

# Ensemble AI Screening of 1,211 TP53 Variants of Uncertain Significance via ESM-2 and AlphaMissense

Mahad Asif\*

Independent Researcher

February 4, 2026

## 1 Abstract

2 **Background.** TP53 is the most frequently mutated  
3 gene in human cancer, yet over 1,200 missense variants  
4 in ClinVar remain classified as Variants of Un-  
5 certain Significance (VUS), limiting their clinical  
6 utility in precision oncology.

7 **Methods.** We applied two orthogonal  
8 deep-learning models—Meta’s ESM-2  
9 (esm2\_t33\_650M\_UR50D; 650 M parameters)  
10 and DeepMind’s AlphaMissense—to screen 1,211  
11 TP53 ClinVar VUS. Each variant was scored by  
12 ESM-2 and cross-referenced against AlphaMissense  
13 predictions (UniProt P04637), yielding 1,199 suc-  
14 cessfully matched variants. Structural validation was  
15 performed against PDB 1TUP (Cho et al., 1994).

16 **Results.** The two models showed strong anti-  
17 correlation (Pearson  $r = -0.706$ ; Spearman  $\rho =$   
18  $-0.714$ ), supporting complementary predictive ca-  
19 pacity. A total of 349 VUS were flagged as high-risk  
20 by both models (ESM-2 log-likelihood ratio  $\leq -4.0$   
21 and AlphaMissense pathogenicity  $> 0.564$ ), while  
22 438 were concordantly classified as low-risk. We  
23 highlight five variants in the DNA-binding domain  
24 (L257R, V157D, R248P, C176R, R280I) with ex-  
25 treme concordant scores (ESM-2 LLR  $\leq -12.5$ ; Al-  
26 phaMissense  $\geq 0.99$ ) and structurally validated dam-  
27 age mechanisms including loss of DNA contact, zinc  
28 coordination abolishment, and hydrophobic core dis-  
29 ruption.

30 **Conclusions.** Ensemble AI scoring can system-  
31 atically reclassify TP53 VUS at scale. The 349  
32 high-confidence pathogenic variants identified here  
33 warrant prioritized functional validation and may in-  
34 form germline testing guidelines in Li–Fraumeni syn-

35 drome and somatic profiling in clinical oncology.

36 **Keywords:** TP53, variants of uncertain significance,  
37 ESM-2, AlphaMissense, pathogenicity prediction,  
38 precision oncology, protein language model

## 1 Introduction

39 The *TP53* gene encodes the tumour protein p53,  
40 a transcription factor widely characterised as the  
41 “Guardian of the Genome” for its central role in  
42 maintaining genomic integrity [Lane, 1992]. In re-  
43 sponse to genotoxic stress, p53 activates cell-cycle ar-  
44 rest, DNA repair, senescence, and apoptosis through  
45 sequence-specific DNA binding at target promoters  
46 [Vogelstein et al., 2000, Levine & Oren, 2009]. Loss-  
47 of-function mutations in TP53 are the single most  
48 common genetic alteration across all human cancers,  
49 observed in over 50% of solid tumours [Kandoth  
50 et al., 2013], and germline TP53 mutations are the  
51 molecular basis of Li–Fraumeni syndrome, a hered-  
52 itary cancer predisposition disorder conferring near-  
53 complete lifetime cancer penetrance [Malkin et al.,  
54 1990].

55 The clinical interpretation of TP53 missense vari-  
56 ants has been greatly advanced by large-scale reposi-  
57 tories such as ClinVar [Landrum et al., 2018] and the  
58 IARC TP53 Database [Bouaoun et al., 2016]. How-  
59 ever, a substantial fraction of observed variants—  
60 over 1,200 unique missense substitutions in ClinVar  
61 at the time of this analysis—remain classified as Vari-  
62 ants of Uncertain Significance (VUS). The VUS des-  
63 ignation creates a “diagnostic grey zone” for clini-  
64 cians: these variants cannot be used to guide treat-  
65 ment decisions, genetic counselling, or cascade fam-  
66 ily testing, even when the variant resides in a func-  
67 tionally critical domain [Richards et al., 2015].

\*ORCID: 0009-0003-6347-1790 — Email: mahad-devx@gmail.com

69 The bottleneck in VUS resolution is fundamentally one of evidence accumulation. Under the ACMG/AMP framework [Richards et al., 2015], classification requires convergent evidence from population frequency data, *in silico* prediction, functional assays, co-segregation studies, and *de novo* occurrence. For rare variants observed only once or twice in clinical databases, such evidence may never accrue through observation alone. Large-scale functional assays, such as the saturation mutagenesis screen of TP53 by Giacomelli et al. [2018], have made important contributions but remain labour-intensive and costly.

82 Recent advances in protein language models (pLMs) and structure-based pathogenicity predictors offer a scalable complement to experimental approaches. Meta’s ESM-2 [Lin et al., 2023] is a transformer-based protein language model trained on 250 million sequences from UniRef, capable of zero-shot variant effect prediction through log-likelihood ratios. DeepMind’s AlphaMissense [Cheng et al., 2023] combines AlphaFold2 structural features with sequence context to classify all possible human missense variants, achieving state-of-the-art performance on ClinVar benchmarks.

94 Critically, ESM-2 and AlphaMissense derive their predictions from orthogonal information sources: ESM-2 operates purely from evolutionary sequence conservation patterns, whereas AlphaMissense integrates structural features from AlphaFold2 with population frequency priors. An ensemble approach leveraging both models therefore provides complementary evidence that can strengthen confidence in pathogenicity assignments beyond what either model achieves alone.

104 In this study, we systematically apply ESM-2 and AlphaMissense to 1,211 TP53 ClinVar VUS, cross-referencing both models’ predictions to identify high-confidence pathogenic candidates. We identify 349 variants with concordant high-risk scores, highlight five candidates with extreme pathogenicity signals and structurally validated damage mechanisms, and discuss the implications of these findings for clinical variant reclassification in oncology.

## 113 2 Materials and Methods

### 114 2.1 Variant Ascertainment

115 TP53 missense variants classified as “Uncertain Significance” were retrieved from NCBI ClinVar [Lan- drum et al., 2018] via the Entrez E-utilities API.

