



Figures and figure supplements

Computational and cellular studies reveal structural destabilization and degradation of MLH1 variants in Lynch syndrome

Amanda B Abildgaard *et al*

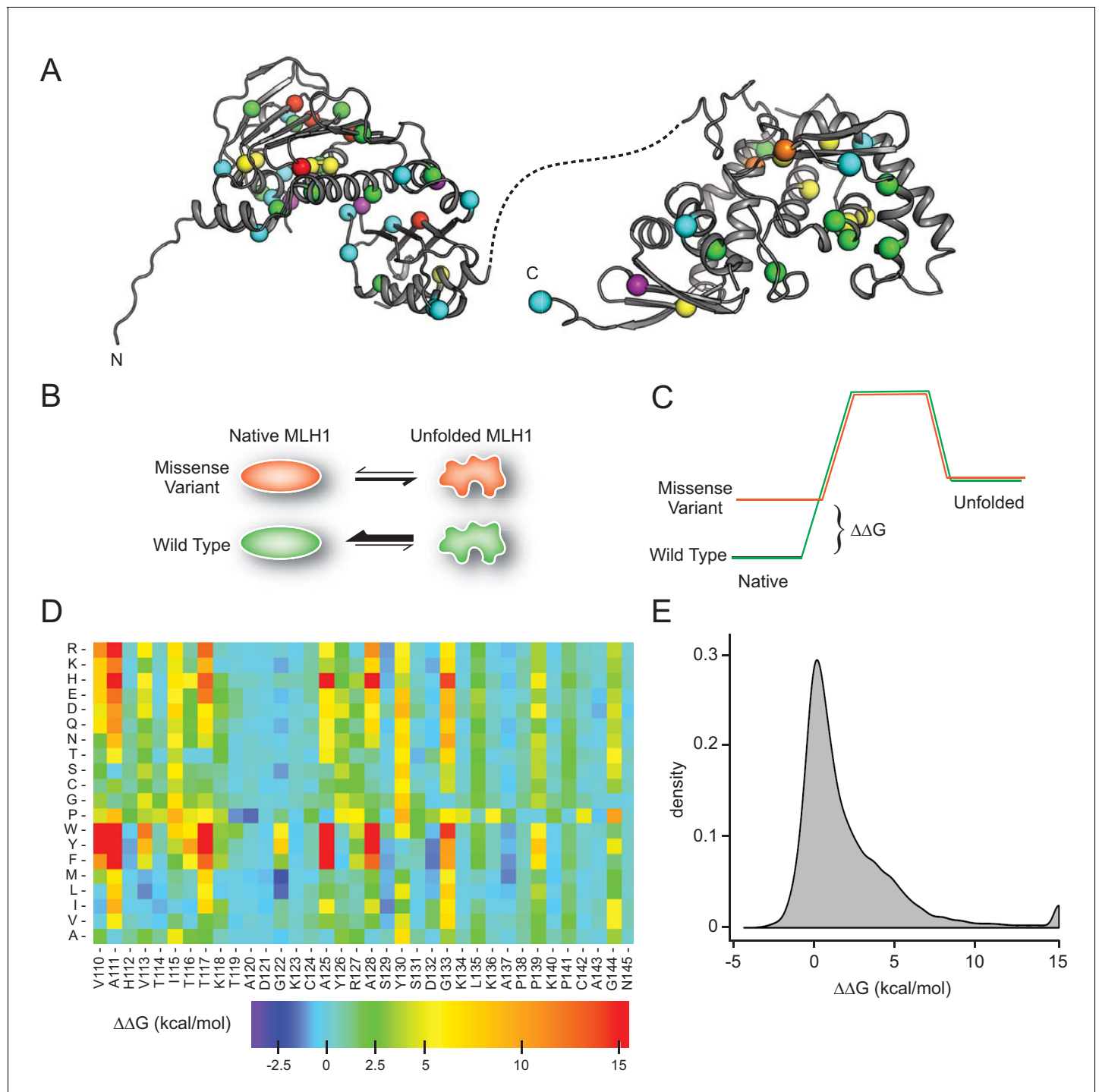


Figure 1. MLH1 structural stability predictions. (A) Structure of MLH1 (PDB IDs 4P7A and 3RBN). Positions of variants tested in this work are highlighted with colored spheres, indicating the predicted $\Delta\Delta G$ (<0.5 kcal/mol, purple, <1, cyan, <3.5 green, <7, yellow, <12, orange, >12, red). (B) Many disease-linked MLH1 missense variants (red) are structurally destabilized and therefore, compared to wild-type MLH1 (green), more likely to unfold. (C) The free energy of the folded conformation of a destabilized missense variant (red) is closer to that of the fully unfolded state. The employed stability calculations predict the difference of the free energy of unfolding ($\Delta\Delta G$) between a missense variant (red) and wild-type MLH1 (green). (D) Excerpt of the in silico saturation mutagenesis map (full dataset provided in **Supplementary file 1**). (E) Distribution of all predicted $\Delta\Delta G$ s from saturation mutagenesis. The peak at 15 kcal/mol contains all variants with $\Delta\Delta G$ values greater than this value.

