

# MitoAtlas: a Domain-Resolved Spatial Map of the Human Mitochondrial Proteome

Jiwoong Kang<sup>1,9</sup>, Sanghee Shin<sup>2,3,6,9</sup>, Chulhwan Kwak<sup>1,7</sup>, Song-Yi Lee<sup>1,8</sup>, Minkyong Jung<sup>4</sup>, Hanseul Lee<sup>5</sup>, Myeong-Gyun Kang<sup>1</sup>, Jiho Sim<sup>1</sup>, Seung-Jae V. Lee<sup>5</sup>, Ji Young Mun<sup>4</sup>, Jong-Seo Kim<sup>2,3</sup>, Hyun-Woo Rhee<sup>1,2,10</sup>

<sup>1</sup>Department of Chemistry, Seoul National University, Seoul 08826, Republic of Korea

<sup>2</sup>School of Biological Sciences, Seoul National University, Seoul 08826, Republic of Korea

<sup>3</sup>Center for RNA Research, Institute of Basic Science (IBS), Seoul 08826, Republic of Korea

<sup>4</sup>Neural Circuit Research Group, Korea Brain Research Institute, Daegu 41062, Republic of Korea

<sup>5</sup>Department of Biological Sciences, Korea Advanced Institute of Science and Technology (KAIST), Daejeon 34141, Republic of Korea

<sup>6</sup>Present address: Department of Cancer Biology, Dana-Farber Cancer Institute, Boston 02215, MA, USA;  
Department of Cell Biology, Harvard Medical School, Boston, MA 02215, USA

<sup>7</sup>Present address: Department of Neurosurgery, Stanford University, Stanford, CA 94305, USA

<sup>8</sup>Present address: Department of New Biology, DGIST, Daegu 42988, Republic of Korea

<sup>9</sup>These authors contributed equally.

<sup>10</sup>Lead contact

\*Correspondence: rhee.hw@snu.ac.kr (H.-W.R.), jongseokim@snu.ac.kr (J.-S.K.)

## SUMMARY

Mitochondrial function depends on the precise sub-organellar localization and topological orientation of its proteome. However, mapping this architecture with domain-level precision across the double-membrane system has remained challenging. Here, we present MitoAtlas, a spatial systems biology resource built on SR-PL (super-resolution proximity labeling) that resolves the sub-mitochondrial architecture of 861 proteins at domain-level resolution. By deploying 42 spatially restricted APEX2 and BioID/BioID2 baits across mitochondrial compartments and cytosol in HEK293T-REx cells, we mapped 13,440 unique proximity-labeled sites. Integrating high-density spatial data with a machine-learning classification pipeline and AlphaFold structural modeling, we resolved membrane topologies for 196 transmembrane proteins and assigned sub-mitochondrial localizations to 160 previously unannotated mitochondrial proteins absent from MitoCarta3.0, including MINPP1 as an intermembrane space resident. Coupling bait-specific proximity labeling signatures with AlphaFold structure prediction and metabolite docking identified 47 protein–protein interactions, including the LETM1–Citrate Synthase complex, and 155 protein–metabolite interactions. MitoAtlas is available at [mitoatlas.org](http://mitoatlas.org) as a domain-resolution spatial map of mitochondrial architecture and metabolism.

**KEYWORDS:** Mitochondrial Proteome, Proximity Labeling, Sub-mitochondrial Localization, AlphaFold2-Multimer, Protein-Protein Interactions, Protein-Metabolite Interactions

## INTRODUCTION

The mitochondrion is defined by its sophisticated double-membrane architecture, which partitions the organelle into distinct physiological environments: the outer mitochondrial membrane (OMM), the intermembrane space (IMS), the inner mitochondrial membrane (IMM), and the matrix. This compartmentalization is not merely structural; it is the foundation for specialized biological processes. For example, LYR domains facilitate respiratory complex assembly in the matrix, while Bcl-2 homology (BH) domains at the OMM mediate apoptotic signaling.<sup>1,3,4,8</sup> Consequently, mapping the mitochondrial proteome with domain-level precision is essential for deciphering the spatial logic of mitochondrial metabolism and regulation.

Despite extensive efforts to catalog mitochondrial proteins, our current spatial understanding remains incomplete.<sup>11</sup> MitoCarta3.0, which catalogs over 1,100 mitochondrial proteins, still assigns 90 of them to ambiguous “membrane” or “unknown” categories without sub-compartment resolution. Furthermore, over 400 proteins lack precise topological classifications, leaving their functional orientation across the IMM or OMM unresolved. Traditional biochemical methods for topology mapping often struggle with high-throughput scaling and are susceptible to methodological artifacts, leading to long-standing topological errors for critical proteins such as RTN4IP1<sup>9</sup> and EXD2.<sup>5,6</sup> Similarly, the topology of LETM1 remains a subject of active debate, complicating our understanding of its role in ion transport and mitochondrial morphology.<sup>7</sup> Such errors fundamentally hinder the modeling of protein-protein interaction networks and the spatial distribution of functional domains across the double-membrane system. Addressing these challenges requires a spatial systems biology approach: one that integrates proximity labeling data from multiple spatially defined baits within a unified computational framework, rather than relying on binary comparisons between individual compartments.

Proximity labeling (PL) coupled with mass spectrometry has emerged as a powerful tool for mapping sub-mitochondrial architectures in live cells.<sup>2-16</sup> However, conventional PL workflows often suffer from high false-positive rates due to the detection of non-specific biotinylation or background contaminants. To overcome these limitations, we previously developed super-resolution proximity labeling (SR-PL).<sup>10,17,18</sup> By focusing exclusively on biotinylated peptides and their specific modification sites, SR-PL provides high spatial accuracy. While our initial efforts with a small number of APEX2 baits validated the precision of this approach, a comprehensive, organelle-wide map required significantly greater proteome coverage. Recent advances in AI-driven protein structure prediction, particularly AlphaFold, now offer the potential to cross-validate and refine such spatial assignments at domain-level resolution by integrating high-confidence

structural models with proximity labeling data.

Here, we describe an expanded SR-PL workflow integrating both APEX2 and BioID enzymes across 42 mitochondrial baits and characterize the sequence selectivity and steric constraints underlying each labeling chemistry. A two-stage classification pipeline assigns sub-mitochondrial localization and membrane topology from multi-bait intensity profiles. This framework reveals a substantial cohort of mitochondrial proteins absent from existing databases, whose identities we validate through imaging, biochemistry, and functional assays. Systematic topology mapping across both mitochondrial membranes uncovers unexpected structural features, including a novel  $\beta$ -barrel architecture. Integration of site-resolved labeling data with AlphaFold-Multimer and metabolite docking further enables prediction of protein–protein and protein–metabolite interactions. All spatial annotations, topology maps, and structural predictions are accessible through MitoAtlas.org.

Altogether, MitoAtlas constitutes the most detailed sub-mitochondrial proteome map to date. Unlike existing databases that classify proteins into broad compartmental categories, it provides domain-level spatial resolution through systems-level integration of multi-bait proximity labeling, enabling membrane topology determination and functional domain mapping inaccessible to fractionation-based approaches. Across 861 proteins and 13,440 uniquely labeled sites, we assigned precise sub-mitochondrial localizations to 160 previously unannotated proteins and proposed 47 protein–protein and 155 protein–metabolite interactions, including the LETM1–Citrate Synthase complex.

## RESULTS

### Multiplexed Baits and Orthogonal Labeling Chemistries Dramatically Expand Mitochondrial Proteome Coverage

To achieve comprehensive organellar coverage, we expanded our proximity labeling (SR-PL) toolkit, building upon the foundational framework established in our previous work.<sup>13,14,18</sup> We utilized two enzymatic systems with distinct chemistries: APEX2<sup>19</sup>, which generates biotin-phenoxy radicals that label tyrosine residues primarily through contact-dependent labeling within ~10 nm, with low-level diffusive labeling extending further<sup>20,21</sup>, and BioID/BioID2, which produces biotin-AMP esters that modify lysine residues (radius ~10 nm)<sup>22</sup> (**Figure 1A**).

We generated 42 stable Flp-In 293T-REx (hereafter HEK293T-REx) cell lines expressing 38 sub-mitochondrial and four non-mitochondrial baits. We strategically biased APEX2 toward the matrix and IMS (17 baits) to leverage its high labeling efficiency in dense compartments. Notably, we previously observed that when APEX2 is fused to OMM transmembrane (TM) proteins such as EXD2, the resulting biotinylation pattern is highly diffusive and cytosolic, raising concerns about the reliability of APEX2 for OMM proteomic analysis.<sup>6</sup> To circumvent this limitation, BioID was primarily deployed to the OMM (11 baits), where its shorter labeling radius produces more confined, neighborhood-specific labeling that faithfully captures the local OMM protein environment (**Figure 1C**).<sup>6,20,22</sup> Bait localization was confirmed via confocal microscopy and transmission electron microscopy (TEM) combined with 3,3'-diaminobenzidine (DAB) polymerization, while enzymatic activity was validated by streptavidin-conjugated horseradish peroxidase (Streptavidin-HRP) western blotting (**Figures 1D, S1A, S1B, S1C, and S1D**).

LC-MS/MS analysis of 129 biological replicates following SR-PL yielded 13,440 unique proximity-labeled sites (6,046 desthiobiotin-phenol (DBP)-labeled Tyr; 7,394 biotin-labeled Lys) across 4,058 proteins (**Figure 1A**). This is a 19-fold increase compared to our prior SR-PL study.<sup>18</sup> UMAP analysis of biological replicates confirmed that samples clustered by sub-mitochondrial compartment for both APEX2 and BioID baits, demonstrating the reproducibility and spatial specificity of the labeling approach (**Figure 1E**). Together, APEX2 and BioID detected 772 MitoCarta3.0 proteins, with 383 identified by both enzyme systems (**Figure 1F**). On average, APEX2 labeled 2.39 tyrosine residues per protein, while BioID labeled 2.54 lysine residues per protein (**Figure 1G**).

Critically, the identification of multiple labeled sites within the same protein not only increases the confidence of sub-mitochondrial localization assignments but also provides the spatial resolution necessary to determine the membrane topology of transmembrane proteins;

that is, which functional domains are exposed to the matrix, IMS, or cytosol across the lipid bilayer. This site-level resolution represents a fundamental advance over existing resources such as MitoCarta3.0, which assigns proteins to broad sub-mitochondrial categories (e.g., "OMM," "IMM," "IMS," "Matrix") without topological detail. By contrast, MitoAtlas provides refined classifications including membrane-spanning orientation and confidence tiers (Gold/Silver), yielding more functionally informative annotations (**Figure 1B**).

The value of our multiplexed approach is further evidenced by the expanded proteome coverage achieved using multiple baits within the same sub-organellar region. For instance, combining data from mitochondrial targeting sequence (MTS)-APEX2, HSPA9-APEX2, and MRM1-APEX2 revealed a vast array of distinct modified peptides, significantly increasing the depth of the matrix proteome (**Figure S1E**). When the same bait proteins were used with both systems (e.g., TOM20, EXD2, and TDRKH), BioID generated a higher proportion of non-overlapping modified peptides compared to APEX2 (**Figure S1E**). This is consistent with the distinct labeling mechanisms of the two enzymes: APEX2 operates through both contact-dependent and diffusive labeling, capturing a broader spatial neighborhood, whereas BioID labels exclusively through a contact-dependent mechanism, detecting only direct interaction partners.<sup>20</sup> This dual-enzyme strategy thus provides both the breadth required for organelle-wide mapping and the specificity needed to define local protein environments.

We observed that APEX2 and BioID display distinct sequence preferences and steric limitations, distinct from previously reported N-hydroxysuccinimide ester modified sites (Figure S1F).<sup>23</sup> Sequence motif analysis revealed that glycine (Gly) is frequently adjacent ( $\pm 1$ ) to APEX2-modified tyrosines, likely because its small side chain minimizes steric interference with the reactive phenol radical. The sequence logo also revealed that electron-rich amino acids capable of radical quenching, including Cys, Trp, Met, Arg, Phe, Leu, Tyr, and His, were negatively enriched at positions flanking the labeled tyrosine (Figure S1F). Among these, cysteine exhibited the strongest depletion, consistent with the high reactivity of sulfur-containing and aromatic amino acid side chains toward electrophilic radicals ( $k = 10^9\text{--}10^{10} \text{ M}^{-1} \text{ s}^{-1}$ , compared with  $\sim 10^7 \text{ M}^{-1} \text{ s}^{-1}$  for aliphatic residues such as glycine).<sup>24</sup> These flanking residues likely intercept the short-lived phenoxy radical before it forms a stable covalent adduct at the target tyrosine, providing mechanistic insight into the sequence selectivity of APEX2 labeling. In contrast, BioID modification sites favored branched amino acids (Val, Leu, Ile) at the +1 position, consistent with the substrate preference of the ancestral *E. coli* BirA enzyme, supporting its contact-dependent labeling mechanism.<sup>20,25</sup>

To further explore the physical constraints of labeling, we analyzed the correlation between solvent-accessible surface area (SASA) and peptide detection using AlphaFold2 structures. APEX2 showed an optimal probe radius near that of water (1.1–1.2 Å), suggesting it can probe deep into small protein clefts. In contrast, BioID exhibited a much larger optimal radius (11.5 Å), indicating that the bulkiness of the biotin-AMP molecule restricts labeling to more exposed surface residues (**Figures S1G and S1H**). These distinct labeling constraints have direct consequences for MitoAtlas: APEX2's small effective probe radius enables it to reach buried residues critical for resolving transmembrane topology (**Figure 3**), while BioID's contact-dependent mechanism, confirmed here by the retained BirA substrate preference, underlies the identification of specific bait–prey interactions (**Figure 6**).

### **A Two-Step Classification Pipeline for Sub-mitochondrial Proteome Mapping**

The high-density labeling profiles generated by our 42-bait library provided a unique fingerprint for each proximity-labeled site. We observed that MS intensities for sites within proteins such as MRPL22 (Matrix), MICU2 (IMS), FKBP8 (OMM), and ACTB (non-mitochondrial) exhibited compartment-specific enrichment patterns across the bait panel, suggesting that these signatures could be used for systematic spatial classification (**Figure S2A**).

To build a classification framework, we developed a two-step computational pipeline (**Figure 2A**). In Step 1 (Site Classification), we evaluated six supervised learning models and selected Logistic Regression (LR) for its accuracy, interpretability, and tunable probability thresholds (**Figure S2B**). Separate models were trained on APEX2 and BioID datasets, with probability thresholds optimized to maintain an overall FDR below 10% (**Figures S2C and S2D**). UMAP visualization of the classified sites confirmed distinct compartment-specific clustering for both APEX2- and BioID-labeled peptides (**Figure 2B**). By integrating data from all 42 baits simultaneously, MitoAtlas outperforms conventional binary enrichment analysis based on a single bait–control comparison. We benchmarked this on the most challenging classification task: discriminating OMM from non-mitochondrial sites, where OMM proteins face the cytosol and share the labeling environment with cytosolic background. MitoAtlas reduced classification error by 43% for APEX2 (AUC = 0.923 vs. 0.866 for TOM20-APEX2 vs. APEX2-NES) and by 58% for BioID (AUC = 0.976 vs. 0.943 for TOM20-BioID vs. mCherry-BioID) (**Figure S2E**).

In Step 2 (Protein Classification), site-level P1 outputs were aggregated into protein-level features and classified by a second LR model (**Figures 2A and S2H**). For non-TM proteins, we implemented a tiered “Gold” and “Silver” hierarchy reflecting the depth of evidence for each

assignment. For Matrix and IMS proteins, those with a single labeled site were designated Silver and those with two or more sites as Gold. Recognizing the higher risk of cytosolic contamination at the outer membrane, we applied more stringent filters for OMM proteins: a minimum of two sites for Silver and three or more for Gold. This pipeline successfully categorized 627 non-TM proteins, including Matrix-Gold/Silver (490), IMS-Gold/Silver (64), and OMM-Gold/Silver (73), while filtering out 2,760 non-mitochondrial proteins (**Figures 2A, S2F, and S2H**). At the selected P2 thresholds, protein-level FDR remained below 5% for Matrix, IMS, and non-mitochondrial compartments; the higher FDR for OMM (13%) reflects the inherent difficulty of discriminating OMM from cytosolic proteins, which motivated the more stringent Gold/Silver site-count requirements described above (**Figure S2G**).

An additional ~1,700 labeled sites did not meet the multi-bait consensus threshold for confident compartment assignment. These inconclusive sites were predominantly BioID-derived lysine modifications (1,323 of 7,394 BioID sites, 17.9%, versus 364 of 6,046 APEX2 sites, 6.0%). Among these inconclusive sites, 96% (APEX2) and 90% (BioID) were detected by at least one mitochondrial bait, indicating genuine mitochondria-proximal labeling events rather than cytosolic background (**Figure S2K**). Within this group, 37% of BioID cases were detected by only a single bait, suggesting that BioID retains a contact-dependent component of its ancestral BirA mechanism, preferentially capturing direct protein–protein interactions. We retained all inconclusive sites with full probability scores in the MitoAtlas database, as they likely reflect contact-dependent bait–prey interactions and represent candidate interactors amenable to targeted validation (see **Figure 6**).

For transmembrane (TM) proteins, we integrated site-level P1 values with transmembrane domain predictions from four independent sources: UniProt curated annotations, DeepTMHMM (sequence-based deep learning), TMbed (protein language model embeddings), and TmDet 4.0 (AlphaFold2 structure-based detection)<sup>26-30</sup> to assign membrane topology (**Figures 2A and S2H**). This workflow uses a topology scoring system that evaluates the concordance between site localizations and predicted TM domain boundaries, as illustrated for ATAD3B, an IMM-TM protein whose mixed Matrix and IMS site assignments are reconciled through integration with the DeepTMHMM-predicted topology and AlphaFold structure (**Figure S2I**). This pipeline successfully assigned membrane topology to 136 IMM-TM, 60 OMM-TM, and 25 MAM-TM proteins, with ambiguous cases classified into “Mito-TM” or “OMM-TM or MAM-TM” categories.

MitoAtlas resolves sub-mitochondrial localizations at a resolution inaccessible to existing databases.<sup>29</sup> MitoCarta3.0 groups all 44 Complex I subunits as “MIM,” whereas MitoAtlas

distinguishes matrix-facing and IMS-facing peripheral subunits from core transmembrane components, with topologies concordant with established cryo-EM structures (**Figure 2D**). More broadly, Sankey diagrams tracing annotation flow across MitoCarta3.0, MitoAtlas, and UniProt show how the pipeline resolves ambiguous categories into precise sub-mitochondrial addresses (**Figure 2C**). In total, MitoAtlas encompasses 861 mitochondrial proteins (matrix/IMM-TM/IMS/OMM-TM/OMM peripheral = 490/136/64/60/73), covering 79% of OXPHOS subunits,<sup>31,32</sup> 91% of the mitochondrial ribosome,<sup>33</sup> and 86% of the TOM complex<sup>34</sup> (**Figure S2J**).

### High-Resolution Mapping of Mitochondrial Membrane Topology and Domain Architecture via SR-PL

Determining the membrane topology and spatial domain architecture of TM proteins is fundamental to understanding their biological functions. Yet current mitochondrial topology information in public databases remains far less resolved than our SR-PL dataset. Although MitoCarta3.0 annotates 469 proteins as localized to the inner (IMM) or outer (OMM) mitochondrial membranes, it provides no topological orientation data.<sup>35</sup> Similarly, other mitochondrial resources such as MitoCOP (high-confidence mitochondrial proteome from subtractive proteomics) and MitCOM (mitochondrial complexome from native gel profiling in yeast) provide protein inventories and complex compositions but do not resolve membrane topology, as their underlying methodologies were not designed to determine the spatial orientation of individual domains.<sup>2,15,36</sup>

To our knowledge, UniProt is the only database that provides mitochondrial topological information through manual curation of experimental evidence.<sup>29</sup> As of January 2026, UniProt contains 134 IMM-TM and 61 OMM-TM entries; however, only a fraction of these are supported by specific PubMed references for topology (52 IMM; 10 OMM), reflecting the historical limitations of high-throughput methods (**Figure 3A**). MitoAtlas, by contrast, resolves the membrane topologies of all 136 IMM-TM and 60 OMM-TM proteins in our dataset by leveraging the inherent spatial specificity of SR-PL. Relative to curated UniProt records, MitoAtlas identified 78 previously uncharacterized mitochondrial TM proteins (52 IMM-TM; 26 OMM-TM) (**Figure 3B**). Of the 196 TM proteins identified in MitoAtlas, 38 are in full agreement with curated experimental evidence from UniProt (**Figure 3C**). Beyond these validated proteins, MitoAtlas resolved the topology of newly identified proteins such as TMEM256. Site-specific labeling indicates a topology in which both the N- and C-termini reside within the mitochondrial matrix, an architecture validated by C-terminal APEX2-EM imaging (**Figure 3D**).

However, our SR-PL data reveal several topological conflicts with existing UniProt annotations and sequence-based predictions. For instance, UniProt reports only 5 TM domains

for SLC30A9 and SFXN2, an architecture incompatible with our labeling patterns. We instead propose a 6-TM domain model for both SLC30A9 (**Figure 3E**) and SFXN2 (**Figure S3A**), a configuration independently supported by AlphaFold structural modeling. Our data also reassigns TMEM11; previously characterized as an IMM protein with two UniProt-predicted TM domains<sup>37</sup>, our results indicate it is an OMM-TM protein with a 3-TM architecture, consistent with AlphaFold predictions (**Figure S3B**). Similarly, whereas protease protection assays previously suggested the TMEM126A N-terminus faces the matrix<sup>38</sup>, our SR-PL results indicate the N-terminus is oriented toward the IMS (**Figure S3C**). These discrepancies underscore the limitations of conventional methodologies. Approaches such as protease protection or antibody-based assays are often conducted under non-physiological conditions or rely on the specificity of reagents that may lead to erroneous assignments. In contrast, our SR-PL strategy conducts proximity labeling in situ within intact living cells, providing a high-resolution, physiologically relevant depiction of the mitochondrial membrane proteome.

Our SR-PL findings demonstrate high concordance with AlphaFold2 (AF2) models: across 29 IMM-TM carrier proteins and 8 multi-pass OMM-TM proteins, 97% of MitoAtlas labeling sites (112/115) matched the expected compartment assignment based on AF2 structures (**Figures 3F-G**). MitoAtlas also provides experimental support for AF2 predictions of novel proteins. For example, UniProt classifies DNAJC11 as a peripheral OMM protein, but MitoAtlas identified both IMS and OMM sites, characterizing it as an OMM-TM protein consistent with an AF2-predicted 17-stranded  $\beta$ -barrel structure harboring an internal pore (**Figures 3H-I, S4E**). Because  $\beta$ -barrel proteins, including VDAC1–3, SAMM50, and TOMM40, are a rare class restricted to the OMM for transport and signaling, our data suggest DNAJC11 is a novel member of this family, with its N-terminal DnaJ domain exposed to the cytosol. Similarly, this integrative approach allowed us to resolve the disputed topology of LETM1.<sup>7</sup> While initially presumed to be an IMM-TM protein, combining our spatial labeling patterns with AF2 models indicates that LETM1, along with LETMD1<sup>39</sup> and FOXRED1, is a peripheral matrix-resident protein (**Figure S3F**).

Beyond individual proteins, MitoAtlas enables the mapping of sub-mitochondrial functional domains. For the m-AAA protease AFG3L2<sup>40</sup>, our data (IMS: Y179, Y211; Matrix: K488–Y756) confirm that the M41 FtsH domain localizes to the IMS while the AAA+ ATPase and lid domains reside in the matrix (**Figure 3J**). This domain-resolution mapping, available for all 861 proteins in our dataset, enables molecular-level analysis of mitochondrial function.

The traditional approach of fusing a single proximity labeling enzyme to a target protein can result in mislocalization artifacts. For instance, COX14 was previously assigned as an OMM

protein in our single-bait studies<sup>18,41</sup>, but multi-bait MitoAtlas SR-PL profiles and AlphaFold-Multimer predictions support its reclassification as an IMM-TM protein (**Figure S3G**). This misclassification likely arose because C-terminal fusion of proximity labeling enzymes prevented this IMM-TM protein from translocating past the OMM, effectively trapping it on the outer membrane. By integrating data from 42 multiplexed spatial baits and cross-validating with AlphaFold structural models, MitoAtlas successfully identifies and corrects such fusion-induced artifacts, demonstrating how a systems-level, multi-bait framework resolves ambiguities that single-enzyme approaches cannot.

### **Expanding the Mitochondrial Map: Characterization of 160 Mito-Orphan Proteins**

Through SR-PL, MitoAtlas identified 160 mitochondrial proteins previously unannotated in MitoCarta3.0,<sup>35</sup> which we termed “Mito-Orphans” (**Figure 4A**). These mito-orphans are distributed across all mitochondrial compartments: 38 matrix, 10 IMS, 11 IMM-TM, 66 OMM (including 7 OMM-TM), 23 MAM-TM, and 12 non-definitive transmembrane proteins (**Figure 4B**).

To validate these spatial assignments, we selected two matrix mito-orphans, GTPBP8 and C17orf80, which exhibited multiple matrix-specific SR-PL sites (**Figure 4C**). Confocal microscopy confirmed that APEX2 fusion constructs for both proteins targeted the mitochondria (**Figure 4D**), and DAB-enhanced electron microscopy established that the C-termini of both proteins reside in the matrix (**Figure 4E**). GTPBP8 has recently been identified as a matrix GTPase required for mitoribosomal biogenesis,<sup>42</sup> while C17orf80 was shown to interact with the mitochondrial nucleoid,<sup>43</sup> providing independent support for their matrix localization.

Among the newly identified mito-orphans, small integral membrane proteins (SMIMs) pose a particular challenge due to their extremely short domains. Using MitoAtlas, we resolved the membrane topologies of seven SMIMs: SMIM4, SMIM8, SMIM12, SMIM20, and SMIM26 as IMM-TM; SMIM19 as MAM-TM; and MTLN (SMIM37) as OMM-TM. MitoCarta3.0<sup>35</sup> previously categorized SMIM8 as “MOM” (OMM), while SMIM12, SMIM19, and SMIM26 were absent entirely. For cases where site-level data showed ambiguous IMS/OMM patterns — as expected for small IMM-TM proteins within the ~10 nm reach of OMM-anchored BioID baits — AlphaFold-Multimer predictions with known partners (SMIM8–AGK, SMIM12–PHB1) confirmed the IMM-TM assignment (**Figures 4F and S4A**). Confocal microscopy and proteinase K protection assays further validated the mitochondrial localization and IMM-TM topology of SMIM8 and SMIM12 (**Figures 4G and 4H**).

Since the IMS hosts a unique disulfide bond relay system mediated by MIA40/CHCHD4,<sup>44</sup>

we validated compartment-specific modifications in newly annotated IMS proteins. MINPP1 and FKBP9 were each identified with a single IMS-classified tyrosine site (**Figure 4I**), and AlphaFold2 predicted disulfide bonds in both<sup>45</sup> (**Figure 4J**). We focused on MINPP1, which contains a putative C-terminal ER-retention signal (SDEL) yet showed dual localization to both the ER lumen and the IMS by confocal microscopy (**Figure 4K**). Proteinase K protection confirmed its IMS localization (**Figure 4L**), and mmPEG24-maleimide modification confirmed the presence of disulfide bonds (**Figure 4M**). Unlike the bacterial homolog BtMinpp, human MINPP1 contains four predicted disulfide bonds with a highly conserved catalytic site (**Figures S4B–S4D**). Given its inositol polyphosphate phosphatase activity,<sup>46</sup> we propose that IMS-localized MINPP1 at ER–mitochondria contact sites can hydrolyze InsP6 to generate IP3 and promote mitochondrial calcium flux (**Figure 4N**).

To assess the functional relevance of mito-orphans, we examined *C. elegans* orthologs with mitochondria-associated phenotypes upon RNAi knockdown (**Figure S4E**). We selected *ykt-6*, the ortholog of YKT6 identified here as a novel OMM protein, for functional characterization. RNAi-mediated knockdown of *ykt-6* resulted in a significant reduction in body-bending frequency (**Figures 4O and 4P**), indicating impaired organismal motility consistent with a role in mitochondrial physiology.

### **Spatial Distribution of Functional Protein Domains Across Mitochondrial Sub-compartments**

SR-PL data enables both membrane topology determination and spatial mapping of functional protein domains and homologous superfamilies across mitochondrial compartments (**Figure 5A**). The mitochondrial matrix is enriched for GTP-binding domains, biotin/lipoyl attachment motifs, LYR domains,<sup>1,3</sup> and the NAD(P)-binding domain superfamily; spatial profiling of all NAD(P)-binding members confirms their predominant matrix sequestration (**Figures 5B and S5A**), enabling refined topological modeling of the inner membrane protein NNT (**Figure S5B**). The IMS is characterized by AAA-type ATPases, CHCH domains, and Tim10-like superfamilies. Notably, the EF-hand domain pair superfamily also shows a strong IMS preference (**Figure 5E**), consistent with the role of calcium signaling in this compartment. SR-PL site mapping of the calcium-binding mitochondrial carriers SLC25A13 and SLC25A24 resolves the spatial orientation of their EF-hand domains relative to the IMS (**Figure 5F**). LETM1, which mediates mitochondrial Ca<sup>2+</sup>/H<sup>+</sup> exchange, is a notable exception; its EF-hand domain faces the matrix rather than the IMS, alongside a LETM1-like ribosome-binding domain (**Figure 5G**). The OMM is predominantly associated with Bcl-2 homology (BH) domains, zinc finger motifs, and TPR domains. TPR-

containing OMM proteins (FIS1, FKBP8, TOMM70, RMDN2, RMDN3) are structurally positioned with cytosolic exposure, likely facilitating organelle–cytosol interactions (**Figures 5H, S5D, and S5E**).

This high-resolution mapping also uncovers atypical spatial partitioning. HSDL1, a NAD(P)-binding protein, localizes to the OMM rather than the matrix (**Figures 5C and S5C**), and RMDN1 resides in the IMS unlike its OMM-localized TPR paralogs RMDN2 and RMDN3 (**Figure 5I**). Such unexpected localizations, combined with structural homology, enable functional inference: CRYZ shares a conserved fold with the matrix oxidoreductase RTN4IP1 (pruned RMSD 1.17 Å; **Figures 5D and S5C**), suggesting analogous activity in CoQ metabolism or redox homeostasis,<sup>9,47,48</sup> and HSDL1 aligns tightly with HSDL2 (pruned RMSD 1.01 Å) despite their distinct compartments (**Figure S5C**). These examples demonstrate that MitoAtlas spatial data, combined with structural homology, can predict protein functions across sub-mitochondrial compartments.

### **MitoAtlas-Guided Prediction of Sub-mitochondrial Protein–Protein and Protein–Metabolite Interactions**

Proximity labeling captures not only protein localization but also the local protein neighborhood. By identifying proteins that share bait-specific labeled sites, MitoAtlas provides spatial constraints for predicting protein–protein interactions. We integrated these signatures with AlphaFold-Multimer to predict novel sub-mitochondrial interactions. Bait-prey pairs were filtered by site specificity and intensity and scored by AlphaFold-Multimer (**Figure 6A**); the resulting pTM + ipTM distribution identified a subset of high-confidence structural models (**Figure 6B**). For example, ss-APEX2-LETM1 yielded three specific labeled sites: CS-Y185, ECH1-Y149, and SLC25A3-Y196, all on proteins central to mitochondrial energy metabolism (**Figure 6C**). AF2-Multimer returned a high-scoring model for the LETM1–CS complex ( $0.8 \times \text{ipTM} + 0.2 \times \text{pTM} = 0.80$ ), and AlphaFold3<sup>49</sup> predicted a LETM1–CS $\times 2$  trimer with similar confidence (ipTM = 0.78, pTM = 0.82; **Figure 6D**). Two interface residues, R380 and R393, correspond to pathogenic variants (R380P, R393H) in LETM1-related mitochondrial disease (**Figure 6E**).<sup>7</sup> Co-immunoprecipitation confirmed that these interface-proximal pathogenic variants, along with structure-guided charge-reversal mutations R380E and R385E, specifically abolished LETM1–CS binding, whereas pathogenic variants distal to the interface (I293N, R294Q, P300S, D358N) retained it (**Figure 6F**; **Figure S6A**). Notably, E222Q, which abolishes LETM1-mediated Ca<sup>2+</sup>/H<sup>+</sup> antiport,<sup>50</sup> also disrupted CS binding,

suggesting that ion transport and CS interaction are structurally coupled. This interaction further supports LETM1's reclassification as a peripheral matrix protein, since its binding partner CS is a matrix enzyme.

This approach generalizes beyond LETM1. For example, EXD2-APEX2 specific sites identified interactions with cytoskeletal proteins including CNN3 and ACTB (**Figures S6B and S6C**), consistent with EXD2's OMM localization and cytosolic exposure. Network visualization of all EXD2 bait-prey connections (**Figure S6D**) and the complete bait-prey landscape across the 42-bait library (**Figure S6E**) illustrate the breadth of spatial interaction data captured by MitoAtlas.

MitoAtlas topology also independently validates external PPI predictions. Applying sub-mitochondrial topology as a spatial filter to 29,257 predicted mitochondrial PPIs,<sup>51,52</sup> MitoAtlas identified 77 topologically incompatible pairs (7.0%), compared to 26 (1.7%) by MitoCarta3.0 (**Figure S6F**). Incompatible interactions exhibited significantly lower prediction confidence (median 0.831 vs. 0.946; Wilcoxon  $p = 6.27 \times 10^{-7}$ ), providing reciprocal validation: MitoAtlas topology confirms computational predictions, while prediction scores validate MitoAtlas spatial assignments.

We next asked whether MitoAtlas could facilitate discovery of protein–metabolite interactions. Intersecting HProteome-BSite predictions<sup>53</sup> with MitoAtlas-localized proteins, we screened for interactions involving mitochondria-specific metabolites (coenzyme Q, heme, cholesterol, lipids, cardiolipins), identifying 155 candidates with high structural similarity to known binding templates (TM-score  $\geq 0.5$ ; **Figure 6G, H**). Among these, GalaxySite docking and AlphaFold3 independently predicted a heme-binding site on MTX1 at the OMM (**Figure 6J**).<sup>53-55</sup> Purified recombinant MBP–MTX1 exhibited tryptophan fluorescence quenching upon heme titration, confirming the interaction (**Figures S6G and 6K**).

### **MitoAtlas.org: An Interactive Database for the Mitochondrial Research Community**

MitoAtlas is accessible at [mitoatlas.org](https://mitoatlas.org) (**Figure 7A**). For each of the 861 proteins, the platform provides a protein overview panel summarizing gene and protein names, synonyms, UniProt accession, database cross-references (MitoAtlas, UniProt/PubMed, MitoCarta), SR-PL labeled sites identified by APEX2 and BioID/BioID2, sub-mitochondrial localizations of both the protein and its individual sites, and transmembrane domain annotations. Five additional interactive

modules enable in-depth exploration: a linear domain architecture map overlaying SR-PL labeled tyrosine and lysine sites onto InterPro domain annotations<sup>56</sup> with compartment-specific color coding; an interactive membrane topology viewer that displays labeled sites across the lipid bilayer with logistic regression–derived topology scores; a multi-track sequence feature viewer integrating domain boundaries, labeled sites, AlphaFold2 pLDDT confidence scores, and biochemical properties in a single zoomable browser; a bait-resolved SR-PL intensity heatmap showing APEX2 and BioID signal profiles across all 42 proximity labeling baits with classification confidence for each site; and projection of labeled sites onto PDB and AlphaFold2 three-dimensional structures via the iCn3D viewer<sup>57</sup>, enabling spatial visualization of compartment assignments on experimentally determined and predicted protein structures. All data are downloadable in tabular format for computational analysis and integration with external datasets (tabular download function not implemented in current pre-submission stage).

Three application scenarios demonstrate how MitoAtlas site-resolved spatial data enhances biological discovery beyond compartment-level annotations (**Figure 7B**; see also **Figure S7**). First, mapping ClinVar<sup>67</sup> pathogenic variants onto MitoAtlas membrane topology reveals topology-dependent disease patterns. For SLC25A4 (ANT1), an IMM carrier with six transmembrane helices, nine pathogenic variant positions segregate by sub-mitochondrial localization: variants within or near IMS-facing TM helices (positions 98, 104, 114, 289) are exclusively associated with autosomal dominant progressive external ophthalmoplegia (adPEO; MIM #609283), whereas variants in matrix-facing loops (positions 141, 235, 236) cause mtDNA depletion syndrome type 12 with cardiomyopathy (MIM #615418, #615438). Across all mitochondrial TM proteins, this approach assigned 3,139 disease variants in 88 genes to specific membrane faces: 556 (17.7%) to the matrix, 666 (21.2%) to the IMS, 502 (16.0%) within TM helices, and 1,415 (45.1%) to the cytosolic face of OMM proteins, a topological resolution unavailable from existing mitochondrial databases. Additional examples, including COX15, SLC25A22, and YME1L1, further illustrate topology-dependent disease segregation across diverse mitochondrial TM proteins (**Figure S7A**). Second, PTM sites show clear compartment specificity when mapped onto MitoAtlas topology. For ANT1, a single phosphoserine (position 7) localizes to the IMS-facing N-terminal segment, while three succinyllysine sites (positions 147, 245, 272) are exclusively found on matrix-facing loops, consistent with IMS kinase accessibility and non-enzymatic acylation driven by matrix succinyl-CoA concentration, respectively. Across all mitochondrial TM proteins, succinylation density is highest on the matrix face of IMM proteins and nearly absent from OMM-TM faces (**Figure 7B**). This compartment-dependent pattern extends to phosphorylation, acetylation, and ubiquitination across all mitochondrial TM proteins

**(Figure S7B).** Third, analysis of gnomAD population variant density<sup>59</sup> across MitoAtlas-defined topological faces reveals that TM domains of IMM proteins are strongly depleted in missense variants compared to soluble loops (43.1 vs. 45.0–51.8 variants per 100 aa; binomial  $p=8.0 \times 10^{-33}$ ), indicating strong purifying selection on membrane-embedded residues. N-terminal regions (positions 1–50), encompassing the mitochondrial targeting sequence, show consistently higher variant density than the remainder of the protein, consistent with reduced constraint on the cleaved MTS **(Figure S7C)**.

In total, MitoAtlas contains 861 proteins, 13,440 SR-PL sites, 196 resolved topologies, 160 mito-orphans, 47 bait–prey interactions, and 155 protein–metabolite interactions, and will be updated as data from additional cell types become available.

## DISCUSSION

In this study, we present MitoAtlas, a spatial systems biology resource that resolves the sub-mitochondrial domain architecture of 861 proteins at high resolution. Rather than analyzing individual bait–prey relationships in isolation, MitoAtlas integrates proximity labeling data from 42 spatially defined APEX2 and BioID/BioID2 baits within a unified logistic regression framework, enabling systems-level classification that accounts for the full spatial context of each labeled site. This integrated approach identified 13,440 unique labeled sites whose MS intensity profiles encode precise spatial information. The logistic regression classifiers decode these profiles to assign sub-mitochondrial localizations at the level of individual tyrosine and lysine residues, a resolution not achievable by conventional fractionation-based approaches or binary bait-control comparisons.<sup>2,12,16,35,36</sup> This site-level resolution enables MitoAtlas to provide not only protein localization but also membrane topology and domain architecture for all 196 transmembrane proteins in our dataset, information that is absent from existing mitochondrial databases. Moreover, MitoAtlas assigns spatial localizations to 160 proteins not previously annotated by MitoCarta3.0, substantially expanding the known mitochondrial proteome.

A key advance of MitoAtlas is the systematic determination of membrane topologies for mitochondrial transmembrane proteins. As of January 2026, UniProt contains curated topological annotations for only 52 IMM and 10 OMM proteins supported by specific experimental references.<sup>29</sup> These annotations derive largely from protease-protection assays conducted on purified mitochondrial fractions under non-physiological conditions, which can yield conflicting results across laboratories. SR-PL overcomes these limitations by conducting proximity labeling in situ within intact living cells, and we demonstrate 97% concordance between our site assignments and AlphaFold2-predicted structures across multi-pass transmembrane carriers. Importantly, this concordance establishes MitoAtlas as an experimental validation platform for AI-predicted protein structures, a capability of growing significance as the field increasingly relies on computational structural models. MitoAtlas thereby provides a comprehensive and experimentally grounded topology resource that complements and extends the existing UniProt annotations.

Our topology analyses revealed several unexpected structural features. DNAJC11, previously classified as a peripheral OMM protein<sup>60</sup>, was reclassified as an OMM-TM protein with a predicted 17-stranded  $\beta$ -barrel architecture. This finding aligns with recent CRISPR screening data implicating DNAJC11 in the mitochondrial protein import pathway.<sup>61</sup> While the original study attributed this phenotype to MICOS complex dysfunction<sup>62</sup>, the predicted pore diameter of DNAJC11 (14 Å) is comparable to that of TOMM40, the primary OMM import channel, raising the

possibility that DNAJC11 may serve as an accessory import pore (**Figures S3E**). Similarly, our reclassification of LETM1 as a peripheral membrane protein rather than a single-pass IMM-TM protein informs its disputed role in Wolf–Hirschhorn Syndrome and its proposed function as a K<sup>+</sup>/H<sup>+</sup> antiporter.<sup>7</sup>Notably, because K<sup>+</sup>/H<sup>+</sup> antiport requires a transmembrane ion-conduction pathway, the reclassification of LETM1 as a peripheral protein raises the possibility that it participates in ion transport indirectly, perhaps as a regulatory subunit of a larger channel complex, rather than functioning as an autonomous antiporter. These examples illustrate how site-resolved proximity labeling can resolve longstanding topological ambiguities that have hindered functional characterization.

It is also noteworthy that our SR-PL findings uniquely demonstrate that BioID retains the BirA enzyme's substrate preference<sup>23</sup> (e.g., branched amino acids (Val, Leu, Ile) at the +1 position, **Figure S1F**), supporting its contact-dependent labeling mechanism.<sup>20,25</sup> We hope that these insights will advance both the mechanistic understanding and rational engineering of proximity labeling enzymes.

Beyond topology determination, MitoAtlas enables systematic analysis of how sub-mitochondrial compartmentalization shapes both post-translational modification landscapes and evolutionary constraint. Mapping UniProt-annotated PTM sites onto MitoAtlas-defined topological faces reveals striking compartment specificity: phosphorylation is enriched on IMS-facing and cytosol-facing surfaces accessible to kinases, acetylation clusters on matrix-facing loops consistent with acetyl-CoA availability and SIRT3-regulated deacetylation, and ubiquitination is almost exclusively confined to the cytosol-facing domains of OMM proteins accessible to the ubiquitin-proteasome system (**Figure S7B**). These patterns demonstrate that the topology assignments are not merely structural descriptors but reflect the distinct biochemical environments of each mitochondrial compartment. Similarly, analysis of gnomAD population variant density<sup>59</sup> across topological faces reveals that transmembrane domains of IMM proteins are under significantly stronger purifying selection than soluble loops (binomial  $p=2.0 \times 10^{-33}$ ), while N-terminal mitochondrial targeting sequences show elevated variant density consistent with reduced constraint on the cleaved MTS (**Figure S7C**). Together, these orthogonal validations, spanning disease genetics, post-translational biochemistry, and population genomics, demonstrate that site-resolved spatial data provides biological insights inaccessible from compartment-level annotations alone.

## Limitations of the Study

While MitoAtlas provides a comprehensive sub-mitochondrial spatial map, several limitations should be noted. First, all SR-PL data were generated in a single cell type (HEK293T-REx), which limits coverage of cell-type-specific proteins such as UCP1 in brown adipocytes or CYP11A1 in steroidogenic tissues; consequently, the 861 proteins resolved here are fewer than the collective inventories of databases that aggregate data across multiple cell types and organisms. However, this single-cell-type design provides an internally consistent spatial map in which all localizations, topologies, and expression levels are measured under identical conditions; information that cannot be extracted from aggregated databases. Second, the computationally predicted protein–protein and protein–metabolite interactions, although constrained by experimentally determined topologies, require systematic experimental validation. Third, proximity labeling coverage depends on bait accessibility, and proteins in poorly accessible sub-compartments may be underrepresented. We address this by providing probability-based confidence scores for all classifications, and the ~1,700 inconclusive sites that did not meet the multi-bait consensus threshold are retained in the database for future investigation. To overcome these limitations, we are expanding MitoAtlas to include SR-PL datasets from additional human cell lines and mammalian tissues using transgenic mouse lines expressing mitochondria-targeted APEX2<sup>9</sup>, with the goal of building a tissue-resolved mitochondrial atlas.

## ACKNOWLEDGMENTS

MitoAtlas project was co-initiated by the laboratories of Prof. Hyun-Woo Rhee and Prof. Jong-Seo Kim. This work was supported by the National Research Foundation of Korea (RS-2025-15782968, RS-2023-KH136879, RS-2023-00260454 and RS-2023-00265581 to H.W.R.; RS-2024-00343424 and RS-2024-00440824 (the Bio&Medical Technology Development Program) to J.-S.K.). H.W.R. was supported by Samsung Science and Technology Foundation (SSTF-BA2201-08). J.-S.K. was supported by the Institute for Basic Science of the Ministry of Science and ICT of Korea (IBS-R008-D1). J.Y.M. and M.J. were supported by the Korea Brain Research Institute (KBRI) Basic Research Program, funded by the Ministry of Science and ICT (Information & Communication Technology) (23-BR-01-03). The EM data was acquired from the Brain Research Core Facilities of KBRI. Yun-Bin Lee helped cloning the novel mitochondrial protein APEX2-conjugated plasmids.

## AUTHOR CONTRIBUTIONS

H.W.R. and J.S.K., conceptualization; H.W.R., J.S.K., S.S., J.K., writing draft; S.Y.L. and C.K., cloning, stable cell generation and proteomics sample preparation; S.S., sample preparation, LC-MS analysis, and data analysis; J.K., data analysis, AF2 modeling, data curation and web application development; M.G.K., confocal microscope imaging; M.J. and J.Y.M., electron microscope imaging; H.L. and S.J.V.L., functional assay using *C. elegans*; J.S., metabolite docking; H.W.R. and J.S.K., supervision.

## DECLARATION OF INTERESTS

The authors declare no competing interests related to this work.

## DECLARATION OF GENERATIVE AI AN AI-ASSISTED TECHNOLOGIES IN THE WRITING PROCESS

During the preparation of this work the author(s) used Claude (Anthropic) in order to language editing and improving manuscript readability. After using this tool/service, the author(s) reviewed and edited the content as needed and take(s) full responsibility for the content of the publication.

## STAR★METHODS

### RESOURCE AVAILABILITY

#### ***Lead contact***

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Hyun-Woo Rhee (rheehw@snu.ac.kr).

#### ***Materials availability***

All unique reagents generated in this study are available from the lead contact with a completed Materials Transfer Agreement.

### EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

#### **Material and Method**

##### **Stable Cell Line Selection and Culturing**

Flp-In T-REx 293 cells (Life Technologies) were transfected using Lipofectamine 2000 (Life Technologies), typically with 20  $\mu$ L Lipofectamine 2000 and 4000 ng plasmid (9:1,pOG44:pCDNA5) per T25 flask. After 24 h, cells were cultured and selected with hygromycin B (100–200  $\mu$ g/mL) for 2–3 weeks. To induce stable protein expression, the cells were treated with 100 ng/mL doxycycline. Cells were desthiobiotin-phenol-labeled or biotin-labeled depending on the bait enzyme used and were lysed 18–24 h after induction by doxycycline.

##### **Desthiobiotin-Phenol Labeling in Stably Expressed APEX Cell Lines**

Stable-expression cells were incubated with 250  $\mu$ M desthiobiotin-phenol at 37 °C under 5% CO<sub>2</sub> for 30 min. Subsequently, 750  $\mu$ L of 10 mM H<sub>2</sub>O<sub>2</sub> (diluted from 30% H<sub>2</sub>O<sub>2</sub>, Sigma-Aldrich, H1009) was added to each flask, for a final concentration of 1 mM H<sub>2</sub>O<sub>2</sub>. The flasks were gently agitated for 1 min at room temperature. The reaction was then quenched by adding 7.5 mL of Dulbecco's

phosphate-buffered saline (DPBS) additionally containing 10 mM Trolox, 20 mM sodium azide, and 20 mM sodium ascorbate to each flask. The cells were then washed three times with cold quenching solution (DPBS containing 5 mM Trolox, 10 mM sodium azide, and 10 mM sodium ascorbate) and lysed with 1.5 mL RIPA lysis buffer (50 mM Tris, 150 mM NaCl, 0.1% SDS, 0.5% sodium deoxycholate, and 1% Triton X-100), 1× protease cocktail, 1 mM phenylmethylsulfonyl fluoride (PMSF), 10 mM sodium azide, 10 mM sodium ascorbate, and 5 mM Trolox for 10 min at 4 °C. Lysates were clarified by centrifugation at 15 000 ×g for 10 min at 4 °C.

### **Biotin Labeling in Stably Expressed BioID and BioID2 Cell Lines**

Stable-expression cell lines of promiscuous biotin ligases (BioID, BioID2) were incubated with 50 μM biotin (Alfa Aesar, A14207) at 37 °C for 16-18 hrs under 5% CO<sub>2</sub> and lysed with RIPA buffer (ELPISBIO, EBA-1149) containing 1× protease cocktail (Invitrogen, 78438) for 30 min at 4°C. The cells were then washed three times with cold quenching solution (DPBS containing 5 mM Trolox, 10 mM sodium azide, and 10 mM sodium ascorbate) and lysed with 1.5 mL RIPA lysis buffer (50 mM Tris, 150 mM NaCl, 0.1% SDS, 0.5% sodium deoxycholate, and 1% Triton X-100), 1× protease cocktail, 1 mM phenylmethylsulfonyl fluoride (PMSF), 10 mM sodium azide, 10 mM sodium ascorbate, and 5 mM Trolox for 10 min at 4 °C. Lysates were clarified by centrifugation at 15 000 ×g for 10 min at 4 °C.

### **Labeled Peptide Enrichment**

For removal of unreacted free desthiobiotin-phenol or biotin-phenol, cell lysates were moved into an Amicon filter (Merck Millipore, 10 kDa cut-off) followed by centrifugation at 7500 ×g for 15 min at 4 °C. The volume was increased to 4 mL by adding phosphate-buffered saline (PBS) containing 1 mM PMSF and 1× protease cocktail, followed by centrifugation three more times. Finally, the cell lysates were transferred to an Eppendorf tube and mixed with 300 μL of streptavidin beads (Pierce). The sample was rotated for 1 h at room temperature and washed twice with PBS. After removing the PBS, 100 μL of denaturing solution (6 M urea, 2 M thiourea, and 10 mM 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES)) was added and reduced using 20 μL of 100 mM dithiothreitol (DTT) in 50 mM ammonium bicarbonate (ABC) buffer for 60 min at 56 °C using a ThermoMixer (Eppendorf). Protein alkylation was performed by adding 35 μL of 300 mM iodoacetamide in 50 mM ABC buffer with shaking in the dark for 30 min. Trypsin gold (Promega) was then added to the solution, which was then incubated at 37 °C with shaking overnight. Formic acid was then added to terminate the trypsin reaction, and the beads were washed with PBS four times. After adding 250 μL of 95% formamide, and 10 mM ethylene-diamine-tetraacetic acid

(EDTA) (pH 8.2), the elute was obtained by boiling at 95 °C for 10 min. The eluted peptide samples were then desalted using Varian Bond ELUT (Agilent, 12109301) and a home-made column.

### **Liquid chromatography tandem mass spectrometry (LC-MS/MS) analysis of enriched peptide samples.**

Analytical capillary columns (100 cm × 75 µm i.d.) and trap columns (2 cm × 150 µm i.d.) were packed in-house with 3 µm Jupiter C18 particles (Phenomenex, Torrance). The long analytical column was placed in a column heater (Analytical Sales and Services) regulated to a temperature of 45°C. The nanoAcquity UPLC system (Waters, Milford) was operated at a flow rate of 300 nL/min over 2 h with a linear gradient ranging from 95% solvent A (CH<sub>2</sub>O with 0.1% formic acid) to 40% of solvent B (acetonitrile with 0.1% formic acid). The enriched samples were analyzed on an Orbitrap Fusion Lumos mass spectrometer (Thermo Fisher Scientific) equipped with an in-house customized nanoelectrospray ion source. Precursor ions were acquired at 120 k resolving power (*m/z* 300–1500), and the isolation of precursor ions for MS/MS analysis was performed with a 1.4 Th. Higher-energy collisional dissociation (HCD) with 30% collision energy was used for sequencing with auto gain control (AGC) target of  $1 \times 10^9$ . The resolving power to acquire the MS<sub>2</sub> spectra was set to 30 k with 200 ms maximum injection time.

### **MS data processing and protein identification.**

All MS/MS data were screened using MaxQuant (v. 1.6.2.3) with the Andromeda search engine at a 10 ppm precursor ion mass tolerance against the SwissProt Homo sapiens proteome database (20,199 entries, UniProt; <http://www.uniprot.org>). Label free quantification (LFQ) and Match Between Runs were used, with the following search parameters for BioID experiments: semi-tryptic digestion, fixed carbamidomethylation of cysteine, dynamic oxidation of methionine, protein N-terminal acetylation, and biotinylation for lysine residues. Following parameters were used for APEX2 experiment: tryptic digestion, fixed carbamidomethylation on cysteine, dynamic oxidation of methionine, and protein N-terminal acetylation with DBP (+331.1896) for tyrosine residues. A false discovery rate (FDR) < 1% was obtained for the uniquely labeled peptides and unique labeled proteins. LFQ intensity values were log-transformed for further analysis.

### **Sub-mitochondrial localization classification by logistic regression**

Sub-mitochondrial localization was assigned through a two-step logistic regression (LR) pipeline implemented in Python (scikit-learn v1.2). In Step 1 (Site Classification, P1), separate one-vs-rest (OVR) LR models were trained for APEX2 and BioID datasets. The APEX2 site LR model was

trained on 660 manually curated sites (Matrix: 135, IMS: 149, OMM: 107, Cytosol: 269) and applied to predict the localization of all 6,046 APEX2-labeled tyrosine sites. Input features consisted of log<sub>10</sub>-transformed MS intensities across all 25 APEX2 baits, calibrated by iBAQ protein abundance. The BioID site LR model was trained on 576 manually curated sites (Matrix: 85, IMS: 80, OMM: 216, Cytosol: 195) and applied to predict the localization of all 7,394 BioID-labeled lysine sites, using log<sub>10</sub>-transformed MS intensities across all 17 BioID baits as features. Training sites were curated from proteins with well-established sub-mitochondrial localizations in the literature, with transmembrane proteins manually annotated at the individual residue level based on known topology. For both models, 10-fold cross-validation accuracy exceeded 90%, and probability thresholds were optimized to maintain an overall false discovery rate (FDR) below 10%.

In Step 2 (Protein Classification, P2), site-level P1 outputs were aggregated into protein-level features. For each protein, the logit-transformed P1 probabilities were summed across all labeled sites, yielding an 8-dimensional feature vector (4 APEX2 compartment scores + 4 BioID compartment scores: Matrix, IMS, OMM, Non-mitochondrial). A second OVR LR model was trained on curated reference protein localizations and applied to classify the 3,825 non-TM proteins. The remaining 233 TM proteins were instead routed to a topology scoring pipeline that integrates site-level P1 classifications with transmembrane domain predictions from DeepTMHMM, TMbed, and AlphaFold2 structural models. In total, 4,058 proteins were classified across both arms of the pipeline. Non-TM proteins were assigned to tiered confidence categories: Gold ( $\geq 2$  concordant sites above the P1 threshold) or Silver (1 site above the P1 threshold) for Matrix and IMS assignments. For OMM, the thresholds were set at  $\geq 3$  sites (Gold) and  $\geq 2$  sites (Silver) to account for the higher background labeling in the outer membrane compartment. Complete training sets and classification outputs are provided in Tables S1–S2.