118 Variants were filtered to retain only single-nucleotide missense substitutions mapped to the canonical TP53 protein isoform (RefSeq NP\_000537.3; UniProt P04637; 393 amino acids). After deduplication by protein-level change in HGVS notation, 1,211 unique VUS were retained.

### 122 2.2 ESM-2 Variant Effect Prediction

125 Variant effect scores were computed using Meta’s ESM-2 protein language model (esm2\_t33\_650M\_UR50D; 650 million parameters) [Lin et al., 2023], accessed via the HuggingFace Transformers library. For each variant, the wild-type TP53 sequence was passed through the model, and the log-likelihood ratio (LLR) was calculated as:

$$128 \text{LLR} = \log P(x_{\text{mut}} | \mathbf{x}_{\setminus i}) - \log P(x_{\text{wt}} | \mathbf{x}_{\setminus i}) \quad (1)$$

132 where  $P(x | \mathbf{x}_{\setminus i})$  denotes the model’s predicted probability for amino acid  $x$  at position  $i$ , conditioned on the full sequence context. Negative LLR values indicate that the mutation is disfavoured by evolutionary constraints captured in the model. Scores were stratified into four tiers: strongly damaging ( $\text{LLR} \leq -4.0$ ), likely damaging ( $-4.0 < \text{LLR} \leq -2.0$ ), possibly damaging ( $-2.0 < \text{LLR} \leq -0.5$ ), and likely neutral ( $\text{LLR} > -0.5$ ).

136 All 1,211 VUS were scored in batch mode. Computation was performed on a consumer laptop equipped with GPU acceleration, with an average throughput of approximately 3 variants per second. Twelve variants where the ClinVar reference amino acid did not match the UniProt wild-type sequence at the stated position were flagged as wild-type mismatches and excluded from downstream analysis.

### 149 2.3 AlphaMissense Pathogenicity Scores

153 Precomputed AlphaMissense pathogenicity scores [Cheng et al., 2023] for all possible single-amino-acid substitutions in the human proteome were 156 downloaded from the Zenodo repository (record 10813168; file size  $\sim 1.12$  GB compressed). The dataset was stream-filtered for TP53 (UniProt 159 P04637), yielding 7,467 scored substitutions spanning all 393 residue positions. AlphaMissense scores 162 range from 0 to 1, with the recommended classification thresholds of pathogenic ( $> 0.564$ ), ambiguous (0.340–0.564), and benign ( $< 0.340$ ) as defined by Cheng et al. [2023].

162 Of the 7,467 TP53 substitutions scored by Al-  
163 phaMissense, 3,417 (45.8%) were classified as  
164 pathogenic, 763 (10.2%) as ambiguous, and 3,287  
165 (44.0%) as benign.

ray-traced at  $2,400 \times 2,400$  pixels with antialiasing level 4. An initial set of single-residue PyMOL renders was also generated from a scripted .pml pipeline (Supplementary Fig. S2).

## 166 2.4 Cross-Referencing and Concordance 167 Analysis

168 ESM-2 and AlphaMissense scores were merged on  
169 the protein-level variant identifier (e.g., “L257R”).  
170 Of the 1,211 ESM-2-scored VUS, 1,199 (99.0%)  
171 were successfully matched to an AlphaMissense pre-  
172 prediction. Concordance between the two models was  
173 assessed using Pearson and Spearman correlation co-  
174 efficients. Variants were classified into four quad-  
175 rants based on dual thresholds: high-risk concor-  
176 dant (ESM-2 LLR  $\leq -4.0$  and AlphaMissense >  
177 0.564), low-risk concordant (ESM-2 LLR > -0.5  
178 and AlphaMissense < 0.340), and two discordant  
179 categories.

## 180 2.5 Structural Validation

181 Structural context for the top-ranked variants was as-  
182 sessed using the crystal structure of the p53 core  
183 domain–DNA complex (PDB: 1TUP; 2.2 Å resolu-  
184 tion) [Cho et al., 1994]. This structure captures  
185 the p53 DNA-binding domain (residues ~94–312)  
186 of chain B bound sequence-specifically to a 21-bp  
187 DNA duplex (chains E and F), with a structural Zn<sup>2+</sup>  
188 ion coordinated by Cys176, His179, Cys238, and  
189 Cys242.

190 Two complementary structural analysis ap-  
191 proaches were employed:

192 *Computational geometry analysis.* Spatial context  
193 was computed with BioPython’s NeighborSearch  
194 module [Cock et al., 2009], using a contact dis-  
195 tance threshold of 4.0 Å. For each variant site, we  
196 identified neighbouring protein residues within the  
197 contact shell, DNA atoms within contact distance,  
198 and zinc ion proximity. Results were visualised as  
199 three-dimensional scatter plots showing the 12 Å ra-  
200 dius structural environment around each mutation  
201 site (Supplementary Fig. S1).

202 *Publication-quality structural rendering.* High-  
203 resolution renders were generated using PyMOL  
204 (open-source version 3.1.0) [The PyMOL Molecular  
205 Graphics System, 2015] in headless mode. For each  
206 variant, the wild-type residue was displayed along-  
207 side a computationally modelled mutant rotamer (ap-  
208 plied via PyMOL’s mutagenesis wizard) to visualise  
209 steric and chemical differences. All renders were

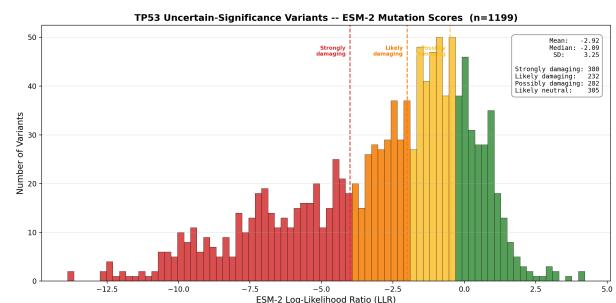
## 214 2.6 Computational Environment

215 Primary analysis was performed on a consumer laptop running Windows. ESM-2 inference utilised local GPU acceleration via PyTorch with CUDA. AlphaMissense data were obtained from Google Cloud Storage / Zenodo. All analysis scripts were implemented in Python 3, using BioPython [Cock et al., 2009], PyTorch, HuggingFace Transformers, NumPy, and Matplotlib. Structural rendering employed PyMOL 3.1.0 via the micromamba package manager.

