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

Amanda B. Abildgaard[1], Amelie Stein[1,*], Sofie V. Nielsen[1], Katrine Schultz-Knudsen[1], Elena Papaleo[1,2], Amruta Shrikhande[3], Eva R. Hoffmann[3], Inge Bernstein[4], Anne-Marie Gerdes[5], Masanobu Takahashi[6], Chikashi Ishioka[6], Kresten Lindorff-Larsen[1,*] and Rasmus Hartmann-Petersen[1,*]

Running title: Degradation of cancer-linked MLH1 variants

Keywords: protein folding, misfolding, protein quality control, proteasome, HSP70

1: The Linderstrøm-Lang Centre, Department of Biology, University of Copenhagen, Ole Maaløes Vej 5, DK-2200 Copenhagen, Denmark.

2: Current address: Computational Biology Laboratory, Danish Cancer Society Research Center, Strandboulevarden 49, DK-2100 Copenhagen, Denmark.

3: Department of Cellular and Molecular Medicine, University of Copenhagen, Blegdamsvej 3B, DK-2200 Copenhagen, Denmark.

4: Department of Surgical Gastroenterology, Aalborg University Hospital, DK-9000 Ålborg, Denmark.

5: Department of Clinical Genetics, Rigshospitalet, Blegdamsvej 9, DK-2100 Copenhagen, Denmark.

6: Tohoku University Hospital, Tohoku University, Sendai, Japan.

*Corresponding authors: A.S. (amelie.stein@bio.ku.dk), K.L.-L. (lindorff@bio.ku.dk) and R.H.-P. (rhpetersen@bio.ku.dk).

Abstract

Defective mismatch repair leads to increased mutation rates, and germline loss-of-function variants in the repair component MLH1 cause the hereditary cancer predisposition disorder known as Lynch syndrome. Early diagnosis is important, but complicated by many variants being of unknown significance. Here we show that a majority of the disease-linked MLH1 variants we studied are present at reduced cellular levels. We show that destabilized MLH1 variants are targeted for chaperone-assisted proteasomal degradation, resulting also in degradation of co-factors PMS1 and PMS2. *In silico* saturation mutagenesis and computational predictions of thermodynamic stability of MLH1 missense variants revealed a correlation between structural destabilization, reduced steady-state levels and loss-of-function. Thus, we suggest that loss of stability and cellular degradation is an important mechanism underlying many *MLH1* variants in Lynch syndrome. Combined with analyses of conservation, the thermodynamic stability predictions separate disease-linked from benign *MLH1* variants, and therefore hold potential for Lynch syndrome diagnostics.

Introduction

The DNA mismatch repair (MMR) pathway corrects mismatched base pairs inserted during replication. The MutSα (MSH2-MSH6) heterodimer initiates repair by detecting the mismatch after which the MutLα (MLH1-PMS2) heterodimer promotes the process by generating a nick in the newly synthesized DNA strand, thereby stimulating downstream repair proteins (Jiricny, 2006; Jun et al., 2006). The MMR pathway is phylogenetically highly conserved, emphasizing its importance as a key DNA repair mechanism of the cell (Jiricny, 2013; Sachadyn, 2010). Loss of MMR activity causes genome instability, and can result in both sporadic and inherited cancer, such as Lynch Syndrome (LS) (OMIM: #609310), also known as hereditary nonpolyposis colorectal cancer (HNPCC). The predominant consequence of LS is colorectal cancer (CRC), making LS the underlying reason for around 4% of all CRC cases (Aarnio et al., 1999; Hampel et al., 2008; Sijmons & Hofstra, 2016; Vasen et al., 1996; Vasen & de Vos Tot Nederveen Cappel WH, 2013; Thompson et al., 2014; Moller et al., 2018). Importantly, the cumulative lifetime cancer risk varies considerably between patients and depends on the specific germline mutation in the genes encoding the key mismatch repair proteins MSH2, MSH6, MLH1, and PMS2 (Barrow et al., 2008; Dowty et al., 2013; Dunlop et al., 1997; Lynch et al., 2015; Peltomaki et al., 1993; Plaschke et al., 2004; Sijmons & Hofstra, 2016).

The majority of LS cases result from *MLH1* and *MSH2* mutations (Peltomaki, 2016), many of which are missense mutations (Heinen, 2010; Palomaki et al., 2009; Peltomaki & Vasen, 1997; Peltomaki, 2016). Evidently, such missense mutations may cause loss-of-function by directly perturbing protein-protein interactions or ablating enzymatic activity. Many missense mutations, however, cause loss-of-function by inducing structural destabilization of the protein (Stein et al., 2019), which in turn may trigger protein misfolding and degradation by the ubiquitin-proteasome system (UPS) (Kampmeyer et al., 2017; Nielsen et al., 2014; Kriegenburg et al., 2014). As a result, the cellular amount of a missense protein may be reduced to an insufficient level, which can ultimately cause disease (Ahner

et al., 2007; Casadio et al., 2011; Matreyek et al., 2018; Nielsen et al., 2017), as we and others have previously shown for LS-linked variants of MSH2 (Gammie et al., 2007; Arlow et al., 2013; Nielsen et al., 2017).

In this study, we investigated whether this is the case for LS-linked variants of the MLH1 protein. We determined cellular abundance for 69 missense variants, and show that several destabilized LS-linked MLH1 variants are targeted for chaperone-assisted proteasomal degradation and are therefore present at reduced cellular amounts. In turn, this lower amount of MLH1 results in degradation of the MLH1-binding proteins PMS1 and PMS2. *In silico* saturation mutagenesis and computational prediction of the thermodynamic stability of all possible MLH1 single site missense variants revealed a correlation between the structural destabilization of MLH1, reduced steady-state levels and the loss-of-function phenotype. Accordingly, the thermodynamic stability predictions accurately separate disease-linked *MLH1* missense mutations from benign *MLH1* variants (area under the curve is 0.82 in a receiver-operating characteristic analysis), and therefore hold potential for classification of *MLH1* missense variants of unknown consequence, and hence for LS diagnostics. Further, by suggesting a mechanistic origin for many LS-causing *MLH1* missense variants our studies provide a starting point for development of novel therapies.

Results

In silico saturation mutagenesis and thermodynamic stability predictions

Most missense proteins are less structurally stable than the wild-type protein (Tokuriki & Tawfik, 2009), and individual missense variants may thus lead to increased degradation and insufficient amounts of protein. To comprehensively assess this effect for MLH1, we performed energy calculations based on crystal structures of MLH1 to predict the consequences of missense mutations in *MLH1* on the thermodynamic stability of the MLH1 protein structure. Full-length human MLH1 is a 756 residue protein which forms two folded units, an N-terminal domain (residues 7-315) and a C-terminal domain (residues 502-756) (Mitchell et al., 2019) separated by a flexible and intrinsically disordered linker (Fig. 1A). Using the structures (Wu et al., 2015) of the two domains (PDB IDs 4P7A and 3RBN) (Fig. 1A), we performed *in silico* saturation mutagenesis, introducing all possible single site amino acid substitutions into the wild-type MLH1 sequence at the 564 structurally resolved residues. We then applied the FoldX energy function (Schymkowitz et al., 2005) to estimate the change in thermodynamic folding stability compared to the wild-type MLH1 protein (ΔΔG) (Fig. 1BC). Negative values indicate mutations that are predicted to stabilize MLH1, while positive values indicate that the mutations may destabilize the protein. Thus, those variants with ΔΔG predictions > 0 kcal/mol are expected to have a larger population of fully or partially unfolded structures that, in turn, may be prone to protein quality control (PQC)-mediated degradation. Our saturation mutagenesis dataset comprises 19 (amino acids, excluding the wild-type residue) * 564 (residues resolved in the N- and C-terminal structures) = 10,716 different MLH1 variants, thus covering 75% of all possible missense variants in MLH1. We illustrate a subsection as a heat map in Fig. 1D (the entire dataset is included in the supplemental material, supplemental material file 1). The predictions reveal that 34% of the substitutions are expected to change the stability of MLH1 by less than 0.7

106 kcal/mol, which is the typical error of the predictions (Guerois et al., 2002) (Fig. 1E). A comparable
107 fraction (32%) are, however, predicted to cause a substantial destabilization (>2.5 kcal/mol) of the
108 MLH1 protein (Fig. 1E).

109

110 **Thermodynamic stability calculations predict severely reduced MLH1 steady-state levels**

111 To test whether the *in silico* stability predictions are predictive of cellular stability, abundancy, and
112 function, we selected 69 naturally occurring MLH1 missense variants with predicted ΔΔGs spanning
113 from -1.6 to >15 kcal/mol (Table 1). We further ensured that the selected mutations were distributed
114 throughout the *MLH1* gene, thus probing the entire structured parts of the MLH1 protein (Fig. 1A).
115 Then, the variants were introduced into *MLH1*-negative HCT116 cells and analyzed by automated
116 immunofluorescence microscopy using a polyclonal antiserum to MLH1.
117 As expected, wild-type MLH1 localized primarily to the nucleus (Fig. 2A). This localization pattern
118 was also observed for all the MLH1 variants, and we did not detect any protein aggregates. We did,
119 however, observe large variations in the fluorescence intensity, and consequently the steady-state
120 protein levels, between the different MLH1 variants (Fig. 2A). To quantify these differences, we first
121 excluded the non-transfected cells using the intensity in the non-transfected control. Then we
122 measured the total intensity of the MLH1 fluorescence in each cell and normalized to the intensity
123 for wild-type MLH1. This analysis revealed up to 12-fold difference in intensity between the variants
124 showing sizable differences in abundance.
125 To examine whether these variations in cellular abundance is correlated with thermodynamic
126 stability, we plotted the normalized values against the predicted structural stabilities (ΔΔGs). This
127 analysis indeed reveals that those MLH1 variants that were predicted to be structurally destabilized
128 (high ΔΔGs) also displayed reduced steady-state levels (Fig. 2B), indicating that the predicted
129 structural destabilization and low steady-state MLH1 levels go hand in hand. Almost all (30 out of

31) variants with steady state > 75% have a ΔΔG < 3 kcal/mol, and similarly most (22/23) with ΔΔG > 3 kcal/mol have steady state levels <75%. A destabilization of 3 kcal/mol is relatively low threshold, but consistent with previous observations of other unrelated proteins (Nielsen et al., 2017; Scheller et al., 2019; Bullock et al., 2000). Such low stability thresholds may indicate that local unfolding (discussed below) plays an important role in the recognition and degradation of pathogenic variants, and/or reflect that wild-type MLH1 is a marginally stable protein. To test this, cell lysates were incubated for 30 min. at a range of temperatures. Then the lysates were separated into soluble (supernatant) and insoluble (pellet) fractions by centrifugation, and the amount of soluble MLH1 was determined by blotting (Fig. 2 – figure supplement 1). Comparison with abundant cellular proteins stained by Ponceau S and blotting for GAPDH, revealed that wild-type MLH1 appears somewhat less thermostable than these other proteins (Fig. 2 – figure supplement 1), supporting that MLH1 may indeed be marginally stable.

Given that decreased levels of MLH1 protein could cause loss of MMR function, we also examined whether cellular abundancy correlated with pathogenicity. Of the 69 variants that we studied, 29 are classified as pathogenic or likely pathogenic in the ClinVar database (Landrum et al., 2018), whereas 12 are (likely) benign, and 28 are variants of unknown significance. We found that all (likely) benign variants appeared stable and had steady-state levels >70% (Fig. 2B). Conversely, 18 out of the 29 pathogenic variants (62%) had steady-state levels < 70% (Fig. 2B), suggesting that protein destabilization is a common feature for more than half of the MLH1 variants linked to LS, and that predictions of stability might be useful for classifying such variants (see below).

Next, we analyzed how the measured steady-state levels and the stability predictions correlated with previously published *in vivo* functional data on MLH1 (Takahashi et al., 2007). In that study, MLH1 function was tested in a number of assays and ranked from 0 (no function) to 3 (full function) based on their dominant mutator effect (DME) when human MLH1 variants are expressed in yeast cells

(Shimodaira et al., 1998). Our comparison revealed that variants with reduced steady-state levels and high risk of destabilization in general are less likely to be functional (Fig. 2CD), which again indicates that the reduced structural stability may be linked to the observed loss-of-function phenotype. For example, while 22/23 variants with DME=3 have steady-state levels >70%, only five of the 23 variants with DME=0 have this high amount of protein. These functional differences are also reflected in the correlation between loss of stability (ΔΔG) and function (Fig. 2D). In particular none of the fully functional proteins (DME=3) are predicted to be destabilized by more than 3 kcal/mol, whereas 18/23 variants with DME=0 are predicted to be destabilized by at least this amount. The unstable and non-functional variants do not appear structurally clustered to a particular site within the protein, and are found throughout both the N- and C-terminal domains of MLH1 (Fig. 2E). These affected positions are, however, closer to one another than random pairs in the respective domain (average pairwise distances of 17.3 Å vs. 24.1 Å in the N-terminal domain, and 14.7 Å vs. 21.0 Å in the C-terminal domain). In contrast, the linker region is depleted in detrimental variants but not in benign variants, hence functional (Takahashi et al., 2007) and benign (Landrum et al., 2018) variants are found both in structured and unstructured regions (Fig. 2, figure supplement 2).

Proteasomal degradation causes reduced steady-state levels of destabilized MLH1 variants

Next, we analyzed why the steady-state levels of certain MLH1 variants were reduced. For this purpose, we carefully selected eight of the 69 missense MLH1 variants for further in-depth analyses (E23D, G67R, R100P, T117M, I219V, R265C, K618A, and R659P). As previously, these variants were chosen so the mutations were distributed across the *MLH1* gene, and to represent a broad range of predicted structural stabilities (ΔΔGs) as well as different pathogenicity annotations from the ClinVar database (Table 1).