## **Comparison of binary enrichment analysis and MitoAtlas classification**

To evaluate the added value of integrating multiple proximity labeling datasets over a conventional single-experiment analysis, we compared binary enrichment analysis with the MitoAtlas logistic regression (LR) classifier for discriminating outer mitochondrial membrane (OMM) from non-mitochondrial (cytosolic) biotinylation sites. For binary enrichment analysis, site-level intensities were extracted for the OMM-targeted bait condition and the cytosolic control condition (APEX2: TOM20-APEX2,  $n = 3$ , versus APEX2-NES,  $n = 6$ ; BioID: TOM20-BioID,  $n = 3$ , versus mCherry-BioID,  $n = 3$ ). Raw intensities were log<sub>2</sub>-transformed and median-centered across all channels to a common global median. Missing values were imputed using a Perseus-style missing-not-at-

random (MNAR) strategy, in which each missing value was drawn from a down-shifted normal distribution (mean = column mean – 1.8 SD, width = 0.3 SD)<sup>63</sup>. For each biotinylation site, the log<sub>2</sub> fold change was calculated as the difference in mean log<sub>2</sub> intensity between bait and control replicates, and statistical significance was assessed by a two-sided Welch's t-test. For MitoAtlas classification, the pre-computed LR posterior probabilities across compartments were used directly. Receiver operating characteristic (ROC) analysis was performed on the subset of sites with known compartment assignments (OMM as positive class; non-mitochondrial as negative class). For the binary approach, log<sub>2</sub> fold change served as the classification score; for MitoAtlas, p(OMM) was used. The classification threshold was swept across 500 evenly spaced values spanning the full range of each score, and the true positive rate and false positive rate were computed at each threshold to generate the ROC curve. The area under the ROC curve (AUC) was calculated by the trapezoidal rule. For APEX2, the analysis included 107 OMM and 269 non-mitochondrial training sites; for BioID, 216 OMM and 195 non-mitochondrial training sites. All analyses were performed in R (v4.5) using ggplot2 for visualization.

## Transmission Electron Microscopy

HEK293T cells were grown in 35-mm glass grid-bottomed culture dishes to 50%–60% confluence. The cells were transfected with APEX2-conjugated plasmid with lipofectamine reagent. The next day, the cells were fixed with 2.5% glutaraldehyde and 2% paraformaldehyde in 0.1 M cacodylate solution (pH 7.0). After washing, DAB staining was performed for 5–45 min until a light brown stain was visible under an inverted light microscope. DAB-stained cells were post-fixed with 2% potassium ferrocyanide-reduced osmium tetroxide for 1 h at 4 °C. The fixed cells were dehydrated using an ethanol series (50%, 60%, 70%, 80%, 90%, and 100%) for 10 min at each concentration, then infiltrated with an embedding medium. The identified samples were cut horizontally to the plane of the block into 60 nm sections (UC7; Leica Microsystems, Germany), mounted on copper slot grids with a specimen support film, then, double stained with UranylLess (Electron Microscopy Sciences, #22409) for 2 min and 3% lead citrate (Electron Microscopy Sciences, #22410) for 1 min. The sections were then analyzed using a transmission electron microscope at 120 kV (Tecnai G2, ThermoFisher, USA).

## AlphaFold2

AlphaFold v2.3.0 was downloaded from GitHub and installed as a dockerized application under Linux (Ubuntu 22, AMD Ryzen 9 5950X, 128 GB RAM, Nvidia RTX 4090 with 24GB RAM). The full databases were stored on a 4TB NVMe SSD using the flag "db\_preset=full\_dbs", and 25

multimer models were generated using the flag "num\_multimer\_predictions\_per\_model=5".

## Data Visualization

Protein structure visualization and PAE visualization were performed using UCSF ChimeraX (<https://www.rbvi.ucsf.edu/chimerax>).

## Web Application

The mitoatlas.org web application is based on the JavaScript frameworks React (<https://react.dev/>) and ExpressJS (<https://expressjs.com/>). The experimental data are stored and accessed using a MySQL relational database (<https://www.mysql.com/>). The iCn3D viewer was embedded using the npm package (<https://www.npmjs.com/package/icn3d>) to visualize the labeled sites on the AlphaFold2 structure. D3.js (<https://d3js.org/>) was used to generate heatmaps of the MS data. PubMed queries were performed using the Entrez Programming Utilities API (<https://www.ncbi.nlm.nih.gov/books/NBK25500/>).

## TM domain annotation sources

UniProt TM domain annotation were downloaded from the website using the UniProt ID-mapping service (<https://www.uniprot.org/id-mapping>). TMbed transmembrane domain annotations were downloaded from the pre-calculated human proteome data ([https://roslab.org/public/tmbed/predictions/human\\_210422\\_tmbed.tar.gz](https://roslab.org/public/tmbed/predictions/human_210422_tmbed.tar.gz)). DeepTMHMM TM domain annotations were obtained by running the BioLib application (<https://dtu.biolib.com/DeepTMHMM/>) on a local machine (Ubuntu 22, AMD Ryzen 9 5950X, 128 GB RAM, NVidia RTX 4090 with 24 GB RAM) as per the instructions on the website, using the UniProt primary protein sequence as input. AF2-TMDET TM domain boundaries were obtained by running TMDET 4.1 via Docker (brgenzim/tmdet:4.1.1) on AlphaFold2-predicted structures (CIF format) for each protein.<sup>30</sup>

## Membrane topology scoring and assignment

For each TM protein classified in Step 2, membrane topology was assigned using a topology scoring algorithm that integrates site-level sub-mitochondrial localizations (from Step 1) with TM domain boundary predictions from four independent sources: UniProt (curated annotations), DeepTMHMM (deep learning), TMbed (protein language model embeddings), and AF2-TMDET (AlphaFold2 structure-based detection). For each source, every SR-PL labeled site was classified relative to the predicted TM segments as residing in the N-terminal loop, within a TM segment, in

an inter-TM loop, or in the C-terminal loop. Sites in loop regions contributed a full count to the corresponding membrane face (lower or upper), while sites within TM segments contributed fractional counts to both faces based on their relative position within the segment (logit-transformed). IMM and OMM hypotheses were then scored: the IMM score rewards matrix sites on one face and IMS sites on the other; the OMM score rewards IMS sites on one face and OMM or cytosolic sites on the other. The source with the highest adjusted score was selected as the adopted TM model, with adjustments including a structural confidence bonus for AF2-TMDET predictions, penalties for short TM segments and TM count outliers, and a boundary-agreement bonus when independent methods (AF2-TMDET and TMbed) converged on the same TM architecture. Ties were resolved by consensus TM count among tied sources, with manual review for unresolved cases. N- and C-terminus positions were then assigned based on the winning orientation (score1 vs. score2) and the determined membrane identity.

For proteins where the topology scoring algorithm could not confidently distinguish between IMM and OMM assignment, typically due to ambiguous site localizations (e.g., IMS-only evidence, which is consistent with both membranes), AlphaFold-Multimer structure predictions with known interacting partners were used as additional evidence to resolve the membrane assignment. For example, the predicted SMIM8–AGK and SMIM12–PHB1 complexes confirmed IMM-TM assignments, while the SMIM20–COX14–MTCO1, and HIGD2A–MTCO3 complexes provided independent structural evidence for IMM localization. In cases where the algorithm returned ambiguous results for proteins with well-established sub-mitochondrial localization (e.g., OMM-resident TOM complex subunits and VDACs), the known sub-mitochondrial context was used to assign membrane identity.

### **Co-immunoprecipitation (CoIP)**

HEK293T-REx cells were transiently transfected with MTS-3xFLAG-LETM1 (WT or mutant) constructs using polyethyleneimine (PEI) and harvested 24 h post-transfection. Cells were lysed in PBS (pH 7.4) containing 1% NP-40 and 1x protease inhibitor cocktail for 10 min at 4 °C. Lysates were cleared by centrifugation at 15,000 xg for 5 min and incubated with anti-FLAG affinity gel (ChemScene) for 2 h at 4 °C with rotation. Beads were washed twice with PBS containing 0.1% NP-40 and eluted by boiling in 2x SDS-PAGE loading buffer at 95 °C for 5 min. Input and elution fractions were analyzed by western blot using anti-FLAG and anti-CS antibodies.

### **C. *elegans* strains and RNAi clones**

*C. elegans* strain N2 (wild-type) was maintained under standard laboratory culture conditions on

nematode growth medium (NGM) seeded with *E. coli* OP50 or HT115, which expressed double-stranded RNA for RNAi experiments. The RNAi clones used in this study were L4440 (control) and *ykt-6*.

### Measurement of *C. elegans* swimming rates

The *C. elegans* swimming rate (body bends per min) was measured as described previously.<sup>58</sup> Wild-type worms were treated with control RNAi or *ykt-6* RNAi for two generations before performing the assay. Fifteen pre-fertile adult worms were transferred to 24-well plates with 1 mL of M9 buffer per well. After stabilization for 1 min, body bends were recorded for 30 s using a digital microscope (DIMIS-M, Siwon Optical Technology, Anyang, South Korea); this number was then converted to the number of bends per min. *P* values were calculated using the unpaired Student's *t*-test (two-tailed), comparing the swimming rates of the control and RNAi-treated worms.

### Measurement of *C. elegans* developmental rates

Images of control RNAi- and *ykt-6* RNAi-treated worms cultured for two generations were obtained using a digital microscope (DIMIS-M, Siwon Optical Technology, Anyang, South Korea). Adult *C. elegans* were washed with M9 buffer and the remaining eggs were incubated for 2 hrs at 20°C. Hatched L1 larvae were transferred to new NGM plates seeded with OP50. Images were obtained at 20 °C, 44 h after transfer.

### Ligand-Binding Site Similarity Search and Docking Analysis

A similarity-based search for mitochondrial protein targets that may interact with key metabolites like heme was done using HProteome-BSite<sup>53</sup>. HProteome-BSite is a database which contains ligand binding sites and putative ligands for human proteins. With protein annotations from MitoAtlas, there were 263 mitochondrial proteins with HEM as their putative binding partner, and 209 with PTY. Findings included novel OMM protein like ARHGAP1. Complex structure of PTY and ARHGAP1, one of the novel OMM protein, was modeled using GalaxyDock3<sup>54</sup>, and CSAIalign<sup>55</sup>. Some novel interactions were found for known mitochondrial proteins such as MTX1. Results for MTX1 suggested interaction with HEM, and its complex structure was modeled using docking algorithm from latest version of GalaxySite<sup>53</sup>.

## REFERENCES

1. Angerer, H. (2015). Eukaryotic LYR Proteins Interact with Mitochondrial Protein Complexes. *Biology (Basel)* 4, 133–150. [10.3390/biology4010133](https://doi.org/10.3390/biology4010133).
2. Morgenstern, M., Peikert, C.D., Lubbert, P., Suppanz, I., Klemm, C., Alka, O., Steiert, C., Naumenko, N., Schendzielorz, A., Melchionda, L., et al. (2021). Quantitative high-confidence human mitochondrial proteome and its dynamics in cellular context. *Cell Metab* 33, 2464–2483 e2418. [10.1016/j.cmet.2021.11.001](https://doi.org/10.1016/j.cmet.2021.11.001).
3. Angerer, H., Radermacher, M., Mankowska, M., Steger, M., Zwicker, K., Heide, H., Wittig, I., Brandt, U., and Zickermann, V. (2014). The LYR protein subunit NB4M/NDUFA6 of mitochondrial complex I anchors an acyl carrier protein and is essential for catalytic activity. *Proc Natl Acad Sci U S A* 111, 5207–5212. [10.1073/pnas.1322438111](https://doi.org/10.1073/pnas.1322438111).
4. Kelekar, A., and Thompson, C.B. (1998). Bcl-2-family proteins: the role of the BH3 domain in apoptosis. *Trends Cell Biol* 8, 324–330. [10.1016/s0962-8924\(98\)01321-x](https://doi.org/10.1016/s0962-8924(98)01321-x).
5. Hensen, F., Moretton, A., van Esveld, S., Farge, G., and Spelbrink, J.N. (2018). The mitochondrial outer-membrane location of the EXD2 exonuclease contradicts its direct role in nuclear DNA repair. *Sci Rep* 8, 5368. [10.1038/s41598-018-23690-y](https://doi.org/10.1038/s41598-018-23690-y).
6. Park, J., Lee, S.Y., Jeong, H., Kang, M.G., Van Haute, L., Minczuk, M., Seo, J.K., Jun, Y., Myung, K., Rhee, H.W., et al. (2019). The structure of human EXD2 reveals a chimeric 3' to 5' exonuclease domain that discriminates substrates via metal coordination. *Nucleic Acids Res* 47, 7078–7093. [10.1093/nar/gkz454](https://doi.org/10.1093/nar/gkz454).
7. Kaiyrzhanov, R., Mohammed, S.E.M., Maroofian, R., Husain, R.A., Catania, A., Torracco, A., Alahmad, A., Dutra-Clarke, M., Gronborg, S., Sudarsanam, A., et al. (2022). Bi-allelic LETM1 variants perturb mitochondrial ion homeostasis leading to a clinical spectrum with predominant nervous system involvement. *Am J Hum Genet* 109, 1692–1712. [10.1016/j.ajhg.2022.07.007](https://doi.org/10.1016/j.ajhg.2022.07.007).
8. Pfanner, N., Warscheid, B., and Wiedemann, N. (2019). Mitochondrial proteins: from biogenesis to functional networks. *Nat Rev Mol Cell Biol* 20, 267–284. [10.1038/s41580-018-0092-0](https://doi.org/10.1038/s41580-018-0092-0).
9. Park, I., Kim, K.E., Kim, J., Kim, A.K., Bae, S., Jung, M., Choi, J., Mishra, P.K., Kim, T.M., Kwak, C., et al. (2023). Mitochondrial matrix RTN4IP1/OPA10 is an oxidoreductase for coenzyme Q synthesis. *Nat Chem Biol*. [10.1038/s41589-023-01452-w](https://doi.org/10.1038/s41589-023-01452-w).

10. Rhee, H.W., Zou, P., Udeshi, N.D., Martell, J.D., Mootha, V.K., Carr, S.A., and Ting, A.Y. (2013). Proteomic mapping of mitochondria in living cells via spatially restricted enzymatic tagging. *Science* 339, 1328–1331. 10.1126/science.1230593.
11. Pagliarini, D.J., and Rutter, J. (2013). Hallmarks of a new era in mitochondrial biochemistry. *Genes Dev* 27, 2615–2627. 10.1101/gad.229724.113.
12. Go, C.D., Knight, J.D.R., Rajasekharan, A., Rathod, B., Hesketh, G.G., Abe, K.T., Youn, J.Y., Samavarchi-Tehrani, P., Zhang, H., Zhu, L.Y., et al. (2021). A proximity-dependent biotinylation map of a human cell. *Nature* 595, 120–124. 10.1038/s41586-021-03592-2.
13. Hung, V., Zou, P., Rhee, H.W., Udeshi, N.D., Cracan, V., Svinkina, T., Carr, S.A., Mootha, V.K., and Ting, A.Y. (2014). Proteomic mapping of the human mitochondrial intermembrane space in live cells via ratiometric APEX tagging. *Mol Cell* 55, 332–341. 10.1016/j.molcel.2014.06.003.
14. Hung, V., Lam, S.S., Udeshi, N.D., Svinkina, T., Guzman, G., Mootha, V.K., Carr, S.A., and Ting, A.Y. (2017). Proteomic mapping of cytosol-facing outer mitochondrial and ER membranes in living human cells by proximity biotinylation. *Elife* 6. 10.7554/eLife.24463.
15. Singh, A.P., Salvatori, R., Aftab, W., Aufschneider, A., Carlstrom, A., Forne, I., Imhof, A., and Ott, M. (2020). Molecular Connectivity of Mitochondrial Gene Expression and OXPHOS Biogenesis. *Mol Cell* 79, 1051–1065 e1010. 10.1016/j.molcel.2020.07.024.
16. Antonicka, H., Lin, Z.Y., Janer, A., Aaltonen, M.J., Weraarpachai, W., Gingras, A.C., and Shoubridge, E.A. (2020). A High-Density Human Mitochondrial Proximity Interaction Network. *Cell Metab* 32, 479–497 e479. 10.1016/j.cmet.2020.07.017.
17. Lee, Y.B., Jung, M., Kim, J., Charles, A., Christ, W., Kang, J., Kang, M.G., Kwak, C., Klingstrom, J., Smed-Sorensen, A., et al. (2023). Super-resolution proximity labeling reveals anti-viral protein network and its structural changes against SARS-CoV-2 viral proteins. *Cell Rep* 42, 112835. 10.1016/j.celrep.2023.112835.
18. Lee, S.Y., Kang, M.G., Shin, S., Kwak, C., Kwon, T., Seo, J.K., Kim, J.S., and Rhee, H.W. (2017). Architecture Mapping of the Inner Mitochondrial Membrane Proteome by Chemical Tools in Live Cells. *J Am Chem Soc* 139, 3651–3662. 10.1021/jacs.6b10418.
19. Martell, J.D., Deerinck, T.J., Sancak, Y., Poulos, T.L., Mootha, V.K., Sosinsky, G.E., Ellisman, M.H., and Ting, A.Y. (2012). Engineered ascorbate peroxidase as a genetically encoded reporter

for electron microscopy. *Nat Biotechnol* 30, 1143–1148. 10.1038/nbt.2375.

20. Yang, Z., Zhang, Y., Fang, Y., Zhang, Y., Du, J., Shen, X., Zhang, K., Zou, P., and Chen, Z. (2025). Spatial barcoding reveals reaction radii and contact-dependent mechanism of proximity labeling. *Nat Chem Biol*. 10.1038/s41589-025-02086-w.

21. Oakley, J.V., Buksh, B.F., Fernandez, D.F., Oblinsky, D.G., Seath, C.P., Geri, J.B., Scholes, G.D., and MacMillan, D.W.C. (2022). Radius measurement via super-resolution microscopy enables the development of a variable radii proximity labeling platform. *Proc Natl Acad Sci U S A* 119, e2203027119. 10.1073/pnas.2203027119.

22. Kim, D.I., Birendra, K.C., Zhu, W., Motamedchaboki, K., Doye, V., and Roux, K.J. (2014). Probing nuclear pore complex architecture with proximity-dependent biotinylation. *Proc Natl Acad Sci U S A* 111, E2453–2461. 10.1073/pnas.1406459111.

23. Ward, C.C., Kleinman, J.I., and Nomura, D.K. (2017). NHS-Esters As Versatile Reactivity-Based Probes for Mapping Proteome-Wide Ligandable Hotspots. *ACS Chem Biol* 12, 1478–1483. 10.1021/acscchembio.7b00125.

24. Buxton, G.V., Greenstock, C.L., Helman, W.P., and Ross, A.B. (1988). Critical review of rate constants for reactions of hydrated electrons, hydrogen atoms and hydroxyl radicals ( $\cdot\text{OH}/\cdot\text{O}^-$ ) in aqueous solution. *J Phys Chem Ref Data* 17, 513–886. 10.1063/1.555805.

25. Schatz, P.J. (1993). Use of peptide libraries to map the substrate specificity of a peptide-modifying enzyme: a 13 residue consensus peptide specifies biotinylation in *Escherichia coli*. *Biotechnology (N Y)* 11, 1138–1143. 10.1038/nbt1093-1138.

26. Hallgren, J., Tsigos, K.D., Pedersen, M.D., Armenteros, J.J.A., Marcatili, P., Nielsen, H., Krogh, A., and Winther, O. (2022). DeepTMHMM predicts alpha and beta transmembrane proteins using deep neural networks. *bioRxiv*. 10.1101/2022.04.08.487609.

27. Bernhofer, M., and Rost, B. (2022). TMbed: transmembrane proteins predicted through language model embeddings. *BMC Bioinformatics* 23, 1–19. 10.1186/s12859-022-04873-x.

28. Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Zidek, A., Potapenko, A., et al. (2021). Highly accurate protein structure prediction with AlphaFold. *Nature* 596, 583–589. 10.1038/s41586-021-03819-2.

29. UniProt, C. (2022). UniProt: the Universal Protein Knowledgebase in 2023. *Nucleic Acids Res*. 10.1093/nar/gkac1052.

30. Tusnády, G.E., and Gerdán, C. (2025). TmDet 4.0: determining membrane orientation of transmembrane proteins from 3D structure. *Nucleic Acids Res* 53, W542–W546. 10.1093/nar/gkaf338.
31. Meldal, B.H.M., Perfetto, L., Combe, C., Lubiana, T., Ferreira Cavalcante, J.V., Bye, A.J.H., Waagmeester, A., Del-Toro, N., Shrivastava, A., Barrera, E., et al. (2022). Complex Portal 2022: new curation frontiers. *Nucleic Acids Res* 50, D578-D586. 10.1093/nar/gkab991.
32. Guo, R., Zong, S., Wu, M., Gu, J., and Yang, M. (2017). Architecture of Human Mitochondrial Respiratory Megacomplex I(2)III(2)IV(2). *Cell* 170, 1247–1257 e1212. 10.1016/j.cell.2017.07.050.
33. Dennerlein, S., Poerschke, S., Oeljeklaus, S., Wang, C., Richter-Dennerlein, R., Sattmann, J., Bauermeister, D., Hanitsch, E., Stoldt, S., Langer, T., et al. (2021). Defining the interactome of the human mitochondrial ribosome identifies SMIM4 and TMEM223 as respiratory chain assembly factors. *Elife* 10. 10.7554/eLife.68213.
34. Su, J., Liu, D., Yang, F., Zuo, M.Q., Li, C., Dong, M.Q., Sun, S., and Sui, S.F. (2022). Structural basis of Tom20 and Tom22 cytosolic domains as the human TOM complex receptors. *Proc Natl Acad Sci U S A* 119, e2200158119. 10.1073/pnas.2200158119.
35. Rath, S., Sharma, R., Gupta, R., Ast, T., Chan, C., Durham, T.J., Goodman, R.P., Grabarek, Z., Haas, M.E., Hung, W.H.W., et al. (2021). MitoCarta3.0: an updated mitochondrial proteome now with sub-organelle localization and pathway annotations. *Nucleic Acids Res* 49, D1541-D1547. 10.1093/nar/gkaa1011.
36. Schulte, U., den Brave, F., Haupt, A., Gupta, A., Song, J., Muller, C.S., Engelke, J., Mishra, S., Martensson, C., Ellenrieder, L., et al. (2023). Mitochondrial complexome reveals quality-control pathways of protein import. *Nature* 614, 153–159. 10.1038/s41586-022-05641-w.
37. Rival, T., Macchi, M., Arnauné-Pelloquin, L., Poidevin, M., Maillet, F., Richard, F., Fatmi, A., Belenguer, P., and Royet, J. (2011). Inner-membrane proteins PMI/TMEM11 regulate mitochondrial morphogenesis independently of the DRP1/MFN fission/fusion pathways. *EMBO Rep* 12, 223–230. 10.1038/embor.2010.214.
38. Hanein, S., Garcia, M., Fares-Taie, L., Serre, V., De Keyzer, Y., Delaveau, T., Perrault, I., Delphin, N., Gerber, S., Schmitt, A., et al. (2013). TMEM126A is a mitochondrial located mRNA (MLR) protein of the mitochondrial inner membrane. *Biochim Biophys Acta* 1830, 3719–3733. 10.1016/j.bbagen.2013.02.025.

39. Park, A., Kim, K.E., Park, I., Lee, S.H., Park, K.Y., Jung, M., Li, X., Sleiman, M.B., Lee, S.J., Kim, D.S., et al. (2023). Mitochondrial matrix protein LETMD1 maintains thermogenic capacity of brown adipose tissue in male mice. *Nat Commun* 14, 3746. 10.1038/s41467-023-39106-z.
40. Ramelot, T.A., Yang, Y., Sahu, I.D., Lee, H.W., Xiao, R., Lorigan, G.A., Montelione, G.T., and Kennedy, M.A. (2013). NMR structure and MD simulations of the AAA protease intermembrane space domain indicates peripheral membrane localization within the hexaoligomer. *FEBS Lett* 587, 3522–3528. 10.1016/j.febslet.2013.09.009.
41. Lee, S.Y., Kang, M.G., Park, J.S., Lee, G., Ting, A.Y., and Rhee, H.W. (2016). APEX Fingerprinting Reveals the Subcellular Localization of Proteins of Interest. *Cell Rep* 15, 1837–1847. 10.1016/j.celrep.2016.04.064.
42. Wang, L., Hilander, T., Liu, X., Tsang, H.Y., Eriksson, O., Jackson, C.B., Varjosalo, M., and Zhao, H. (2023). GTPBP8 is required for mitoribosomal biogenesis and mitochondrial translation. *Cell Mol Life Sci* 80, 361. 10.1007/s00018-023-05014-0.
43. Potter, A., Hangan, A., Goffart, S., Huynen, M.A., Cabrera-Orefice, A., and Spelbrink, J.N. (2023). Uncharacterized protein C17orf80 - a novel interactor of human mitochondrial nucleoids. *J Cell Sci* 136. 10.1242/jcs.260822.
44. Mesecke, N., Terziyska, N., Kozany, C., Baumann, F., Neupert, W., Hell, K., and Herrmann, J.M. (2005). A disulfide relay system in the intermembrane space of mitochondria that mediates protein import. *Cell* 121, 1059–1069. 10.1016/j.cell.2005.04.011.
45. Willems, P., Huang, J., Messens, J., and Van Breusegem, F. (2023). Functionally annotating cysteine disulfides and metal binding sites in the plant kingdom using AlphaFold2 predicted structures. *Free Radic Biol Med* 194, 220–229. 10.1016/j.freeradbiomed.2022.12.001.
46. Stentz, R., Osborne, S., Horn, N., Li, A.W., Hautefort, I., Bongaerts, R., Rouyer, M., Bailey, P., Shears, S.B., Hemmings, A.M., et al. (2014). A bacterial homolog of a eukaryotic inositol phosphate signaling enzyme mediates cross-kingdom dialog in the mammalian gut. *Cell Rep* 6, 646–656. 10.1016/j.celrep.2014.01.021.
47. Angebault, C., Guichet, P.O., Talmat-Amar, Y., Charif, M., Gerber, S., Fares-Taie, L., Gueguen, N., Halloy, F., Moore, D., Amati-Bonneau, P., et al. (2015). Recessive Mutations in RTN4IP1 Cause Isolated and Syndromic Optic Neuropathies. *Am J Hum Genet* 97, 754–760. 10.1016/j.ajhg.2015.09.012.

48. Lulli, M., Nencioni, D., Papucci, L., and Schiavone, N. (2020). Zeta-crystallin: a moonlighting player in cancer. *Cell Mol Life Sci* 77, 965–976. 10.1007/s00018-019-03301-3.
49. Abramson, J., Adler, J., Dunger, J., Evans, R., Green, T., Pritzel, A., Ronneberger, O., Willmore, L., Ballard, A.J., Bambrick, J., et al. (2024). Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature* 630, 493–500. 10.1038/s41586-024-07487-w.
50. Shao, J., Fu, Z., Ji, Y., Guan, X., Guo, S., Ding, Z., Yang, X., Cong, Y., and Shen, Y. (2016). Leucine zipper-EF-hand containing transmembrane protein 1 (LETM1) forms a Ca<sup>2+</sup>/H<sup>+</sup> antiporter. *Sci Rep* 6, 34174. 10.1038/srep34174.
51. Zhang, J., Humphreys, I.R., Pei, J., Kim, J., Choi, C., Yuan, R., Durham, J., Liu, S., Choi, H.J., Baek, M., et al. (2025). Predicting protein-protein interactions in the human proteome. *Science* 390, eadt1630. 10.1126/science.adt1630.
52. Evans, R., O'Neill, M., Pritzel, A., Antropova, N., Senior, A., Green, T., Židek, A., Bates, R., Blackwell, S., Yim, J., et al. (2022). Protein complex prediction with AlphaFold-Multimer. *bioRxiv*, 2021.2010.2004.463034. 10.1101/2021.10.04.463034.
53. Sim, J., Kwon, S., and Seok, C. (2023). HProteome-BSite: predicted binding sites and ligands in human 3D proteome. *Nucleic Acids Res* 51, D403-D408. 10.1093/nar/gkac873.
54. Yang, J., Baek, M., and Seok, C. (2019). GalaxyDock3: Protein-ligand docking that considers the full ligand conformational flexibility. *J Comput Chem* 40, 2739–2748. 10.1002/jcc.26050.
55. Kwon, S., and Seok, C. (2023). CSAAlign and CSAAlign-Dock: Structure alignment of ligands considering full flexibility and application to protein-ligand docking. *Comput Struct Biotechnol J* 21, 1–10. 10.1016/j.csbj.2022.11.047.
56. Paysan-Lafosse, T., Blum, M., Chuguransky, S., Grego, T., Pinto, B.L., Salazar, G.A., Bileschi, M.L., Bork, P., Bridge, A., Colwell, L., et al. (2023). InterPro in 2022. *Nucleic Acids Res* 51, D483–D491. 10.1093/nar/gkac993.
57. Wang, J., Youkharibache, P., Zhang, D., Lanczycki, C.J., Geer, R.C., Madej, T., Phan, L., Ward, M., Lu, S., Marchler, G.H., et al. (2022). iCn3D, a web-based 3D viewer for sharing 1D/2D/3D representations of biomolecular structures. *Bioinformatics* 36, 131–135. 10.1093/bioinformatics/btz502.
58. Hwang, S.E., Lee, H., Kim, M., Chung, C.H., and Lee, S.V. (2025). Mitochondrial unfolded protein response triggered by ykt-6 knockdown in *Caenorhabditis elegans*. *Mol. Cells* 48, 100182.

59. Karczewski, K.J., Francioli, L.C., Tiao, G., Cummings, B.B., Alfoldi, J., Wang, Q., Collins, R.L., Laricchia, K.M., Ganna, A., Birnbaum, D.P., et al. (2020). The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature* 581, 434–443. 10.1038/s41586-020-2308-7.
60. Ioakeimidis, F., Ott, C., Kozjak-Pavlovic, V., Violitzi, F., Rinotas, V., Makrinou, E., Eliopoulos, E., Fasseas, C., Kollias, G., and Douni, E. (2014). A splicing mutation in the novel mitochondrial protein DNAJC11 causes motor neuron pathology associated with cristae disorganization, and lymphoid abnormalities in mice. *PLoS One* 9, e104237. 10.1371/journal.pone.0104237.
61. Michaelis, J.B., Brunstein, M.E., Bozkurt, S., Alves, L., Wegner, M., Kaulich, M., Pohl, C., and Munch, C. (2022). Protein import motor complex reacts to mitochondrial misfolding by reducing protein import and activating mitophagy. *Nat Commun* 13, 5164. 10.1038/s41467-022-32564-x.
62. An, J., Shi, J., He, Q., Lui, K., Liu, Y., Huang, Y., and Sheikh, M.S. (2012). CHCM1/CHCHD6, novel mitochondrial protein linked to regulation of mitofilin and mitochondrial cristae morphology. *J Biol Chem* 287, 7411–7426. 10.1074/jbc.M111.277103.
63. Tyanova, S., Temu, T., Sinitcyn, P., Carlson, A., Hein, M.Y., Geiger, T., Mann, M., and Cox, J. (2016). The Perseus computational platform for comprehensive analysis of (prote)omics data. *Nat Methods* 13, 731–740. 10.1038/nmeth.3901.
64. Lomize, A.L., and Pogozheva, I.D. (2022). Membranome 3.0: Database of single-pass membrane proteins with AlphaFold models. *Protein Sci* 31, e4318. 10.1002/pro.4318.
65. Palmieri, L., Pardo, B., Lasorsa, F.M., del Arco, A., Kobayashi, K., Iijima, M., Runswick, M.J., Walker, J.E., Saheki, T., Satrustegui, J., and Palmieri, F. (2001). Citrin and aralar1 are Ca<sup>2+</sup>-stimulated aspartate/glutamate transporters in mitochondria. *EMBO J* 20, 5060–5069. 10.1093/emboj/20.18.5060.
66. Fiermonte, G., De Leonardis, F., Todisco, S., Palmieri, L., Lasorsa, F.M., and Palmieri, F. (2004). Identification of the Mitochondrial ATP-Mg/Pi Transporter. *J Biol Chem* 279, 30722–30730. 10.1074/jbc.M400445200.
67. Landrum, M.J., Chitipiralla, S., Brown, G.R., Chen, C., Gu, B., Hart, J., Li, D., O'Daniel, J.M., Rubinstein, W.S., et al. (2020). ClinVar: improvements to accessing data. *Nucleic Acids Res* 48, D835–D844. 10.1093/nar/gkz972.

## FIGURE LEGENDS

### **Figure 1. Construction of MitoAtlas via Multiplexed SR-PL Mapping (see also Figure S1 and Table S1).**

(A) Overview of the MitoAtlas workflow. APEX2 generates biotin-phenoxy radicals that label tyrosine residues through contact-dependent labeling (radius ~10 nm), while BioID/BioID2 produces biotin-AMP esters that modify lysine residues (radius ~10 nm). Proximity-labeled peptides are enriched by streptavidin pulldown and identified via LC-MS/MS as desthiobiotin-phenol (DBP)-modified tyrosines or biotin-modified lysines (Super-Resolution Proximity Labeling, SR-PL). Site-specific MS intensity profiles across matrix, IMS, OMM, and non-mitochondrial baits are then classified by logistic regression (LR) to assign sub-mitochondrial localizations, enabling downstream analyses of membrane topologies (Figure 3), domain distributions (Figure 5), mitochorphan characterization (Figure 4), protein–protein/metabolite interactions (Figure 6), and the MitoAtlas web database (Figure 7). All data are accessible through the MitoAtlas web database ([mitoatlas.org](http://mitoatlas.org)). Representative examples show site-level classification for MRPL3 (Matrix-gold), UQCRCQ (IMM-TM), and MINPP1 (IMS-silver).

(B) Comparison of sub-mitochondrial annotation resolution between MitoCarta3.0 and MitoAtlas. MitoCarta3.0 assigns proteins to broad categories (e.g., “OMM,” “IMS,” “IMM,” “Matrix”) without topological detail, whereas MitoAtlas provides refined classifications including membrane-spanning orientation (e.g., OMM-TM) and confidence tiers (Gold/Silver), yielding more functionally informative annotations.

(C) Sub-mitochondrial distribution of the 42-bait library across the four mitochondrial compartments. APEX2 fusions (filled circles) were preferentially deployed to the matrix (9 baits) and IMS (8 baits) to leverage high labeling efficiency in luminal compartments, while BioID/BioID2 fusions (open circles) were primarily targeted to the OMM (11 baits) for confined, neighborhood-specific labeling. Soluble proteins are depicted as circles and transmembrane proteins as rectangles. Colored arrowheads indicate targeting sequences.

(D) Validation of bait localization and enzymatic activity. Confocal immunofluorescence microscopy images show colocalization of representative APEX2 (V5-tagged) and BioID/BioID2 (HA-tagged) baits with a mitochondrial marker, and streptavidin staining confirms biotinylation activity. Biotinylation pattern images show the spatial distribution of labeling for matrix-targeted (OXA1L), IMS-targeted (EMRE), OMM-targeted (EXD2), and non-mitochondrial (GBP, mCherry-

BioID) baits.

(E) UMAP analysis of biological replicates from APEX2 baits (left) and BioID baits (right). Each bait is represented by colored dots: matrix (red), IMS (blue), OMM (gold), and non-mitochondrial (purple). Shaded polygons connect biological replicates of the same bait.

(F) Venn diagram showing the overlap between APEX2- and BioID-identified proteins and MitoCarta3.0. APEX2 and BioID together identified 861 mitochondrial proteins, with 383 detected by both enzyme systems. Examples of uniquely identified proteins are shown (e.g., MINPP1 by APEX2 only; SMIM12 by BioID only).

(G) Distribution of the number of labeled sites per protein for APEX2 (DBP-modified tyrosines; average 2.39 sites per protein) and BioID (biotin-modified lysines; average 2.54 sites per protein).

**Figure 2. A Two-Step Classification Pipeline for Sub-mitochondrial Proteome Mapping (see also Figure S2 and Tables S2 and S3).**

(A) Schematic overview of the two-step classification pipeline. In Step 1 (Site Classification), 13,440 modified sites from APEX2 and BioID experiments are classified into four sub-mitochondrial compartments (Matrix, IMS, OMM, and Non-mitochondrial) by logistic regression (workflow detailed in Figure S2H). In Step 2 (Protein-level Integration and Classification), site-level assignments are aggregated per protein. Non-TM proteins are assigned gold or silver confidence tiers based on the number of supporting sites, while TM proteins undergo additional topology validation to resolve membrane orientation (IMM-TM, OMM-TM, MAM-TM).

(B) UMAP visualization of APEX2-labeled (top) and BioID-labeled (bottom) peptides classified by the site-level logistic regression model. Each point represents a single modified site, colored by its predicted sub-mitochondrial compartment (Matrix, red; IMS, blue; OMM, yellow; Non-mitochondrial, purple; Inconclusive, gray). Distinct clustering of compartment-specific sites demonstrates the discriminative power of the multi-bait labeling profiles.

(C) Sankey diagrams illustrating the flow of protein annotations across MitoCarta3.0, MitoAtlas, and UniProt. Flows highlight how MitoAtlas resolves previously ambiguous categories (e.g., “MIM,” “Membrane,” “Unknown”) from MitoCarta3.0 into precise sub-mitochondrial localizations with topological detail, and how these assignments compare with UniProt annotations. Notably, MitoAtlas classifies 160 proteins not annotated in MitoCarta3.0 and provides topology information for proteins listed only broadly as “Mitochondrion” in UniProt.

(D) Comparative spatial mapping of respiratory Complex I subunits. MitoCarta3.0 (left) assigns all 44 subunits broadly as “MIM” using non-SR-PL data. MitoAtlas (right) resolves 34 of 44 subunits with membrane topology information, distinguishing Matrix-facing, IMS-facing, and IMM-TM components consistent with high-resolution cryo-EM structural data.

**Figure 3. High-Resolution Membrane Topology and Domain Architecture of the Mitochondrial Proteome (see also Figure S3 and Tables S3 and S4).**

(A) Comparison of the number of IMM-TM and OMM-TM proteins with identified topologies across databases. MitoAtlas (left) resolves the membrane topologies of all 136 IMM-TM and 60 OMM-TM proteins detected. UniProt (middle) contains curated topological annotations for a subset of entries. MitoCarta3.0 (right) lacks topological orientation data (0/0). Other databases without topology information (Human Cell Map, MitoCoP) are also indicated.

(B) Venn diagrams illustrating the overlap and novel identification of IMM-TM (left) and OMM-TM (right) proteins in MitoAtlas relative to UniProt.

(C) Sankey plot depicting the consistency and discrepancies in membrane topology assignments between MitoAtlas and UniProt experimental records.

(D) Validation of IMM-TM topologies. (Left): Proposed membrane topologies of TMEM223 and TMEM256 at the IMM based on SR-PL data. (Right): Validation of C-terminal orientation via APEX2-enhanced electron microscopy (APEX-EM) imaging.

(E) Comparative analysis of SLC30A9 TM domain count and orientation as determined by MitoAtlas integrated with TMbed (left), MitoAtlas integrated with AlphaFold2 (AF2) structures (middle), and current UniProt annotations (right).

(F) Structural validation of mitochondrial carriers. Concordance between SR-PL-derived labeling sites and AF2-predicted TM domain positions for the mitochondrial carrier protein family (SLC25).

(G) Topology of multi-pass and  $\beta$ -barrel proteins. Proposed membrane topologies for 24 SLC25 family proteins and 5 OMM  $\beta$ -barrel proteins identified in MitoAtlas.

(H) AF2-predicted structure of DNAJC11 highlighting the 17-stranded  $\beta$ -barrel TM domain and corresponding SR-PL labeling sites.

(I) Schematic representation of the DNAJC11 domain organization and sub-mitochondrial localization based on MitoAtlas SR-PL data.

(J) Domain architecture and membrane topology of ATAD3B. (Left) Schematic of ATAD3B domain organization and sub-mitochondrial localization based on MitoAtlas SR-PL data, with the N-terminal domain facing the IMS and the C-terminal ATPase domain in the matrix. (Right) Membranome-adjusted AlphaFold2 structure<sup>64</sup> with SR-PL site annotations mapped onto the 3D model, confirming the IMM-TM topology.

**Figure 4. Expanding the Mitochondrial Map: Characterization of 160 Mito-Orphan Proteins and MINPP1 as an IMS Protein (see also Figure S4 and Table S5).**

(A) Venn diagram comparing protein coverage between MitoCarta3.0 and MitoAtlas. MitoAtlas identifies 160 mitochondrial proteins not annotated in MitoCarta3.0, termed “Mito-Orphans.”

(B) Sub-mitochondrial distribution of the 160 mito-orphans across compartments: matrix (Gold and Silver), IMS (Gold and Silver), OMM-TM, IMM-TM, MAM-TM, mito-TM, and OMM/MAM-TM categories, with representative examples.

(C) SR-PL site maps for two validated matrix mito-orphans, GTPBP8 and C17orf80, showing the distribution of labeled residues and their site-level classifications. Sites are colored by assigned compartment (matrix in red; inconclusive in gray).

(D) Confocal immunofluorescence microscopy validating mitochondrial targeting of V5-APEX2-tagged GTPBP8 and C17orf80, co-stained with Mitotracker and streptavidin.

(E) Transmission electron microscopy (TEM) of APEX2-mediated DAB polymerization for GTPBP8-APEX2 and C17orf80-APEX2, confirming matrix localization of the C-terminus. Control (non-transfected) cells are shown for comparison.

(F) Proposed membrane topologies for SMIM8 and SMIM12 at the IMM, based on SR-PL site assignments. MitoCarta3.0 previously classified SMIM8 as “MOM” (OMM), while SMIM12 was not annotated. Both are reclassified as IMM-TM proteins with N-termini facing the matrix. For SMIM12, site-level P1 classification of four BioID-labeled lysine residues is shown: two sites are assigned to IMS and two to OMM. This mixed assignment pattern is consistent with an IMM-TM topology, as the short IMS-exposed loop of small single-pass TM proteins lies in close proximity to OMM-anchored BioID baits, falling within their ~10 nm labeling radius.

(G) Confocal microscopy validation of SMIM8 and SMIM12 localization using V5-ALFAx2-tagged constructs expressed in HEK293T-REx cells, detected with NbALFA-mScarlet and co-stained with anti-TOM20.

(H) Proteinase K protection assay confirming IMM-TM topology of SMIM8 (left, V5-ALFA-SMIM8) and SMIM12 (right, SMIM12-V5-ALFAx2) using isolated mitochondria from HEK293T-REx cells, with the same assay layout as in (L).

(I) SR-PL site maps for MINPP1 and FKBP9, two IMS mito-orphans. Each displays a single IMS-classified tyrosine site, with domain architecture shown (histidine phosphatase superfamily for MINPP1; FKBP-type PPlase and EF-hand domains for FKBP9).

(J) AlphaFold2/3-predicted structures of MINPP1 and FKBP9, highlighting predicted disulfide bonds consistent with IMS residence where the MIA40/CHCHD4 disulfide relay system operates. Structural comparison with the bacterial homolog BtMinpp (PDB: 4FDU), which lacks disulfide bonds, is shown.

(K) Construct design for MINPP1-V5-ALFAx2-SDEL (retaining the ER retention signal) and confocal microscopy in COS-7 cells showing dual localization of MINPP1 to both the ER (mCherry-Sec61b) and mitochondria (anti-TOM20), establishing MINPP1 as a dual-localizing ER lumen/IMS protein.

(L) Proteinase K protection assay using isolated mitochondria from MINPP1-V5-ALFAx2-SDEL-expressing HEK293T-REx cells. Digitonin titration shows progressive loss of the MINPP1 signal, consistent with IMS localization. EXD2 (OMM) and HSP60 (matrix) serve as controls.

(M) mmPEG24-maleimide modification assay confirming the presence of disulfide bonds in MINPP1. Treatment with the reducing agent TCEP prior to mmPEG24 labeling reveals a mobility shift, demonstrating that MINPP1 contains reducible disulfide bonds consistent with proper folding in the IMS environment. Detection by both immunofluorescence (IF: EGFP-NbALFA) and immunoblotting (IB: anti-V5).

(N) Proposed model for MINPP1 function. In the ER lumen, MINPP1 cannot access its substrate InsP6. At ER-mitochondria contact sites, IMS-localized MINPP1 can hydrolyze InsP6 to generate IP3, which may promote calcium release through IP3R and subsequent mitochondrial calcium uptake via VDAC and MCU.

(O) Representative images of *C. elegans* body morphology following RNAi knockdown of *ykt-6*, a mito-orphan identified as a novel OMM protein.

(P) Quantification of body-bending frequency in *C. elegans* upon *ykt-6* RNAi knockdown, showing a significant reduction in motility compared to control RNAi (\*\*\*\*p < 0.0001), confirming the functional importance of YKT6 at the OMM.

**Figure 5. Spatial Partitioning of Functional Domains and Homologous Superfamilies Across Mitochondrial Sub-compartments. (See also Figure S5 and Table S6)**

(A) Comparative analysis of sub-mitochondrial enrichment for protein domains (left) and homologous superfamilies (right) based on InterPro<sup>56</sup> classifications

(B) Sub-mitochondrial distribution of proteins containing the NAD(P)H-binding domain superfamily, highlighting a predominant matrix localization.

(C) Schematic representation of SR-PL sites and domain architectures for HSDL1 and HSDL2, illustrating the non-canonical OMM localization of HSDL1.

(D) Schematic representation of SR-PL sites and domain architectures for RTN4IP1 and CRYZ.

(E) Sub-mitochondrial distribution of the EF-hand domain pair superfamily, demonstrating a preference for the intermembrane space.

(F) Schematic representation of SR-PL sites and domain architectures for the calcium-binding transporters SLC25A1, SLC25A13<sup>65</sup>, and SLC25A24<sup>66</sup>.

(G) Schematic representation of SR-PL sites and domain architecture of LETM1.

(H) Sub-mitochondrial distribution of the Tetratricopeptide-like (TPR) helical domain superfamily across mitochondrial compartments.

(I) Labeled sites and domain architectures for RMDN1, RMDN2, and RMDN3, highlighting the atypical IMS localization of RMDN1 compared to its paralogs.

**Figure 6. Prediction of sub-mitochondrial protein-protein and protein-metabolite interactions (see also Figures S6 and S7 and Table S7).**

(A) Workflow to find high pTM + ipTM scoring bait-prey models by filtering specific bait-prey sites.

(B) Histogram of pTM + ipTM scores of AlphaFold models generated from bait-prey pairs with high specificity and intensity (left), and refined structures of score over 0.4 (right).

(C) Quantile normalized intensity and specificity of ss-APEX2-LETM1 labeled sites.

(D) AF3-Multimer prediction of trimer complex of one LETM1 and two CS with PAE values of residues near 4 angstroms between LETM1 and CS#1 or CS#2 highlighted (left). The LETM1 APEX2 specific site CS-Y185 is also depicted in red. PAE matrix is in the right.

(E) Alternative view of AF3-Multimer prediction of LETM1-CSx2 trimer. Two residues at the

LETM1-CS interfaces with known pathogenic consequences of mutation, R380 and R393, are highlighted.

(F) Co-immunoprecipitation (CoIP) validation of the LETM1–CS interaction. MTS-3×FLAG-LETM1 (WT or indicated mutants) was transiently expressed in HEK293T-REx cells and immunoprecipitated with anti-FLAG beads; immunoblots were probed with anti-FLAG and anti-CS antibodies. Three classes of LETM1 variants were tested: (i) six pathogenic missense variants (I293N, R294Q, P300S, D358N, R380P, R393H) from bi-allelic patients with mitochondrial disease<sup>7</sup>; (ii) two structure-guided charge-reversal mutations (R380E, R385E) designed to disrupt the predicted LETM1–CS interaction interface; and (iii) E222Q, corresponding to mouse E221Q, which abolishes LETM1-mediated Ca<sup>2+</sup>/H<sup>+</sup> antiport<sup>50</sup>.

(G) Workflow used to find novel protein-metabolite interactions using MitoAtlas and GalaxySite.

(H) Sequence identity and TM-score of predicted protein-metabolite interaction structures.

(I) The sub-mitochondrial localizations of already known heme binding proteins FECH (matrix), PTGES2 (IMM-TM) and the newly discovered MTX1 (OMM).

(J) Predicted complex structure of MTX1 and haem (HEM) modeled using GalaxySite docking with template restrictions (top) and AlphaFold3 (bottom). Tryptophan residues were colored orange in the AlphaFold3 model.

(K) Absorbance spectrum of 10 μM MBP-MTX1 protein treated with heme compared to heme only (top) and Fluorescence spectrum of 100 nM MBP-MTX1 protein only, and 50 nM and 100 nM heme treated 100 nM MBP-MTX1 protein (bottom).

**Figure 7. MitoAtlas.org: an interactive web platform for the mitochondrial spatial proteome (see also Figure S7).**

(A) Overview of MitoAtlas.org interface components, shown using ATAD3B as a representative example. The platform comprises six integrated modules: (i) Protein overview panel, displaying gene name, protein name, synonyms, UniProt ID, database cross-references (MitoAtlas, UniProt/PubMed sources, MitoCarta), SR-PL labeled sites identified by APEX2 and BioID, sub-mitochondrial localizations of the protein and individual sites, and transmembrane domain annotations; (ii) Linear domain architecture map, showing SR-PL labeled tyrosine and lysine sites along the full protein sequence with InterPro domain annotations<sup>56</sup> and compartment assignments (matrix, IMS, OMM, or non-mitochondrial); (iii) Interactive membrane topology viewer, displaying SR-PL labeled sites across the membrane bilayer with color-coded compartment assignments

and topology scores derived from logistic regression classification; (iv) Sequence feature viewer, a multi-track browser integrating domain boundaries, labeled sites, pLDDT confidence scores, and biochemical properties in a single linear view; (v) SR-PL intensity heatmap, showing bait-resolved APEX2 and BioID intensity profiles across all proximity labeling baits with logistic regression classification confidence for each site; and (vi) Projection onto PDB/AlphaFold2 3D structures, where TM domain and SR-PL site annotations are mapped onto experimentally determined or predicted structures via the interactive iCn3D viewer<sup>57</sup>.

(B) Application scenarios demonstrating the utility of MitoAtlas for biological discovery (see Figure S7 for detailed examples). Three representative use cases illustrate how site-resolved spatial data enhances mitochondrial protein analysis: (i) Disease variant interpretation: membrane topology of SLC25A4 (ANT1) with ClinVar pathogenic<sup>67</sup> variant mapping. Nine pathogenic variant positions are mapped onto the six-TM-helix topology and color/shape-coded by disease category, revealing a topology-dependent disease pattern: IMS-facing variants associate with progressive external ophthalmoplegia (adPEO), whereas matrix-facing variants cause cardiomyopathic mtDNA depletion syndrome. (ii) Post-translational modification mapping: PTM topology of SLC25A4 (ANT1) showing compartment-specific PTM distribution. A single phosphoserine (pos 7) localizes to the IMS-facing N-terminal segment, while three succinyllysine sites (pos 147, 245, 272) are exclusively on matrix-facing loops, consistent with IMS kinase accessibility and matrix succinyl-CoA-driven non-enzymatic acylation. (iii) Population variant density: gnomAD missense variant density<sup>59</sup> across MitoAtlas-defined topological faces reveals strong purifying selection on TM domains of IMM proteins and reduced constraint on N-terminal mitochondrial targeting sequences.

## **SUPPLEMENTAL INFORMATION**

Figure S1. Validation of APEX2 and BioID/BioID2 stably expressing cells, related to Figure 1.

Figure S2. Site and Protein Classification Model Details, related to Figure 2.

Figure S3. Validation and Structural Refinement of Mitochondrial Membrane Topologies, related to Figure 3.

Figure S4. Structural Validation and Functional Analysis of Mito-Orphan Proteins, related to Figure 4.

Figure S5. Spatial Enrichment and Membrane Topology of NAD(P)-binding and Tetratricopeptide Helical Domain Superfamilies, related to Figure 5.

Figure S6. CoIP optimization, specific bait–prey site analysis, and topology validation of predicted interactome, related to Figure 6.

Figure S7. Application Scenarios: Enhanced Biological Discovery with MitoAtlas Site-Resolved Spatial Data, related to Figure 7.

## SUPPLEMENTAL VIDEOS AND SPREADSHEETS

Table S1. LC-MS/MS detection of biotin or DBP labeled SR-PL sites, related to Figure 1.

Table S2. Classification of SR-PL sites into sub-mitochondrial spaces, related to Figure 2.

Table S3. Classification of proteins into sub-mitochondrial spaces, related to Figures 2 and 3.

Table S4. Membrane topology information from SR-PL, related to Figure 3.

Table S5. Further study on mito-orphans, related to Figure 4.

Table S6. Interpro domain annotations of MitoAtlas proteins, related to Figure 5.

Table S7. Protein-protein interactions and protein-metabolite interactions of MitoAtlas proteins, related to Figure 6.

## SUPPLEMENTAL FIGURE LEGENDS

**Figure S1. Validation of APEX2 and BioID/BioID2 stably expressing cells (related to Figure 1).**

(A) Architecture of all APEX2- and BioID/BioID2-fused bait constructs used in this study. APEX2 baits are tagged with V5; BioID/BioID2 baits are tagged with HA. Inclusion of mCherry as a linker in specific constructs was utilized to enhance protein solubility or adjust bait size. Bait proteins are color-coded by their targeted sub-mitochondrial compartment: matrix (red), IMS (blue), OMM (yellow), and non-mitochondrial (purple).

(B) Confocal immunofluorescence microscopy of all HEK293T-REx stable cell lines. APEX2 baits (left, anti-V5) and BioID/BioID2 baits (right, anti-HA) were co-stained with a mitochondrial marker and streptavidin to confirm sub-organellar targeting and biotinylation activity. Pearson's correlation coefficients for colocalization with the mitochondrial marker are shown. Merged images display the overlay of the POI signal and the mitochondrial marker.

(C) Transmission electron microscopy (TEM) images of APEX2-mediated 3,3'-diaminobenzidine (DAB) polymerization for representative matrix- and IMS-targeted baits, providing ultrastructural validation of bait positioning.

(D) Western blot analysis using Streptavidin-HRP (top rows) and anti-V5 or anti-HA antibodies (bottom rows) to confirm expression levels and total protein biotinylation across the stable cell lines.

(E) Venn diagrams illustrating the overlap and unique coverage of modified peptides identified by multiple baits within the same sub-compartment. Top: three matrix APEX2 baits (Matrix-APEX2, MRM1-APEX2, HSPA9-APEX2) show extensive non-overlapping peptide sets, demonstrating the value of multiplexed baits for expanding proteome depth. Bottom left: three OMM-targeted APEX2 baits (TDRKH-APEX2, TOM20-APEX2, EXD2-APEX2). Bottom right: three OMM-targeted BioID baits (TDRKH-BioID, TOM20-BioID, EXD2-BioID). Comparison of the bottom panels reveals that BioID generates a higher proportion of non-overlapping peptides relative to APEX2, consistent with its shorter labeling radius capturing more spatially restricted protein neighborhoods.

(F) Sequence logos highlighting the local amino acid preferences surrounding APEX2-labeled tyrosine residues (left) and BioID/BioID2-labeled lysine residues (right), compared with previously reported N-hydroxysuccinimide (NHS)-ester-modified lysine sites (bottom). APEX2-modified sites show enrichment of glycine at  $\pm 1$  positions, while BioID sites favor branched amino acids (Val, Leu, Ile) at the +1 position. Conversely, radical-quenching amino acids (Cys, Trp, Met, Arg, Phe, Leu, Tyr, His) are negatively enriched around APEX2 sites, with cysteine showing the strongest depletion, consistent with its favorable one-electron reduction potential for phenoxy radical scavenging.