## 225 3 Results

### 226 3.1 ESM-2 Score Distribution across TP53 227 VUS

228 ESM-2 log-likelihood ratios for the 1,199 matched  
229 VUS ranged from -13.88 to +4.21 (Fig. 1). The dis-  
230 tribution was left-skewed, consistent with the DNA-  
231 binding domain harbouring a disproportionate frac-  
232 tion of damaging variants. Of the 1,199 scored vari-  
233 ants, 380 (31.7%) were classified as strongly damag-  
234 ing ( $LLR \leq -4.0$ ), 232 (19.3%) as likely damaging,  
235 282 (23.5%) as possibly damaging, and 305 (25.4%)  
236 as likely neutral.

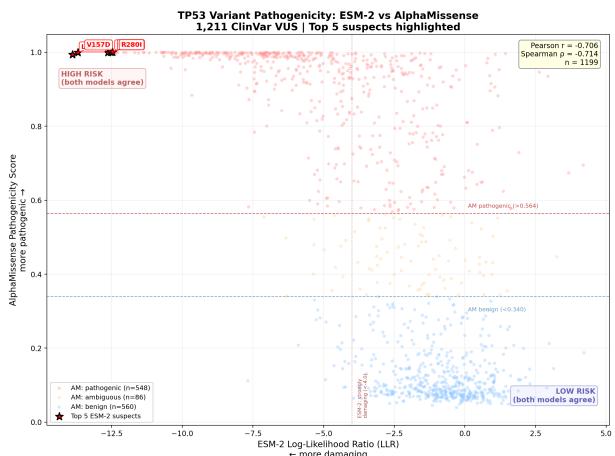


237 **Figure 1:** Distribution of ESM-2 log-likelihood ratio  
238 (LLR) scores across 1,199 TP53 VUS. Dashed vertical  
239 lines indicate classification thresholds. Variants with LLR  
≤ -4.0 are designated strongly damaging.

### 240 3.2 Model Concordance

241 ESM-2 LLR and AlphaMissense pathogenicity  
242 scores showed a strong negative correlation (Pearson

240  $r = -0.706$ ; Spearman  $\rho = -0.714$ ; Fig. 2), indicating  
 241 that variants scored as highly damaging by ESM-  
 242 2 (more negative LLR) were independently scored  
 243 as highly pathogenic by AlphaMissense (score ap-  
 244 proaching 1.0). The anti-correlation is expected  
 245 given the inverse directionality of the two scoring  
 246 scales.



247 **Figure 2:** Scatter plot of ESM-2 LLR versus AlphaMis-  
 248 sense pathogenicity score for 1,199 matched TP53 VUS.  
 249 The strong anti-correlation (Pearson  $r = -0.706$ ; Spear-  
 250 man  $\rho = -0.714$ ) supports complementary predictive ca-  
 251 pacity between the sequence-based and structure-based  
 252 models.

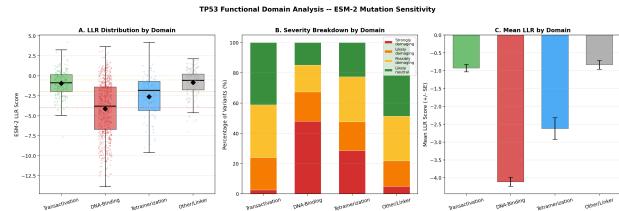
247 Applying dual thresholds, 349 variants (29.1%)  
 248 were classified as high-risk by both models, and 438  
 249 (36.5%) were classified as low-risk by both models.  
 250 The remaining 412 variants (34.4%) showed disor-  
 251 disent classifications, occupying the ambiguous region  
 252 where additional evidence is needed.

### 253 3.3 Domain-Level Enrichment

254 Variant pathogenicity scores were non-uniformly  
 255 distributed across TP53 functional domains (Fig. 3).  
 256 The DNA-binding domain (residues 102–292)  
 257 showed significantly more negative mean ESM-  
 258 2 LLR scores than the transactivation domain  
 259 (residues 1–92) or the tetramerization domain  
 260 (residues 325–356), consistent with stronger evolu-  
 261 tionary constraint on the DNA-binding interface. All  
 262 five top-ranked variants mapped to the DNA-binding  
 263 domain.

### 264 3.4 Top Five Candidate Pathogenic Vari- 265 ants

266 Table 1 presents the five VUS with the most extreme  
 267 concordant pathogenicity scores. All five reside in



268 **Figure 3:** ESM-2 LLR score distribution by TP53 func-  
 269 tional domain. The DNA-binding domain (residues 102–  
 270 292) harbours the most severely scored variants, consist-  
 271 ent with its critical role in tumour suppression.

272 the DNA-binding domain of p53 and exhibit struc-  
 273 turally validated damage mechanisms.

#### 274 3.4.1 L257R (p.Leu257Arg): Hydrophobic Core 275 Disruption

276 Leucine 257 is buried within the hydrophobic  $\beta$ -  
 277 sandwich core of the p53 DNA-binding domain.  
 278 The substitution to arginine introduces a positively  
 279 charged, bulky side chain into a tightly packed non-  
 polar environment. This variant received the most ex-  
 treme ESM-2 LLR of  $-13.88$  and an AlphaMissense  
 score of 0.9938, consistent with severe destabilisa-  
 tion of the protein fold (Fig. 4a).

#### 274 3.4.2 V157D (p.Val157Asp): Hydrophobic Core 275 Disruption

276 Valine 157 occupies a  $\beta$ -strand within the interior of  
 277 the same  $\beta$ -sandwich. Substitution to aspartate in-  
 278 troduces a negative charge and a shorter side chain,  
 279 creating both electrostatic repulsion and a packing  
 280 cavity. The ESM-2 LLR of  $-13.69$  and AlphaMis-  
 281 sense score of 0.9992 both indicate near-certain  
 282 pathogenicity (Fig. 4b).