177 The variants were transiently transfected into HCT116 cells. Indeed, six of the variants (G67R,
178 R100P, T117M, R265C, K618A, R659P) displayed reduced steady-state levels, while wild type-like
179 levels were observed for two variants (E23D, I219V), in agreement with the fluorescence-based
180 observations (Fig. 3AB). Co-transfection with a GFP-expression vector revealed that this was not
181 caused by differences between transfection efficiencies since the amount of GFP was unchanged (Fig.
182 3A).
183 Next, in order to investigate if the reduced MLH1 levels were caused by degradation, we monitored
184 the amounts of MLH1 over time in cultures treated with the translation inhibitor cycloheximide
185 (CHX). This revealed that those variants with reduced steady-state levels were indeed rapidly
186 degraded (half-life between 3 and 12 hours), whereas wild-type MLH1 and the other variants
187 appeared stable (estimated half-life >> 12 hours) (Fig. 3CD). Treating the cells with the proteasome-
188 inhibitor bortezomib (BZ) significantly increased the steady-state levels of the unstable variants,
189 whereas the levels of the wild-type and stable MLH1 variants were unaffected (Fig. 3E). Separating
190 cell lysates into soluble and insoluble fractions by centrifugation revealed that the destabilized MLH1
191 variants appeared more insoluble than the stable MLH1 variants, and bortezomib treatment mainly
192 caused an increase in the amount of insoluble MLH1 (Fig. 3, figure supplement). Based on these
193 results, we conclude that certain missense MLH1 variants are structurally destabilized, which in turn
194 leads to proteasomal degradation and reduced steady-state protein levels, and a loss-of-function
195 phenotype as scored by the DME.
196
197 **PMS1 and PMS2 are destabilized when MLH1 is degraded**
198 In order to carry out its role in MMR, it is essential that MLH1 associates with PMS2 to form the
199 active MutLα complex (Li & Modrich, 1995; Raschle et al., 2002; Tomer et al., 2002). Additionally,

MLH1 can bind the PMS1 protein and form the MutLβ complex, the function of which remains unknown (Cannavo et al., 2007; Kondo et al., 2001; Wu et al., 2003).

As typical for "orphan" proteins lacking their binding partners (Yanagitani et al., 2017; McShane et al., 2016), PMS2 has been found to be unstable in the absence of MLH1 (Hinrichsen et al., 2017; Lynch et al., 2015; Mohd et al., 2006; Perera & Bapat, 2008). To test the mechanism underlying this instability, we measured the stability of endogenous PMS1 and PMS2 in HCT116 cells with or without introducing wild-type MLH1. In cells treated with cycloheximide, the absence of MLH1 led to rapid degradation of both PMS1 and PMS2 ($t_{½}$ ~ 3-5 hours). However, when wild-type MLH1 was present, PMS1 and PMS2 were dramatically stabilized ($t_{½}$ ~ 12 hours) (Fig. 4AB). Treating untransfected HCT116 cells with bortezomib led to an increase in the amount of endogenous PMS1 and PMS2, showing that their degradation is proteasome-dependent (Fig. 4C). The stabilizing effect of MLH1 on PMS1 and PMS2 was also observed for the stable MLH1 variants (Fig. 4DE). Accordingly, we found that the MLH1 levels correlated with the PMS1 and PMS2 levels (Fig. 4F). Collectively, these results suggest that either there is not enough MLH1 variant in the cells to form complexes with PMS1/2 or that only stable MLH1 variants are able to bind PMS1 and PMS2, and that this binding in turn protects PMS1 and PMS2 from proteasomal degradation. To test these possibilities, we proceeded to assess the PMS2-binding activity of the selected MLH1 variants. To this end, HCT116 cells were co-transfected with both MLH1 and YFP-tagged PMS2. Importantly, the overexpressed YFP-PMS2 protein did not affect the MLH1 level and appeared stable in the absence of MLH1 (Fig. 4G), allowing us to directly compare the PMS2-binding activity of the selected MLH1 variants. To ensure that the cells contained sufficient levels of the unstable MLH1 variants, the cells were treated with bortezomib prior to lysis. We found that the wild-type and stable MLH1 variants (E23D, I219V) were efficiently co-precipitated with the YFP-tagged PMS2 (Fig. 4H). Several of the unstable MLH1 variants did not display appreciable affinity for PMS2, even after

blocking their degradation, suggesting that these MLH1 variants are structurally perturbed or unfolded to an extent that disables complex formation with PMS2. Interestingly, the K618A variant displayed a strong interaction with PMS2 (Fig. 4H), indicating that this unstable variant retains the ability to bind PMS2, and therefore potentially engage in mismatch repair. We note that this result is supported by the K618A variant's ability to stabilize PMS2 (Fig. 4E) and the distal positioning of K618 to the PMS2 binding site (Gueneau et al., 2013).

HSP70 is required for degradation of some destabilized MLH1 variants

Since structurally destabilized proteins are prone to expose hydrophobic regions that are normally buried in the native protein conformation, molecular chaperones, including the prominent HSP70 and HSP90 enzymes, often engage such proteins in an attempt to refold them or to target them for proteasomal degradation (Arndt et al., 2007). Indeed, both HSP70 and HSP90 are known to interact with many missense variants though with different specificities and cellular consequences (Karras et al., 2017), and a previous study has linked HSP90 to MLH1 function (Fedier et al., 2005).

To test the involvement of molecular chaperones in degradation of the selected MLH1 variants, we analyzed their interaction with HSP70 and HSP90 by co-immunoprecipitation and Western blotting. Similar to above, the cells were treated with bortezomib to ensure detectable amounts of MLH1. Interestingly, four of the destabilized MLH1 variants (G67R, R100P, T117M and R265C) displayed a strong interaction with HSP70, up to approximately 7-fold greater compared to wild-type MLH1 (Fig. 5AB). Conversely, in the case of HSP90 we observed binding to all the tested MLH1 variants, including the wild-type (Fig. 5CD), suggesting that HSP90 may be involved in the *de novo* folding of wild-type MLH1 or assembly of MLH1-containing MMR complexes, while HSP70 may be involved in regulation of certain destabilized MLH1 variants, potentially playing a role in their degradation.

To test this hypothesis, we measured the steady-state levels of the MLH1 variants following inhibition of HSP70 and HSP90, respectively. We treated cells with the HSP70 inhibitor YM01 or the HSP90 inhibitor geldanamycin (GA) and compared with the MLH1 levels in untreated cells. In comparison to the HSP70 binding, the effect of YM01 appeared more subtle. The levels of three variants (G67R, R100P, T117M) were, however, increased (Fig. 5EF), and all three were also found to bind HSP70 (Fig. 5A) and had the lowest steady-state levels of the eight tested variants prior to HSP70 inhibition. Together, these results suggest that HSP70 actively partakes in detecting and/or directing certain destabilized MLH1 variants for degradation. We did not observe any effect of HSP90 inhibition on the MLH1 protein levels for any of the tested variants (Fig. 5GH).

Structural stability calculations for predicting pathogenic mutations

Our results show that unstable protein variants are likely to be rapidly degraded, suggesting that predictions of changed thermodynamic stability of missense MLH1 variants could be used to estimate whether a particular MLH1 missense variant is pathogenic or not. In comparison with the sequence-based tools (e.g. PolyPhen2, PROVEAN) that are currently employed in the clinic (Adzhubei et al., 2010; Choi & Chan, 2015), the FoldX energy predictions provide an orthogonal structure-based and sequence-conservation-independent prediction of whether a mutation is likely to be pathogenic. Unlike most variant consequence predictors, FoldX was not trained on whether mutations were benign or pathogenic, but solely on biophysical stability measurements (Guerois et al., 2002). This considerably reduces the risk of overfitting to known pathogenic variants. More importantly, because of the mechanistic link to protein stability, FoldX predictions enable insights into why a particular mutation is problematic (Kiel & Serrano, 2014; Kiel et al., 2016; Pey et al., 2007; Nielsen et al., 2017; Stein et al., 2019).

As a first test for utilizing the biophysical calculations, we analyzed the predicted protein stabilities of MLH1 variants reported in the >140.000 exomes available in the Genome Aggregation Database (gnomAD) (Lek et al., 2016; Karczewski et al., 2019). Gratifyingly, this revealed that those variants reported to occur at a high frequency in the population all displayed low ΔΔG values (Fig. 6A), suggesting that these MLH1 proteins are stable. Accordingly, with only a few exceptions, the most common MLH1 alleles reported in gnomAD also appeared functional (high DME scores) (Fig. 6A). To further test the performance of the structural stability calculations for identifying pathogenic MLH1 variants, we then compared the ΔΔG values for ClinVar-annotated MLH1 variants. This revealed that the benign MLH1 variants all appeared structurally stable, while many pathogenic variants appeared destabilized (Fig. 6B). For example, 15 of the 28 pathogenic variants (54%) for which we could calculate a stability change have ΔΔG >3 kcal/mol and 20 (71%) have ΔΔG >2 kcal/mol, whereas none of the 11 benign variants have ΔΔG >1.5 kcal/mol. Applying this calculation to all MLH1 missense variants in ClinVar revealed similar trends, with 55/95 (57%) of the pathogenic variants having ΔΔG >3 kcal/mol, while only 2/21 benign variants have ΔΔG >1.5 kcal/mol.

Overall, while many pathogenic variants are severely destabilized, a subset of these are predicted to be as stable as non-pathogenic variants (e.g. 7/28 have predicted ΔΔG <1.0 kcal/mol). This observation could be explained e.g. by inaccuracies of our stability calculations or by loss of function via other mechanisms such as direct loss of enzymatic activity, post-translational modifications or protein-protein interactions (Wagih et al., 2018). Thus, as a separate method for predicting the biological consequences of mutations, we explored if sequence analysis of the MLH1 protein family across evolution would reveal differences in selective pressure between benign and pathogenic variants. We performed an analysis of a multiple sequence alignment of MLH1 homologs, which considers both conservation at individual sites, but also non-trivial, co-evolution between pairs of residues (Balakrishnan et al., 2011; Stein et al., 2019). Turning this data into a statistical model

allowed us to score all possible missense MLH1 variants. As this statistical sequence model is based on homologous sequences shaped by evolutionary pressures, it is expected to capture which residues, and pairs of residues, are tolerated (Balakrishnan et al., 2011). As opposed to stability calculations via e.g. FoldX, this approach is not directly linked to an underlying mechanistic model. Thus, we generally expect destabilizing residues to be recognized as detrimental by both FoldX and the evolutionary statistical energies, while variants in functionally active sites might only be recognized by the latter, if they do not affect protein stability (Stein et al., 2019). On the other hand, stability calculations could capture effects specific to human MLH1 that are more difficult to disentangle through the sequence analyses. In our implementation, low scores indicate mutations that during evolution appear tolerated, while high scores mark amino acid substitutions that are rare and therefore more likely to be detrimental to protein structure and/or function. Indeed, the average sequence-based score for the benign variants is lower (variations more likely to be tolerated) than the average for the ClinVar-curated pathogenic variants (Fig. 6C). The full matrix of evolutionary statistical energies is included in the supplemental material (supplemental material file 2).

Next, we compared the structure-based stability calculations and evolutionary statistical energies in a two-dimensional landscape of variant tolerance (Fig. 6D), which we find largely agrees with the functional classification by Takahashi *et al.* (Takahashi et al., 2007). There are, however, three variants with low evolutionary statistical energies (typically indicating tolerance), but predicted and experimentally-confirmed to be destabilized relative to wild-type MLH1 (T662P, I565F, G244V). Further, a number of stable variants have high DME scores (indicating wild-type-like function), but also high evolutionary statistical energies, indicating likely loss of function, and indeed several of these are classified as pathogenic in ClinVar (Fig. 6, figure supplements 1 and 2). One possible explanation for these discrepancies is a different sensitivity in the employed yeast assays (Takahashi et al., 2007), i.e., these variants may be sufficiently functional under assay conditions, but their

319 impaired function relative to wild-type MLH1 may nevertheless render them pathogenic in human
320 variant carriers.
321 To exploit the complementary nature that ΔΔG and evolutionary sequence energies display for a
322 subset of the variants, we applied logistic regression to combine these two metrics. In a jackknife test
323 (to avoid overfitting) we found that 99 of 116 (85%) ClinVar (Landrum et al., 2018) missense variants
324 were classified correctly by the regression model (Fig. 6E).
325 To compare the capability of the above-described evolutionary statistical energies, FoldX, our
326 regression model, and more traditional sequence-based methods (PolyPhen2, PROVEAN and
327 REVEL) in separating pathogenic and non-pathogenic variants, we applied these approaches to the
328 set of 116 known benign and disease-causing MLH1 variants. We then used receiver-operating
329 characteristic (ROC) analyses to compare how well the different methods are able to distinguish the
330 21 benign variants from the 95 known pathogenic variants (Fig. 6F and Fig. 6, figure supplement 3).
331 The results show that although all predictors perform fairly well, the logistic regression model
332 performs best (AUC: 0.90±0.03), and evolutionary statistical energies alone (AUC: 0.88±0.03) are
333 slightly better at distinguishing disease-linked missense variants from harmless variants than REVEL
334 (AUC: 0.83±0.03), PolyPhen2 (AUC: 0.84±0.03) and PROVEAN (AUC: 0.76±0.04). Structure-
335 based ΔΔG calculations show similar performance to these sequence-based predictors (AUC:
336 0.82±0.04), but, as the shape of the ROC curve illustrates, are particularly informative in the region
337 of high specificity.
338 Lastly, we assessed whether the underlying genomic changes could affect splicing and thus have
339 pathogenic potential, rather than directly acting on the protein level. Using SpliceAI (Jaganathan et
340 al., 2019) (REF) we predicted that 19/116 (16%) of the ClinVar variants may affect splicing. All 19
341 are pathogenic variants; no benign variants were predicted to affect splicing. Interestingly, 4 of these
342 variants are in have scores in our regression model that suggest them to be non-detrimental, and thus

343 predicted not to affect protein function or stability (Fig. 6G). Thus, more than half (4/7) of these
344 erroneously classified variants may affect splicing rather that protein function, a substantially higher
345 fraction than the 17% (15/88) that are predicted to affect splicing among those variants that are
346 correctly classified by our regression model integrating ΔΔG and evolutionary sequence energies.
347 The overall fraction predicted to affect splicing is similar to that reported in a recent genome-editing-
348 based study on BRCA1 variants (Findlay et al., 2018).
349

Discussion

Missense variants in the *MLH1* gene are a leading cause of Lynch syndrome (LS) and colorectal cancer (Peltomaki, 2016). In recent years, germline mutations that cause structural destabilization and subsequent protein misfolding have surfaced as the cause of several diseases, including cystic fibrosis (Ahner et al., 2007), phenylketonuria (Pey et al., 2007; Scheller et al., 2019), early onset Parkinson's disease (Mathiassen et al., 2015; Olzmann et al., 2004) and MSH2-linked LS (Arlow et al., 2013; Nielsen et al., 2017). Although previous studies have shown some MLH1 variants to be destabilized (Takahashi et al., 2007; Perera & Bapat, 2008), this has not been systematically addressed and the contribution of MLH1 protein stability for LS remains to be resolved.

In this study, we performed *in silico* saturation mutagenesis and stability predictions of all single-site MLH1 missense variants in the structurally-resolved regions of MLH1. Comparisons with a selected group of naturally occurring MLH1 variants revealed that those variants that are predicted to be destabilized indeed display substantially reduced steady-state protein levels. The decreased cellular amounts are caused by rapid proteasomal degradation. In turn, the loss of MLH1 causes a dramatic destabilization and proteasomal degradation of both PMS1 and PMS2. This effect suggests that the MutLα and MutLβ heterodimers are likely to be rather stable protein complexes, as the PMS1 and PMS2 proteins would otherwise be required to be stable in the absence of MLH1. These observations are in line with several previous studies on individual MLH1 variants (Cravo et al., 2002; Kosinski et al., 2010; Perera & Bapat, 2008; Raevaara et al., 2005) including a thorough analysis by Takahashi *et al*. (Takahashi et al., 2007), and also agree with tissue staining of tumor cells from patients with germline *MLH1* mutations (Hampel et al., 2008; de Jong et al., 2004).

