RESEARCH ARTICLE

Figures and figure supplements

Comprehensive fitness landscape of SARS-CoV-2 M^{pro} reveals insights into viral resistance mechanisms

Julia M Flynn *et al*

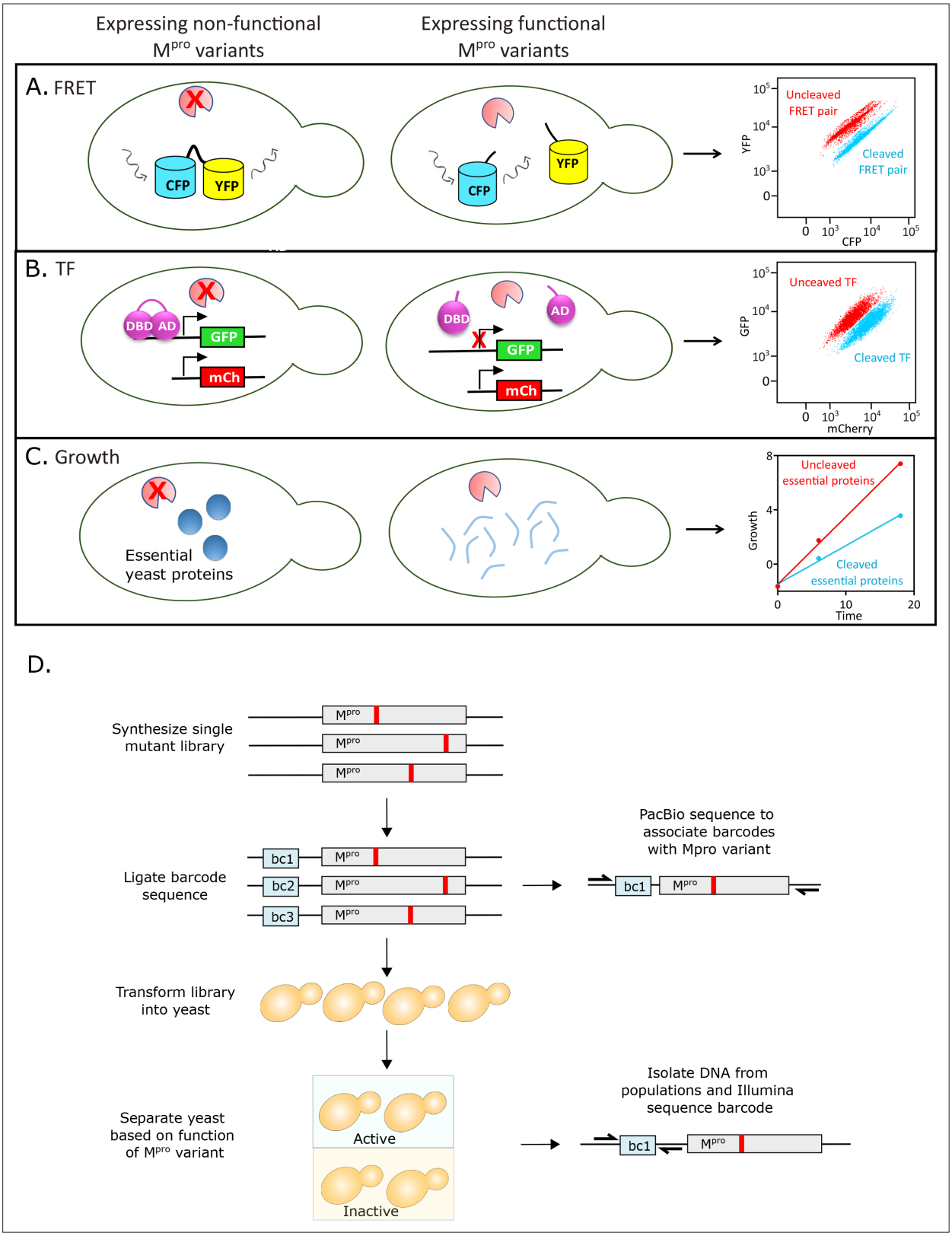


Figure 1. Experimental strategy to measure the function of all individual mutations of main protease (M^{pro}). (**A**) Fluorescence resonance energy transfer (FRET)-based reporter screen. The M^{pro} variants were sorted based on their ability to cleave at the M^{pro} cut-site, separating the YFP-CFP FRET pair. Cells were separated by fluorescence-activated single cell sorting (FACS) into cleaved (low FRET) and uncleaved (high FRET) populations. (**B**) Split transcription factor screen. M^{pro} variants were sorted based on their ability to cleave at the M^{pro} cut-site, separating the DNA binding domain (DBD) and

Figure 1 continued on next page

Figure 1 continued

activation domain (AD) of the Gal4 transcription factor. The transcription factor drives GFP expression from a galactose promoter. Cells were separated by FACS into cleaved (low GFP expression) and uncleaved (high GFP expression) populations. (**C**) Growth screen. Yeast cells expressing functional M^{pro} variants that cleave essential yeast proteins grow slowly and are depleted in bulk culture, while yeast cells expressing non-functional M^{pro} variants are enriched. (**D**) Barcoding strategy to measure frequency of all individual mutations of M^{pro} in a single experiment.

