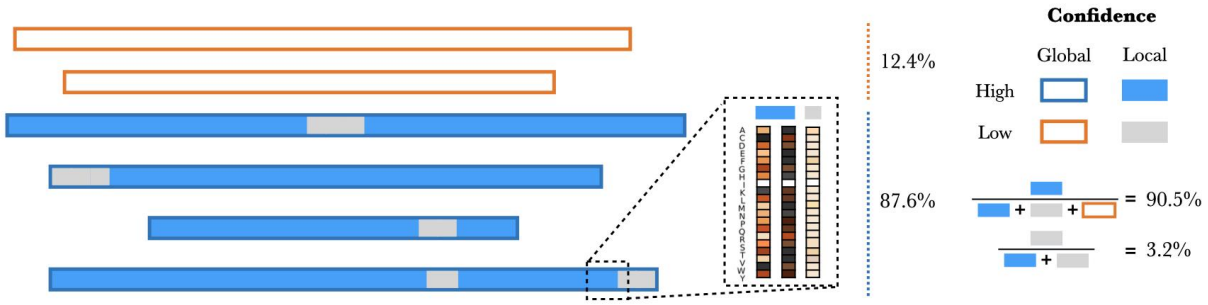
Extended Data

Method	Threshold	Specificity		Recall	Balanced accuracy		
		DGRP	DEST2	Lethal	DGRP	DEST2	All
ProteoCast classification	-	0.88	0.82	0.85	0.86	0.83	0.83
GEMME raw scores	-2.4	0.87	0.81	0.81	0.84	0.81	0.81

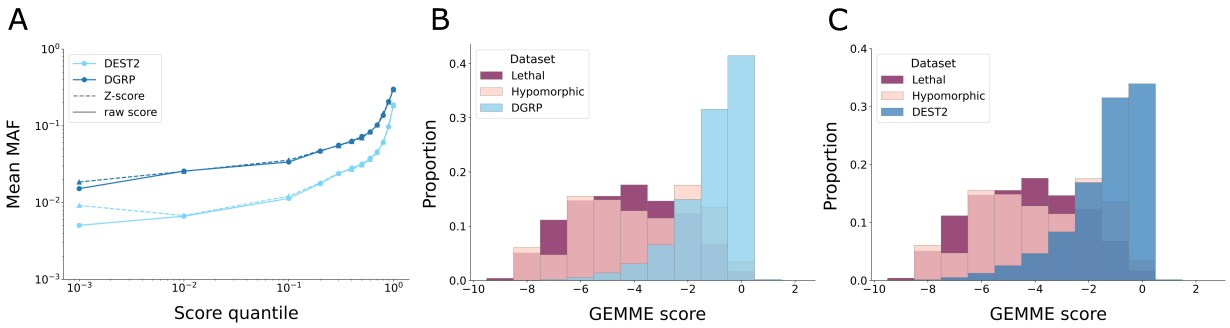
Extended Data Table 1: **ProteoCast performance on the custom benchmark (DEST2, DGRP, Lethal).** Balanced accuracy is computed as the average of the Specificity and Recall. We compare ProteoCast classification to binary classifications obtained by applying a universal threshold to GEMME raw scores.

