Disentangling the mutational effects on protein stability and interaction of human MLH1

Sven Larsen-Ledet[1,*], Aleksandra Panfilova[1] and Amelie Stein[1,*]

[1] Department of Biology, University of Copenhagen, Ole Maaløes Vej 5, DK-2200 Copenhagen, Denmark.

* Corresponding authors

E-mail: sven.larsenledet@bio.ku.dk, amelie.stein@bio.ku.dk

Abstract

Missense mutations can have diverse effects on proteins, depending on their location within the protein and the specific amino acid substitution. Mutations in the DNA mismatch repair gene *MLH1* are associated with Lynch syndrome, yet the underlying mechanism of most disease-causing mutations remains elusive. To address this gap, we aim to disentangle the mutational effects on two essential properties for MLH1 function: protein stability and protein-protein interaction. We systematically examine the cellular abundance and interaction with PMS2 of 4839 (94%) MLH1 variants in the C-terminal domain. Our combined data shows that most MLH1 variants lose interaction with PMS2 due to reduced cellular abundance. However, substitutions to charged residues in the canonical interface lead to reduced interaction with PMS2. Unexpectedly, we also identify a distal region in the C-terminal domain of MLH1 where substitutions cause both decreased and increased binding with PMS2, and propose a region in PMS2 as the binding site. Our data successfully distinguish benign from pathogenic MLH1 variants and correlate with thermodynamic stability predictions and evolutionary conservation. This work provides mechanistic insights into variant consequences and may help interpret MLH1 variants.

Author Summary

Mutations in proteins, including single amino acid substitutions, can disrupt normal cellular function and lead to diseases. Hence, there is great interest in understanding their consequences from both basic science and clinical genetics perspectives. Loss-of-function variants in the human mismatch repair protein MLH1 are associated with Lynch syndrome, a hereditary condition that significantly increases the risk of certain cancers. Detailed insights into the molecular consequences of MLH1 variants are essential for understanding the mechanisms underlying Lynch syndrome. In this study, we use high-throughput experimental approaches to test how missense mutations affect the cellular abundance of MLH1 and the interaction with its partner, PMS2. We find that many variants disrupting the interaction with PMS2 also lead to reduced cellular abundance levels, suggesting a strong relationship between protein stability and interaction in MLH1. Additionally, we identify a potential novel interaction site between MLH1 and PMS2, where mutations in MLH1 either increase or decrease interaction without affecting abundance. Our datasets are in line with the classification of clinically annotated MLH1 variants and may aid in interpreting variants of uncertain significance, while also providing new insights into the molecular details of the interaction with PMS2.

Introduction

Proteins constitute the functional units of the cell and perform essential cellular processes to sustain cell function and health. To obtain biological function, most proteins must fold into a marginally stable native structure, which is largely determined by the amino acid sequence [1]. This makes proteins extremely sensitive to missense mutations that disrupt the structural stability, leading to protein degradation and reduced cellular abundance. Thus, protein stability is fundamental for protein function and, indeed, the most frequent mechanism of disease-causing missense mutations is through destabilization and degradation [2–9]. However, in addition to protein stability, proteins often have other properties important for their function. The function of most proteins relies on interactions with other proteins to form complexes and networks, making protein-protein interactions (PPIs) a central property of protein function [10–12].

In general, PPIs can be grouped into obligate or non-obligate interactions based on whether the proteins function only as a complex or can also function as monomers [13,14]. Non-obligate interactions are often of lower affinity, transient, and involved in regulatory and signaling pathways, whereas obligate PPIs are typically stronger with larger and more hydrophobic interfaces [14–16]. Several studies have shown that disease-causing missense mutations are enriched in PPI interfaces [17–20], and it has been estimated that the underlying mechanism of around half of all disease-causing missense mutations in proteins that interact with other proteins is through the perturbation of specific interfaces [21–23]. This suggests that numerous clinically relevant protein variants affecting PPI interfaces remain stable in the cell, which highlights that mutations may affect protein stability and PPIs differently, depending on the type and site of the mutation. Mechanistic insights into mutational effects on protein stability and interactions have important implications for our understanding of protein function and diseases, and may provide avenues towards developing treatments.

MutL homolog 1 (MLH1) is a 756-residue protein that folds into an N- and a C-terminal domain, separated by a long intrinsically disordered linker [24]. MLH1 forms an obligate heterodimer with postmeiotic segregation increased 2 (PMS2), termed MutLα. Although the structure of the human MutLα complex remains unresolved, it is known that MLH1 and PMS2 dimerize through their C-terminal domains [25]. MutLα participates in the conserved DNA mismatch repair (MMR) pathway that detects and repairs base-base mismatches during DNA replication [26,27]. Defects in the MMR pathway increase the risk of spontaneous somatic mutations and sporadic cancers. Loss of function mutations in MLH1 cause Lynch syndrome, an inherited disease that predisposes individuals to various cancer types, such as colorectal and endometrial cancer [28,29]. A large proportion of Lynch syndrome patients suffer from pathogenic missense mutations in the *MLH1* gene [30,31]. However, the mechanistic effects of most of these variants are still unknown, and most of the MLH1 variants observed in the population are still categorized as variants of uncertain significance (VUS) [32–34].

Here, we use orthogonal yeast-based growth assays to disentangle the mutational effects on the abundance of MLH1 and on its interaction with PMS2. We systematically screen site-saturated libraries comprising 4839 (94%) missense and nonsense variants in the C-terminal domain of human MLH1. We find that MLH1 variants predominantly affect protein abundance compared to interaction and that these positions often are confined to buried residues. However, substitutions to charged residues in positions in the suggested interface between MLH1 and PMS2 are detrimental to interaction. Surprisingly, we find that a region distal to the interface confers improved interaction without affecting abundance. We find that the experimental results correlate with computational predictions and can distinguish clinically classified pathogenic variants from benign population variants. Our data suggest that single missense mutations are more likely to affect abundance than interaction but do pinpoint specific positions critical for interaction in known and novel interfaces with PMS2.

Results

Yeast-based assays to measure MLH1 abundance and interaction with PMS2

To disentangle the mutational effects on protein abundance and interaction, we selected the interaction between the human proteins MLH1 and PMS2. MLH1 and PMS2 form the MutLα complex through their C-terminal domains, an essential component of the DNA mismatch repair pathway (**Fig. 1A**). The structure of the human C-terminal MutLα complex is unknown, and the C-terminal domain of PMS2 has not been structurally resolved, making it challenging to accurately assign the interface between the two proteins. However, despite the moderate sequence identity of the human C-terminal domains (30% for MLH1 and 41% for PMS2) with their yeast orthologs, MLH1 and PMS1, an AlphaFold2-Multimer-predicted model of the human C-terminal complex exhibits relatively high structural similarity (RMSD = 2.08) to the yeast complex (PDB: 4E4W) [35]. Therefore, we used AlphaFold2-Multimer to model the human C-terminal MutLα complex (**Fig. 1A**, **Supplementary Fig. 1**). The structure predicted that MLH1 and PMS2 dimerize through four β-strands, three of which are close in sequence, while the fourth is located towards the C-terminus.

To measure the abundance of MLH1 variants, we used a previously described dihydrofolate reductase protein-fragment complementation assay (DHFR-PCA) (**Fig. 1B**) [36–39]. This assay is based on a methotrexate-resistant variant of murine DHFR, which is split into two fragments, F[1,2] and F[3], that are expressed separately but from the same centromere-based vector in wild-type *S. cerevisiae* cells (**Supplementary Fig. 2**). The F[3] is fused to MLH1, while the F[1,2] remains free (untagged). Plating of cells on medium with methotrexate (MTX) inhibits the essential endogenous yeast Dfr1 and makes growth dependent on reconstitution of the methotrexate-resistant variant of murine DHFR. A stable, high-abundant MLH1 variant enables reconstitution, allowing growth in the presence of MTX, whereas an unstable, low-abundance MLH1 variant prevents reconstitution, resulting in reduced growth in the presence

124 of MTX. Hence, the DHFR-PCA couples cellular abundance of MLH1 to yeast growth.
125 However, this assay solely measures the abundance of MLH1 variants and hence is not
126 amenable to report on other properties important to MLH1 function, such as its interaction with
127 PMS2. Therefore, we used the yeast two-hybrid (Y2H) assay to assess the binding of MLH1
128 variants to PMS2 (**Fig. 1C**) [40]. In this assay, MLH1 and PMS2 are fused to the activation
129 domain (AD) and DNA binding domain (DBD) of the GAL4 transcription factor, respectively.
130 The two fusion proteins are expressed from their own centromere-based vectors in the MaV203
131 yeast strain (**Supplementary Fig. 3A**, **Supplementary Fig. 3B**). If MLH1 and PMS2 interact,
132 the AD and DBD reconstitute the Gal4 transcription factor and initiate the expression of the
133 *HIS3* reporter gene, enabling growth on medium without histidine.
134 In order to examine the sensitivity and dynamic range of the DHFR-PCA and Y2H assay, we
135 performed small-scale yeast growth assays of the wild-type and five MLH1 variants (**Fig. 2A**).
136 The Q542L and L550P variants locate to the C-terminal domain of MLH1, the S406N variant
137 locates to the intrinsically disordered linker and the I219V and T117M variants locate to the N-
138 terminal domain of MLH1. The S406N and I219V variants are classified as benign in the
139 ClinVar database [41] and frequently observed in the population, as reported in the Genome
140 Aggregation Database (gnomAD) [42]. The L550P and T117M variants are known to cause
141 Lynch syndrome, are classified as pathogenic, and are extremely rare in the population. The
142 Q542L variant is classified as a VUS and has been shown to exhibit wild-type-like abundance,
143 interaction with PMS2, and MMR activity [25,31,43], although some studies suggest reduced
144 PMS2 interaction and MMR activity [44,45]. In both assays, the empty vector displayed a
145 strong growth defect in the presence of MTX and absence of histidine, respectively, compared
146 to the wild-type (**Fig. 2A**). The Q542L, S406N, and I219V variants showed similar growth to
147 the wild-type, whereas the L550P and T117M variants exhibited a similar growth defect to the
148 empty vector (**Fig. 2A**). This suggests that the L550P and T117M variants likely lose PMS2

149 interaction due to reduced cellular abundance. In conclusion, this data confirmed that the
150 DHFR-PCA and Y2H assay could report on MLH1 abundance and its interaction with PMS2,
151 respectively, and distinguish benign population variants from clinically pathogenic ones.
152
153 ***Saturated abundance and interaction maps of MLH1 variants***
154 Next, we used the DHFR-PCA and Y2H assay to determine the abundance and PMS2
155 interaction of all possible missense variants in the C-terminal domain (positions 487-756) of
156 human MLH1 (**Fig. 2B**). We decided to focus only on the C-terminal domain of MLH1, as it
157 mediates the interaction with PMS2, and we expected variants in the N-terminal domain to
158 have little impact on this interaction. We divided the C-terminal domain into seven tiles
159 (**Supplementary Fig. 4A**) and designed site-saturation mutagenesis libraries in each tile,
160 which were synthesized and cloned into a Gateway-compatible vector to produce the entry
161 libraries. The entry libraries were cloned into destination vectors tailored to the DHFR-PCA
162 and Y2H assay, respectively, and subsequently transformed into appropriate yeast strains. The
163 transformation cultures were grown until saturation and then cells were plated in triplicates
164 onto control and selection medium, and incubated at 30 °C. In the DHFR-PCA, cells were
165 grown on medium with MTX to select for MLH1 abundance, and in the Y2H assay, cells were
166 grown on medium without histidine to select for PMS2 interaction. Following incubation, cells
167 were scraped off the plates and directly used for plasmid isolation and amplicon preparation.
168 The amplicons were sequenced using paired-end next-generation sequencing and the resulting
169 sequencing data was analyzed using Enrich2 [46] to calculate an abundance score and a PMS2
170 interaction score for each MLH1 variant (**Fig. 2B**). The variant scores were calculated by
171 comparing the change in frequency of each variant to that of synonymous wild-type variants
172 in both the control and selection conditions. The MLH1 abundance scores displayed a bimodal
173 distribution with many variants centered around 0 (wild-type-like) and a peak of low-