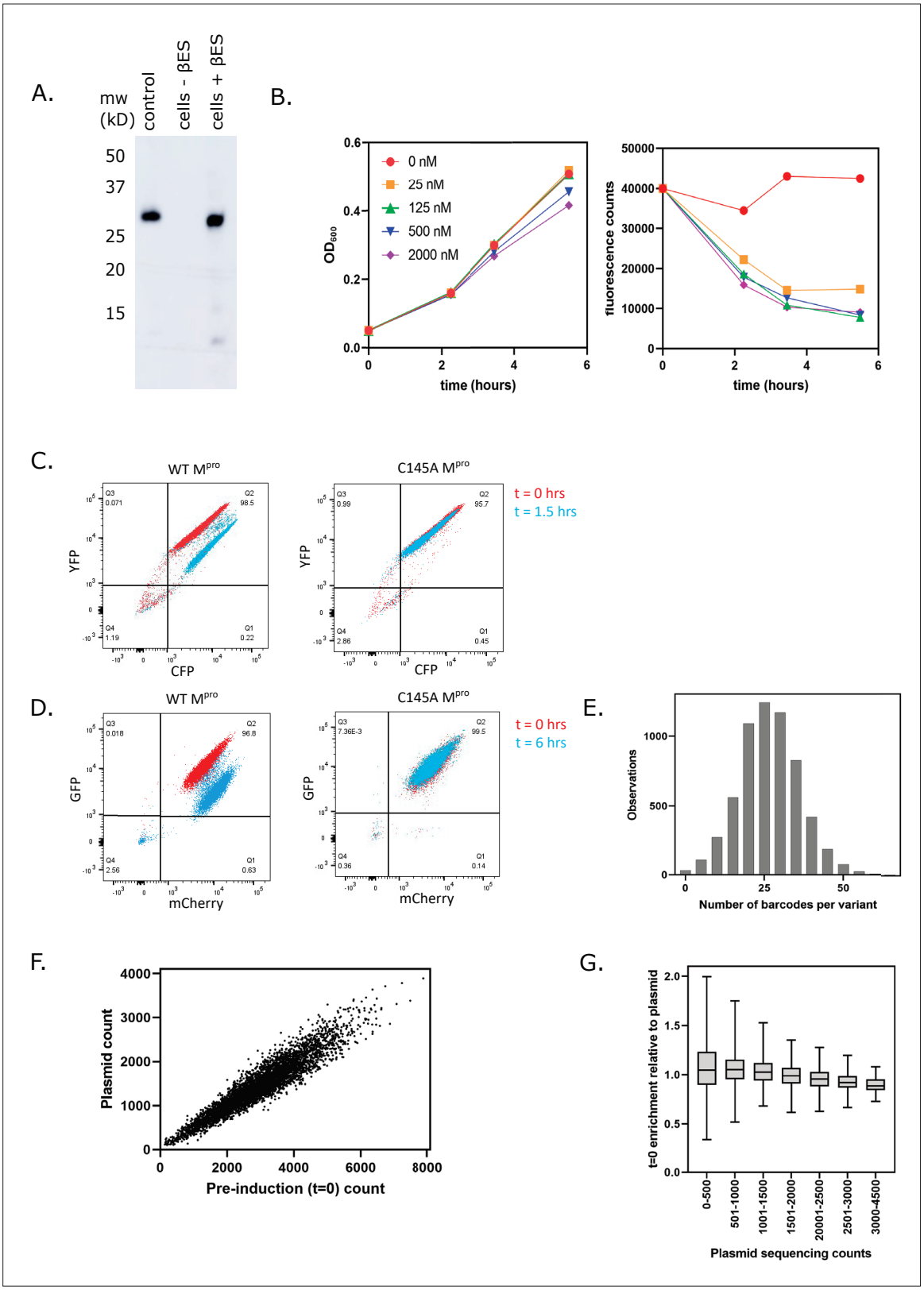


Figure 1—figure supplement 1. Main protease (M^{pro}) expression in cells harboring the LexA-UbM^{pro} plasmid construct. (**A**) Yeast cells transformed with a plasmid expressing C145A Ub-M^{pro}-his_6 under the LexA promoter were grown to exponential phase followed by the addition of 2 μM β-estradiol to induce expression for 8 hr. The M^{pro} levels were monitored by Western blot with an anti-his_6 antibody and the correct size was measured against purified M^{pro}-his_6 protein (control). (**B**) The plasmid expressing wild-type (WT) Ub-M^{pro} under control of the LexA promoter was transformed into cells expressing