Our observation that single amino acid changes in MLH1 are sufficient to cause degradation is similar to results from multiple other proteins including recent deep mutational scans on PTEN and TPMT (Matreyek et al., 2018), our previous results on MSH2 in human cells (Nielsen et al., 2017), and

374 earlier observations on Lynch syndrome MSH2 variants in yeast (Gammie et al., 2007; Arlow et al.,
375 2013). Our structural stability calculations predict that a relatively mild destabilization of just a few
376 (~3 kcal/mol) is sufficient to trigger MLH1 degradation, an observation in line with previous studies
377 on other proteins (Bullock et al., 2000; Nielsen et al., 2017; Suri et al., 2017; Scheller et al., 2019;
378 Jepsen et al., 2019; Caswell et al., 2019). Although the absolute thermodynamic stability of MLH1 is
379 unknown, both *in vitro* and in a cellular context, it is possible that the 3 kcal/mol destabilization
380 necessary to trigger degradation is lower than that required to reach the fully unfolded state. Although
381 we show that wild-type MLH1 is perhaps only marginally stable, we propose that these effects of
382 slightly destabilizing amino acid substitutions could be the result of local unfolding events rather than
383 a global unfolding. Accordingly, it appears that the PQC system is tightly tuned to detect increased
384 amounts of minor or transient structural defects, which is supported by observations on MSH2
385 showing that in some cases the destabilized variants are even functional when degradation is blocked
386 (Arlow et al., 2013), and results showing that the PQC system preferably targets folding intermediates
387 (Bershtein et al., 2013). As a result, it is surprising that predictions of changes to the global folding
388 stability (using the fully unfolded state as a reference) are so effective in distinguishing variants that
389 are stable in the cell from those that are more rapidly degraded. These observations also suggest that
390 improved understanding of the biophysical basis for cellular quality control might lead to even better
391 predictions of cellular abundance.
392 In our cellular studies, the MLH1 variants were expressed from a constitutive promoter, thus
393 bypassing any potential transcriptional regulation of MLH1 expression involving, for example,
394 amounts of the MLH1 protein or its function. Hence, in other experimental setups where MLH1
395 variants are generated at the endogenous locus, additional layers of control may affect the observed
396 correlation between structural stability and steady-state amounts. Another main difference between
397 the cellular context and the predicted stability is the multiple roles played by the PQC system. For

example, on one hand, chaperones may aid in folding or stabilizing proteins, but may at the same time act as sensors of misfolded proteins and help target them for degradation. Increased amounts of unfolded or misfolded proteins may affect and titrate PQC components, thus complicating the relationship between protein stability and abundance. Nevertheless, since the predicted protein stabilities can, to a large extent, discriminate between pathogenic and benign variants, we expect that multiple disease-linked MLH1 variants will indeed display reduced cellular levels.

Similar to our previous observations for MSH2 (Nielsen et al., 2017), these results indicate that structural destabilization appears to be a common result of many disease-linked *MLH1* missense variants. Supported by earlier functional studies (Takahashi et al., 2007), we suggest that the loss-of-function phenotype in many cases can be explained by structural destabilization and subsequent degradation. Indeed, 20 out of the 31 variants with DME=0 or DME=1 have steady-state levels less than 70% of wildtype MLH1. Our data also include examples of loss-of-function variants with high steady-state levels, which is expected, as variants can affect function without modifying stability, e.g. by changing binding interfaces or active sites (Gueneau et al., 2013), and these would therefore be interesting to analyze biochemically in more detail. For example, M35R, N64S and F80V are all close to the ATP binding site in the N-terminal domain of MLH1 (Fig. 6, figure supplement 2), and might thus interfere with the catalytic activity. This would be consistent with a loss-of-function phenotype (Takahashi et al., 2007) but wild-type-like cellular protein levels (Table 1).

The correlation between the predicted structural stability with both cellular stability and protein function, suggests that the stability predictions may be used for classifying MLH1 missense variants. This is particularly relevant for LS, where according to the ClinVar database (Landrum et al., 2018) 711 out of 851 (~84%) reported MLH1 missense variants are assigned as so-called variants of uncertain significance (VUS) (Manolio et al., 2017). Of note, the fact that we did not observe the E23D VUS as destabilized does not preclude this variant from being pathogenic, since it may affect

function without being structurally perturbed. Although the variant appears functional in yeast cells (Takahashi et al., 2007), the evolutionary statistical energy of 0.9 indicates that this change is rare across the MLH1 protein family evolution and thus might be detrimental. Our results for the K618A VUS suggest that this variant, albeit being unstable, is still able to associate with PMS2 and may therefore be functional, but untimely degraded. Of the 69 MLH1 variants that we analyzed, 30 have status as VUSs. Of these, our analysis identified several (e.g. G54E, G244V, and L676R) that hold characteristics indicating that they are pathogenic: steady-state levels below 50% of WT, high ΔΔG and low functionality score *in vivo* (Takahashi et al., 2007) (Table 1).

The potential use of stability predictions for LS diagnostics is supported by the predicted MLH1 stabilities clearly separating into disease-linked and benign MLH1 variants. Moreover, since we observe that those MLH1 alleles that occur more frequently in the population are in general predicted as stable, this suggests that these common MLH1 alleles are either benign or at least only disease causing with a low penetrance. However, certainly not all the unstable variants were accurately detected by the structural predictions. For instance, out of our eight selected variants, three (R100P, R265C, K618A) appeared unstable, but were not predicted to be so (Table 1). As described above, it is important to note that the stability predictions report on the global stability of the protein, while in a cellular context it is unlikely that any of the variants are fully unfolded. Instead, it is likely that local elements unfold (Stein et al., 2019), and although refolding may occur, the locally unfolded state allows chaperones and other protein quality control components to associate and target the protein for proteasomal degradation (Fig. 7). A better understanding of the importance of local unfolding events for cellular stability is an important area for further research. Indeed, in addition to the utility for classifying potential LS variants, we believe that an important aspect of our work is that it suggests a single mechanistic origin of ~60% of the LS MLH1 variants that we have studied. This observation is also supported by our ROC analysis (Fig. 6D), which demonstrates that the stability calculations

446 can identify ~60% of the pathogenic variants at very high specificity (few false positives) exactly
447 because these are the variants that appear to cause disease via this mechanism. The ability to
448 separately analyze effects on protein stability and other effects that might be captured by the sequence
449 analysis is also an advantage of applying the predictors individually, rather than relying on combined
450 predictors such as our regression model or published ensemble-based meta-predictors. While those
451 have slightly higher overall accuracy, they do not directly indicate the underlying molecular reason
452 for pathogenicity. Incidentally, we note that by analyzing both calculations and multiplexed assays
453 of variant effects, we recently found that ~60% of disease-causing variants in the protein PTEN were
454 caused by destabilization and a resulting drop in cellular abundance (Jepsen et al., 2019). On the other
455 hand, while stability prediction is very useful for accurate identification of many pathogenic variants,
456 it may have lower overall sensitivity. We speculate that this is due to stability being a necessary, but
457 not sufficient criterion, i.e., destabilized variants are likely pathogenic, while a variant being stable
458 does not necessary imply it being functional.
459 Lastly, we predicted the effects variants with known consequences (benign or pathogenic) may have
460 on splicing, which may also lead to pathogenic changes. We found that, overall, about 16% of the
461 variants are predicted to affect splicing, though none of the 19 benign variants are in this category.
462 Interestingly, 4/7 pathogenic variants in the region that the logistic regression model predicts to be
463 benign may affect splicing (Fig. 6G), indicating that integration of effects on the genomic level are
464 likely to boost overall predictive power, and should be considered in future developments of
465 pathogenicity predictors.
466 In line with observations on other destabilized proteins, we found that the degradation of some
467 structurally destabilized MLH1 variants depends on the molecular chaperone HSP70. This suggests
468 that HSP70 recognizes the destabilized MLH1 variants and targets them for proteasomal degradation.
469 Accordingly, we observed that several destabilized MLH1 variants associate with HSP70.

Involvement of molecular chaperones in protein degradation is a well-established phenomenon (Samant et al., 2018; Arndt et al., 2007; Kandasamy & Andreasson, 2018). Moreover, our findings are consistent with recent developments in the field, showing that degradation signals, so-called degrons, are buried within the native structure of most globular proteins. Upon exposure when the protein structure is destabilized, the degrons are recognized by chaperones and other protein quality control components, which in turn guide the target protein for degradation (Enam et al., 2018; Geffen et al., 2016; Kim et al., 2013; Maurer et al., 2016; Ravid & Hochstrasser, 2008). Thus, ultimately the degradation of a protein will depend on both the structural destabilization (ΔΔG) as well as the exposed degrons, and how efficiently these are recognized by the degradation machinery. The results presented here suggest that biophysical calculations are able to predict the structural destabilization (ΔΔG), however, since the nature of protein quality control degrons is still largely undefined (Geffen et al., 2016; Maurer et al., 2016; Rosenbaum et al., 2011; van der Lee et al., 2014), *in silico* prediction of these is currently not possible.

In conclusion, our results support a model (Fig. 7) where missense mutations can cause destabilization of the MLH1 protein, leading to exposure of degrons which, in turn, trigger HSP70-assisted proteasomal degradation, causing disruption of the MMR pathway and ultimately leading to an increased cumulative lifetime risk of cancer development in LS patients. Potentially, this opens up for new therapeutic approaches, including inhibiting the PQC-mediated clearance of marginally stable MLH1 variants, or small molecule stabilizers of MLH1.

490 **Materials and Methods**

491

Key Resources Table				
Reagent type (species) or resource	**Designation**	**Source or reference**	**Identifiers**	**Additional information**
gene (*Homo sapiens*)	*MLH1*	-	UniProt identifier: P40692-1	-
cell line (*Homo sapiens*)	HCT116	ATCC	CCL-247EMT	-
antibody	anti-MLH1 (rabbit polyclonal)	Santa Cruz Biotechnology	sc-11442	Dilution: 1:100 (IF) 1:1000 (WB)
antibody	anti-β-actin (mouse monoclonal)	Sigma-Aldrich	A5441	Dilution: 1:20000
antibody	anti-PMS2 (mouse monoclonal)	BD Biosciences	556415	Dilution: 1:2500
antibody	anti-PMS1 (rabbit polyclonal)	Invitrogen	PA5-35952	Dilution: 1:2500
antibody	anti-GFP (rat monoclonal)	ChromoTek	3H9	Dilution: 1:2000
antibody	anti-myc (rat monoclonal)	ChromoTek	9E1	Dilution: 1:1000

antibody	anti-HA (rat monoclonal)	Roche	3F10	Dilution: 1:2000
antibody	anti-GAPDH (rabbit monoclonal)	Cell Signaling Technologies	14C10	Dilution: 1:2000
antibody	anti-PMCA (mouse monoclonal)	Invitrogen	MA3-914	Dilution: 1:2000
recombinant DNA reagent	pCMV6-MYC-DDK-HSP70-1A (HSPA1A)	OriGene	RC200270	-
recombinant DNA reagent	pcDNA3-HA-HSP90	Addgene	22487	-
recombinant DNA reagent	pEYFP-C2-PMS2	Prof. Lene J. Rasmussen	(Andersen et al., 2012)	-
recombinant DNA reagent	pEGFP-C1	Clontech	Discontinued by supplier	Available from NovoPro Labs (Cat. No. V12024)
recombinant DNA reagent	pcDNA3.1-V5-His	Invitrogen	V81020	-
recombinant DNA reagent	pCMV-MLH1 & MLH1 variants	Prof. Chikashi Ishioka	(Takahashi et al., 2007)	-
commercial assay or kit	FuGENE HD	Promega	E2311	-

chemical compound, drug	Bortezomib	LC Laboratories	B-1408	-
chemical compound, drug	YM01	StressMarq	SIH-121	-
chemical compound, drug	Geldanamycin	Sigma-Aldrich	G3381	-
chemical compound, drug	Cycloheximide	Sigma-Aldrich	C1988	-
software, algorithm	UnScanIt gel	Silk Scientific	V6.1	-
software, algorithm	FoldX	http://foldxsuite.crg.eu/	January 2017	Details see Methods
software, algorithm	Gremlin	https://github.com/sokrypton/GREMLIN	V2.01	Details see Methods
software, algorithm	Custom R script	-	-	-
software, algorithm	SpliceAI	https://github.com/Illumina/SpliceAI	V1.2.1	-

Plasmids

Plasmids for expression of wild-type and mutant MLH1 variants have been described before (Takahashi et al., 2007). pCMV-MYC-DDK-HSP70 and pcDNA3-HA-HSP90 were kindly provided by Dr. Kenneth Thirstrup (H. Lundbeck A/S). The pEYFP-C2-PMS2 plasmid was kindly provided by Prof. Lene J. Rasmussen (University of Copenhagen). pEGFP was purchased from Clontech. A pcDNA3-V5 vector served as a negative control.

499

500 *Cell culture*

501 HCT116 cells (kindly provided by Prof. Mads Gyrd-Hansen (University of Oxford)) were maintained
502 in McCoy's 5A medium (Gibco) supplemented with 10% fetal calf serum (Invitrogen), 2 mM
503 glutamine, 5000 IU/ml penicillin and 5 mg/ml streptomycin at 37 °C in a humidified atmosphere with
504 5% CO_2. Cell line authentication by STR analysis was performed by Eurofins. The cells were not
505 contaminated by mycoplasma.
506 Transfections were performed using FugeneHD (Promega) as described by the manufacturer in
507 reduced serum medium OptiMEM (Gibco). Cells were harvested no later than 72 hours after
508 transfection at a final confluence around 90%.About 24 hours after transfection, cells were treated
509 with serum-free growth medium containing 25 µg/mL cycloheximide (Sigma) for a duration of 4, 8
510 or 12 hours, 10 µM bortezomib (LC laboratories) for 8 or 16 hours, 5 µM YM01 (StressMarq) for 24
511 hours or 1 µM geldanamycin (Sigma) for 24 hours. Cells were lysed in SDS sample buffer (94 mM
512 Tris/HCl pH 6.8, 3% SDS, 19% glycerol and 0.75% β-mercaptoethanol) and protein levels were
513 analyzed by SDS-PAGE and Western blotting.