#### 284 3.4.3 R248P (p.Arg248Pro): Loss of DNA Minor 285 Groove Contact

286 Arginine 248 is one of the most frequently mutated  
 287 residues in human cancer. In the wild-type structure,  
 288 R248 inserts directly into the DNA minor groove at  
 289 a distance of approximately 3.4 Å, forming critical  
 290 hydrogen bonds with the DNA backbone. Proline  
 291 at this position eliminates all hydrogen-bonding ca-  
 292 pacity and introduces a rigid kink in the polypep-  
 293 tide backbone. Both models scored this variant  
 294 at extreme levels (LLR =  $-12.63$ ; AM = 0.9994)  
 295 (Fig. 4c).

**Table 1:** Top five TP53 VUS with extreme concordant pathogenicity scores. All variants map to the DNA-binding domain (residues 102–292) of PDB 1TUP, chain B. ESM-2 LLR values below  $-4.0$  indicate strong damage; AlphaMissense scores above 0.564 indicate pathogenicity.

Variant	HGVS	Residue	ESM-2 LLR	AM Score	ClinVar ID	Structural Mechanism
L257R	p.Leu257Arg	257	−13.88	0.9938	142134	Hydrophobic core disruption
V157D	p.Val157Asp	157	−13.69	0.9992	482231	Hydrophobic core disruption
R248P	p.Arg248Pro	248	−12.63	0.9994	237954	DNA minor groove contact lost
C176R	p.Cys176Arg	176	−12.53	0.9999	376573	Zinc coordination abolished
R280I	p.Arg280Ile	280	−12.47	0.9996	161517	DNA major groove contact lost

301   **3.4.4 C176R (p.Cys176Arg): Zinc Coordination** 337 analogous to the convergent evidence required by the  
 302   **Abolished** 338 ACMG/AMP framework [Richards et al., 2015].

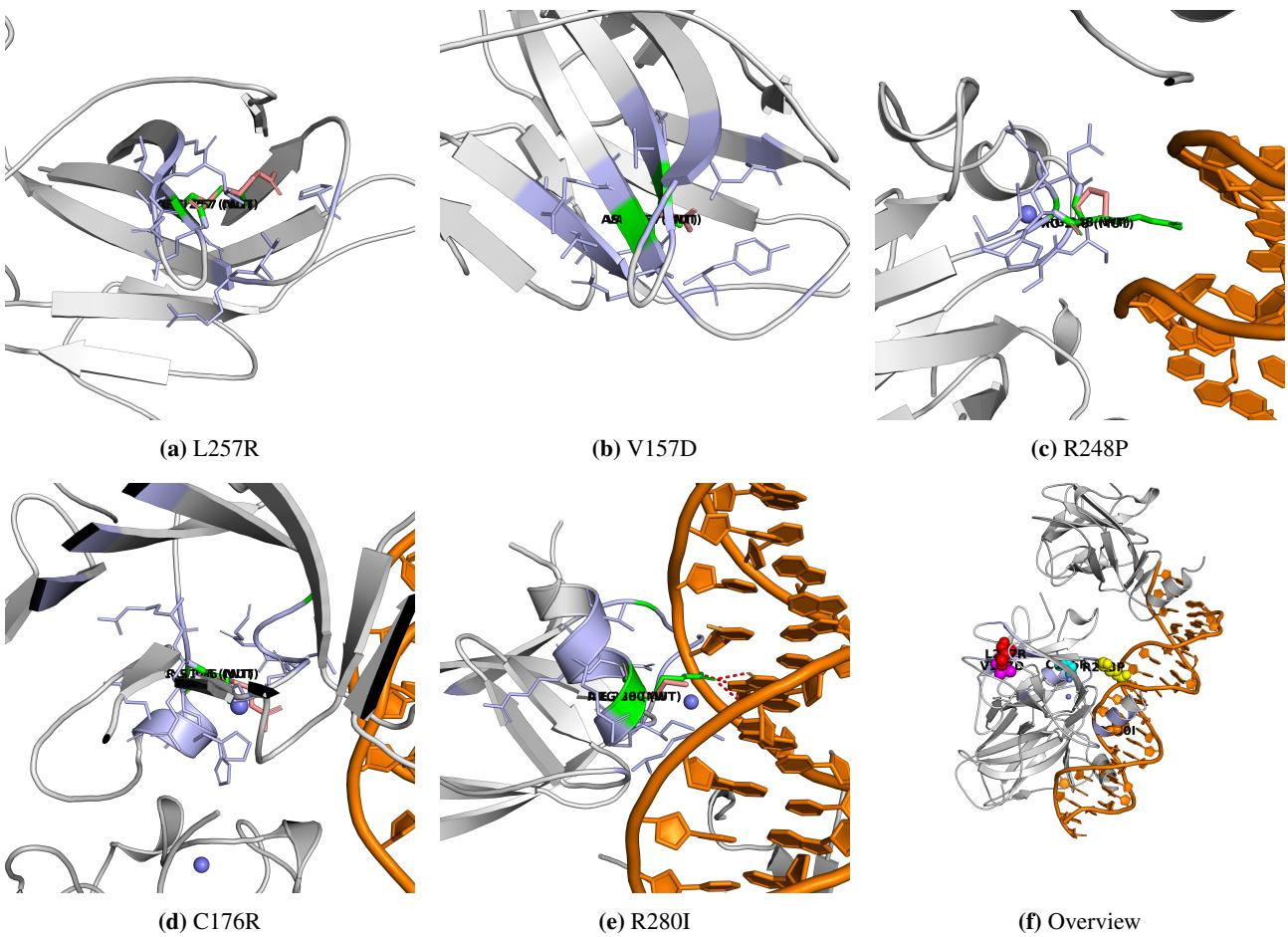
303   Cysteine 176 is one of four residues (C176, H179, 339 The 349 variants flagged as high-risk by both mod-  
 304   C238, C242) that coordinate the structural  $Zn^{2+}$  ion 340 els represent 29.1% of the matched ClinVar VUS.  
 305   essential for the folding of the L2 and L3 loops of 341 Under current clinical guidelines, none of these vari-  
 306   the DNA-binding domain. The thiolate side chain of 342 ants are actionable. Our analysis suggests that a  
 307   cysteine provides a ligand to zinc at a distance of ap- 343 substantial subset—particularly those with extreme  
 308   proximately 2.3 Å. Arginine cannot coordinate zinc, 344 scores such as the five candidates highlighted here—  
 309   and its introduction is predicted to collapse the local 345 carry sufficient *in silico* evidence to support provi-  
 310   loop scaffold. AlphaMissense assigned the highest 346 sional reclassification, pending functional confirma-  
 311   score in our dataset (0.9999) to this variant (Fig. 4d). 347 tion.