Figure 1—figure supplement 1 continued on next page

Figure 1—figure supplement 1 continued

the split transcription factor. Cells were grown to exponential phase followed by addition of the indicated concentration of β-estradiol. Cell density was monitored based on absorbance at 600 nm at the times indicated (left panel). At the same time points, cells were washed, diluted to equal cell number, and GFP fluorescence was monitored at 525 nm (right panel). (**C**) Fluorescence-activated single cell sorting (FACS) analysis of cells expressing the CFP-M^{pro}CS-YFP FRET pair and either WT Ub-M^{pro} (left) or C145A Ub-M^{pro} (right). Cell samples were collected before and after induction of M^{pro} expression with 125 nM β-estradiol for 1.5 hr. (**D**) The FACS analysis of cells expressing the split transcription factor separated by the M^{pro} cut-site and either WT Ub-M^{pro} (left) or C145A Ub-M^{pro} (right). Cell samples were collected before and after induction of M^{pro} expression with 125 nM β-estradiol for 6 hr. (**E**) Distribution of number of barcodes associated with all M^{pro} variants. (**F**) Correlation between total counts of each variant in the M^{pro} plasmid library (plasmid count) and the total counts of that variant before M^{pro} induction (pre-induction count). (**G**) M^{pro} variants present at low frequency in the library showed a wider variance between plasmid library counts and counts in the pre-induction sample, consistent with lower sampling.