DOI: <https://doi.org/10.7554/eLife.49138.002>

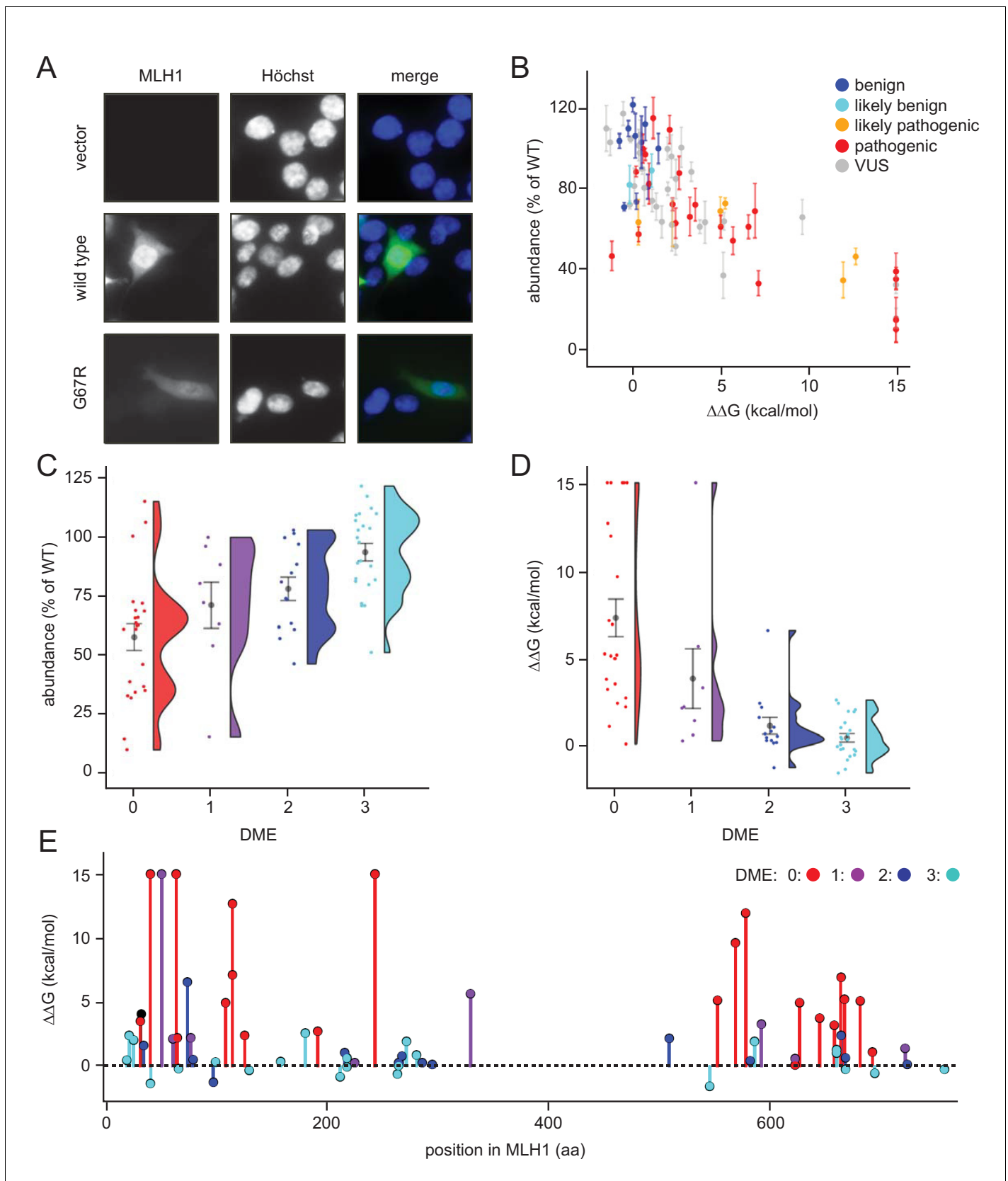


Figure 2. Steady-state levels of MLH1 variants correlate with structural stability predictions. (A) Example of the immunofluorescence imaging of HCT116 cells using antibodies to MLH1. Hoechst staining was used to mark the nucleus. Note the reduced steady-state levels of the G67R MLH1 variant

Figure 2 continued on next page

Figure 2 continued

compared to wild-type MLH1. (B) The total fluorescent intensity for each of the 69 different MLH1 variants was determined after excluding the non-transfected cells and normalizing the intensities to that for wild-type MLH1. The intensities were then plotted vs. the predicted $\Delta\Delta G$ values. Between 200 and 1,000 cells were included for each quantification. The error bars indicate the standard error of the mean ($n = 5$ experiments). Each variant is color-coded according to the ClinVar disease category. (C) Distribution of steady-state levels by DME category – 0 is loss-of-function in all assays by **Takahashi et al. (2007)**, 3 represents function in all these assays (for details see the Materials and Methods). Raincloud plot visualization as described in **Allen et al. (2018)**. Colored surface, smoothed density estimate. Gray dots represent means within each DME category, with bars for standard error. (D) Distribution of FoldX $\Delta\Delta G$ s across DME categories (as in (C)). (E) FoldX $\Delta\Delta G$ s for all variants tested in this work, indicating their position in the MLH1 sequence. As elsewhere, values above 15 kcal/mol were here set to this value.

DOI: <https://doi.org/10.7554/eLife.49138.005>

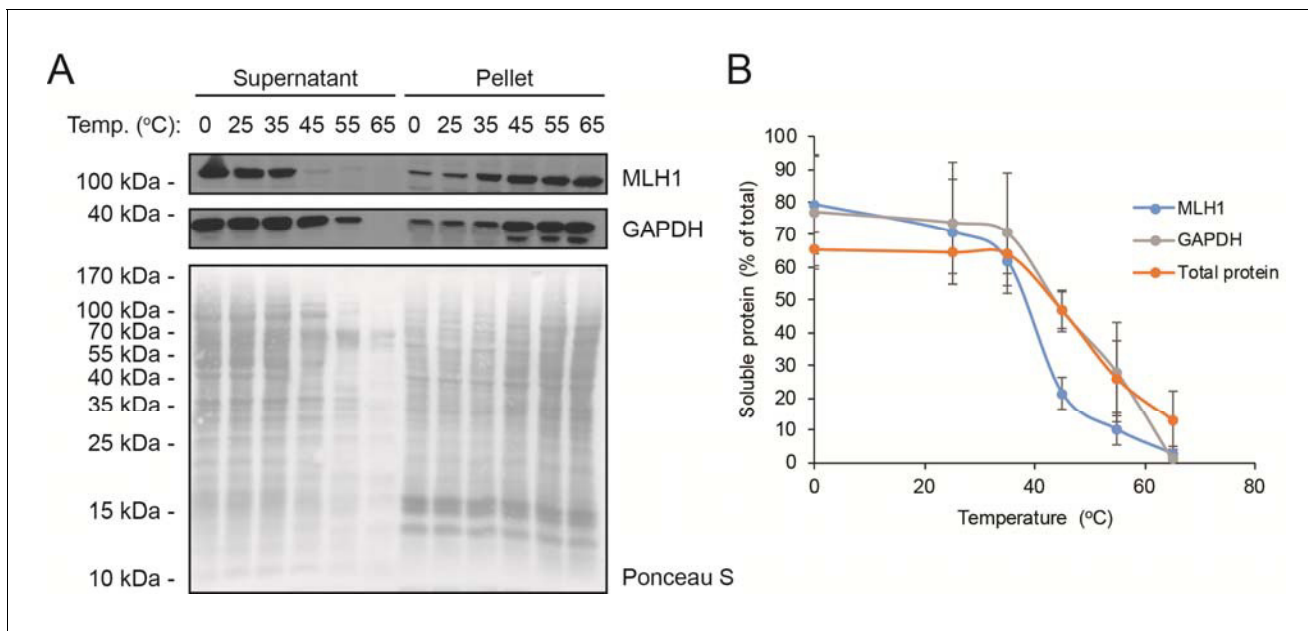


Figure 2—figure supplement 1. Solubility of wild-type MLH1 at a range of temperatures. **(A)** Transfected cells were lysed by sonication and incubated for 30 min. at the indicated temperatures. Then the lysates were separated into supernatant and pellet fractions by centrifugation. Western blotting using antibodies to MLH1 was employed to determine the amount of MLH1 in the fractions. Ponceau S staining and blotting with antibodies to GAPDH served as controls. **(B)** Quantification of blots as shown in panel **(A)** showing the amount of soluble protein (supernatant) normalized to total (supernatant + pellet). The total protein quantification is based on the full lanes of the Ponceau S stainings. The error bars indicate the standard deviation ($n = 3$).