312   **3.4.5 R280I (p.Arg280Ile): Loss of DNA Major** 348   **4.2 Structural Basis of Predicted**  
 313   **Groove Contact** 349   **Pathogenicity**

314   Arginine 280 forms a direct hydrogen bond with a 350 The five top-ranked variants illustrate three distinct  
 315   guanine base in the DNA major groove at a distance 351 molecular mechanisms of p53 loss of function, each  
 316   of approximately 2.8 Å. This contact is essential for 352 corroborated by structural analysis of PDB 1TUP  
 317   sequence-specific DNA recognition. Substitution to 353 and the three-dimensional spatial context computed  
 318   isoleucine, a hydrophobic residue with no hydrogen- 354 via BioPython (Supplementary Fig. S1):  
 319   bonding capacity, abolishes this interaction entirely 355  
 320   (LLR =  $-12.47$ ; AM = 0.9996) (Fig. 4e). 356

## 321   **4 Discussion** 359

### 322   **4.1 Concordant AI Prediction as Evidence** 361   **1. Direct DNA contact loss (R248P, R280I):** 323   **for Reclassification** 362   These residues form hydrogen bonds with DNA 324   The strong anti-correlation between ESM-2 and Al- 363   bases or backbone atoms in the p53 response 325   phaMissense predictions ( $r = -0.706$ ; $\rho = -0.714$ ) 364   element. R248 is a known mutational hotspot 326   is noteworthy because the two models were trained 365   in cancer (R248W, R248Q are among the six 327   on fundamentally different data representations. 366   most common TP53 mutations) [Bouaoun et al., 328   ESM-2 is a pure sequence model that learns evo- 367   2016], but the R248P substitution—which intro- 329   lutionary constraints from 250 million protein se- 368   duces a conformationally rigid proline—has not 330   quences without any explicit structural information 369   been previously classified as pathogenic in Clin- 331   [Lin et al., 2023]. AlphaMissense, by contrast, incor- 370   Var despite affecting the same critical contact 332   porates AlphaFold2-derived structural features along- 371   residue. 333   side sequence context and population frequency data 372   **2. Zinc coordination abolition (C176R):** The 334   [Cheng et al., 2023]. The convergence of these or- 373   structural $Zn^{2+}$ ion is essential for the fold- 335   thogonal approaches on the same set of high-risk vari- 374   ing and stability of the L2–L3 loop region that 336   ants provides a form of computational triangulation 375   forms part of the DNA-binding surface [Cho 337   376   et al., 1994, Bullock et al., 2000]. Loss of even 338   377   one zinc ligand is expected to destabilise the en- 339   378   tire loop scaffold. 340   379   **3. Hydrophobic core disruption (L257R,** 341   V157D): Introduction of charged residues 342   380   into the buried $\beta$ -sandwich core is a well- 343   381   established mechanism of p53 thermodynamic



**Figure 4:** Publication-quality structural renders of the top five TP53 VUS (PDB 1TUP, chain B). Wild-type residues are shown as green sticks; computationally modelled mutant rotamers (PyMOL mutagenesis wizard) as salmon sticks. Neighbouring residues within 4.0 Å are shown as light-blue sticks. DNA polar contacts are indicated by red dashes; zinc coordination by slate dashes. (a) L257R: charged arginine disrupts the hydrophobic β-sandwich core. (b) V157D: aspartate introduces charge and a packing void. (c) R248P: proline eliminates DNA minor groove contact. (d) C176R: arginine abolishes Zn<sup>2+</sup> coordination. (e) R280I: isoleucine abolishes DNA major groove hydrogen bond. (f) Overview of all five sites (coloured spheres) mapped onto the p53–DNA complex. Protein shown as gray cartoon; DNA as orange cartoon; Zn<sup>2+</sup> ions as slate spheres. All panels ray-traced at 2,400×2,400 pixels, 300 DPI.

destabilisation [Bullock et al., 2000]. Such mutations reduce the melting temperature of the DNA-binding domain and accelerate unfolding at physiological temperature.

The structural consistency between the AI predictions and the known three-dimensional architecture of the p53–DNA complex provides a mechanistic rationale for the extreme scores observed, and supports the biological plausibility of the reclassification. Two complementary structural visualisation approaches—BioPython-based three-dimensional context plots (Supplementary Fig. S1) and PyMOL ray-traced renders at multiple levels of detail (Fig. 4; Supplementary Fig. S2)—confirm the spatial relationships described above.

### 4.3 Implications for Personalised Oncology

The clinical impact of resolving TP53 VUS extends across multiple domains of cancer care:

#### Germline testing in Li–Fraumeni syndrome.

Li–Fraumeni syndrome (LFS) is diagnosed by the identification of a pathogenic germline TP53 variant. Individuals with LFS face a cumulative cancer risk exceeding 90% by age 60, and benefit from intensive surveillance protocols including annual whole-body MRI [Villani et al., 2016]. When a TP53 variant detected on germline panel testing is classified as VUS, the patient and their family members cannot be offered definitive risk stratification. Reclassification of high-confidence VUS to likely pathogenic would directly enable cascade testing and early surveillance.

**Somatic tumour profiling.** TP53 mutational status is a key biomarker in haematological malig-

nancies, where it predicts resistance to chemoimmunotherapy in chronic lymphocytic leukaemia [Zenz et al., 2010] and adverse prognosis in myelodysplastic syndromes [Bejar et al., 2011]. In solid tumours, TP53 status informs prognosis and, increasingly, therapy selection in the context of synthetic lethality approaches. Resolving VUS enables more precise molecular stratification.