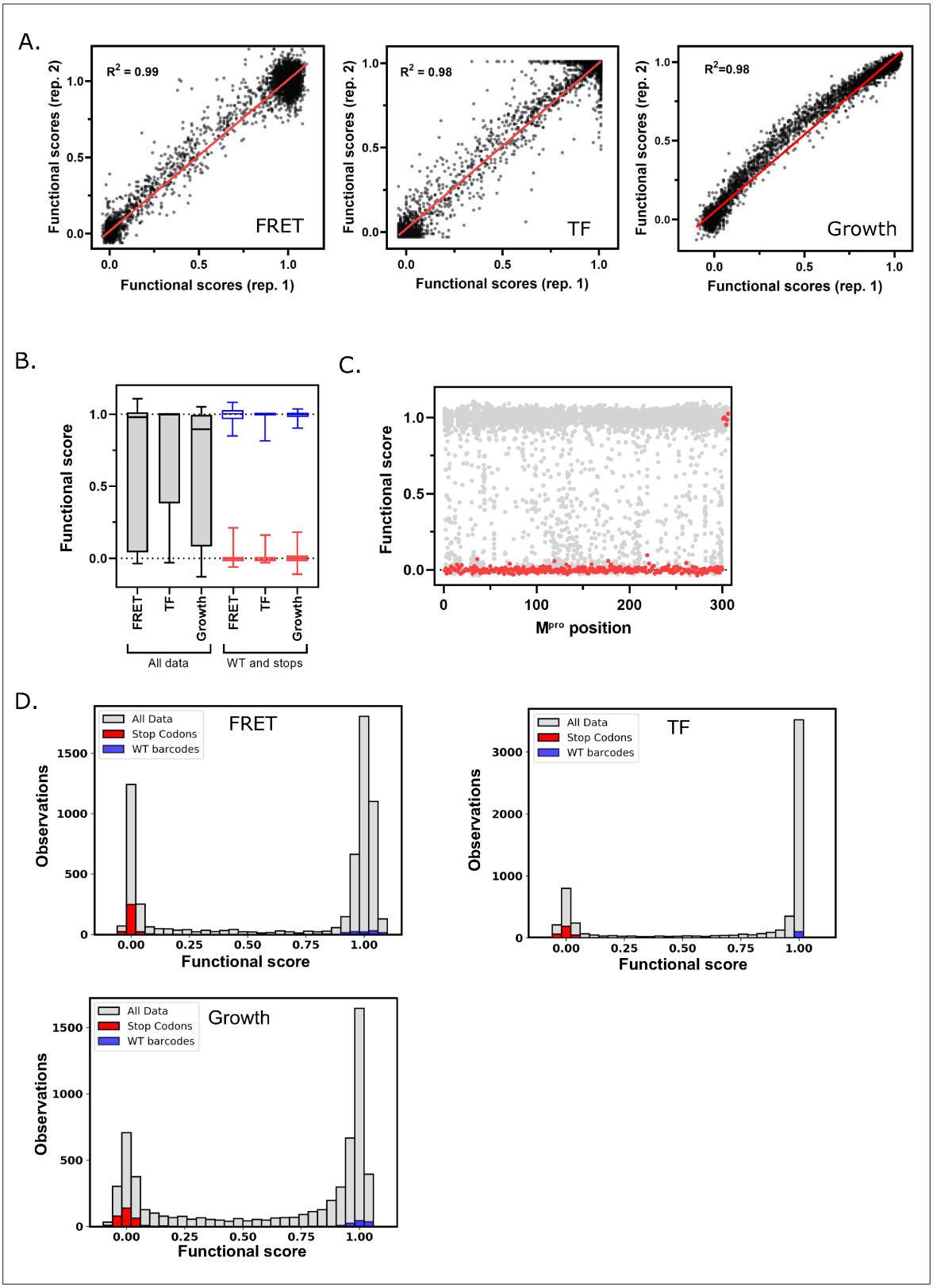


Figure 2. Main protease (M^{pro}) functional scores are reproducible, and variants can be clearly distinguished based on function. (**A**) Correlation between biological replicates of functional scores of all M^{pro} variants for each screen. Red line indicates best fit. (**B**) Distribution of functional scores for all variants (gray), stop codons (red), and wild-type (WT) barcodes (blue) in each screen. (**C**) The functional scores for all variants (gray) and stop codons (red) at each position of M^{pro} in the fluorescence resonance energy transfer (FRET) screen. (**D**) Distribution of all functional scores (gray) in each screen. Functional

Figure 2 continued on next page

Figure 2 continued

scores are categorized as WT-like, intermediate, or null based on the distribution of WT barcodes (blue) and stop codons (red) in each screen. See ***Figure 2—source data 1***.