(G) Receiver Operating Characteristic (ROC) curves evaluating the correlation between solvent-accessible surface area (SASA) and peptide detection using AlphaFold2 structures. Top: APEX2 site prediction at varying probe radii, showing an optimal probe radius near that of water ( $\sim 1.2$  Å), suggesting APEX2 can probe deep into small protein clefts. Bottom: BioID site prediction, exhibiting a much larger optimal radius ( $\sim 11.5$  Å), consistent with the bulkiness of the biotin-AMP molecule restricting labeling to more exposed surface residues.

(H) Mechanistic model illustrating the differences in reactive probe size between APEX2 and BioID. The small desthiobiotin-phenol (DBP) radical generated by APEX2 ( $\sim 1.2$  Å probe radius) can access sterically restricted tyrosine residues, whereas the larger biotin-AMP intermediate produced by BioID ( $\sim 11.5$  Å probe radius) is limited to more solvent-exposed lysine residues.

### **Figure S2. Site and Protein Classification Model Details (Related to Figure 2).**

(A) Comparative MS intensity heatmaps of representative peptides containing DBP-modified tyrosine (top, APEX2 baits) and biotin-modified lysine (bottom, BioID baits) residues. Reference proteins include MRPL22 (Matrix), MICU2 (IMS), FKBP8 (OMM), and ACTB (Non-mitochondrial), demonstrating the compartment-specific enrichment patterns used as training features for the logistic regression model.

(B) Comparison of 10-fold cross-validation accuracy for six supervised learning algorithms (CART, KNN, LDA, LR, NB, and SVM) trained separately on APEX2 (left) and BioID (right) datasets. Logistic Regression (LR) was selected for its high accuracy, interpretability, and tunable probability thresholds.

(C) True Positive Rate (TPR) and False Discovery Rate (FDR) curves across a range of P1 probability thresholds for APEX2 (top) and BioID (bottom), evaluated separately for Matrix, IMS, OMM, and Non-mitochondrial compartments. Optimal thresholds were selected to maintain an overall FDR below 10%.

(D) Visualization of the trained site-level logistic regression coefficient weights for each APEX2 bait (left) and BioID/BioID2 bait (right). Rows represent baits; columns represent the four sub-mitochondrial compartment predictions. Positive weights (yellow) indicate baits that serve as strong predictors for a specific localization.

(E) Comparison of binary enrichment analysis and MitoAtlas multi-bait classification. Top: APEX2. Left: Volcano plot of log<sub>2</sub> fold change (TOM20-APEX2 vs. APEX2-NES) versus  $-\log_{10}(\text{p-value})$  for training-set biotinylation sites. OMM training sites (gold) cluster at positive fold changes, while non-mitochondrial training sites (dark gray) cluster near zero or negative values. Center: MitoAtlas classification probabilities for the same sites, plotted as p(non-mito) versus p(OMM), showing clear separation. Right: ROC curves comparing binary fold change (teal; AUC = 0.866) and MitoAtlas p(OMM) (red; AUC = 0.923) for discriminating 107 OMM from 269 non-mitochondrial training sites. Bottom: BioID. Left: Volcano plot of log<sub>2</sub> fold change (TOM20-BioID vs. mCherry-BioID) versus  $-\log_{10}(\text{p-value})$ . Center: MitoAtlas probability scatter. Right: ROC curves comparing binary fold change (teal; AUC = 0.943) and MitoAtlas p(OMM) (red; AUC = 0.976) for discriminating 216 OMM from 195 non-mitochondrial training sites. Both comparisons demonstrate that the multi-bait MitoAtlas classifier consistently outperforms conventional binary enrichment analysis, particularly for the APEX2 dataset where the dual-bait approach shows substantial overlap between OMM and non-mitochondrial sites.

(F) Visualization of the trained protein-level (P2) logistic regression coefficients used to aggregate site-level P1 outputs into protein-level localization probabilities. The heatmap shows the relative contribution of APEX2 and BioID site classifications for each compartment assignment.

(G) Protein-level classification performance. True Positive Rate (TPR) and False Discovery Rate (FDR) curves across a range of P2 probability thresholds, evaluated separately for Matrix, IMS, OMM, and Non-mitochondrial compartments, demonstrating the accuracy of the second-step

protein-level classifier.

(H) Schematic overview of the two-step classification pipeline. Step 1 (Site Classification): raw MS intensities from 25 APEX2 and 17 BioID baits are log-transformed and iBAQ-calibrated, then classified by one-vs-rest (OvR) logistic regression into four sub-mitochondrial compartments (Matrix, IMS, OMM, and Non-mitochondrial), yielding a site-level probability ( $P_1$ ) for each of the 13,440 labeled sites. Step 2 (Protein Classification): proteins are first screened for transmembrane domains using four independent predictors (UniProt TM, TMbed, DeepTMHMM, AF2-TMDET). The 233 TM proteins are routed to a topology scoring pipeline that integrates site-level classifications with predicted TM domain boundaries to assign membrane orientation (IMM-TM, OMM-TM, MAM-TM). The remaining 3,825 non-TM proteins are classified by aggregating site-level  $P_1$  outputs into an 8-dimensional protein-level feature vector and applying a second OvR LR model, with results assigned to tiered confidence categories (Gold/Silver) based on  $P_2$  probability and the number of supporting sites.

(I) Worked examples of the classification pipeline. Left: non-TM protein MRPL45 (Matrix-gold). Raw MS intensities across all baits for a representative tyrosine site (Y46), followed by log-transformation and iBAQ calibration. OvR logistic regression assigns  $P_1$  probabilities to all seven MRPL45 sites (6 Tyr + 1 Lys), with all sites classified as Matrix (one site, K126, shown in yellow, falls below the confidence threshold). Aggregation of site-level  $P_1$  values into a protein-level  $P_2$  probability yields a high-confidence Matrix-gold assignment. Right: TM protein ATAD3B (IMM-TM). Raw MS intensities and log-transformed/iBAQ-calibrated values for a representative lysine site (K104) across all BioID baits.  $P_1$  site-level classification of all 16 ATAD3B sites (7 Tyr + 9 Lys) reveals a mixture of Matrix and IMS assignments, characteristic of an IMM-spanning protein. Topology scoring integrates site localizations with DeepTMHMM transmembrane domain predictions and AlphaFold structural models to determine the membrane orientation, assigning ATAD3B as IMM-TM with its N-terminus facing the matrix.

(J) Metabolic pathway map including glycolysis, the TCA cycle, and the electron transport chain (OXPHOS), with enzymes and complexes color-coded by their MitoAtlas sub-mitochondrial localization assignments. Numbers beneath OXPHOS complexes indicate the fraction of subunits detected by MitoAtlas.

(K) Characterization of inconclusive sites. Sites that failed to meet the multi-bait consensus threshold for any compartment ( $p(\text{matrix}) < 0.5$ ,  $p(\text{IMS}) < 0.5$ ,  $p(\text{OMM}) < 0.5$ , and  $p(\text{non-mito}) < 0.7$ ) were designated inconclusive. (i) Stacked bar chart showing the number of inconclusive sites

by method: APEX2 (364 of 6,046 sites, 6.0%) and BioID (1,323 of 7,394 sites, 17.9%), split by subtype. Mito-proximal inconclusive sites (gold) were detected by at least one mitochondrial bait but lacked multi-bait consensus; non-mito background sites (gray) were detected exclusively by cytosolic control baits. For APEX2, 96% of inconclusive sites (350 of 364) are mito-proximal; for BioID, 90% (1,185 of 1,323), indicating that the vast majority represent biologically relevant mitochondria-proximal labeling events rather than cytosolic background. (ii) Mirrored histogram of the number of baits detecting each inconclusive site (APEX2 upward, BioID downward). For BioID, 37% of inconclusive sites (492 of 1,323) were detected by only a single bait, consistent with contact-dependent bait-prey interactions rather than compartment-resident labeling. (iii) Box plots comparing the number of detecting baits for confidently classified OMM sites versus inconclusive sites, faceted by method. OMM-classified sites are detected by significantly more baits (BioID: median 5 vs. 2; APEX2: median 6 vs. 4), confirming that multi-bait consensus underlies confident spatial assignment.

**Figure S3. Validation and Structural Refinement of Mitochondrial Membrane Topologies (related to Figure 3).**

(A) Revised 6-TM architecture of SFXN2. Comparison of the transmembrane (TM) domain count and topological orientation of the IMM protein SFXN2. The MitoAtlas data, integrated with the TMbed model (left) and AlphaFold2 (AF2) structure (middle), supports a 6-TM model, contradicting the 5-TM model currently annotated in UniProt (right).

(B) Reclassification and topology of TMEM11. Comparison of TM domain predictions for TMEM11. MitoAtlas and AF2 evidence (left, middle) demonstrate that TMEM11 is an OMM protein with a 3-TM architecture, challenging the UniProt annotation of a 2-TM IMM protein (right).

(C) Orientation of the TMEM126A N-terminus. Topological analysis of the IMM protein TMEM126A. While UniProt (right) suggests a matrix-facing N-terminus based on historical protease protection assays, the integration of MitoAtlas SR-PL data with TMbed and AF2 (left, middle) indicates an intermembrane space (IMS) orientation.

(D) Structural modeling of the MTX1-DNAJC11-SAMM50 complex. AlphaFold3 (AF3) predicted model of the OMM complex (left), with SR-PL labeling sites projected onto the structure (middle). The Predicted Alignment Error (PAE) matrix (right) indicates the relative orientations of the subunits.

(E) Pore analysis of the DNAJC11  $\beta$ -barrel. Visualization of the internal pore of DNAJC11 (green)

generated via MOLE 2.5 software from side (left) and bottom (middle) perspectives. The corresponding pore radius profile is shown on the right.

(F) Reclassification of UniProt-predicted TM proteins to the matrix. AF2 structures of proteins with legacy TM annotations in UniProt that were reclassified as matrix-resident proteins in MitoAtlas. Structures are colored by pLDDT confidence scores, with UniProt-predicted TM domains highlighted in green to illustrate their lack of characteristic transmembrane structural features.

(G) AlphaFold3-Multimer predicted structure of the SMIM20–COX14–MT–CO1 complex, rendered in ChimeraX (ipTM = 0.69, pTM = 0.83).

**Figure S4. Structural Validation and Functional Analysis of Mito-Orphan Proteins (related to Figure 4).**

(A) AlphaFold3-Multimer predicted structures for SMIM8–AGK and SMIM12–PHB1 complexes, supporting the IMM-TM localization of SMIM8 and SMIM12. Predicted Aligned Error (PAE) matrices and interface confidence scores (ipTM, pTM) are shown for each complex.

(B) Structural alignment of MINPP1 (residues 41–478) with InsP6 substrate, showing the predicted binding site. The distance matrix (bottom) and PAE plot indicate high confidence in the core catalytic domain structure.

(C) Structural superposition of MINPP1 AlphaFold3 model (human, IMS) with the bacterial homolog BtMinpp (B. thetaiotaomicron, PDB: 4FDU), showing a pruned RMSD of 1.08 Å over 175 aligned residues, highlighting the conserved histidine phosphatase fold despite divergent disulfide bond content.

(D) Active site comparison between human MINPP1 (top) and bacterial BtMinpp (bottom), with key catalytic residues labeled. The conserved catalytic architecture supports enzymatic activity of MINPP1 in the IMS.

(E) Table of mito-orphan orthologs in *C. elegans* showing mitochondria-related phenotypes upon RNAi knockdown (top), compared with known mitochondrial genes exhibiting similar phenotypes (bottom). Phenotypes include reduced brood size, altered lifespan, and slow growth, supporting the functional relevance of the newly identified mito-orphans.

**Figure S5. Spatial Enrichment and Membrane Topology of NAD(P)-binding and Tetratricopeptide Helical Domain Superfamilies. (related to Figure 5)**

(A) Sub-mitochondrial localization profiles for all members of the NAD(P)-binding superfamily,

indicating a predominant matrix enrichment.

(B) Refined topological model of the inner mitochondrial membrane (IMM) protein Nicotinamide Nucleotide Transhydrogenase (NNT), illustrating the orientation of its NAD(P)-binding domains relative to the matrix and IMS.

(C) AlphaFold2-based structural alignments of representative protein pairs RTN4IP1-CRYZ and HSDL1-HSDL2 using ChimeraX matchmaker, demonstrating conserved folding patterns despite divergent primary sequences. Sequence homology was calculated using LALIGN.

(D) Sub-mitochondrial distribution of the Tetratricopeptide-like (TPR) helical domain superfamily.

(E) Membrane topology models for OMM-resident proteins FIS1, FKBP8, TOMM70, RMDN2, and RMDN3, depicting the cytosolic exposure of their respective tetratricopeptide repeat domains.

**Figure S6. CoIP optimization, specific bait–prey site analysis, and topology validation of predicted interactome (related to Figure 6).**

(A) Optimization of co-immunoprecipitation (CoIP) lysis conditions for the LETM1–CS interaction. Six buffer conditions containing 1% NP-40 were tested: PBS at pH 7.4 and 6.8 (left), and 50 mM HEPES at pH 8.0, 7.5, 7.0, and 6.5 (right). Four western blots are shown (top to bottom): anti-FLAG input (overexpressed MTS-3×FLAG-LETM1), anti-FLAG elution, anti-CS input (endogenous), and anti-CS elution. PBS pH 7.4 with 1% NP-40 was selected as the optimal condition for the main figure CoIP experiment (Figure 6F).

(B) Quantile normalized intensity and specificity of EXD2-APEX2 labeled sites.

(C) AF2-Multimer prediction of the EXD2 N-terminus-domain (NTD) and CNN3 complex and EXD2 C-terminus-domain (CTD) and ACTB complex.

(D) Network graph representation of all EXD2-APEX2 specific bait–prey sites identified by the intensity and specificity filtering pipeline.

(E) Network graph representation of all MitoAtlas baits connected to their specific bait–prey sites, showing the complete landscape of high-confidence proximity interactions across the 42-bait library.

(F) Validation of computationally predicted mitochondrial protein–protein interactions using MitoAtlas sub-mitochondrial topology. (i) Coverage: grouped bar chart showing the number of predicted PPIs (from 29,257 total<sup>27</sup>) assessable by MitoAtlas (1,104) versus MitoCarta3.0 (1,507).

(ii) Compartment co-occurrence heatmap of MitoAtlas-assessable interactions across sub-mitochondrial compartment pairs (Matrix, IMM-TM, IMS, OMM-TM, OMM, MAM). Green borders indicate topologically compatible pairs; red borders indicate incompatible pairs in non-adjacent compartments. (iii) Compartment co-occurrence heatmap using MitoCarta3.0's coarser four-compartment scheme (Matrix, MIM, IMS, MOM). (iv) Topology compatibility rate: stacked bar chart comparing the fraction of assessed PPIs that are topologically compatible versus incompatible for MitoAtlas (7.0% incompatible, 77 pairs) versus MitoCarta3.0 (1.7% incompatible, 26 pairs). MitoAtlas's finer granularity, particularly the distinction between OMM-TM and peripheral OMM proteins, enables more stringent spatial filtering. (v) Box plots comparing computational prediction confidence scores (best of AFprob, CFprob, AFMprob) between topologically compatible and incompatible interactions. Incompatible interactions have significantly lower prediction scores (median 0.831 vs. 0.946; Wilcoxon  $p = 6.27 \times 10^{-7}$ ), providing reciprocal validation of both MitoAtlas topology and the computational predictions. (vi) Representative table of PPIs that pass MitoCarta3.0's coarser topology check but are flagged as incompatible by MitoAtlas, demonstrating the added value of domain-level sub-mitochondrial resolution for interactome filtering.

(G) Coomassie staining result of immobilized metal affinity chromatography of recombinant protein His10-MBP-TEVs-MTX1<sup>152-393</sup> expressed in *E. coli*.

**Figure S7. Application Scenarios: Enhanced Biological Discovery with MitoAtlas Site-Resolved Spatial Data (related to Figure 7).**

(A) Disease variant interpretation: additional examples of topology-dependent disease patterns. Top: COX15 (UniProt Q7KZN9), a heme A farnesyltransferase with eight TM helices (N-terminus in matrix). Nine ClinVar pathogenic variants are mapped. All are associated with fatal infantile cardioencephalomyopathy due to cytochrome c oxidase deficiency type 2 (MIM #615119); positions 102 and 199 are additionally linked to Leigh syndrome. Variants are distributed across TM helices (pos 70, 151, 199, 229, 239, 280) and IMS-facing loops (pos 102, 107, 217) with no clear topology-dependent segregation, consistent with a uniformly essential role of COX15 in heme A biosynthesis. Middle: SLC25A22 (GC1; UniProt Q9H936), an IMM carrier with six TM helices (N-terminus in IMS). Eleven variants are mapped, all associated with early-onset epileptic encephalopathy (DEE type 3, MIM #609304; early myoclonic encephalopathy). Variants span IMS (pos 91, 220), TM helices (pos 110, 206, 236), and matrix-facing loops (pos 132, 140, 147, 151, 254, 273). Matrix-facing variants (6/11) tend to present as early myoclonic encephalopathy, while

IMS/TM variants are more frequently annotated as DEE type 3. Bottom: YME1L1 (i-AAA protease; UniProt Q96TA2), a single-pass IMM protein with a large IMS-facing AAA+ protease domain and matrix-facing N-terminal domain. Two pathogenic variants show clear topology-dependent disease segregation: position 149 (matrix-facing) causes optic atrophy 11 (MIM #617302), while position 667 (IMS-facing catalytic domain) causes 3-methylglutaconic aciduria (MIM #617248).

(B) PTM density and cysteine content across mitochondrial compartments and membrane faces. PTM site density and amino acid composition were analyzed across five compartments: matrix (non-TM, n=489), IMM-TM (n=134), IMS (non-TM, n=64), OMM-TM (n=57), and OMM (non-TM, n=73). For TM proteins, residues and PTM sites were assigned to membrane faces based on MitoAtlas topology. PTM annotations from UniProt. Dots, mean; error bars, SEM. Gray lines connect three faces within each TM group. Brackets, one-way ANOVA (\*p<0.05, \*\*p<0.01, \*\*\*p<0.001). First panel, phosphorylation (sites/1,000 S/T/Y): TM domains show near-zero phosphorylation in both IMM (0.2/1k) and OMM (0.5/1k). Within IMM-TM, IMS-facing loops of IMM-TM proteins have the highest density (20.5/1k vs. matrix 10.8/1k, TM 0.2/1k; p=0.031). OMM-TM cytosol-facing surfaces are most modified (38.6/1k; p=0.004) across all faces. Second panel, acetylation (N6-acetylysine/1,000 K): IMM-TM matrix-facing loops are enriched (63.3/1k vs. TM 7.9/1k, IMS 11.5/1k; ANOVA \*\*\*), consistent with matrix acetyl-CoA availability and SIRT3 deacetylation. OMM-TM shows minimal face-dependent variation (n.s.). Third panel, ubiquitination (sites/1,000 K): absent from IMM-TM faces (0-0.2/1k; n.s.). OMM-TM cytosol-facing domains show elevated ubiquitination (101.5/1k vs. TM 28.0/1k, IMS 19.4/1k; p=0.049), reflecting cytosolic ubiquitin-proteasome access. Fourth panel, cysteine content (%): IMM-TM faces differ significantly (ANOVA p=1.4×10<sup>-5</sup>); matrix faces 1.07%, TM 1.67%, IMS faces 0.65% (matrix vs. IMS Welch p=0.032). OMM-TM faces also differ (p=0.040); cytosol-facing 1.85%, TM 1.22%, IMS 0.79%. Soluble IMS proteins are cysteine-enriched (2.98% vs. matrix 1.73%, OMM 1.59%), consistent with the Mia40/CHCHD4 disulfide relay pathway.

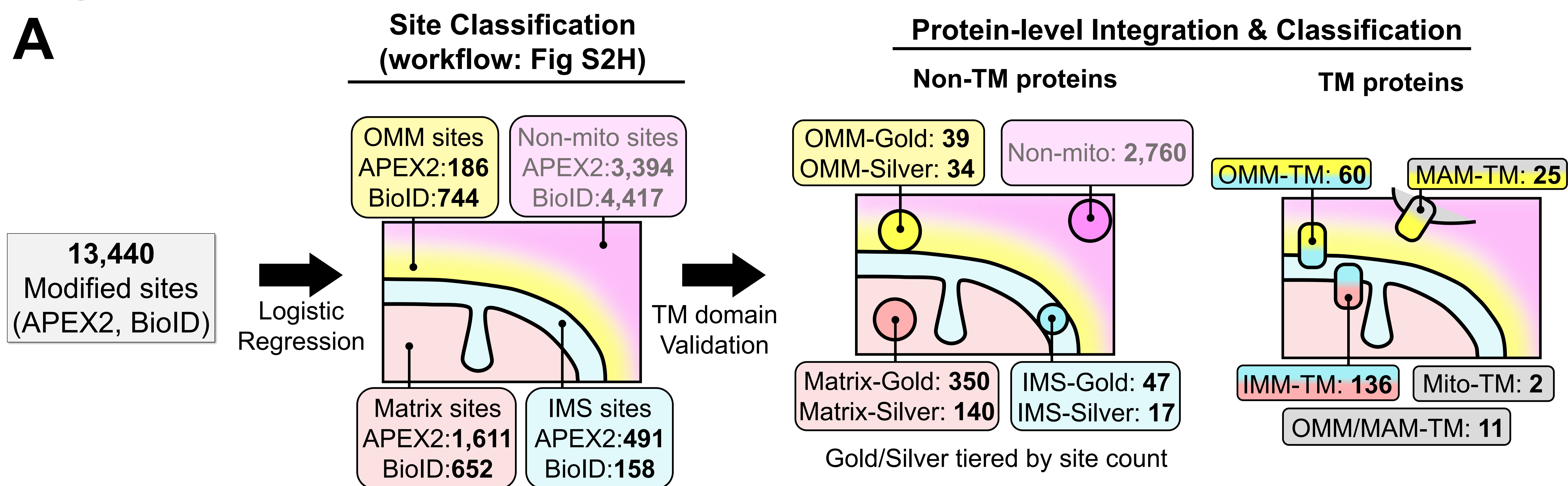
(C) gnomAD population variant density across mitochondrial compartments. Missense variant density was analyzed across 822 MitoAtlas proteins in five compartments: matrix (non-TM, n=489), IMM-TM (n=122–134), IMS (non-TM, n=64), OMM-TM (n=55–57), and OMM (non-TM, n=73); ranges reflect per-face protein counts. Variant positions from gnomAD via EBI Proteins API. Overall density: 51.8 variants/100 aa (182,774 unique missense positions across 352,895 residues). Dashed line, overall mean. Brackets, binomial test (TM vs. non-TM; \*\*\*p<0.001). Left, N-terminal (pos 1–50) vs. remainder: N-terminal residues, encompassing the mitochondrial targeting sequence in most matrix and IMM proteins, show higher variant density across all

compartments (matrix 61.0 vs. 52.8/100aa, \*\*\*; IMM-TM 56.3 vs. 46.8, \*\*\*; IMS 52.7 vs. 49.3, \*\*\*; OMM-TM 50.3 vs. 47.6, \*\*; OMM 51.6 vs. 50.0, n.s.), indicating reduced constraint on the cleaved MTS. Right, density per topological face: IMM-TM transmembrane domains are depleted in variants (43.1/100aa) compared to matrix-facing (51.8) and IMS-facing (45.0) loops (binomial  $p=8.0 \times 10^{-33}$ , \*\*\*). OMM-TM shows a similar but non-significant trend (TM 46.9 vs. non-TM 47.8–50.6;  $p=df0.20$ ). Soluble matrix proteins show the highest overall density (53.7/100aa).

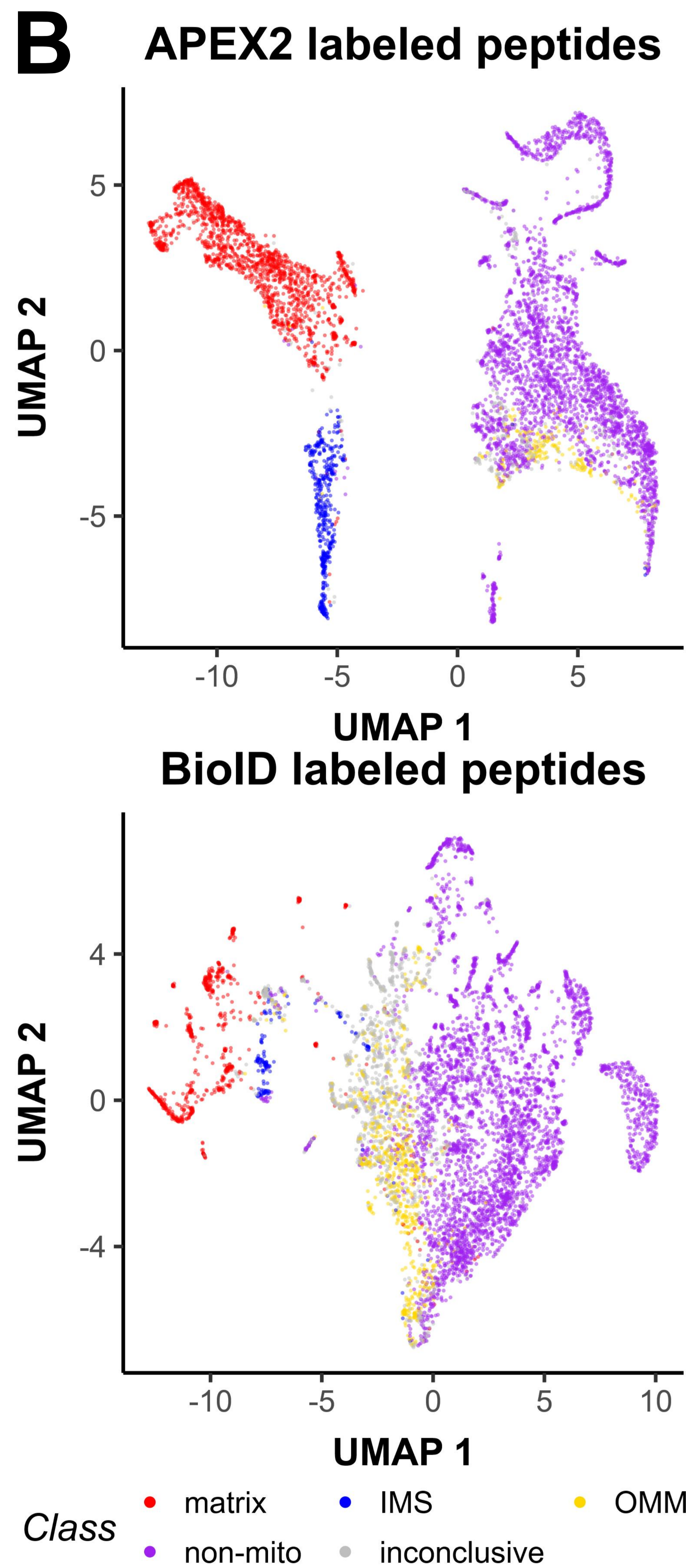


# Figure 2

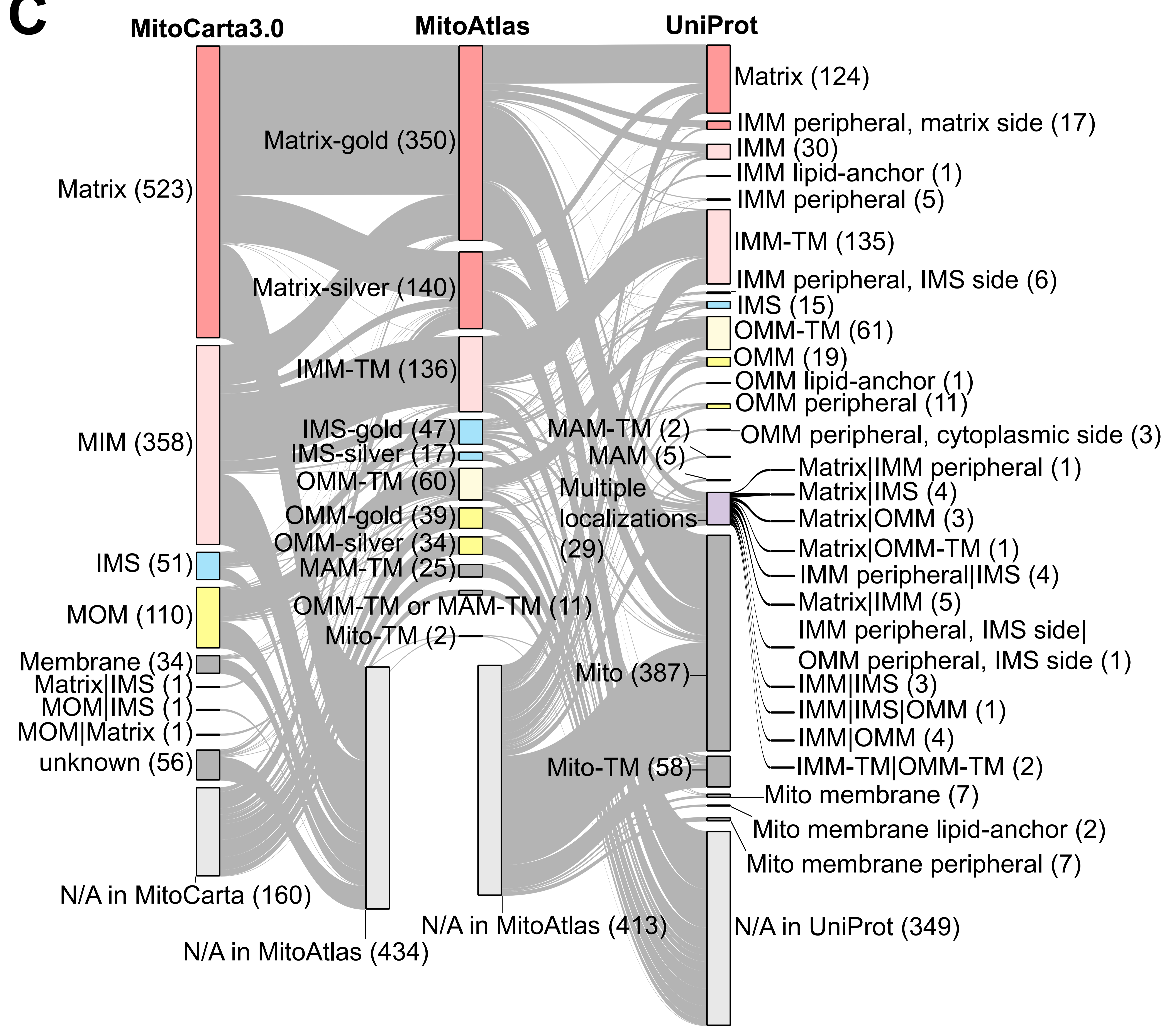
## A



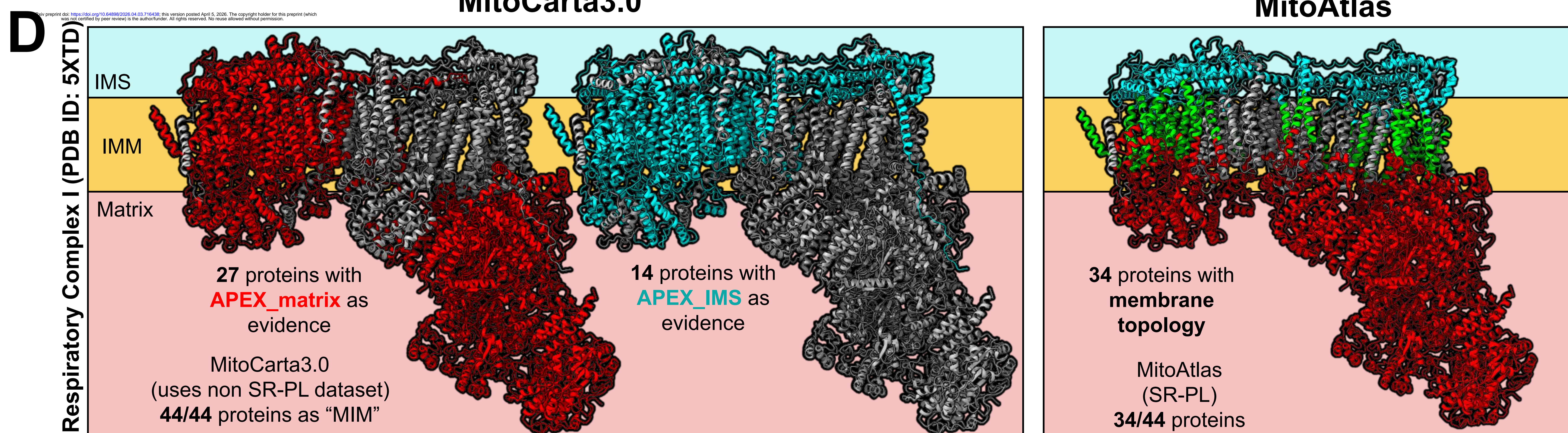
## B



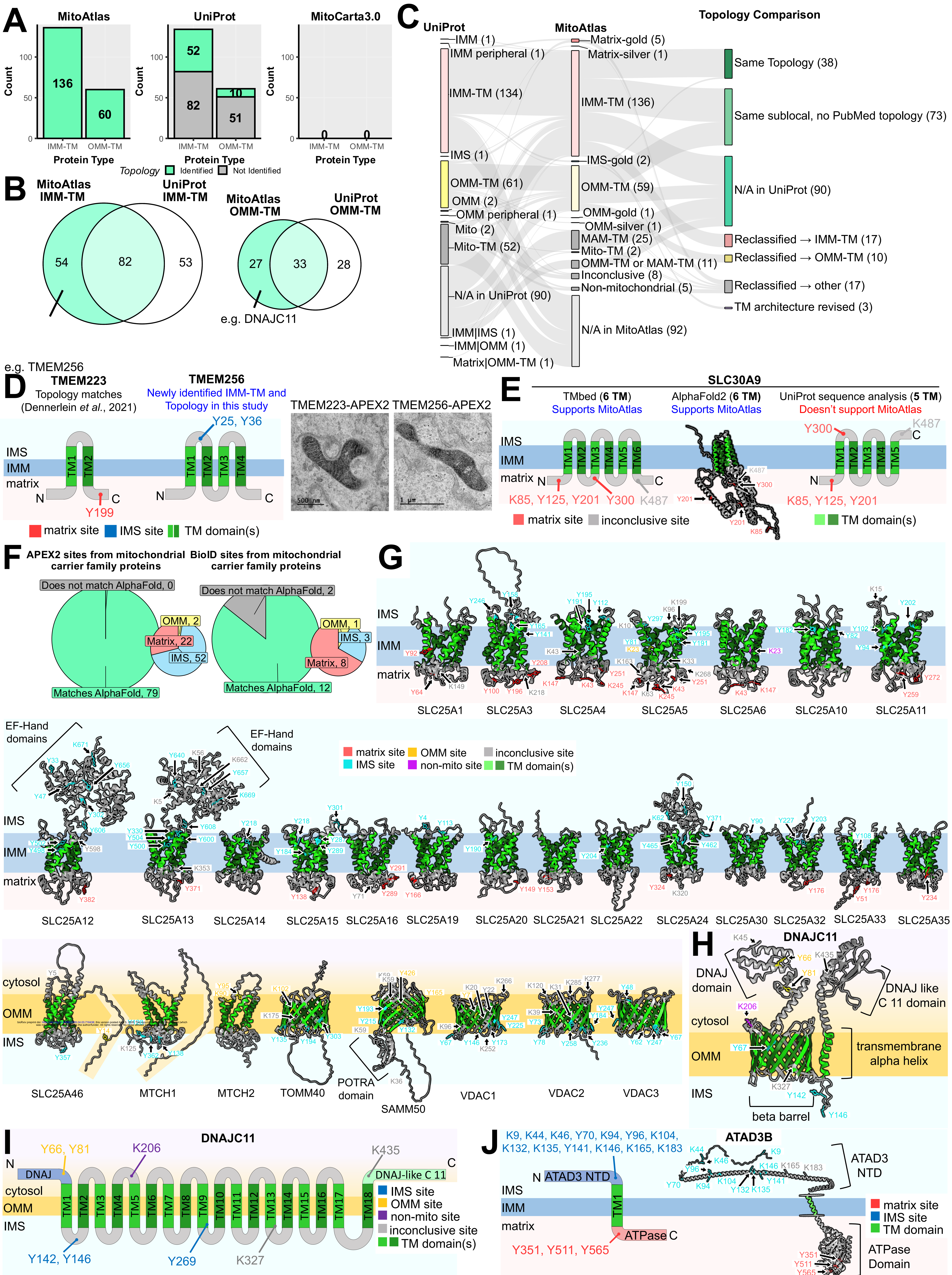
## C



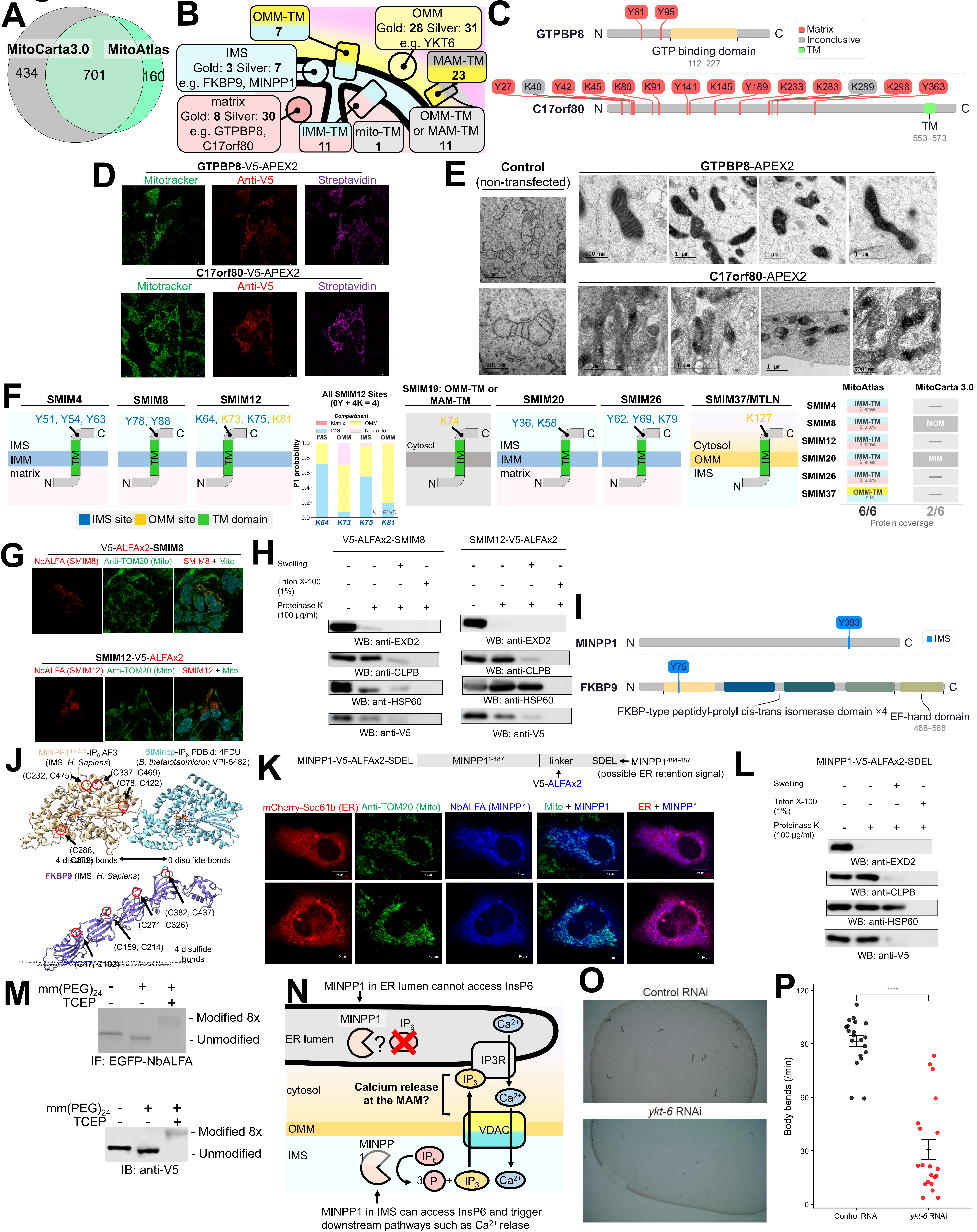
## D



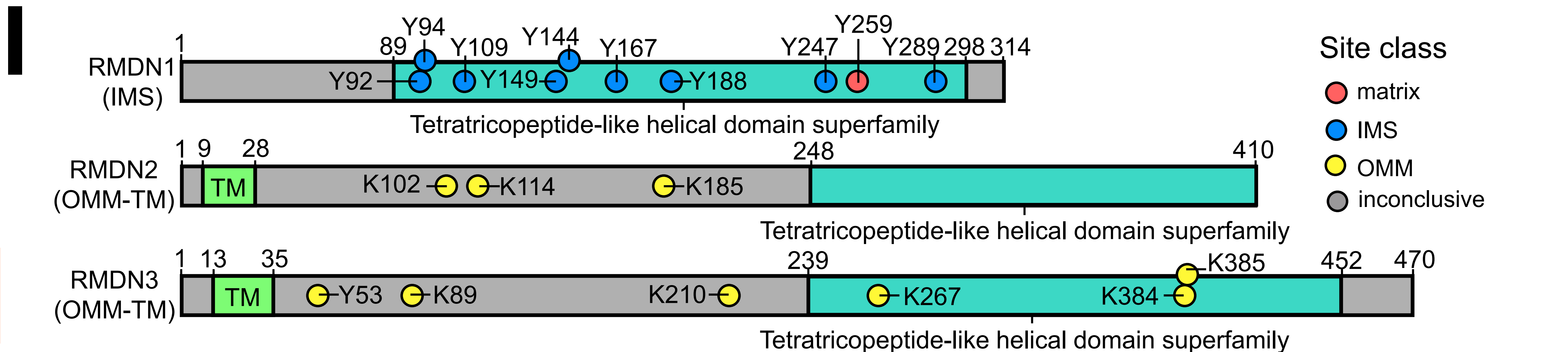
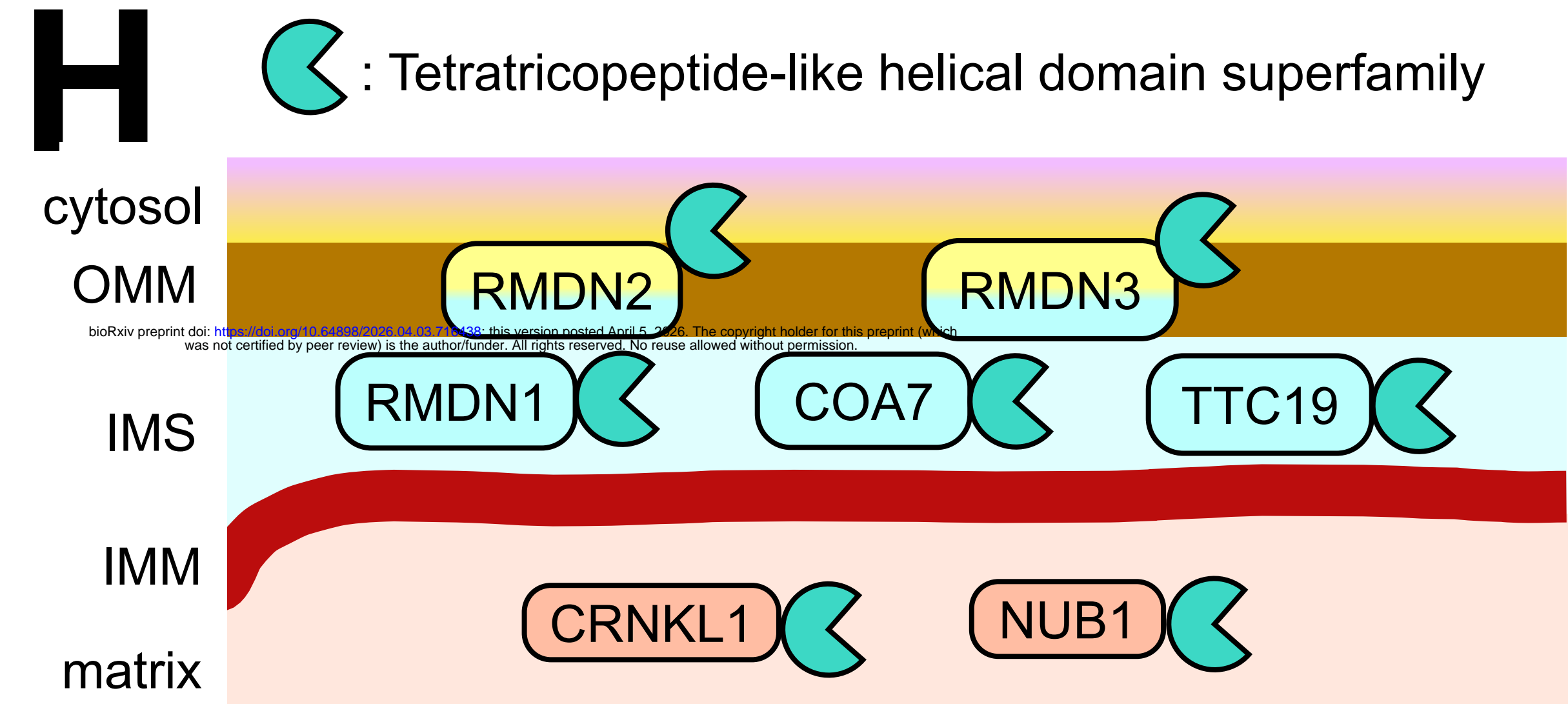
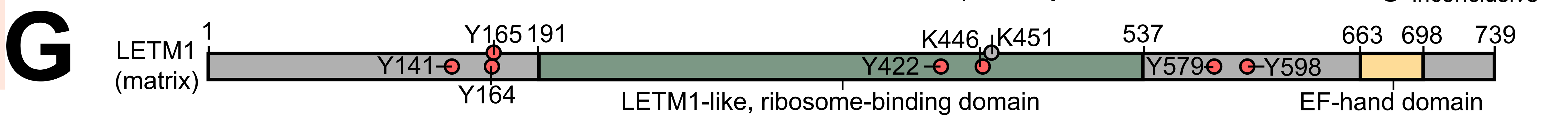
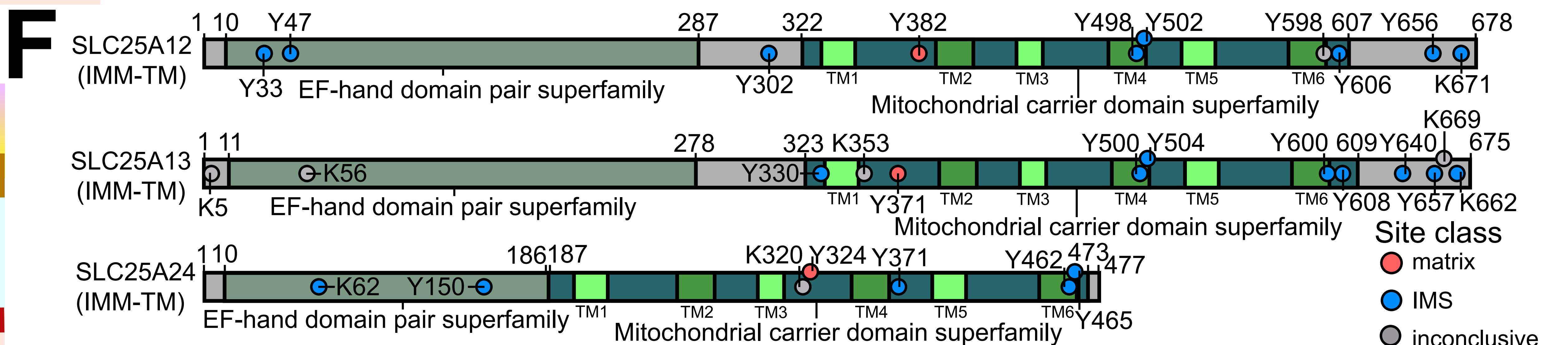
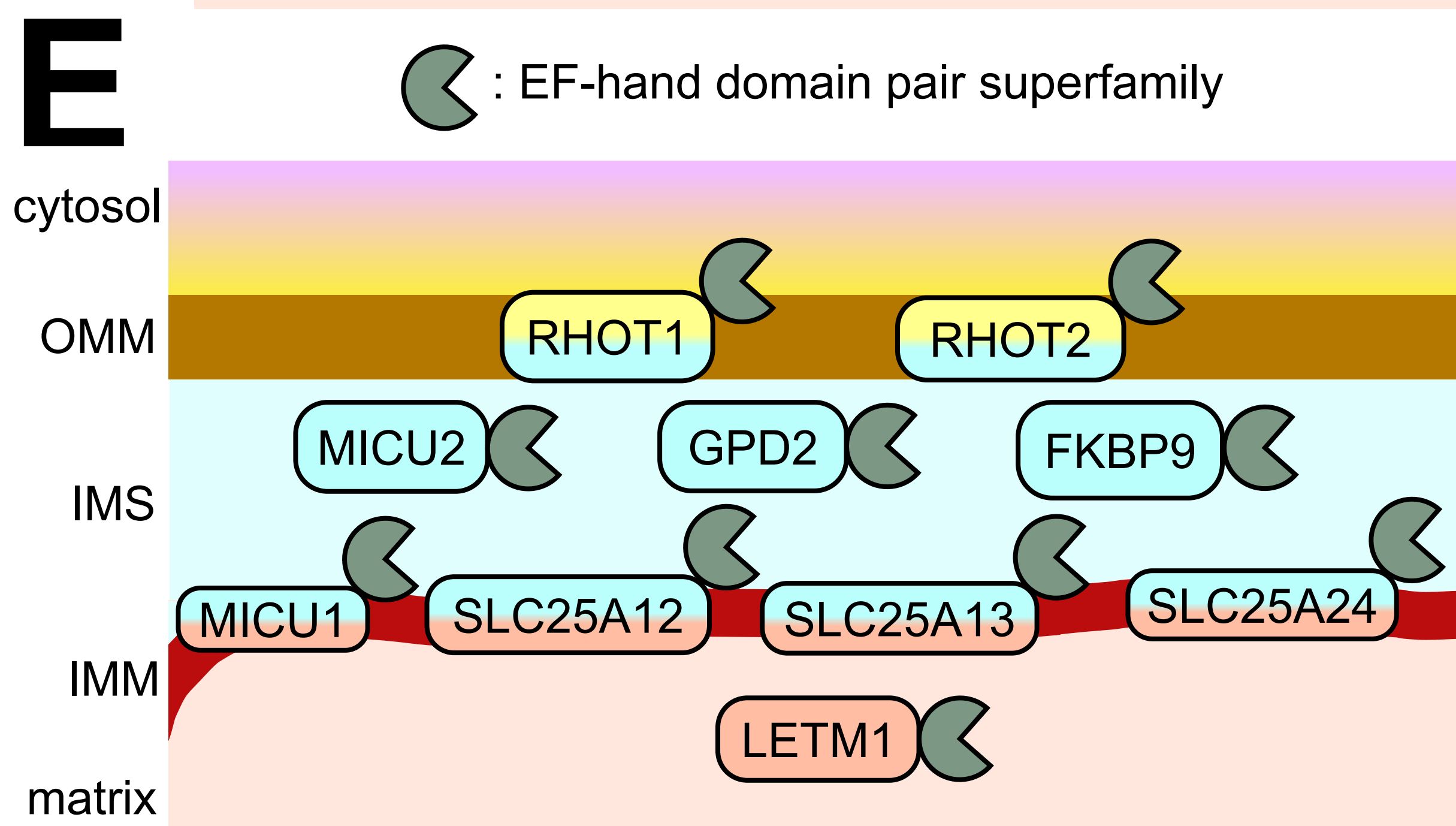
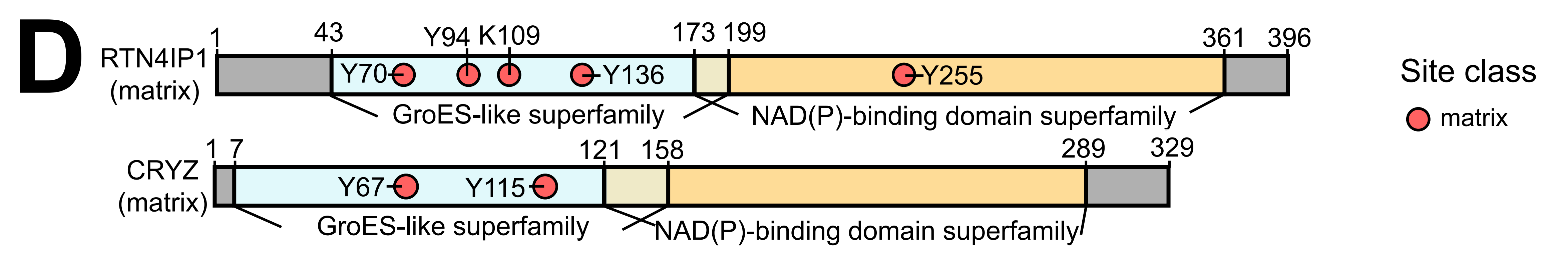
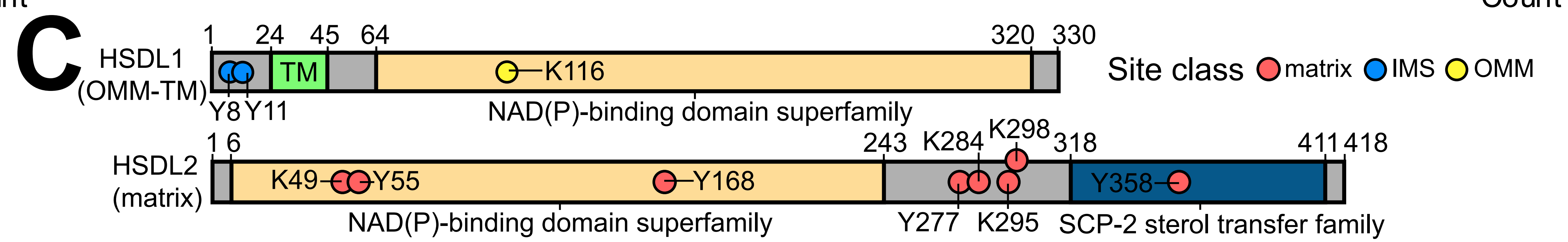
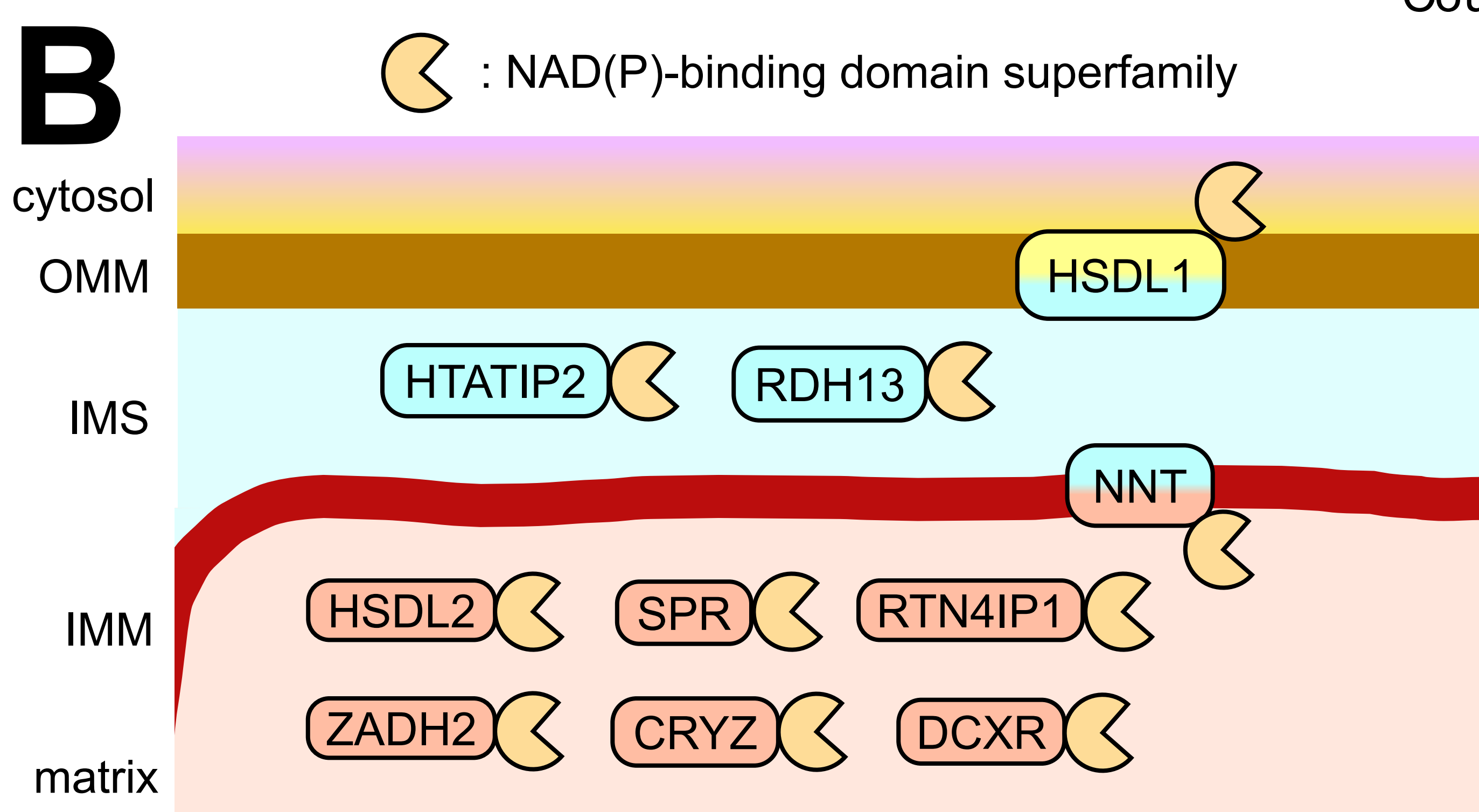
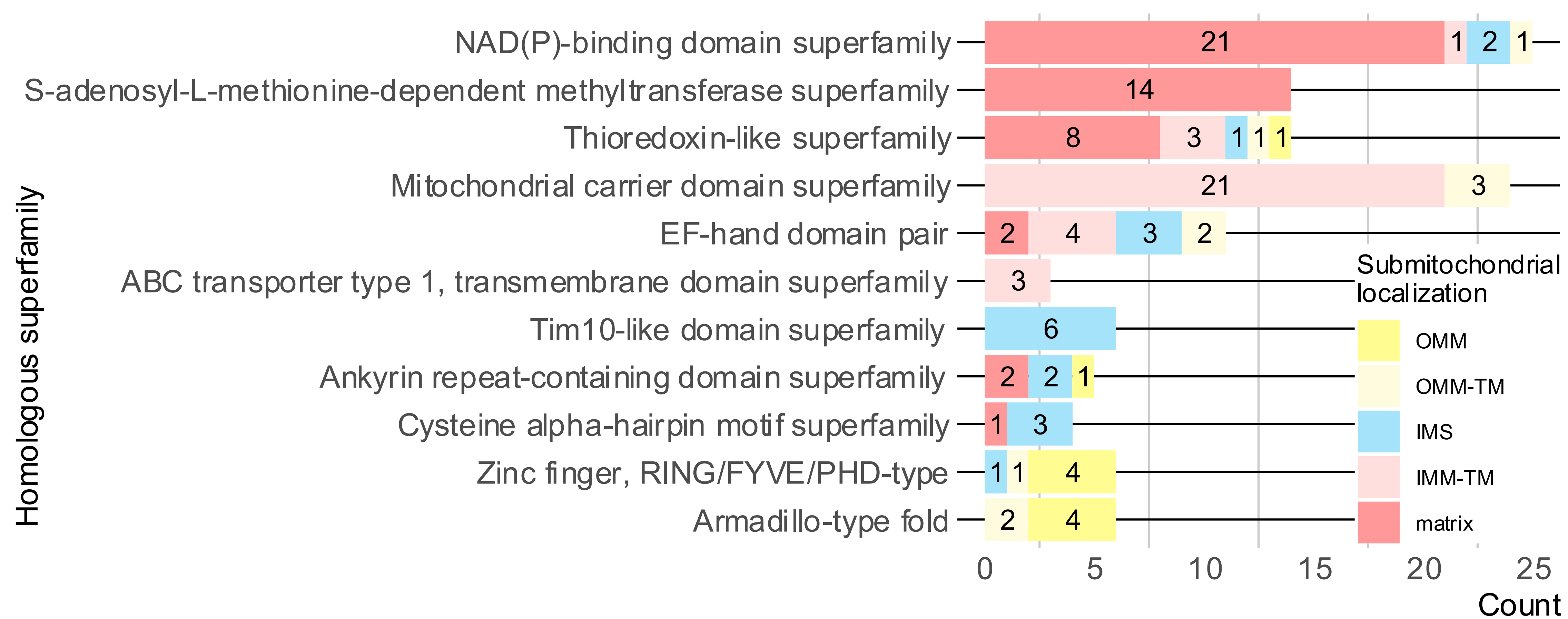
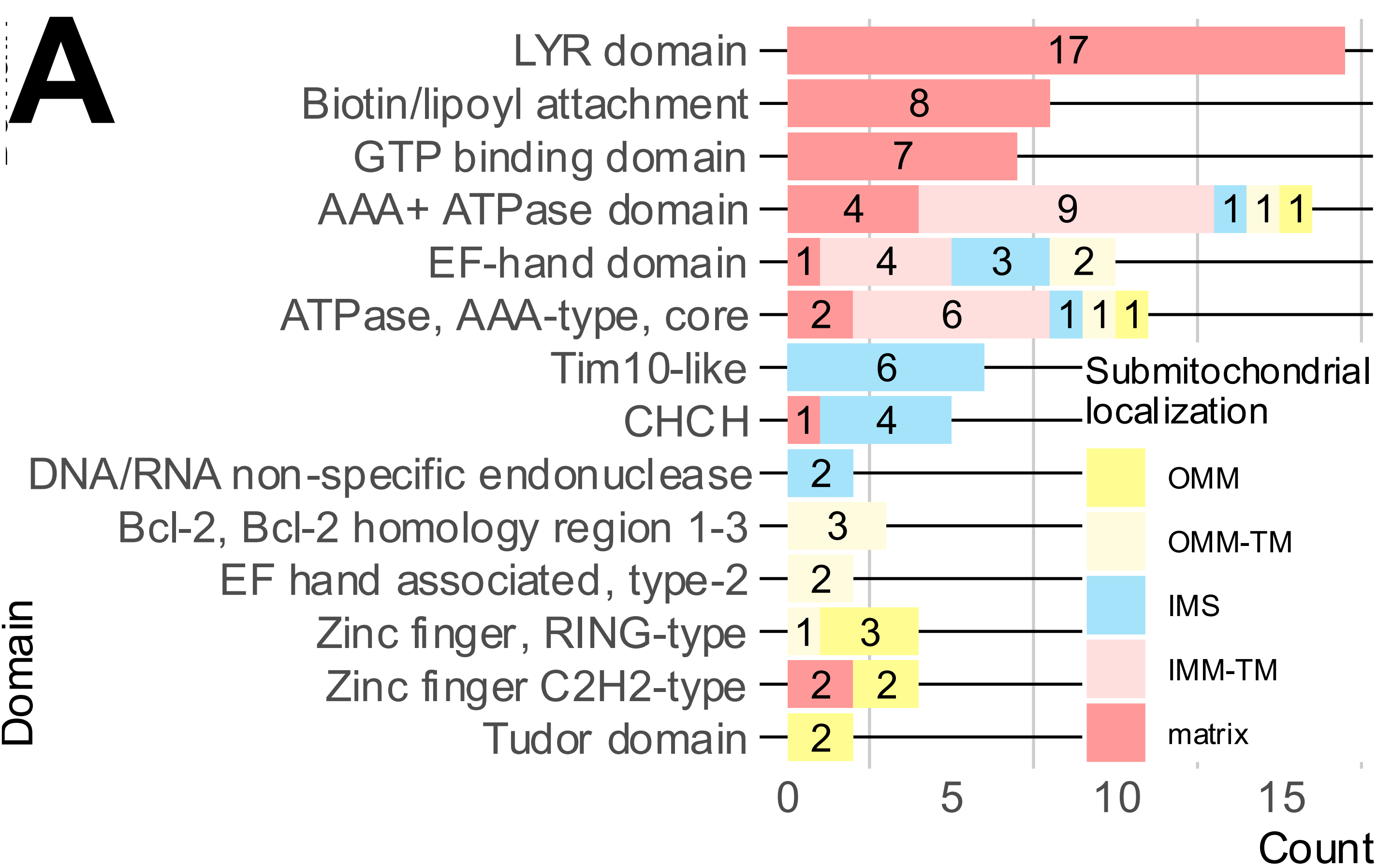
# Figure 3