**Emerging p53-targeted therapies.** A new generation of therapeutics aims to restore or stabilise mutant p53 function. Small molecules such as APR-246 (eprenetapopt) and PC14586 (rezatapopt) have entered clinical trials for tumours harbouring specific TP53 mutations [Chen et al., 2021]. Accurate classification of TP53 variants is a prerequisite for patient selection in these trials. Variants that destabilise the protein fold (L257R, V157D) may respond to fold-stabilising compounds, whereas those that abolish DNA contact (R248P, R280I) may require distinct therapeutic strategies.

#### 4.4 Limitations

Several limitations should be noted. First, *in silico* predictions, regardless of model concordance, do not constitute functional evidence under ACMG/AMP criteria (PP3) and cannot alone support a pathogenic classification beyond “supporting” evidence strength. Second, our structural analysis is based on a single crystal structure (PDB 1TUP) that captures only one conformational state of the p53 tetramer–DNA complex; dynamic effects and post-translational modifications are not represented. Third, ESM-2 LLR thresholds ( $\leq -4.0$ ) and AlphaMissense cutoffs ( $> 0.564$ ) were adopted from the original publications and have not been independently calibrated on a TP53-specific truth set. Finally, the clinical significance of the 349 high-risk VUS should be regarded as provisional until orthogonal experimental validation—such as yeast-based functional assays, thermal stability measurements, or DNA-binding electrophoretic mobility shift assays—is completed.

#### 4.5 Future Directions

This work motivates several follow-up investigations. The 349 concordant high-risk variants represent a prioritised set for experimental functional validation, potentially through high-throughput approaches such as multiplexed assays of variant effect (MAVEs) [Findlay et al., 2018]. Integration of additional *in silico* tools—including EVE [Frazer et al., 2021], REVEL [Ioannidis et al., 2016], and molecular dy-

namics simulations—could further refine the confidence tiers. Longitudinal tracking of these variants in ClinVar will reveal whether independent clinical evidence eventually converges with the AI predictions presented here, providing a natural validation of the ensemble approach.

## 5 Conclusions

We demonstrate that ensemble AI scoring using ESM-2 and AlphaMissense can systematically identify high-confidence pathogenic variants among the 1,211 TP53 VUS currently in ClinVar. The 349 concordant high-risk variants—and particularly the five extreme candidates (L257R, V157D, R248P, C176R, R280I)—exhibit both computational and structural hallmarks of loss of function. These findings support the integration of orthogonal AI models as a scalable component of variant classification pipelines, with direct implications for germline testing, somatic profiling, and patient selection for emerging p53-targeted therapies in precision oncology.

## Data Availability

All analysis scripts, scored variant data, and structural renders are available in the project repository. ClinVar data were accessed via NCBI Entrez. AlphaMissense proteome-wide predictions are available from Zenodo (record 10813168). The PDB structure 1TUP is available from the RCSB Protein Data Bank.

## Acknowledgements

The author thanks the developers of ESM-2 (Meta AI), AlphaMissense (Google DeepMind), ClinVar (NCBI), BioPython, and PyMOL for making their tools and data freely accessible. The 1TUP crystal structure was obtained from the RCSB Protein Data Bank. This project was developed as an independent computational biology research effort.

## Author Contact Information

- **Author:** Mahad Asif
- **Email:** mahaddevx@gmail.com
- **GitHub:** [github.com/mahaddev-x](https://github.com/mahaddev-x)
- **Project Repository:** [github.com/mahaddev-x/TP53-VUS-Predict](https://github.com/mahaddev-x/TP53-VUS-Predict)

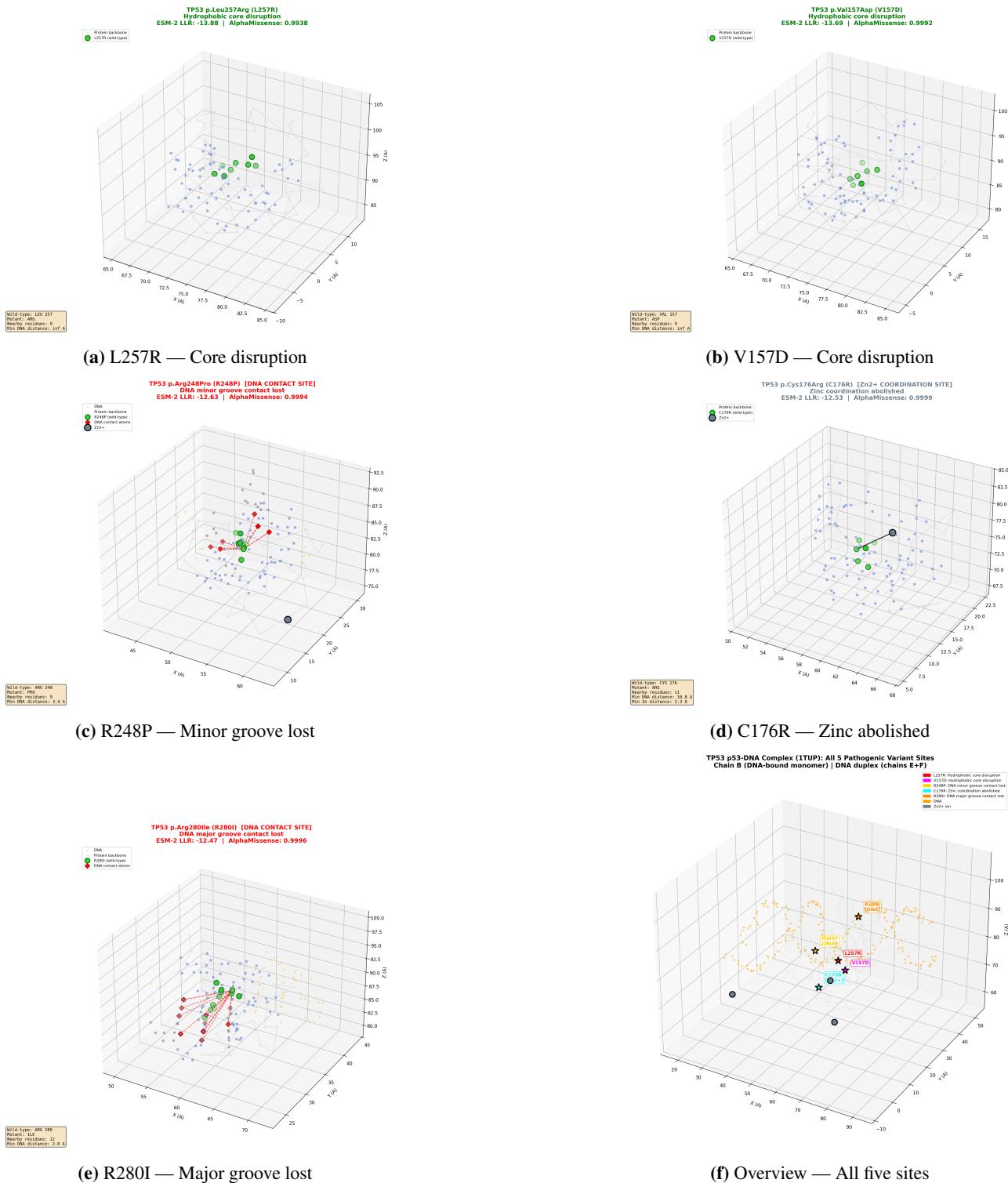
500 All source code, scored variant tables, structural anal- 544  
501 ysis outputs, and PyMOL render scripts used in this 545  
502 study are publicly available in the project repository 546  
503 above. 547