174 abundance variants with negative scores (**Supplementary Fig. 4B**). However, due to the
175 observed variance in negative scores across tiles within the low-abundance peak
176 (**Supplementary Fig. 4B**) we rescaled the abundance scores to range from -1 to 1 (**Fig. 2C**),
177 which standardized the abundance scores and ensured uniformity across tiles. The distribution
178 of interaction scores largely appeared centered around 0 (wild-type-like) with some variants
179 showing reduced interaction with negative interaction scores, while a few variants exhibited
180 improved interaction with positive interaction scores (**Fig. 2C**, **Supplementary Fig. 4C**). The
181 interaction scores were not rescaled. As expected, the L550P variant displayed reduced
182 abundance and interaction, in line with the results from the small-scale yeast growth assays and
183 underscoring the validity of the high-throughput screens (**Fig. 2A**, **Fig. 2C**).
184 The abundance and interaction datasets contained variant effects for 4839 (94%) out of 5170
185 (19 substitutions/position*270 residues + 40 nonsense mutations) possible programmed MLH1
186 variants in the C-terminal domain. The abundance and interaction scores showed high
187 correlation between the three replicates, but the DHFR-PCA had relatively lower correlations
188 compared to the Y2H assay (minimum R = 0.92 vs. minimum R = 0.96) (**Supplementary Fig.
189 5**, **Supplementary Fig. 6**). Overall, substitutions in the C-terminal domain of MLH1 more
190 often conferred reduced abundance than affecting the interaction with PMS2 (**Fig. 3A**, **Fig.
191 3C**). The abundance map revealed that substitutions were detrimental in several positions and
192 particularly in secondary structural elements (**Fig. 3A**). The internal loop spanning position
193 591-605, as well as some positions close to the N- and C-termini were tolerant to substitutions
194 (**Fig. 3A**). As expected, the disordered region with low AlphaFold2-Multimer-predicted
195 pLDDT scores at position 692-711 was insensitive to substitutions, whereas substitutions to
196 proline appeared detrimental to abundance, and to a lesser extent to interaction, at most
197 positions (**Fig. 3**). Interestingly, nonsense mutations caused reduced abundance and interaction
198 up until position 742, but from position 745-756, stop codon variants displayed increased

199 abundance compared to the wild-type (**Fig. 3**). Additionally, several point mutations at
200 positions toward the C-terminus led to increased abundance and interaction (**Fig. 3**). We
201 mapped the median abundance scores onto the C-terminal domain of MLH1 in the AlphaFold2-
202 Multimer-predicted structure of the human C-terminal MutLα complex (**Fig. 3B**). This revealed
203 that most of the sensitive positions were pointing into the core of the protein, while exposed
204 and flexible positions were tolerant to substitutions. Indeed, the median abundance scores
205 correlated with relative solvent accessible surface area (rSASA), and indicated that
206 substitutions in buried, core positions were more detrimental than in surface-exposed positions
207 (**Supplementary Fig. 7A**).
208 The interaction map revealed that the impact of MLH1 variants on the interaction with PMS2
209 was confined to specific positions rather than being widespread (**Fig. 3C**). In general, most
210 variants that displayed reduced interaction with PMS2 also appeared low-abundant, suggesting
211 that loss of stability was the primary mechanism of reduced interaction with PMS2 (**Fig. 3A**,
212 **Fig. 3C**). In line with this, we observed that low median interaction scores were only observed
213 for buried positions (**Supplementary Fig. 7B**). Large regions at the beginning and end of the
214 C-terminal domain of MLH1 were completely tolerant to substitutions, with most variants
215 exhibiting wild-type-like interaction with PMS2 (**Fig. 3C**). It was evident that substitutions
216 from hydrophobic wild-type residues to both negatively and positively charged amino acids in
217 the β-strands constituting the canonical interface at position 539-541 and 547-550 were
218 detrimental to the interaction (**Fig. 3C**). Interestingly, at position 540, the substitution to
219 glutamic acid had no effect on the interaction with PMS2, whereas the substitution to aspartic
220 acid displayed complete loss-of-interaction (**Fig. 3C**). We noticed that the introduction of
221 tryptophan at position 544, which is located in the short loop between the second and third β-
222 strand of the canonical interface, substantially improved the binding with PMS2 (**Fig. 3C**). In
223 addition, MLH1 variants immediately upstream of the central α-helix, specifically at position

601-604, showed increased binding with PMS2 (**Fig. 3C**). This effect was most pronounced for substitutions to hydrophobic or aromatic residues and disappeared with substitutions to charged residues (**Fig. 3C**). To validate the effects on PMS2 interaction observed in this screen, we performed low-throughput Y2H growth assays on four loss-of-interaction variants and three gain-of-interaction variants. These experiments confirmed the effects observed in the high-throughput screen (**Supplementary Fig. 8**).

To visualize the variant effects on PMS2 interaction, we mapped the median interaction scores onto MLH1 in the predicted C-terminal MutLα complex (**Fig. 3D**). The mapping revealed that substitutions of surface-exposed residues in the central α-helix spanning position 605-627 exhibited slightly decreased interaction with PMS2, without affecting the MLH1 abundance (**Fig. 3B**, **Fig. 3D**). This led us to speculate that MLH1 and PMS2 interact through an additional interface, distal to the canonical four β-strands known from the yeast complex [35] (**Supplementary Fig. 9**). Indeed, when we predicted the full structure of the human MutLα complex using AlphaFold3, we identified an α-helix (positions 409-431) within the intrinsically disordered linker region of PMS2 that directly interacts with the distal central α-helix in MLH1 (**Supplementary Fig. 10A**, **Supplementary Fig. 10B**). Furthermore, mapping the median interaction scores to the structure revealed that several contact residues in MLH1 within 6 Å of the PMS2 409-431 α-helix located to positions with both loss- and gain-of-interaction effects in our Y2H assay (**Supplementary Fig. 10C**). These findings support the additional interface in the AlphaFold 3-predicted model and suggest that the disordered linker region in PMS2 harbors an α-helix forming a previously uncharacterized interface with MLH1. The mapping of the median interaction scores to the structure also revealed that several positions in the helices neighboring the central α-helix in MLH1 were highly sensitive to substitutions (**Fig. 3D**). However, these residues pointed into the protein core, and substitutions

248 also resulted in reduced abundance, suggesting that the loss of interaction was likely due to
249 decreased stability (**Fig. 3B**, **Fig. 3D**).
250
251 ***Computational predictions and clinical data correlate with variant effects***
252 The above data indicate that MLH1 abundance and its interaction with PMS2 are tightly
253 connected and loss of interaction is often caused by loss of stability. To further examine this
254 relationship, we used predictive computational tools to evaluate the impact of MLH1 variants
255 on thermodynamic folding stability and their likelihood of being tolerated based on the
256 evolutionary history. We used Rosetta [47] to predict the changes in thermodynamic stability
257 (ΔΔG) of all possible missense variants in the C-terminal domain of MLH1. However, in
258 addition to stability, certain residues may be important for other aspects of MLH1 function,
259 such as its interaction with PMS2. Such residues are often conserved across species, and
260 therefore analysis of homologous sequences can report on the mutational tolerance at these
261 positions. To this end, we used GEMME [48] to model the evolutionary conservation of MLH1
262 sequences and then predict the required adaptation of the MLH1 sequence to accommodate all
263 possible missense variants in the C-terminal domain of MLH1. GEMME returned a ΔΔE score
264 for each MLH1 variant, ranging from 0 (tolerated substitution) to -6 (non-tolerated substitution
265 based on evolutionary conservation). Rosetta and GEMME separated the MLH1 variants
266 according to their predicted effects on MLH1 stability and other properties important for
267 function, respectively (**Fig. 4A**, **Supplementary Fig. 11**, **Supplementary Fig. 12**). We
268 observed that the majority of variants were predicted to have a ΔΔG score below 2 and ΔΔE
269 score above -2, suggesting that they exhibited similar stability and functional properties as the
270 wild-type (**Fig. 4A**). Several variants were predicted to be incompatible with MLH1 function
271 (strongly negative ΔΔE scores) and some of them were also predicted to be unstable (positive
272 ΔΔG score) (**Fig. 4A**). We mapped the experimental abundance scores to the predictions, which

273 revealed that variants with wild-type-like abundance clustered in the upper left corner, whereas
274 low-abundant MLH1 variants mapped towards the bottom right corner (**Fig. 4A**). An analogous
275 pattern appeared when we mapped the experimental interaction scores to the predictions (**Fig.**
276 **4B**, **Supplementary Fig. 13**). This indicates that most of the loss-of-interaction variants were
277 predicted to be detrimental based on the evolutionary history while also losing structural
278 stability and displaying reduced abundance levels, confirming the close relationship between
279 MLH1 abundance and its interaction with PMS2.
280 To assess the clinical relevance of the generated datasets, we next compared the experimental
281 abundance and interaction scores with the clinical classifications in the ClinVar database [41]
282 and population allele frequencies of known MLH1 variants that have been reported in gnomAD
283 [42] (**Fig. 4C**, **Fig. 4D**). This showed that variants that are common in the human population
284 tended to behave like the wild-type, whereas several of the rare population MLH1 variants
285 displayed reduced abundance and interaction with PMS2. To evaluate if our abundance and
286 interaction scores could differentiate harmless benign variants from disease-associated
287 pathogenic variants, we colored the MLH1 variants based on their classification in ClinVar [41]
288 (**Fig. 4C**, **Fig. 4D**). The five pathogenic MLH1 variants were extremely rare in the population,
289 and all displayed much reduced abundance (**Fig. 4C**). Conversely, most of the benign and likely
290 benign MLH1 variants were frequently found among individuals and exhibited wild-type-like
291 protein levels (**Fig. 4C**). Additionally, several rare VUS exhibited similarly low abundance
292 levels as the pathogenic variants, indicating that these variants are likely to be disease-causing
293 (**Fig. 4C**). The K618T variant had an abundance score of -1, despite being classified as benign
294 (**Fig. 4C**). This variant has previously been shown to have reduced cellular abundance, nuclear
295 localization and MMR activity [43,49–52], while other studies have suggested this variant to
296 form a stable protein [31]. Thus, despite the variant being too frequent in the population to be
297 a fully penetrant Lynch syndrome variant, its clinical interpretation remains inconclusive [53].

298 The interaction scores were also able to distinguish most benign from pathogenic variants (**Fig.**
299 **4D**). The pathogenic R687T variant showed wild-type-like interaction with PMS2 but
300 dramatically decreased abundance, suggesting that the disease-mechanism of this variant is
301 through loss of stability (**Fig. 4C**, **Fig. 4D**). Interestingly, some of the MLH1 variants that
302 displayed improved interaction with PMS2 were classified as VUS but were extremely rare in
303 the population (**Fig. 4D**). This could suggest that these variants are pathogenic, and thus gain
304 of interaction may represent an additional underlying disease-mechanism of Lynch syndrome.
305 Lastly, we performed receiver operating characteristic (ROC) analyses and calculated the area
306 under the curve (AUC) to assess whether our experimental scores and computational
307 predictions could distinguish pathogenic variants from benign ones (**Supplementary Fig. 14**).
308 This analysis revealed that both our abundance and interaction scores accurately predict
309 disease-causing MLH1 variants, both achieving AUCs of 0.89 (**Supplementary Fig. 14)**.
310 Similarly, Rosetta (AUC = 0.84) and GEMME (AUC = 0.91) demonstrated comparable
311 predictive power, suggesting that these computational prediction models capture MLH1
312 pathogenicity almost as effectively as our experimental data (**Supplementary Fig. 14**).
313 However, it is important to emphasize that although GEMME slightly outperforms our
314 experimental data, it only generally predicts MLH1 function and does not provide mechanistic
315 insights into variant effects. It should also be noted that the test set only consists of 46 clinically
316 classified variants, so additional data might reveal difference in performance.

317 **Discussion**

318 Most proteins have several properties that contribute to their function. Here, we have combined
319 two multiplexed assays of variant effects (MAVEs), one probing stability and another probing
320 interaction, to disentangle which variants affect either or both properties in the C-terminal
321 domain of human MLH1.