DOI: <https://doi.org/10.7554/eLife.49138.006>

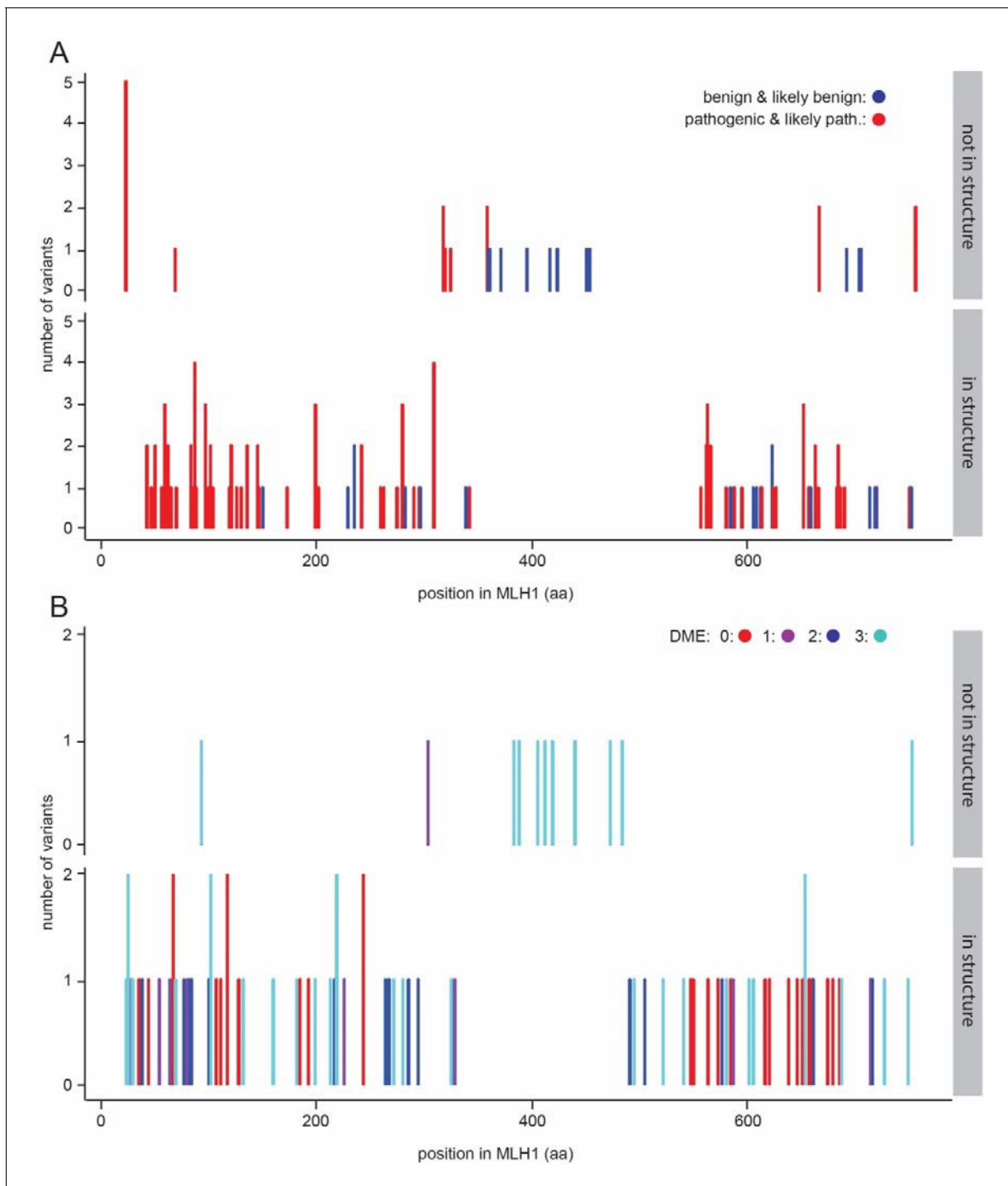


Figure 2—figure supplement 2. Variants within the central disordered region. (A) ClinVar benign and likely benign variants (blue) are found throughout the protein, while pathogenic and likely pathogenic variants (red) are enriched in the structured domains. (B) The variants previously tested for function (Takahashi et al., 2007) are distributed throughout the MLH1 protein. Similar to the observations from ClinVar, functional variants (DME score 3) are seen throughout the protein, while loss-of-function in multiple assays (DME scores.

DOI: <https://doi.org/10.7554/eLife.49138.007>

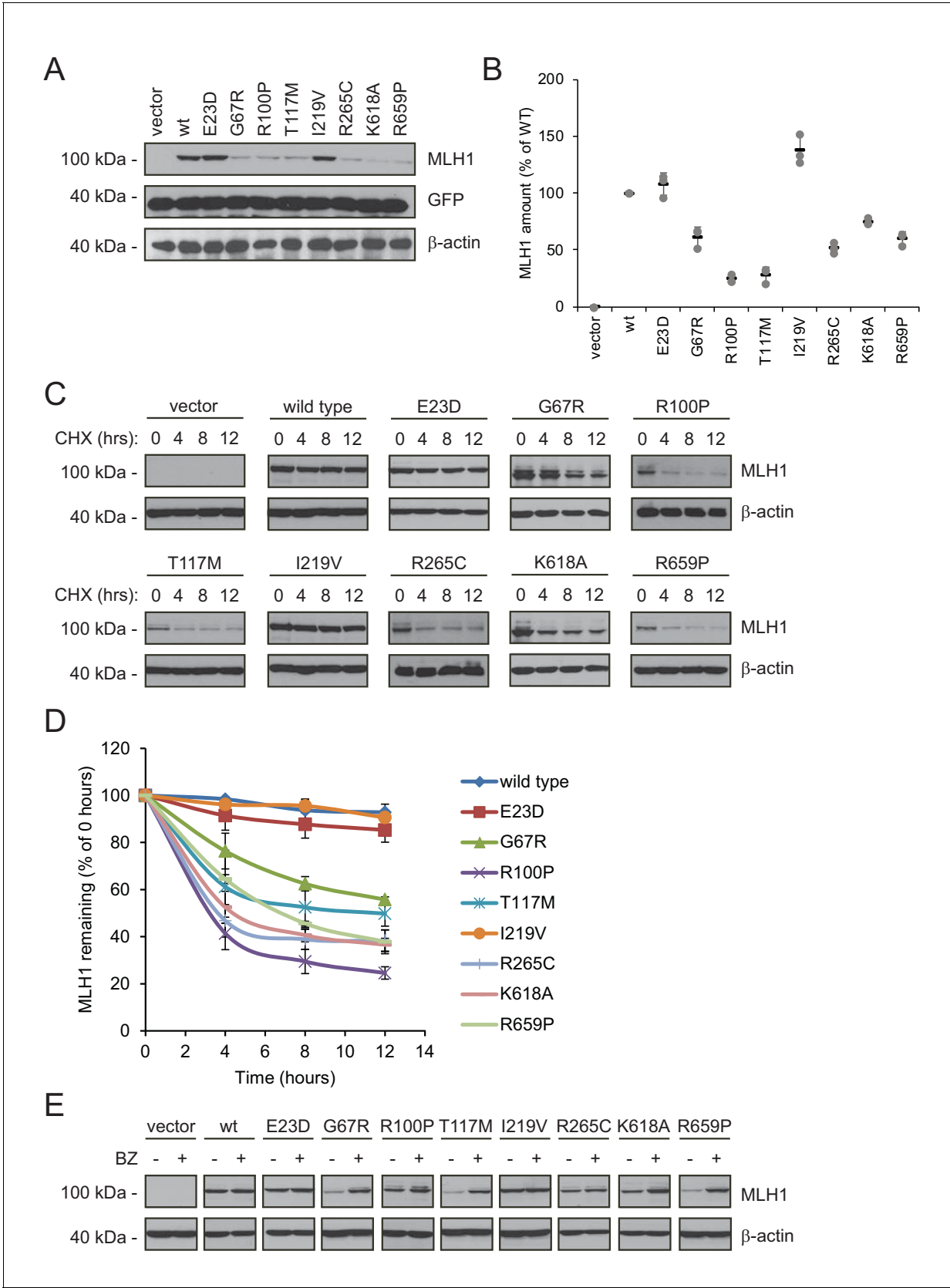


Figure 3. Many MLH1 variants are degraded by the proteasome. (A) HCT116 cells transfected with the indicated variants were analyzed by blotting with antibodies to MLH1. Co-transfection with a plasmid expressing GFP was included to test the transfection efficiencies between the MLH1 variants. β -

Figure 3 continued on next page

Figure 3 continued

actin served as a loading control. **(B)** Quantification of blots as in **(A)** normalized to the steady-state level of wild-type (WT) MLH1. The error bars show the standard deviation ($n = 3$). **(C)** MLH1-transfected HCT116 cells were treated with 25 $\mu\text{g/mL}$ cycloheximide (CHX) for 0, 4, 8 or 12 hr, and lysates were analyzed by blotting using antibodies to MLH1. β -actin was used as a loading control. **(D)** Quantification of blots as in panel **(C)**, normalized to the steady-state levels at $t = 0$ hr. The error bars indicate the standard deviation ($n = 3$). **(E)** Western blotting with antibodies to MLH1 of whole cell lysates from transfected cells either untreated or treated for 16 hr with 10 μM bortezomib (BZ). Blotting for β -actin was included as a loading control.