# Figure 4

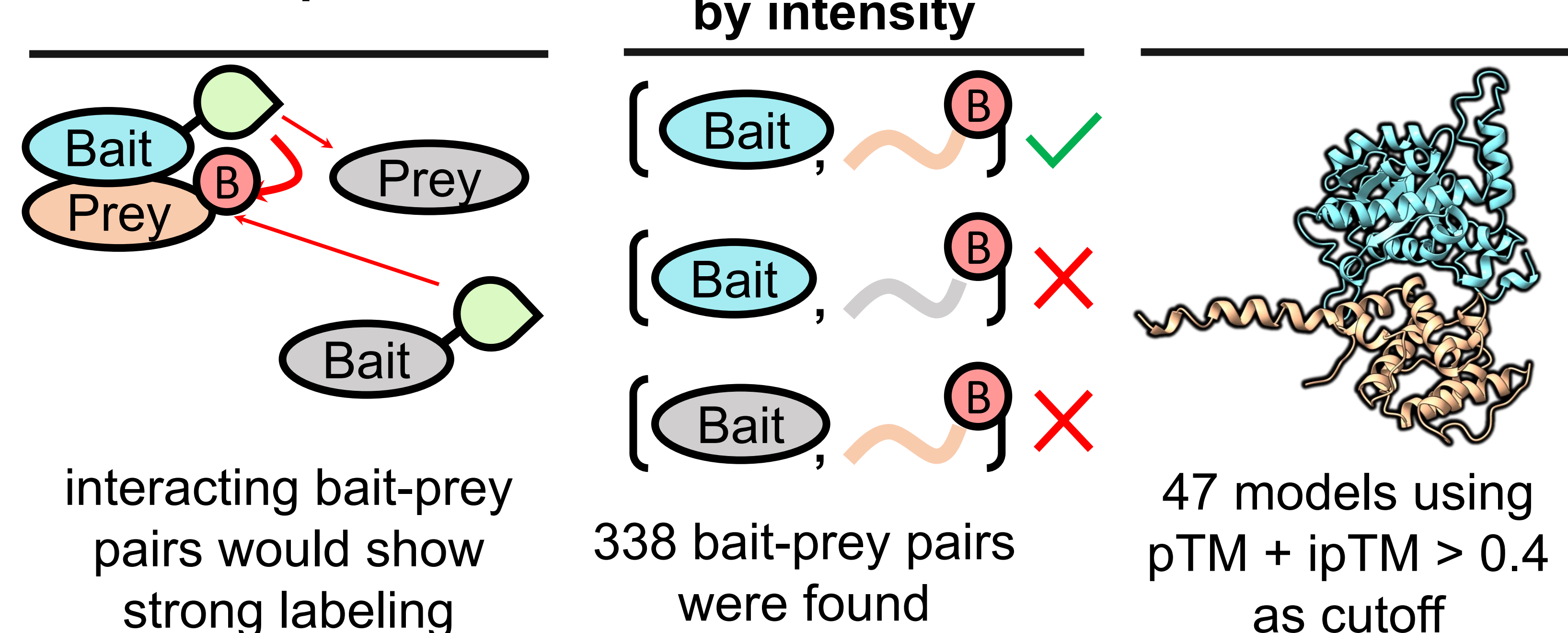


# Figure 5

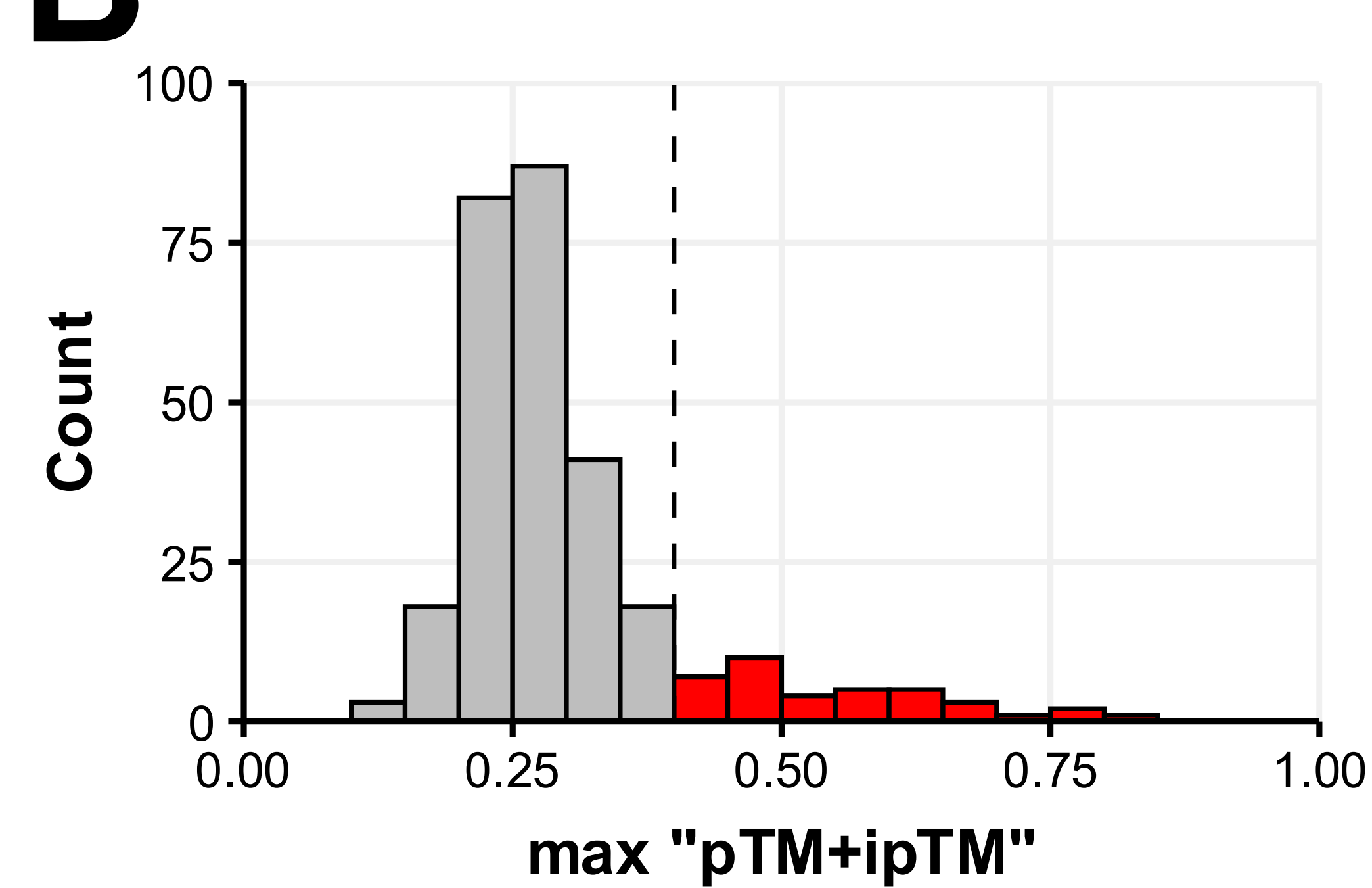


# Figure 6

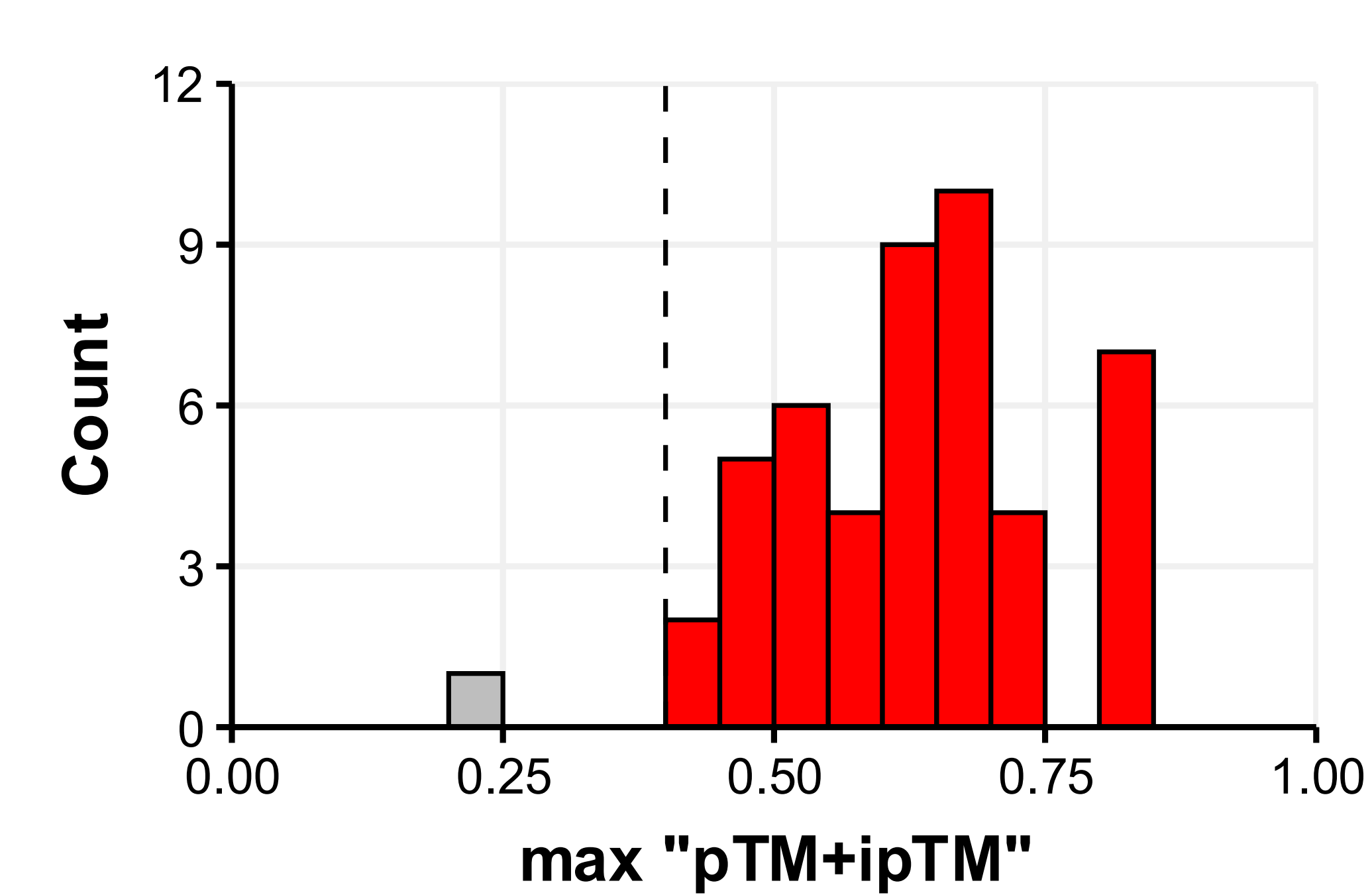
## A SR-PL analysis with multiple baits



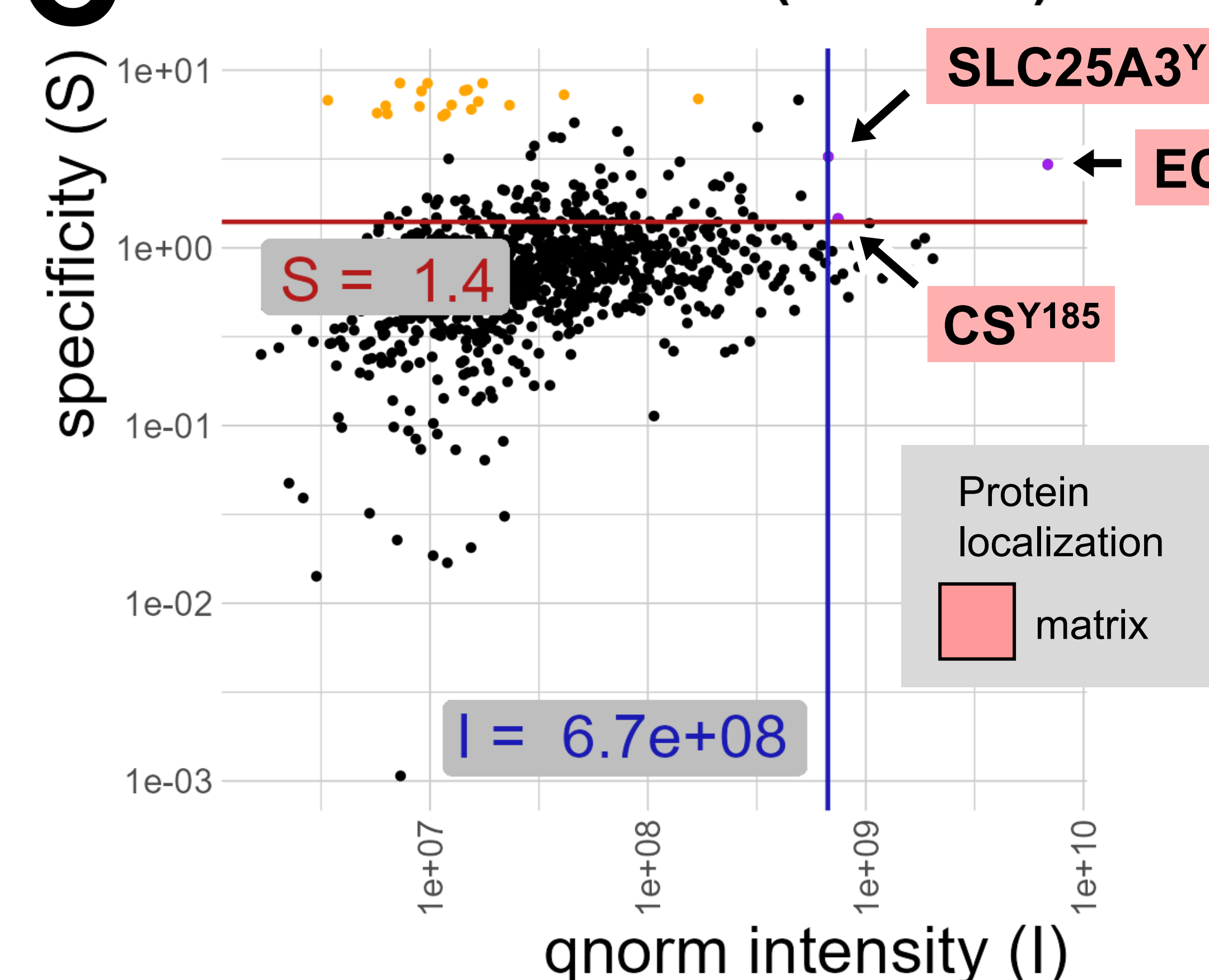
## B Initial run (n = 338)



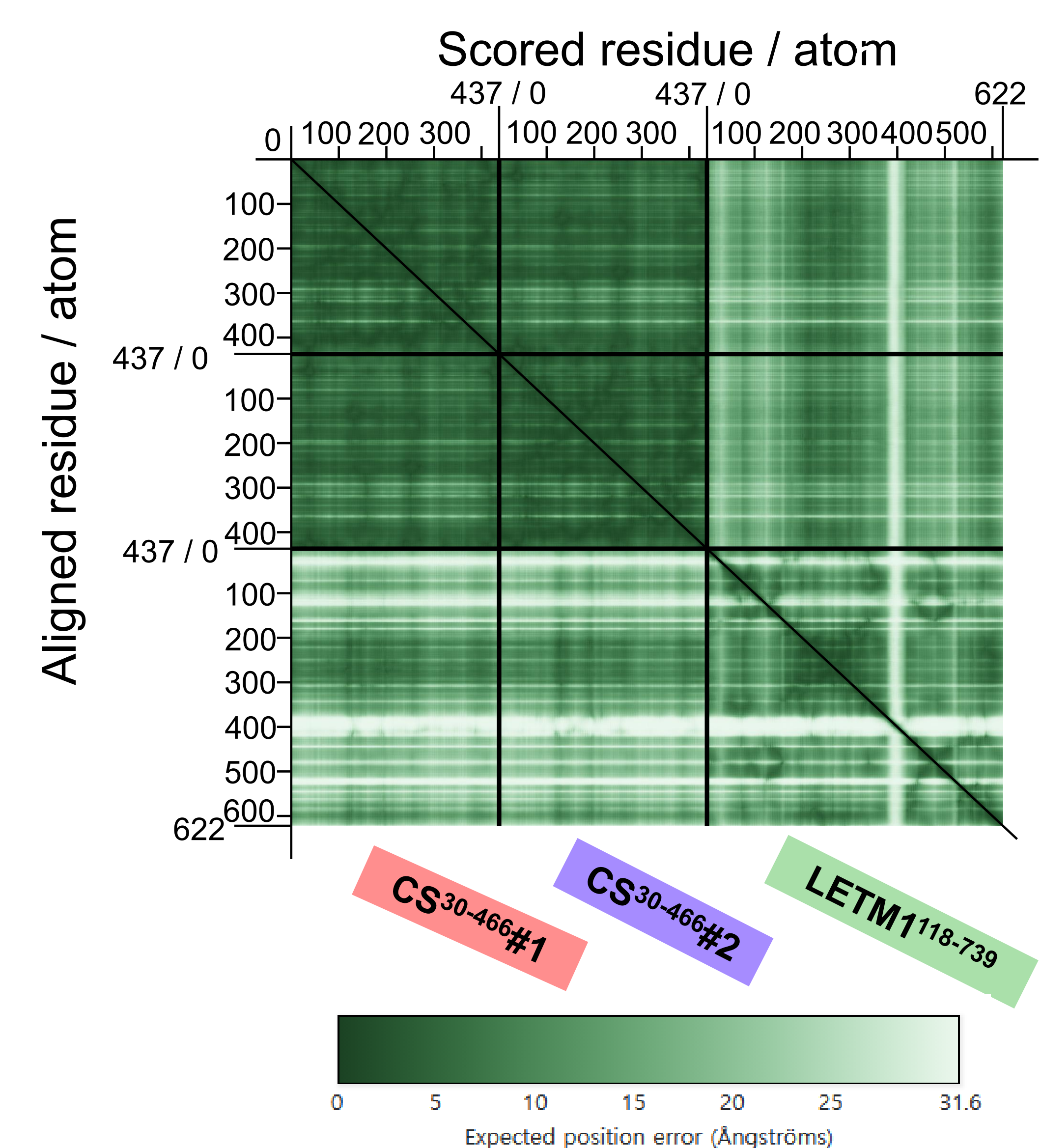
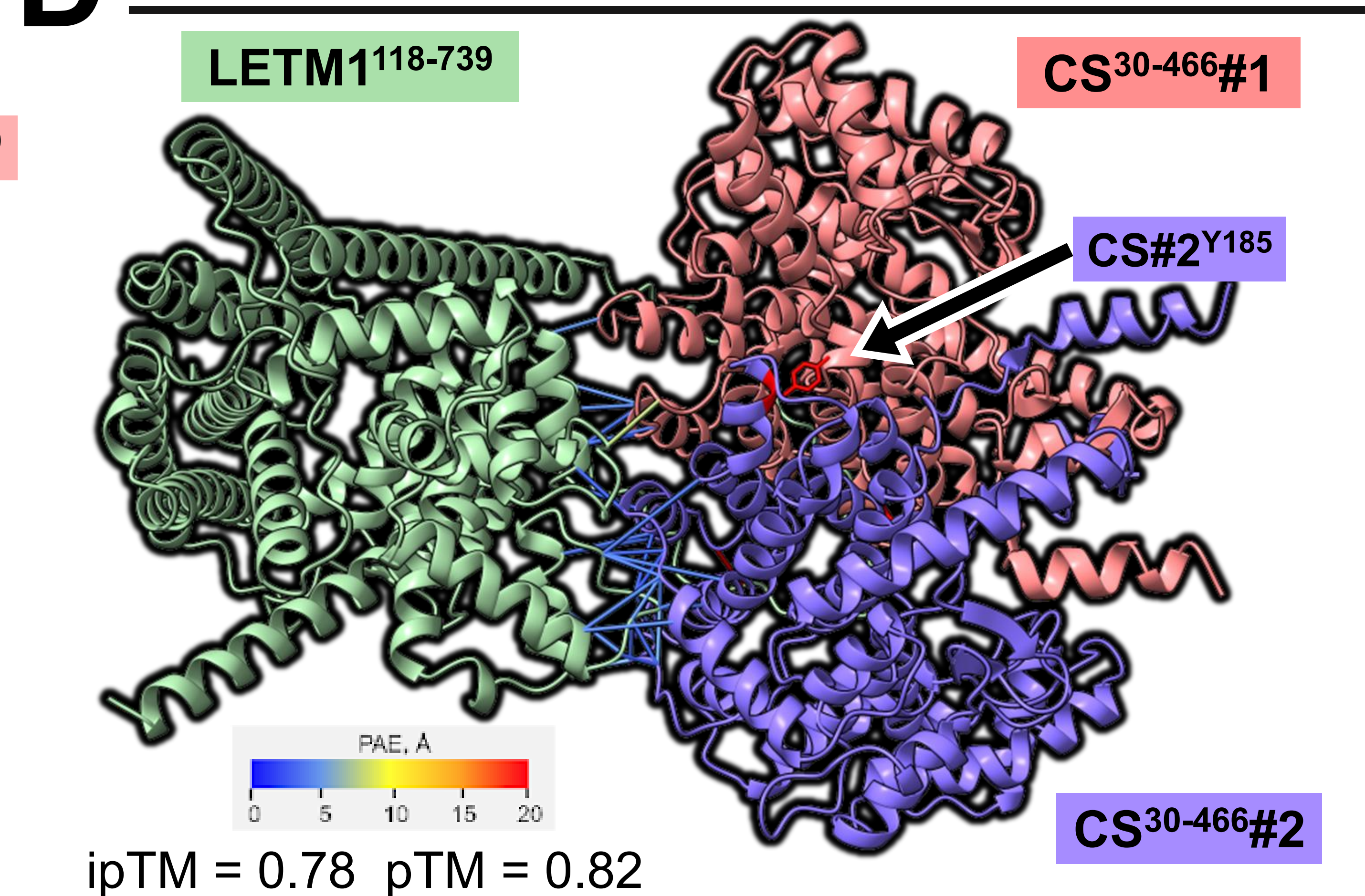
## Refinement run (n = 48)



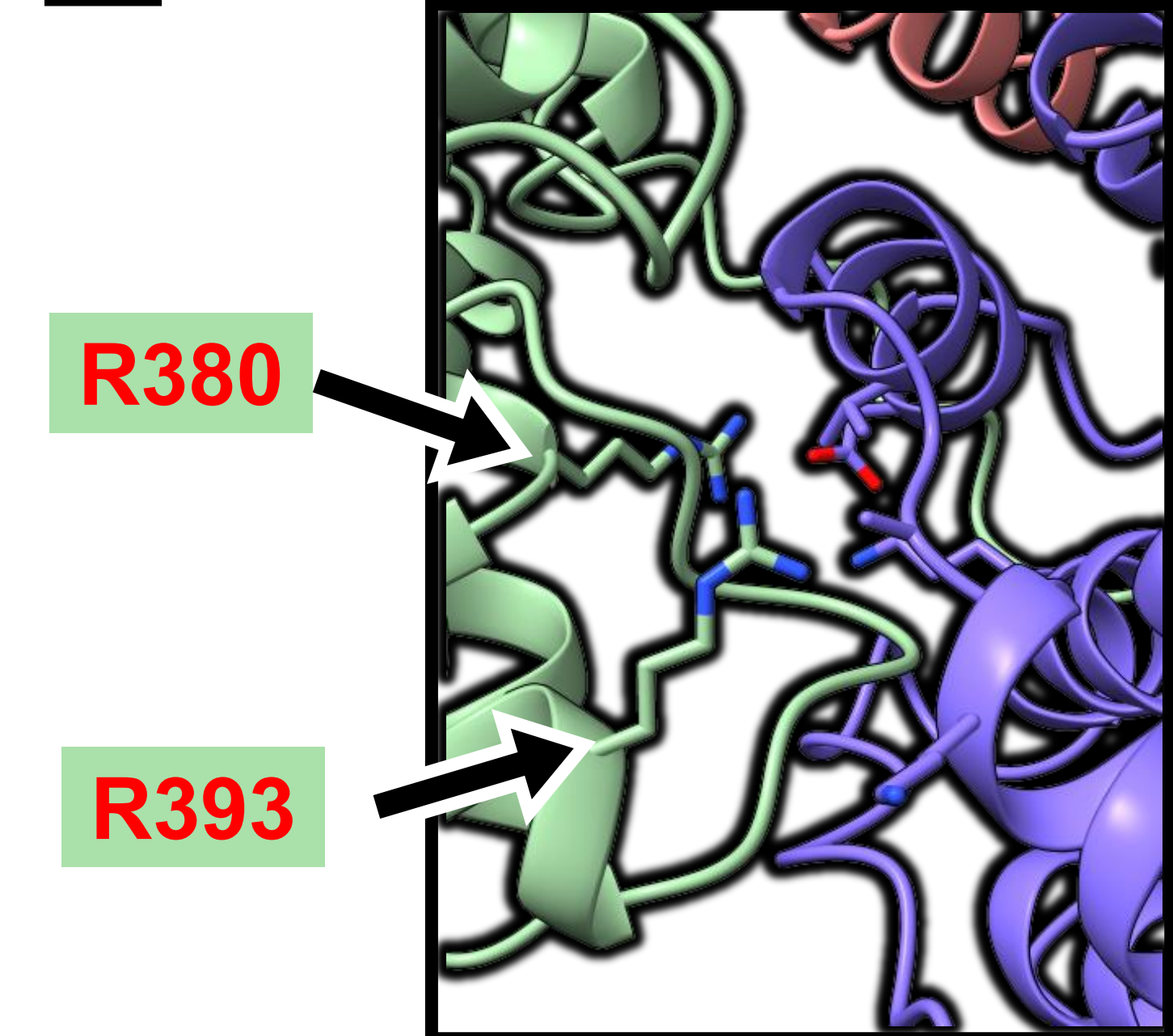
## C ss-APEX2-LETM1(152-739) sites



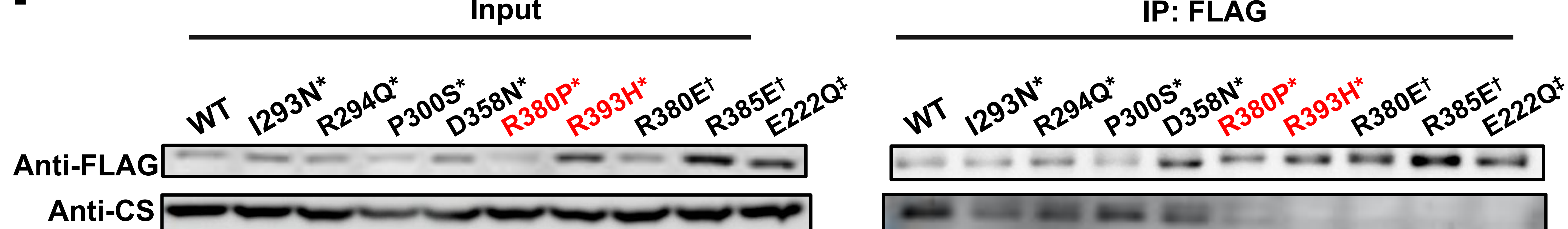
## D LETM1-CSx2



## E LETM1<sup>118-739</sup> CS<sup>30-466</sup>#1



## F MTS-3xFLAG-LETM1 variants Co-IP

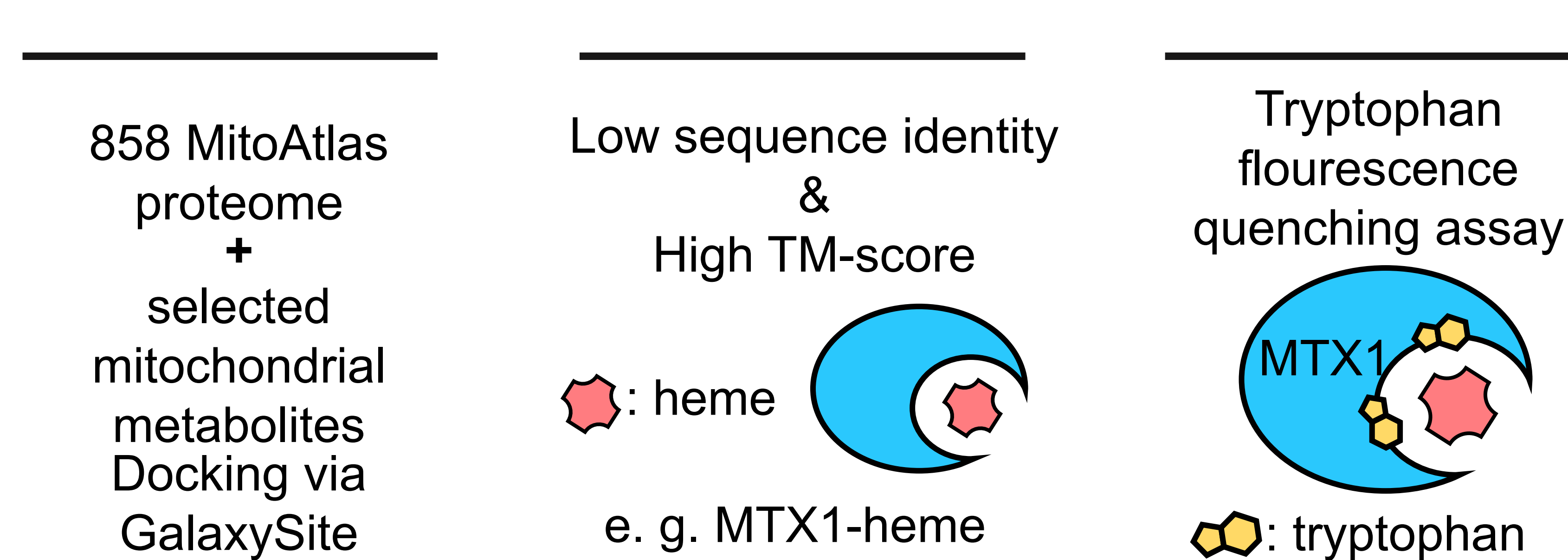


\*Pathogenic variants reported in bi-allelic LETM1 patients (Kaiyrzhanov et al., 2022)

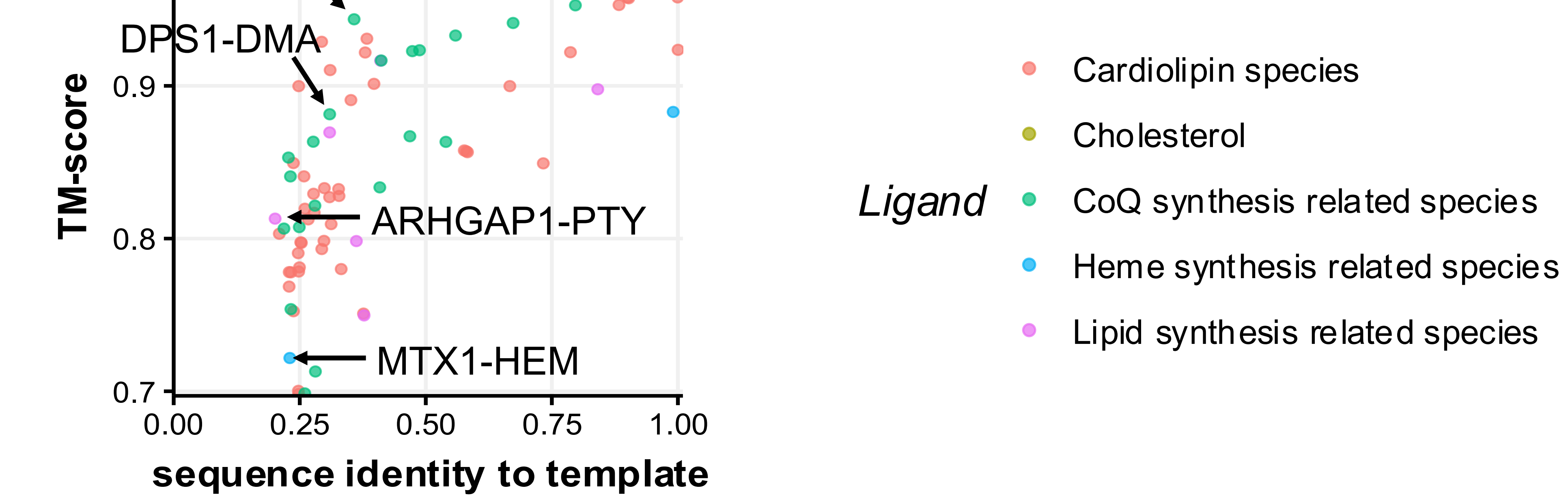
†Mutations designed to disrupt the predicted LETM1-CS interaction interface.

‡Ca<sup>2+</sup> transport-dead mutation; (corresponding to mouse E221Q) (Shao et al., 2016)

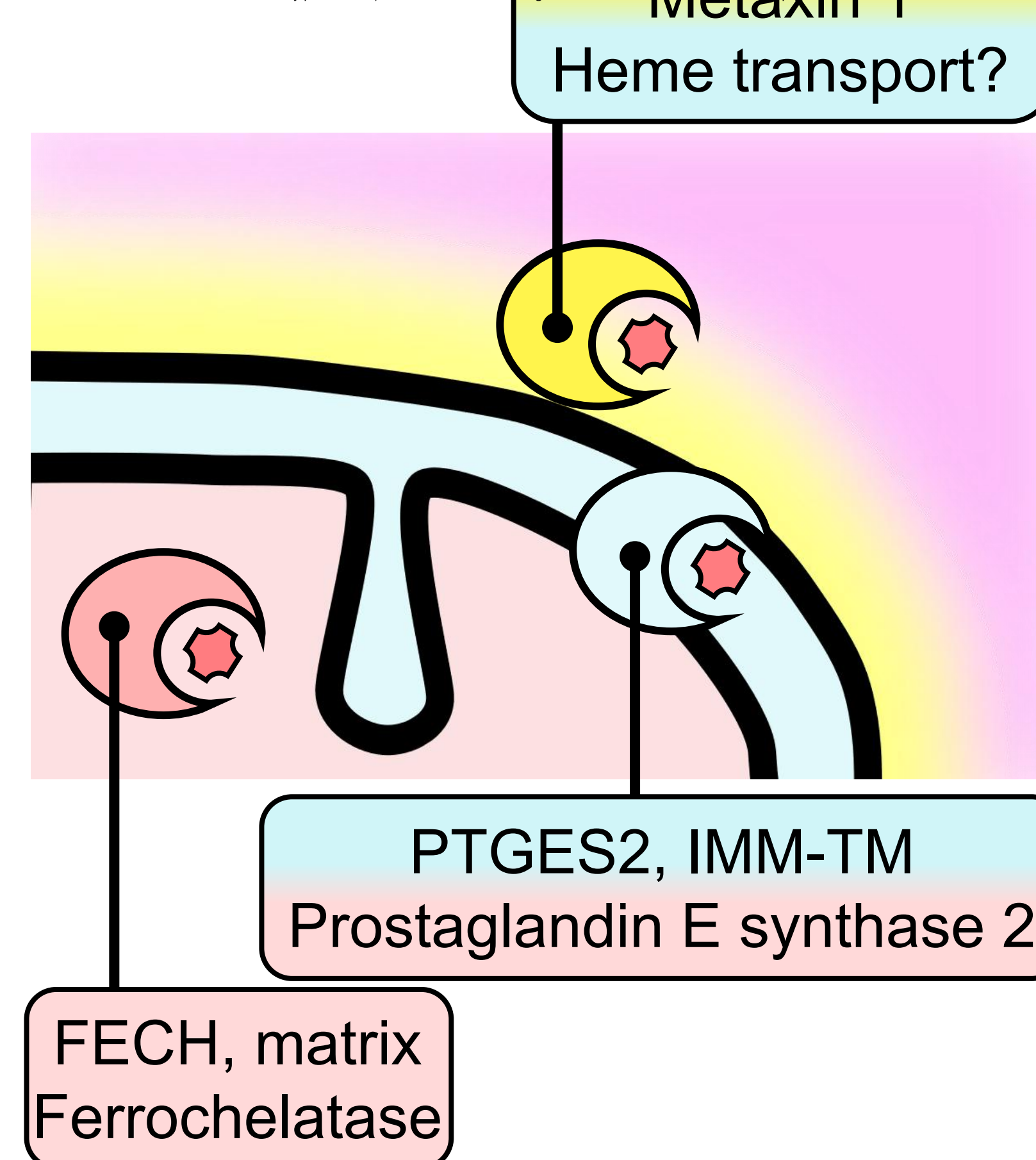
## G Metabolite docking



## H TM-score vs sequence identity to template



## I heme MTX1, OMM-TM Metaxin 1 Heme transport?



## J GalaxySite

MTX1<sup>152-393</sup>-HEM

Template: 2PBJ chain B

Sequence identity: 0.23

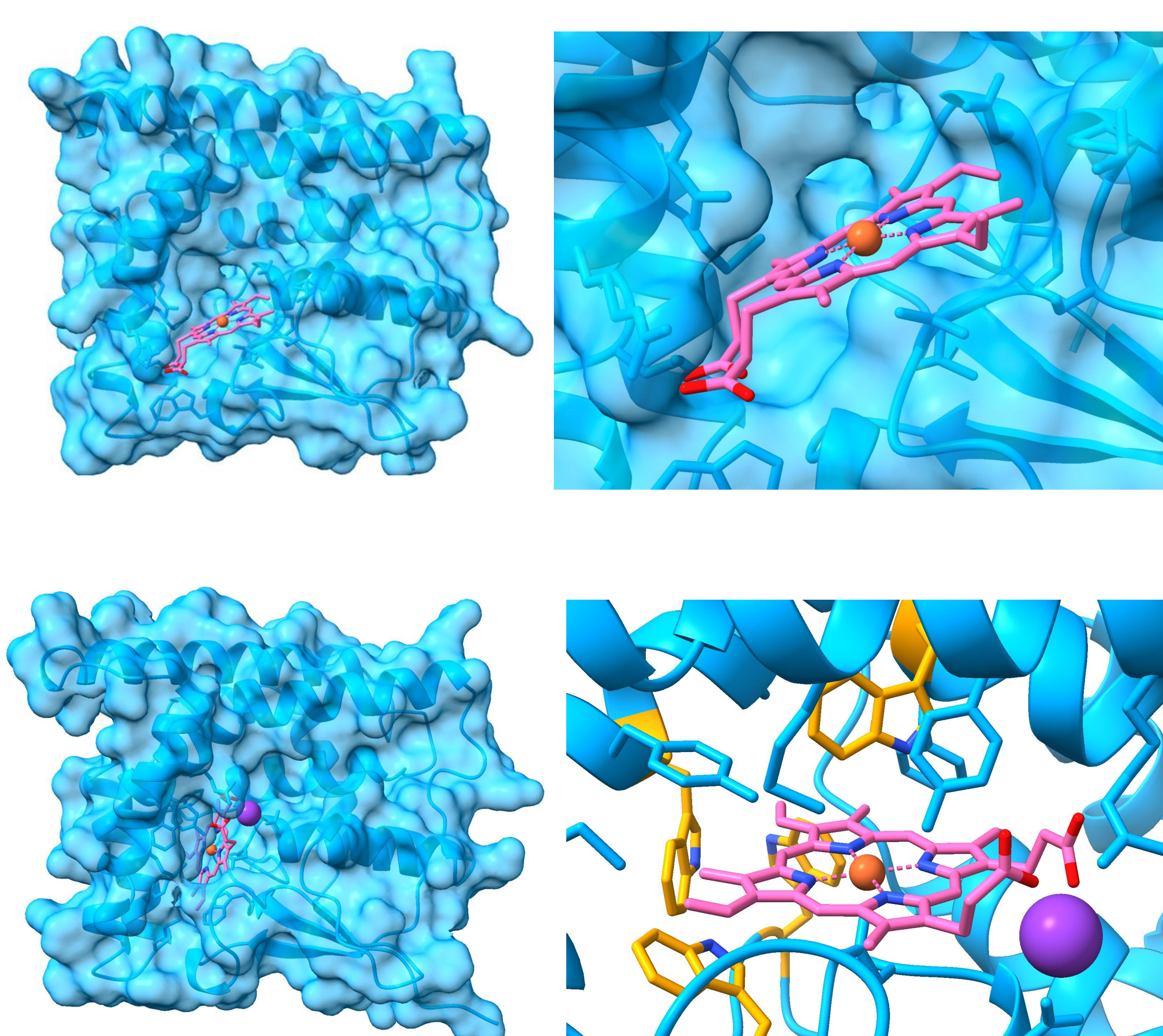
TM-score: 0.72

## AlphaFold3

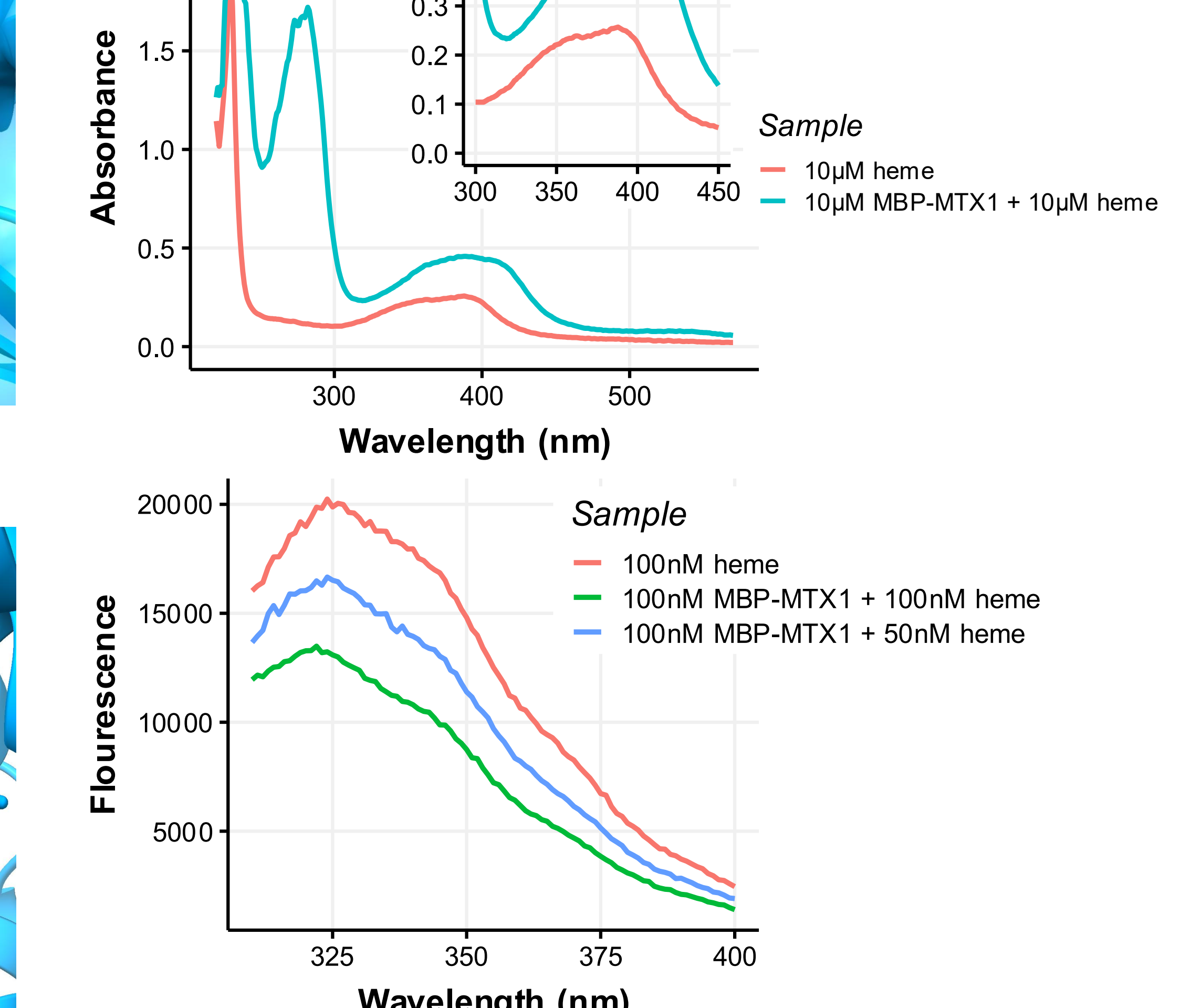
MTX1<sup>152-393</sup>-HEM, K<sup>+</sup>