322 Loss-of-function mutations in the *MLH1* gene lead to Lynch syndrome, a cancer predisposing
323 disease. Loss of stability has been shown to be a common mechanism of loss-of-function
324 phenotypes [3,31,32,54,55], and pathogenic variants are enriched in interaction interfaces [17–
325 20]. Our data helps pinpoint whether the molecular mechanism underlying pathogenicity of a
326 given variant is loss of stability, loss of interaction with PMS2, or yet other properties of the
327 protein. Our results indicate that several mutations reduce protein stability, while far fewer
328 variants affect the interaction. At the same time, variants that result in a loss of interaction often
329 display reduced abundance levels. These observations suggest that it is difficult to perturb the
330 interaction with PMS2 without also compromising the structure of the MLH1 protein. We
331 envision that this intricate relationship between structural protein stability and protein-protein
332 interactions (PPIs) reflects the nature of large and structured interfaces, such as the interaction
333 between MLH1 and PMS2. Thus, for single MLH1 missense mutations to affect the interaction
334 with PMS2, it requires strong effects that are unlikely to occur without also structurally
335 affecting MLH1 stability. This contrasts with PPIs mediated through smaller interfaces and
336 involving unstructured components, such as short linear motifs (SLiMs) in intrinsically
337 disordered regions [56,57]. In those cases, missense mutations are expected to have local
338 effects on specific binding sites rather than globally perturbing protein stability. We speculate
339 that the novel interaction between the α-helix in PMS2 (positions 409-431) and the distal
340 central α-helix in the C-terminal domain of MLH1 represents such a motif-mediated interaction

341 (**Supplementary Fig. 10**), which explains why individual substitutions specifically disrupt the
342 interaction without compromising protein stability.
343 Despite the difficulty in disentangling protein stability and PMS2 interaction for MLH1
344 variants, we observe that substitutions to charged residues in two of the four β-strands
345 constituting the canonical interface with PMS2 are detrimental to the interaction. We note that
346 this interface is enriched with hydrophobic residues and completely depleted of charged
347 residues, which may explain why the introduction of charged residues in these β-strands is
348 unfavorable for the interaction with PMS2 (**Supplementary Fig. 15**) [58]. Interestingly, at
349 position 540, only substitutions to aspartic acid and arginine are detrimental, whereas
350 substitutions to glutamic acid and lysine exhibit wild-type-like interaction. This suggests that
351 charge is not the only determinant of the effect on PMS2 interaction; spatial constraints and the
352 local environment of the interface also influence the effect of substitutions. We identify a single
353 gain-of-interaction variant, Q544W, in the canonical interface, located in the short loop
354 between the second and third β-strand. We speculate that introducing a tryptophan residue at
355 this site might alter the loop's conformation, and that the inherent stickiness of tryptophan could
356 promote new molecular interactions with PMS2. In addition to the canonical interface,
357 substitutions within and immediately before a central α-helix of MLH1 lead to reduced and
358 increased binding with PMS2, respectively. This effect is presumed to occur via an α-helix that
359 extends towards the disordered linker region of PMS2. We hypothesize that this is a novel
360 interaction site between MLH1 and PMS2.
361 Counterintuitively, some MLH1 variants that appear unstable still exhibit wild-type-like
362 interaction with PMS2. We confirm this observation for selected variants in low-throughput
363 (**Supplementary Fig. 16A**, **Supplementary Fig. 16B**). This suggests differences in the
364 dynamic range and sensitivity of the DHFR-PCA and Y2H assay. We believe there are several
365 possible explanations for this. First, the expression of the *HIS3* gene in the Y2H assay likely

366 provides a much more sensitive read-out compared to the reconstitution of the murine DHFR
367 in the DHFR-PCA. This sensitivity difference could potentially be mitigated by using a higher
368 concentration of 3-AT to increase the selection pressure on the interaction, thereby shifting the
369 dynamic range to select for variants that mildly affect the interaction without perturbing the
370 protein stability. Second, it has been shown that the DHFR-PCA underestimates the abundance
371 of proteins localized to subcellular compartments (e.g., the nucleus), as the binding of the
372 F[1,2] and F[3] fragments of murine DHFR occurs in the cytosol [38]. However, studies have
373 shown that translocation of MLH1 into the nucleus depends on its heterodimerization with
374 PMS2 [59–62]. Since the DHFR-PCA is performed without co-expression of PMS2, a large
375 fraction of MLH1 is expected to be present in the cytosol. We hypothesize that this ectopic
376 cytosolic localization and absence of PMS2 negatively affect MLH1 stability. Despite these
377 limitations, we observe a reasonable correlation with existing literature on MLH1 abundance
378 [31,63] and PMS2 interaction [44], especially given differences in cell types and experimental
379 techniques used (**Supplementary Fig. 17**).
380 Our abundance and interaction maps reveal that removal of residues 745-756 of MLH1 leads
381 to both enhanced stability and interaction with PMS2. We find it unlikely that this effect is due
382 to a decrease in degron potency, as this region is not enriched with hydrophobic residues nor
383 depleted in acidic residues, which are characteristic of protein quality control degrons [64–66].
384 Additionally, since the region near the C-terminus does not contain any known C-degrons [67–
385 70], deleting these residues should not lead to increased abundance by removing a C-degron.
386 However, the C-terminus of MLH1 contains an extremely conserved FERC motif, which is
387 suggested to have important regulatory and activity functions [35,71–74]. Therefore, we
388 hypothesize that truncating the C-terminus stabilizes the protein and creates additional space
389 around the canonical interface, thereby enhancing binding with PMS2, while potentially
390 affecting other properties of MLH1 function mediated by the FERC motif. In general, our

391 experimental abundance and interaction effects are in line with clinical variant classification,
392 as well as with thermodynamic stability predictions and evolutionary conservation. It is,
393 however, important to emphasize that this is not a symmetric relationship. Our assays can
394 identify residues that are detrimentally affected and hence at high risk of being pathogenic. If
395 a variant behaves like wild-type in both assays, though, this does not necessarily imply that
396 these variants have functional MMR activity or should be considered benign. As an example,
397 beyond the truncated C-terminus, some wild-type-like variants are predicted to be extremely
398 rare or absent from the family of MLH1 homologs ($\Delta\Delta E < -3$). These variants could be in
399 positions that are critical for other functions of the C-terminal domain of MLH1, such as its
400 interaction with EXO1, which occurs via a different binding site than its interaction with PMS2
401 [35,75]. It has also been observed that mutational consequences may be coupled to other
402 properties than the interaction partner binding to the site of mutation [76] and it will be
403 interesting to revisit this when data on additional MLH1 binding partners is available.
404 In conclusion, our data suggests that the abundance of MLH1 and its interaction with PMS2
405 are closely related, with impaired interaction often caused by reduced cellular abundance. We
406 identify detrimental mutations in the canonical interface with PMS2 and uncover a potential
407 distal binding site between MLH1 and PMS2. Our abundance and interaction maps may aid in
408 interpreting clinically observed MLH1 variants and enhance our understanding of their effects
409 in Lynch syndrome.

410 **Materials and Methods**

411

412 **Plasmids and cloning**

413 The DNA sequence of human wild-type MLH1 (UniProt ID: P40692) and human wild-type
414 PMS2 (UniProt ID: P54278) were codon-optimized for yeast expression and cloned into
415 pDONR221 (Genscript). Missense variants were created by Genscript. For the Y2H assay,
416 entry clones were cloned into pDEST22 (prey) or pDEST32 (bait) from the ProQuest Two-
417 Hybrid System (Invitrogen) using Gateway cloning (Invitrogen). For the DHFR-PCA, entry
418 clones were cloned into pDEST-DHFR-PCA using Gateway cloning (Invitrogen).

419

420 **Yeast strains and media**

421 The Y2H assay was performed in the MaV203 yeast strain from the ProQuest Two-Hybrid
422 System (Invitrogen) and the DHFR-PCA was performed in the BY4741 wild-type yeast strain.
423 Single variant transformations were done as described in [77]. Yeast cells were grown in yeast
424 extract, peptone, dextrose (YPD) medium (2% glucose, 2% tryptone, 1% yeast extract) and in
425 synthetic complete (SC) medium (2% glucose, 0.67% yeast nitrogen base without amino acids
426 (Sigma), 0.2% synthetic drop-out supplement (Sigma)). For the Y2H assay, SC medium
427 without histidine was prepared with 20 mM 3-amino-1,2,4-triazole (3-AT) (Sigma-Aldrich),
428 unless otherwise specified. For the DHFR-PCA yeast growth assays, SC medium without uracil
429 was prepared using drop-out supplement. For the DHFR-PCA library screen, SC medium
430 without uracil was prepared using 76 mg/L histidine, 380 mg/L leucine, 76 mg/L methionine.
431 Selection medium for the DHFR-PCA was prepared using 200 μM methotrexate (MTX)
432 (Sigma-Aldrich) and 1 mM sulfanilamide (Sigma-Aldrich). Control medium for the DHFR-
433 PCA was prepared using a corresponding volume of DMSO. Solid media was prepared using
434 2% agar.

435 **Yeast growth assays**

436 For growth assays on solid media, cultures were grown overnight (170 rpm, 30 °C) to
437 exponential phase. The cultures were harvested and washed in sterile water (3000 rpm, 5
438 minutes, room temperature) and the OD_{600nm} was adjusted to 0.4. Then, cultures were used for
439 five-fold dilution series in sterile water and 5 µL of each dilution was spotted onto agar plates.
440 The plates were briefly air-dried before they were incubated at 30 °C for 2-3 days.

441

442 **Library design and cloning**

443 Full-length human MLH1 was divided into seven tiles of 28, 40 or 42 residues. The residues
444 included in each tile were: tile 1: 487-526, tile 2: 527-568, tile 3: 569-596, tile 4: 597-636, tile
445 5: 637-676, tile 6: 677-716 and tile 7: 717-756. For each tile, the library was synthesized and
446 cloned into pDONR221 by Twist Bioscience. The libraries comprised all possible single amino
447 acid substitutions at each position as well as the wild-type sequence and a synonymous wild-
448 type mutation at each position. Tile 7 also included a nonsense mutation at each position.
449 Mutagenesis of 15 positions and a total of 307 variants failed during library preparation by
450 Twist Bioscience. The final libraries comprised 4839 out of 5170 (94 %) designed possible
451 single substitution variants and nonsense variants (19 substitutions/per residue*270 residues +
452 40 nonsense variants). From this point forward, libraries in each tile were handled individually.
453 Entry libraries were cloned into appropriate destination vectors using Gateway cloning
454 (Invitrogen). Briefly, LR reactions were prepared with 243 ng (for the Y2H assay) or 260 ng
455 (for the DHFR-PCA) pDONR221-library, 450 ng pDEST22 (for the Y2H assay) or 450 ng
456 pDEST-DHFR-PCA (for the DHFR-PCA), 6 µL Gateway LR Clonase II enzyme mix
457 (ThermoFisher) and TE buffer to 30 µL. The LR reactions were incubated overnight at room
458 temperature and the following day, the reactions were stopped by adding 3 µL proteinase K (10
459 minutes, 37 °C). The cloned libraries were electroporated into NEB 10-beta electrocompetent

460 *E. coli* cells (New England Biolabs) using 4 µL LR reaction and 100 µL cells at 1.8 kV. Cells
461 were recovered in 3.9 mL NEB 10-beta stable outgrowth medium for 1 hour (250 rpm, 37 °C).
462 Then, cells were transferred to 36 mL LB medium and 4 µL was plated on LB medium with
463 ampicillin and incubated overnight (37 °C) for estimating the transformation efficiency. The
464 remaining cells were plated on large BioAssay plates (500 cm^2) containing LB medium with
465 ampicillin and incubated overnight (37 °C). For each tile, the number of single colonies
466 exceeded the expected number of variants by at least 100-fold. The following day, cells were
467 scraped off the plates with sterile water and library DNA was isolated and purified using the
468 NucleoBond Xtra Midi kit (Macherey-Nagel).
469
470 **Library transformation and selection**
471 The cloned MLH1 libraries in each tile were transformed and plated for selection separately.
472 The libraries cloned into pDEST22 were transformed into the MaV203 yeast strain with
473 pDEST32-PMS2 to perform the Y2H assay. The libraries cloned into pDEST-DHFR-PCA were
474 transformed into the BY4741 wild-type yeast strain to perform the DHFR-PCA.
475 Transformations were done using 30× scale-up as described in [78]. For the Y2H assay, the
476 MaV203 yeast strain with pDEST32-PMS2 was grown overnight (170 rpm, 30 °C) in SC-
477 leucine medium. For the DHFR-PCA, the BY4741 wild-type yeast strain was grown overnight
478 (170 rpm, 30 °C) in YPD medium. Upon reaching late-exponential phase, the cultures were
479 diluted in at least 150 mL SC-leucine medium (for the Y2H assay) or YPD medium (for the
480 DHFR-PCA) to an OD_{600nm} of 0.3 and incubated (170 rpm, 30 °C) until an OD_{600nm} of 1.2 was
481 reached. The cells were harvested and washed in sterile water (3000 rpm, 5 minutes, room
482 temperature). The cell pellets were resuspended in a transformation mix and incubated in a
483 water bath (40 minutes, 42 °C). The transformation mixed contained: 7.2 mL 50% PEG-3350
484 (Sigma), 1.08 mL 1 M LiAc (Sigma), 300 µL 10 mg/mL freshly denatured carrier DNA

