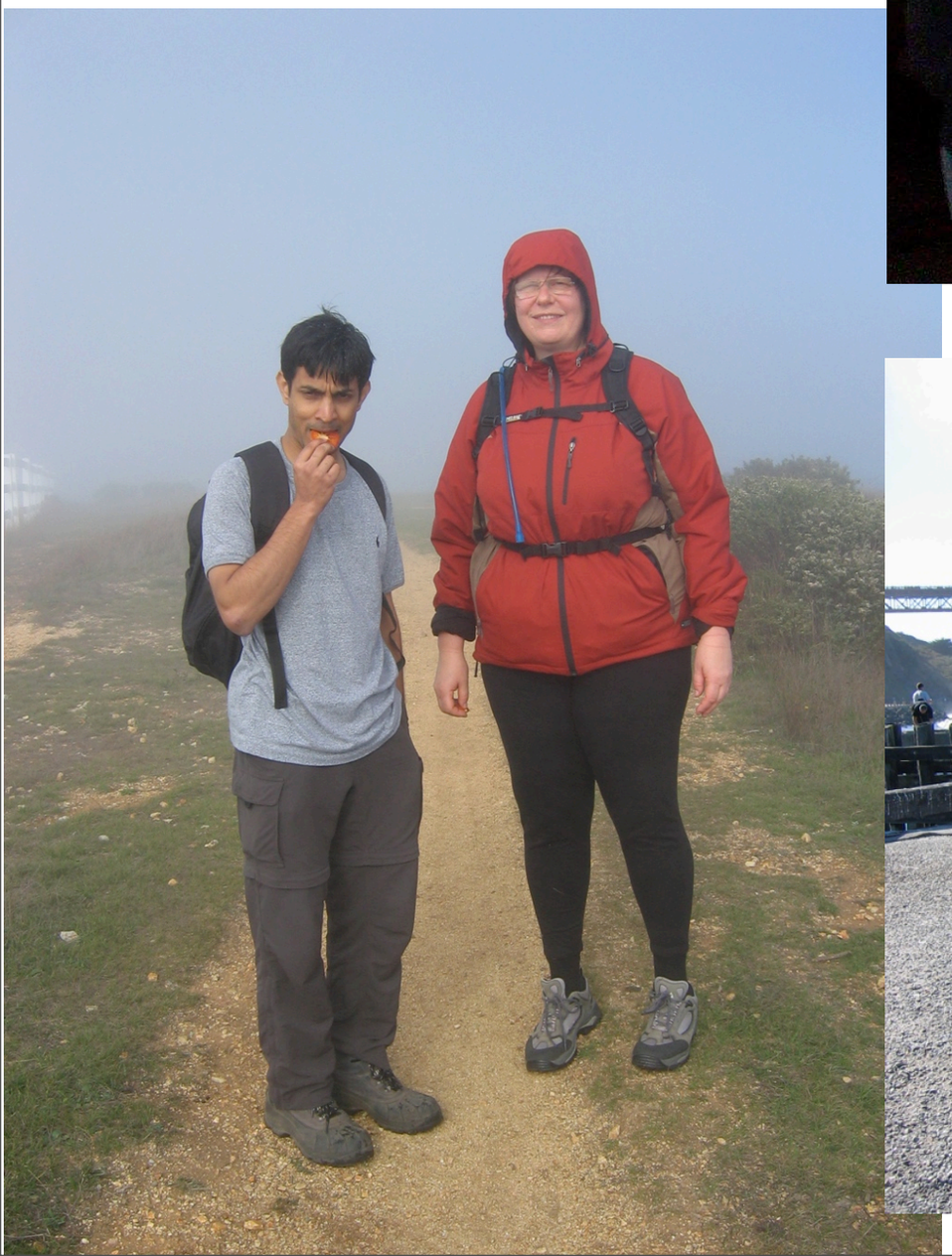

Deanna Kroetz

# We're not usually this organized


Coca-Cola







# Deanna Kroetz lab

Leslie
Jason
Hisa

# UCSF and BMI!



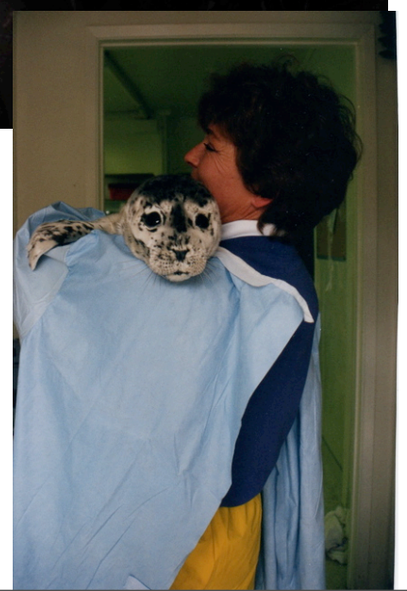



Rebecca, Julia, Patsy and Tom


The girls. :)