DOI: <https://doi.org/10.7554/eLife.49138.008>

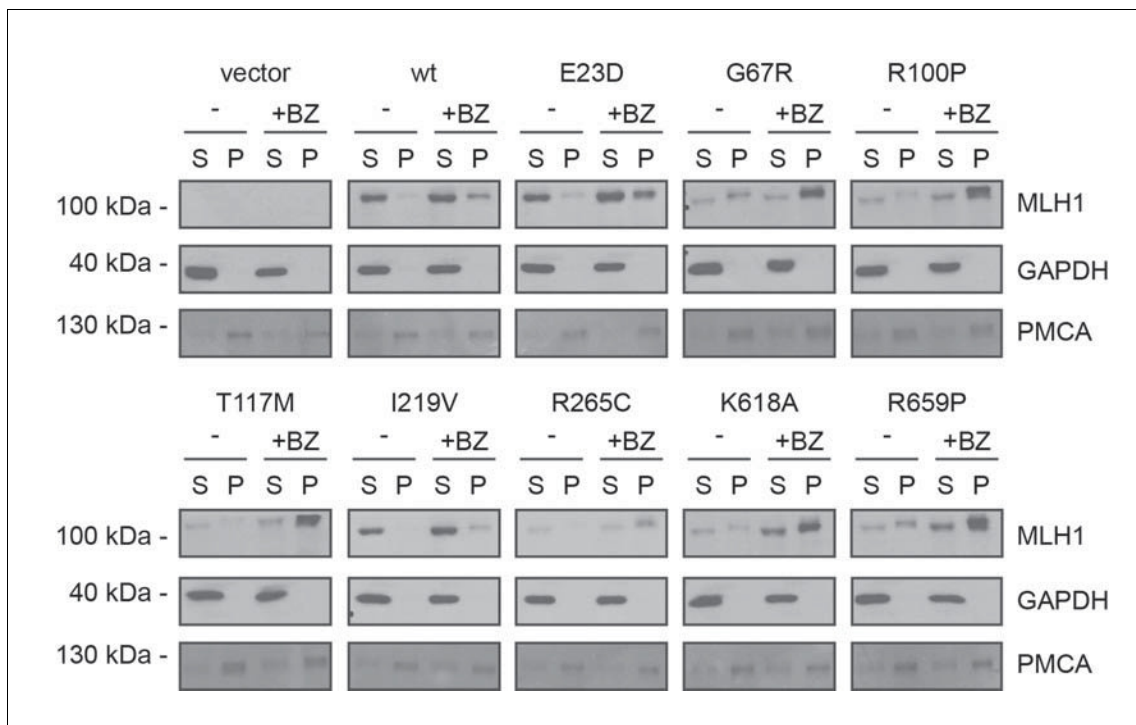


Figure 3—figure supplement 1. Solubility of selected MLH1 variants. Transfected cells, either untreated (-) or treated for 16 hr with 10 μ M bortezomib (+BZ) were lysed by sonication and immediately separated into supernatant (S) and pellet (P) fractions by centrifugation. Western blotting using antibodies to MLH1 was employed to determine the amount of MLH1 in the fractions. Blotting with antibodies to GAPDH and PMCA served as loading controls for the soluble and insoluble fraction, respectively.

DOI: <https://doi.org/10.7554/eLife.49138.009>

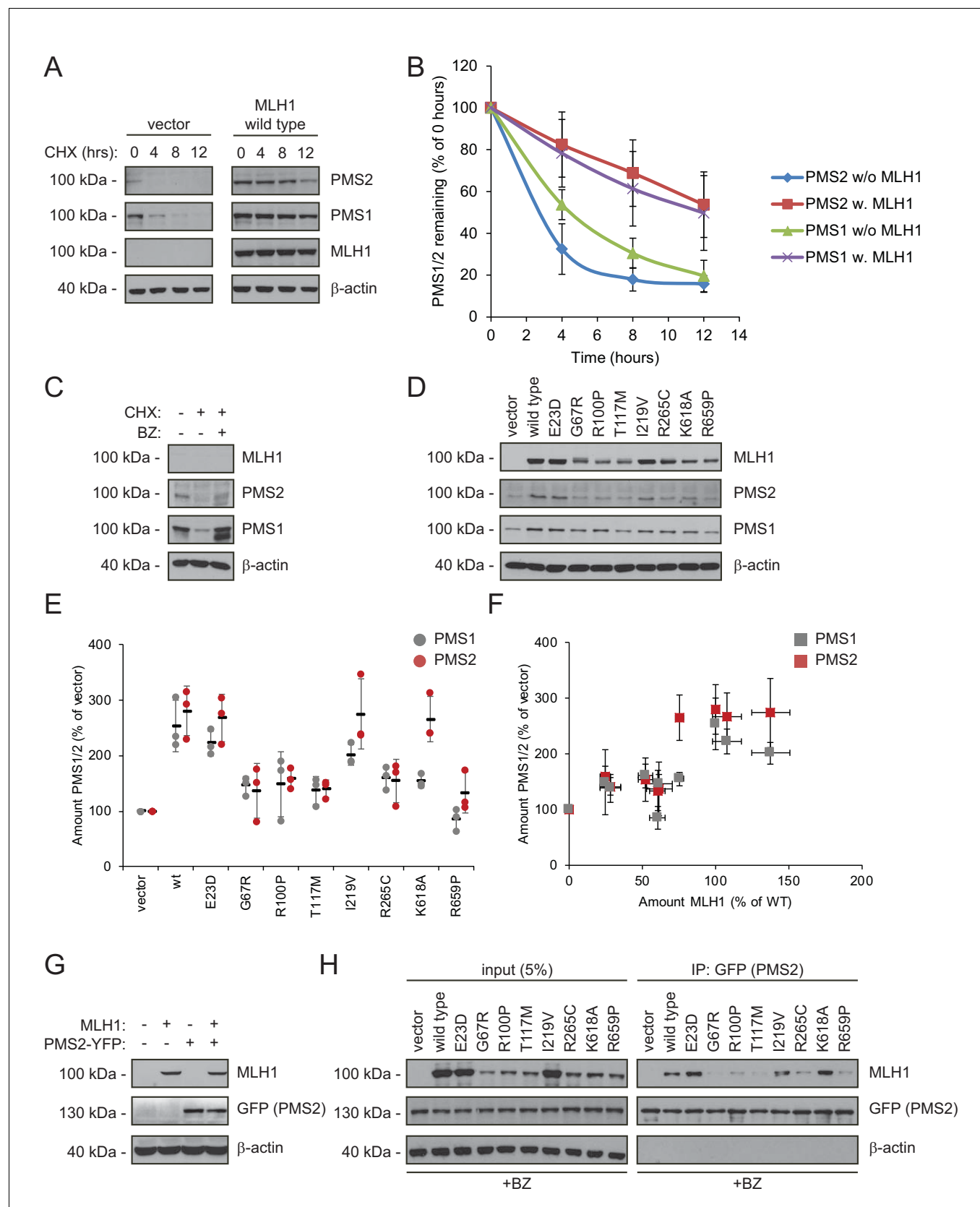


Figure 4. Stable MLH1 variants increase steady-state levels of PMS1 and PMS2. **(A)** The levels of endogenous PMS1 and PMS2 were determined by blotting of whole-cell lysates of HCT116 cells transfected with either empty vector or with wild-type MLH1 and treated with 25 μ g/mL cycloheximide