485 (Sigma), 30 μg cloned library plasmid DNA and sterile water to 10.8 mL. Then, the cells were
486 harvested (3000 rpm, 5 minutes, room temperature) and the transformation mix was discarded.
487 The cell pellets were resuspended in 30 mL sterile water and 5 μL and 10 μL were plated on
488 SC-leucine-tryptophan medium (for the Y2H assay) or SC-uracil medium (for the DHFR-PCA)
489 and incubated for two days (30 °C) to evaluate the efficiency of the transformations. The
490 remaining cells were diluted in SC-leucine-tryptophan medium (for the Y2H assay) or SC-
491 uracil medium (for the DHFR-PCA) to an OD_{600nm} of 0.2 and incubated for two days (170 rpm,
492 30 °C). For each tile, for both the DHFR-PCA and Y2H assay, every variant was represented
493 at least 100 times. After two days, the transformation cultures reached saturation. For the Y2H
494 assay, 2×4.5 OD_{600nm} units were harvested (17,000 g, 1 minute, room temperature) in triplicates,
495 washed in sterile water and resuspended in 600 μL sterile water. Then, 300 μL was plated on
496 large BioAssay plates (500 cm^2) containing SC-leucine-tryptophan medium to serve as the
497 control condition. The remaining 300 μL was plated on large BioAssay plates (500 cm^2)
498 containing SC-leucine-tryptophan-histidine medium with 20 mM 3-amino-1,2,4-triazole (3-
499 AT) to serve as the selection condition. The plates were air-dried and incubated at 30 °C for
500 two days. For the DHFR-PCA, 2×22.5 OD_{600nm} units were harvested (17,000 g, 1 minute, room
501 temperature) in triplicates, washed in sterile water and resuspended in 600 μL sterile water.
502 Then, 300 μL was plated on large BioAssay plates (500 cm^2) containing
503 SC+histidine+leucine+methionine medium to serve as the control condition. The remaining
504 300 μL was plated on large BioAssay plates (500 cm^2) containing
505 SC+histidine+leucine+methionine medium with 200 μM MTX and 1 mM sulfanilamide to
506 serve as the selection condition. The plates were air-dried and incubated at 30 °C for three days.
507 Following the incubation, yeast cells were scraped off the plates using 30 mL sterile water and
508 9 OD_{600nm} units from each plate was harvested (17,000 g, 1 minute, room temperature) and

509 used for plasmid purification. Library plasmid DNA was isolated from yeast cells using the
510 ChargeSwitch Plasmid Yeast Mini kit (Invitrogen).
511
512 **Amplicon preparation and library sequencing**
513 Amplicons for each tile were prepared via two rounds of PCR, with the product cleaned-up
514 after each reaction. For the first PCR, 50 µL reactions were prepared by mixing 25 µL of 2×
515 Phusion High-Fidelity PCR Master Mix (New England Biolabs), 5 ng of purified library
516 plasmid DNA, 2.5 µL of 10 µM forward and reverse tile-specific primers with Illumina adapter
517 sequences, and nuclease-free water to reach a total volume of 50 µL. The primer sequences and
518 annealing temperatures are listed in the supplementary material (**Supplementary File 1**). The
519 reactions were subjected to the following PCR program: 98 °C for 30 seconds, followed by 8
520 cycles of 98 °C for 10 seconds, 57-64 °C for 30 seconds, 72 °C for 20 seconds, followed by 72
521 °C for 2 minutes and 4 °C hold. The product from the first PCR was cleaned-up using Ampure
522 XP beads (Beckman Coulter). Briefly, 15 µL of the PCR product were thoroughly mixed with
523 14.25 µL Ampure XP beads in 1.5 mL Eppendorf tubes and incubated at room temperature for
524 5 minutes. Following the incubation, the tubes were placed on a magnetic stand and the
525 supernatant was carefully removed before the beads were washed twice with 200 µL 70%
526 ethanol. The beads were air-dried at room temperature for 15 minutes to allow residual ethanol
527 to evaporate. Finally, the cleaned-up PCR product was eluted from the beads with nuclease-
528 free water and mixed thoroughly. The tubes were placed on a magnetic stand and the cleared
529 supernatant was transferred to a fresh 1.5 mL Eppendorf tube. Then, 50 µL reactions were
530 prepared for the second PCR by mixing 25 µL of 2× Phusion High-Fidelity PCR Master Mix
531 (New England Biolabs), 15 µL of cleaned-up PCR product from the first PCR and 10 µL
532 Nextera DNA CD Indexes (96-well format from Illumina) for dual indexing. The reactions
533 were subjected to the following PCR program: 98 °C for 30 seconds, followed by 12 cycles of

534 98 °C for 10 seconds, 62 °C for 30 seconds, 72 °C for 20 seconds, followed by 72 °C for 2
535 minutes and 4 °C hold. The product from the second PCR was cleaned-up as described for the
536 first PCR, using 11.25 µL Ampure XP beads. Finally, the concentration of the amplicons was
537 measured using the Qubit 2.0 Fluorometer (Invitrogen) with the Qubit dsDNA HS Assay kit
538 (Thermo Fisher Scientific) and the integrity was assessed using the 2100 Bioanalyzer (Agilent
539 Technologies) with the High Sensitivity DNA kit. The amplicons were volumetrically pooled
540 and diluted before they were subjected to pair-end next-generation sequencing using the
541 Illumina NextSeq 550 System with the Mid Output v2.5 300 cycle kit (Illumina).
542
543 **Analysis of sequencing data**
544 FASTQ files with read 1 and read 2 were downloaded from BaseSpace and sequences from the
545 four lanes were merged. The quality and length of sequences in the merged read 1 and read 2
546 FASTQ files were assessed using FastQC [79]. Then, Trimmomatic [80] paired-end mode was
547 used to obtain sequences of 151 base pairs using the following flags: MINLEN:151 and
548 CROP:151. The FASTQ files with trimmed sequences were used as input to Enrich2 [46] to
549 count reads and calculate a score for each variant. Enrich2 calculated log ratios and standard
550 errors for each variant based on three replicates using the control condition as time point 1 and
551 the selection condition as time point 2. The scores were calculated by normalization to
552 synonymous wild-type variants. Enrich2 was run using the overlap mode with default
553 parameters, but with a minimum and average quality of 1, maximum N's of 0, maximum
554 mismatches of 0, and 'remove unresolvable overlaps' and 'overlap only' set to TRUE. The start
555 and end were appropriately indicated for each tile. The wild-type sequence was included with
556 the correct offset for each tile. For the DHFR-PCA, the scores appeared clearly bimodal with
557 a wild-type-like peak centered around 0 and a peak of unstable variants with negative scores
558 (**Supplementary Fig. 4B**). However, the low-abundance peak displayed varying scores

559 between tiles. Hence, to put all tiles on a similar scale, the DHFR-PCA scores were rescaled
560 per tile by dividing all scores with the score of the low-abundance peak, giving scores ranging
561 from -1 to 1, with wild-type centered at 0. For the Y2H assay, the score distributions did not
562 appear bimodal, but instead all tiles had a sharp wild-type-like peak centered around 0 and few
563 loss- or gain-of-interaction variants, and therefore, we omitted rescaling the scores
564 (**Supplementary Fig. 4C**). The heatmaps were produced in R using the packages
565 ComplexHeatmap [81] and circlize [82].
566

567 **Modeling and structural analysis**

568 The C-terminal MutLα complex was predicted using AlphaFold2-Multimer. AlphaFold2-
569 Multimer was run using ColabFold [83]. The C-terminal domain of MLH1 (UniProt ID:
570 P40692) included residue 481-756 and the C-terminal domain of PMS2 (UniProt ID: P54278)
571 included residue 611-862. The N-terminal domains of MLH1 and PMS2 shown in **Fig. 1A** are
572 PDB: 4P7A [84] and 1H7S [85], respectively.
573 The surface hydrophobicity of the C-terminal domain of MLH1 was assessed using the YRB
574 script [58] in PyMOL with the AlphaFold2-predicted structure. We used the predicted structure
575 of the C-terminal domain of MLH1 due to the homomeric nature of the solved structure (PDB:
576 3RBN), which rendered it unsuitable for this analysis.
577 The full structure of the human MutLα complex was predicted using AlphaFold 3 [86]
578 https://alphafoldserver.com. We used the best-scoring model, which is available in the GitHub
579 repository for this project.
580

581 **Thermodynamic stability predictions**

582 Rosetta with the Cartesian ΔΔG protocol [47] (GitHub SHA1
583 99d33ec59ce9fcecc5e4f3800c778a54afdf8504) was used on the crystal structure of the C-

terminal domain of MLH1 (PDB: 3RBN) to predict the changes in thermodynamic stability of MLH1 variants. The MLH1 structures were prepared and relaxed and inputted for ΔΔG calculations using an in-house pipeline (https://github.com/KULL-Centre/PRISM/tree/main/software/rosetta_ddG_pipeline, v0.2.1). The change in folding stability was calculated by subtracting the wild-type ΔG value from the variant ΔG value. The ΔΔG values in Rosetta Energy Units were divided by 2.9 to convert them to kcal/mol [47].

Evolutionary conservation predictions

The evolutionary conservation ΔΔE score was calculated using GEMME [48]. The first isoform of MLH1 (UniProt: P40692-1) was input into HHblits [87] to build an MSA of MLH1 homologs using an E-value threshold of 10^{-10}. The resulting MSA was refined by first including only positions present in the wild-type MLH1 sequence and then by excluding sequences with more than 50% gaps. This process resulted in a final MSA containing 683 MLH1 homologs, which was input into GEMME to estimate the evolutionary distance of the MLH1 variants from the wild-type MLH1 sequence.

Acknowledgements

We acknowledge the use of the sequencing and computing core facilities at the Department of Biology, University of Copenhagen. We thank Vasileios Voutsinos for assistance with Illumina sequencing and Matteo Cagiada for running Rosetta and GEMME predictions. pDEST-DHFR-PCA was a gift from Professor Rasmus Hartmann-Petersen (University of Copenhagen).

Declaration of interests

The authors declare no competing interests.

Supplemental files

This article includes the following supplemental information.

- Supplemental figures (SupplementalFigures.pdf).
- Supplemental dataset (SupplementalFile1.xlsx).

Data and code availability

All data generated are included in the figures and supplemental files. Sequencing reads are available at the NCBI Gene Expression Omnibus (GEO) repository (accession number: GSE273652). The data and code are available at GitHub (https://github.com/KULL-Centre/_2024_Larsen-Ledet_MLH1).

Author contributions

Conceptualization: SLL, AS

Formal Analysis: SLL, AP

Funding Acquisition: AS

Investigation: SLL

Visualization: SLL

Writing – Original Draft Preparation: SLL

Writing – Review & Editing: SLL, AS

Funding

The present work was supported by grants from the Lundbeck Foundation (R272-2017–452) and the Novo Nordisk Foundation challenge program PRISM (NNF18OC0033950). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