	Total number of sites*	pLDDT <70				
		Number of sites	Number of sites classified as sensitive		Number of sites standing out from their background	
PTMs	60,428	53,737(89%)	20,923 (39%)		33,866 (63%)	
			With at least 1 residue detected	Fully included	With at least 1 residue detected	Fully included
SLiMs	91	72 (79%)	59 (82%)	34 (47%)	61 (85%)	44 (61%)

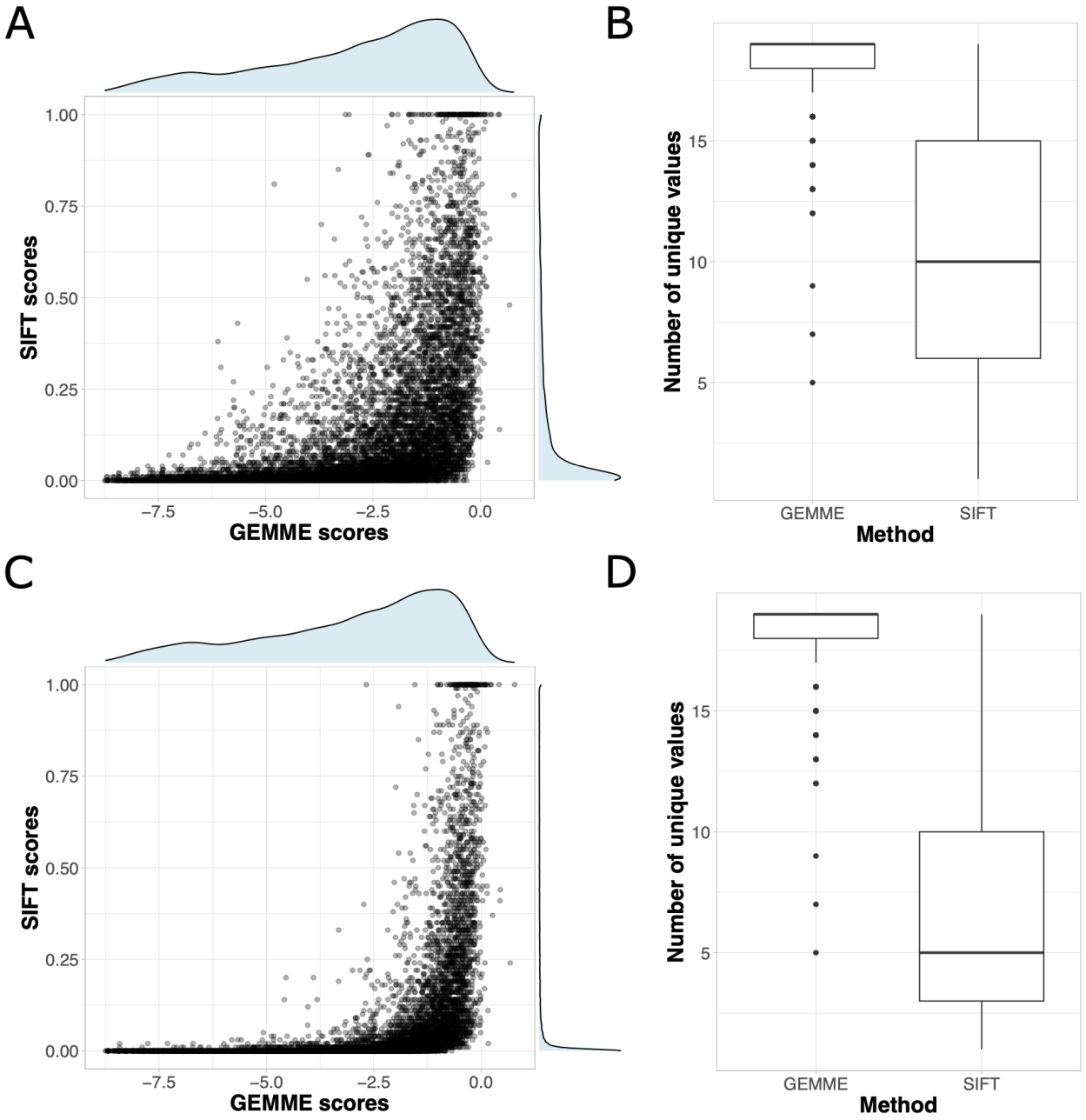
Extended Data Table 2: **Summary table of functional signal retrieved by ProteoCast for post-translational modifications (PTMs)** [36] **and Short Linear Motifs (SLiMs)** [41] **in unstructured regions (AlphaFold pLDDT<70).** *in proteoforms with 3D structure information.