514

515 *SDS-PAGE and Western blotting*

516 Proteins were resolved by SDS-PAGE on 7x8 cm 12.5% acrylamide gels, and transferred to 0.2 µm
517 nitrocellulose membranes (Advantec, Toyo Roshi Kaisha Ltd.). Blocking was performed using PBS
518 (8 g/L NaCl, 0.2 g/L KCl, 1.44 g/L Na_2HPO_4, 0.24 g/L KH_2PO_4, pH 7.4) with 5% dry milk powder
519 and 0.05% Tween-20. Membranes were probed with primary antibodies (see Key Resources Table)
520 at 4 °C overnight. HRP-conjugated secondary antibodies were purchased from DAKO. ECL detection
521 reagent (GE Life Sciences) was used for development.

522

Immunofluorescence and imaging

Transfected cells were seeded 24 hours prior to fixing with 4 % formaldehyde in PBS in thin-bottomed 384-well plates. The fixed cells were then washed three times in PBS and permeabilized with 0.25 % Triton-X-100 in PBS for 5 minutes at room temperature (RT). After washing with PBS, 5 % bovine serum albumin (BSA, Sigma) in PBS was used for 45 minutes at RT for blocking. The cells were then washed with PBS and incubated with a 1:100 dilution of the anti-MLH1 antibody (Santa Cruz Biotechnology, Product no.: sc-11442) in 1 % BSA in PBS for 1 hour at RT. The cells were washed with PBS and incubated with a 1:1000 dilution Alexa Fluor 568 anti-rabbit antibody (Invitrogen) in 1 % BSA in PBS for 1 hour at RT. After additional washing with PBS, the DNA was stained with Höchst 33342 (Sigma) for 10 minutes. Microscopy was performed using an InCell2200 microscope (GE Healthcare). The filters were Höchst (ex 390 nm, em 432 nm) and TexasRed (ex 575 nm, em 620 nm). The InCell Developer Toolbox (GE Healthcare) was used for image analysis. To determine the abundance of the MLH1 variants, the total intensity of the red channel in each cell was measured after excluding the non-transfected cells.

Co-immunoprecipitation

Transfected cells were lysed in buffer A (50 mM Tris/HCl pH 7.5, 150 mM NaCl, 1 mM EDTA and 0.5% NP-40 supplemented with Complete Mini EDTA-free Protease inhibitor cocktail tablets (Roche)) and left to incubate for 20 minutes on ice. The lysates were cleared by centrifugation (13000 g, 30 min) and the proteins were captured with GFP-trap (Chromotek), Myc-trap (Chromotek) or HA-agarose beads (Sigma) by tumbling end-over-end overnight at 4℃. The beads were washed three times by centrifugation (1000 g, 10 sec) in buffer A. Finally, the beads were resuspended in SDS sample buffer (94 mM Tris/HCl pH 6.8, 3% SDS, 19% glycerol and 0.75% β-mercaptoethanol) and analyzed by SDS-PAGE and Western blotting.

Solubility assays

For analyses of wild-type MLH1, HCT116 cells were transfected as described above. Cells were harvested in buffer B (50 mM Tris/HCl pH 7.5, 150 mM NaCl, 1 mM EDTA) supplemented with complete protease inhibitor cocktail tablets (Roche), and lysed by sonication (3 x 10 sec). Then the lysate was distributed into different tubes that were incubated at 30 min. at the indicated temperatures. The soluble and insoluble fractions were separated by centrifugation (15000 g, 4 ºC, 30 min), after which the supernatant was removed and the pellet washed once in buffer B. Subsequently, SDS sample buffer was added, and the final sample volume was kept identical between the pellet and the supernatant. Finally, fractions were analyzed by SDS-PAGE and Western blotting as described.

For comparison of the variants, HCT116 cells were transfected and treated with bortezomib as described above. Cells were harvested in buffer B supplemented with complete protease inhibitor cocktail tablets (Roche), and lysed by sonication (3 x 10 sec). The soluble and insoluble fractions were then immediately separated by centrifugation and analyzed SDS-PAGE and Western blotting as described above.

Stability calculations

The changes in folding stability (ΔΔG) were calculated using FoldX (Guerois et al., 2002) based on PDB IDs 4P7A (Wu et al., 2015) for the N-terminal domain of MLH1, and 3RBN for the C-terminal domain. The ΔΔGs were calculated from each structure individually by first applying the RepairPDB function to fix minor issues in the original coordinates, and then the BuildModel function to generate each individual amino acid variant. Each calculation was repeated 5 times, and the average difference in stability between wild type and variant is reported, such that values <0 kcal/mol indicate stabilized variants, and values >0 kcal/mol indicate destabilized variants with respect to the wild type MLH1

571 protein. Values > 15 kcal/mol likely indicate clashes in the model FoldX generated. While this does
572 indicate that major destabilization is likely, the actual values are less meaningful for these clashing
573 variants, and we thus truncated them to 15 kcal/mol in our figures.
574
575 *Evolutionary statistical energy calculations*
576 To assess the likelihood of finding any given variant in the protein family, we created a multiple
577 sequence alignment of human MLH1 using HHblits (Zimmermann et al., 2018) and then calculated
578 a sequence log-likelihood score combining site-conservation and pairwise co-variation using Gremlin
579 (Balakrishnan et al., 2011). Scores were normalized to a range of (0,1), with low scores indicating
580 tolerated sequences and high scores indicating variants that are rare or unobserved across the multiple
581 sequence alignment. Positions at which the number of distinct homologous sequences was too small
582 to extract meaningful evolutionary statistical energies are set to NA (supplemental material file 2).
583 Other sequence-based predictions of functional variant consequences were retrieved from the
584 webservers of PROVEAN (Choi & Chan, 2015) and PolyPhen2 (Adzhubei et al., 2010), and extracted
585 from the pre-calculated scores for REVEL (Ioannidis et al., 2016).
586
587 *Dominant mutator effect (DME)*
588 We grouped the functional classification observations from Takahashi *et al.* (Takahashi et al., 2007)
589 into 4 categories by summarizing the number of assays each variant showed functional behavior in.
590 Thus, variants in group 0 were non-functional in all three assays, those in group 3 were functional in
591 all three assays, and the rest were functional in some, but not in other assays.
592

Acknowledgements

The authors thank Prof. Lene J. Rasmussen, Prof. Mads Gyrd-Hansen, and Dr. Kenneth Thirstrup for sharing cells and reagents, and Dr. Elin J. Pietras, Dr. Cornelia Steinhauer, Cecilie Søltoft, Anne-Marie Lauridsen, and the Core Facility for Integrated Microscopy for excellent technical assistance.

Competing interests

No competing interests declared.

Author contributions

A.B.A., A.S., S.V.N., K.S.K., E.P., and A.Sh. conducted the experiments. A.B.A., A.S., S.V.N., M.T., C.I., K.L.-L. and R.H.-P. analyzed the data. A.B.A., A.S., K.L.-L. and R.H.-P. designed the experiments. E.R.H., I.B., A.-M.G., K.L.-L. and R.H.-P. conceived the study. M.T. and C.I. contributed reagents. A.B.A, A.S., K.L.-L. and R.H.-P. wrote the paper.

Funding

This work was supported by grants from the Novo Nordisk Foundation Challenge Program PRISM (to A.S., K.L.-L., and R.H.-P.), the Lundbeck Foundation (to A.S., K.L.-L., and R.H.-P.), the Danish Cancer Society (to R.H.-P.), the A.P. Møller Foundation (to R.H.-P.), Novo Nordisk Foundation Young Investigator Award (NNF15OC0016662; to E.R.H.) and Hallas-Møller programme (to K.L.-L.) and the Danish Council for Independent Research (to R.H.-P).

Table 1

Characteristics of the selected naturally occurring MLH1 variants

Variant*	Steady-state level (% of WT)	FoldX ΔΔG (kcal/mol)	ClinVar annotation¤	DME#
E23D	**90.0**	**0.49**	**VUS**	**3**
I25T	51.3	2.40	VUS	3
A29S	109.6	2.06	(likely) pathogenic	3
M35R	72.1	3.52	(likely) pathogenic	0
I36S	63.2	4.08	VUS	NA
N38D	63.5	1.61	VUS	2
S44F	9.8	>15	(likely) pathogenic	0
S44A	103.2	-1.35	VUS	3
G54E	15.3	>15	VUS	1
N64S	96.3	2.16	VUS	1
G67R	**35.0**	**>15**	**(likely) pathogenic**	**0**
G67W	14.5	>15	(likely) pathogenic	0
I68N	62.5	2.22	(likely) pathogenic	0
R69K	104.9	-0.18	VUS	3
C77Y	61.0	6.57	(likely) pathogenic	2
F80V	72.3	2.22	(likely) pathogenic	1
T82I	100.0	0.54	(likely) pathogenic	2
R100P	**46.4**	**-1.25**	**(likely) pathogenic**	**2**
E102D	97.9	0.34	(likely) pathogenic	3
A111V	68.8	4.96	(likely) pathogenic	0
T117M	**32.8**	**7.14**	**(likely) pathogenic**	**0**
T117R	46.2	12.70	(likely) pathogenic	0
A128P	62.8	2.40	(likely) pathogenic	0
D132H	110.2	-0.30	(likely) benign	3
A160V	107.3	0.38	VUS	3
R182G	87.9	2.60	(likely) pathogenic	3
S193P	100.7	2.73	VUS	0
V213M	103.9	-0.81	(likely) benign	3
R217C	74.1	1.06	VUS	2
I219V	**112.5**	**0.66**	**(likely) benign**	**3**
I219L	121.9	-0.05	(likely) benign	3
R226L	63.3	0.27	(likely) pathogenic	1
G244V	32.0	>15	VUS	0
G244D	38.8	>15	(likely) pathogenic	0
H264R	117.6	-0.60	VUS	3
R265C	**57.2**	**0.28**	**(likely) pathogenic**	**2**
R265H	81.4	0.04	VUS	3
E268G	81.1	0.81	(likely) benign	2
L272V	80.0	1.95	VUS	3
A281V	82.5	0.87	(likely) pathogenic	3
K286Q	101.8	0.28	VUS	2
S295G	88.6	0.13	(likely) pathogenic	2
H329P	54.1	5.67	(likely) pathogenic	1
V506A	62.1	2.18	VUS	2
Q542L	110.2	-1.56	VUS	3
L549P	63.7	5.17	VUS	0
I565F	65.9	9.64	VUS	0
L574P	34.4	11.97	(likely) pathogenic	0
E578G	103.2	0.45	(likely) benign	2
L582V	100.0	1.93	VUS	3
L588P	88.6	3.30	VUS	1
K618A	**80.4**	**0.61**	**VUS**	**1**
K618T	106.5	0.09	(likely) benign	0
L622H	61.1	4.97	(likely) pathogenic	0
P640T	61.1	3.78	VUS	0

L653R	66.1	3.22	(likely) pathogenic	0
I655V	89.1	1.03	(likely) benign	3
I655T	71.1	1.29	VUS	3
R659P	**69.0**	**6.93**	**(likely) pathogenic**	**0**
R659Q	84.9	2.41	VUS	2
T662P	72.7	5.23	(likely) pathogenic	0
E663G	72.0	-0.23	VUS	3
E663D	97.2	0.66	(likely) pathogenic	2
L676R	36.8	5.12	VUS	0
R687W	115.4	1.12	(likely) pathogenic	0
Q689R	71.0	-0.54	(likely) benign	3
V716M	100.1	1.41	(likely) benign	1
H718Y	73.7	0.16	(likely) benign	2
K751R	82.1	-0.23	(likely) benign	3

*: boldfaced variants studied in detail; ¤: VUS: variant of unknown significance; #DME: dominant mutator effect

618 **References**

619

620 Aarnio, M., Sankila, R., Pukkala, E., Salovaara, R., Aaltonen, L. A., de la Chapelle, A., Peltomaki,
621 P., Mecklin, J. P., & Jarvinen, H. J. (1999). Cancer risk in mutation carriers of DNA-
622 mismatch-repair genes. *Int.J.Cancer, 81,* 214-218.

623 Adzhubei, I. A., Schmidt, S., Peshkin, L., Ramensky, V. E., Gerasimova, A., Bork, P., Kondrashov,
624 A. S., & Sunyaev, S. R. (2010). A method and server for predicting damaging missense
625 mutations. *Nat.Methods, 7,* 248-249.

626 Ahner, A., Nakatsukasa, K., Zhang, H., Frizzell, R. A., & Brodsky, J. L. (2007). Small heat-shock
627 proteins select deltaF508-CFTR for endoplasmic reticulum-associated degradation.
628 *Mol.Biol.Cell, 18,* 806-814.

629 Allen, M., Poggiali, D., Whitaker, K., Marshall, T. R., & Kievit, R. (2018). Raincloud plots: a multi-
630 platform tool for robust data visualization. *PeerJ Preprints, 6,* e27137v1.

631 Andersen, S. D., Liberti, S. E., Lutzen, A., Drost, M., Bernstein, I., Nilbert, M., Dominguez, M.,
632 Nystrom, M., Hansen, T. V., Christoffersen, J. W., Jager, A. C., de, W. N., Nielsen, F. C.,
633 Torring, P. M., & Rasmussen, L. J. (2012). Functional characterization of MLH1 missense
634 variants identified in Lynch syndrome patients. *Hum.Mutat., 33,* 1647-1655.

635 Arlow, T., Scott, K., Wagenseller, A., & Gammie, A. (2013). Proteasome inhibition rescues clinically
636 significant unstable variants of the mismatch repair protein Msh2. *Proc.Natl.Acad.Sci.U.S.A,*
637 *110,* 246-251.

638 Arndt, V., Rogon, C., & Hohfeld, J. (2007). To be, or not to be--molecular chaperones in protein
639 degradation. *Cell Mol.Life Sci., 64,* 2525-2541.

640 Balakrishnan, S., Kamisetty, H., Carbonell, J. G., Lee, S. I., & Langmead, C. J. (2011). Learning
641 generative models for protein fold families. *Proteins, 79,* 1061-1078.

642 Barrow, E., Alduaij, W., Robinson, L., Shenton, A., Clancy, T., Lalloo, F., Hill, J., & Evans, D. G.
643 (2008). Colorectal cancer in HNPCC: cumulative lifetime incidence, survival and tumour
644 distribution. A report of 121 families with proven mutations. *Clin.Genet., 74,* 233-242.

645 Bershtein, S., Mu, W., Serohijos, A. W., Zhou, J., & Shakhnovich, E. I. (2013). Protein quality control
646 acts on folding intermediates to shape the effects of mutations on organismal fitness. *Mol.Cell,*
647 *49,* 133-144.

648 Bullock, A. N., Henckel, J., & Fersht, A. R. (2000). Quantitative analysis of residual folding and
649 DNA binding in mutant p53 core domain: definition of mutant states for rescue in cancer
650 therapy. *Oncogene, 19,* 1245-1256.