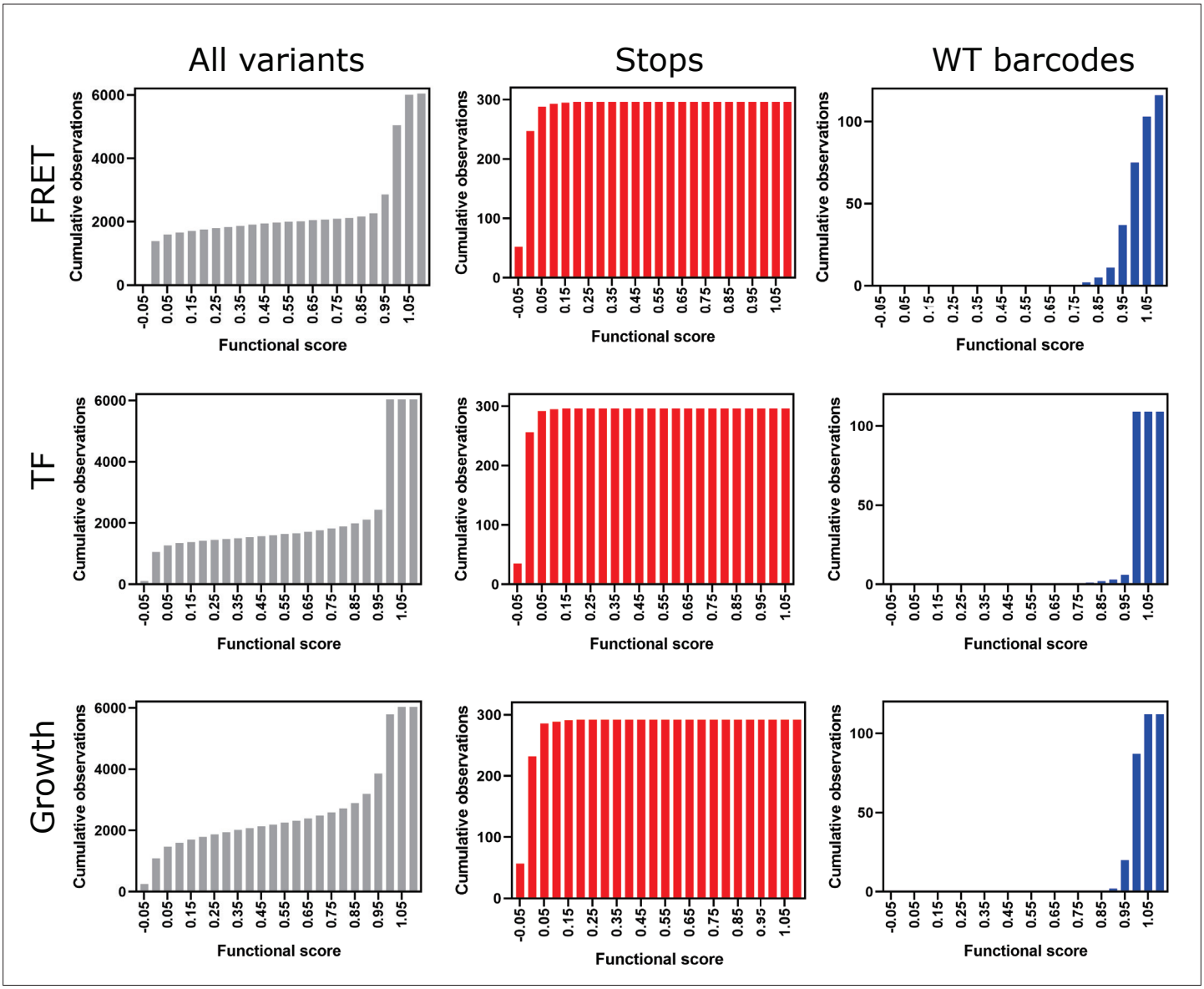


Figure 2—figure supplement 1. Cumulative frequency distributions for all variants (gray), stops (red), and wild-type (WT) barcodes (blue) for all three screens.