Figure 4 continued on next page

Figure 4 continued

(CHX) for 0, 4, 8 or 12 hr. The antibodies used were to PMS1 and PMS2, and as a control to MLH1. β -actin served as loading control. **(B)** Quantification of blots as in panel **(A)** normalized to protein levels at 0 hr. The error bars indicate the standard deviation ($n = 3$). **(C)** The levels of endogenous MLH1, PMS1 and PMS2 were compared by blotting of cell lysates of HCT116 cells either untreated, or treated with cycloheximide (CHX) or with bortezomib (BZ) and CHX. β -actin served as loading control. **(D)** The levels of endogenous PMS1 and PMS2 and transfected MLH1 were compared by western blotting using antibodies to PMS1, PMS2 and MLH1. β -actin served as loading control. **(E)** Quantification of blots as in panel **(C)** normalized to the level of endogenous PMS1 (grey) or PMS2 (red) in untransfected HCT116 cells. The error bars show the standard deviation ($n = 3$). **(F)** Plotting the levels of the MLH1 variants vs. the levels of endogenous PMS1 (grey) and PMS2 (red). The error bars show the standard deviation ($n = 3$). **(G)** The levels of MLH1 and YFP-tagged PMS2 were analyzed by SDS-PAGE and blotting of whole-cell lysates of HCT116 cells transfected with the indicated expression plasmids. β -actin was included as loading control. **(H)** Co-transfected PMS2-YFP was immunoprecipitated (IP) using GFP-trap beads, and the precipitated material was analyzed by electrophoresis and blotting. Bortezomib was added to all cultures 16 hr prior to cell lysis to ensure ample amounts of the unstable MLH1 variants.

DOI: <https://doi.org/10.7554/eLife.49138.010>

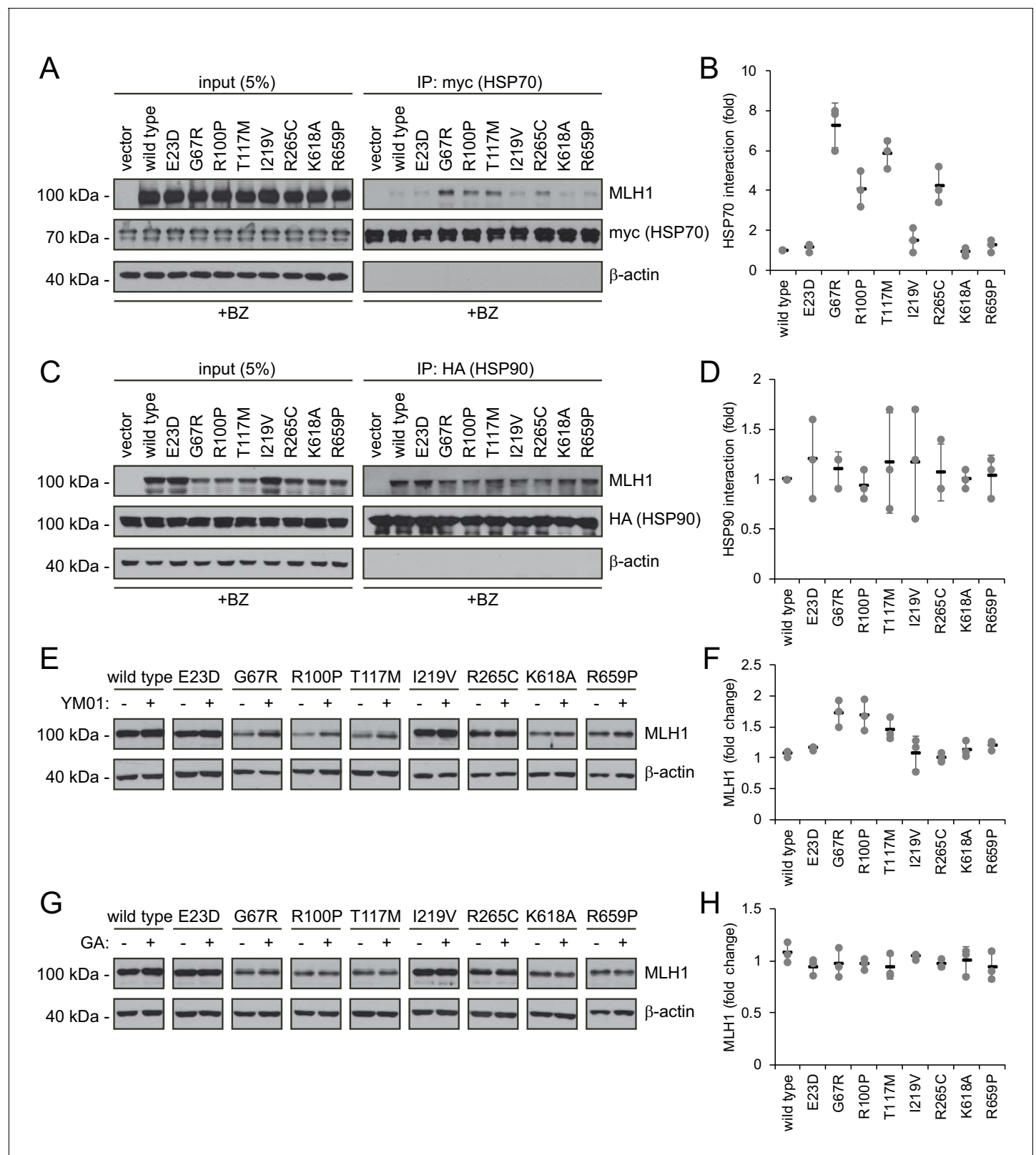


Figure 5. Molecular chaperones play a role in the proteasomal degradation of MLH1. (A) Co-transfected HSP70-myc was immunoprecipitated (IP) using myc-trap beads and analyzed by blotting with antibodies to the myc-tag (HSP70) and MLH1. Bortezomib was added to all cultures 8 hr prior to cell lysis to ensure ample amounts of the unstable MLH1 variants. (B) Quantification of blots as shown in panel (A) normalized to level of wild-type MLH1. The Figure 5 continued on next page

Figure 5 continued

error bars indicate the standard deviation ($n = 3$). (C) Co-transfected HSP90-HA was immunoprecipitated (IP) with anti-HA resin, and the precipitated material analyzed by electrophoresis and western blotting using antibodies to the HA-tag (HSP90) and MLH1. As above, bortezomib was added to all cultures prior to cell lysis. (D) Quantification of blots as in panel (C) normalized to amount of precipitated wild-type MLH1. The error bars show the standard deviation ($n = 3$). (E) Western blotting using antibodies to MLH1 of whole-cell lysates from transfected cells treated with 5 μM YM01 for 24 hr as indicated. (F) Quantification of blots as shown in panel (E) normalized to level of MLH1 without YM01. The error bars indicate the standard deviation ($n = 3$). (G) Western blotting using antibodies to MLH1 of whole-cell lysates from transfected cells treated 1 μM geldanamycin (GA) for 24 hr. (H) Quantification of blots as shown in panel (G) normalized to level of MLH1 without GA. The error bars indicate the standard deviation ($n = 3$).