pTM: 0.87

ipTM: 0.82

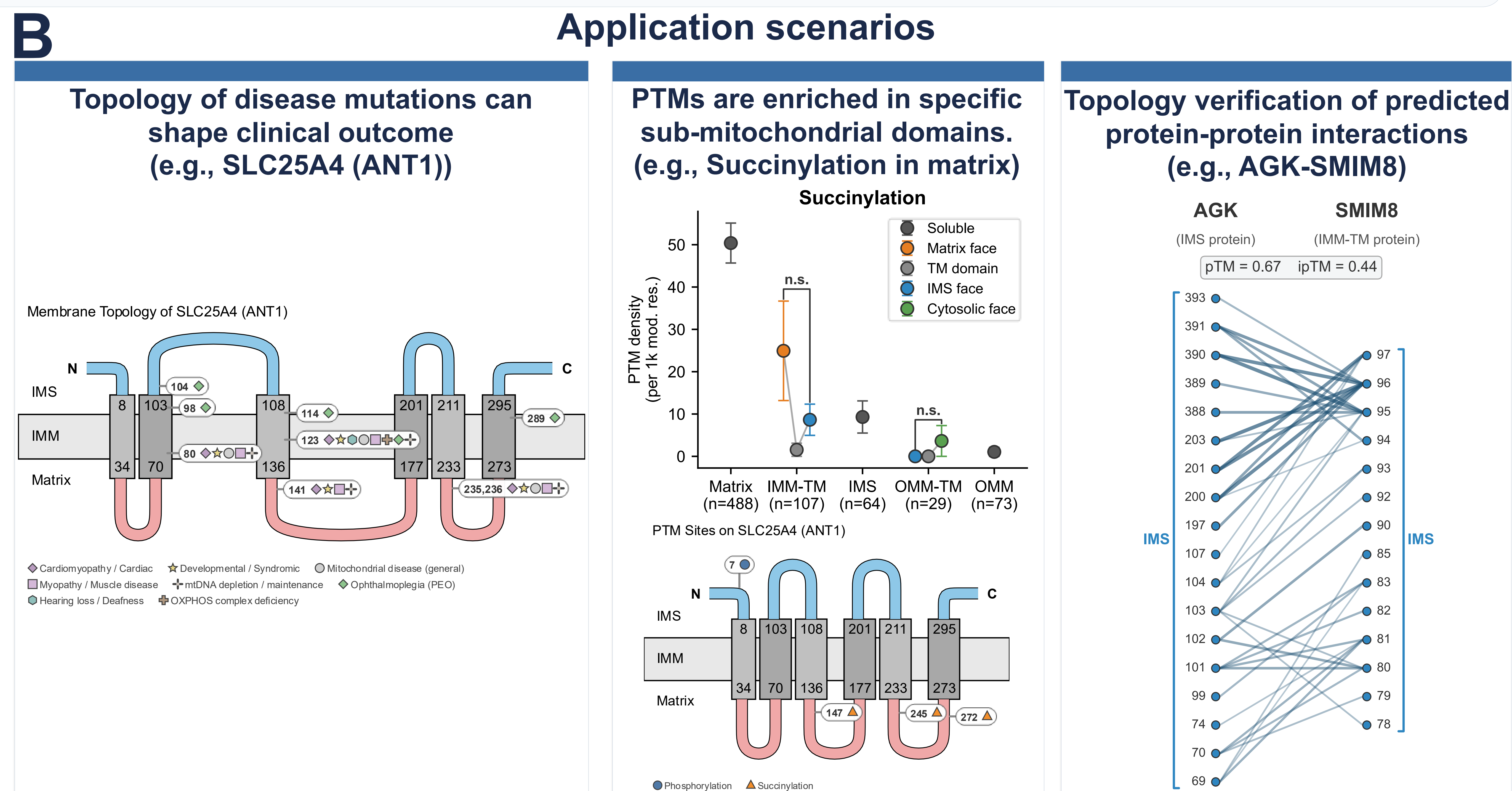
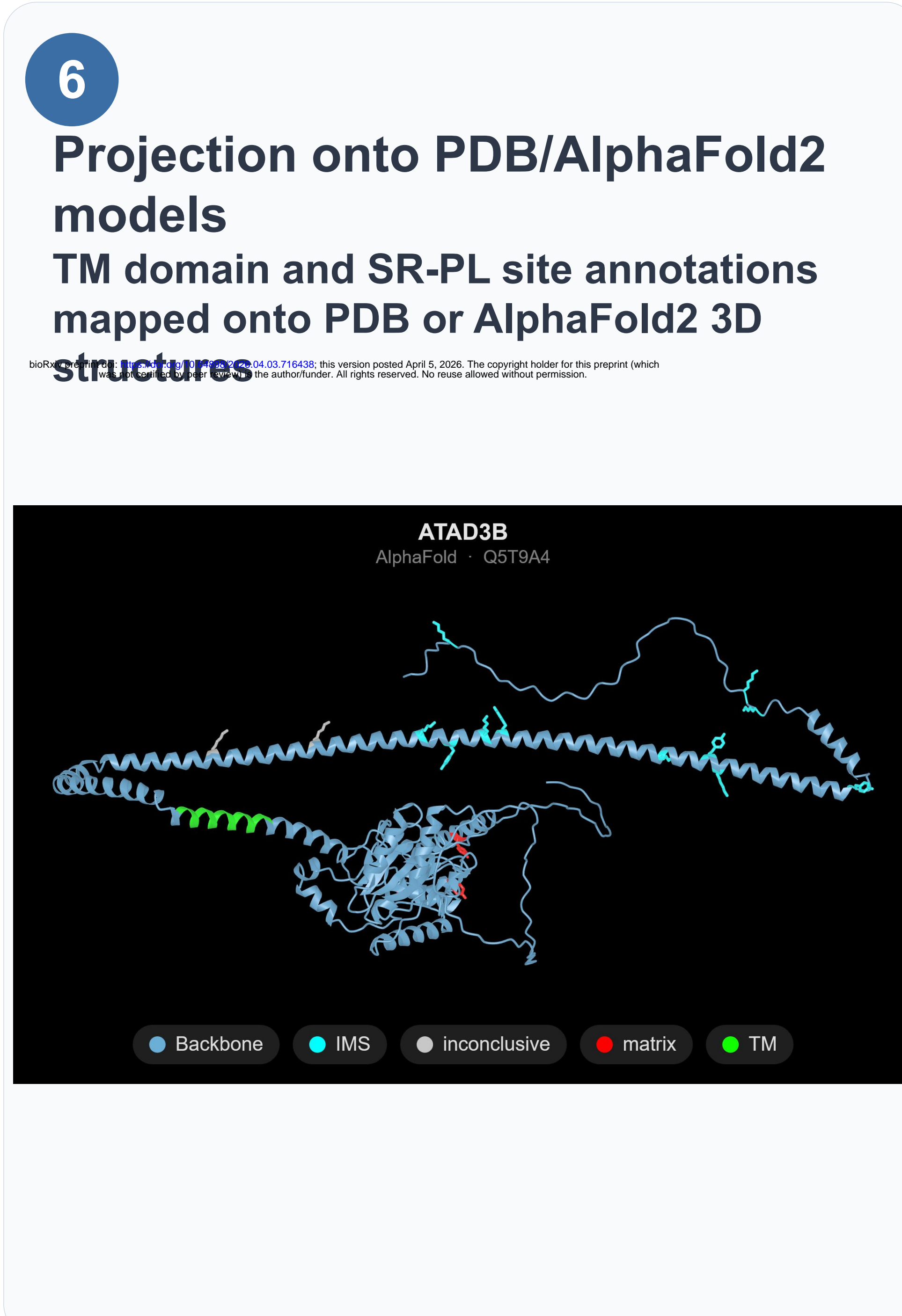
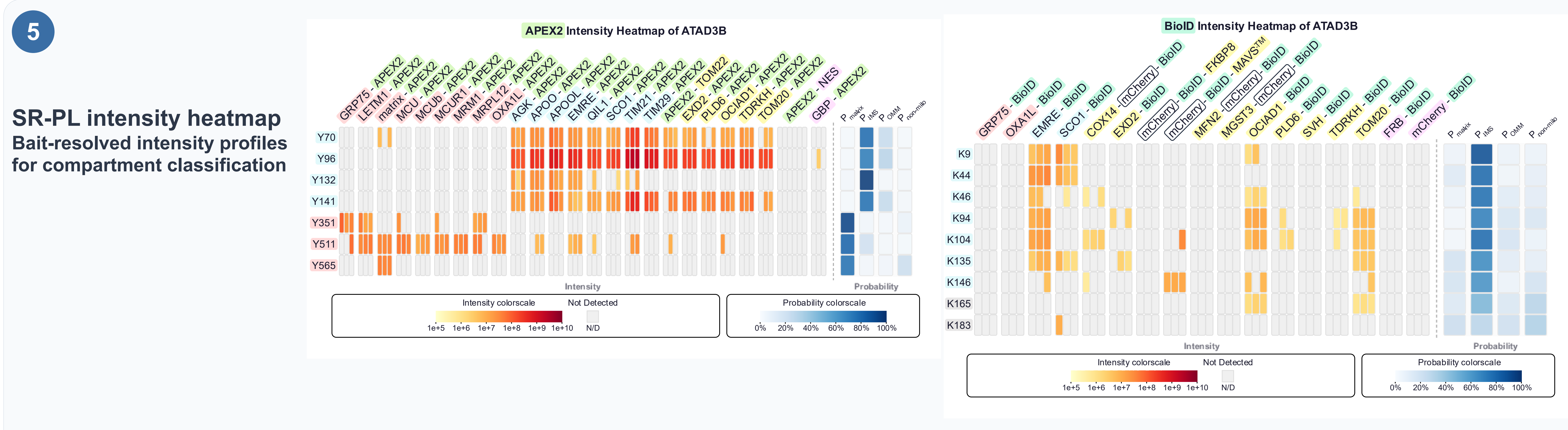
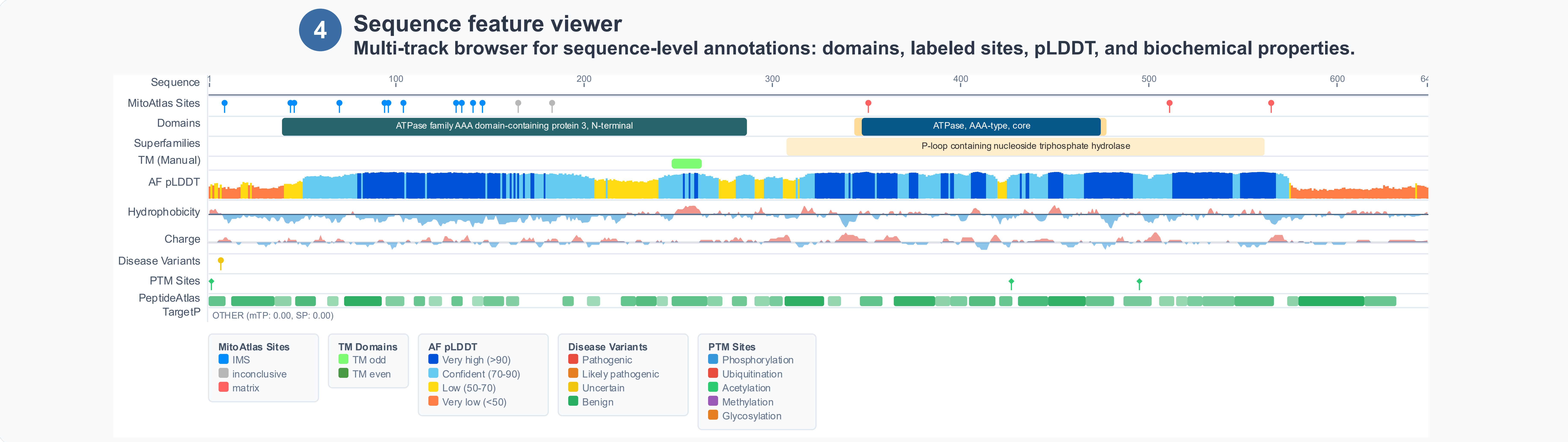
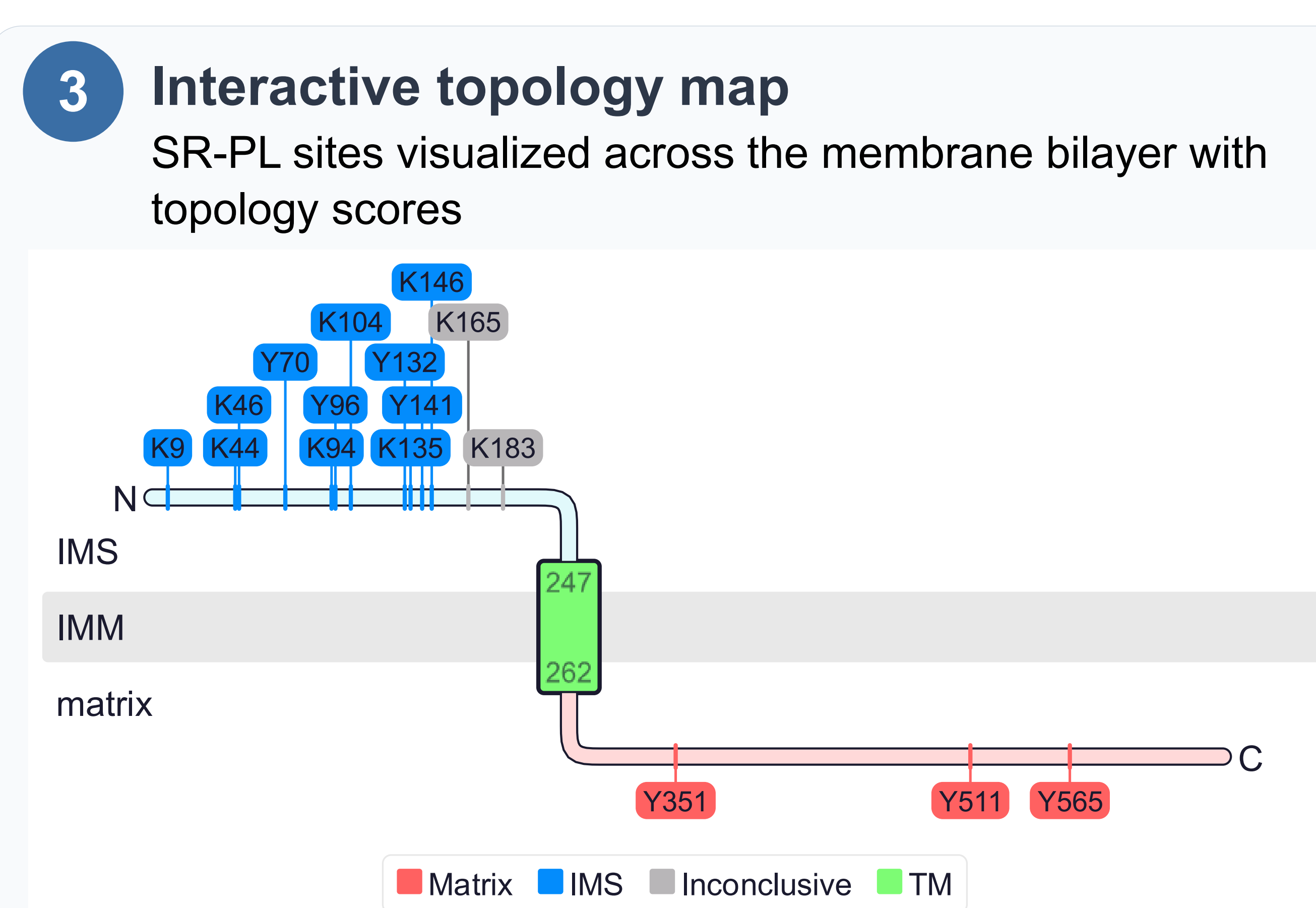
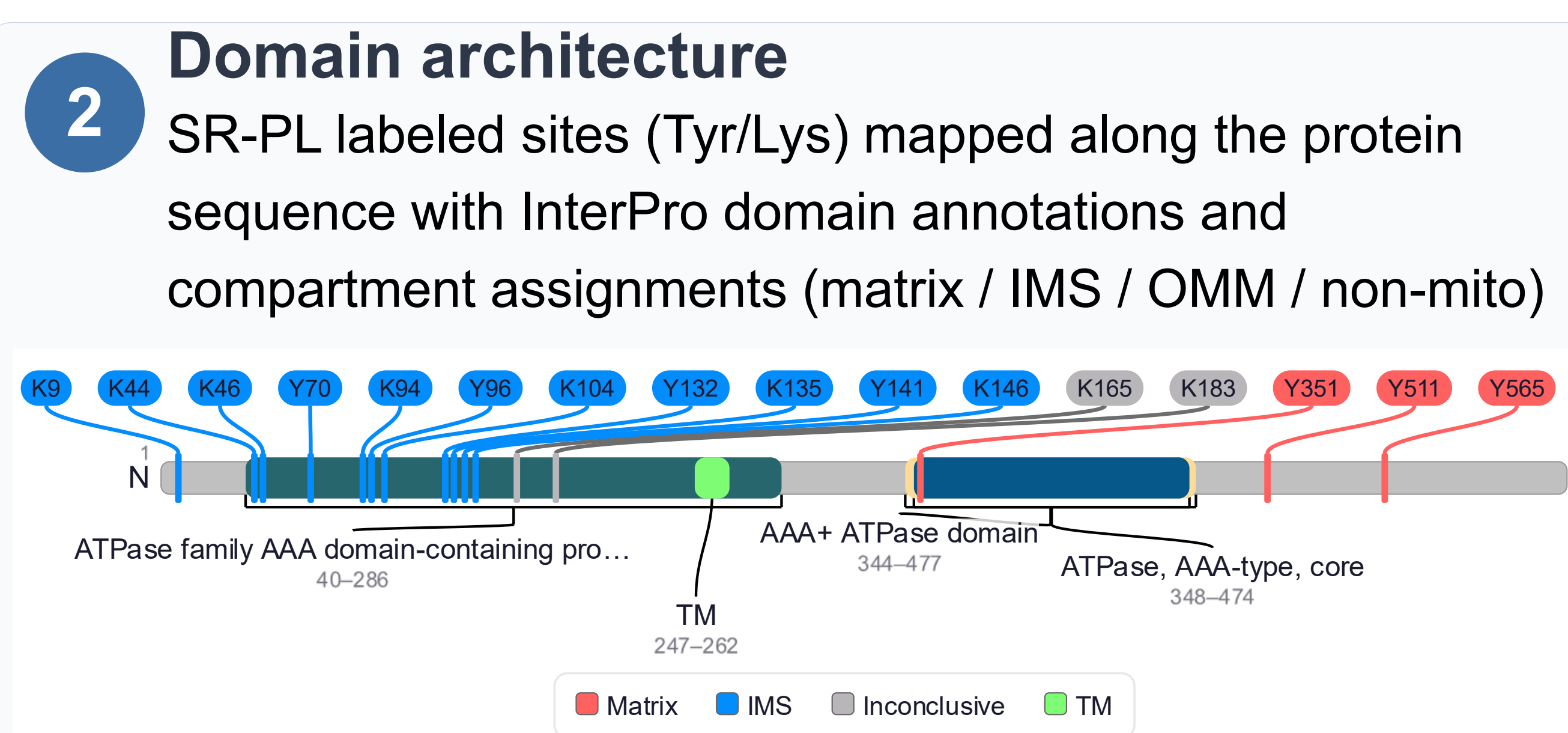
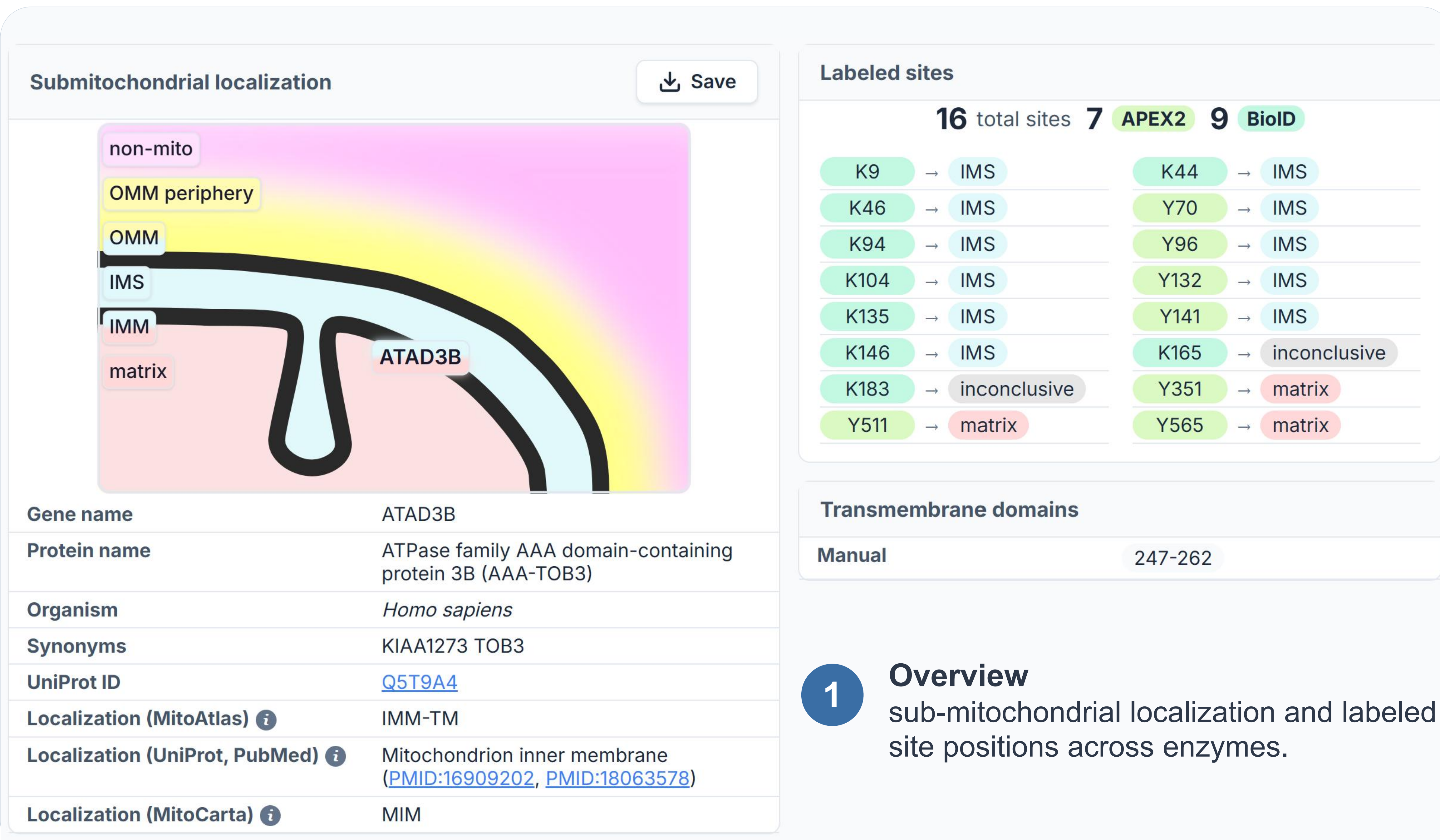


## K Absorbance vs Wavelength (nm)



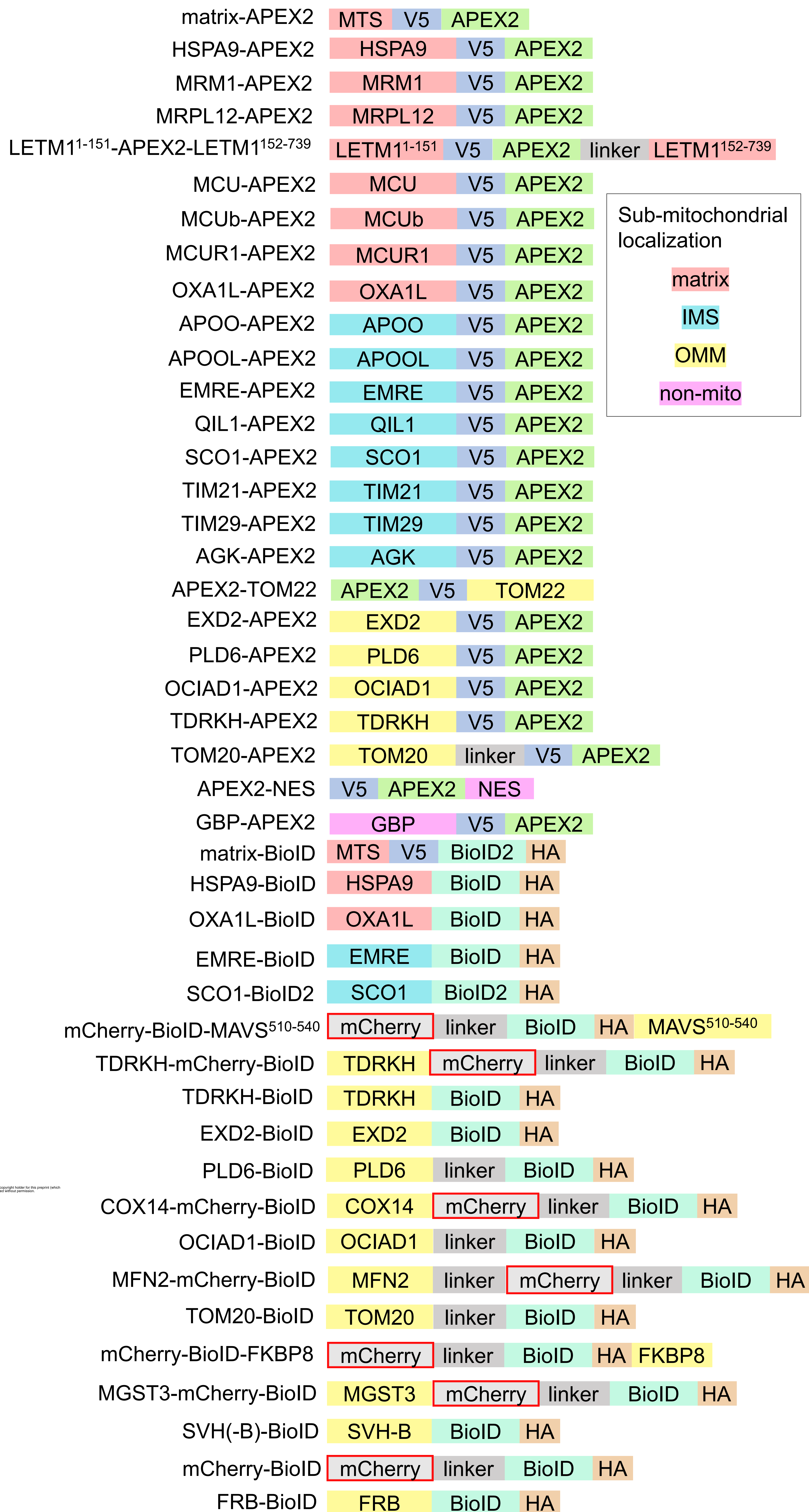
# Figure 7

## A MitoAtlas.org: Protein search and discovery Example: ATAD3B



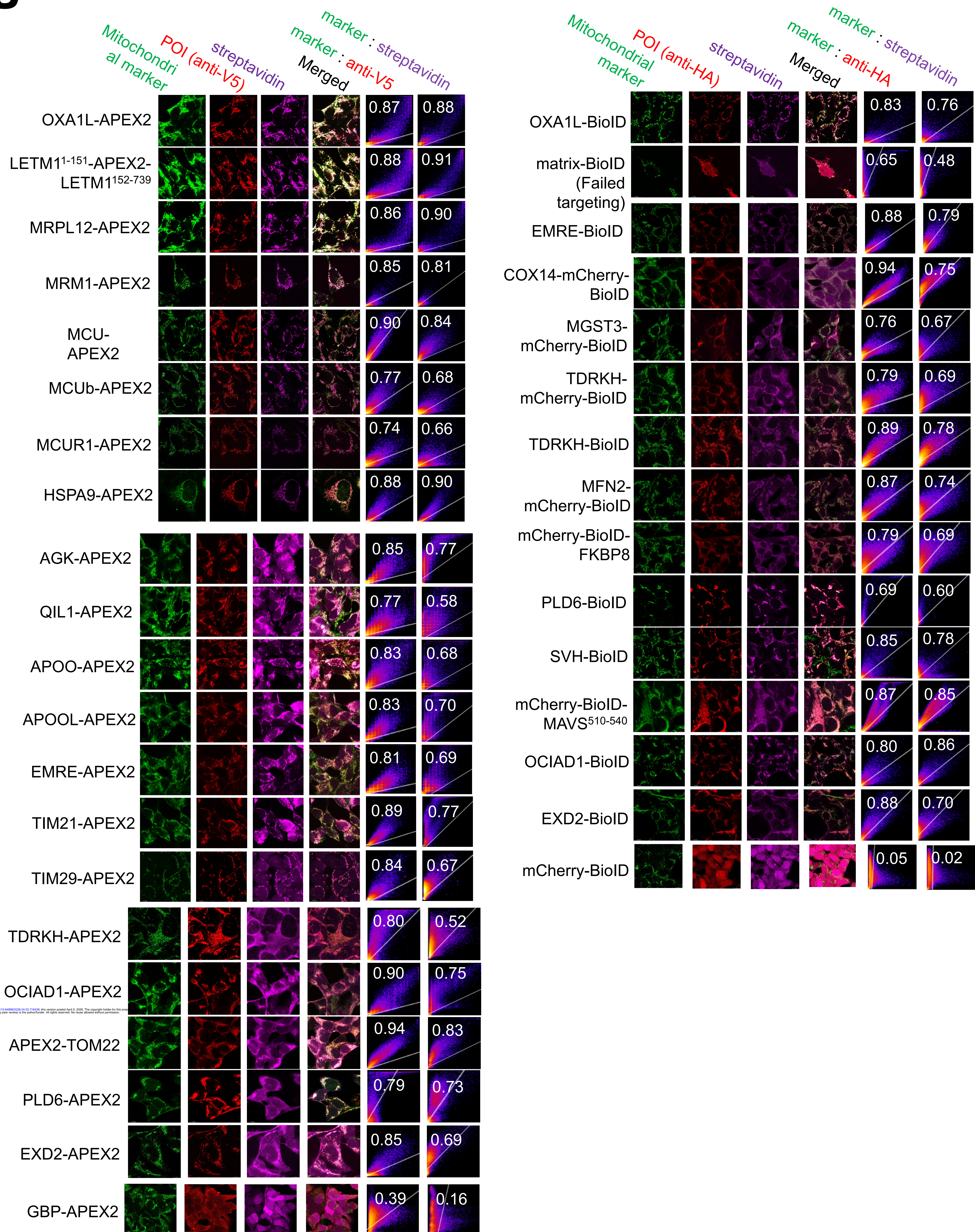
# Figure S1

## A



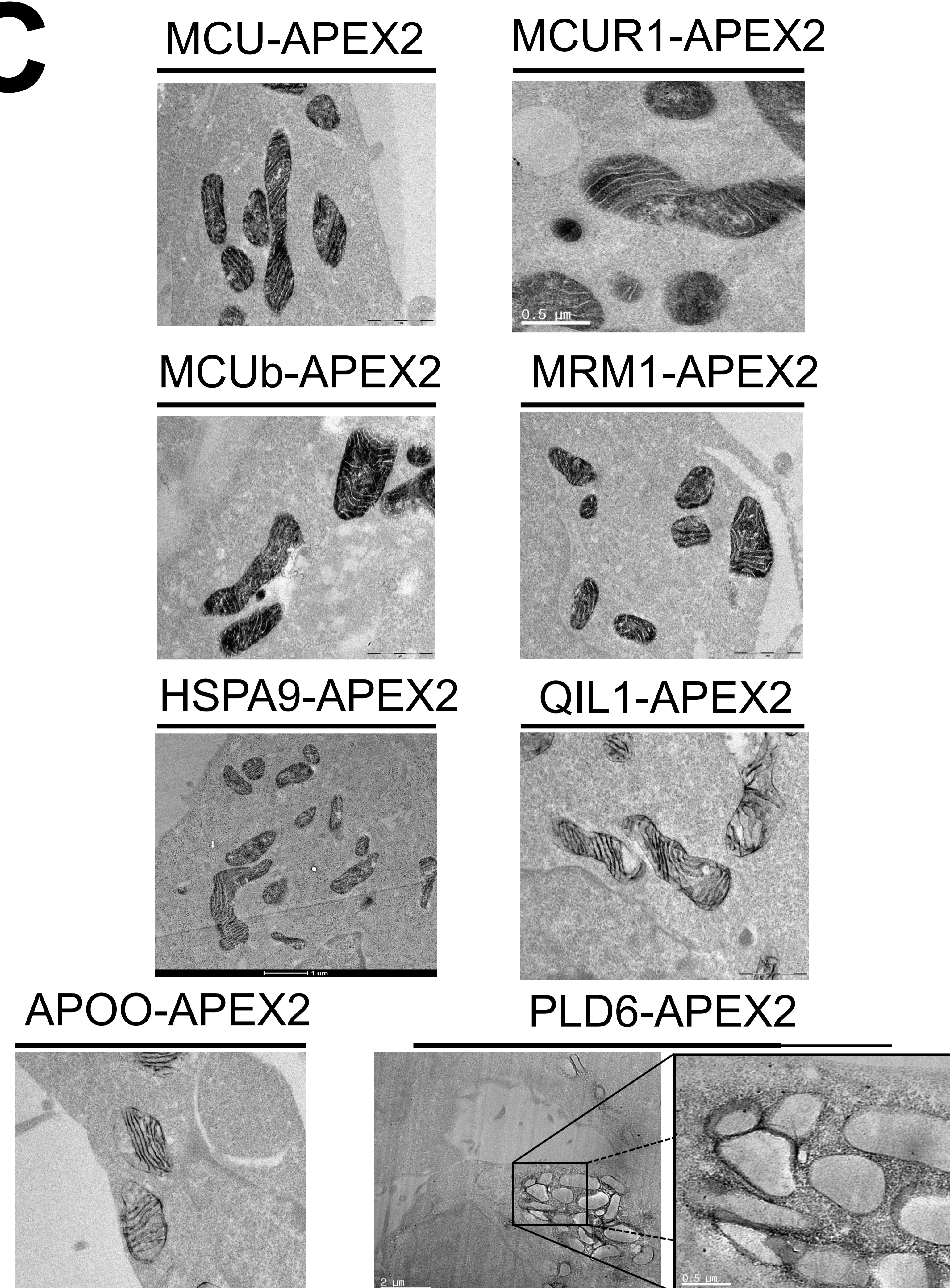
# Figure S1

## B

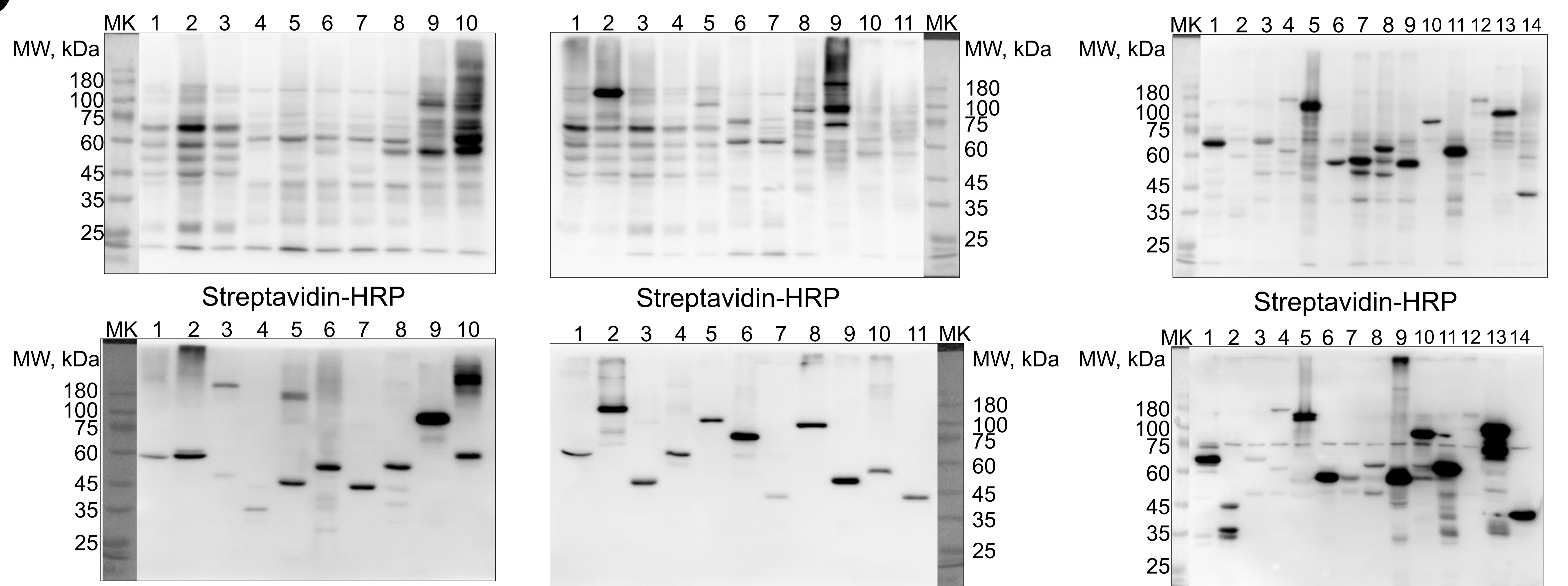


# Figure S1

## C



## D



Anti-V5

Lanes

- 1: MCUb-APEX2
- 2: MCU-APEX2
- 3: MCUR1-APEX2
- 4: EMRE-APEX2
- 5: TIM21-APEX2
- 6: TIM29-APEX2
- 7: APOO-APEX2
- 8: APOOL-APEX2
- 9: TDRKH-APEX2
- 10: OCIAD1-APEX2

Anti-V5

Lanes

- 1: OXA1L-APEX2
- 2: LETM1<sup>11-151</sup>-APEX2-  
LETM1<sup>152-739</sup>
- 3: MRPL12-APEX2
- 4: MRM1-APEX2
- 5: HSPA9-APEX2
- 6: AGK-APEX2
- 7: QIL1-APEX2
- 8: EXD2-APEX2
- 9: APEX2-TOM22
- 10: PLD6-APEX2
- 11: GBP-APEX2

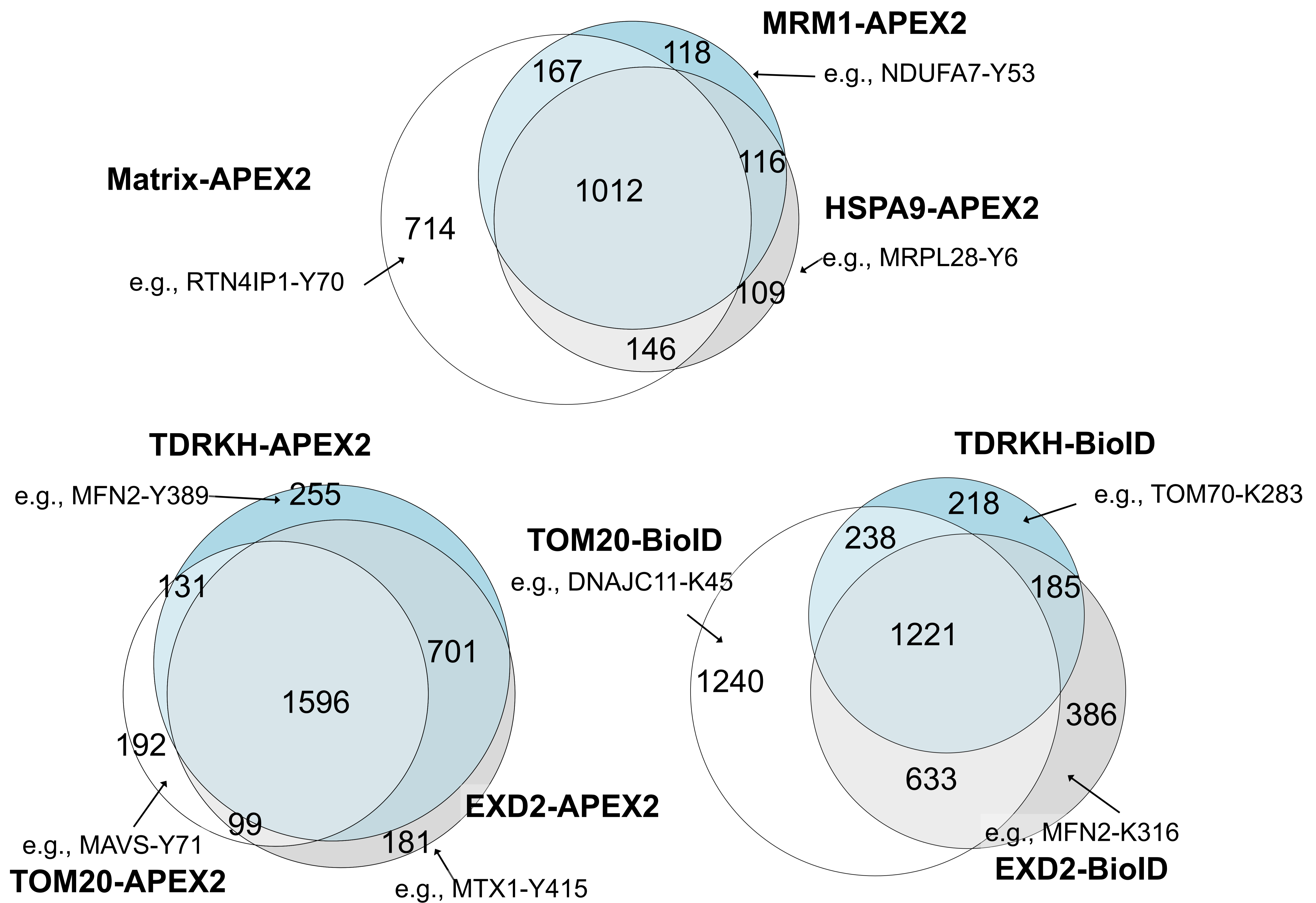
Anti-HA

Lanes

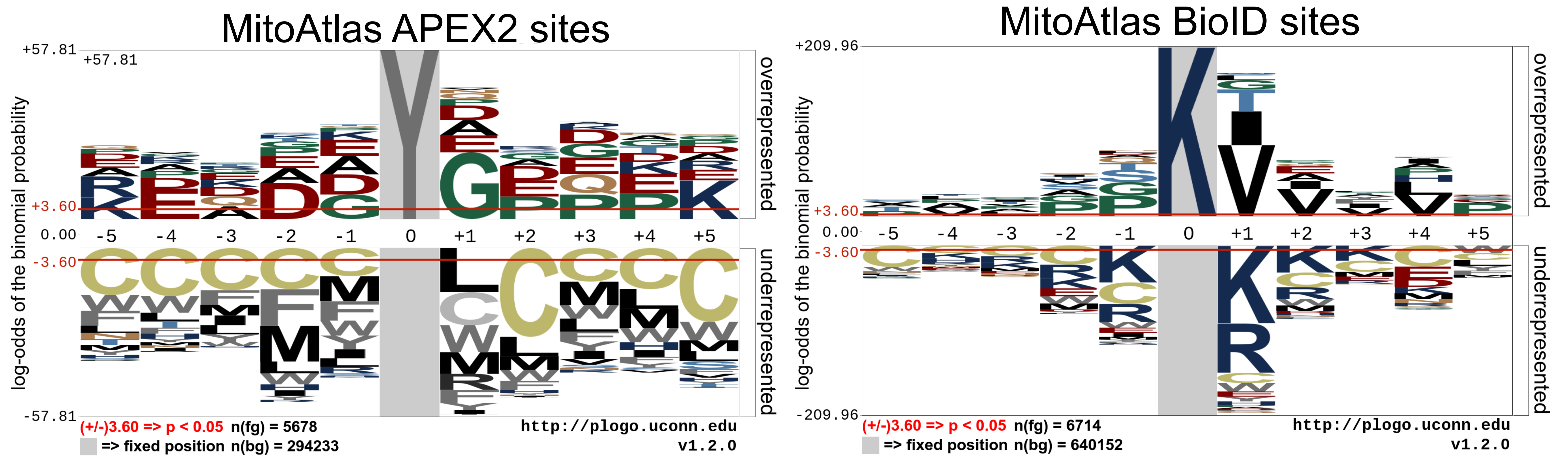
- 1: OXA1L-BioID
- 2: EMRE-BioID
- 3: MGST3-mCherry-BioID
- 4: MFN2-mCherry-BioID
- 5: mCherry-BioID-FKBP8
- 6: SVH(-B)-BioID
- 7: mCherry-BioID-MAVS<sup>510-540</sup>
- 8: COX14-mCherry-BioID
- 9: PLD6-BioID
- 10: EXD2-BioID
- 11: OCIAD1-BioID
- 12: TDRKH-mCherry-BioID
- 13: TDRKH-BioID
- 14: FRB-BioID

# Figure S1

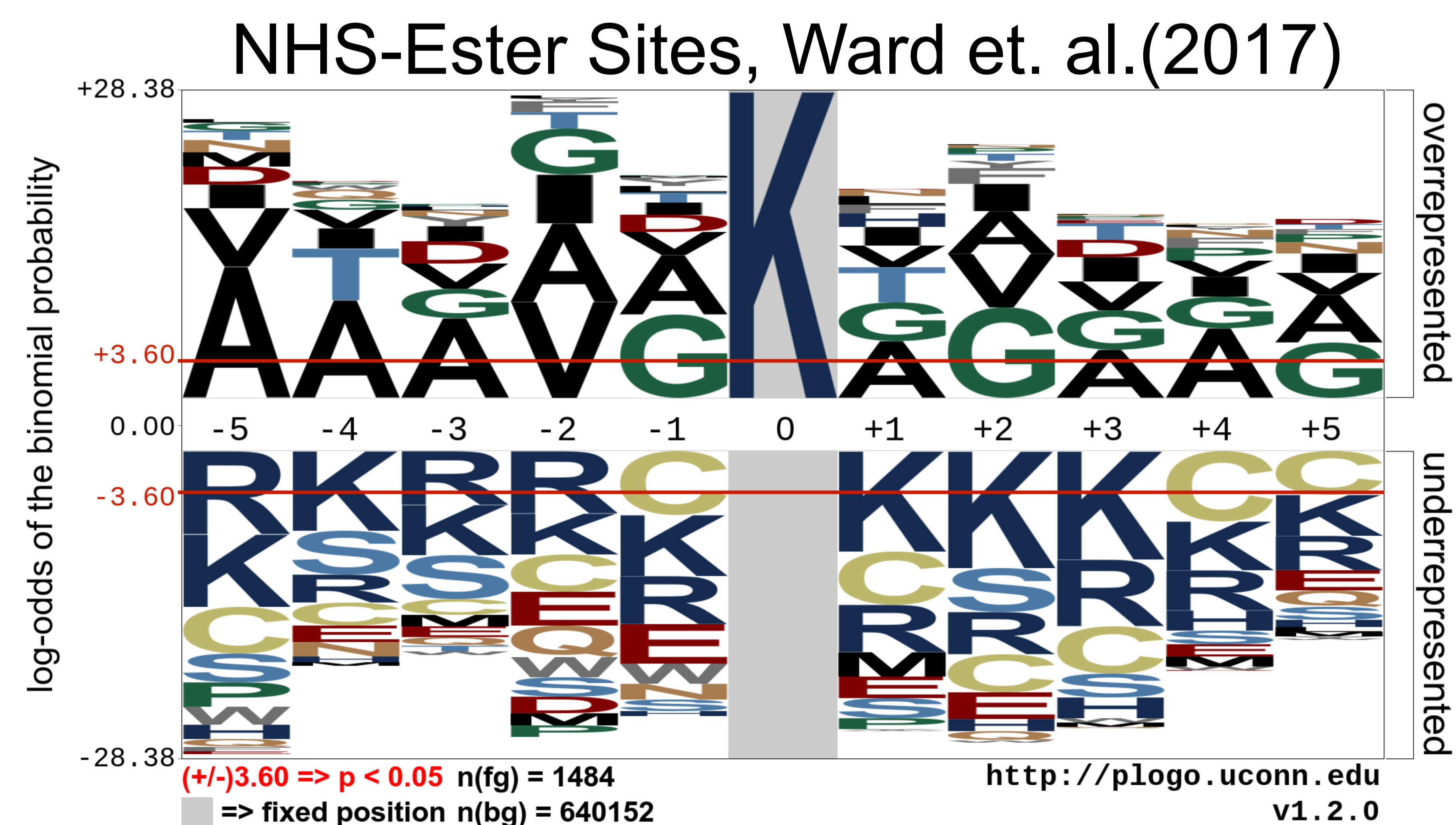
## E



## F

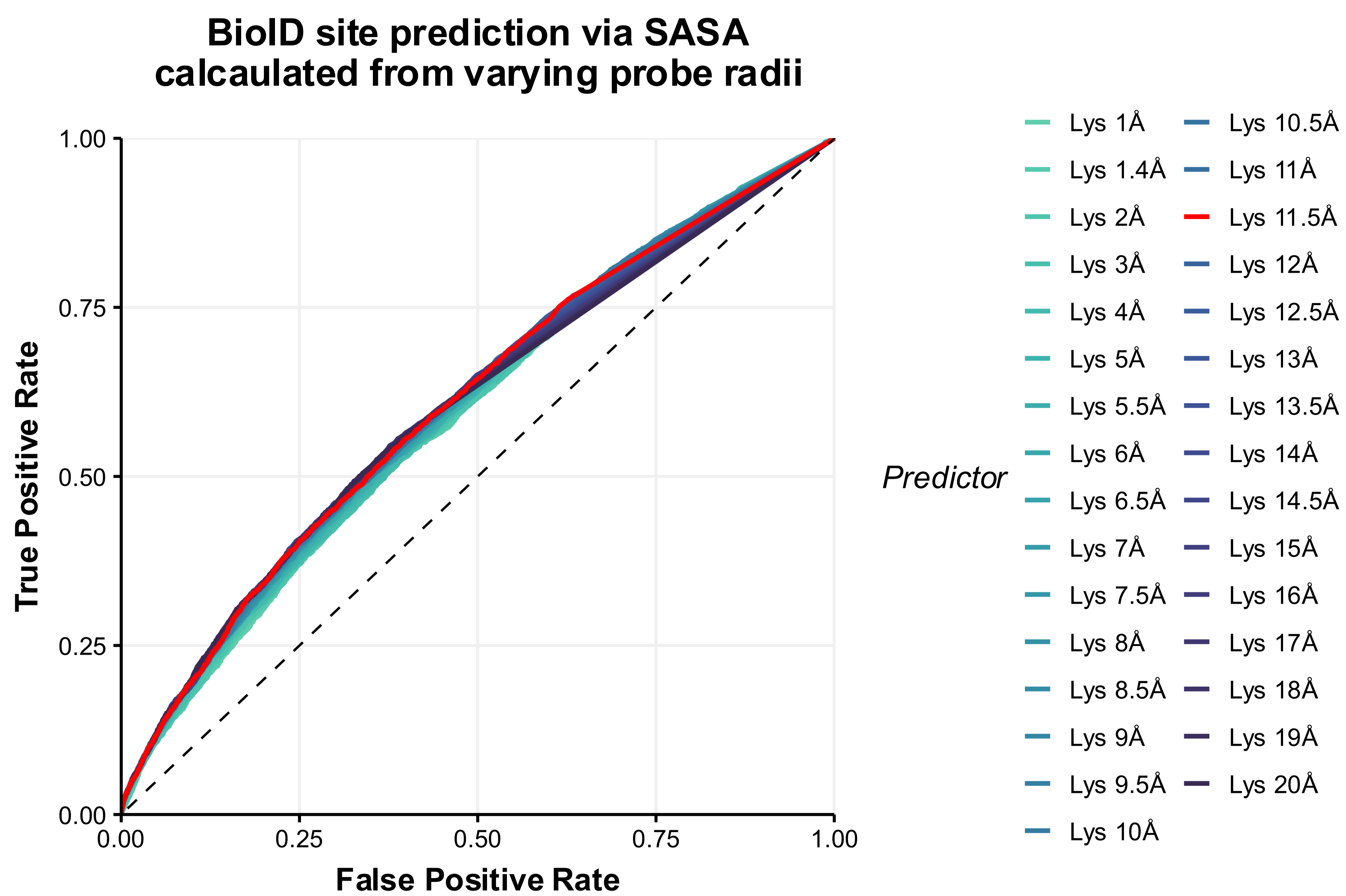
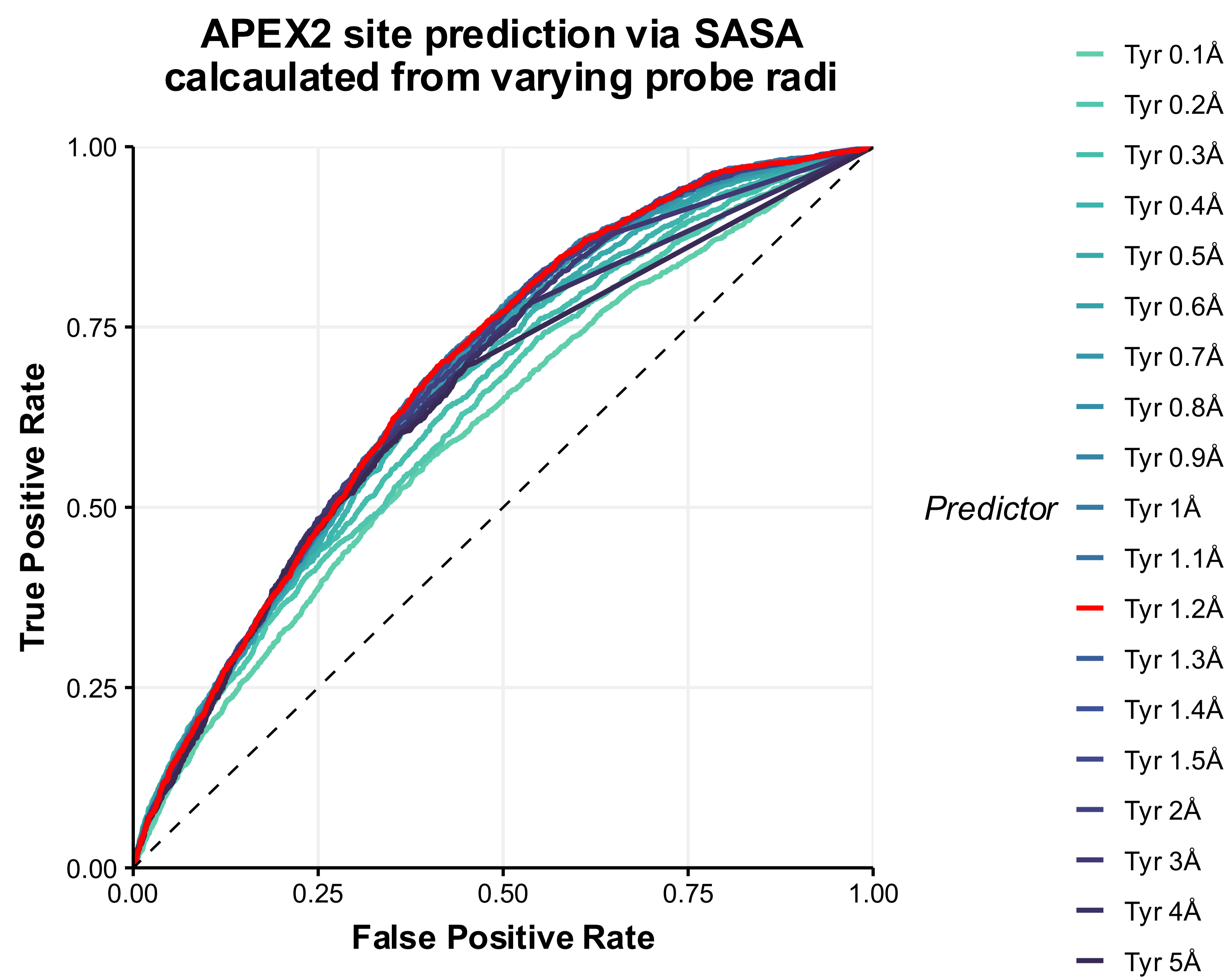


bioRxiv preprint doi: <https://doi.org/10.1101/2020.04.02.116430>; this version posted April 6, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