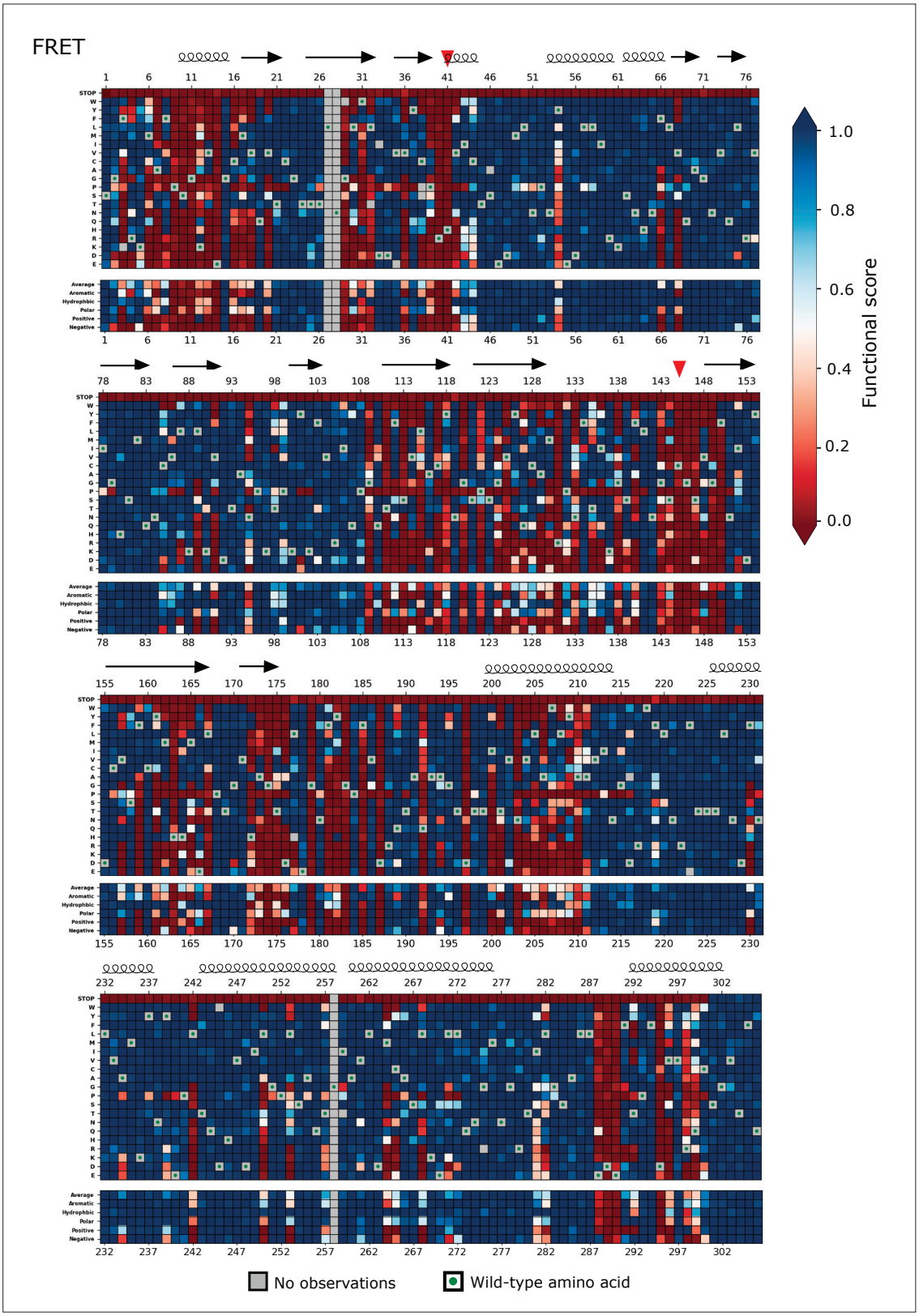


Figure 3. Heatmap representation of the main protease (M^{pro}) functional scores measured in the fluorescence resonance energy transfer (FRET) screen (replicate 1). Arrows represent positions that form β-sheets, coils represent α-helices, and red triangles indicate the catalytic dyad residues H41 and C145.

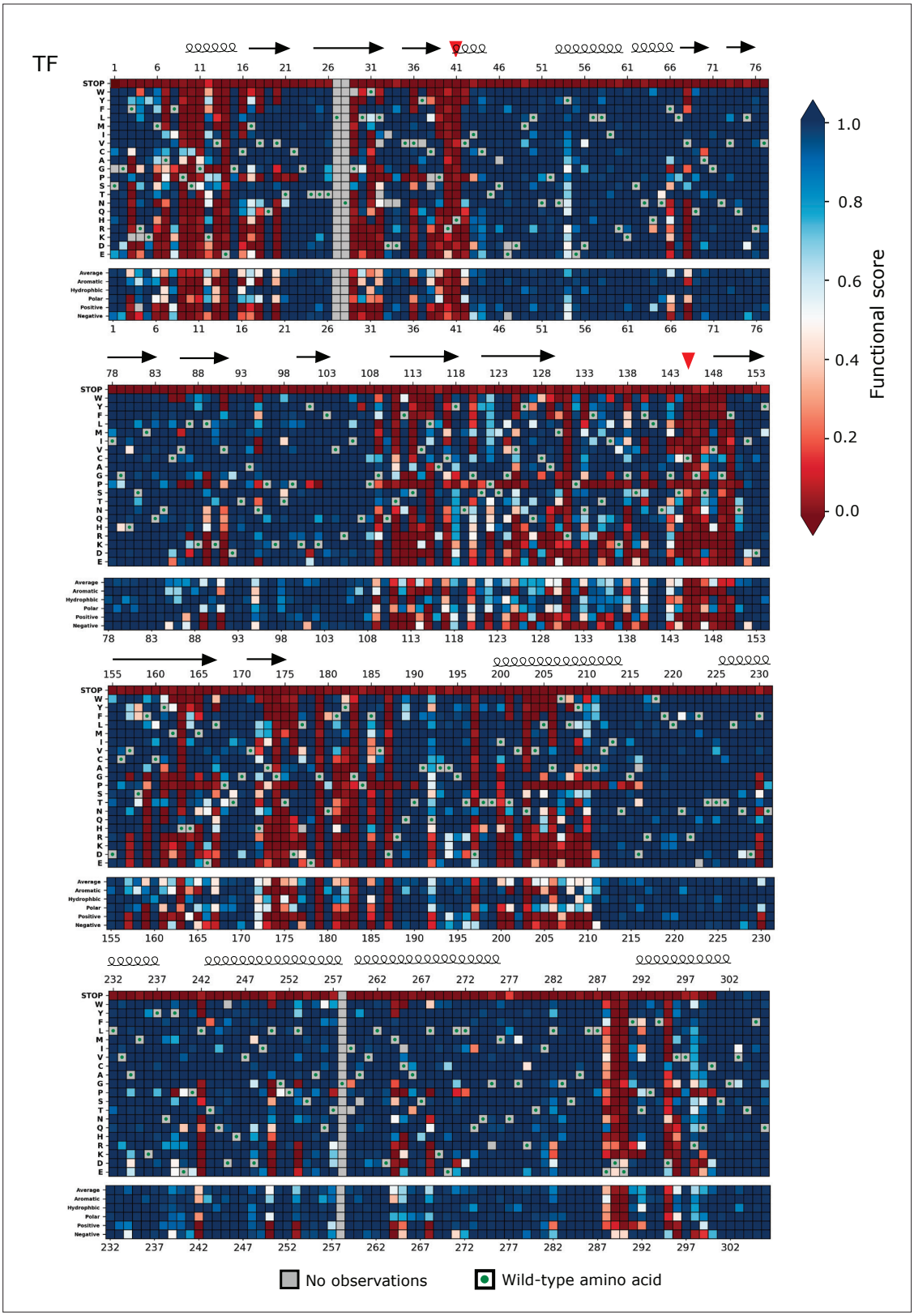


Figure 3—figure supplement 1. Heatmap representation of scores from the transcription factor (TF) screen (replicate 1). Arrows represent positions that form beta sheets, coils represent α-helices, and red triangles indicate the catalytic dyad residues H41 and C145.