DOI: <https://doi.org/10.7554/eLife.49138.011>

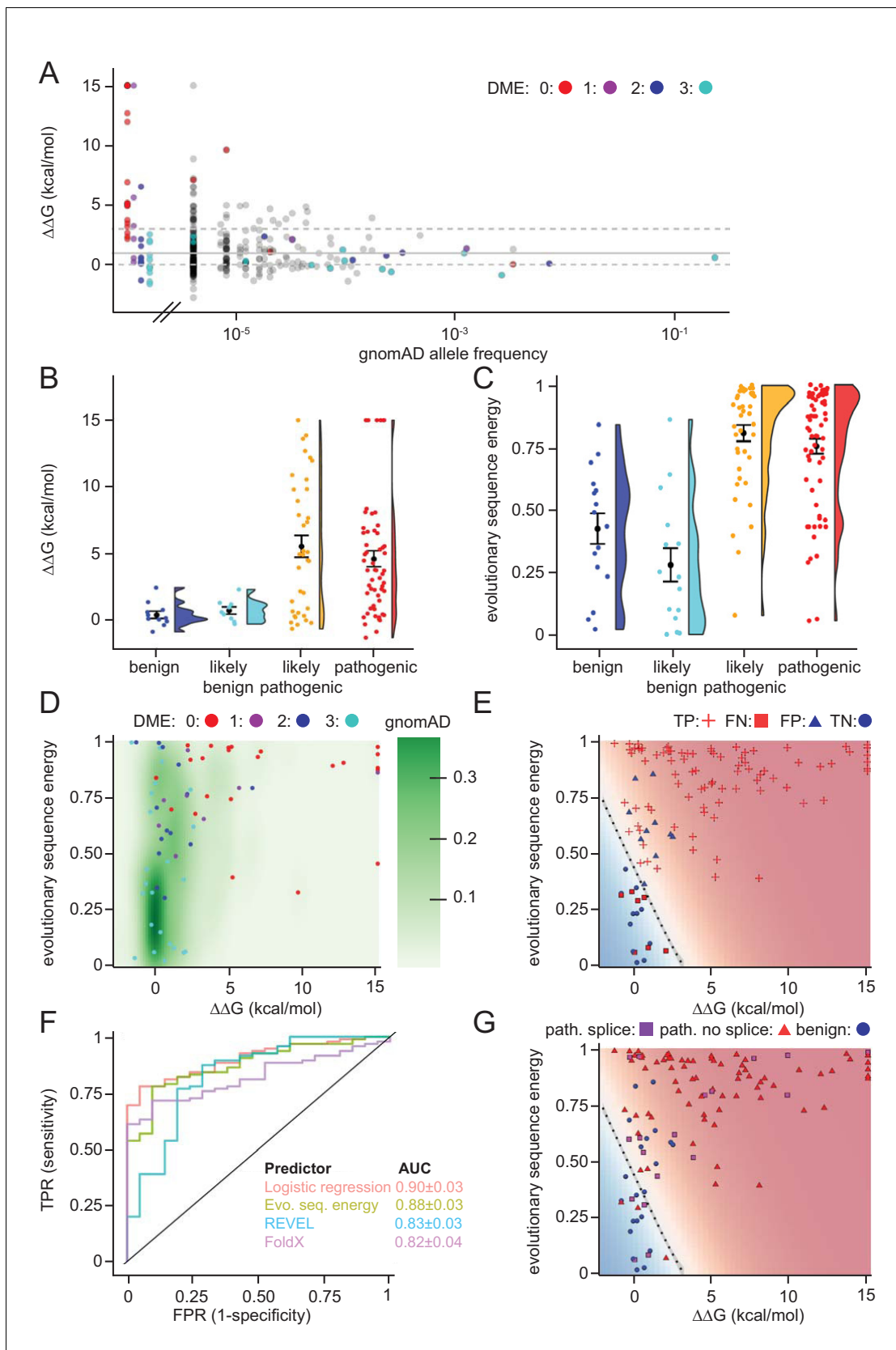


Figure 6. Assessing stability calculations for predicting pathogenicity. (A) 'Fishtail plot' of $\Delta\Delta G$ -values vs. allele frequencies for all variants listed in gnomAD (gray), as well as those analyzed by *Takahashi et al. (2007)*; the latter are color-coded by DME. Note that the leftmost group of colored dots

Figure 6 continued on next page

Figure 6 continued

are variants that have been reported in patients, but are not recorded in gnomAD (thus their allele frequency in gnomAD is zero). Variants with common to intermediate frequencies are all predicted to be stable, while some rare variants are predicted to be destabilized. $\Delta\Delta G$ s for gnomAD variants are provided as source data (**Figure 6—source data 1**), those for variants characterized by **Takahashi et al. (2007)** in **Table 1** and source data (**Table 1—source data 1**). (B) FoldX $\Delta\Delta G$ for benign (blue), likely benign (cyan), likely pathogenic (orange), and pathogenic (red) variants that are reported in ClinVar with ‘at least one star’ curation. The whiskers represent the mean and standard error of the mean. (C) Evolutionary sequence energies for ClinVar-reported variants, color scheme as in (B). The whiskers represent the mean and standard error of the mean. (D) Landscape of variant tolerance by combination of changes in protein stability (x axis) and evolutionary sequence energies (y axis), such that the upper right corner indicates most likely detrimental variants, while those in the lower left corner are predicted stable and observed in MLH1 homologs. The green background density illustrates the distribution of all variants listed in gnomAD. The combination of metrics captures most non-functional variants (DME scores 0 or 1). Outliers are discussed in the main text. (E) Logistic regression model of FoldX $\Delta\Delta G$ s and evolutionary sequence energies. Pathogenic variants in red, benign in blue. Dot shape indicates whether pathogenicity of the respective variant was correctly predicted by a regression model trained on all but this data point (‘jackknife’, TP, true positives, FN, false negatives, FP, false positives, TN, true negatives). Parameters for a model trained on the full dataset are: FoldX $\Delta\Delta G$ weight 0.52, evolutionary sequence energy weight 3.50, intercept -1.55 . (F) ROC curves for logistic regression model, FoldX $\Delta\Delta G$ s, evolutionary sequence energies, and the ensemble-predictor REVEL to assess their performance in separating benign from pathogenic variants. TPR, true positive rate. FPR, false positive rate. Standard deviations in AUC were determined by performing 100 ROC analyses on randomly sampled but balanced subsets, so that there are equal numbers of positive and negative cases. (G) Integrating potential effects these variants may have on splicing in the genomic context. Purple squares indicate pathogenic variants that are predicted to affect splicing (SpliceAI, threshold 0.5). No benign variants are predicted to affect splicing. Mapping to genomic loci (**Yates et al., 2015**) and SpliceAI Scores for ClinVar entries used in this work are provided as source data (**Figure 6—source data 2**).