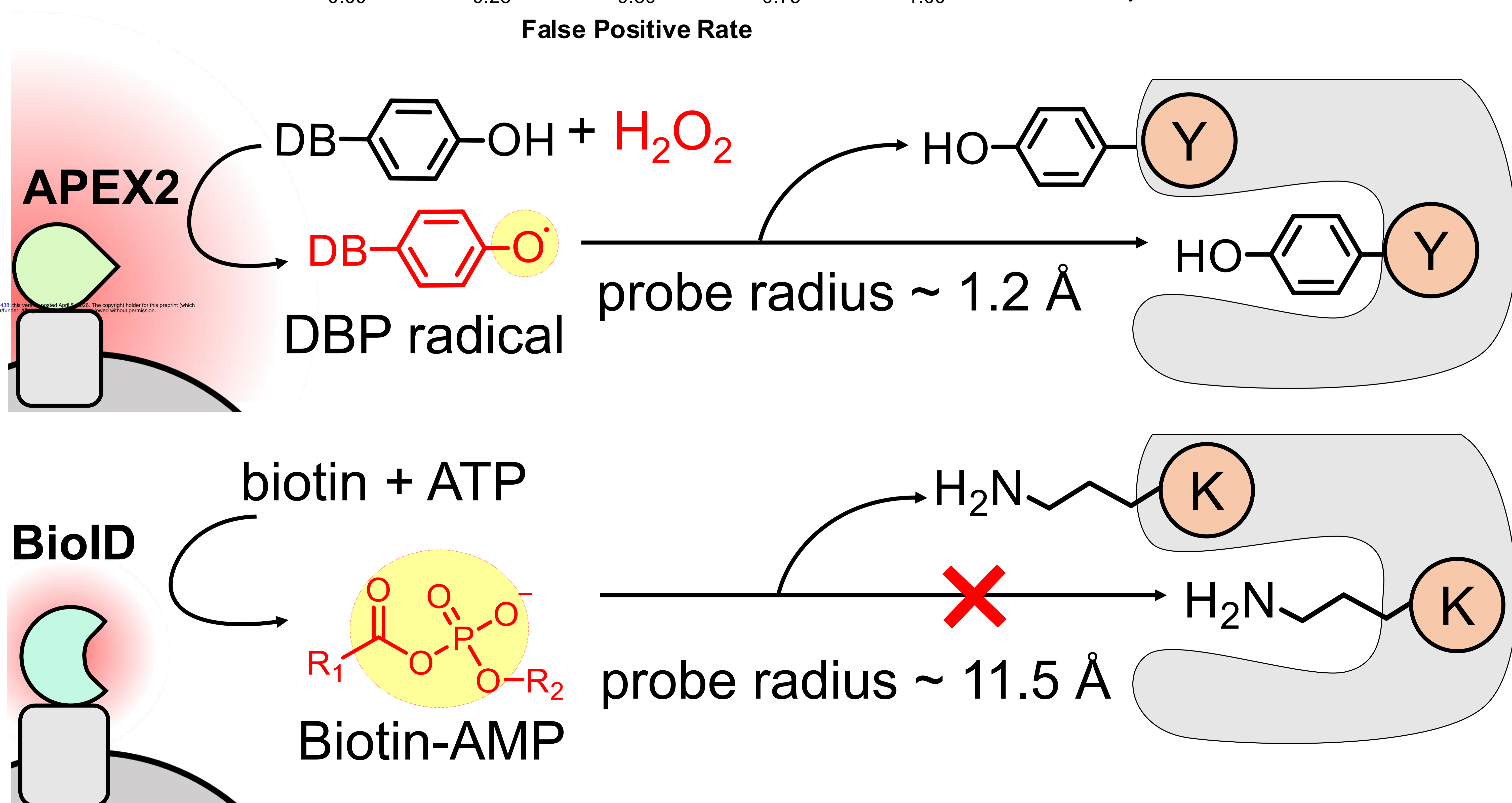


# Figure S1

## G

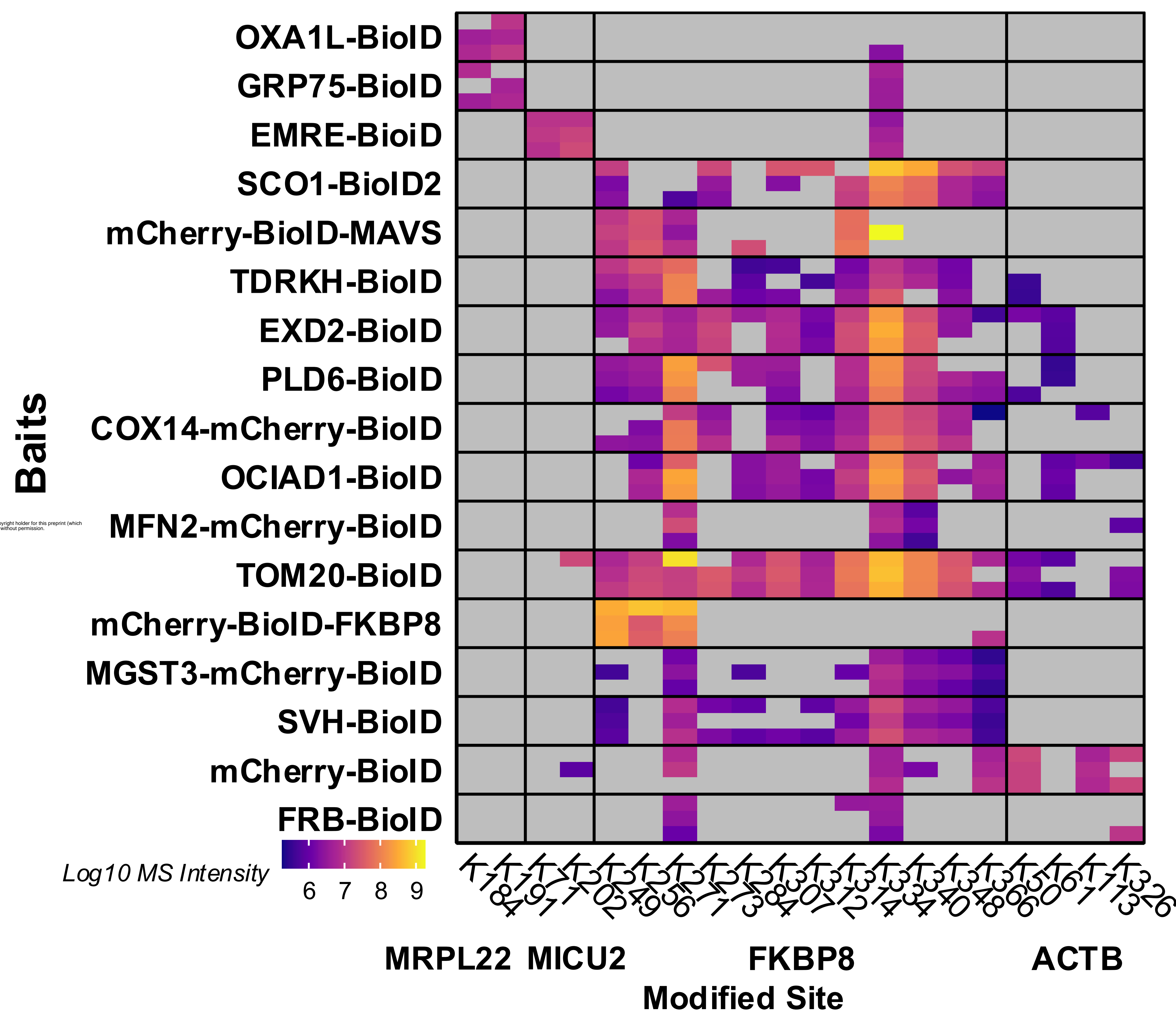
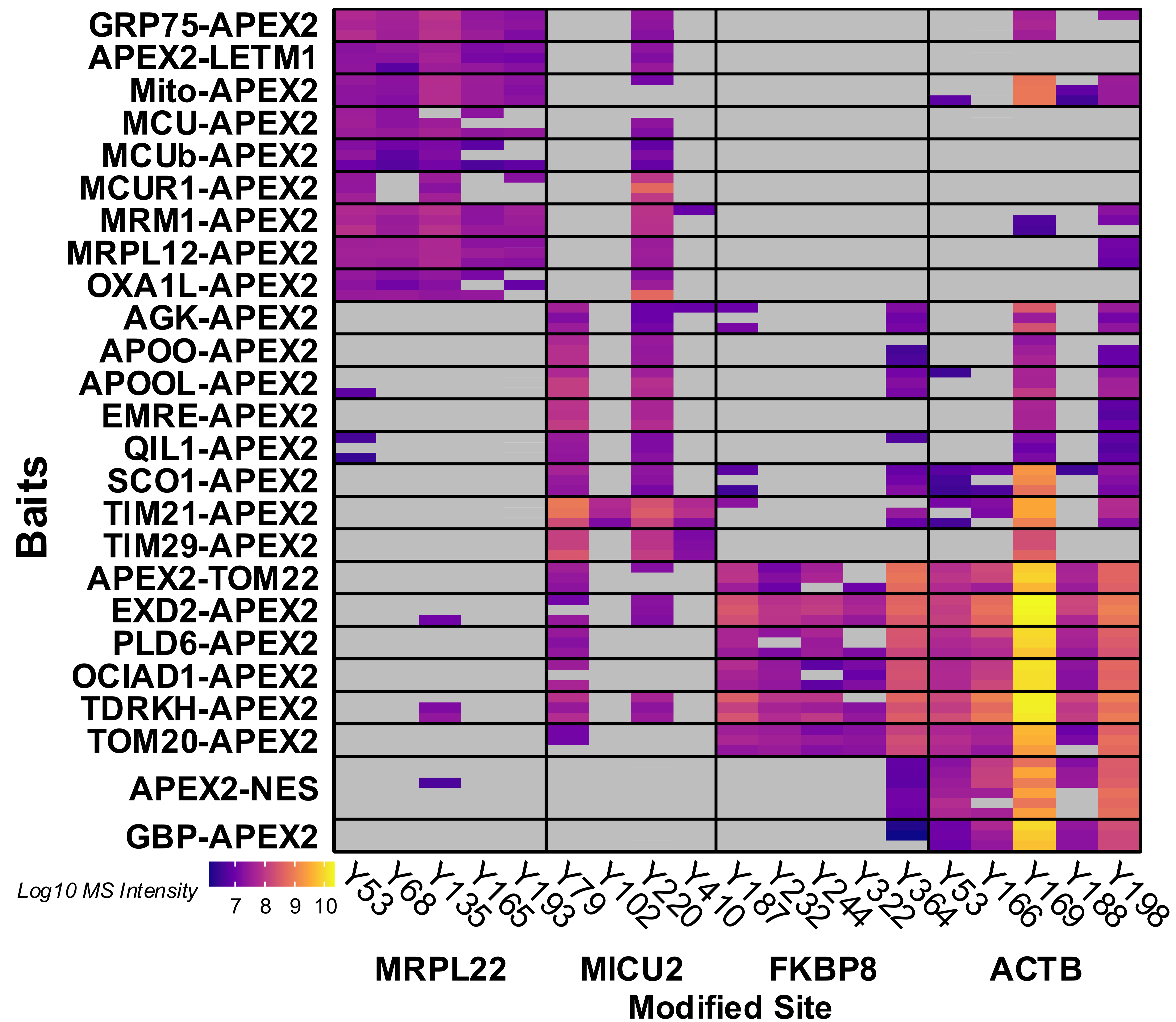


## H



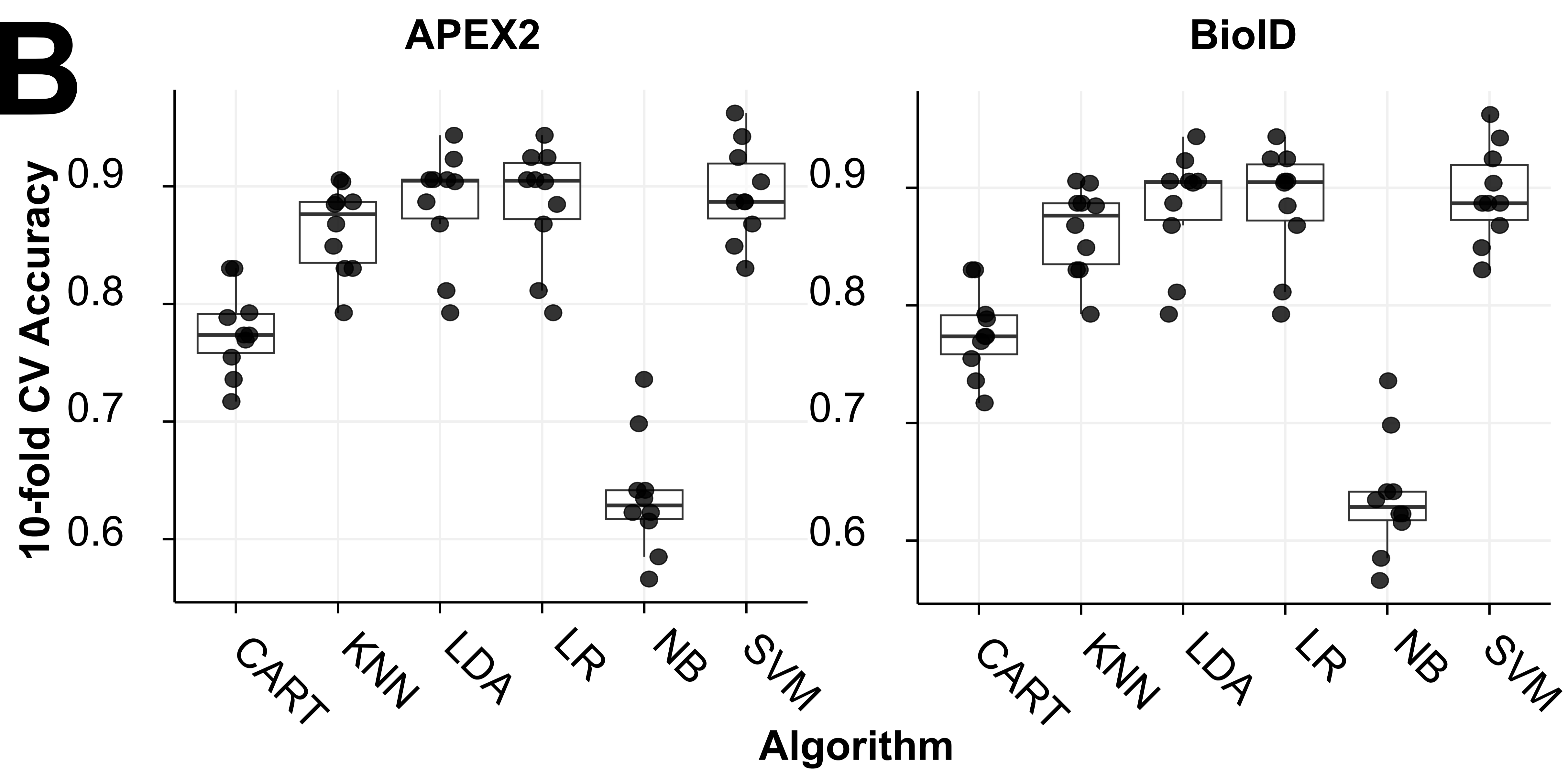
# Figure S2

## A

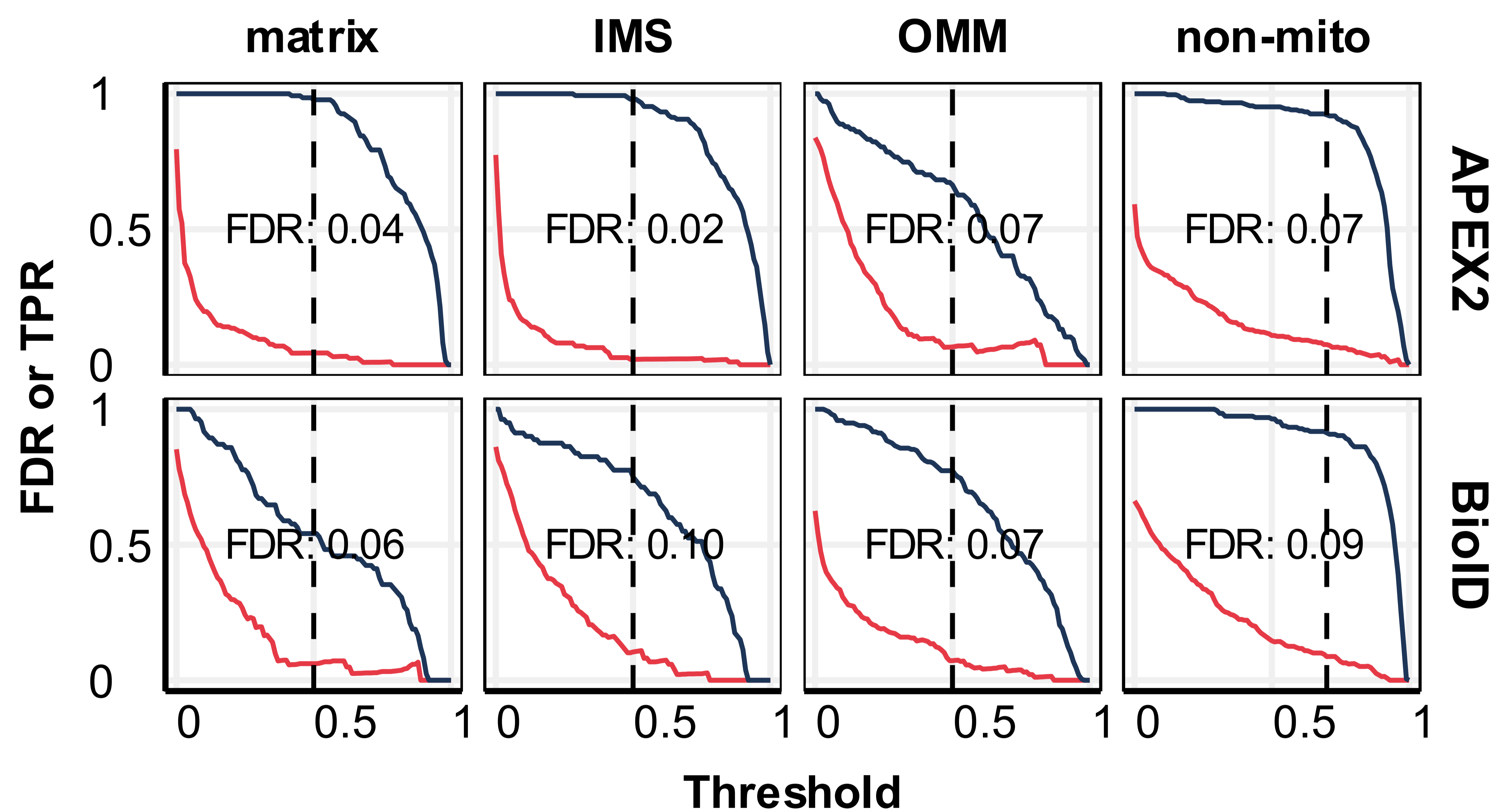


# Figure S2

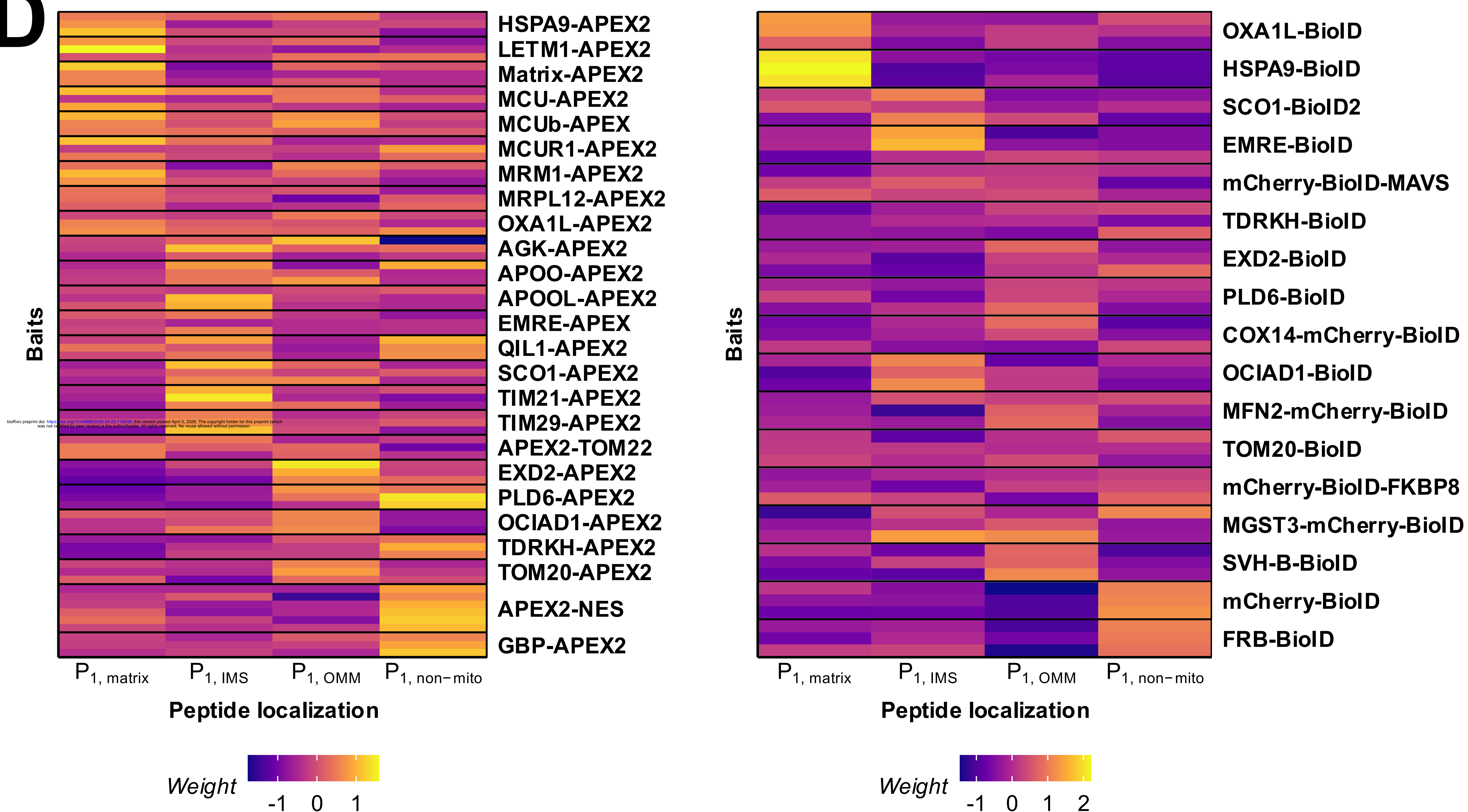
## B



## C



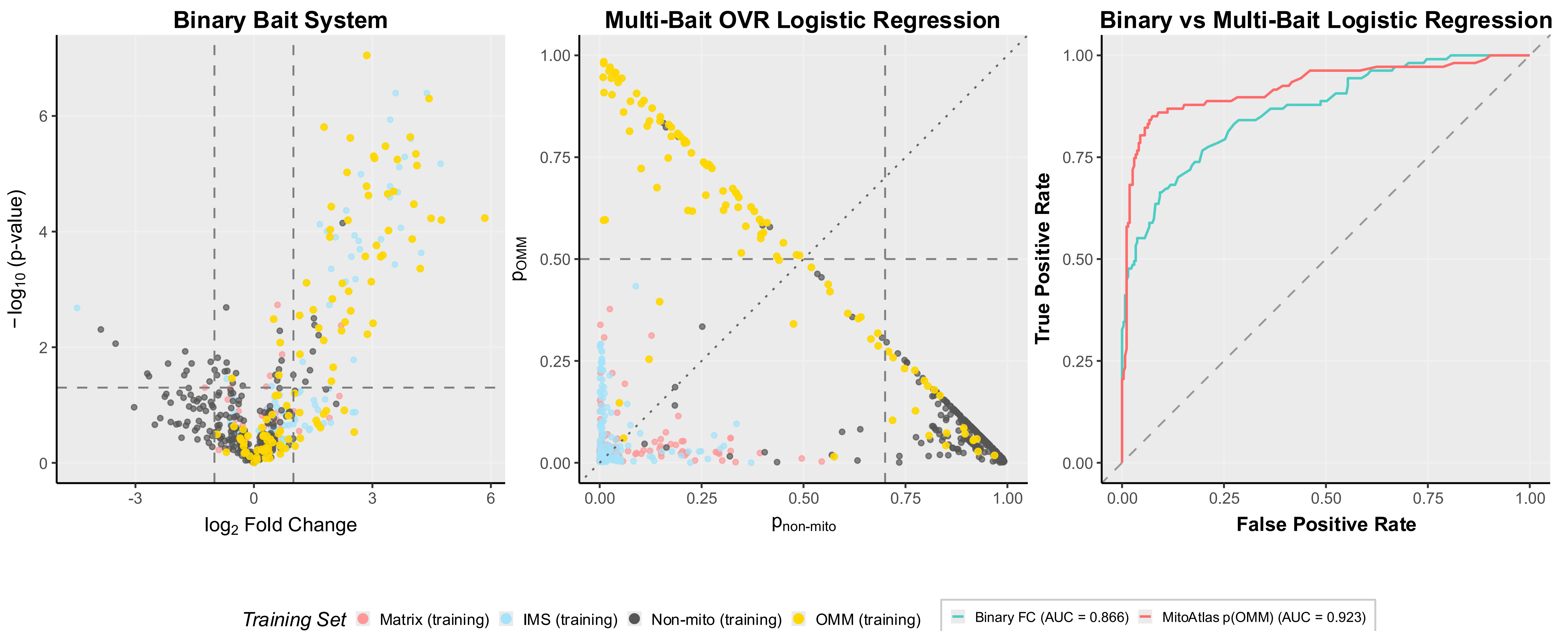
## D



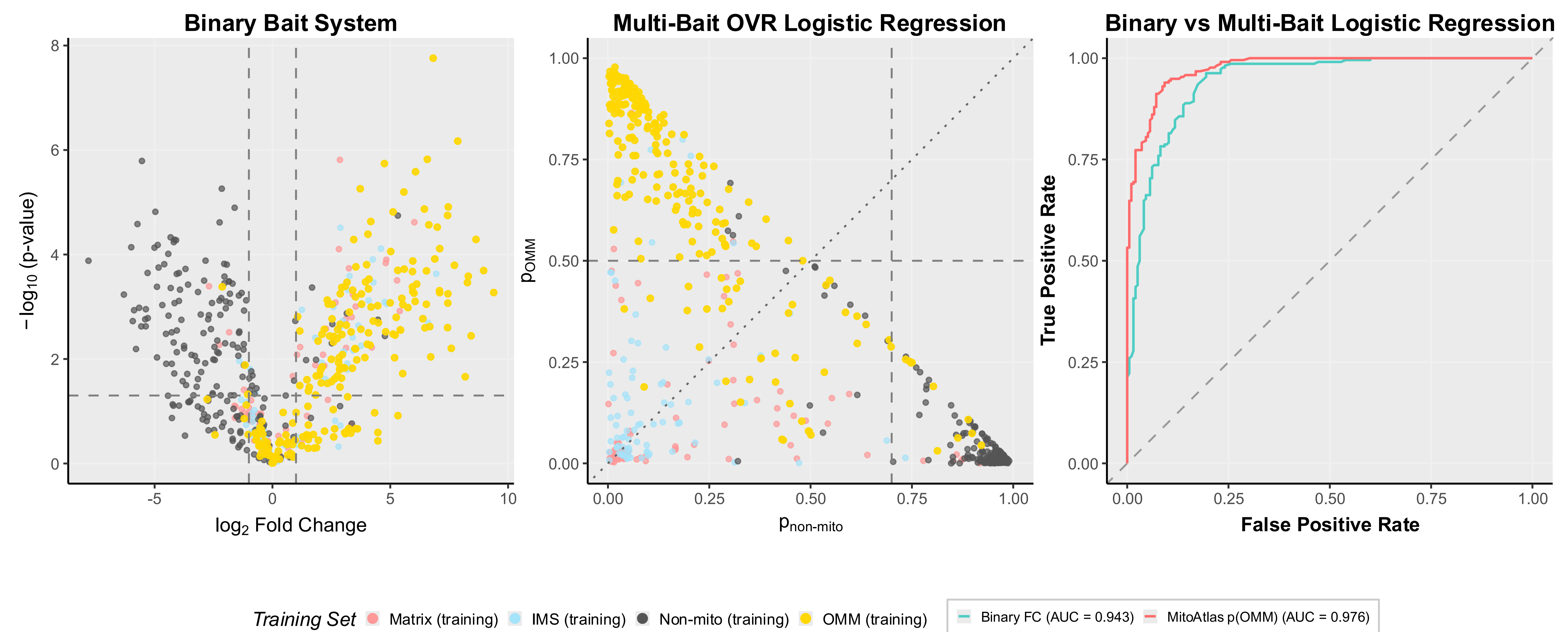
# Figure S2

## E

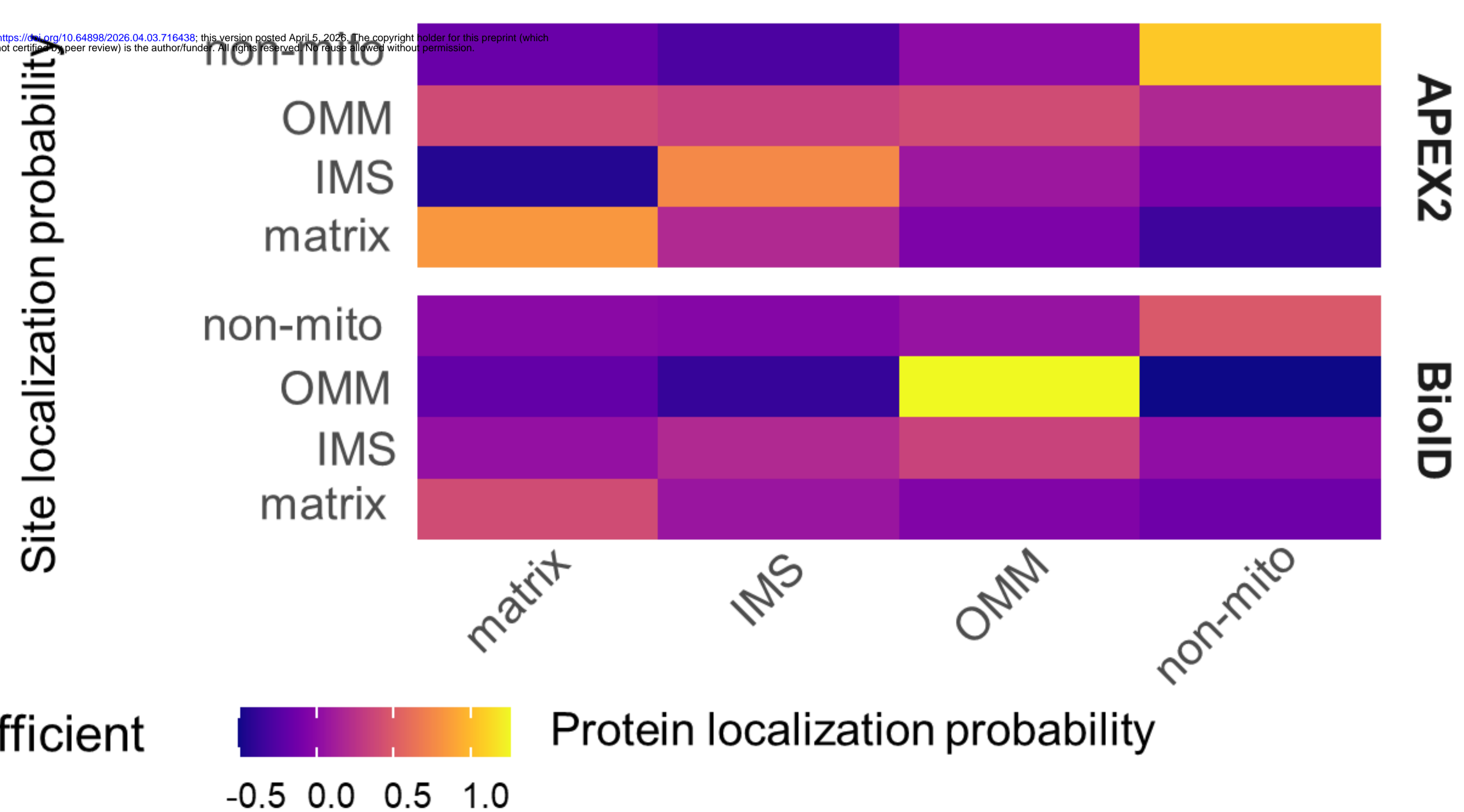
### APEX2 Site OMM and Non-mito Classification Performance Comparison



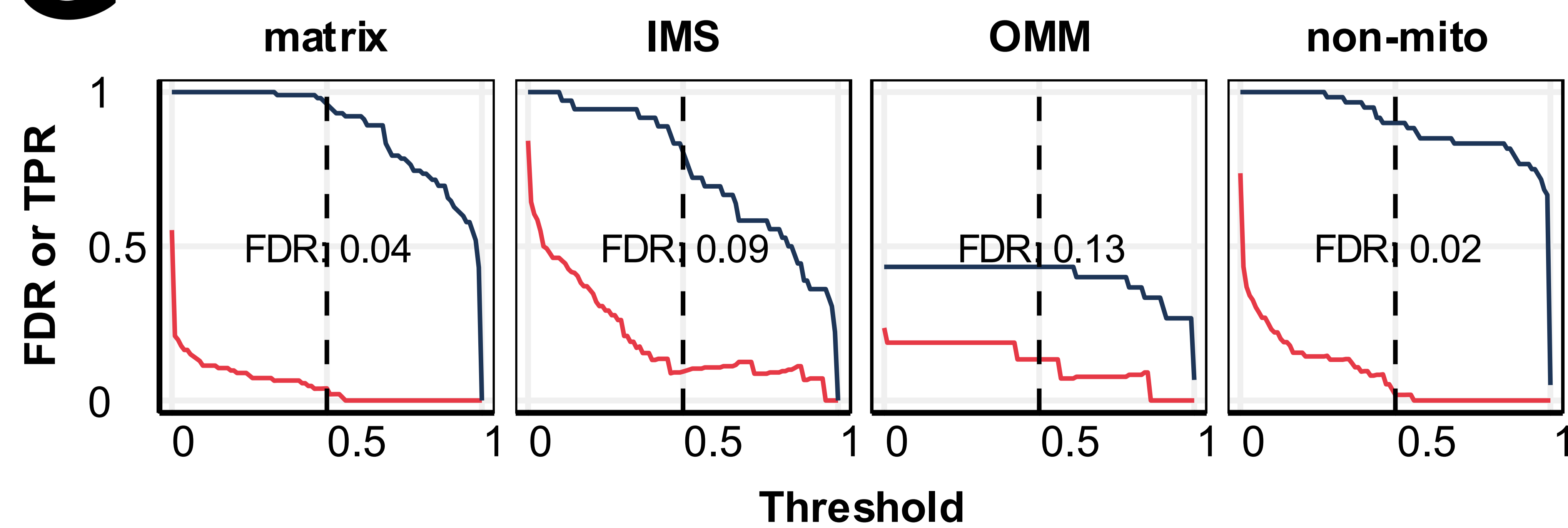
### Biold Site OMM and Non-mito Classification Performance Comparison



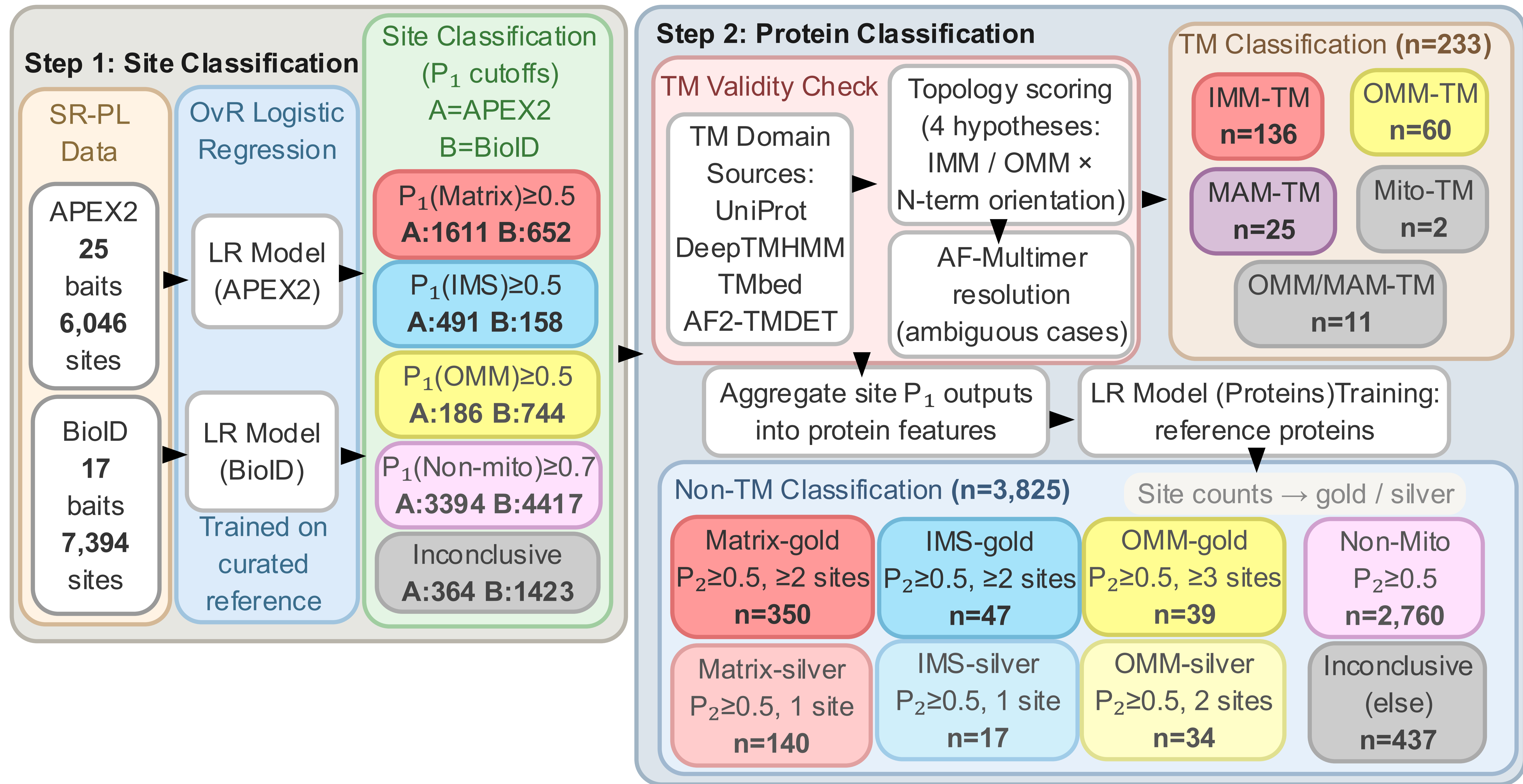
## F



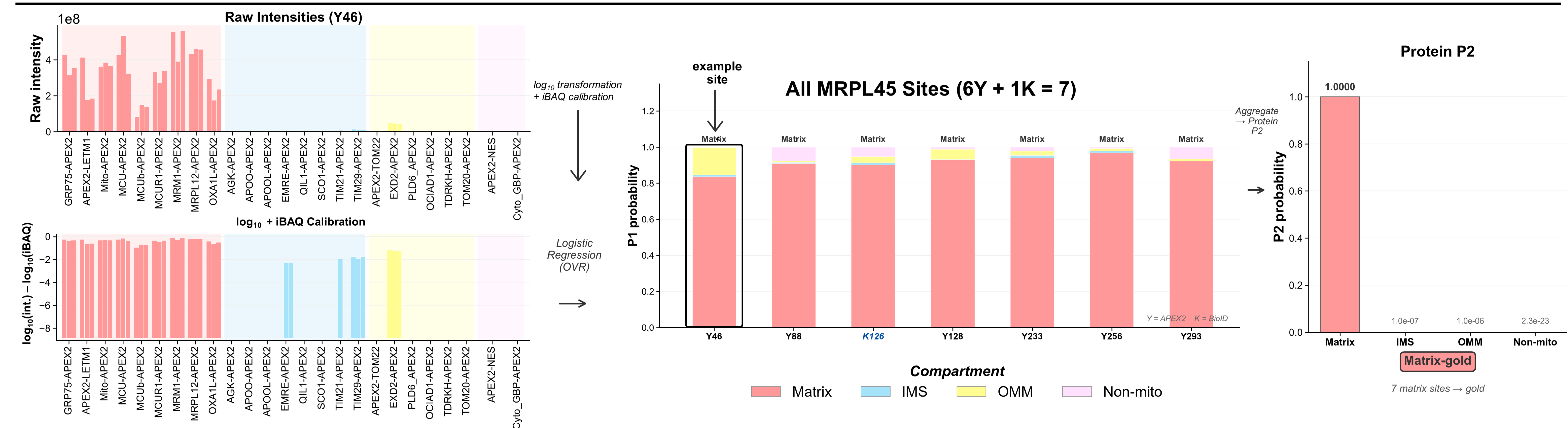
## G



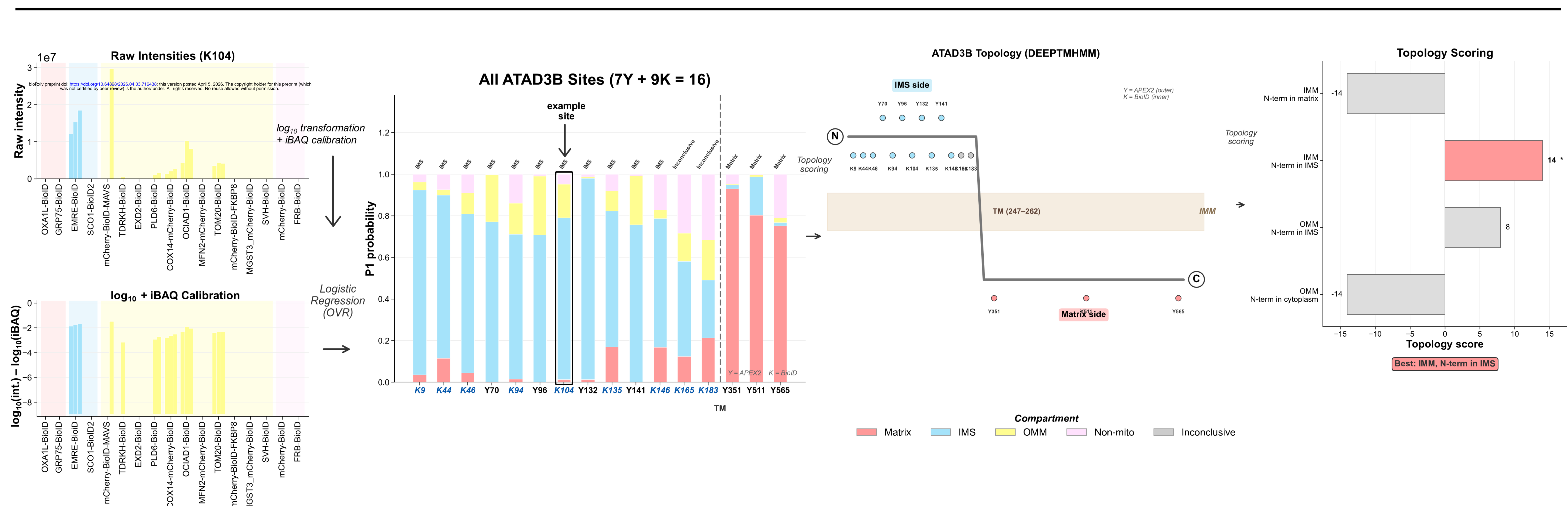
# Figure S2



**Example 1: Classification pipeline of MRPL45 (Matrix-gold)**



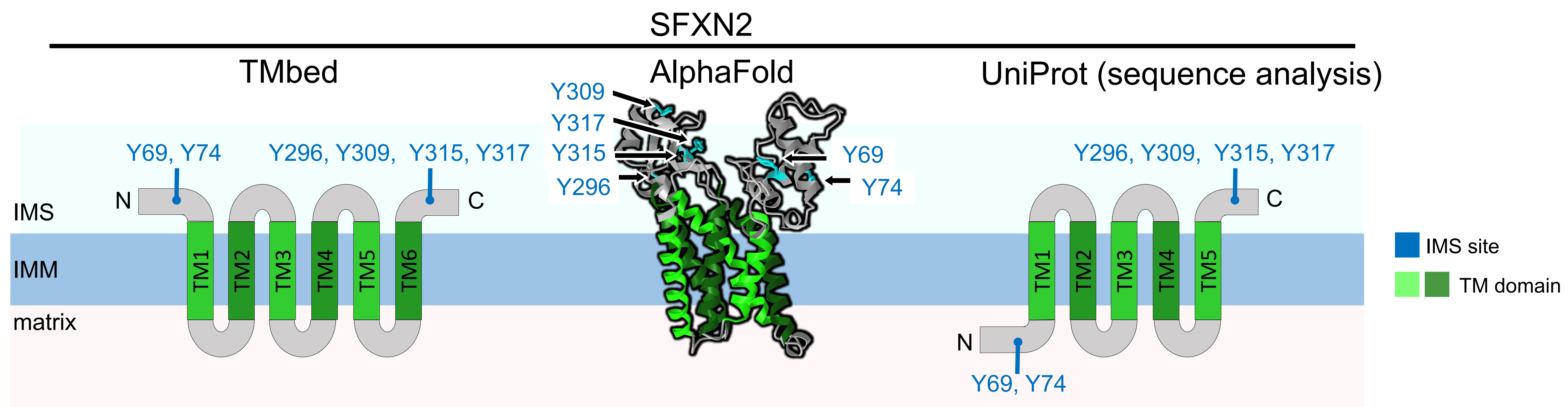
**Example 2: Classification pipeline of ATAD3B (IMM-TM)**