651 Cannavo, E., Gerrits, B., Marra, G., Schlapbach, R., & Jiricny, J. (2007). Characterization of the
652 interactome of the human MutL homologues MLH1, PMS1, and PMS2. *J.Biol.Chem., 282,*
653 2976-2986.

654 Casadio, R., Vassura, M., Tiwari, S., Fariselli, P., & Luigi, M. P. (2011). Correlating disease-related
655 mutations to their effect on protein stability: a large-scale analysis of the human proteome.
656 *Hum.Mutat., 32,* 1161-1170.

657 Caswell, R. C., Owens, M. M., Gunning, A. C., Ellard, S., & Wright, C. F. (2019). Using structural
658 analysis *in silico* to assess the impact of missense variants in MEN1. *bioRxiv, 661512.*

659 Choi, Y. & Chan, A. P. (2015). PROVEAN web server: a tool to predict the functional effect of amino
660 acid substitutions and indels. *Bioinformatics., 31,* 2745-2747.

661 Cravo, M., Afonso, A. J., Lage, P., Albuquerque, C., Maia, L., Lacerda, C., Fidalgo, P., Chaves, P.,
662 Cruz, C., & Nobre-Leitao, C. (2002). Pathogenicity of missense and splice site mutations in
663 hMSH2 and hMLH1 mismatch repair genes: implications for genetic testing. *Gut, 50,* 405-
664 412.

665 de Jong, A. E., van, P. M., Hendriks, Y., Tops, C., Wijnen, J., Ausems, M. G., Meijers-Heijboer, H.,
666 Wagner, A., van Os, T. A., Brocker-Vriends, A. H., Vasen, H. F., & Morreau, H. (2004).
667 Microsatellite instability, immunohistochemistry, and additional PMS2 staining in suspected
668 hereditary nonpolyposis colorectal cancer. *Clin.Cancer Res., 10,* 972-980.

669 Dowty, J. G., Win, A. K., Buchanan, D. D., Lindor, N. M., Macrae, F. A., Clendenning, M., Antill,
670 Y. C., Thibodeau, S. N., Casey, G., Gallinger, S., Marchand, L. L., Newcomb, P. A., Haile,
671 R. W., Young, G. P., James, P. A., Giles, G. G., Gunawardena, S. R., Leggett, B. A., Gattas,
672 M., Boussioutas, A., Ahnen, D. J., Baron, J. A., Parry, S., Goldblatt, J., Young, J. P., Hopper,
673 J. L., & Jenkins, M. A. (2013). Cancer risks for MLH1 and MSH2 mutation carriers.
674 *Hum.Mutat., 34,* 490-497.

675 Dunlop, M. G., Farrington, S. M., Carothers, A. D., Wyllie, A. H., Sharp, L., Burn, J., Liu, B., Kinzler,
676 K. W., & Vogelstein, B. (1997). Cancer risk associated with germline DNA mismatch repair
677 gene mutations. *Hum.Mol.Genet., 6,* 105-110.

678 Enam, C., Geffen, Y., Ravid, T., & Gardner, R. G. (2018). Protein Quality Control Degradation in
679 the Nucleus. *Annu.Rev.Biochem., 87,* 725-749.

680 Fedier, A., Stuedli, A., & Fink, D. (2005). Presence of MLH1 protein aggravates the potential of the
681 HSP90 inhibitor radicicol to sensitize tumor cells to cisplatin. *Int.J.Oncol., 27,* 1697-1705.

682 Findlay, G. M., Daza, R. M., Martin, B., Zhang, M. D., Leith, A. P., Gasperini, M., Janizek, J. D.,
683 Huang, X., Starita, L. M., & Shendure, J. (2018). Accurate classification of BRCA1 variants
684 with saturation genome editing. *Nature, 562,* 217-222.

685 Gammie, A. E., Erdeniz, N., Beaver, J., Devlin, B., Nanji, A., & Rose, M. D. (2007). Functional
686 characterization of pathogenic human MSH2 missense mutations in Saccharomyces
687 cerevisiae. *Genetics, 177,* 707-721.

688 Geffen, Y., Appleboim, A., Gardner, R. G., Friedman, N., Sadeh, R., & Ravid, T. (2016). Mapping
689 the Landscape of a Eukaryotic Degronome. *Mol.Cell, 63,* 1055-1065.

690 Gueneau, E., Dherin, C., Legrand, P., Tellier-Lebegue, C., Gilquin, B., Bonnesoeur, P., Londino, F.,
691 Quemener, C., Le Du, M. H., Marquez, J. A., Moutiez, M., Gondry, M., Boiteux, S., &
692 Charbonnier, J. B. (2013). Structure of the MutLalpha C-terminal domain reveals how Mlh1
693 contributes to Pms1 endonuclease site. *Nat.Struct.Mol.Biol., 20,* 461-468.

694 Guerois, R., Nielsen, J. E., & Serrano, L. (2002). Predicting changes in the stability of proteins and
695 protein complexes: a study of more than 1000 mutations. *J.Mol.Biol., 320,* 369-387.

696 Hampel, H., Frankel, W. L., Martin, E., Arnold, M., Khanduja, K., Kuebler, P., Clendenning, M.,
697 Sotamaa, K., Prior, T., Westman, J. A., Panescu, J., Fix, D., Lockman, J., LaJeunesse, J.,
698 Comeras, I., & de la Chapelle, A. (2008). Feasibility of screening for Lynch syndrome among
699 patients with colorectal cancer. *J.Clin.Oncol., 26,* 5783-5788.

700 Heinen, C. D. (2010). Genotype to phenotype: analyzing the effects of inherited mutations in
701 colorectal cancer families. *Mutat.Res., 693,* 32-45.

702 Hinrichsen, I., Wessbecher, I. M., Huhn, M., Passmann, S., Zeuzem, S., Plotz, G., Biondi, R. M., &
703 Brieger, A. (2017). Phosphorylation-dependent signaling controls degradation of DNA
704 mismatch repair protein PMS2. *Mol.Carcinog., 56,* 2663-2668.

705 Ioannidis, N. M., Rothstein, J. H., Pejaver, V., Middha, S., McDonnell, S. K., Baheti, S., Musolf, A.,
706 Li, Q., Holzinger, E., Karyadi, D., Cannon-Albright, L. A., Teerlink, C. C., Stanford, J. L.,
707 Isaacs, W. B., Xu, J., Cooney, K. A., Lange, E. M., Schleutker, J., Carpten, J. D., Powell, I.
708 J., Cussenot, O., Cancel-Tassin, G., Giles, G. G., MacInnis, R. J., Maier, C., Hsieh, C. L.,
709 Wiklund, F., Catalona, W. J., Foulkes, W. D., Mandal, D., Eeles, R. A., Kote-Jarai, Z.,
710 Bustamante, C. D., Schaid, D. J., Hastie, T., Ostrander, E. A., Bailey-Wilson, J. E., Radivojac,
711 P., Thibodeau, S. N., Whittemore, A. S., & Sieh, W. (2016). REVEL: An Ensemble Method
712 for Predicting the Pathogenicity of Rare Missense Variants. *Am.J.Hum.Genet., 99,* 877-885.

713 Jaganathan, K., Kyriazopoulou, P. S., McRae, J. F., Darbandi, S. F., Knowles, D., Li, Y. I., Kosmicki,
714 J. A., Arbelaez, J., Cui, W., Schwartz, G. B., Chow, E. D., Kanterakis, E., Gao, H., Kia, A.,
715 Batzoglou, S., Sanders, S. J., & Farh, K. K. (2019). Predicting Splicing from Primary
716 Sequence with Deep Learning. *Cell, 176,* 535-548.

717 Jepsen, M. M., Fowler, D. M., Hartmann-Petersen, R., Stein, A., & Lindorff-Larsen, K. (2019).
718 Classifying disease-associated variants using measures of protein activity and stability.
719 *bioRxiv, 688234.*

720 Jiricny, J. (2006). The multifaceted mismatch-repair system. *Nat.Rev.Mol.Cell Biol., 7,* 335-346.

721 Jiricny, J. (2013). Postreplicative mismatch repair. *Cold Spring Harb.Perspect.Biol., 5,* a012633.

722 Jun, S. H., Kim, T. G., & Ban, C. (2006). DNA mismatch repair system. Classical and fresh roles.
723 *FEBS J., 273,* 1609-1619.

724 Kampmeyer, C., Nielsen, S. V., Clausen, L., Stein, A., Gerdes, A. M., Lindorff-Larsen, K., &
725 Hartmann-Petersen, R. (2017). Blocking protein quality control to counter hereditary cancers.
726 *Genes Chromosomes.Cancer, 56,* 823-831.

727 Kandasamy, G. & Andreasson, C. (2018). Hsp70-Hsp110 chaperones deliver ubiquitin-dependent
728 and -independent substrates to the 26S proteasome for proteolysis in yeast. *J.Cell Sci., 131.*

729 Karczewski, K. J., Francioli, L. C., Tiao, G., Cummings, B. B., Alfoldi, J., Wang, Q., Collins, R. L.,
730 Laricchia, K. M., Ganna, A., Birnbaum, D. P., Gauthier, L. D., Brand, H., Solomonson, M.,
731 Watts, N. A., Rhodes, D., Singer-Berk, M., Seaby, E. G., Kosmicki, J. A., Walters, R. K.,
732 Tashman, K., Farjoun, Y., Banks, E., Poterba, T., Wang, A., Seed, C., Whiffin, N., Chong, J.
733 X., Samocha, K. E., Pierce-Hoffman, E., Zappala, Z., Donnell-Luria, A. H., Minikel, E. V.,
734 Weisburd, B., Lek, M., Ware, J. S., Vittal, C., Armean, I. M., Bergelson, L., Cibulskis, K.,
735 Connolly, K. M., Covarrubias, M., Donnelly, S., Ferriera, S., Gabriel, S., Gentry, J., Gupta,
736 N., Jeandet, T., Kaplan, D., Llanwarne, C., Munshi, R., Novod, S., Petrillo, N., Roazen, D.,
737 Ruano-Rubio, V., Saltzman, A., Schleicher, M., Soto, J., Tibbetts, K., Tolonen, C., Wade, G.,
738 Talkowski, M. E., The Genome Database Consortium, Neale, B. M., Daly, M. J., &
739 MacArthur, D. G. (2019). Variation across 141,456 human exomes and genomes reveals the
740 spectrum of loss-of-function intolerance across human protein-coding genes. *bioRxiv,*
741 *531210.*

742 Karras, G. I., Yi, S., Sahni, N., Fischer, M., Xie, J., Vidal, M., D'Andrea, A. D., Whitesell, L., &
743 Lindquist, S. (2017). HSP90 Shapes the Consequences of Human Genetic Variation. *Cell,*
744 *168,* 856-866.

745 Kiel, C., Benisty, H., Llorens-Rico, V., & Serrano, L. (2016). The yin-yang of kinase activation and
746 unfolding explains the peculiarity of Val600 in the activation segment of BRAF. *Elife., 5,*
747 e12814.

748 Kiel, C. & Serrano, L. (2014). Structure-energy-based predictions and network modelling of
749 RASopathy and cancer missense mutations. *Mol.Syst.Biol., 10,* 727.

750 Kim, I., Miller, C. R., Young, D. L., & Fields, S. (2013). High-throughput analysis of in vivo protein
751 stability. *Mol.Cell Proteomics., 12,* 3370-3378.

752 Kondo, E., Horii, A., & Fukushige, S. (2001). The interacting domains of three MutL heterodimers
753 in man: hMLH1 interacts with 36 homologous amino acid residues within hMLH3, hPMS1
754 and hPMS2. *Nucleic Acids Res., 29,* 1695-1702.

755 Kosinski, J., Hinrichsen, I., Bujnicki, J. M., Friedhoff, P., & Plotz, G. (2010). Identification of Lynch
756 syndrome mutations in the MLH1-PMS2 interface that disturb dimerization and mismatch
757 repair. *Hum.Mutat., 31,* 975-982.

758 Kriegenburg, F., Jakopec, V., Poulsen, E. G., Nielsen, S. V., Roguev, A., Krogan, N., Gordon, C.,
759 Fleig, U., & Hartmann-Petersen, R. (2014). A chaperone-assisted degradation pathway targets
760 kinetochore proteins to ensure genome stability. *PLoS.Genet., 10,* e1004140.

761 Landrum, M. J., Lee, J. M., Benson, M., Brown, G. R., Chao, C., Chitipiralla, S., Gu, B., Hart, J.,
762 Hoffman, D., Jang, W., Karapetyan, K., Katz, K., Liu, C., Maddipatla, Z., Malheiro, A.,
763 McDaniel, K., Ovetsky, M., Riley, G., Zhou, G., Holmes, J. B., Kattman, B. L., & Maglott,
764 D. R. (2018). ClinVar: improving access to variant interpretations and supporting evidence.
765 *Nucleic Acids Res., 46,* D1062-D1067.

766 Lek, M., Karczewski, K. J., Minikel, E. V., Samocha, K. E., Banks, E., Fennell, T., O'Donnell-Luria,
767 A. H., Ware, J. S., Hill, A. J., Cummings, B. B., Tukiainen, T., Birnbaum, D. P., Kosmicki, J.
768 A., Duncan, L. E., Estrada, K., Zhao, F., Zou, J., Pierce-Hoffman, E., Berghout, J., Cooper,
769 D. N., Deflaux, N., DePristo, M., Do, R., Flannick, J., Fromer, M., Gauthier, L., Goldstein, J.,
770 Gupta, N., Howrigan, D., Kiezun, A., Kurki, M. I., Moonshine, A. L., Natarajan, P., Orozco,
771 L., Peloso, G. M., Poplin, R., Rivas, M. A., Ruano-Rubio, V., Rose, S. A., Ruderfer, D. M.,
772 Shakir, K., Stenson, P. D., Stevens, C., Thomas, B. P., Tiao, G., Tusie-Luna, M. T., Weisburd,
773 B., Won, H. H., Yu, D., Altshuler, D. M., Ardissino, D., Boehnke, M., Danesh, J., Donnelly,
774 S., Elosua, R., Florez, J. C., Gabriel, S. B., Getz, G., Glatt, S. J., Hultman, C. M., Kathiresan,
775 S., Laakso, M., McCarroll, S., McCarthy, M. I., McGovern, D., McPherson, R., Neale, B. M.,
776 Palotie, A., Purcell, S. M., Saleheen, D., Scharf, J. M., Sklar, P., Sullivan, P. F., Tuomilehto,
777 J., Tsuang, M. T., Watkins, H. C., Wilson, J. G., Daly, M. J., & MacArthur, D. G. (2016).
778 Analysis of protein-coding genetic variation in 60,706 humans. *Nature, 536,* 285-291.

779 Li, G. M. & Modrich, P. (1995). Restoration of mismatch repair to nuclear extracts of H6 colorectal
780 tumor cells by a heterodimer of human MutL homologs. *Proc.Natl.Acad.Sci.U.S.A, 92,* 1950-
781 1954.