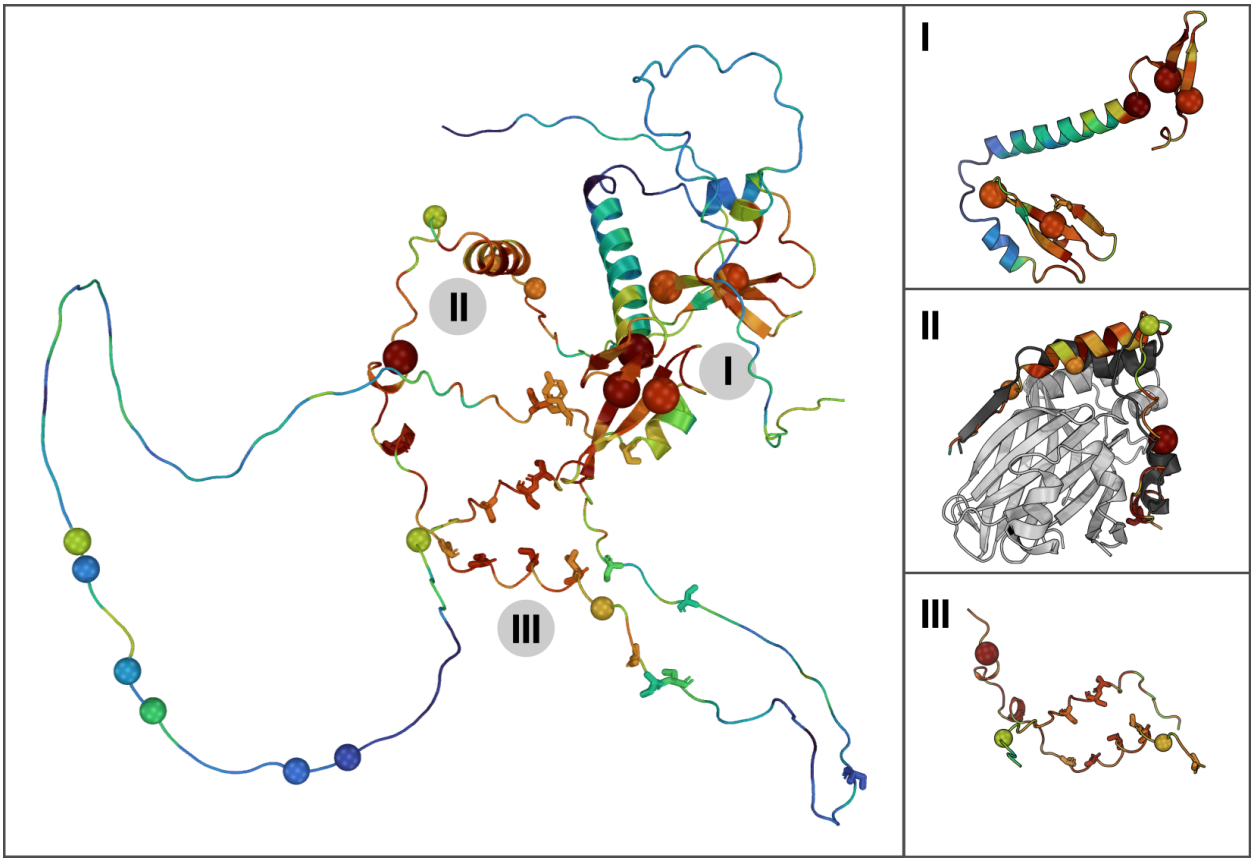
Extended Data Figure 1: **Overview of global and local confidence metrics.** The rectangular outlines (orange or blue) indicate global confidence while the internal coloring (blue or grey) indicates local confidence. The insert illustrates archetypal scenarios for high and low residue confidence. The left column has high resolution, meaning that GEMME can distinguish the effects of the different substitutions at that protein position. By contrast, the two other columns show a narrow distribution of scores. In the middle, all substitutions are predicted to induce strong effects, implying high evolutionary conservation of the wild-type amino acid. On the right however, the low-resolution and weak predicted values may indicate that the amount of input evolutionary information is not sufficient to make a reliable prediction.