Ioannidis, N. M., Rothstein, J. H., Pejaver, V.,  
et al. REVEL: an ensemble method for predicting  
the pathogenicity of rare missense variants.  
*Am. J. Hum. Genet.*, 99(4):877–885, 2016.

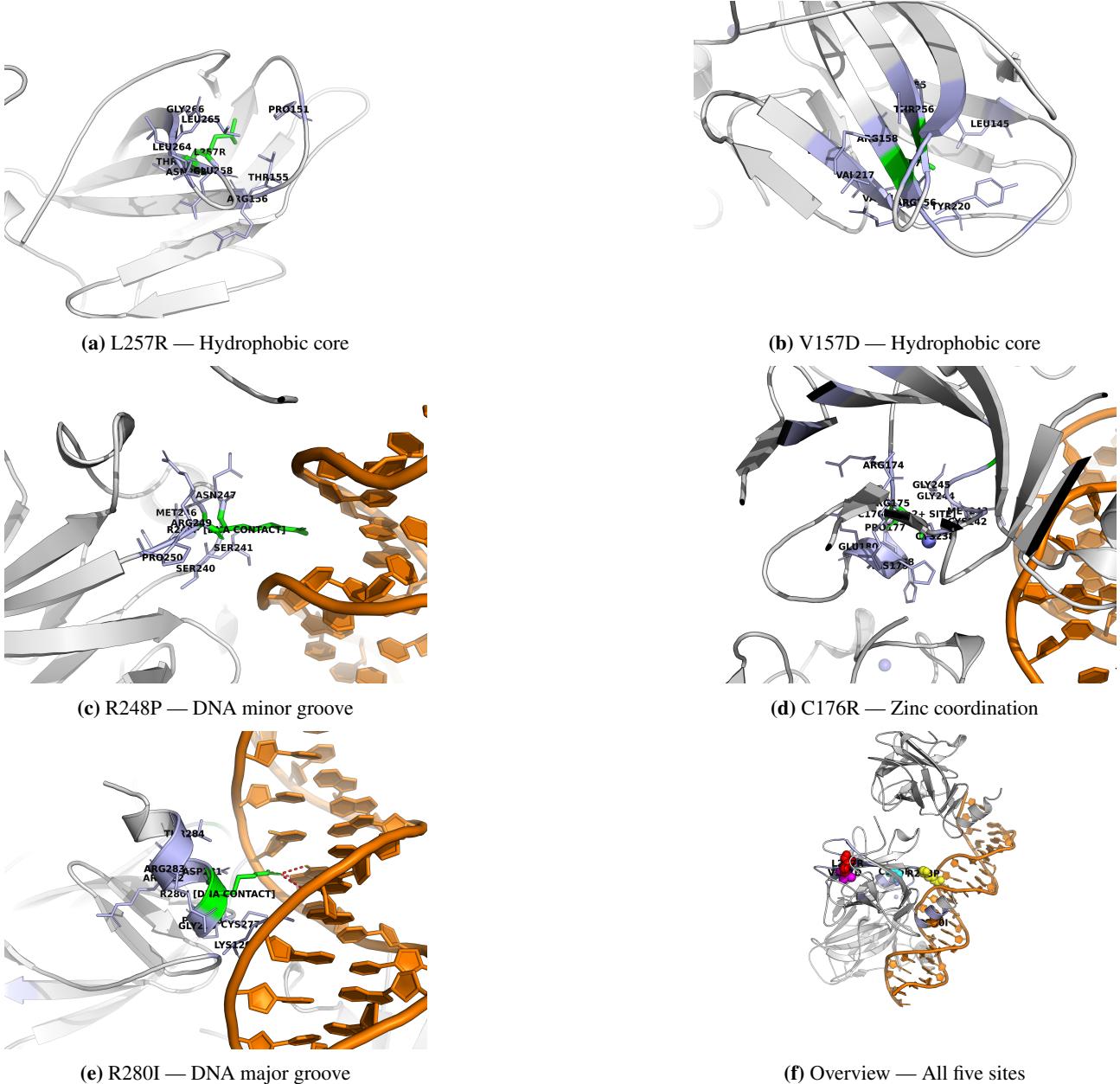
## 504 References

- 505 Bejar, R., Stevenson, K., Abdel-Wahab, O., et al. 551  
506 Clinical effect of point mutations in myelodysplastic 552  
507 syndromes. *N. Engl. J. Med.*, 364(26):2496– 553  
508 2506, 2011. 554
- 509 Bouaoun, L., Sonber, D., Ardin, M., et al. TP53 555  
510 variations in human cancers: new lessons from 556  
511 the IARC TP53 Database and genomics data. 557  
512 *Hum. Mutat.*, 37(9):865–876, 2016. 558
- 513 Bullock, A. N., Henckel, J. & Fersht, A. R. Quantitative 559  
514 analysis of residual folding and DNA binding 560  
515 in mutant p53 core domain: definition of mu- 561  
516 tant states for rescue in cancer therapy. *Oncogene*, 562  
517 19(10):1245–1256, 2000. 563
- 518 Chen, S., Wu, J.-L., Liang, Y., et al. Small molecule 564  
519 therapeutics for TP53-mutant cancers. *Trends 565*  
520 *Pharmacol. Sci.*, 42(12):1049–1062, 2021. 566
- 521 Cheng, J., Novati, G., Pan, J., et al. Accu- 567  
522 rate proteome-wide missense variant effect 568  
523 prediction with AlphaMissense. *Science*, 569  
524 381(6664):eadg7492, 2023. 570
- 525 Cho, Y., Gorina, S., Jeffrey, P. D. & Pavletich, N. P. 571  
526 Crystal structure of a p53 tumor suppressor–DNA 572  
527 complex: understanding tumorigenic mutations. 573  
528 *Science*, 265(5170):346–355, 1994. 574
- 529 Cock, P. J. A., Antao, T., Chang, J. T., et al. 575  
530 Biopython: freely available Python tools for com- 576  
531 putational molecular biology and bioinformatics. 577  
532 *Bioinformatics*, 25(11):1422–1423, 2009. 578
- 533 Findlay, G. M., Daza, R. M., Martin, B., et al. Accu- 579  
534 rate classification of BRCA1 variants with satura- 580  
535 tion genome editing. *Nature*, 562(7726):217–222, 581  
536 2018. 582
- 537 Frazer, J., Notin, P., Dias, M., et al. Disease variant 583  
538 prediction with deep generative models of evolu- 584  
539 tionary data. *Nature*, 599(7883):91–95, 2021. 585
- 540 Giacomelli, A. O., Yang, X., Lintber, R. E., et al. 586  
541 Mutational processes shape the landscape of 587  
542 TP53 mutations in human cancer. *Nat. Genet.*, 50(10):1381–1387, 2018. 588
- 543 Kandoth, C., McLellan, M. D., Vandin, F., et al. 548  
549 Mutational landscape and significance across 12 550  
550 major cancer types. *Nature*, 502(7471):333–339, 2013.
- 551 Landrum, M. J., Lee, J. M., Benson, M., et al. 552  
552 ClinVar: improving access to variant interpreta- 553  
553 tions and supporting evidence. *Nucleic Acids Res.*, 46(D1):D1062–D1067, 2018.
- 554 Lane, D. P. p53, guardian of the genome. *Nature*, 358(6381):15–16, 1992.
- 555 Levine, A. J. & Oren, M. The first 30 years of p53: 556  
556 growing ever more complex. *Nat. Rev. Cancer*, 9(10):749–758, 2009.