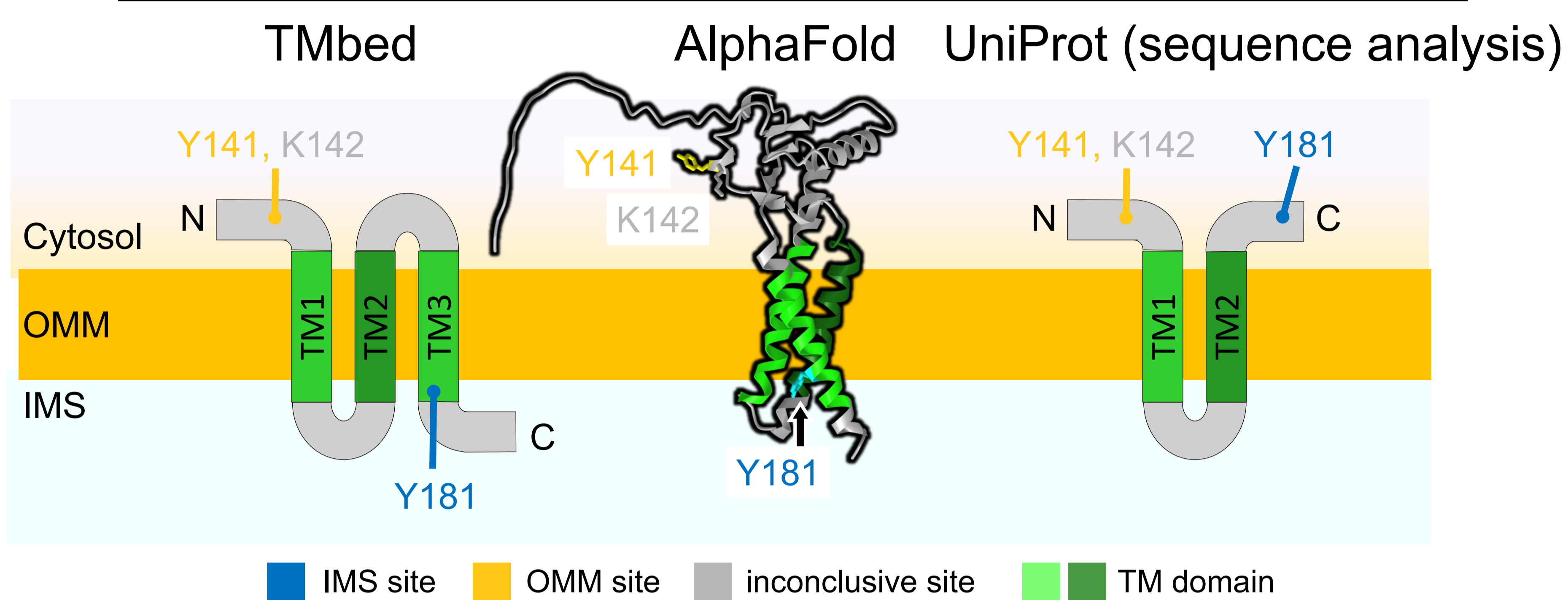
# Figure S3

## A



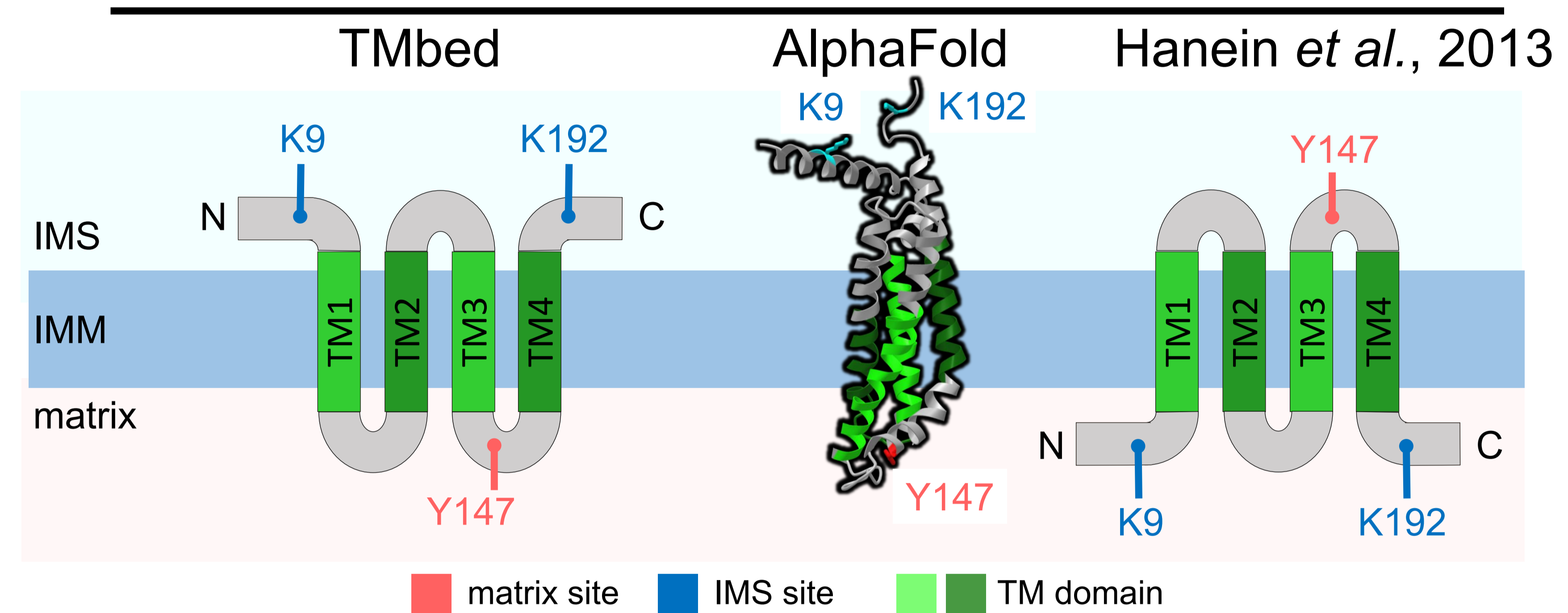
## B

**TMEM11**, Reported as IMM-TM by Rival *et al.*, 2011

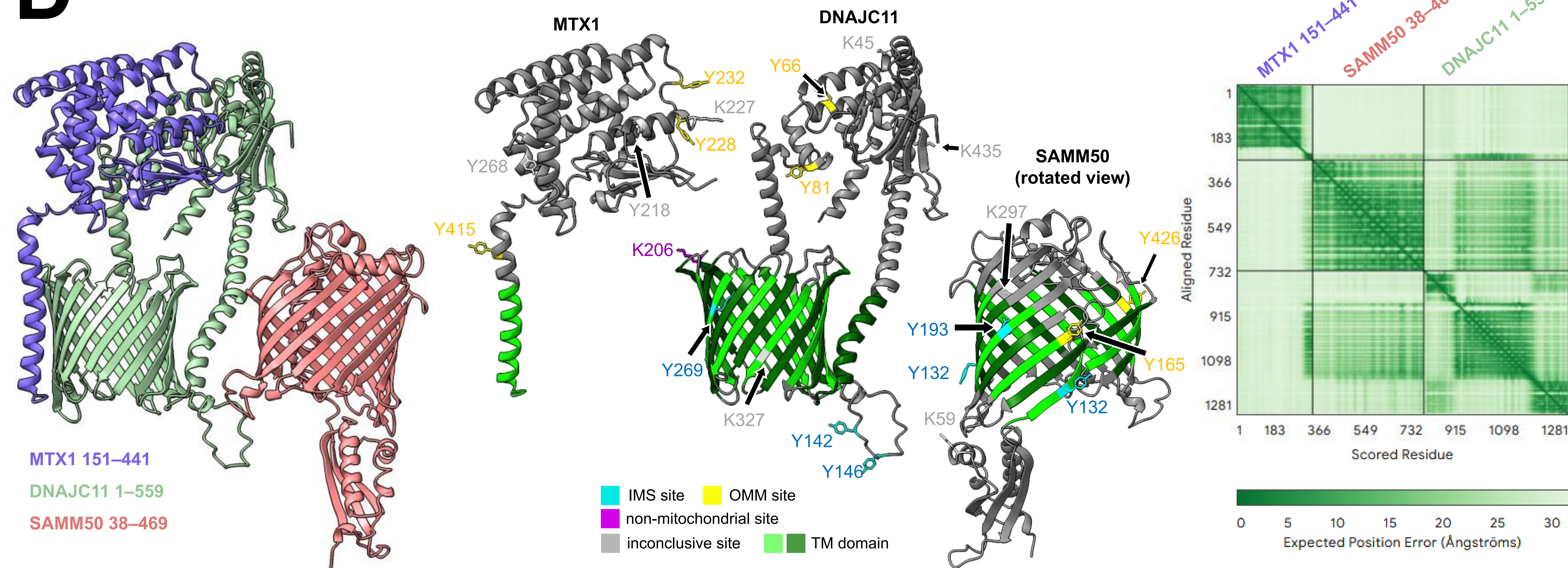


## C

**TMEM126A**

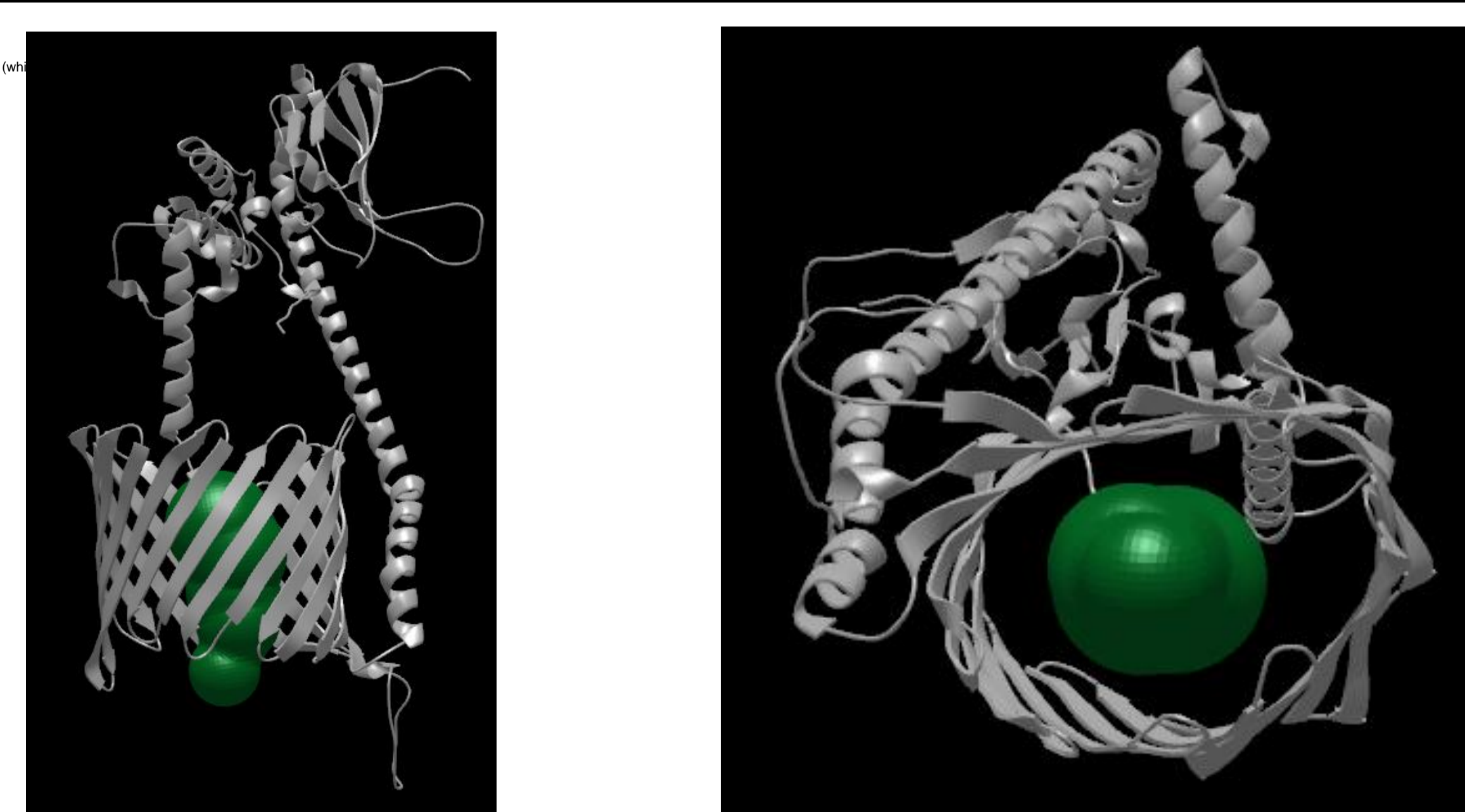


## D

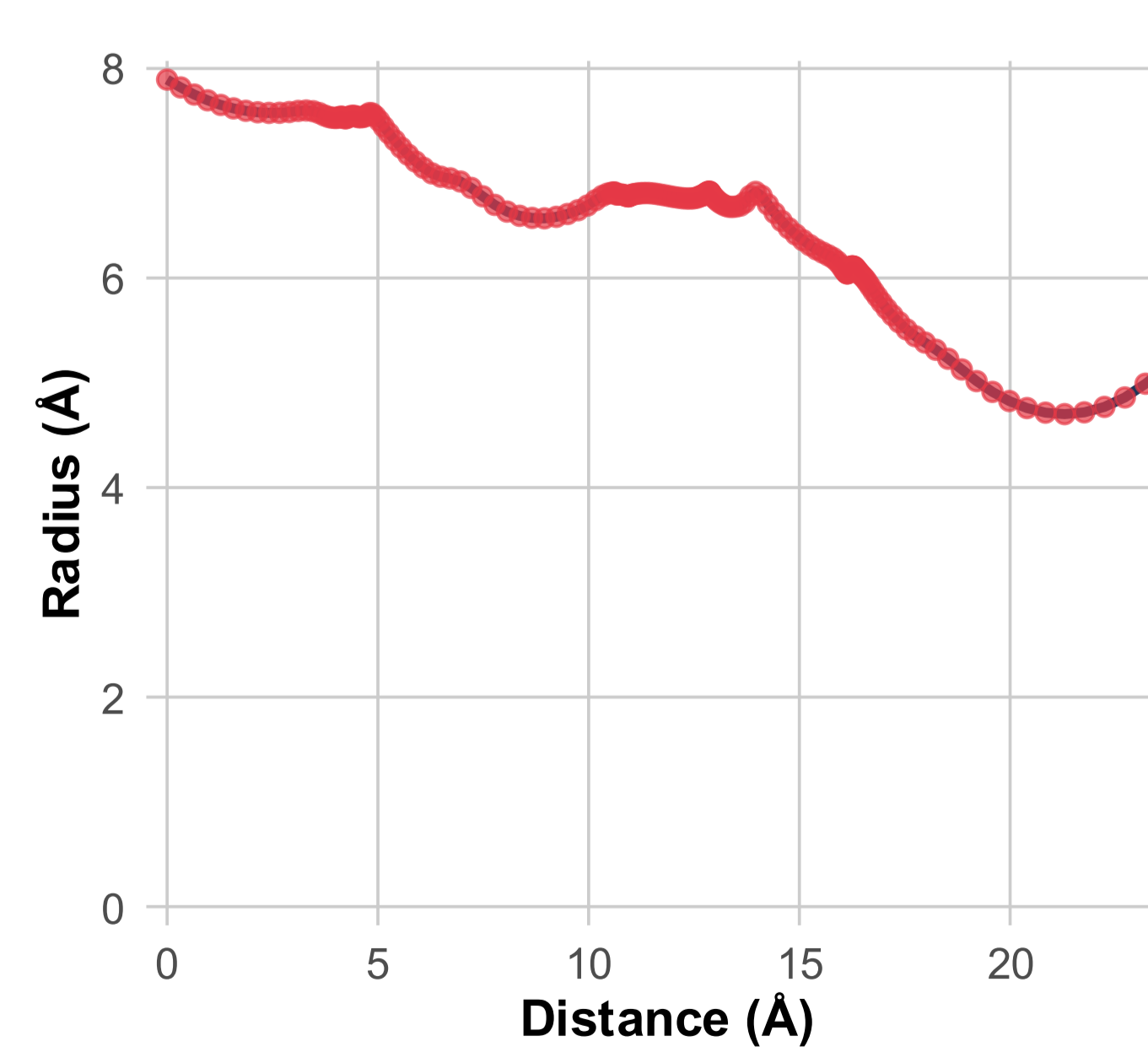


## E

**DNAJC11**

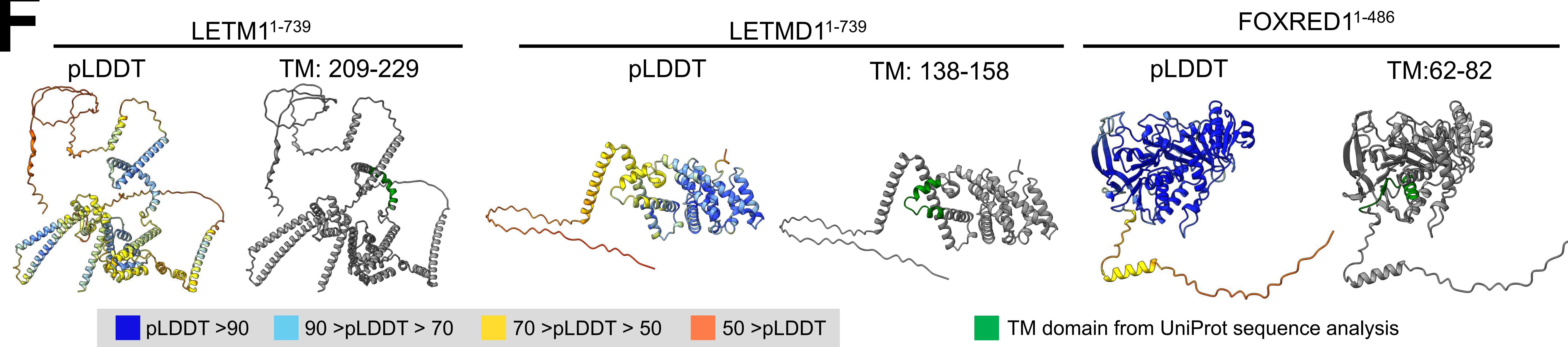


**Pore Radius Profile**



# Figure S3

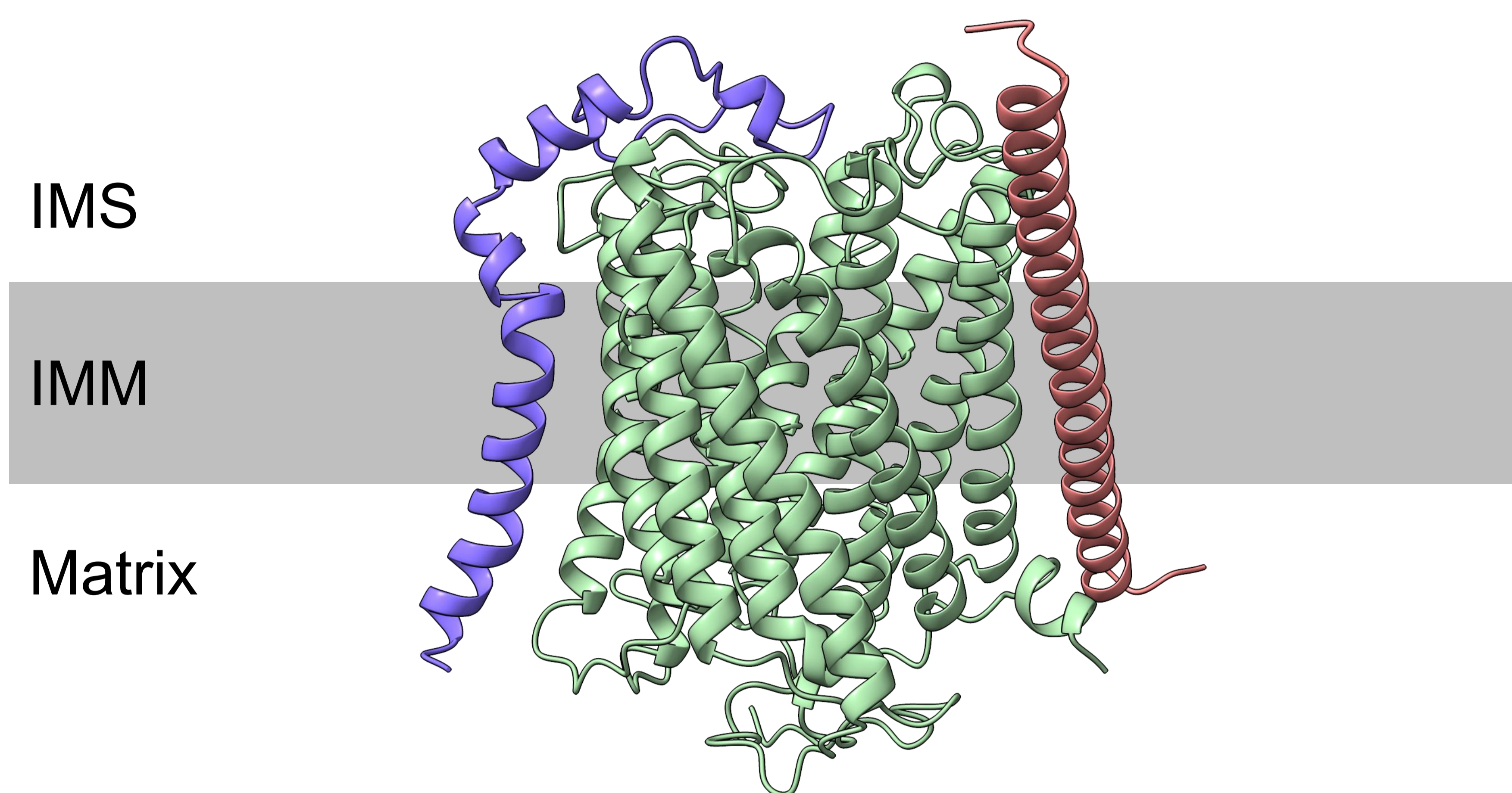
## F



## G

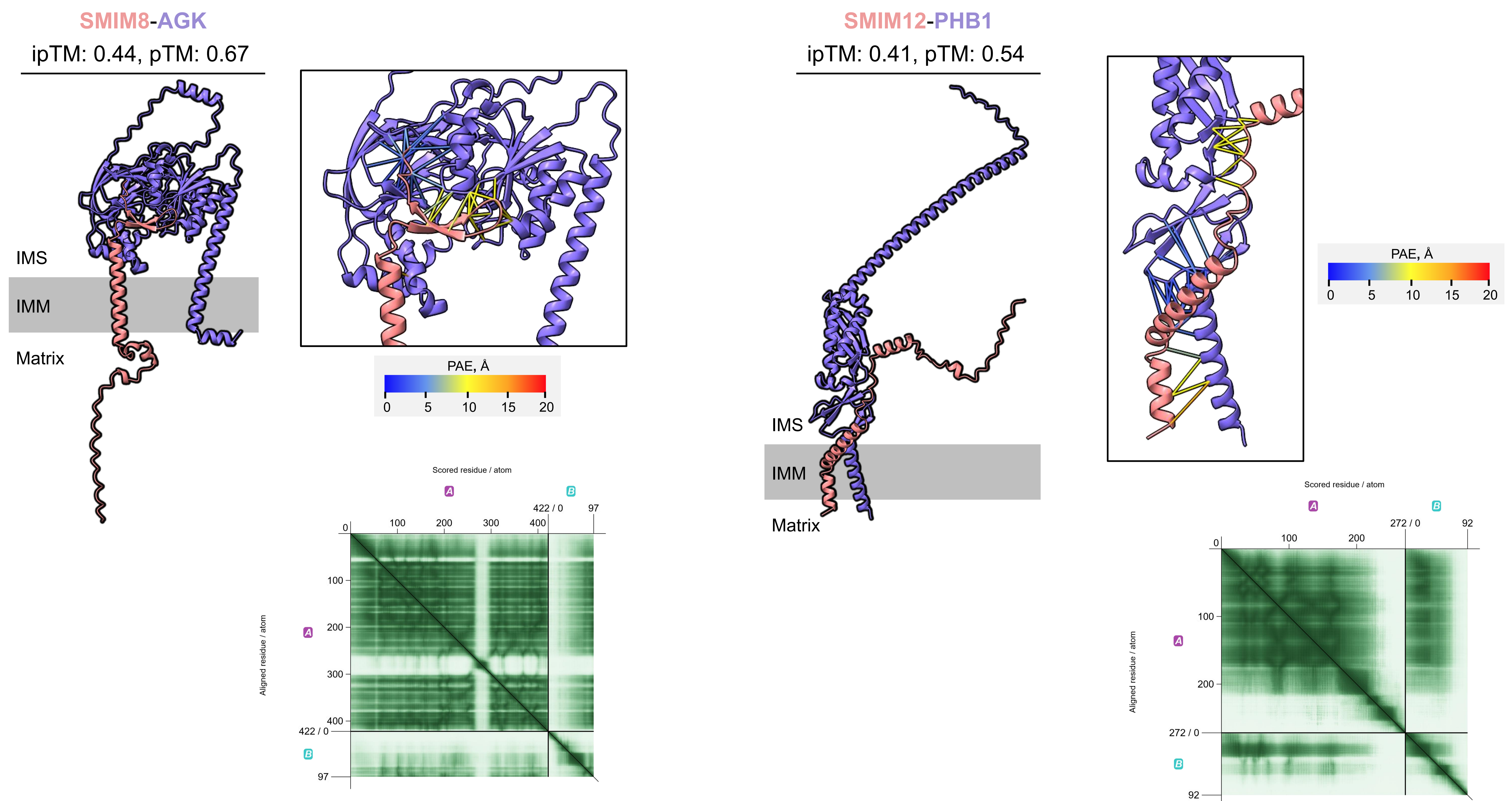
SMIM20-COX14-MT-CO1

ipTM: 0.69, pTM: 0.83

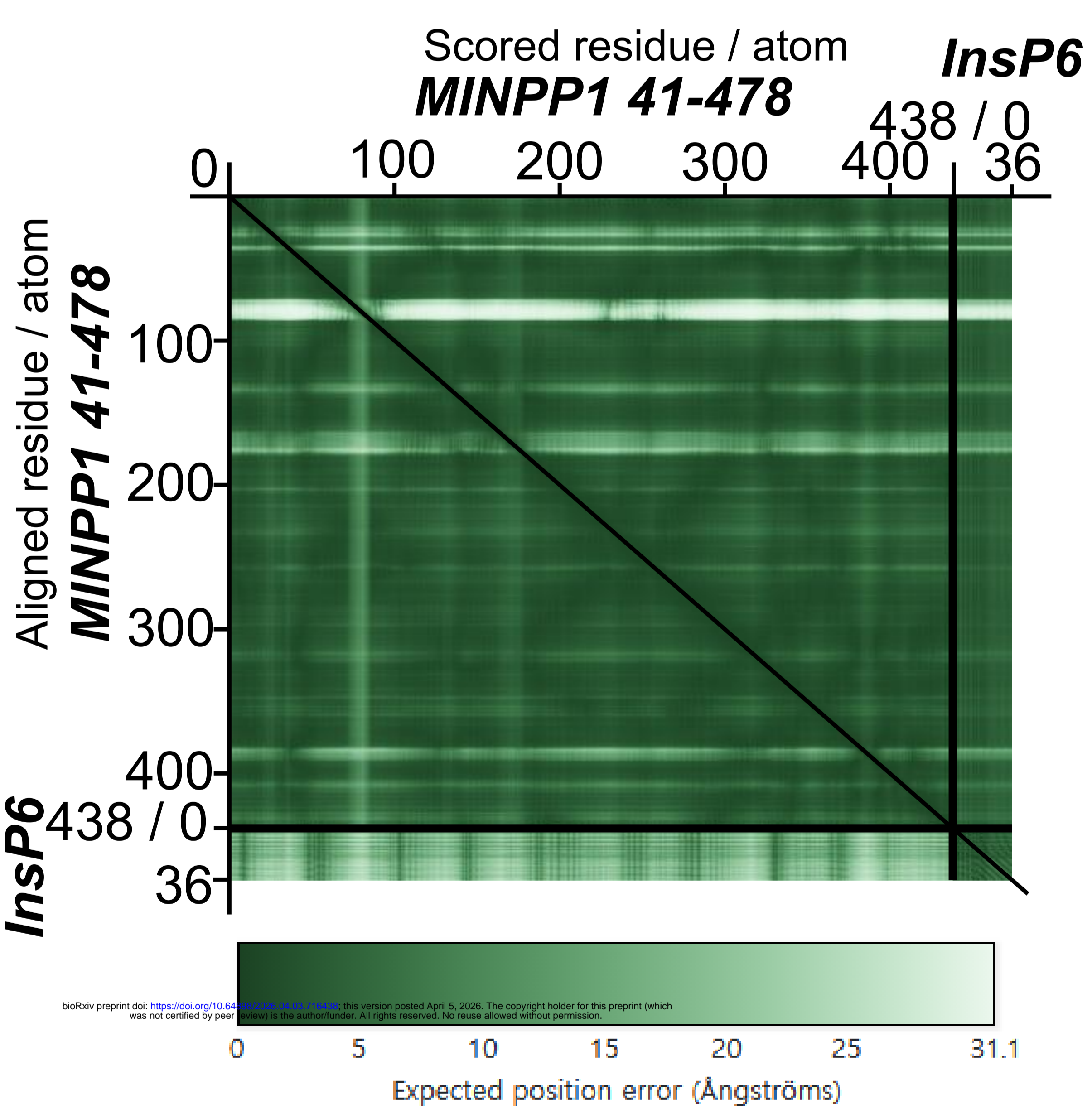


# Figure S4

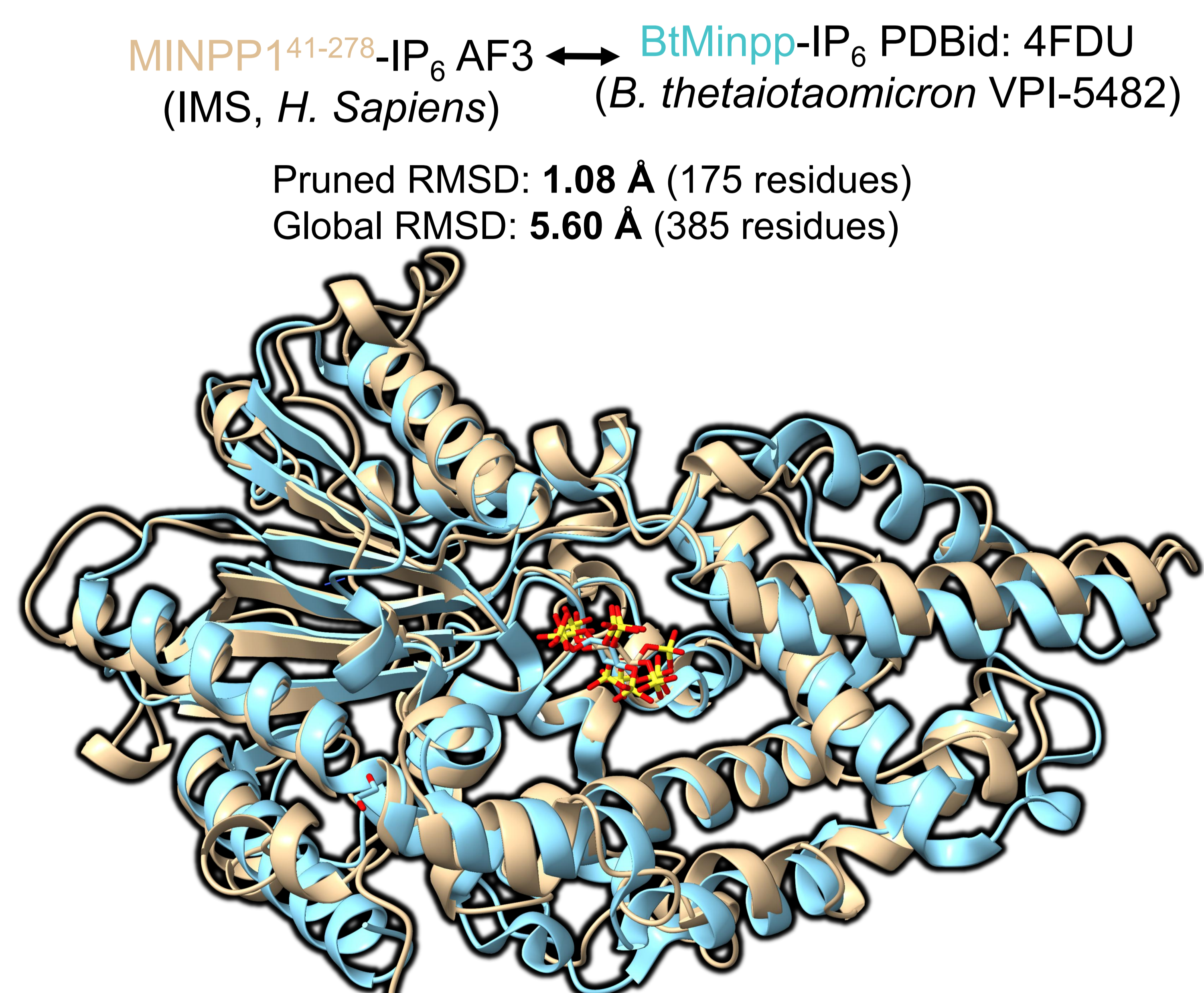
## A



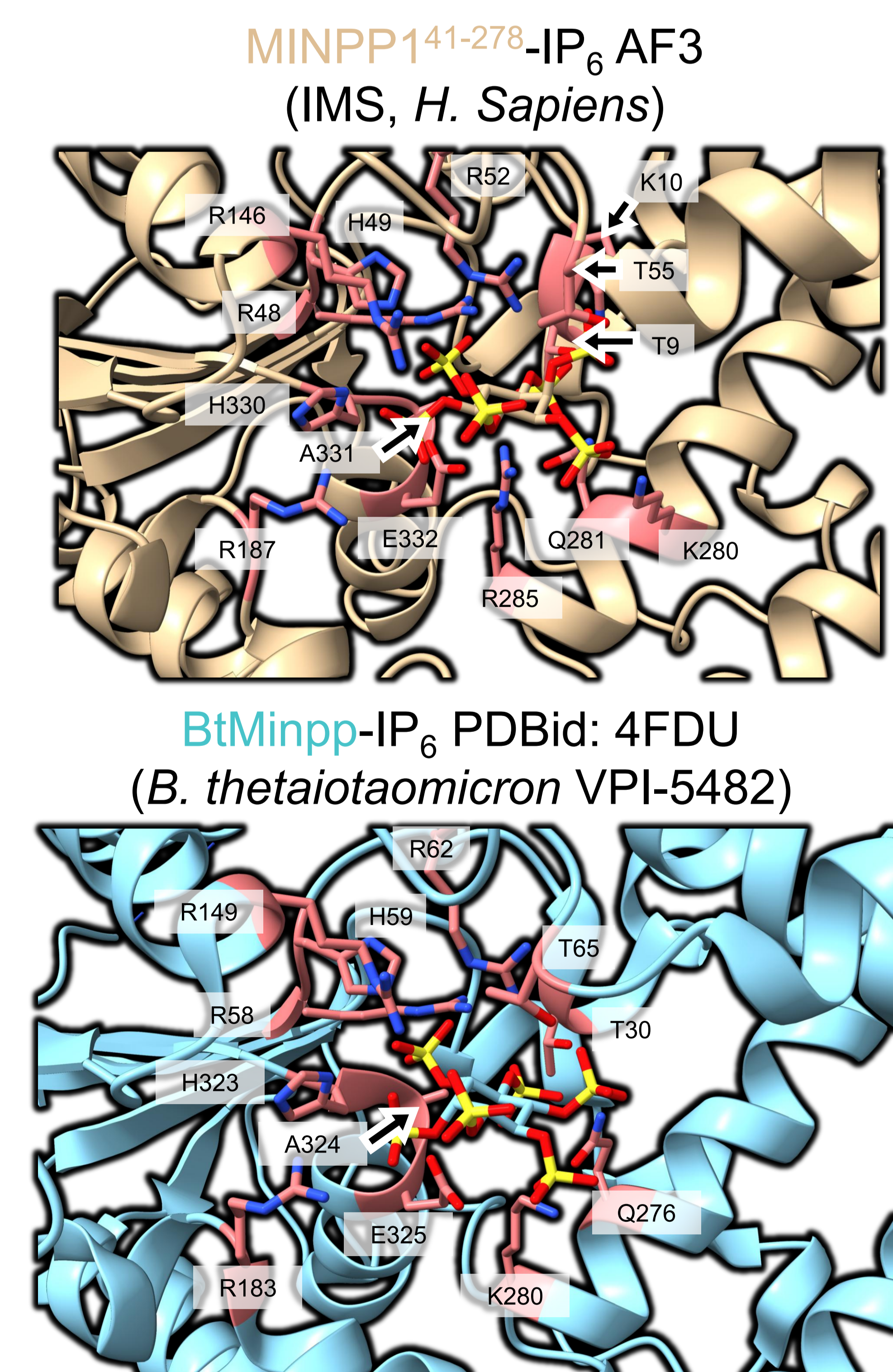
## B



## C



## D



# Figure S4

## E Mito-orphans with mitochondria related phenotypes (life span, brood size, growth)


Genename	<i>C. elegans</i> ortholog	Phenotype
Ep300	cbp-1	reduced brood size (RNAi, WBRNAi00113848, PMID:29099917) shortened life span (cbp-1/daf-16 double RNAi, WBRNAi00084531, PMID:19924292)
FKBP9	fkp-3	extended lifespan (RNAi, WBRNAi00065435, PMID:11743719)
ZFP36L1	oma-1	reduced brood size (RNAi, WBRNAi00110296, PMID:21071832) reduced brood size (RNAi, WBRNAi00078448, PMID:12296824) reduced brood size (RNAi, WBRNAi00078447, PMID:12296824)
ZFP36L1	oma-2	reduced brood size (RNAi, WBRNAi00110296, PMID:21071832) reduced brood size (RNAi, WBRNAi00078448, PMID:12296824) reduced brood size (RNAi, WBRNAi00078447, PMID:12296824)
ZFP36L1	pos-1	extended lifespan (RNAi, WBRNAi00111037, PMID:29063084) extended lifespan (RNAi, WBRNAi00066186, PMID:17411345)
ZFP36L1	mex-5	reduced brood size (RNAi, WBRNAi00036167, PMID:15489339) shortened life span (post-development RNAi, WBRNAi00111751,23144747)
ZFP36L1	moe-3	reduced brood size (RNAi, WBRNAi00078448, PMID:12296824)
YKT6	ykt-6	extended life span (post-development RNAi, WBRNAi00111608, PMID:23144747) reduced brood size (RNAi, WBRNAi00072884, PMID:17417969) slow growth (RNAi, WBRNAi00008301, PMID:12529635)

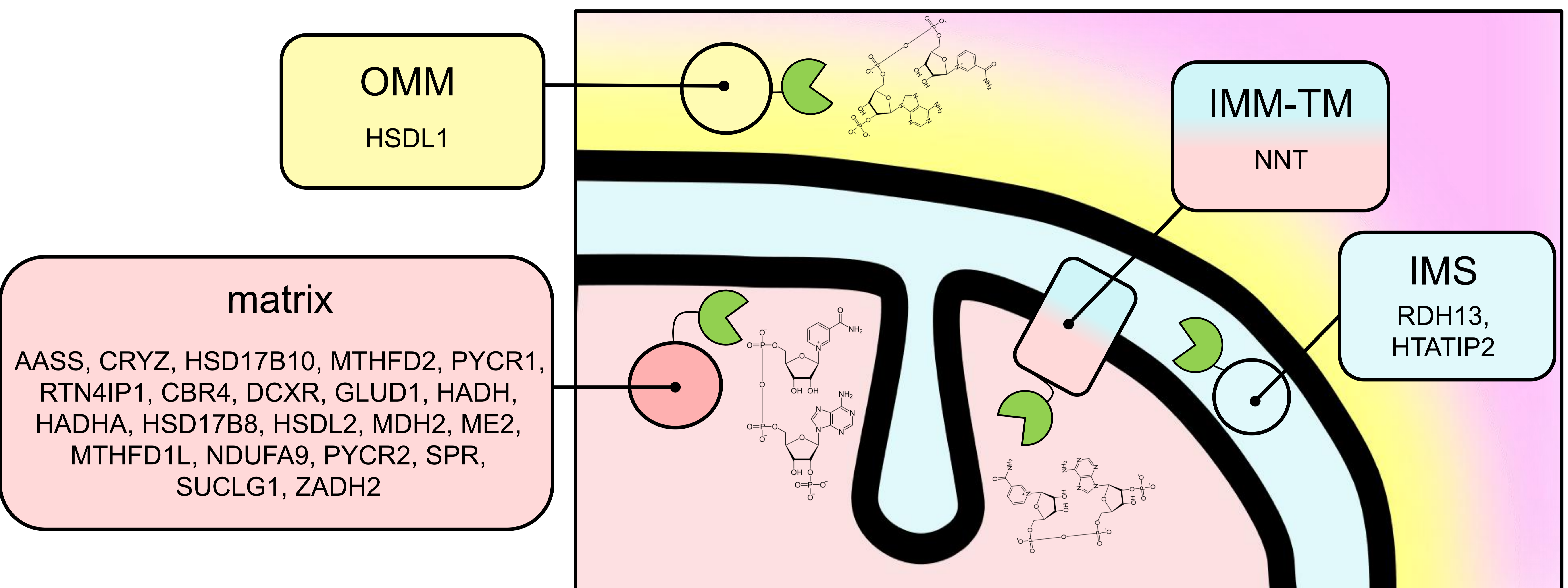
### Known mitochondrial genes

Genename	<i>Caenorhabditis elegans</i> ortholog	Phenotype
COX5B	cox-5B	extended life span (RNAi, WBRNAi00110260, PMID:27133166) reduced brood size (RNAi, WBRNAi00031451, PMID:15489339) slow growth (RNAi, WBRNAi00003489, PMID:11099033)
NDUFB4	nuo-6	extended life span (deletion, allele qm200, PMID:23555681) reduced brood size (RNAi, WBRNAi00072675, PMID:17417969)
COQ7 isoform 2	clk-1	extended life span (deletion, allele qm30, PMID:21531333) reduced brood size (deletion, allele qm30, PMID:20382831) slow growth (deletion, allele qm30, PMID:22719267)
SDHC	mev-1	shortened life span (deletion, allele kn1, PMID:21893079) reduced brood size (deletion, allele kn1, PMID:19583935) slow growth (RNAi, WBRNAi00009138, PMID:12529635)

# Figure S5

## A

NAD(P)-binding domain superfamily  
(IPR036291, )



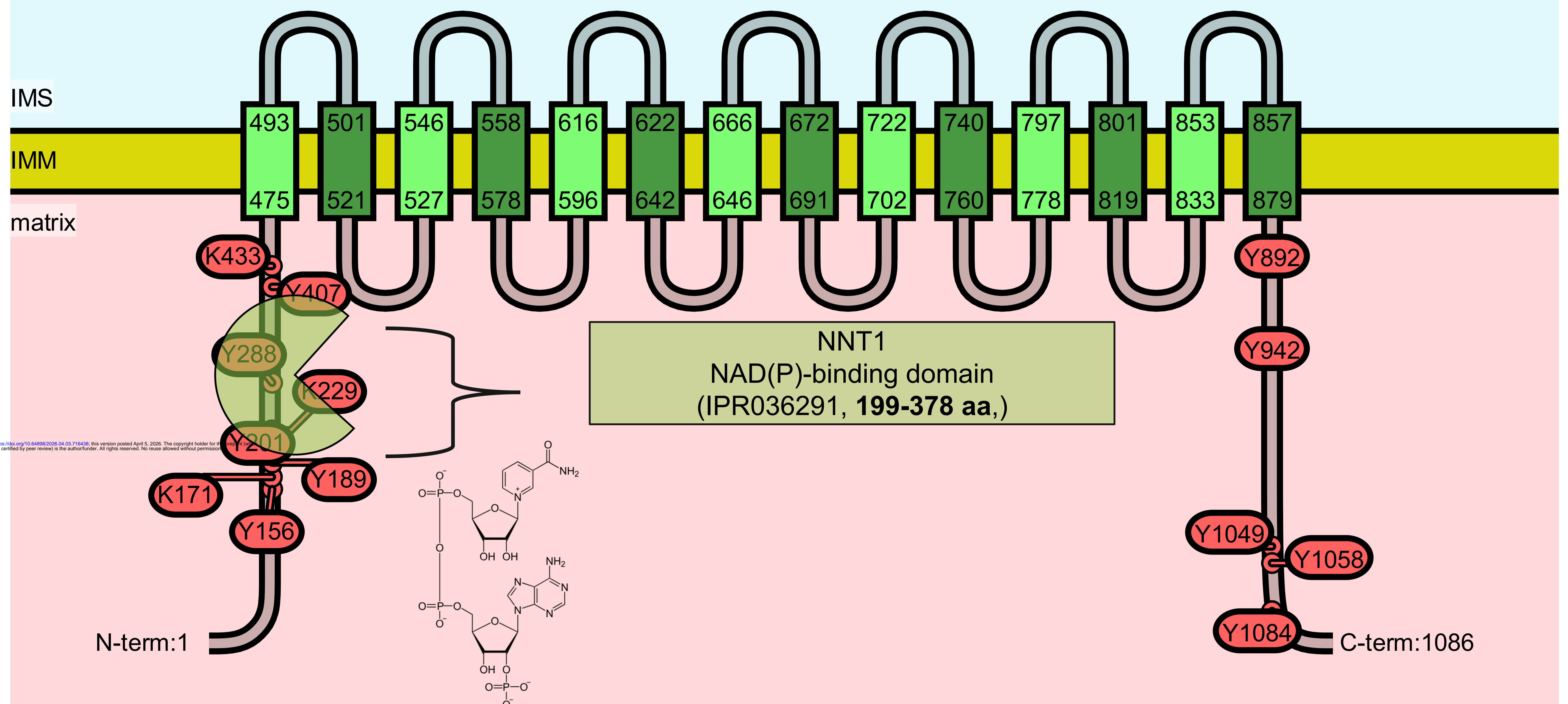
## B

NAD(P)-binding domain topology of NNT

Membrane Topology of NNT

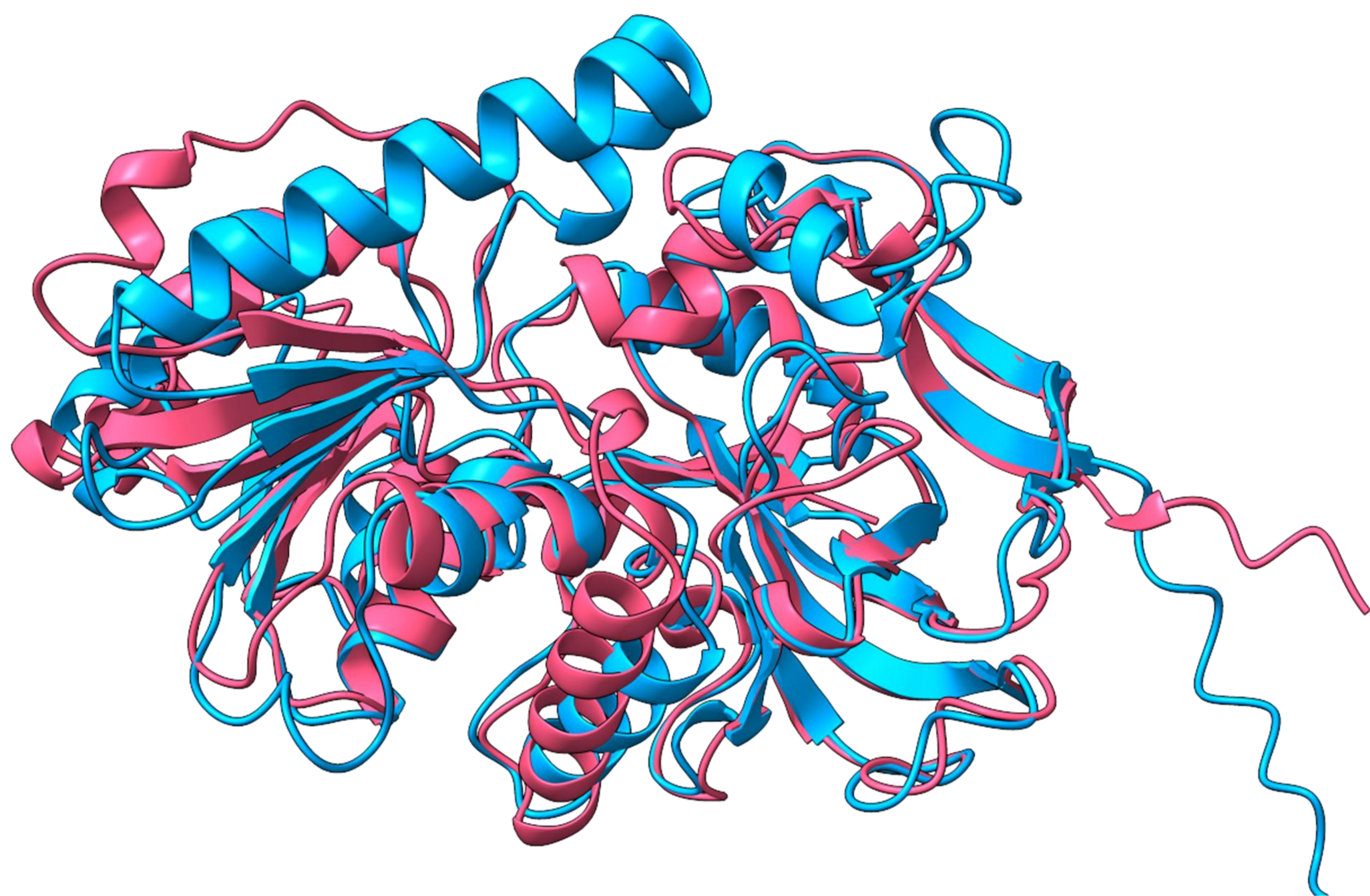
TM annotation from uniprot

Topology Score : 13



# Figure S5

## C



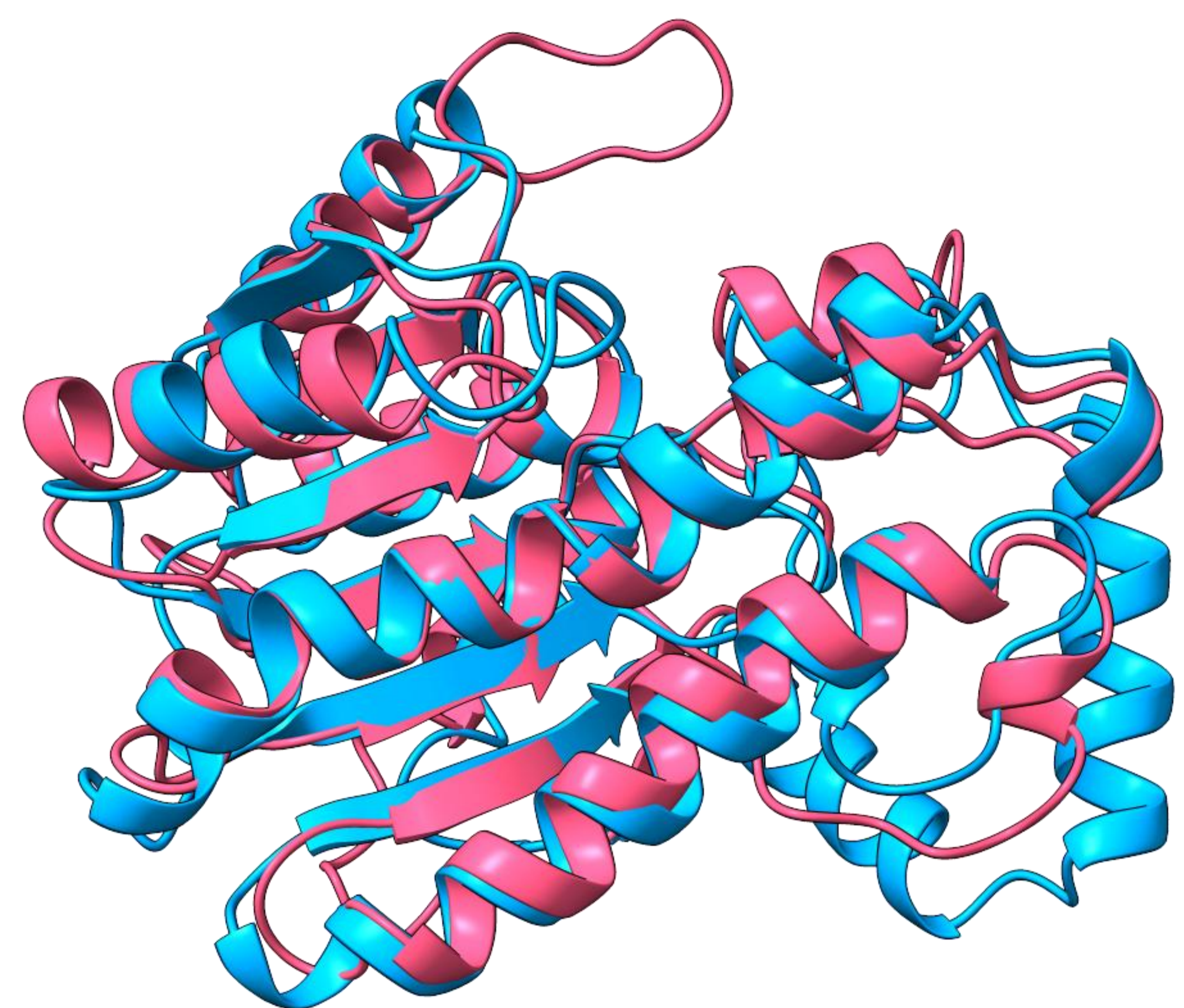
RTN4IP1 1-396

CRYZ 1-329

Sequence homology:

Pruned RMSD: 1.173 Å (178 residues)

Global RMSD: 5.704 Å (318 residues)



HSDL1 64-320

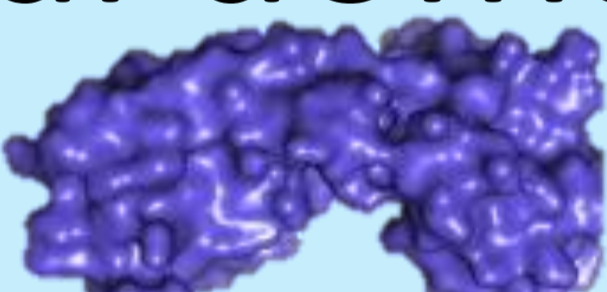
HSDL2 6-243

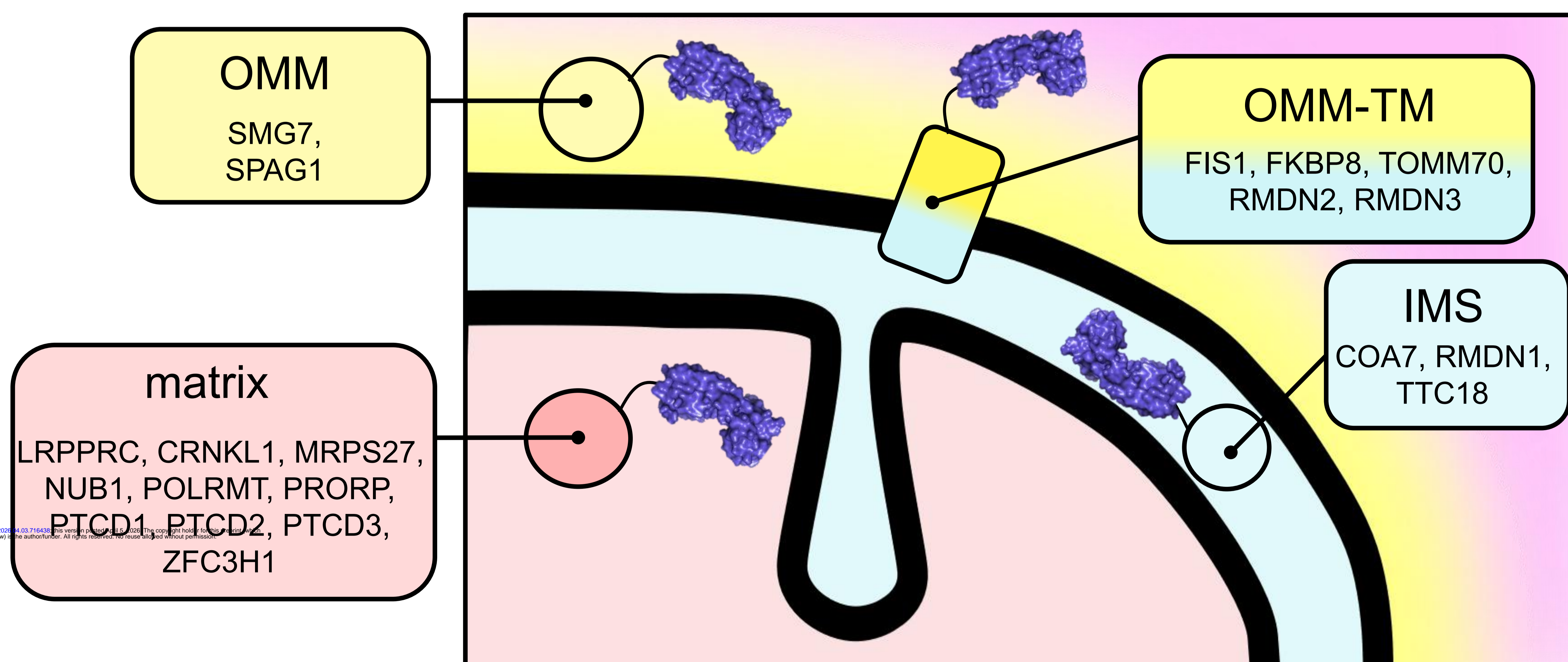
Sequence homology:

Pruned RMSD: 1.007 Å (142 residues)

Global RMSD: 6.418 Å (257 residues)