782 Lynch, H. T., Snyder, C. L., Shaw, T. G., Heinen, C. D., & Hitchins, M. P. (2015). Milestones of
783 Lynch syndrome: 1895-2015. *Nat.Rev.Cancer, 15,* 181-194.

784 Manolio, T. A., Fowler, D. M., Starita, L. M., Haendel, M. A., MacArthur, D. G., Biesecker, L. G.,
785 Worthey, E., Chisholm, R. L., Green, E. D., Jacob, H. J., McLeod, H. L., Roden, D.,
786 Rodriguez, L. L., Williams, M. S., Cooper, G. M., Cox, N. J., Herman, G. E., Kingsmore, S.,
787 Lo, C., Lutz, C., MacRae, C. A., Nussbaum, R. L., Ordovas, J. M., Ramos, E. M., Robinson,
788 P. N., Rubinstein, W. S., Seidman, C., Stranger, B. E., Wang, H., Westerfield, M., & Bult, C.

(2017). Bedside Back to Bench: Building Bridges between Basic and Clinical Genomic Research. *Cell, 169,* 6-12.

Mathiassen, S. G., Larsen, I. B., Poulsen, E. G., Madsen, C. T., Papaleo, E., Lindorff-Larsen, K., Kragelund, B. B., Nielsen, M. L., Kriegenburg, F., & Hartmann-Petersen, R. (2015). A Two-step Protein Quality Control Pathway for a Misfolded DJ-1 Variant in Fission Yeast. *J.Biol.Chem., 290,* 21141-21153.

Matreyek, K. A., Starita, L. M., Stephany, J. J., Martin, B., Chiasson, M. A., Gray, V. E., Kircher, M., Khechaduri, A., Dines, J. N., Hause, R. J., Bhatia, S., Evans, W. E., Relling, M. V., Yang, W., Shendure, J., & Fowler, D. M. (2018). Multiplex assessment of protein variant abundance by massively parallel sequencing. *Nat.Genet., 50,* 874-882.

Maurer, M. J., Spear, E. D., Yu, A. T., Lee, E. J., Shahzad, S., & Michaelis, S. (2016). Degradation Signals for Ubiquitin-Proteasome Dependent Cytosolic Protein Quality Control (CytoQC) in Yeast. *G3.(Bethesda.).*

McShane, E., Sin, C., Zauber, H., Wells, J. N., Donnelly, N., Wang, X., Hou, J., Chen, W., Storchova, Z., Marsh, J. A., Valleriani, A., & Selbach, M. (2016). Kinetic Analysis of Protein Stability Reveals Age-Dependent Degradation. *Cell, 167,* 803-815.

Mitchell, A. L., Attwood, T. K., Babbitt, P. C., Blum, M., Bork, P., Bridge, A., Brown, S. D., Chang, H. Y., El-Gebali, S., Fraser, M. I., Gough, J., Haft, D. R., Huang, H., Letunic, I., Lopez, R., Luciani, A., Madeira, F., Marchler-Bauer, A., Mi, H., Natale, D. A., Necci, M., Nuka, G., Orengo, C., Pandurangan, A. P., Paysan-Lafosse, T., Pesseat, S., Potter, S. C., Qureshi, M. A., Rawlings, N. D., Redaschi, N., Richardson, L. J., Rivoire, C., Salazar, G. A., Sangrador-Vegas, A., Sigrist, C. J. A., Sillitoe, I., Sutton, G. G., Thanki, N., Thomas, P. D., Tosatto, S. C. E., Yong, S. Y., & Finn, R. D. (2019). InterPro in 2019: improving coverage, classification and access to protein sequence annotations. *Nucleic Acids Res., 47,* D351-D360.

Mohd, A. B., Palama, B., Nelson, S. E., Tomer, G., Nguyen, M., Huo, X., & Buermeyer, A. B. (2006). Truncation of the C-terminus of human MLH1 blocks intracellular stabilization of PMS2 and disrupts DNA mismatch repair. *DNA Repair (Amst), 5,* 347-361.

Moller, P., Seppala, T. T., Bernstein, I., Holinski-Feder, E., Sala, P., Gareth, E. D., Lindblom, A., Macrae, F., Blanco, I., Sijmons, R. H., Jeffries, J., Vasen, H. F. A., Burn, J., Nakken, S., Hovig, E., Rodland, E. A., Tharmaratnam, K., de Vos Tot Nederveen Cappel WH, Hill, J., Wijnen, J. T., Jenkins, M. A., Green, K., Lalloo, F., Sunde, L., Mints, M., Bertario, L., Pineda, M., Navarro, M., Morak, M., Renkonen-Sinisalo, L., Valentin, M. D., Frayling, I. M., Plazzer, J. P., Pylvanainen, K., Genuardi, M., Mecklin, J. P., Moeslein, G., Sampson, J. R., & Capella, G. (2018). Cancer risk and survival in path_MMR carriers by gene and gender up to 75 years of age: a report from the Prospective Lynch Syndrome Database. *Gut, 67,* 1306-1316.

824 Nielsen, S. V., Poulsen, E. G., Rebula, C. A., & Hartmann-Petersen, R. (2014). Protein quality control
825 in the nucleus. *Biomolecules., 4,* 646-661.

826 Nielsen, S. V., Stein, A., Dinitzen, A. B., Papaleo, E., Tatham, M. H., Poulsen, E. G., Kassem, M.
827 M., Rasmussen, L. J., Lindorff-Larsen, K., & Hartmann-Petersen, R. (2017). Predicting the
828 impact of Lynch syndrome-causing missense mutations from structural calculations.
829 *PLoS.Genet., 13,* e1006739.

830 Olzmann, J. A., Brown, K., Wilkinson, K. D., Rees, H. D., Huai, Q., Ke, H., Levey, A. I., Li, L., &
831 Chin, L. S. (2004). Familial Parkinson's disease-associated L166P mutation disrupts DJ-1
832 protein folding and function. *J.Biol.Chem., 279,* 8506-8515.

833 Palomaki, G. E., McClain, M. R., Melillo, S., Hampel, H. L., & Thibodeau, S. N. (2009). EGAPP
834 supplementary evidence review: DNA testing strategies aimed at reducing morbidity and
835 mortality from Lynch syndrome. *Genet.Med., 11,* 42-65.

836 Peltomaki, P. (2016). Update on Lynch syndrome genomics. *Fam.Cancer, 15,* 385-393.

837 Peltomaki, P., Aaltonen, L. A., Sistonen, P., Pylkkanen, L., Mecklin, J. P., Jarvinen, H., Green, J. S.,
838 Jass, J. R., Weber, J. L., Leach, F. S., & . (1993). Genetic mapping of a locus predisposing to
839 human colorectal cancer. *Science, 260,* 810-812.

840 Peltomaki, P. & Vasen, H. F. (1997). Mutations predisposing to hereditary nonpolyposis colorectal
841 cancer: database and results of a collaborative study. The International Collaborative Group
842 on Hereditary Nonpolyposis Colorectal Cancer. *Gastroenterology, 113,* 1146-1158.

843 Perera, S. & Bapat, B. (2008). The MLH1 variants p.Arg265Cys and p.Lys618Ala affect protein
844 stability while p.Leu749Gln affects heterodimer formation. *Hum.Mutat., 29,* 332.

845 Pey, A. L., Stricher, F., Serrano, L., & Martinez, A. (2007). Predicted effects of missense mutations
846 on native-state stability account for phenotypic outcome in phenylketonuria, a paradigm of
847 misfolding diseases. *Am.J.Hum.Genet., 81,* 1006-1024.

848 Plaschke, J., Engel, C., Kruger, S., Holinski-Feder, E., Pagenstecher, C., Mangold, E., Moeslein, G.,
849 Schulmann, K., Gebert, J., von Knebel, D. M., Ruschoff, J., Loeffler, M., & Schackert, H. K.
850 (2004). Lower incidence of colorectal cancer and later age of disease onset in 27 families with
851 pathogenic MSH6 germline mutations compared with families with MLH1 or MSH2
852 mutations: the German Hereditary Nonpolyposis Colorectal Cancer Consortium.
853 *J.Clin.Oncol., 22,* 4486-4494.

854 Raevaara, T. E., Korhonen, M. K., Lohi, H., Hampel, H., Lynch, E., Lonnqvist, K. E., Holinski-Feder,
855 E., Sutter, C., McKinnon, W., Duraisamy, S., Gerdes, A. M., Peltomaki, P., Kohonen-Ccorish,
856 M., Mangold, E., Macrae, F., Greenblatt, M., de la Chapelle, A., & Nystrom, M. (2005).

Functional significance and clinical phenotype of nontruncating mismatch repair variants of MLH1. *Gastroenterology, 129,* 537-549.

Raschle, M., Dufner, P., Marra, G., & Jiricny, J. (2002). Mutations within the hMLH1 and hPMS2 subunits of the human MutLalpha mismatch repair factor affect its ATPase activity, but not its ability to interact with hMutSalpha. *J.Biol.Chem., 277,* 21810-21820.

Ravid, T. & Hochstrasser, M. (2008). Diversity of degradation signals in the ubiquitin-proteasome system. *Nat.Rev.Mol.Cell Biol., 9,* 679-690.

Rosenbaum, J. C., Fredrickson, E. K., Oeser, M. L., Garrett-Engele, C. M., Locke, M. N., Richardson, L. A., Nelson, Z. W., Hetrick, E. D., Milac, T. I., Gottschling, D. E., & Gardner, R. G. (2011). Disorder targets misorder in nuclear quality control degradation: a disordered ubiquitin ligase directly recognizes its misfolded substrates. *Mol.Cell, 41,* 93-106.

Sachadyn, P. (2010). Conservation and diversity of MutS proteins. *Mutat.Res., 694,* 20-30.

Samant, R. S., Livingston, C. M., Sontag, E. M., & Frydman, J. (2018). Distinct proteostasis circuits cooperate in nuclear and cytoplasmic protein quality control. *Nature, 563,* 407-411.

Scheller, R., Stein, A., Nielsen, S. V., Marin, F. I., Gerdes, A. M., Di, M. M., Papaleo, E., Lindorff-Larsen, K., & Hartmann-Petersen, R. (2019). Toward mechanistic models for genotype-phenotype correlations in phenylketonuria using protein stability calculations. *Hum.Mutat., 40,* 444-457.

Schymkowitz, J., Borg, J., Stricher, F., Nys, R., Rousseau, F., & Serrano, L. (2005). The FoldX web server: an online force field. *Nucleic Acids Res., 33,* W382-W388.

Shimodaira, H., Filosi, N., Shibata, H., Suzuki, T., Radice, P., Kanamaru, R., Friend, S. H., Kolodner, R. D., & Ishioka, C. (1998). Functional analysis of human MLH1 mutations in Saccharomyces cerevisiae. *Nat.Genet., 19,* 384-389.

Sijmons, R. H. & Hofstra, R. M. (2016). Review: Clinical aspects of hereditary DNA Mismatch repair gene mutations. *DNA Repair (Amst), 38,* 155-162.

Stein, A., Fowler, D. M., Hartmann-Petersen, R., & Lindorff-Larsen, K. (2019). Biophysical and Mechanistic Models for Disease-Causing Protein Variants. *Trends Biochem.Sci.*.

Suri, M., Evers, J. M. G., Laskowski, R. A., O'Brien, S., Baker, K., Clayton-Smith, J., Dabir, T., Josifova, D., Joss, S., Kerr, B., Kraus, A., McEntagart, M., Morton, J., Smith, A., Splitt, M., Thornton, J. M., & Wright, C. F. (2017). Protein structure and phenotypic analysis of pathogenic and population missense variants in STXBP1. *Mol.Genet.Genomic.Med., 5,* 495-507.

889 Takahashi, M., Shimodaira, H., Andreutti-Zaugg, C., Iggo, R., Kolodner, R. D., & Ishioka, C. (2007).
890 Functional analysis of human MLH1 variants using yeast and in vitro mismatch repair assays.
891 *Cancer Res., 67,* 4595-4604.

892 Thompson, B. A., Spurdle, A. B., Plazzer, J. P., Greenblatt, M. S., Akagi, K., Al-Mulla, F., Bapat,
893 B., Bernstein, I., Capella, G., den Dunnen, J. T., du, S. D., Fabre, A., Farrell, M. P., Farrington,
894 S. M., Frayling, I. M., Frebourg, T., Goldgar, D. E., Heinen, C. D., Holinski-Feder, E.,
895 Kohonen-Corish, M., Robinson, K. L., Leung, S. Y., Martins, A., Moller, P., Morak, M.,
896 Nystrom, M., Peltomaki, P., Pineda, M., Qi, M., Ramesar, R., Rasmussen, L. J., Royer-
897 Pokora, B., Scott, R. J., Sijmons, R., Tavtigian, S. V., Tops, C. M., Weber, T., Wijnen, J.,
898 Woods, M. O., Macrae, F., & Genuardi, M. (2014). Application of a 5-tiered scheme for
899 standardized classification of 2,360 unique mismatch repair gene variants in the InSiGHT
900 locus-specific database. *Nat.Genet., 46,* 107-115.

901 Tokuriki, N. & Tawfik, D. S. (2009). Stability effects of mutations and protein evolvability.
902 *Curr.Opin.Struct.Biol., 19,* 596-604.

903 Tomer, G., Buermeyer, A. B., Nguyen, M. M., & Liskay, R. M. (2002). Contribution of human mlh1
904 and pms2 ATPase activities to DNA mismatch repair. *J.Biol.Chem., 277,* 21801-21809.

905 van der Lee, R., Lang, B., Kruse, K., Gsponer, J., Sanchez de, G. N., Huynen, M. A., Matouschek,
906 A., Fuxreiter, M., & Babu, M. M. (2014). Intrinsically disordered segments affect protein half-
907 life in the cell and during evolution. *Cell Rep., 8,* 1832-1844.

908 Vasen, H. F. & de Vos Tot Nederveen Cappel WH (2013). A hundred years of Lynch syndrome
909 research (1913-2013). *Fam.Cancer, 12,* 141-142.

910 Vasen, H. F., Wijnen, J. T., Menko, F. H., Kleibeuker, J. H., Taal, B. G., Griffioen, G., Nagengast, F.
911 M., Meijers-Heijboer, E. H., Bertario, L., Varesco, L., Bisgaard, M. L., Mohr, J., Fodde, R.,
912 & Khan, P. M. (1996). Cancer risk in families with hereditary nonpolyposis colorectal cancer
913 diagnosed by mutation analysis. *Gastroenterology, 110,* 1020-1027.

914 Wagih, O., Galardini, M., Busby, B. P., Memon, D., Typas, A., & Beltrao, P. (2018). A resource of
915 variant effect predictions of single nucleotide variants in model organisms. *Mol.Syst.Biol., 14,*
916 e8430.