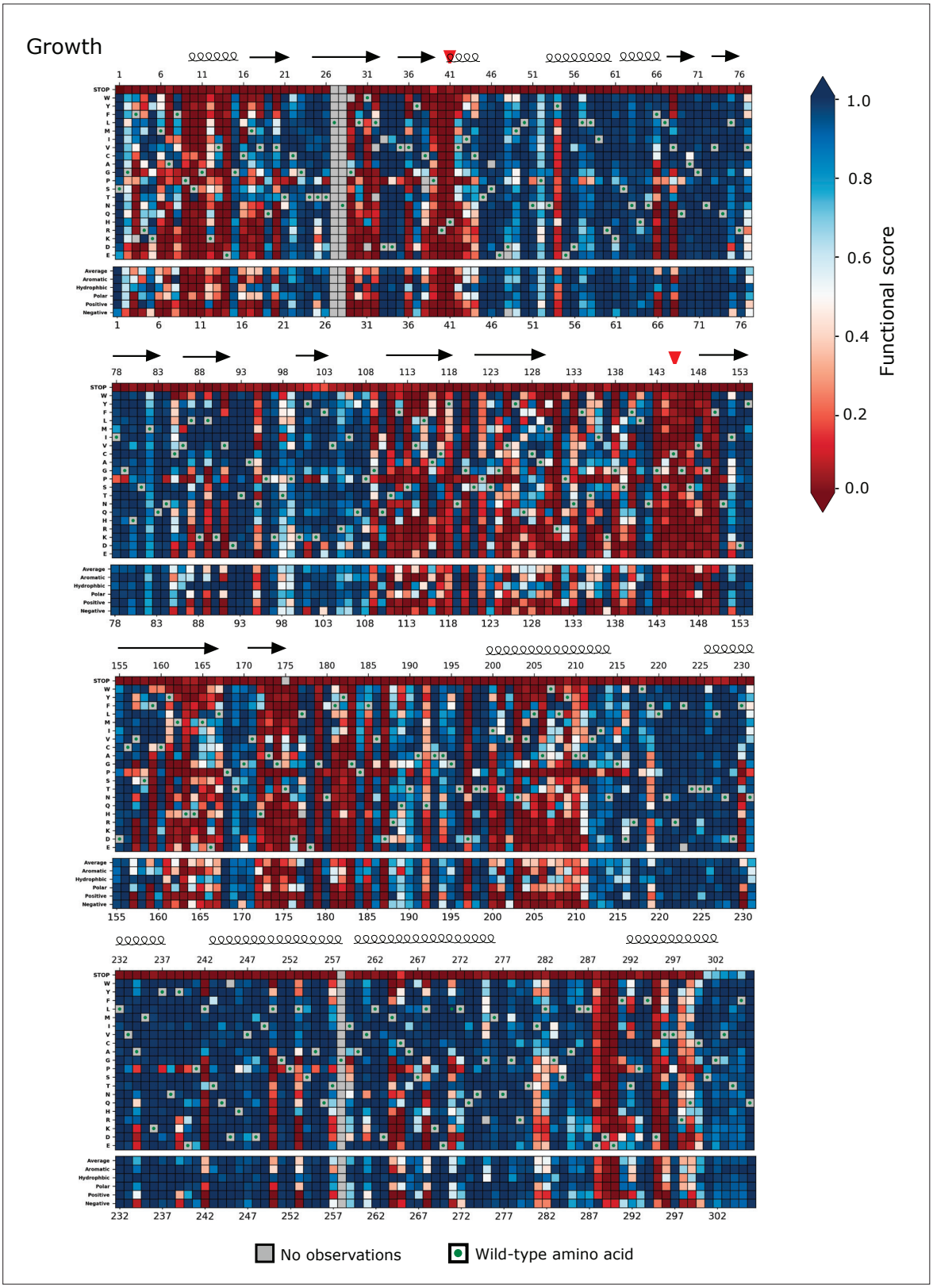


Figure 3—figure supplement 2. Heatmap representation of scores from the growth screen (replicate 1). Arrows represent positions that form beta sheets, coils represent α-helices, and red triangles indicate the catalytic dyad residues H41 and C145.