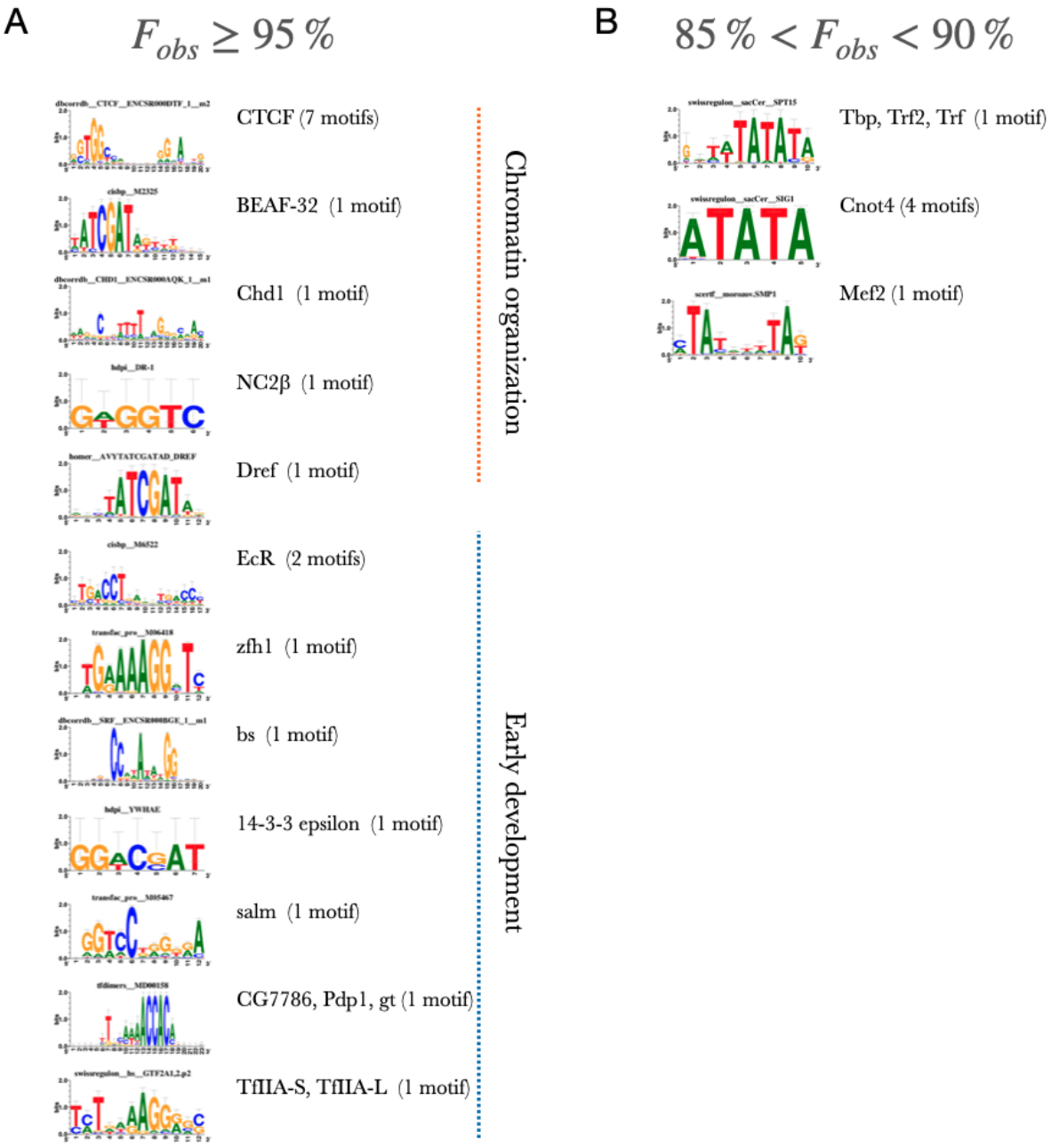
Extended Data Figure 2: **A. Minor Allele Frequencies (MAF) in function of predicted raw- and Z-scores for inbred and natural polymorphisms.** We report the MAF mean values in y-axis for different GEMME raw- and Z-score quantile bins ([0, 10^{-3}), (10^{-3}, 10^{-2}], …, (10^{-1}, 1]) in x-axis. **B-C. Distributions of GEMME raw scores** for the four missense mutation subsets in our custom benchmark: lethal (deep purple, B-C), hypomorphic alleles (pale pink, B-C), DGRP (sky blue, B), and DEST2 (steel blue, C).