DOI: <https://doi.org/10.7554/eLife.49138.012>

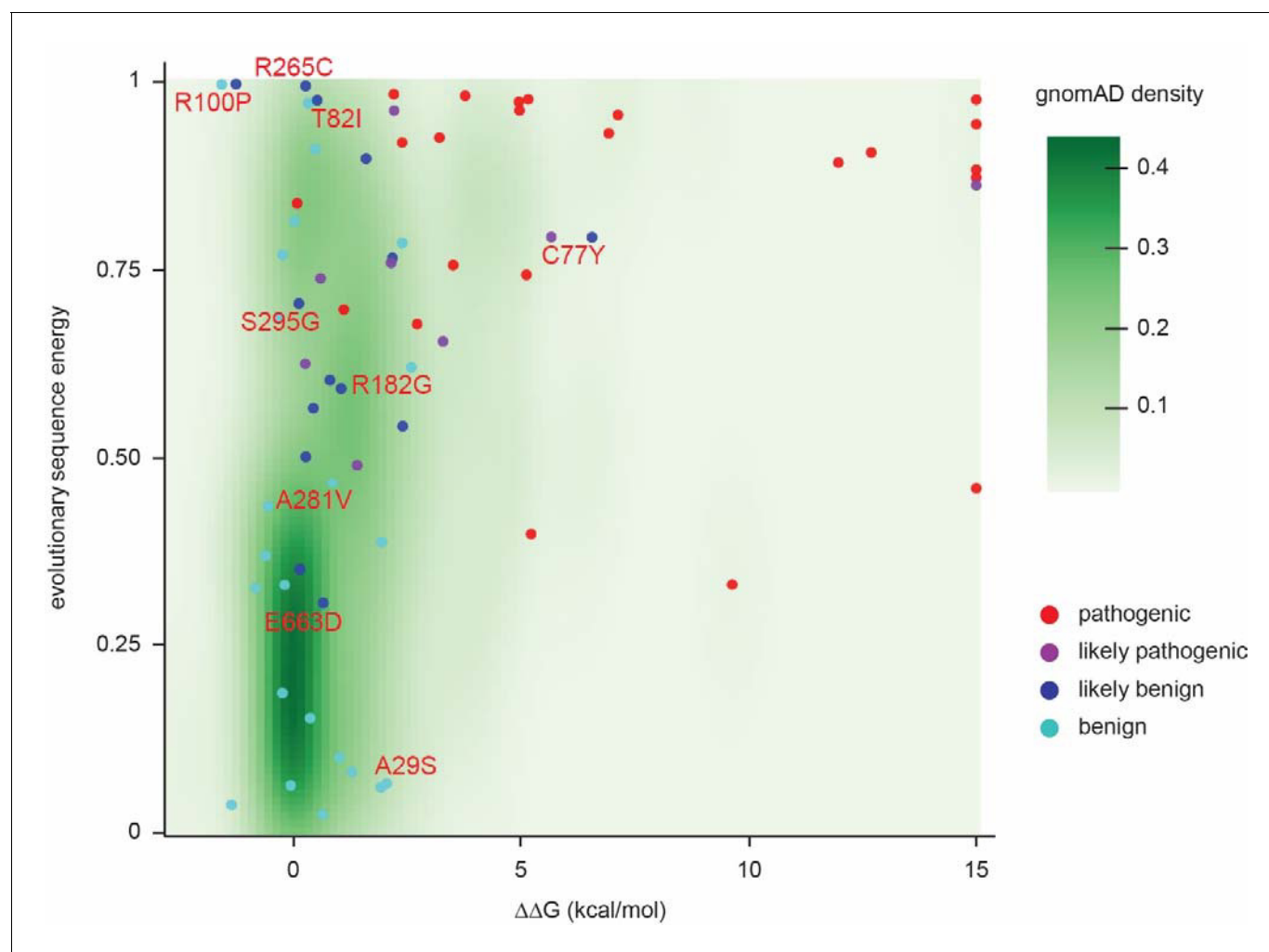


Figure 6—figure supplement 1. Landscape of ClinVar MLH1 variant tolerance. Landscape of variant tolerance by combination of changes in protein stability (x axis) and evolutionary sequence energies (y axis), such that the upper right corner indicates most likely detrimental variants, while those in the lower left corner are predicted stable and observed in MLH1 homologs. The green background density illustrates the distribution of all variants listed in gnomAD. Outliers are discussed in the main text.

DOI: <https://doi.org/10.7554/eLife.49138.013>

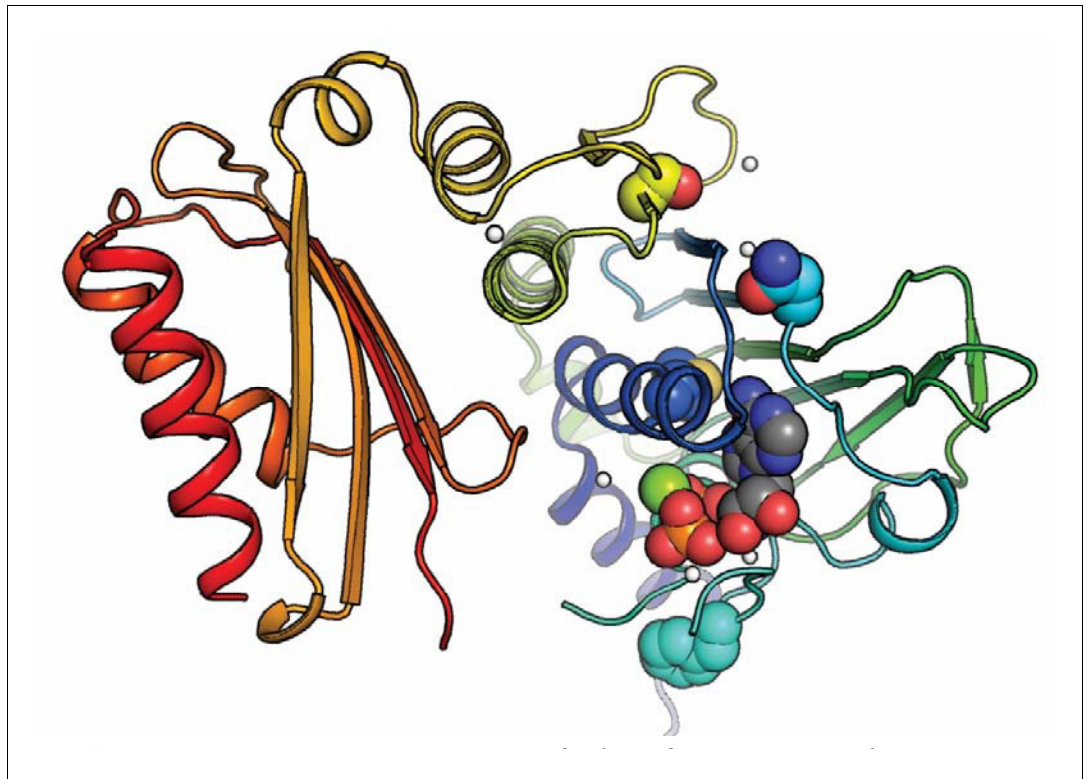


Figure 6—figure supplement 2. Positioning of selected variants near the active site in the N-terminal domain. The MLH1 structure (PDB: 4P7A) (Wu et al., 2015) with highlighted stable loss-of-function MLH1 variants (M35R, N64S, F80V, S193P) as spheres, as well as ADP. The domain is colored in a rainbow color scheme, with blue at the N-terminus and red at the C-terminal end of the N-terminal domain (MLH1 sequence position ~300); sidechain and ligand oxygen atoms are red, nitrogens blue, ligand carbons gray. Three of the stable loss-of-function positions (M35R, N64S, F80V) are very close to the ligand. Variation at these sites may thus interfere with ligand binding, which could explain why they lead to loss of function despite wild-type-like cellular protein levels.