917 Wu, H., Zeng, H., Lam, R., Tempel, W., Kerr, I. D., & Min, J. (2015). Structure of the human MLH1
918 N-terminus: implications for predisposition to Lynch syndrome. *Acta*
919 *Crystallogr.F.Struct.Biol.Commun., 71,* 981-985.

920 Wu, X., Platt, J. L., & Cascalho, M. (2003). Dimerization of MLH1 and PMS2 limits nuclear
921 localization of MutLalpha. *Mol.Cell Biol., 23,* 3320-3328.

922 Yanagitani, K., Juszkiewicz, S., & Hegde, R. S. (2017). UBE2O is a quality control factor for orphans
923 of multiprotein complexes. *Science, 357,* 472-475.

924 Zimmermann, L., Stephens, A., Nam, S. Z., Rau, D., Kubler, J., Lozajic, M., Gabler, F., Soding, J.,
925 Lupas, A. N., & Alva, V. (2018). A Completely Reimplemented MPI Bioinformatics Toolkit
926 with a New HHpred Server at its Core. *J.Mol.Biol., 430,* 2237-2243.
927
928

Figure legends

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 coloured spheres, indicating the predicted ΔΔG (<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 (ΔΔG) between a missense variant (red) and wild-type MLH1 (green). (D) Excerpt of the *in silico* saturation mutagenesis map (full dataset provided in supplemental material file 1). (E) Distribution of all predicted ΔΔGs from saturation mutagenesis. The peak at 15 kcal/mol contains all variants with ΔΔG values greater than this value.

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 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 ΔΔG 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.* (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) Coloured surface, smoothed density estimate. Gray dots represent means within each DME category, with bars for standard error. (D) Distribution of FoldX ΔΔGs across DME categories (as in (C)). (E) FoldX ΔΔGs 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. See also Figure 2 - figure supplements 1 and 2.

Figure 3. Many MLH1 variants are degraded by the proteasome. (A) HCT116 cells transfected with the indicated MLH1 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-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 µg/mL cycloheximide (CHX) for 0, 4, 8 or 12 hours, 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 hours. 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

970 treated for 16 hours with 10 μM bortezomib (BZ). Blotting for β-actin was included as a loading
971 control. See also Figure 3 – figure supplement.
972
973 **Figure 4. Stable MLH1 variants increase steady-state levels of PMS1 and PMS2.** (A) The levels
974 of endogenous PMS1 and PMS2 were determined by blotting of whole-cell lysates of HCT116 cells
975 transfected with either empty vector or with wild-type MLH1 and treated with 25 μg/mL
976 cycloheximide (CHX) for 0, 4, 8 or 12 hours. The antibodies used were to PMS1 and PMS2, and as
977 a control to MLH1. β-actin served as loading control. (B) Quantification of blots as in panel (A)
978 normalized to protein levels at 0 hours. The error bars indicate the standard deviation (n=3). (C) The
979 levels of endogenous MLH1, PMS1 and PMS2 were compared by blotting of cell lysates of HCT116
980 cells either untreated, or treated with cycloheximide (CHX) or with bortezomib (BZ) and CHX. β-
981 actin served as loading control. (D) The levels of endogenous PMS1 and PMS2 and transfected
982 MLH1 were compared by Western blotting using antibodies to PMS1, PMS2 and MLH1. β-actin
983 served as loading control. (E) Quantification of blots as in panel (C) normalized to the level of
984 endogenous PMS1 (grey) or PMS2 (red) in untransfected HCT116 cells. The error bars show the
985 standard deviation (n=3). (F) Plotting the levels of the MLH1 variants vs. the levels of endogenous
986 PMS1 (grey) and PMS2 (red). The error bars show the standard deviation (n=3). (G) The levels of
987 MLH1 and YFP-tagged PMS2 were analyzed by SDS-PAGE and blotting of whole-cell lysates of
988 HCT116 cells transfected with the indicated expression plasmids. β-actin was included as loading
989 control. (H) Co-transfected PMS2-YFP was immunoprecipitated (IP) using GFP-trap beads, and the
990 precipitated material was analyzed by electrophoresis and blotting. Bortezomib was added to all
991 cultures 16 hours prior to cell lysis to ensure ample amounts of the unstable MLH1 variants.
992
993 **Figure 5. Molecular chaperones play a role in the proteasomal degradation of MLH1.** (A) Co-
994 transfected HSP70-myc was immunoprecipitated (IP) using myc-trap beads and analyzed by blotting
995 with antibodies to the myc-tag (HSP70) and MLH1. Bortezomib was added to all cultures 8 hours
996 prior to cell lysis to ensure ample amounts of the unstable MLH1 variants. (B) Quantification of blots
997 as shown in panel (A) normalized to level of wild-type MLH1. The error bars indicate the standard
998 deviation (n=3). (C) Co-transfected HSP90-HA was immunoprecipitated (IP) with anti-HA resin, and
999 the precipitated material analyzed by electrophoresis and Western blotting using antibodies to the
1000 HA-tag (HSP90) and MLH1. As above, bortezomib was added to all cultures prior to cell lysis. (D)
1001 Quantification of blots as in panel (C) normalized to amount of precipitated wild-type MLH1. The
1002 error bars show the standard deviation (n=3). (E) Western blotting using antibodies to MLH1 of
1003 whole-cell lysates from transfected cells treated with 5 μM YM01 for 24 hours as indicated. (F)
1004 Quantification of blots as shown in panel (E) normalized to level of MLH1 without YM01. The error
1005 bars indicate the standard deviation (n=3). (G) Western blotting using antibodies to MLH1 of whole-
1006 cell lysates from transfected cells treated 1 μM geldanamycin (GA) for 24 hours. (H) Quantification
1007 of blots as shown in panel (G) normalized to level of MLH1 without GA. The error bars indicate the
1008 standard deviation (n=3).
1009
1010 **Figure 6. Assessing stability calculations for predicting pathogenicity.** (A) Plot of ΔΔG-values
1011 vs. allele frequencies for all variants listed in gnomAD (gray), as well as those analyzed by Takahashi

et al.; the latter are color-coded by DME. Note that the leftmost group of colored dots 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. (B) FoldX ΔΔG 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 ΔΔGs and evolutionary sequence energies. Pathogenic variants in red, benign in blue. Dot shape indicates whether pathogenicity of the respective variant was correctly predicted by regression model trained on all but this data point ("jackknife", TP, true positives, FN, false negatives, FP, false positives, TN, true negatives). (F) ROC curves for logistic regression model, FoldX ΔΔGs, evolutionary sequence energies, and 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. See also Figure 6 - figure supplements 1, 2 and 3.

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.