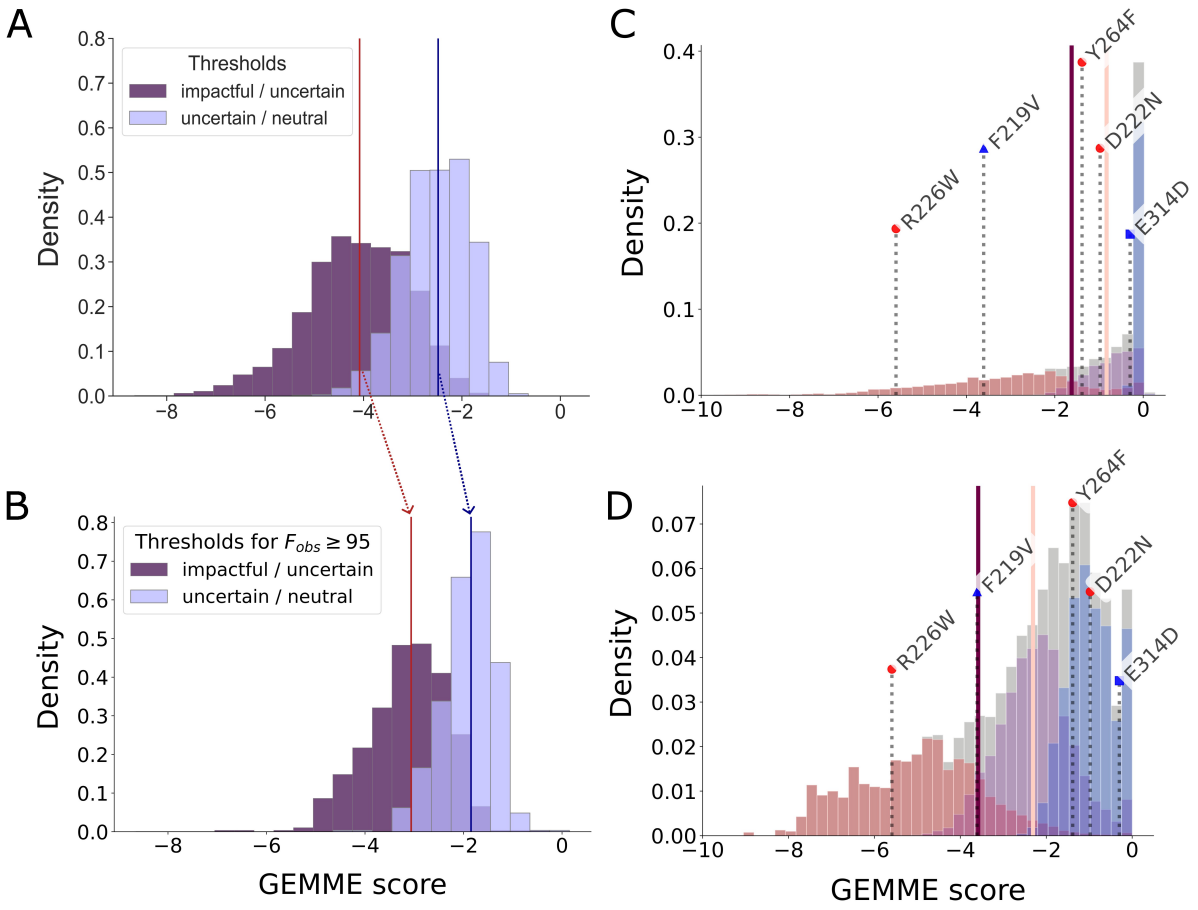
Extended Data Figure 3: **Comparison between ProteoCast and SIFT for Naprt-PH.** We ran SIFT through the webserver https://sift.bii.a-star.edu.sg/ in single-sequence mode (A-B) or inputting the same alignment as that used for ProteoCast (C-D). **A,C**. SIFT scores in function of GEMME raw scores. The respective distributions are shown on top and on the right. SIFT scores below 0.05 indicate impactful variants. **B,D**. Distributions of the number of unique values per residue for the two methods. In all plots, the values obtained for the poorly conserved unstructured regions 1-21 and 301-407 are excluded to reduce the risk of bias they might introduce. The full ProteoCast predictions for Naprt-PH can be accessed at: https://proteocast.ijm.fr/results/results/?q=FBpp0306840.



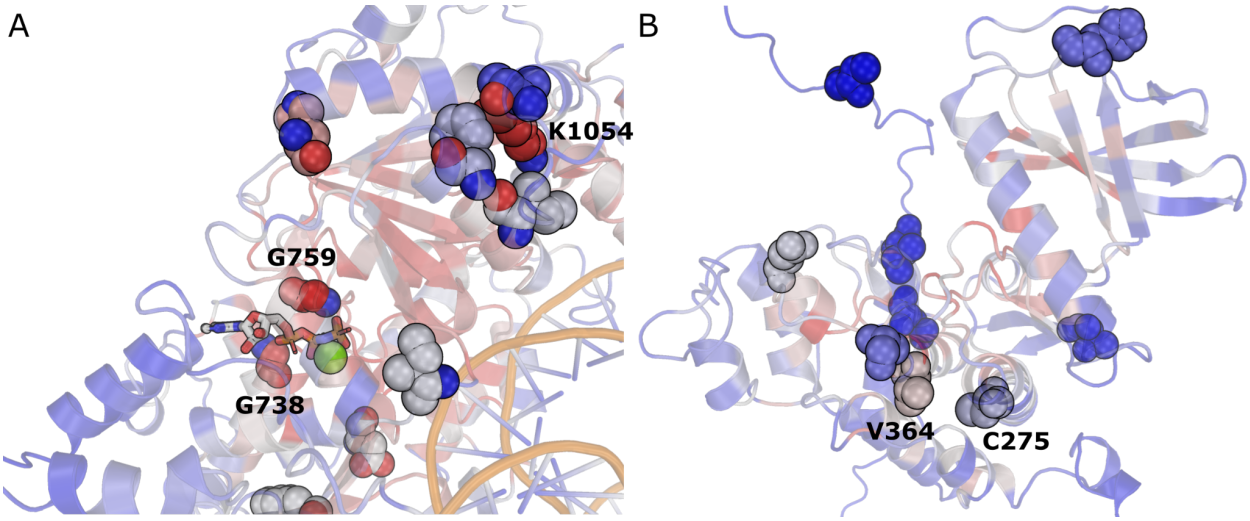
Extended Data Figure 4: **Per-residue mutational sensitivity scores mapped onto the AlphaFold2-predicted 3D model for Yorkie protein** (*yki* gene, FlyBase proteoform id: FBpp0301274, Uniprot id: Q45VV3). The color gradient goes from blue for close-to-zero scores (high tolerance to mutations) to red for very negative scores (low tolerance or high sensitivity). Annotated lethal mutations are highlighted with big spheres while natural and inbred polymorphisms are depicted with small spheres. The known PTM sites are shown as sticks. Three closeup views are shown on the right. I. Region from residue 241 to 343 encompassing the two annotated WW domains. II. Region from residue 31 to 80 sharing similarity with the human YAP1 segment interacting with TEAD. The experimental structure of the YAP1-TEAD complex (PDB id: 3KYS) is superimposed, with YAP1 in dark grey and TEAD in light grey. III. Segments 61-98 and 135-160 identified by ProteoCast as standing out from their surrounding background.