References

1. Anfinsen CB, Haber E, Sela M, White FH. THE KINETICS OF FORMATION OF NATIVE RIBONUCLEASE DURING OXIDATION OF THE REDUCED POLYPEPTIDE CHAIN. Proceedings of the National Academy of Sciences. 1961 Sep;47(9):1309–14.
2. Stein A, Fowler DM, Hartmann-Petersen R, Lindorff-Larsen K. Biophysical and Mechanistic Models for Disease-Causing Protein Variants. Trends Biochem Sci. 2019 Jul;44(7):575–88.
3. Høie MH, Cagiada M, Beck Frederiksen AH, Stein A, Lindorff-Larsen K. Predicting and interpreting large-scale mutagenesis data using analyses of protein stability and conservation. Cell Rep. 2022 Jan;38(2):110207.
4. Hernández-Ramírez LC, Martucci F, Morgan RML, Trivellin G, Tilley D, Ramos-Guajardo N, et al. Rapid Proteasomal Degradation of Mutant Proteins Is the Primary Mechanism Leading to Tumorigenesis in Patients With Missense *AIP* Mutations. J Clin Endocrinol Metab. 2016 Aug;101(8):3144–54.
5. Taipale M. Disruption of protein function by pathogenic mutations: common and uncommon mechanisms. Biochemistry and Cell Biology. 2019 Feb;97(1):46–57.
6. Blaabjerg LM, Kassem MM, Good LL, Jonsson N, Cagiada M, Johansson KE, et al. Rapid protein stability prediction using deep learning representations. Elife. 2023 May 15;12.
7. Scheller R, Stein A, Nielsen S V., Marin FI, Gerdes A, Marco M, et al. Toward mechanistic models for genotype–phenotype correlations in phenylketonuria using protein stability calculations. Hum Mutat. 2019 Apr 25;40(4):444–57.
8. Yue P, Li Z, Moult J. Loss of Protein Structure Stability as a Major Causative Factor in Monogenic Disease. J Mol Biol. 2005 Oct;353(2):459–73.
9. Cagiada M, Johansson KE, Valanciute A, Nielsen S V, Hartmann-Petersen R, Yang JJ, et al. Understanding the Origins of Loss of Protein Function by Analyzing the Effects of Thousands of Variants on Activity and Abundance. Mol Biol Evol. 2021 Jul 29;38(8):3235–46.
10. Stelzl U, Worm U, Lalowski M, Haenig C, Brembeck FH, Goehler H, et al. A Human Protein-Protein Interaction Network: A Resource for Annotating the Proteome. Cell. 2005 Sep;122(6):957–68.
11. Burke DF, Bryant P, Barrio-Hernandez I, Memon D, Pozzati G, Shenoy A, et al. Towards a structurally resolved human protein interaction network. Nat Struct Mol Biol. 2023 Feb 23;30(2):216–25.
12. Luck K, Kim DK, Lambourne L, Spirohn K, Begg BE, Bian W, et al. A reference map of the human binary protein interactome. Nature. 2020 Apr 16;580(7803):402–8.
13. Schreiber G. Protein–Protein Interaction Interfaces and their Functional Implications. In: Protein – Protein Interaction Regulators. The Royal Society of Chemistry; 2020. p. 1–24.
14. Mintseris J, Weng Z. Structure, function, and evolution of transient and obligate protein–protein interactions. Proceedings of the National Academy of Sciences. 2005 Aug 2;102(31):10930–5.
15. Perkins JR, Diboun I, Dessailly BH, Lees JG, Orengo C. Transient Protein-Protein Interactions: Structural, Functional, and Network Properties. Structure. 2010 Oct;18(10):1233–43.
16. Ivarsson Y, Jemth P. Affinity and specificity of motif-based protein–protein interactions. Curr Opin Struct Biol. 2019 Feb;54:26–33.
17. David A, Razali R, Wass MN, Sternberg MJE. Protein-protein interaction sites are hot spots for disease-associated nonsynonymous SNPs. Hum Mutat. 2012 Feb;33(2):359–63.

18. David A, Sternberg MJE. The Contribution of Missense Mutations in Core and Rim Residues of Protein–Protein Interfaces to Human Disease. J Mol Biol. 2015 Aug;427(17):2886–98.
19. Xiong D, Lee D, Li L, Zhao Q, Yu H. Implications of disease-related mutations at protein–protein interfaces. Curr Opin Struct Biol. 2022 Feb;72:219–25.
20. Cheng F, Zhao J, Wang Y, Lu W, Liu Z, Zhou Y, et al. Comprehensive characterization of protein–protein interactions perturbed by disease mutations. Nat Genet. 2021 Mar 8;53(3):342–53.
21. Sahni N, Yi S, Taipale M, Fuxman Bass JI, Coulombe-Huntington J, Yang F, et al. Widespread Macromolecular Interaction Perturbations in Human Genetic Disorders. Cell. 2015 Apr;161(3):647–60.
22. Zhong Q, Simonis N, Li Q, Charloteaux B, Heuze F, Klitgord N, et al. Edgetic perturbation models of human inherited disorders. Mol Syst Biol. 2009 Jan 3;5(1).
23. Sahni N, Yi S, Zhong Q, Jailkhani N, Charloteaux B, Cusick ME, et al. Edgotype: a fundamental link between genotype and phenotype. Curr Opin Genet Dev. 2013 Dec;23(6):649–57.
24. Mitchell AL, Attwood TK, Babbitt PC, Blum M, Bork P, Bridge A, et al. InterPro in 2019: improving coverage, classification and access to protein sequence annotations. Nucleic Acids Res. 2019 Jan 8;47(D1):D351–60.
25. Guerrette S, Acharya S, Fishel R. The Interaction of the Human MutL Homologues in Hereditary Nonpolyposis Colon Cancer. Journal of Biological Chemistry. 1999 Mar;274(10):6336–41.
26. Pećina-Šlaus N, Kafka A, Salamon I, Bukovac A. Mismatch Repair Pathway, Genome Stability and Cancer. Front Mol Biosci. 2020 Jun 26;7.
27. Jun S, Kim TG, Ban C. DNA mismatch repair system. FEBS J. 2006 Apr 31;273(8):1609–19.
28. Momma T, Gonda K, Akama Y, Endo E, Ujiie D, Fujita S, et al. MLH1 germline mutation associated with Lynch syndrome in a family followed for more than 45 years. BMC Med Genet. 2019 Dec 2;20(1):67.
29. Bronner CE, Baker SM, Morrison PT, Warren G, Smith LG, Lescoe MK, et al. Mutation in the DNA mismatch repair gene homologue hMLH 1 is associated with hereditary non-polyposis colon cancer. Nature. 1994 Mar;368(6468):258–61.
30. Plazzer JP, Sijmons RH, Woods MO, Peltomäki P, Thompson B, Den Dunnen JT, et al. The InSiGHT database: utilizing 100 years of insights into Lynch Syndrome. Fam Cancer. 2013 Jun 27;12(2):175–80.
31. Abildgaard AB, Stein A, Nielsen S V, Schultz-Knudsen K, Papaleo E, Shrikhande A, et al. Computational and cellular studies reveal structural destabilization and degradation of MLH1 variants in Lynch syndrome. Elife. 2019 Nov 7;8.
32. Abildgaard AB, Nielsen S V., Bernstein I, Stein A, Lindorff-Larsen K, Hartmann-Petersen R. Lynch syndrome, molecular mechanisms and variant classification. Br J Cancer. 2023 Mar 23;128(5):726–34.
33. Peltomäki P. Update on Lynch syndrome genomics. Fam Cancer. 2016 Jul 12;15(3):385–93.
34. Pérez-Palma E, Gramm M, Nürnberg P, May P, Lal D. Simple ClinVar: an interactive web server to explore and retrieve gene and disease variants aggregated in ClinVar database. Nucleic Acids Res. 2019 Jul 2;47(W1):W99–105.
35. Gueneau E, Dherin C, Legrand P, Tellier-Lebegue C, Gilquin B, Bonnesoeur P, et al. Structure of the MutLα C-terminal domain reveals how Mlh1 contributes to Pms1 endonuclease site. Nat Struct Mol Biol. 2013 Apr 24;20(4):461–8.
36. Remy I, Michnick SW. Clonal selection and *in vivo* quantitation of protein interactions with protein-fragment complementation assays. Proceedings of the National Academy of Sciences. 1999 May 11;96(10):5394–9.

37. Tarassov K, Messier V, Landry CR, Radinovic S, Molina MMS, Shames I, et al. An in Vivo Map of the Yeast Protein Interactome. Science (1979). 2008 Jun 13;320(5882):1465–70.
38. Levy ED, Kowarzyk J, Michnick SW. High-Resolution Mapping of Protein Concentration Reveals Principles of Proteome Architecture and Adaptation. Cell Rep. 2014 May;7(4):1333–40.
39. Faure AJ, Domingo J, Schmiedel JM, Hidalgo-Carcedo C, Diss G, Lehner B. Mapping the energetic and allosteric landscapes of protein binding domains. Nature. 2022 Apr 7;604(7904):175–83.
40. Fields S, Song O kyu. A novel genetic system to detect protein–protein interactions. Nature. 1989 Jul;340(6230):245–6.
41. Landrum MJ, Lee JM, Benson M, Brown GR, Chao C, Chitipiralla S, et al. ClinVar: improving access to variant interpretations and supporting evidence. Nucleic Acids Res. 2018 Jan 4;46(D1):D1062–7.
42. Karczewski KJ, Francioli LC, Tiao G, Cummings BB, Alföldi J, Wang Q, et al. The mutational constraint spectrum quantified from variation in 141,456 humans. Nature. 2020 May 28;581(7809):434–43.
43. Takahashi M, Shimodaira H, Andreutti-Zaugg C, Iggo R, Kolodner RD, Ishioka C. Functional Analysis of Human *MLH1* Variants Using Yeast and *In vitro* Mismatch Repair Assays. Cancer Res. 2007 May 15;67(10):4595–604.
44. Kosinski J, Hinrichsen I, Bujnicki JM, Friedhoff P, Plotz G. Identification of Lynch syndrome mutations in the MLH1-PMS2 interface that disturb dimerization and mismatch repair. Hum Mutat. 2010 May 17;31(8):975–82.
45. Rath A, Radecki AA, Rahman K, Gilmore RB, Hudson JR, Cenci M, et al. A calibrated cell-based functional assay to aid classification of *MLH1* DNA mismatch repair gene variants. Hum Mutat. 2022 Dec 12;43(12):2295–307.
46. Rubin AF, Gelman H, Lucas N, Bajjalieh SM, Papenfuss AT, Speed TP, et al. A statistical framework for analyzing deep mutational scanning data. Genome Biol. 2017 Dec 7;18(1):150.
47. Park H, Bradley P, Greisen P, Liu Y, Mulligan VK, Kim DE, et al. Simultaneous Optimization of Biomolecular Energy Functions on Features from Small Molecules and Macromolecules. J Chem Theory Comput. 2016 Dec 13;12(12):6201–12.
48. Laine E, Karami Y, Carbone A. GEMME: A Simple and Fast Global Epistatic Model Predicting Mutational Effects. Mol Biol Evol. 2019 Nov 1;36(11):2604–19.
49. Kansikas M, Kariola R, Nyström M. Verification of the three-step model in assessing the pathogenicity of mismatch repair gene variants. Hum Mutat. 2011 Jan;32(1):107–15.
50. Dominguez-Valentin M, Plazzer JP, Sampson JR, Engel C, Aretz S, Jenkins MA, et al. No Difference in Penetrance between Truncating and Missense/Aberrant Splicing Pathogenic Variants in MLH1 and MSH2: A Prospective Lynch Syndrome Database Study. J Clin Med. 2021 Jun 28;10(13):2856.
51. Drost M, Tiersma Y, Thompson BA, Frederiksen JH, Keijzers G, Glubb D, et al. A functional assay–based procedure to classify mismatch repair gene variants in Lynch syndrome. Genetics in Medicine. 2019 Jul;21(7):1486–96.
52. Cravo M. Pathogenicity of missense and splice site mutations in hMSH2 and hMLH1 mismatch repair genes: implications for genetic testing. Gut. 2002 Mar 1;50(3):405–12.
53. Singh P, Fragoza R, Blengini CS, Tran TN, Pannafino G, Al-Sweel N, et al. Human MLH1/3 variants causing aneuploidy, pregnancy loss, and premature reproductive aging. Nat Commun. 2021 Aug 18;12(1):5005.
54. Clausen L, Voutsinos V, Cagiada M, Johansson KE, Grønbæk-Thygesen M, Nariya S, et al. A mutational atlas for Parkin proteostasis. Nat Commun. 2024 Feb 20;15(1):1541.