Figure 1

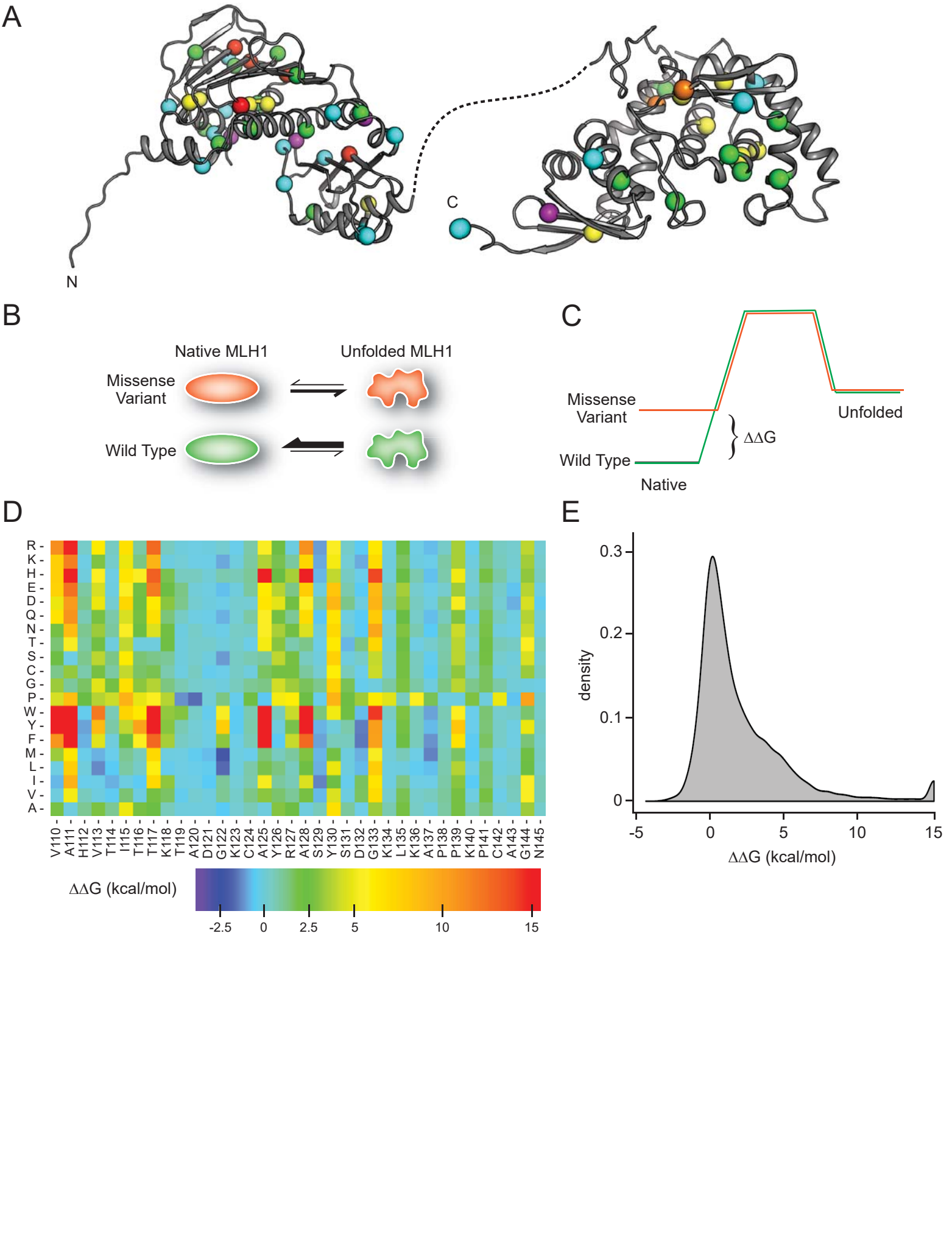

Figure 2

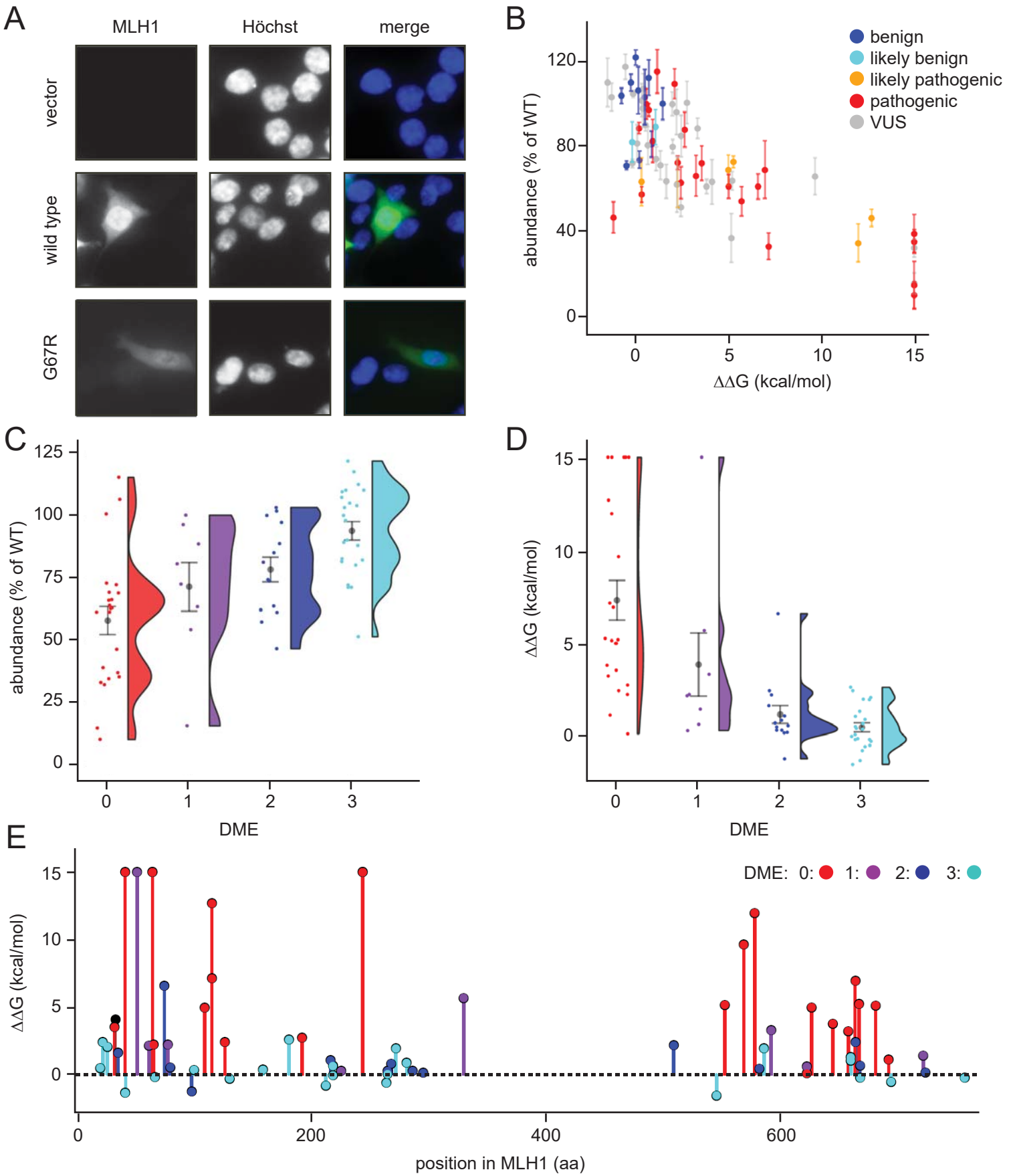

Figure 3

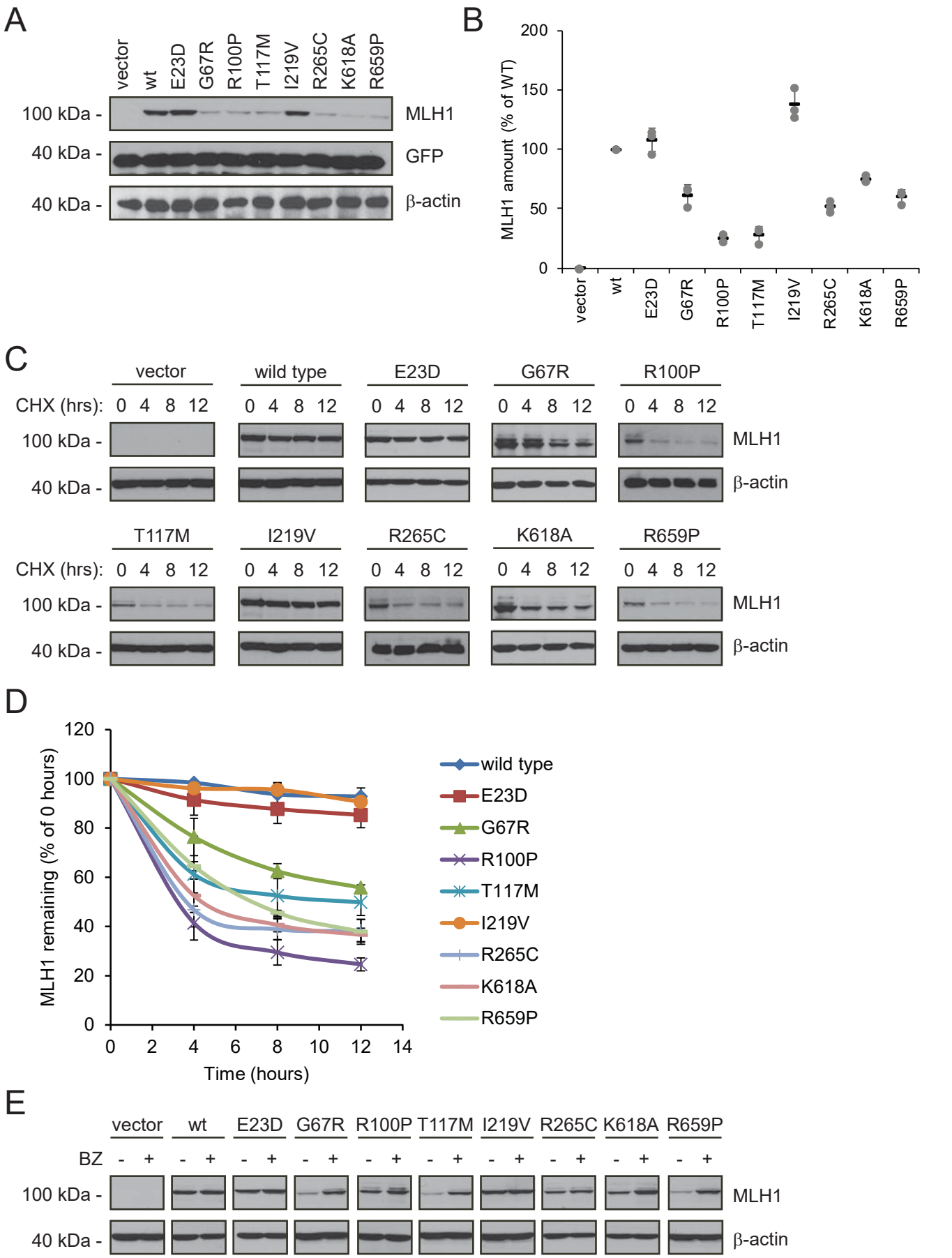

Figure 4

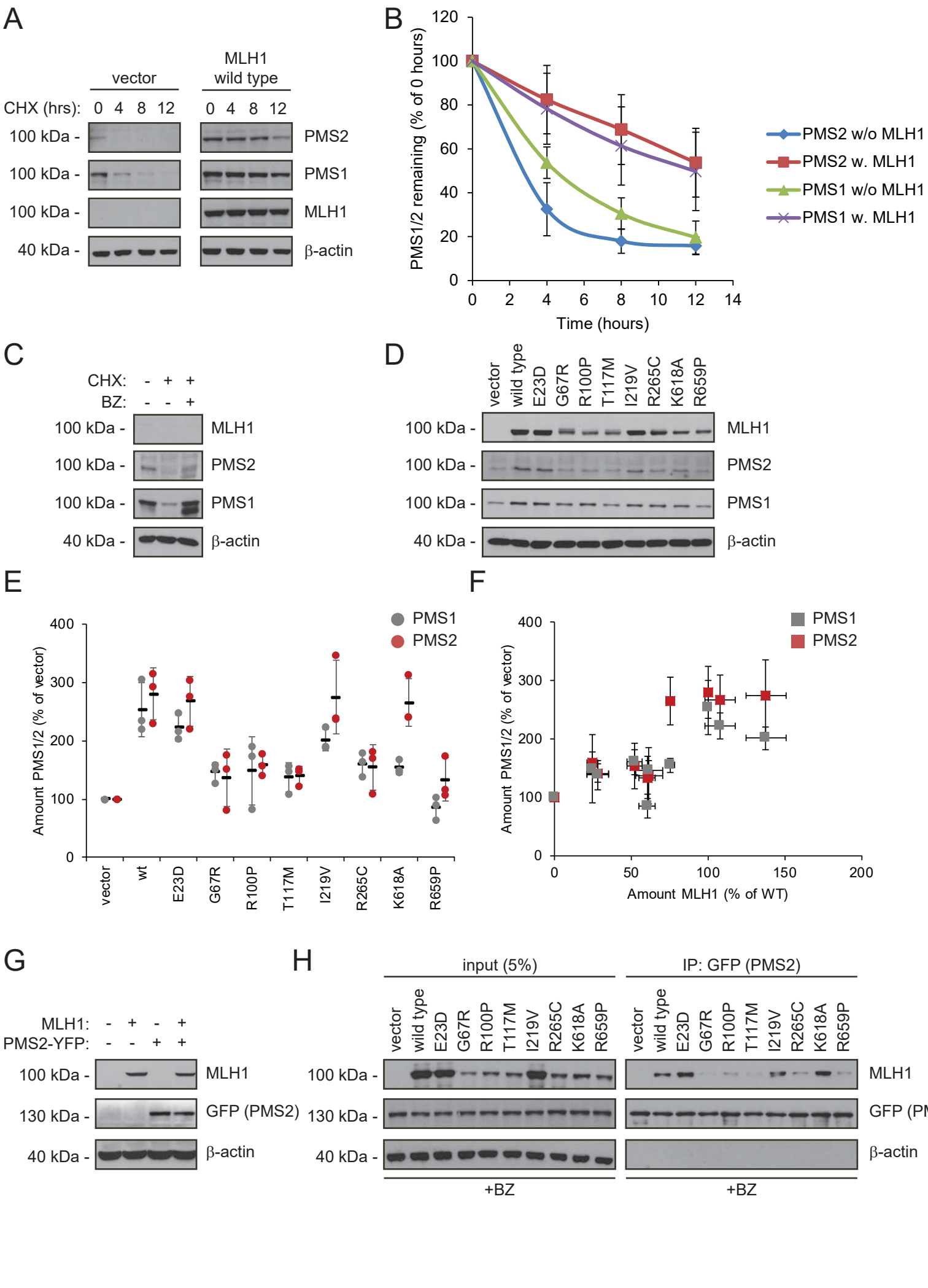

Figure 5

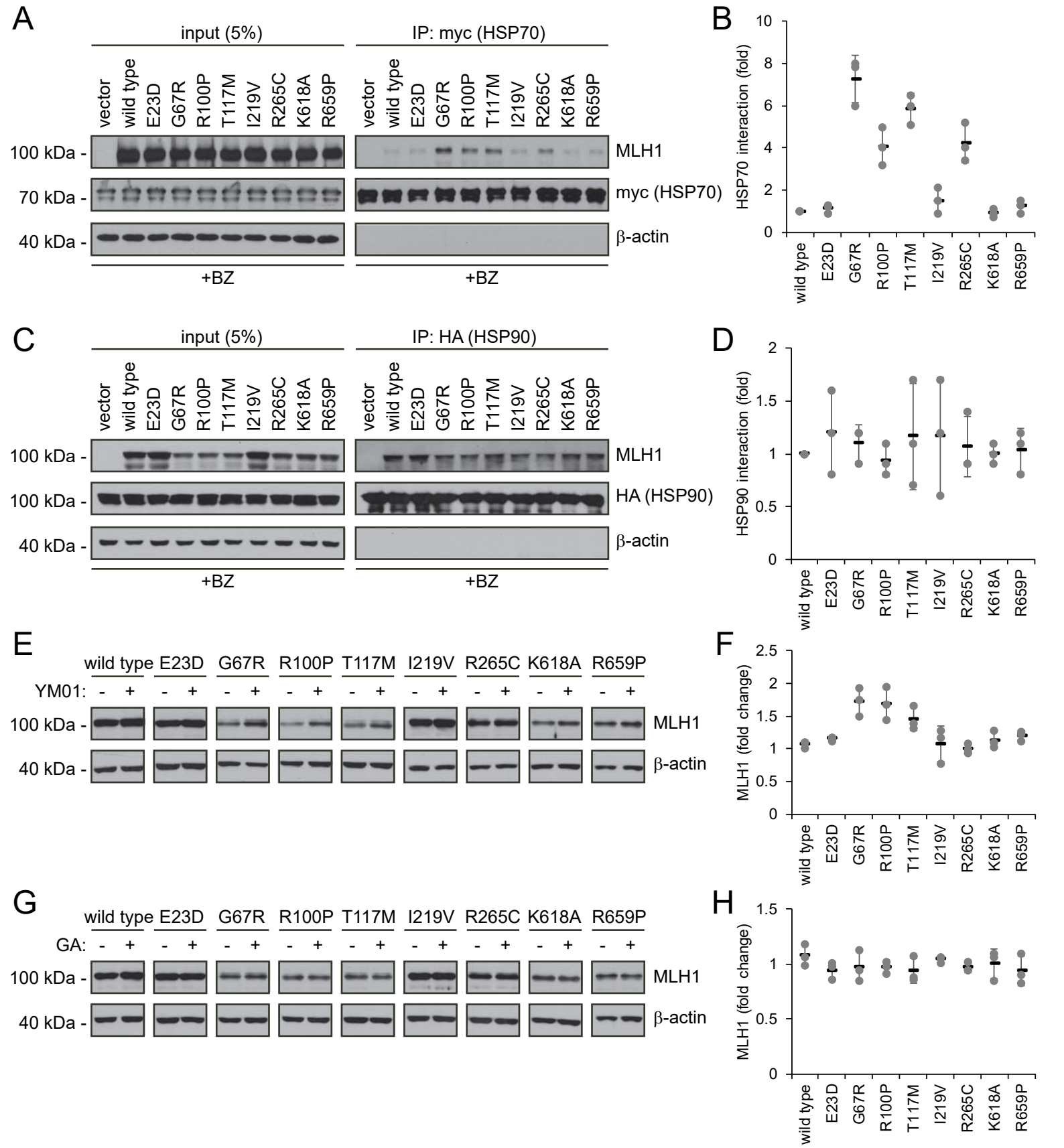

Figure 6

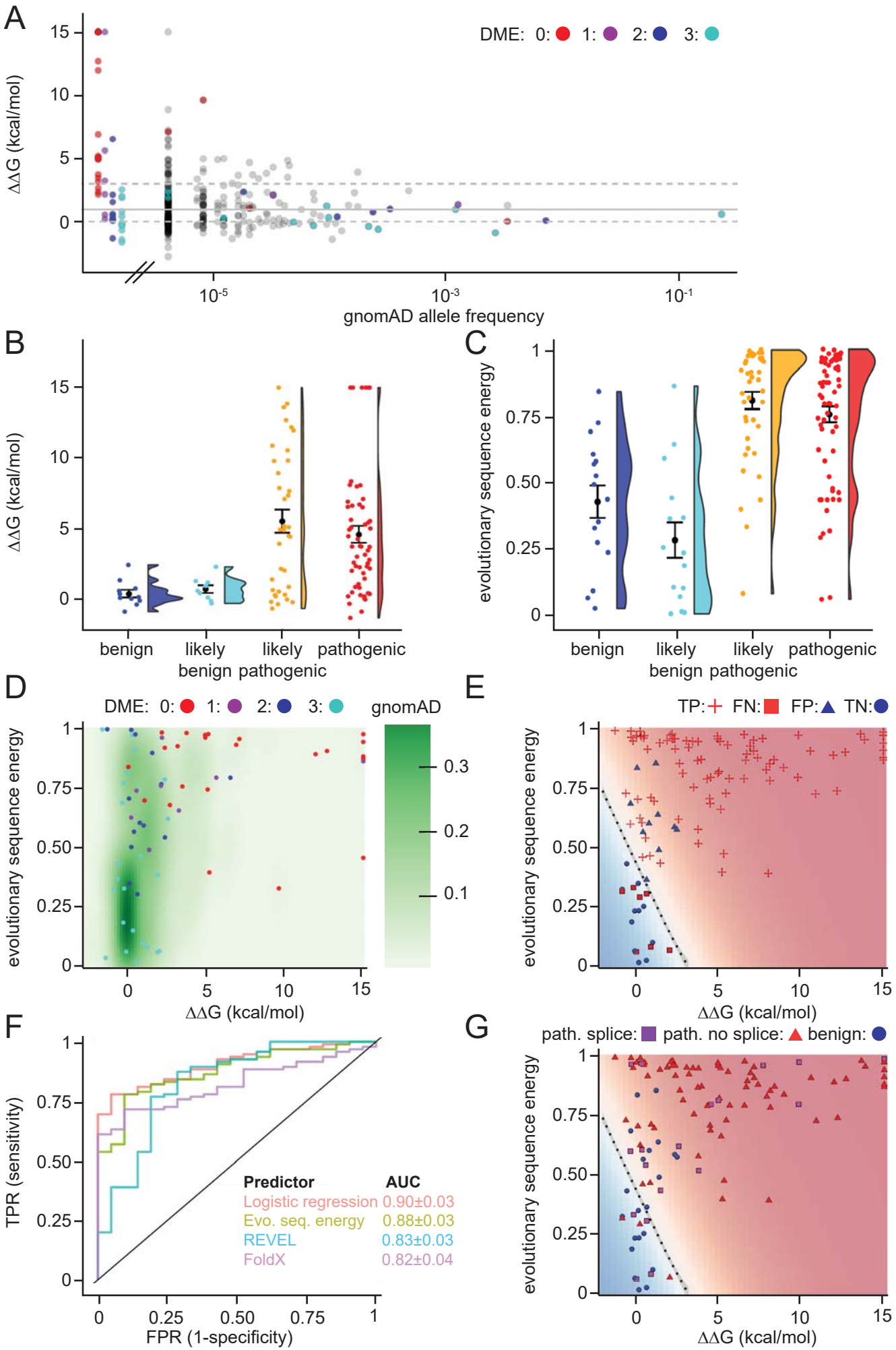

Figure 7