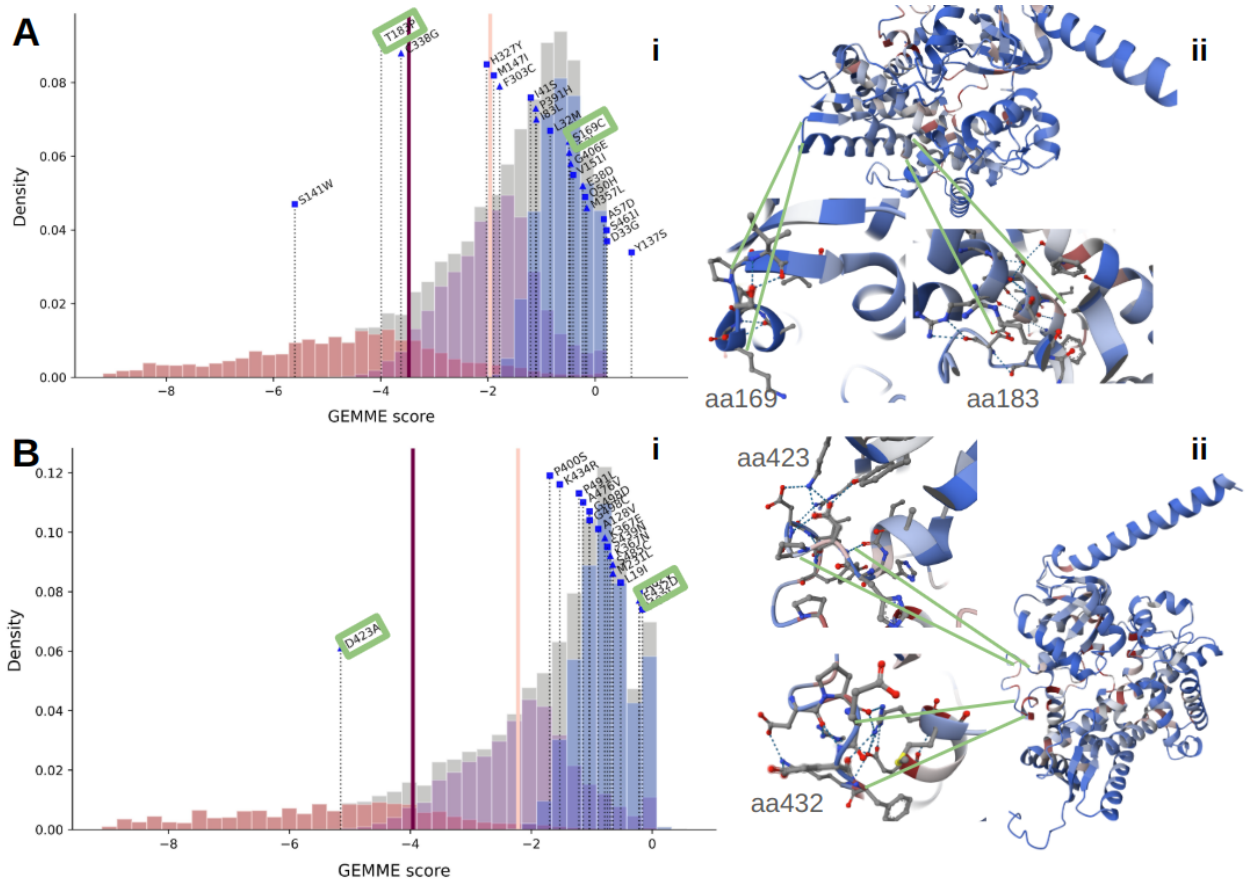
Extended Data Figure 5: **Motif enrichment in subsets of genes with different fractions of observed mutations. Motif enrichment NES > 5.**



Extended Data Figure 6: **A-B. Shift in GMM-based classification thresholds upon increasing the fraction of observed mutations. A.** Entire proteome. The median values for the impactful-uncertain and uncertain-neutral thresholds are of -4.08 and -2.48, respectively. **B.** Proteins with nearly complete coverage of their mutational landscape in the alignment ($F_{obs} \geq 95$). The median values increase up to -3.06 and -1.85. **C-D. ProteoCast classification for the protein-serine/threonine phosphatase Flapwing mutations (flw-PA proteoform, FBpp0071381).** The five known mutations are highlighted on the raw score distributions of Flapwing itself (C) and of Yorkie (D). The three Gaussians fitted to each distribution are colored in red, purple, and blue respectively. The pale pink vertical line represents the threshold above which mutations are considered *neutral*, the deep purple vertical line marks the threshold below which mutations are considered as *impactful*. Mutations with scores between the two threshold values are considered *uncertain*. Two lethal mutations that would appear as neutral in Yorkie (D) are classified as uncertain in Flapwing (C) highlighting the need for adaptive thresholds for variant classification.



Extended Data Figure 7: **3D mapping of natural polymorphism observed in wild flies and predicted as having strong effects on the protein function.** The 3D models for the query proteins were retrieved from the AlphaFold Database or predicted with ColabFold. They are depicted as cartoons and colored according to their per-residue predicted mutational sensitivity. **A.