55. Grønbæk-Thygesen M, Voutsinos V, Johansson KE, Schulze TK, Cagiada M, Pedersen L, et al. Deep mutational scanning reveals a correlation between degradation and toxicity of thousands of aspartoacylase variants. Nat Commun. 2024 May 13;15(1):4026.
56. Davey NE, Van Roey K, Weatheritt RJ, Toedt G, Uyar B, Altenberg B, et al. Attributes of short linear motifs. Mol BioSyst. 2012;8(1):268–81.
57. Teilum K, Olsen JG, Kragelund BB. On the specificity of protein–protein interactions in the context of disorder. Biochemical Journal. 2021 Jun 11;478(11):2035–50.
58. Hagemans D, van Belzen IAEM, Morán Luengo T, Rüdiger SGD. A script to highlight hydrophobicity and charge on protein surfaces. Front Mol Biosci. 2015 Oct 13;2.
59. Leong V, Lorenowicz J, Kozij N, Guarné A. Nuclear import of human MLH1, PMS2, and MutLα: Redundancy is the key. Mol Carcinog. 2009 Aug 15;48(8):742–50.
60. de Barros AC, Takeda AAS, Dreyer TR, Velazquez-Campoy A, Kobe B, Fontes MRM. DNA mismatch repair proteins MLH1 and PMS2 can be imported to the nucleus by a classical nuclear import pathway. Biochimie. 2018 Mar;146:87–96.
61. Brieger A, Plotz G, Raedle J, Weber N, Baum W, Caspary WF, et al. Characterization of the nuclear import of human MutLα. Mol Carcinog. 2005 May 7;43(1):51–8.
62. Wu X, Platt JL, Cascalho M. Dimerization of MLH1 and PMS2 Limits Nuclear Localization of MutLα. Mol Cell Biol. 2003 May 1;23(9):3320–8.
63. Hinrichsen I, Brieger A, Trojan J, Zeuzem S, Nilbert M, Plotz G. Expression Defect Size among Unclassified *MLH1* Variants Determines Pathogenicity in Lynch Syndrome Diagnosis. Clinical Cancer Research. 2013 May 1;19(9):2432–41.
64. Mashahreh B, Armony S, Johansson KE, Chappleboim A, Friedman N, Gardner RG, et al. Conserved degronome features governing quality control associated proteolysis. Nat Commun. 2022 Dec 8;13(1):7588.
65. Abildgaard AB, Voutsinos V, Petersen SD, Larsen FB, Kampmeyer C, Johansson KE, et al. HSP70-binding motifs function as protein quality control degrons. Cellular and Molecular Life Sciences. 2023 Jan 7;80(1):32.
66. Johansson KE, Mashahreh B, Hartmann-Petersen R, Ravid T, Lindorff-Larsen K. Prediction of Quality-control Degradation Signals in Yeast Proteins. J Mol Biol. 2023 Jan;435(2):167915.
67. Timms RT, Mena EL, Leng Y, Li MZ, Tchasovnikarova IA, Koren I, et al. Defining E3 ligase–substrate relationships through multiplex CRISPR screening. Nat Cell Biol. 2023 Oct 21;25(10):1535–45.
68. Kong KYE, Shankar S, Rühle F, Khmelinskii A. Orphan quality control by an SCF ubiquitin ligase directed to pervasive C-degrons. Nat Commun. 2023 Dec 15;14(1):8363.
69. Koren I, Timms RT, Kula T, Xu Q, Li MZ, Elledge SJ. The Eukaryotic Proteome Is Shaped by E3 Ubiquitin Ligases Targeting C-Terminal Degrons. Cell. 2018 Jun;173(7):1622-1635.e14.
70. Timms RT, Koren I. Tying up loose ends: the N-degron and C-degron pathways of protein degradation. Biochem Soc Trans. 2020 Aug 28;48(4):1557–67.
71. Reyes GX, Schmidt TT, Kolodner RD, Hombauer H. New insights into the mechanism of DNA mismatch repair. Chromosoma. 2015 Dec 11;124(4):443–62.
72. Wu T, Yang S, Fang J, Ye Y, Zhang Y, Gao J, et al. *MutL homolog 1* participates in interference-sensitive meiotic crossover formation in soybean. Plant Physiol. 2024 Mar 16;
73. Smith CE, Mendillo ML, Bowen N, Hombauer H, Campbell CS, Desai A, et al. Dominant Mutations in S. cerevisiae PMS1 Identify the Mlh1-Pms1 Endonuclease Active Site and an Exonuclease 1-Independent Mismatch Repair Pathway. PLoS Genet. 2013 Oct 31;9(10):e1003869.
74. Putnam CD, Kolodner RD. Insights into DNA cleavage by MutL homologs from analysis of conserved motifs in eukaryotic Mlh1. BioEssays. 2023 Jul 9;

75. Dherin C, Gueneau E, Francin M, Nunez M, Miron S, Liberti SE, et al. Characterization of a Highly Conserved Binding Site of Mlh1 Required for Exonuclease I-Dependent Mismatch Repair. Mol Cell Biol. 2009 Feb 1;29(3):907–18.
76. Perica T, Mathy CJP, Xu J, Jang GM, Zhang Y, Kaake R, et al. Systems-level effects of allosteric perturbations to a model molecular switch. Nature. 2021 Nov 4;599(7883):152–7.
77. Gietz RD, Schiestl RH. High-efficiency yeast transformation using the LiAc/SS carrier DNA/PEG method. Nat Protoc. 2007 Jan 31;2(1):31–4.
78. Gietz RD, Schiestl RH. Large-scale high-efficiency yeast transformation using the LiAc/SS carrier DNA/PEG method. Nat Protoc. 2007 Jan 31;2(1):38–41.
79. Andrews S. FastQC: A Quality Control Tool for High Throughput Sequence Data [Online]. Available online at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc/. 2010;
80. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. 2014 Aug 1;30(15):2114–20.
81. Gu Z, Eils R, Schlesner M. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. Bioinformatics. 2016 Sep 15;32(18):2847–9.
82. Gu Z, Gu L, Eils R, Schlesner M, Brors B. *circlize* implements and enhances circular visualization in R. Bioinformatics. 2014 Oct 1;30(19):2811–2.
83. Mirdita M, Schütze K, Moriwaki Y, Heo L, Ovchinnikov S, Steinegger M. ColabFold: making protein folding accessible to all. Nat Methods. 2022 Jun 30;19(6):679–82.
84. Wu H, Zeng H, Lam R, Tempel W, Kerr ID, Min J. Structure of the human MLH1 N-terminus: implications for predisposition to Lynch syndrome. Acta Crystallogr F Struct Biol Commun. 2015 Aug 1;71(8):981–5.
85. Guarne A. Structure and function of the N-terminal 40 kDa fragment of human PMS2: a monomeric GHL ATPase. EMBO J. 2001 Oct 1;20(19):5521–31.
86. Abramson J, Adler J, Dunger J, Evans R, Green T, Pritzel A, et al. Accurate structure prediction of biomolecular interactions with AlphaFold 3. Nature. 2024 Jun 13;630(8016):493–500.
87. Remmert M, Biegert A, Hauser A, Söding J. HHblits: lightning-fast iterative protein sequence searching by HMM-HMM alignment. Nat Methods. 2012 Feb 25;9(2):173–5.

863 **Legends for Supporting Information**

864

865 **Fig. S1** – *AlphaFold2-Multimer prediction of the C-terminal MutLα complex.* High confidence
866 model of the human C-terminal MutLα complex predicted by AlphaFold2-Multimer. The
867 structure is colored by the pLDDT score using the reversed rainbow color palette in PyMOL.

868

869 **Fig. S2** – *Vector map of pDEST-DHFR-PCA.* Vector map of the CEN-based expression vector
870 pDEST-DHFR-PCA with a *URA3* marker. The mDHFR[F3]-MLH1 fusion protein and
871 mDHFR[F1,2] were expressed from the CYC1 and GAP promoter, respectively. The vector
872 map was generated using SnapGene®.

873

874 **Fig. S3** – *Vector maps of pDEST22 and pDEST32.* Vector maps of the CEN-based expression
875 vectors pDEST22 and pDEST32 with a *TRP1* and *LEU2* marker, respectively. MLH1 was fused
876 to the activation domain and PMS2 was fused to the DNA binding domain. The vector maps
877 were generated using SnapGene®.

878

879 **Fig. S4** – *Distributions of abundance and interaction scores per tile.* (A) The predicted
880 structure of the C-terminal MutLα complex with the C-terminal domain (CTD) of MLH1
881 colored by the seven tiles. The tiles are colored according to the legend in (C). The C-terminal
882 domain of PMS2 is shown in orange. (B) Density plot showing the distribution of abundance
883 scores for each tile in the C-terminal domain of MLH1 prior to rescaling. The tiles are colored
884 according to the legend in (C). (C) Same as in (B), but for interaction scores.

885

886 **Fig. S5** – *Score correlation between replicates*. Scatter plots showing the score correlation
887 between the three replicates for DHFR-PCA (abundance) and Y2H (interaction). The Pearson
888 correlation coefficient (R) is shown for each plot.

889

890 **Fig. S6** – *Heatmaps of the standard error (SE) of abundance and interaction scores.* (A)
891 Heatmaps of the SE for the abundance scores, calculated using Enrich2. The SE has been
892 rescaled in accordance with the rescaling of the abundance scores. The SE range from 0 (white)
893 to 0.45 (dark green). The wild-type residue at each position is marked in yellow, while missing
894 variants are marked in gray. The position in the full-length MLH1 protein is shown below the
895 heatmap. A linear representation of the secondary structure of MLH1 is displayed above the

heatmap, colored by AlphaFold2 pLDDT as a measure of disorder. (B) Same as in (A), but for interaction scores. Note, the color scale is different from in (A).

Fig. S7 – *Median score correlation with relative solvent accessible surface area (rSASA).* (A) Scatter plot showing the rSASA plotted against the median abundance score per position. Residues are colored by the secondary structural element according to the legend in (B). A rSASA value of 0 corresponds to a completely buried residue, while 1 corresponds to a completely exposed residue. The Spearman's R correlation coefficient is shown. (B) Same as in (A), but for median interaction scores.

Fig. S8 – *Validation of loss- and gain-of-interaction variants.* (A) Scatter density plot showing the interaction score plotted against the abundance score. The color scale is indicated in the legend. Selected loss-of-interaction and gain-of-interaction variants are highlighted in red. (B) The structure of the predicted C-terminal MutLα complex with the selected MLH1 variants highlighted. (C) Y2H growth assays comparing the growth of the vector, wild-type MLH1 and the indicated MLH1 variant. Cells were grown on medium with histidine and without histidine with 20 mM (the concentration used in the high-throughput screen), 40 mM or 60 mM 3-AT.

Fig. S9 – *Interfaces of the MutLα complex.* The AlphaFold2-Multimer predicted structure of the human C-terminal MutLα complex shows the interaction between MLH1 (purple) and PMS2 (orange). The canonical interface known from the yeast complex and the novel interaction site mediated via the distal central α-helix are indicated by black arrows.

Fig. S10 – *Prediction of a novel interaction site with PMS2.* (A) The full structure of the human MutLα complex as predicted by AlphaFold 3. As in other figures, MLH1 is colored purple, and PMS2 is colored orange. The intrinsically disordered linker regions between the N- and C-terminal domains are colored magenta and yellow, respectively. The N-terminal domains and linker regions are displayed as transparent. AlphaFold3 predicts an α-helix within the disordered linker of PMS2, spanning residues 409-431 (yellow), which binds to the distal central α-helix in MLH1. (B) A zoomed-in view of the interaction between the α-helix in PMS2 (yellow) and the central α-helix in MLH1 (purple). Residues facing MLH1 in the PMS2 α-helix are displayed as sticks, with nitrogen atoms in sidechains highlighted in blue. (C) The same view as in (B), with MLH1 colored by the per-position median interaction score. The color scale corresponds to that shown in Fig. 3C. Contact residues in MLH1 within 6 Å of the

PMS2 α-helix are shown as sticks. Notably, many of these residues locate to positions where our experiments showed both loss- and gain-of-interaction variants. For visual clarity, the coloring of nitrogen and oxygen atoms in the sidechains has been omitted.

Fig. S11 – *Correlation of Rosetta and GEMME predictions with experimental scores.* (A) Scatter density plot comparing the predicted Rosetta ΔΔG scores and the abundance scores. The color scale is indicated in the legend. The Spearman's R correlation coefficient is shown. (B) Same as in (A), but for predicted GEMME ΔΔE scores. (C) Scatter density plot comparing the predicted Rosetta ΔΔG scores and the interaction scores. The color scale is indicated in the legend. The Spearman's R correlation coefficient is shown. (D) Same as in (C), but for predicted GEMME ΔΔE scores.

Fig. S12 – *Predicting abundance and interaction scores using Rosetta and GEMME scores.* ROC curves to assess how well Rosetta and GEMME perform in predicting the abundance and interaction scores. A cutoff of -0.5 was applied to the abundance and interaction scores. Variants with a score below -0.5 were categorized as detrimental, while those with a score above -0.5 were categorized as wild-type-like. The AUC for each predictor is reported in the legend. The dashed diagonal line denotes performance of a random classifier. Note that variants with scores above -0.5 include both wild-type-like variants and variants with increased abundance or PMS2 interaction.