DOI: <https://doi.org/10.7554/eLife.49138.014>

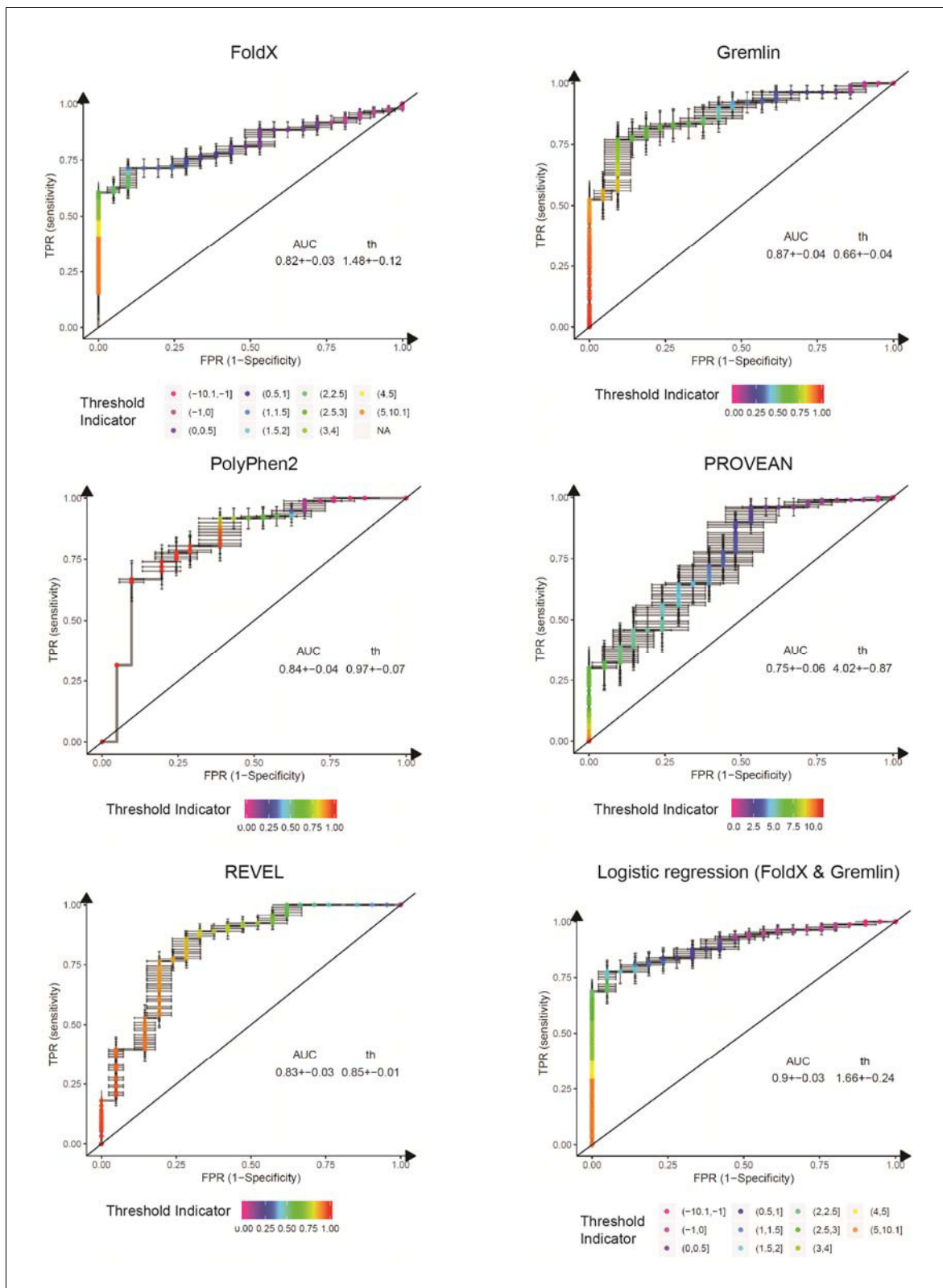


Figure 6—figure supplement 3. ROC curves for variant consequence predictors tested in this work. The indicated predictors were used to score all benign, likely benign, likely pathogenic and pathogenic missense variants reported in ClinVar (Aug 2018). We merged benign and likely benign variants into the negative control set and pathogenic and likely pathogenic into the positive set and generated ROC curves for each predictor, assessing their power in separating benign from pathogenic variants. In addition to the individual predictors, we also assessed a logistic regression model combining Figure 6—figure supplement 3 continued on next page

Figure 6—figure supplement 3 continued

FoldX $\Delta\Delta G$ and Gremlin evolutionary sequence energy (\tilde{E}). AUC, area under the curve. th, threshold (units specific to the respective predictor). Error bars indicate standard deviation from 100 randomly sub-sampled balanced ROC curves. See also **Figure 6E** in the main manuscript.

DOI: <https://doi.org/10.7554/eLife.49138.015>

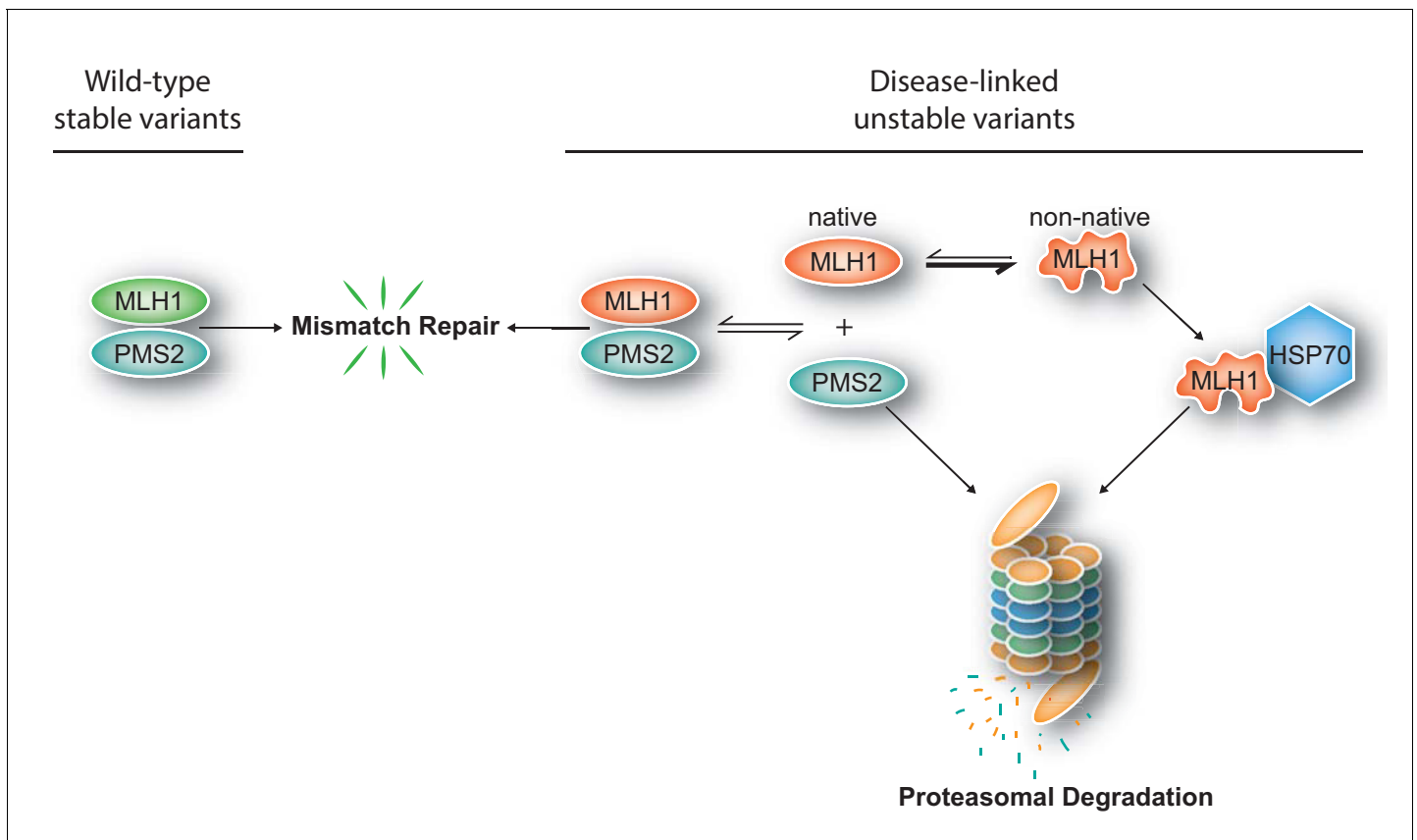


Figure 7. Model for how structural destabilization of MLH1 contributes to disease. The wild-type (green) MLH1-PMS2 heterodimer promotes DNA mismatch repair. Disease-linked missense MLH1 variants (red) may also promote DNA repair, but are at risk of dissociating from PMS2 due to structural destabilization. The structural destabilization of MLH1 may also cause a partial unfolding of MLH1 which is recognized by the molecular chaperone HSP70 and causes proteasomal degradation of the MLH1 variant. In turn, the degradation of MLH1 leaves PMS2 without a partner protein, resulting in proteasomal degradation of PMS2.

DOI: <https://doi.org/10.7554/eLife.49138.018>