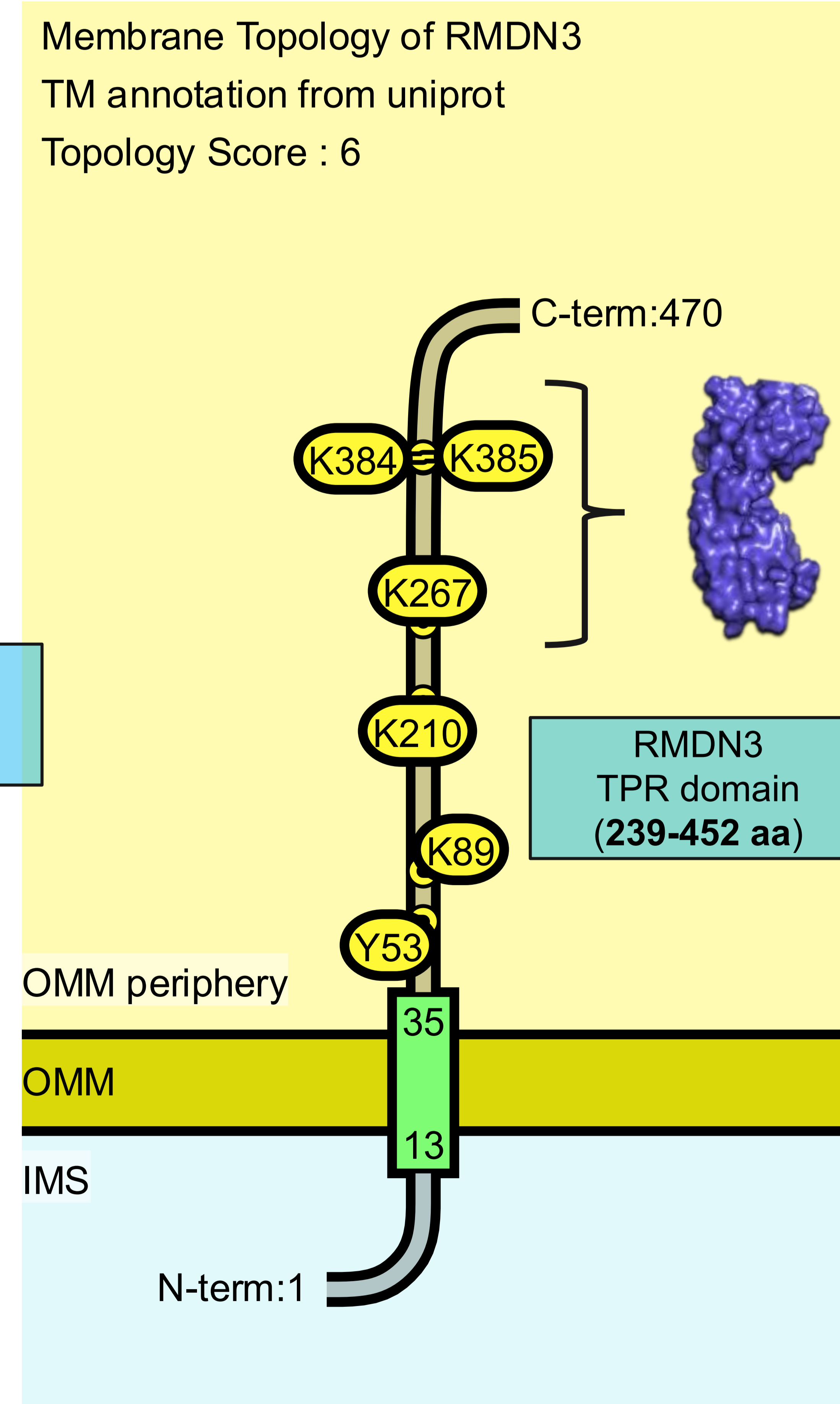
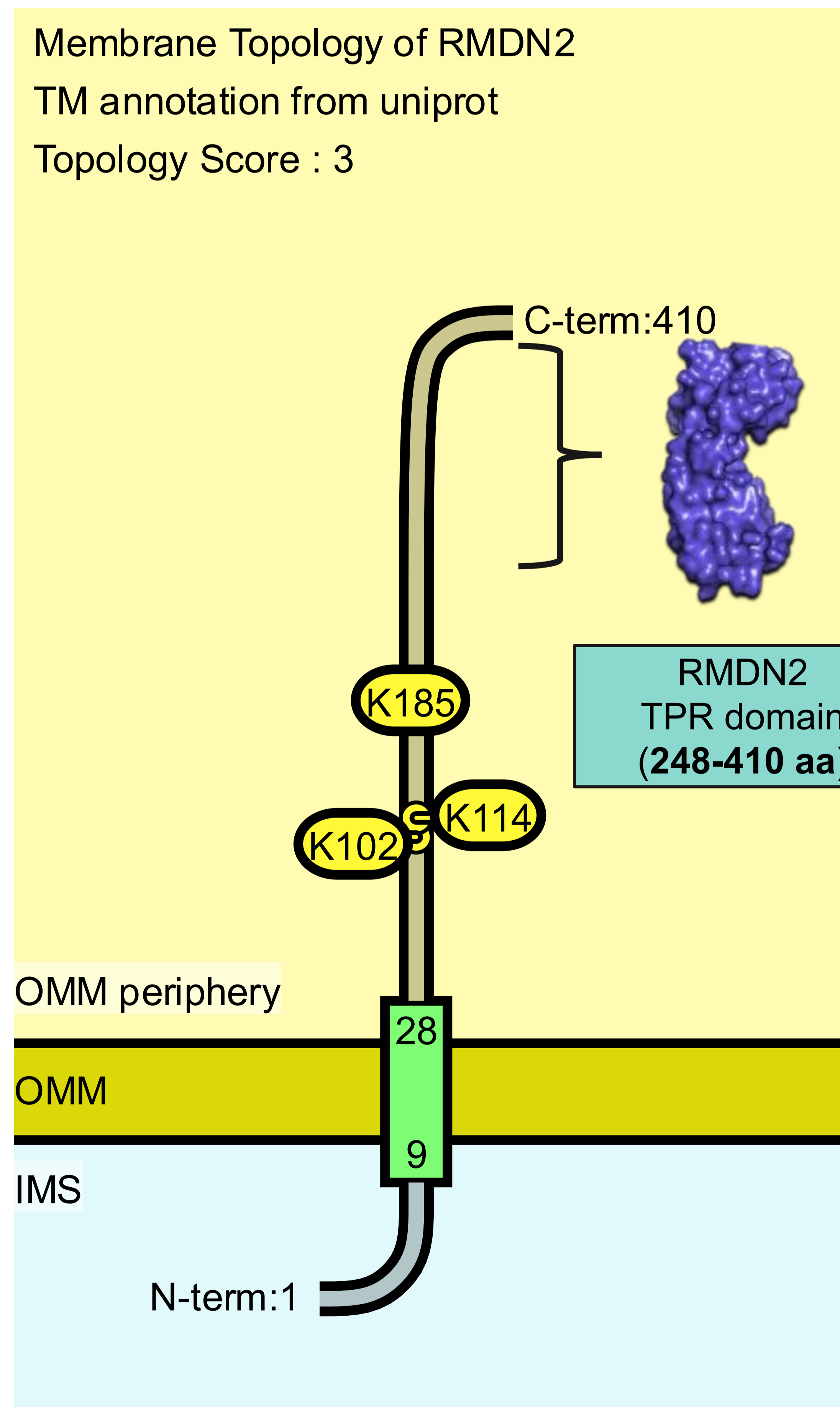
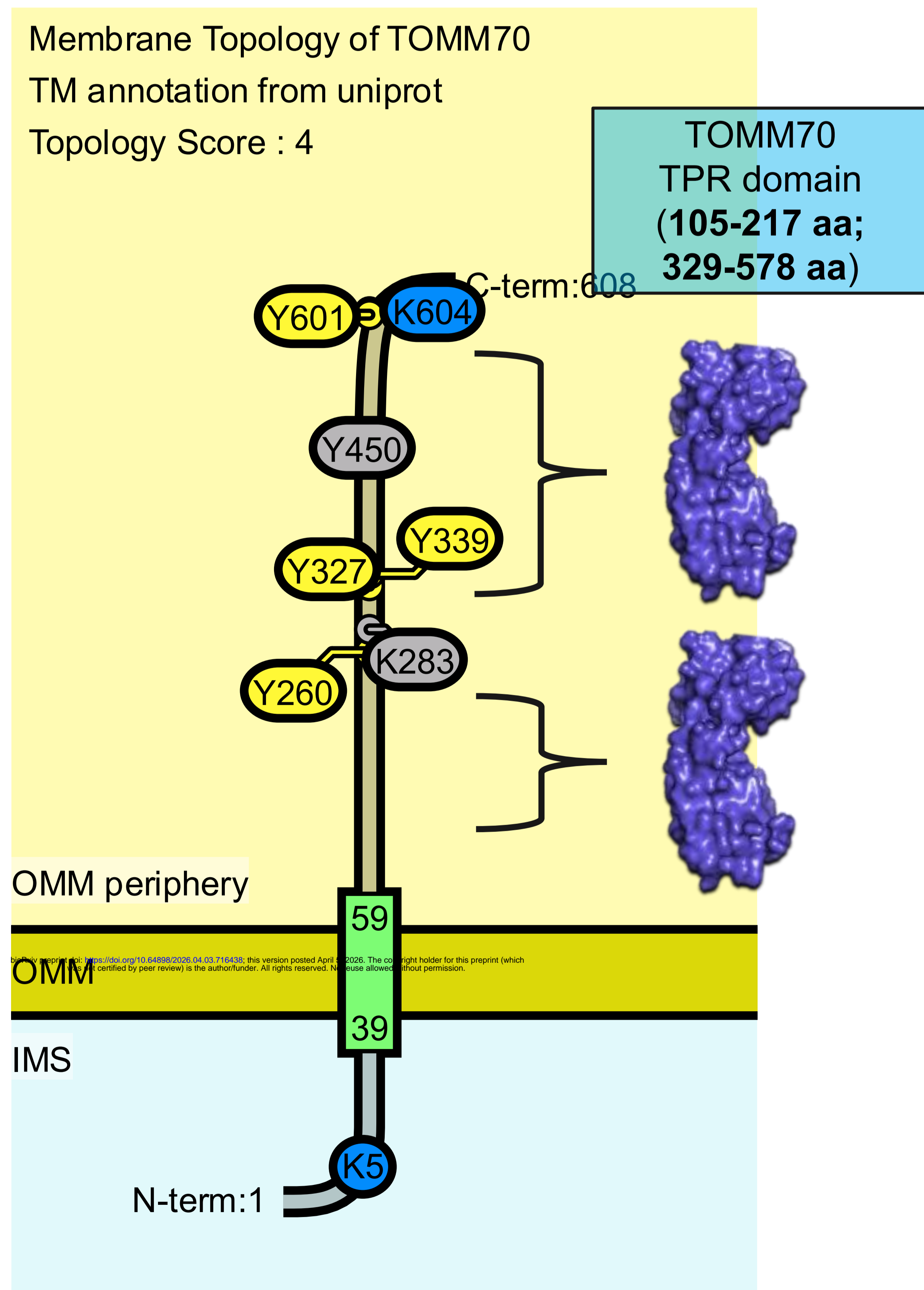
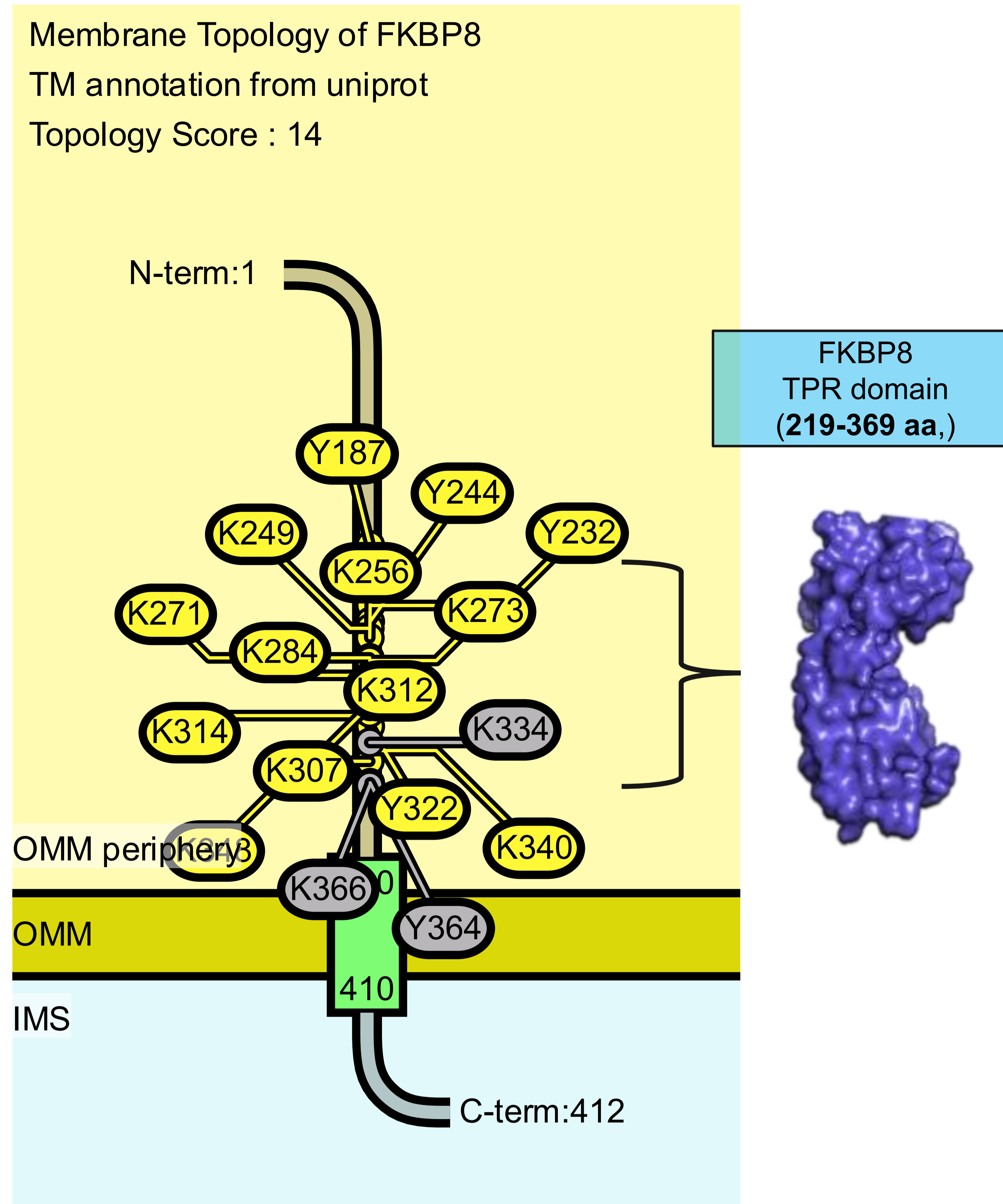
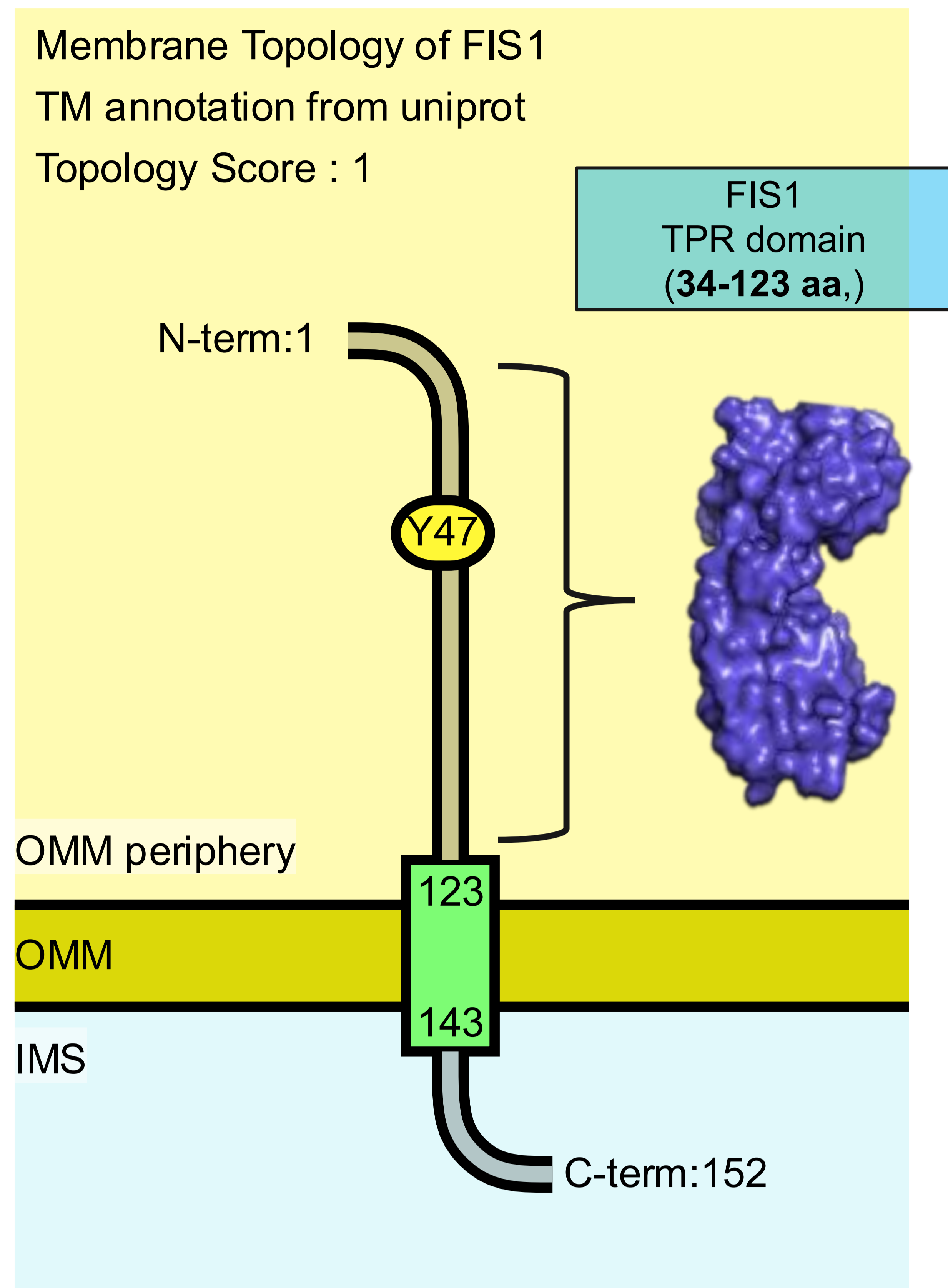
## D

Tetratricopeptide-like helical domain superfamily  
(IPR011990, )

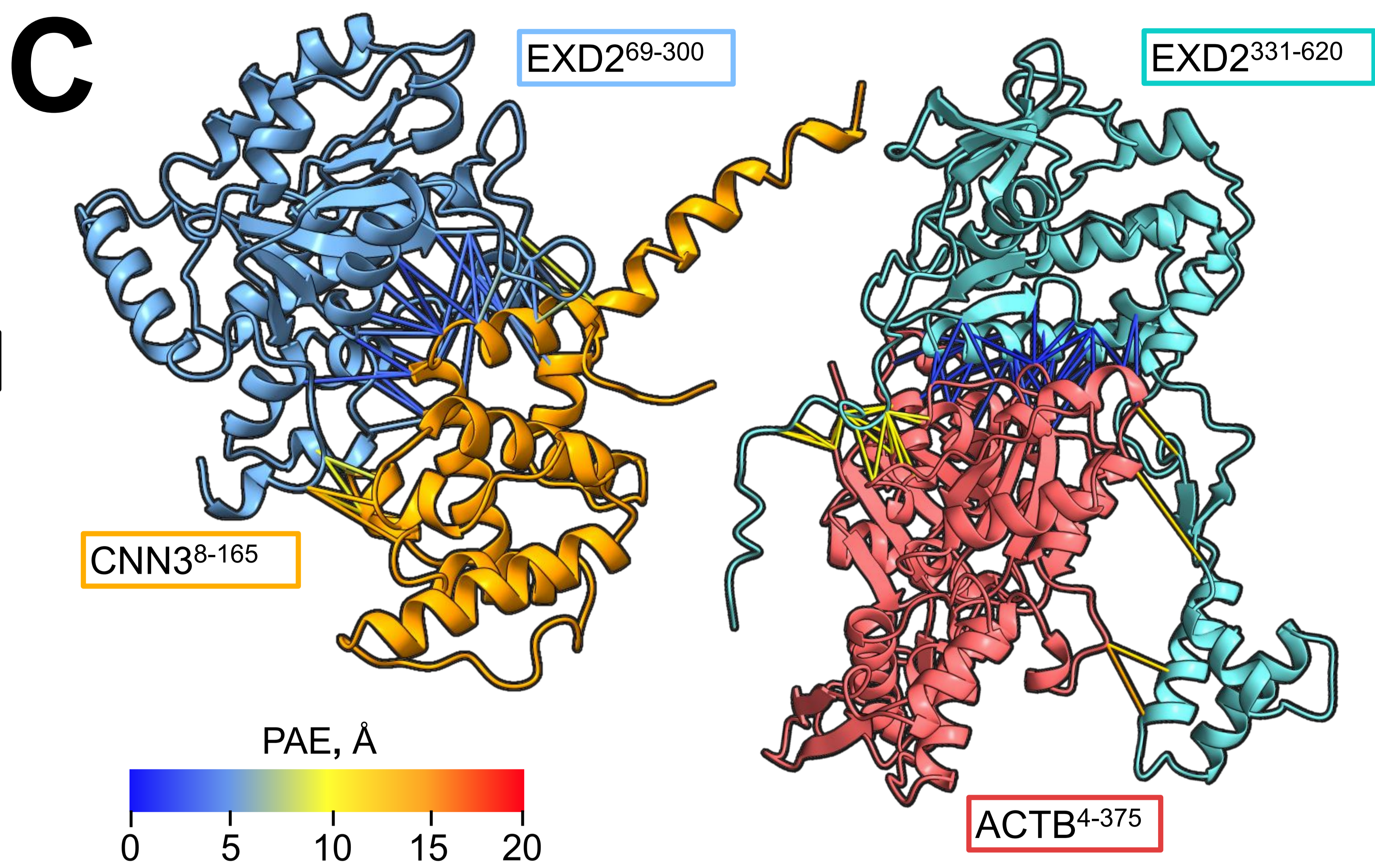
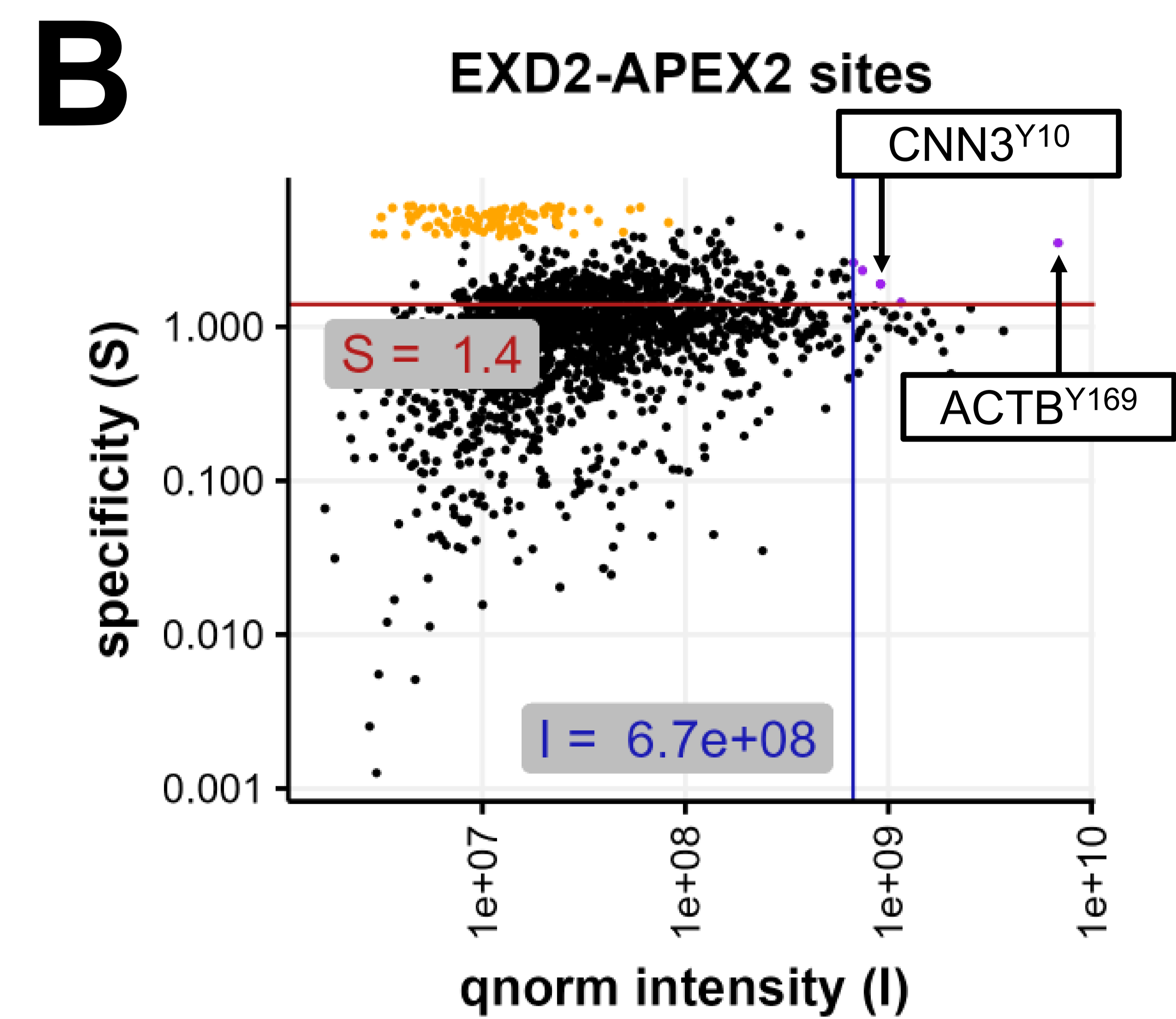
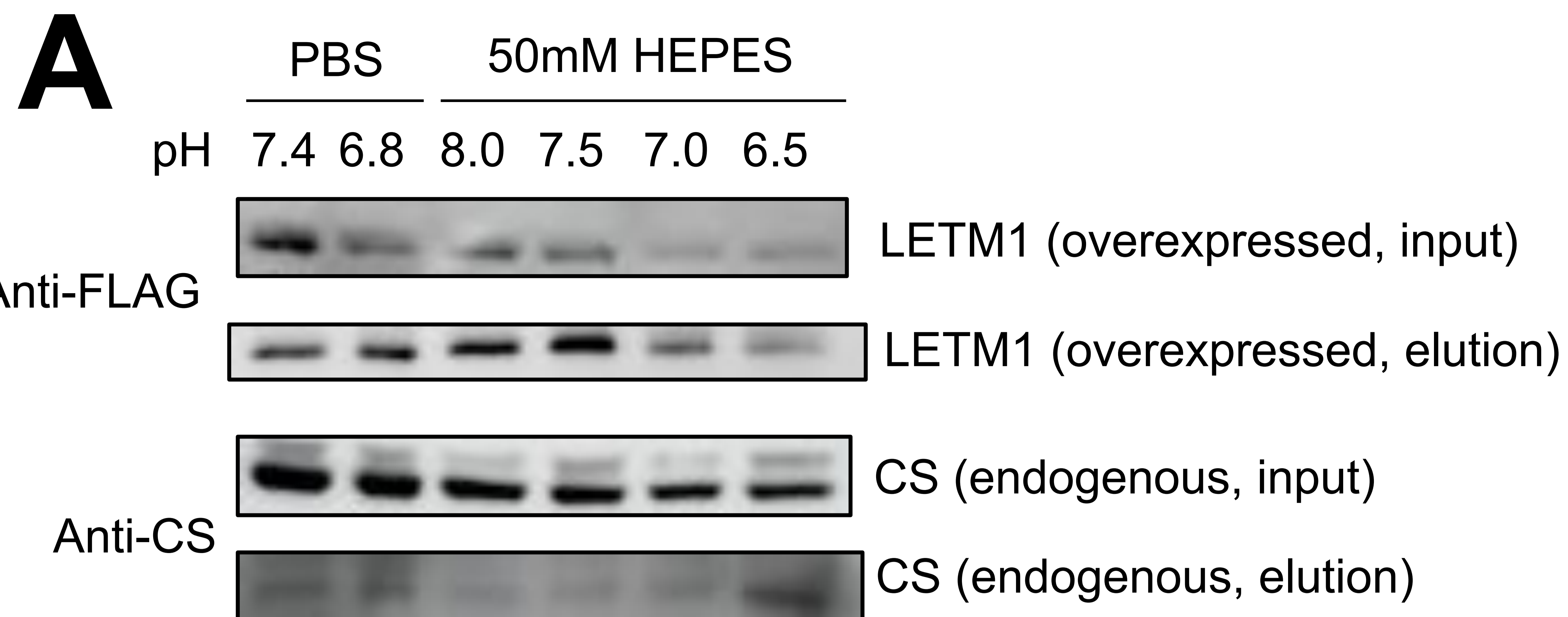


# Figure S5

## E

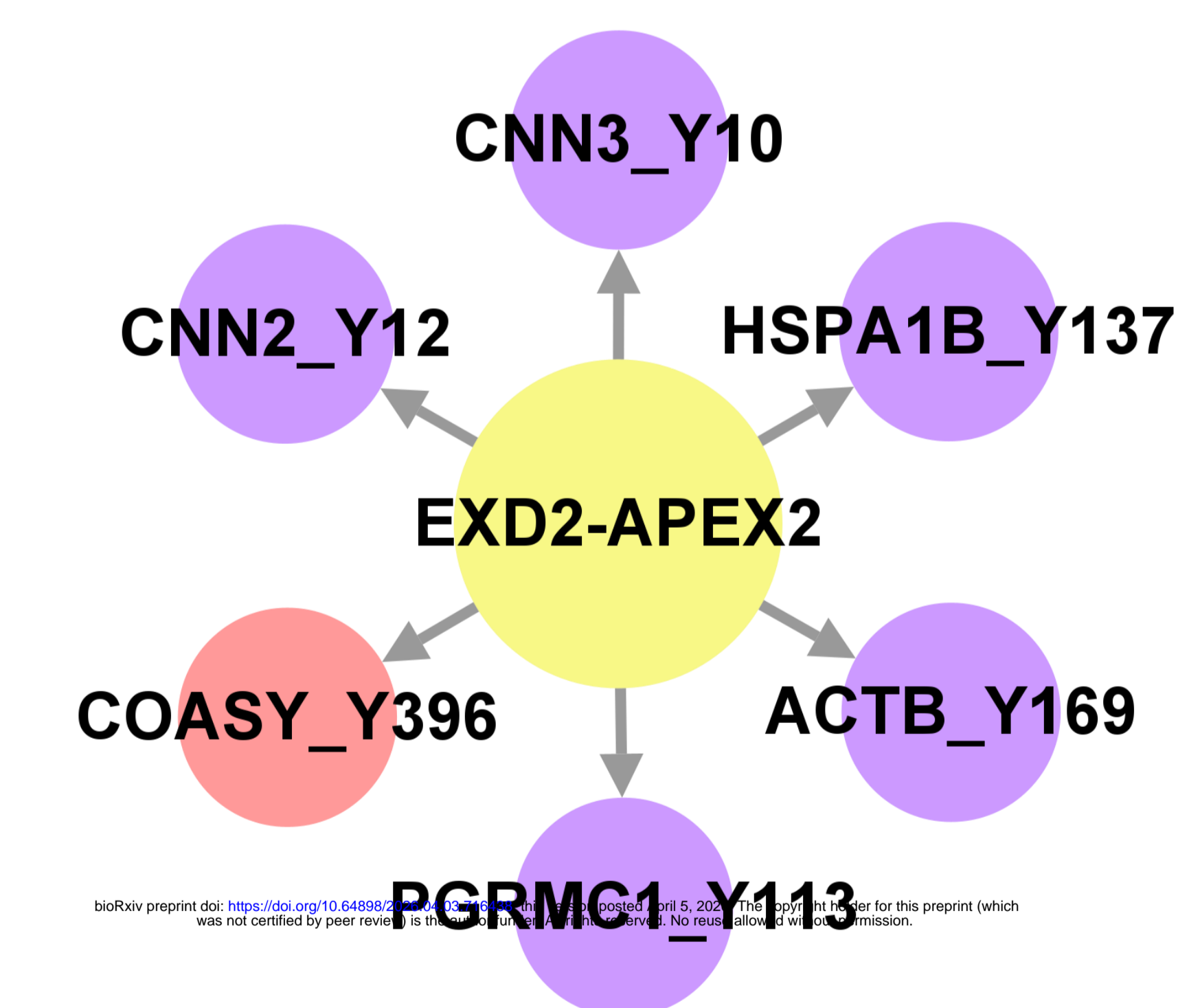


# Figure S6



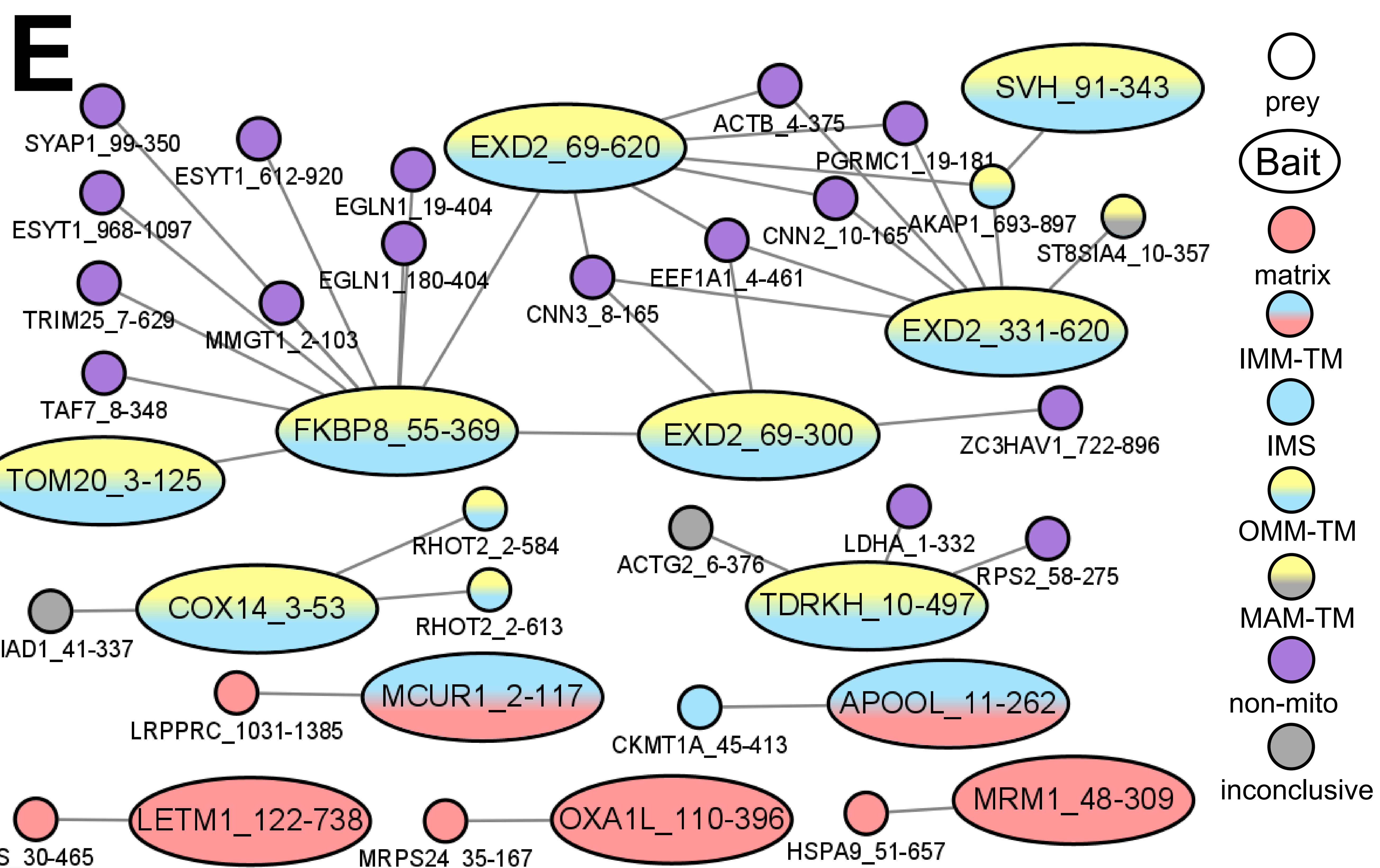
**D**

EXD2-APEX2 specific sites



Protein localization

- matrix
- OMM
- non-mito



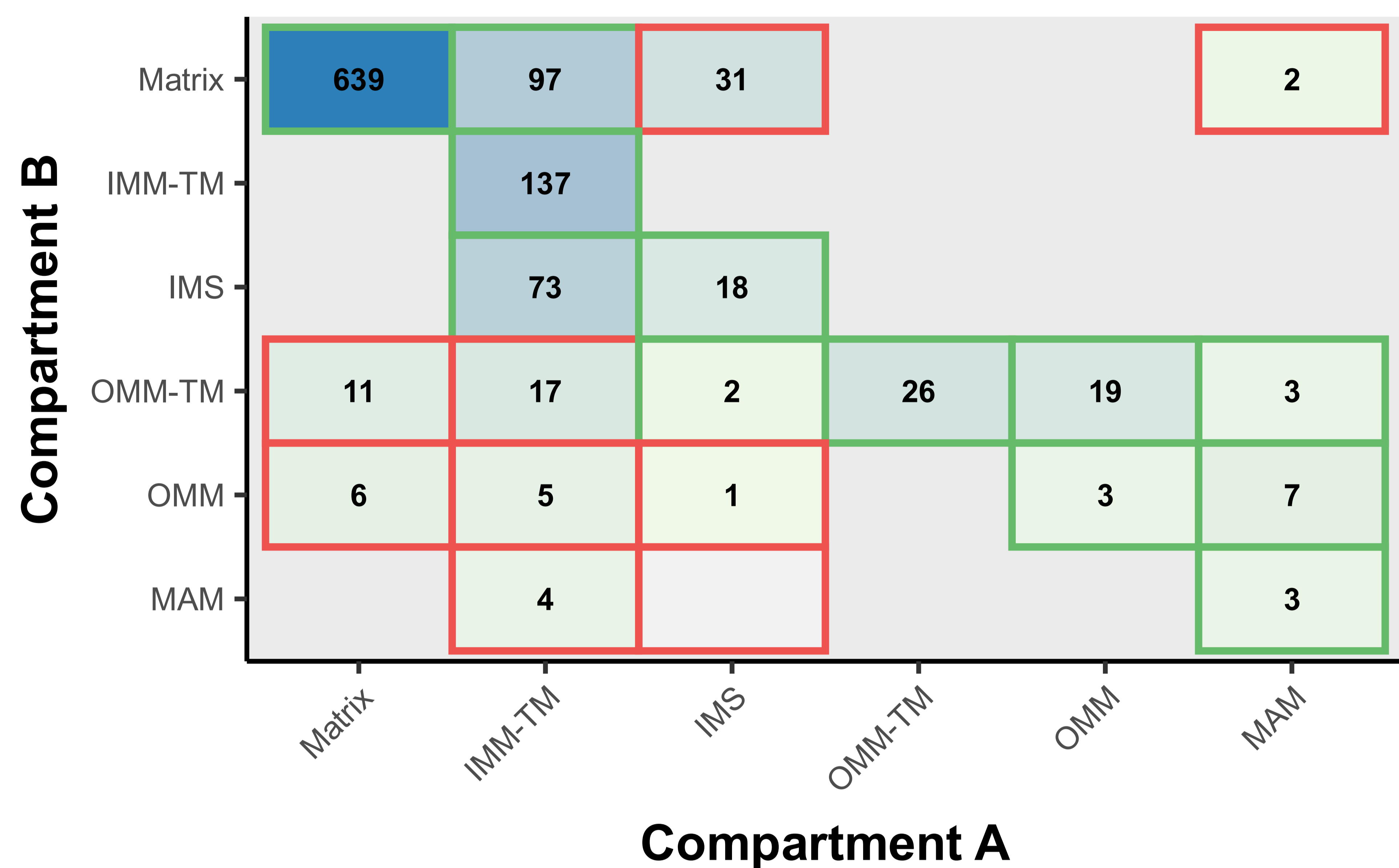
# Figure S6

## F

### Topology validation of predicted interactome

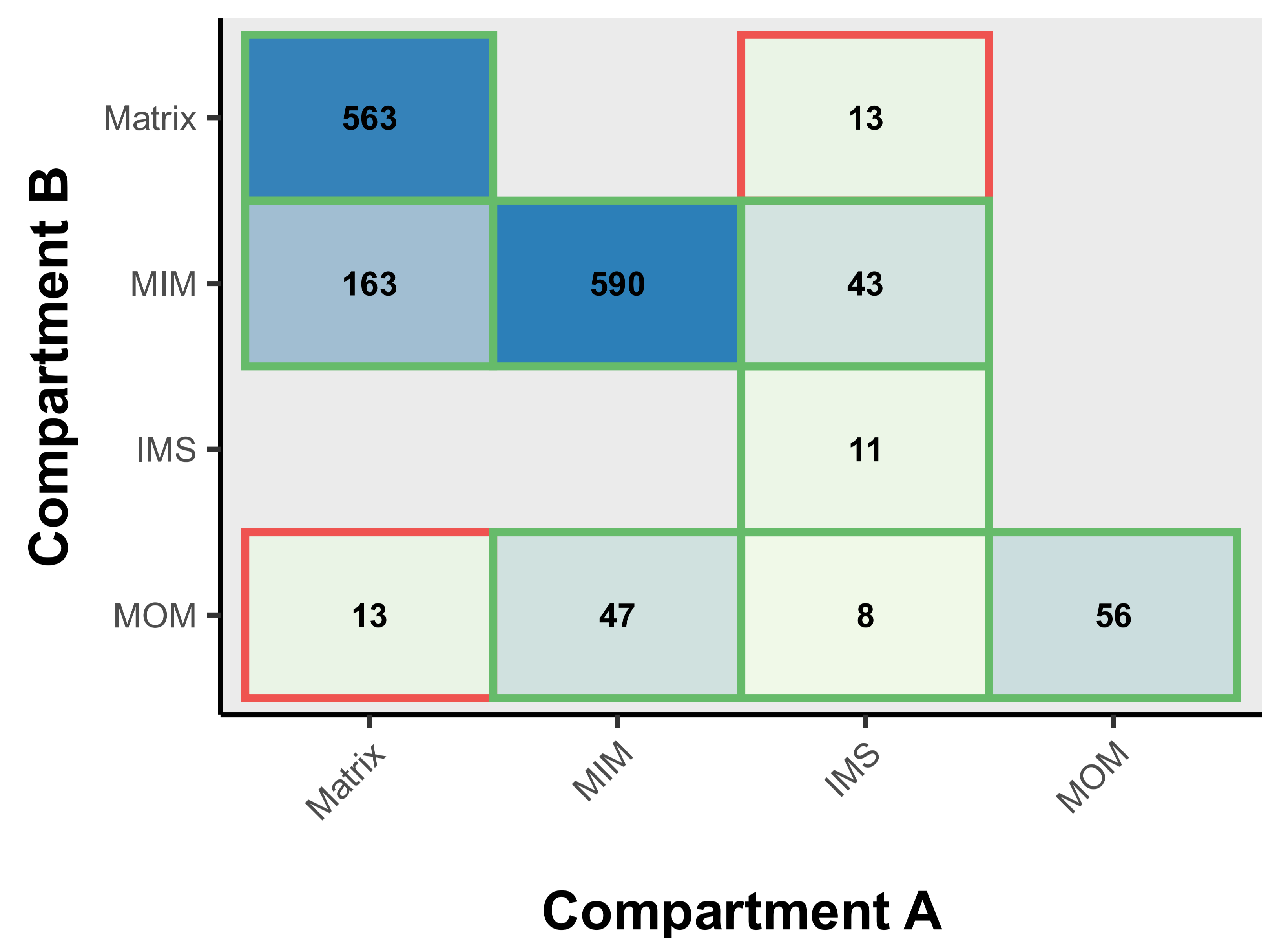
(i)

#### MitoAtlas



(ii)

#### MitoCarta3.0

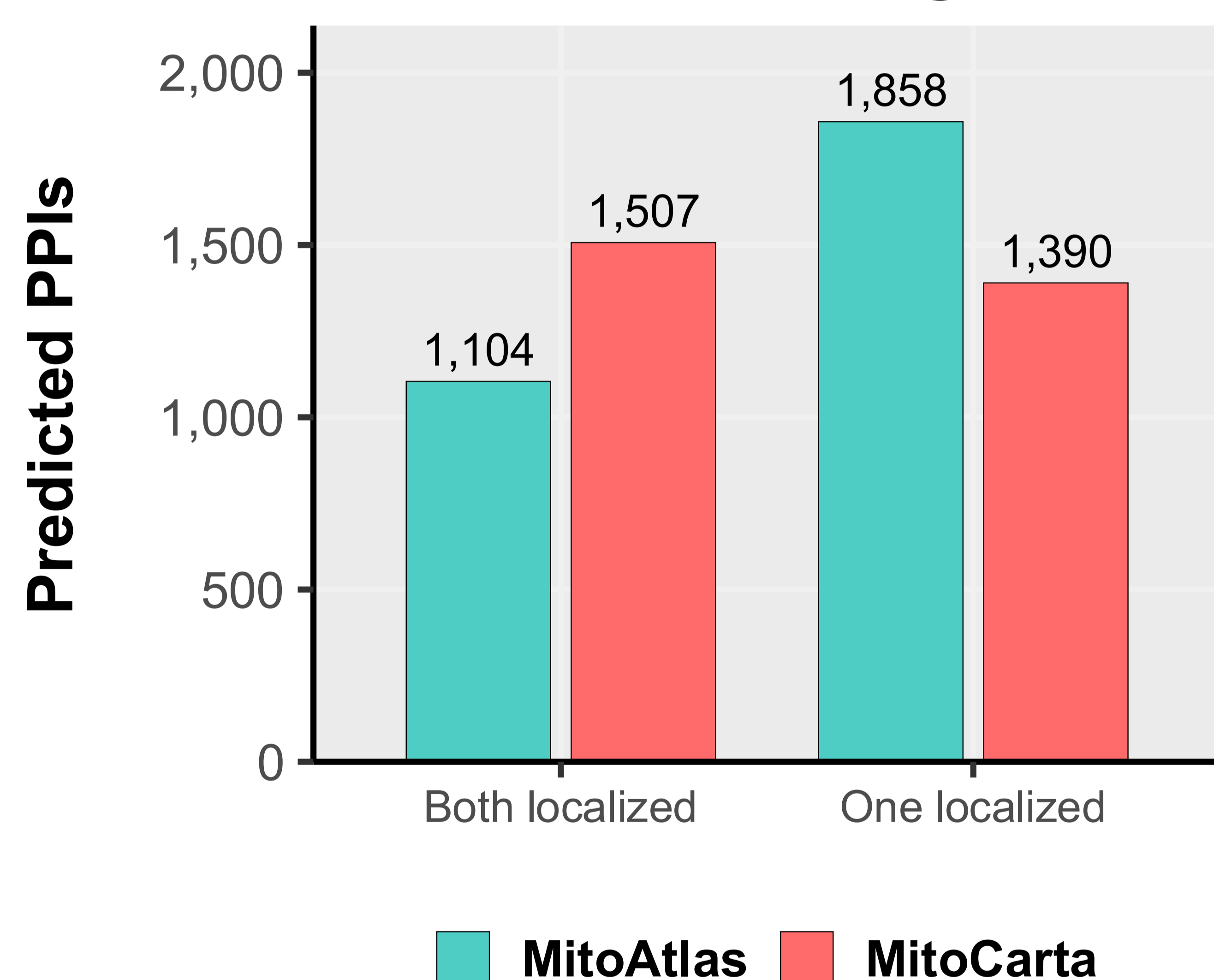


Count 1 100 500 Topology Compatible Incompatible

Count 100 500 Topology Compatible Incompatible

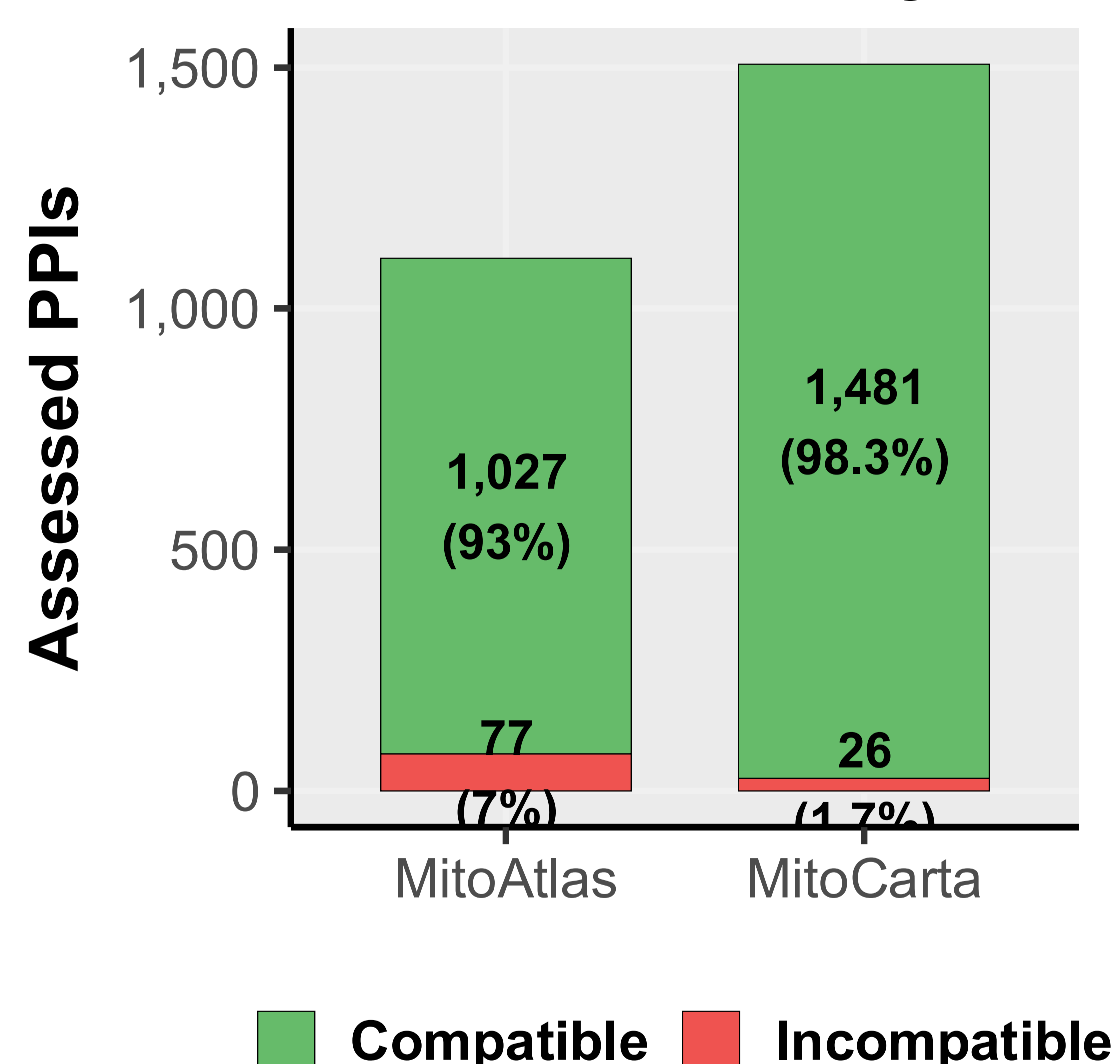
(iii)

#### PPI coverage



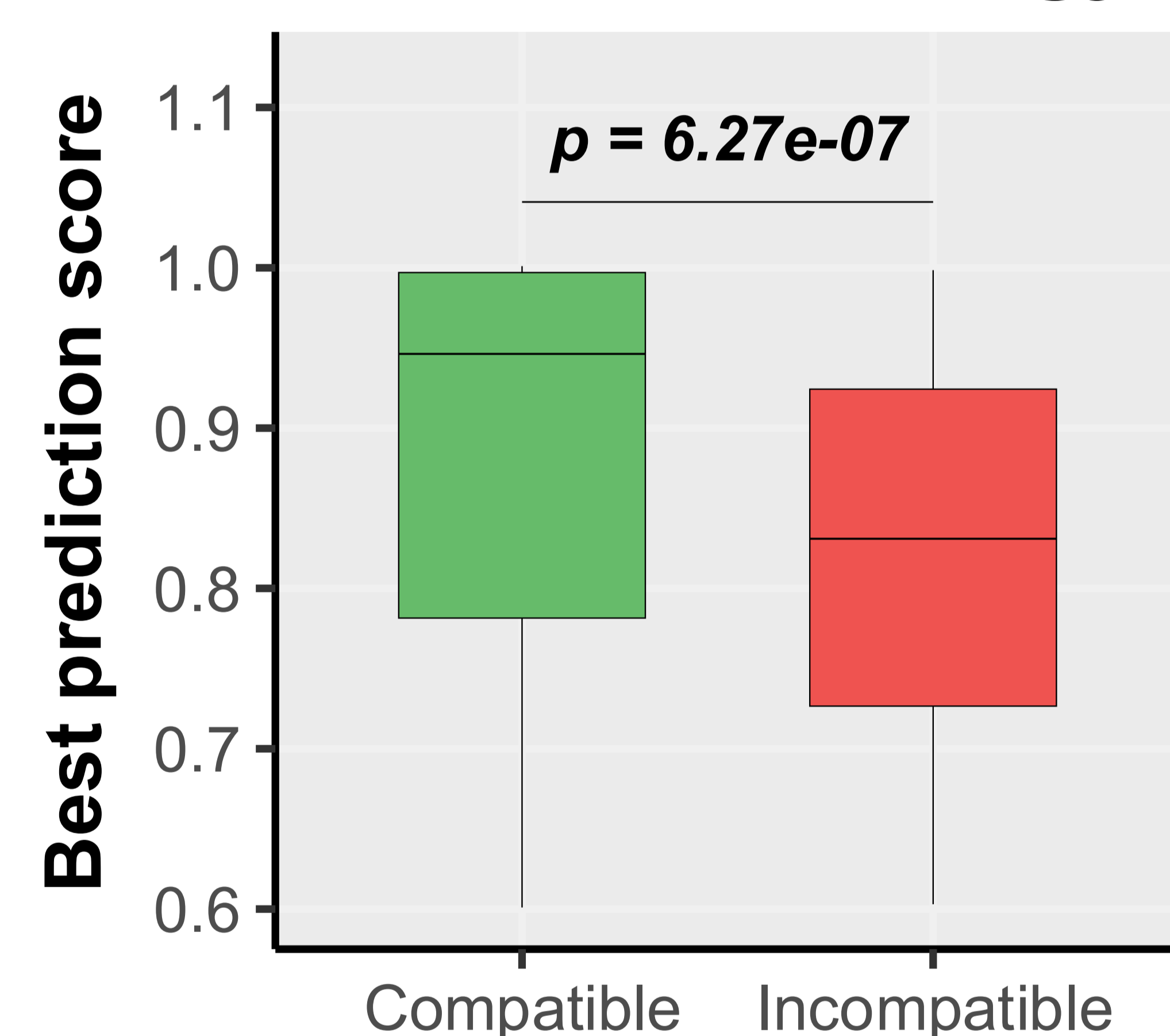
(iv)

#### Compatibility



(v)

#### Score vs topology

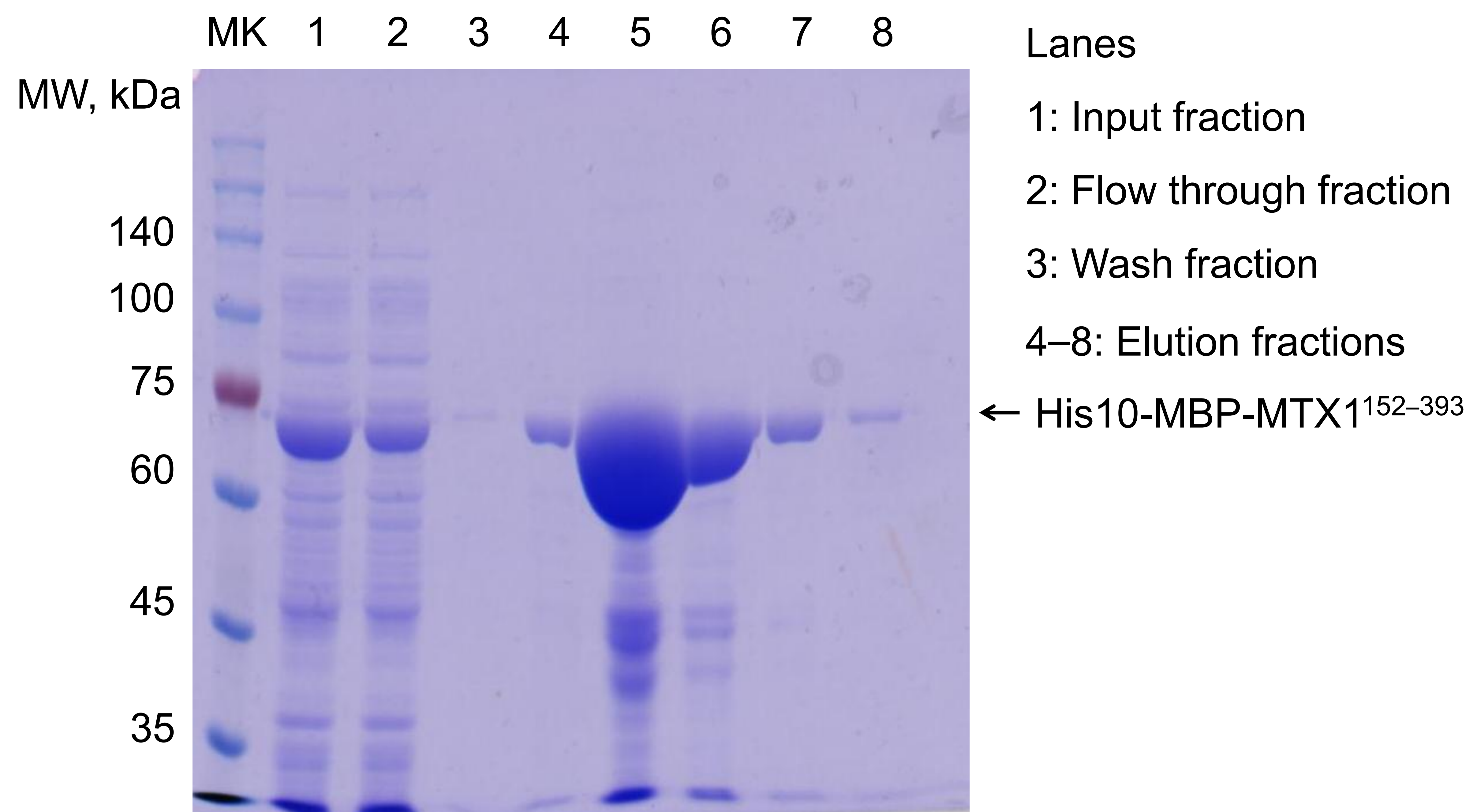


(vi)

#### Discordant PPIs

Interaction	MitoAtlas	MitoCarta 3.0
MTX2 – ATP5MK	OMM ✗	IMM-TM ✓
IMMT – MTX3	IMM-TM ✗	MIM ✓
TOMM7 – MPC1	OMM-TM ✗	MOM ✓
ATP5MK – OCIAD1	IMM-TM ✗	MIM ✓
NDUFS5 – NDUFA5	IMS ✗	MIM ✓
COX6B1 – NDUFA5	IMS ✗	MIM ✓
MTX2 – NDUFAF1	OMM ✗	MOM ✓
NDUFAB1 – DNAJC11	Matrix ✗	MIM ✓
FOXRED1 – TOMM22	Matrix ✗	MIM ✓

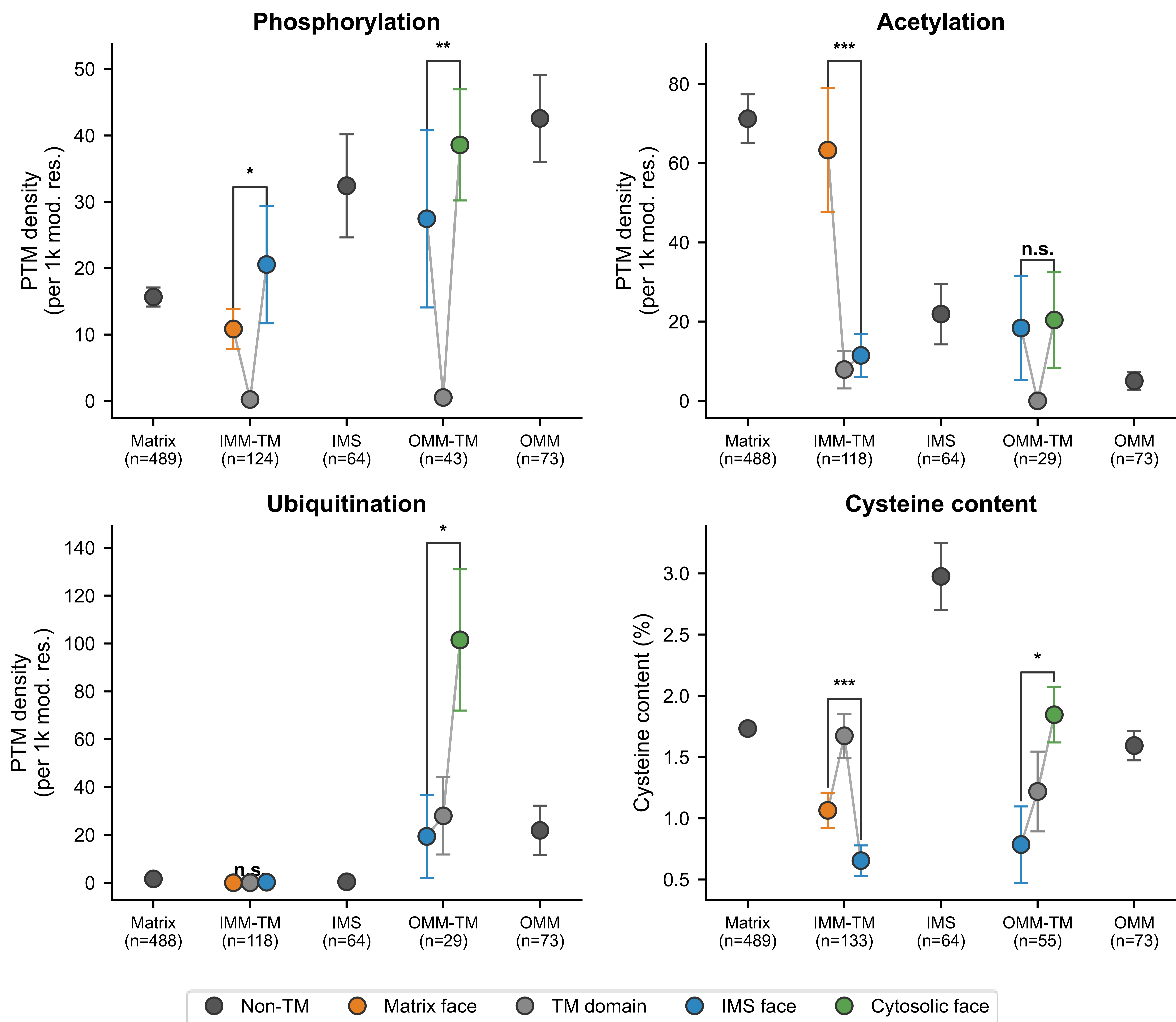
# G





# Figure S7

## B



## C