** Close-up view of the putative ATP- and DNA-binding domain of Mi-2 (FBpp0099808), a nuclear ATP-dependent nucleosome remodeler of the CHD family. The ATP and DNA were positioned by superimposing the bound 3D structure of its human homolog CHD4 (PDB id: 6RYR). The location of known mutations predicted as impactful are highlighted with spheres. The lethal mutation sites G738 and G759 are located near ATP while some of the sites exhibiting naturally occurring mutations form a 3D cluster near the DNA-binding interface. The lethal mutations G738D and G759S, and the DEST2 mutations K1054T have the most negative raw scores (below -6). **B.** Close-up view of the kinase domain of Doa (FBpp0289026), involved in somatic sex determination. The known mutation sites are highlighted with spheres. The residue V364, whose substitution into G is observed in DEST2 despite being predicted as having a strong effect (raw score < -6), is located very close to the lethal mutation site C275.



Extended Data Figure 8: **ProteoCast-informed discrimination between co-segregating natural polymorphisms. Panels i** show the ProteoCast score distributions. **Panels ii** show the AlphaFold-predicted structures colored with the ProteoCast-predicted mutational sensitivity. Cyp6a9 (**A**) and Cyp317a1 (**B**) belong to a family of detoxifying proteins catabolising insecticides amongst other xenobiotics that have been identified as showing natural polymorphism regionally affected, likely as a function of local agricultural uses. The proximity between the two missense mutation sites highlighted for each protein makes them likely to co-segregate and undistinguishable using a GWAS approach. The ProteoCast predictions can significantly help to prioritise one mutation over another, whether the two mutation sites are located in different secondary structure elements – T183P in an α-helix versus S169C in a loop (**Aii**), or they are both part of an unstructured region and one amino acid substitution implies more drastic physico-chemical changes than the other – D423A more drastic than E432D (**Bii**).