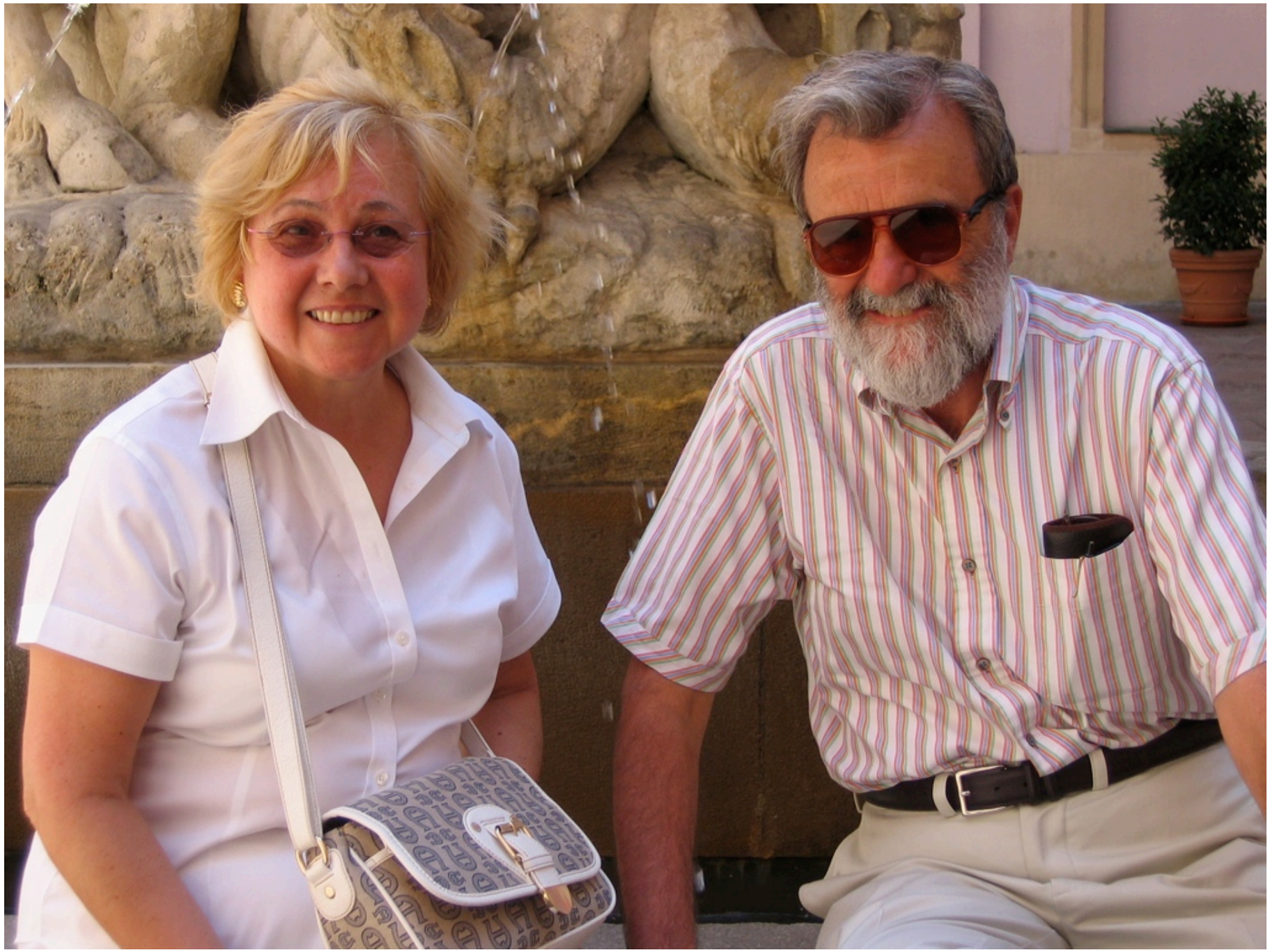
# Thursday night crew





# Marty, Nan, Sylvia, Fred, Sue, Jeff and Mie


# Mama and Papa



- 40 more years!

# Partner in crime