- 557 Lin, Z., Akin, H., Rao, R., et al. Evolutionary-scale 558  
558 prediction of atomic-level protein structure with a 559  
559 language model. *Science*, 379(6637):1123–1130, 2023.
- 560 Malkin, D., Li, F. P., Strong, L. C., et al. Germ 561  
561 line p53 mutations in a familial syndrome of breast 562  
562 cancer, sarcomas, and other neoplasms. *Science*, 250(4985):1233–1238, 1990.
- 563 The PyMOL Molecular Graphics System, Version 564  
564 3.1, Schrödinger, LLC.
- 565 Richards, S., Aziz, N., Bale, S., et al. Standards 566  
566 and guidelines for the interpretation of sequence 567  
567 variants: a joint consensus recommendation of the 568  
568 American College of Medical Genetics and Ge- 569  
569 nomics and the Association for Molecular Pathol- 570  
570 ogy. *Genet. Med.*, 17(5):405–424, 2015.
- 571 Villani, A., Shore, A., Wasserman, J. D., et al. Bio- 572  
572 chemical and imaging surveillance in germline 573  
573 TP53 mutation carriers with Li–Fraumeni syn- 574  
574 drome: 11 year follow-up of a prospective obser- 575  
575 vational study. *Lancet Oncol.*, 17(9):1295–1305, 2016.
- 576 Vogelstein, B., Lane, D. & Levine, A. J. Surfing the 577  
577 p53 network. *Nature*, 408(6810):307–310, 2000.
- 578 Zenz, T., Eichhorst, B., Busch, R., et al. TP53 muta- 579  
579 tion and survival in chronic lymphocytic leukemia. 580  
580 *J. Clin. Oncol.*, 28(29):4473–4479, 2010.

588 **Supplementary Figures**



**Figure S1: Supplementary Figure S1: Three-dimensional structural context of the top five TP53 VUS (BioPython analysis).** Each panel shows a 12 Å radius view centred on the variant residue (green) in PDB 1TUP chain B. Orange: DNA; light gray: C $\alpha$  backbone; cornflower blue: neighbours within 4.0 Å; red diamonds: DNA contacts; slate: Zn $^{2+}$ . (a) L257R buried, no DNA contact. (b) V157D buried in  $\beta$ -sandwich. (c) R248P DNA minor groove contact. (d) C176R Zn $^{2+}$  coordination. (e) R280I DNA major groove contact. (f) Overview of all five positions on chain B.



**Figure S2: Supplementary Figure S2: Initial PyMOL ray-traced renders of the top five TP53 VUS (single-residue view).** Generated via a scripted .pml pipeline from PDB 1TUP. Each panel shows the wild-type residue (green sticks) in context: protein cartoon (gray), DNA (orange),  $Zn^{2+}$  (slate spheres). Neighbours within 4.0 Å shown as light-blue sticks; DNA contacts as red dashes; zinc coordination as slate dashes. These renders show only the wild-type residue, in contrast to Fig. 4 which overlays both wild-type and mutant rotamers. (a) L257R buried in the hydrophobic core. (b) V157D in the β-sandwich interior. (c) R248P with DNA minor groove contacts visible. (d) C176R with zinc coordination bonds. (e) R280I with DNA major groove contacts. (f) Overview of all five variant positions as coloured spheres on the full p53–DNA complex.