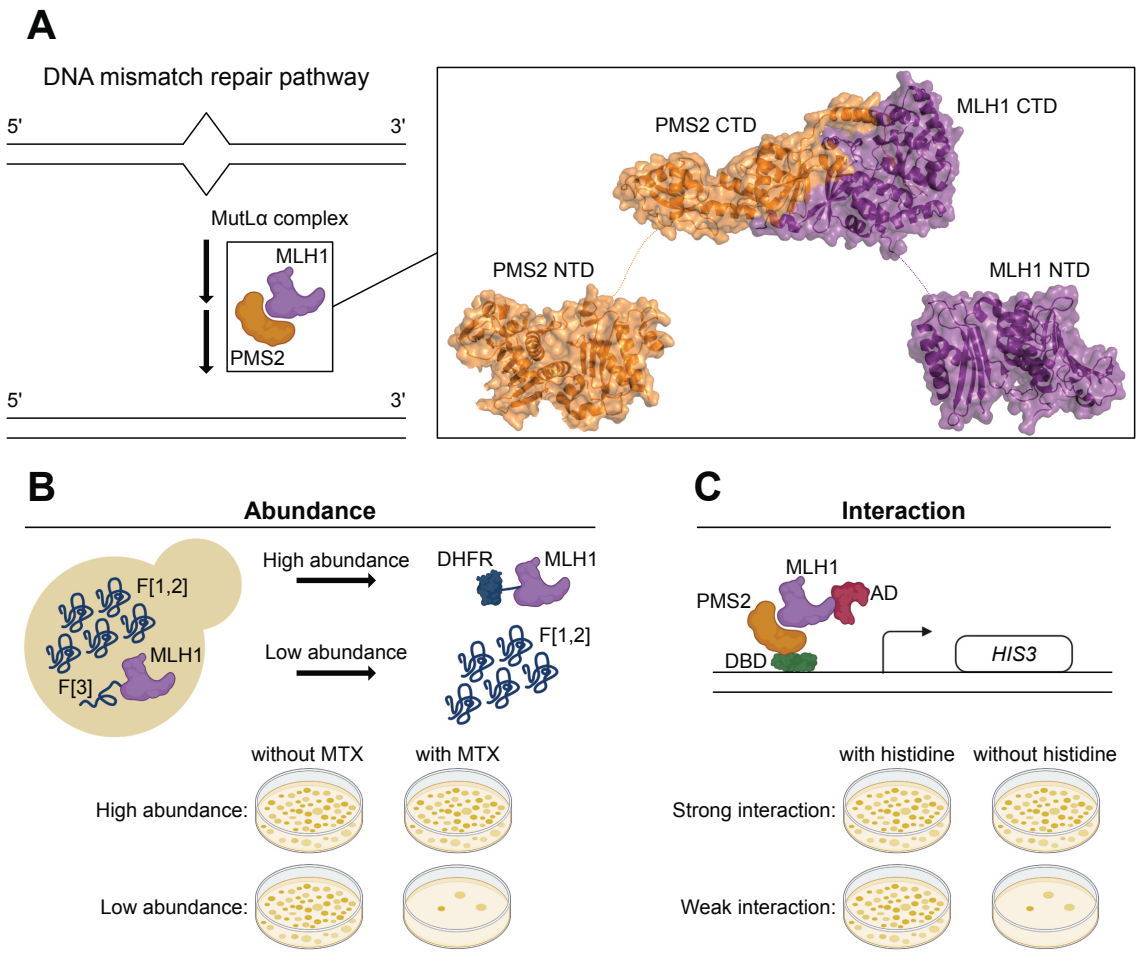
Fig. S13 – *Separation of interaction scores using GEMME predictions and abundance scores.* Scatter plots showing the correlation between the predicted GEMME ΔΔE score and abundance scores. The MLH1 variants are colored by the interaction score and the color scale corresponds to that shown in Fig. 3A.

Fig. S14 – *Classification of MLH1 pathogenicity using experimental and computational scores*. ROC curves to assess how well our experimental abundance and interaction scores, as well as computational Rosetta and GEMME predictions perform in separating pathogenic (and likely pathogenic) from benign (and likely benign) MLH1 variants. The AUC for each predictor is reported in the legend. The dashed diagonal line denotes a random classifier.

Fig. S15 – *Hydrophobicity of the canonical PMS2 interface.* (A) Cartoon representation of the C-terminal domain of MLH1 predicted by AlphaFold2. Surface residues are colored according

964 to their chemical properties using the YRB script (see Materials and Methods). Carbon atoms
965 not bound to oxygen or nitrogen atoms are colored yellow (hydrophobic amino acids),
966 negatively charged oxygen atoms are colored red (negatively charged amino acids), positively
967 charged nitrogen atoms are colored blue (positively charged amino acids) and all other atoms
968 are colored gray. The four β-strands in the canonical interface with PMS2 are enriched with
969 hydrophobic amino acids (yellow) and depleted in charged residues (red and blue). (B) Same
970 as in (A), but a different view of the structure highlighting the four β-strands at the bottom
971 front.
972
973 **Fig. S16** – *Dynamic range and sensitivity of the DHFR-PCA and Y2H assay.* (A) Scatter density
974 plot comparing the interaction scores and the abundance scores. The color scale is indicated in
975 the legend. Selected MLH1 variants that displayed reduced abundance but wild-type-like
976 interaction with PMS2 are highlighted in red. (B) Yeast growth assays comparing the growth
977 of a vector, wild-type or the indicated MLH1 variant in the DHFR-PCA (top panel) and the
978 Y2H assay (bottom panel). In the DHFR-PCA, cells were grown on medium with or without
979 MTX. In the Y2H assay, cells were grown on medium with or without histidine.
980
981 **Fig. S17** – *Benchmarking abundance and interaction scores against existing literature.* (A)
982 Comparison of our experimental abundance scores with protein expression levels reported in
983 [63]. In their study, HEK293T cells were transiently transfected with an MLH1 variant, and
984 cellular extracts were analyzed using SDS-PAGE and immunoblotting with an anti-MLH1
985 antibody. Statistical analysis of MLH1 expression levels relative to WT was performed using
986 a t-test, with p-values calculated for each variant. Variants shown in red indicate $p < 0.05$. As
987 expected, most of the variants with significantly reduced expression levels were also measured
988 to have reduced abundance scores in our work. (B) Comparison of our experimental abundance
989 scores with protein expression levels reported in [31]. In their study, HCT116 cells were
990 transiently transfected with an MLH1 variant and analyzed using immunofluorescence
991 microscopy with an anti-MLH1 antibody. Expression levels were quantified and reported as a
992 percentage relative to WT. As expected, most of the variants with reduced expression levels
993 were also measured to have reduced abundance scores in our work. (C) Comparison of our
994 experimental interaction scores with PMS2 interaction data reported in [44]. In their study,
995 HEK293T cells were transiently co-transfected with an MLH1 variant and PMS2, and cellular
996 extracts were analyzed using SDS-PAGE and immunoblotting with anti-MLH1 and anti-PMS2
997 antibodies. Based on PMS2 expression levels, MLH1 variants were classified as either

interacting or not interacting with PMS2. Variants with expression levels too low to measure PMS2 expression were denoted as NA. As expected, most MLH1 variants that interacted with PMS2 in their study also exhibited positive interaction scores or scores close to 0 in our work. Additionally, variants reported in [44] to have dramatically reduced expression levels also displayed reduced interaction scores in our work.

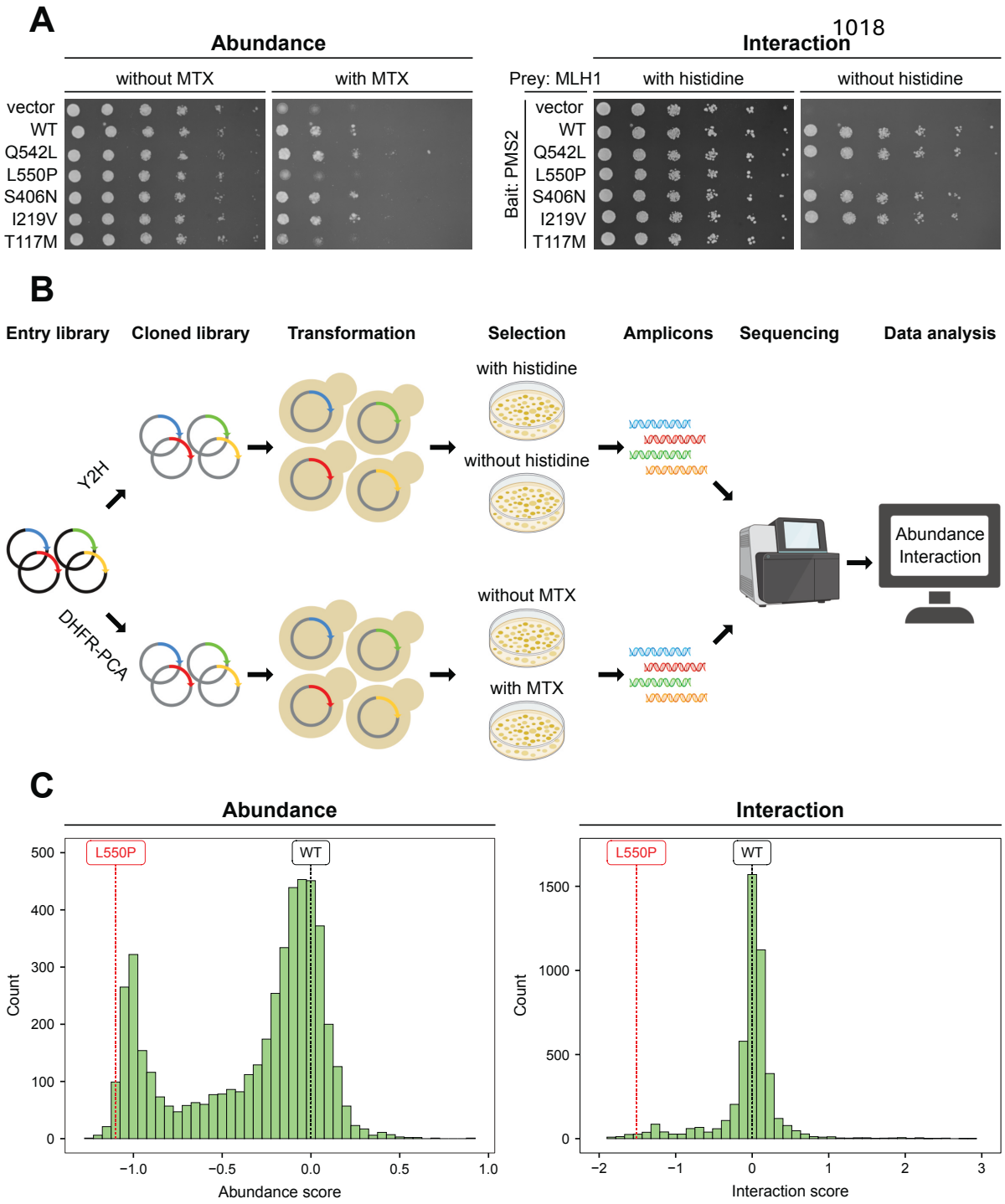

A
DNA mismatch repair pathway
5'
3'
MutLα complex
MLH1
PMS2
5'
3'
PMS2 CTD
MLH1 CTD
PMS2 NTD
MLH1 NTD
B
Abundance
F[1,2]
MLH1
F[3]
High abundance
DHFR
MLH1
Low abundance
F[1,2]
without MTX
with MTX
High abundance:
Low abundance:
C
Interaction
MLH1
PMS2
AD
DBD
HIS3
with histidine
without histidine
Strong interaction:
Weak interaction:

1003

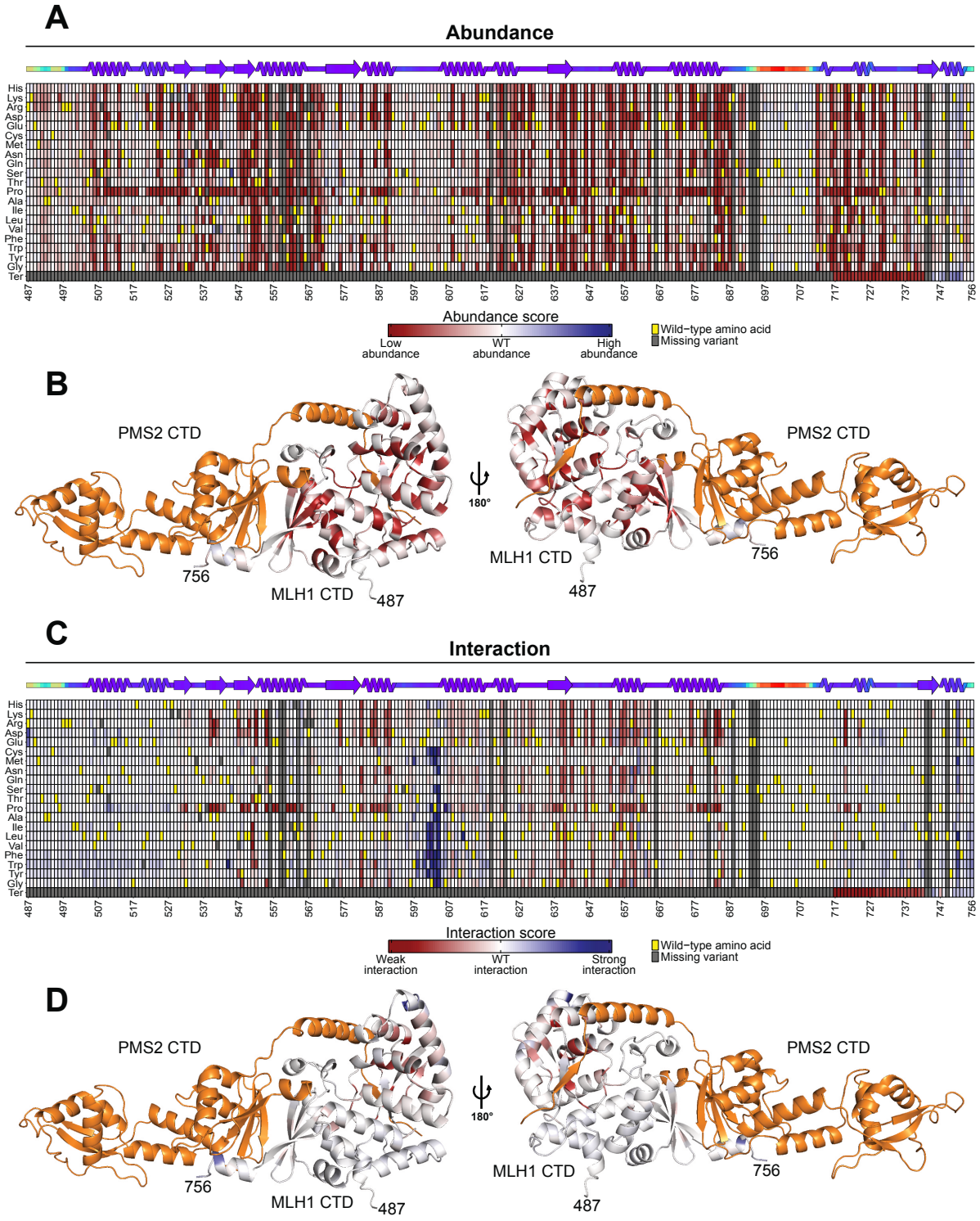
1004 **Fig. 1** – *Yeast-based abundance and interaction assays to measure variant effects in MLH1.*
1005 (A) The MutLα complex (MLH1:PMS2) participates in the DNA mismatch repair (MMR)
1006 pathway together with several other components to recognize and correct mismatches during
1007 DNA replication. MLH1 (purple) and PMS2 (orange) fold into an N-terminal domain (NTD)
1008 and a C-terminal domain (CTD), separated by a disordered linker. The two proteins interact
1009 through their C-terminal domains. The C-terminal MutLα complex was predicted using
1010 AlphaFold2-Multimer, whereas the N-terminal domain of MLH1 is PDB: 4P7A and the N-
1011 terminal domain of PMS2 is PDB: 1H7S. The disordered linkers between the domains were
1012 hand-drawn for illustrative purposes. (B) Schematic illustration of the dihydrofolate reductase
1013 protein-fragment complementation assay (DHFR-PCA) used to measure the cellular
1014 abundance of MLH1 variants. (C) Schematic illustration of the yeast two-hybrid (Y2H) assay
1015 used to measure the interaction of MLH1 variants with PMS2. Parts of Fig. 1 were created
1016 using BioRender.com.

1017

1018


A
Abundance
without MTX
with MTX
vector
WT
Q542L
L550P
S406N
I219V
T117M
Interaction
Prey: MLH1
with histidine
without histidine
Bait: PMS2
B
Entry library
Cloned library
Transformation
Selection
Amplicons
Sequencing
Data analysis
Y2H
DHFR-PCA
with histidine
without histidine
without MTX
with MTX
Abundance
Interaction
C
Abundance
Interaction
L550P
WT
Count
Abundance score
Interaction score

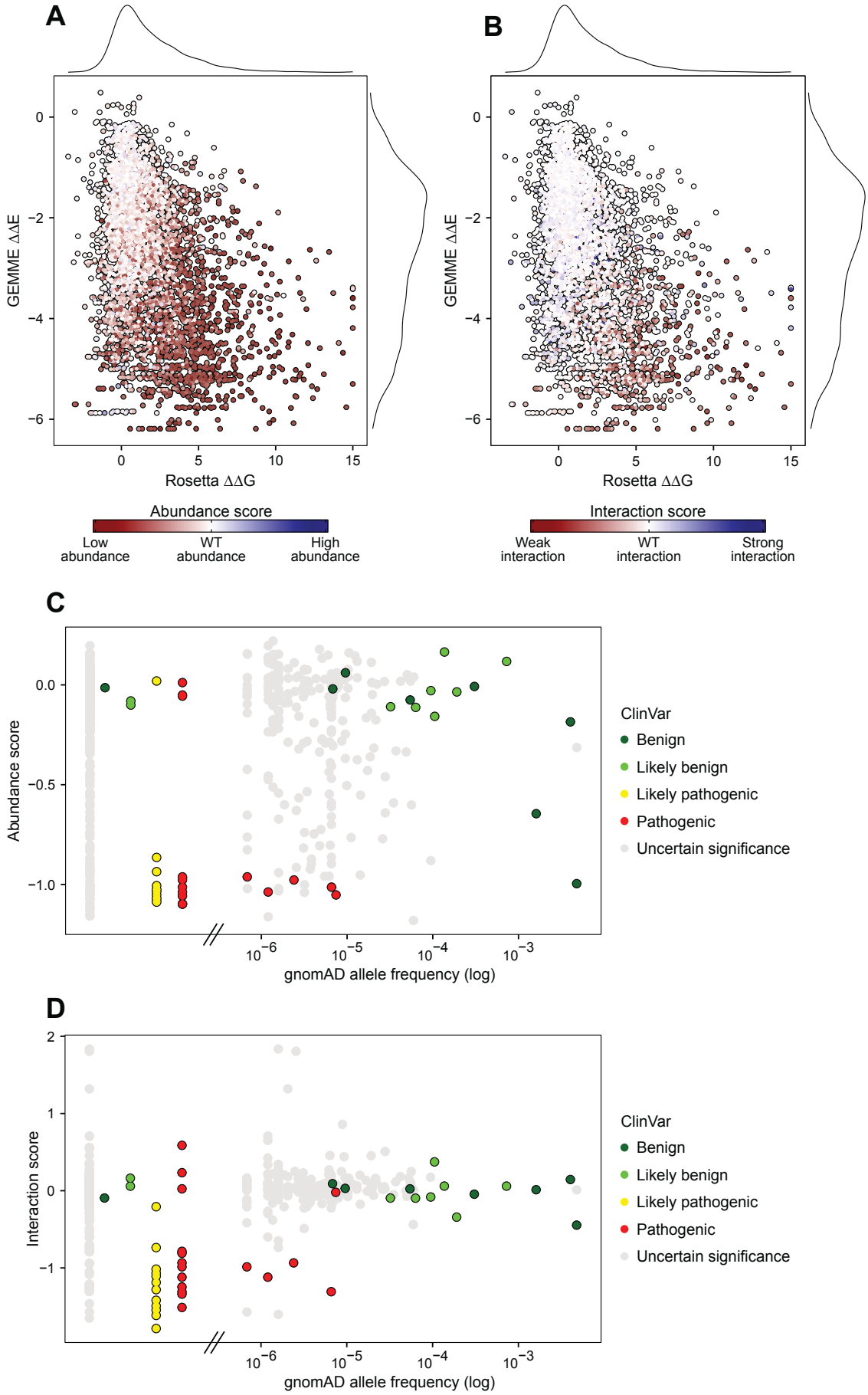
1019 **Fig. 2** – *Screening of site-saturation mutagenesis libraries in the C-terminal domain of MLH1.*
1020 (A) Small-scale yeast growth assays to test the DHFR-PCA (left) and the Y2H assay (right).
1021 Yeast cells expressed either a vector, wild-type MLH1, or the indicated MLH1 variant. For the
1022 DHFR-PCA the vector was an empty vector, while for the Y2H assay the vector was the
1023 activation domain (AD) expressed without MLH1. For the DHFR-PCA, cells were grown on
1024 medium with or without MTX at 30 °C. For the Y2H assay, cells co-expressed PMS2 as bait
1025 fused to the DNA binding domain and were grown on medium with or without histidine at 30
1026 °C. Q542L and L550P are in the C-terminal domain, S406N is in the disordered linker, and
1027 I219V and T117M are in the N-terminal domain. (B) Overview of the experimental design to
1028 screen libraries of MLH1 variants for abundance and interaction with PMS2. Parts of this figure
1029 were created using BioRender.com. (C) Histograms showing the score distributions of MLH1
1030 variants in the DHFR-PCA (left) and the Y2H assay (right). A score of 0 corresponds to the
1031 wild-type in both assays, while a negative or positive score indicates reduced or increased
1032 abundance or interaction, respectively. As a reference, the abundance and interaction score of
1033 the wild-type MLH1 (black) and the L550P variant (red) are highlighted with dashed lines. The
1034 abundance scores were rescaled to range from -1 to 1 (see Materials and Methods).


A
Abundance
His
Lys
Arg
Asp
Glu
Cys
Met
Asn
Gln
Ser
Thr
Pro
Ala
Ile
Leu
Val
Phe
Trp
Tyr
Gly
Ter
487 497 507 517 527 537 547 557 567 577 587 597 607 617 627 637 647 657 667 677 687 697 707 717 727 737 747 756
Abundance score
Low abundance
WT abundance
High abundance
Wild−type amino acid
Missing variant
B
PMS2 CTD
MLH1 CTD
756
487
180°
PMS2 CTD
MLH1 CTD
756
487
C
Interaction
His
Lys
Arg
Asp
Glu
Cys
Met
Asn
Gln
Ser
Thr
Pro
Ala
Ile
Leu
Val
Phe
Trp
Tyr
Gly
Ter
487 497 507 517 527 537 547 557 567 577 587 597 607 617 627 637 647 657 667 677 687 697 707 717 727 737 747 756
Interaction score
Weak interaction
WT interaction
Strong interaction
Wild−type amino acid
Missing variant
D
PMS2 CTD
MLH1 CTD
756
487
180°
PMS2 CTD
MLH1 CTD
756
487

1035

1036 **Fig. 3** – *MLH1 abundance and interaction heatmaps*. (A) Heatmaps showing the abundance
1037 scores of MLH1 variants. The scores range from low abundance (red), through wild-type-like
1038 abundance (white), to high abundance (blue). The wild-type residue at each position is marked
1039 in yellow, while missing variants are marked in gray. Substitutions that lead to a stop codon
1040 (nonsense mutations) are indicted by 'Ter'. The position in the full-length MLH1 protein is
1041 shown below the heatmap. A linear representation of the secondary structure of MLH1 is
1042 displayed above the heatmap, colored by AlphaFold2 pLDDT as a measure of disorder. (B)
1043 The per-position median abundance scores are mapped to the C-terminal domain of MLH1 in
1044 the AlphaFold2-Multimer predicted C-terminal MutLα complex. The color scale corresponds
1045 to that shown in Fig. 3A. The C-terminal domain of PMS2 is shown in orange. The N- and C-
1046 terminus of the C-terminal domain of MLH1 are annotated with the position in the full-length
1047 MLH1 protein. (C) Same as in (A), but for interaction scores. (D) Same as in (B), but for
1048 interaction scores.

1049


A
B
GEMME ΔΔE
Rosetta ΔΔG
Abundance score
Low abundance
WT abundance
High abundance
Interaction score
Weak interaction
WT interaction
Strong interaction
C
D
Abundance score
Interaction score
gnomAD allele frequency (log)
ClinVar
Benign
Likely benign
Likely pathogenic
Pathogenic
Uncertain significance

1050 **Fig. 4** – *Experimental effects correlate with computational predictions and clinical*
1051 *classifications*. (A) Scatter plots showing the correlation between the predicted GEMME ΔΔE
1052 score and predicted Rosetta ΔΔG score. A negative ΔΔE score suggests that the variant may
1053 not be compatible with MLH1 function as encoded in the sequence family. A positive ΔΔG
1054 score suggests that the variant is less stable than the wild-type MLH1. The MLH1 variants are
1055 colored by the abundance score and the color scale corresponds to that shown in Fig. 3A. (B)
1056 Same as in (A), but for interaction scores. (C) Scatter plots showing the correlation between
1057 the abundance score and the gnomAD allele frequency (log scale). The MLH1 variants are
1058 colored based on their classification in ClinVar as either benign (dark green), likely benign
1059 (light green), likely pathogenic (yellow), pathogenic (red) or uncertain significance (gray).
1060 Variants observed in patients but not in gnomAD are indicated to the left of the dashed lines
1061 on the x-axis. These variants were retrieved from Simple ClinVar [34] (accessed July 25, 2024).
1062 (D) Same as in (C), but for interaction scores.