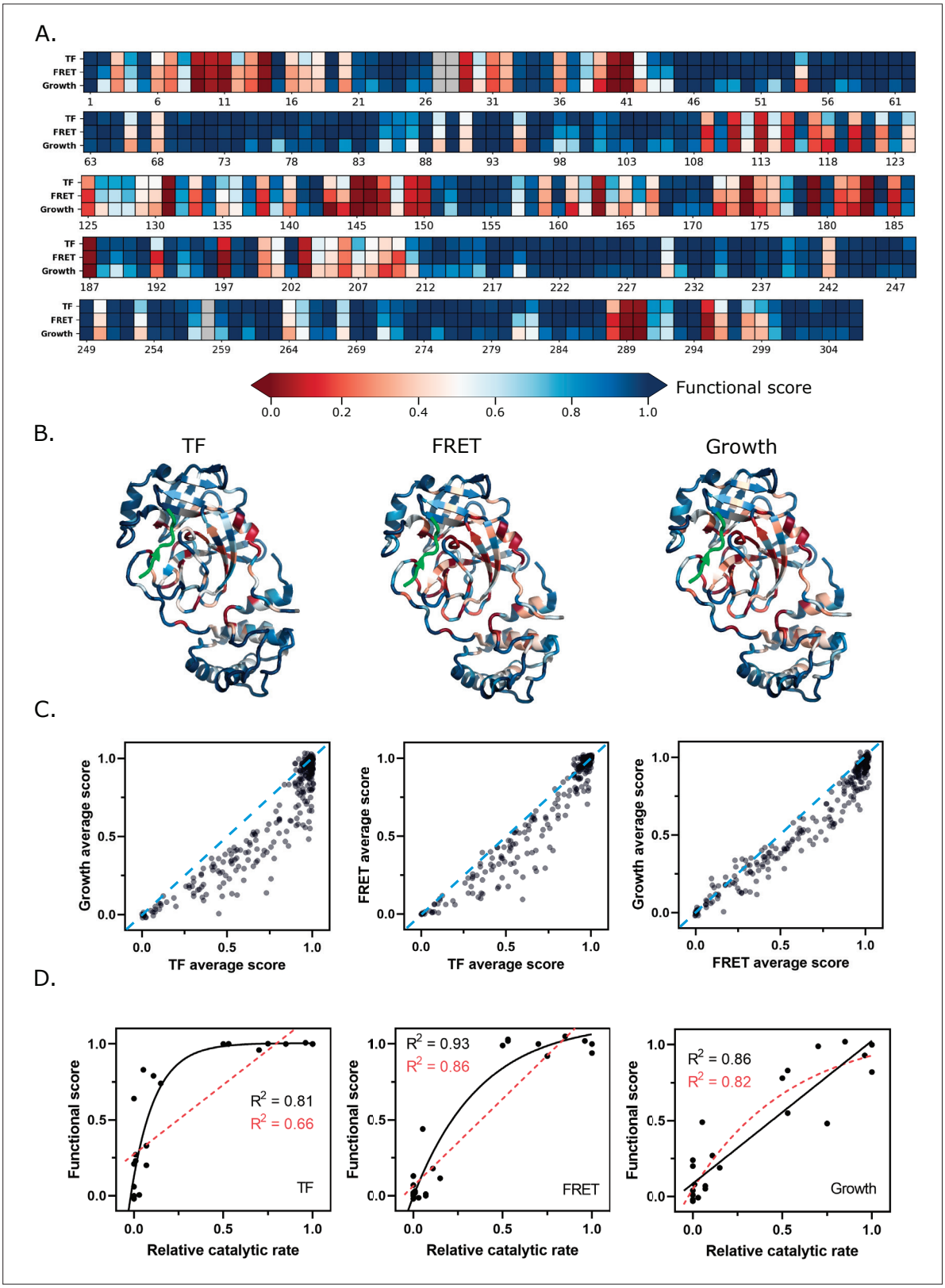


Figure 4. Functional scores reflect fundamental biophysical constraints of main protease (M^{pro}). (**A**) Heatmap representation of the average functional score at each position (excluding stops) in replicate 1 of each screen (see ***Figure 4—source data 1***). (**B**) The average functional score at each position mapped to M^{pro} structure for each screen. The Nsp4/5 substrate peptide is shown in green (PDB 7T70). (**C**) The average functional score at each position compared between the three screens. The diagonal is indicated with a blue dashed line. (**D**) Comparison between relative catalytic rates measured

Figure 4 continued on next page

Figure 4 continued

independently in various studies and functional scores measured in each screen (see ***Figure 4—source data 2***). Each graph is fit with a non-linear and linear regression with the best of the two fits represented with a black solid line and the worst fit represented with a red dashed line. The non-linear regression is fit to the equation $Y = Y_m - (Y_0 - Y_m)\, e^{-kx}$.

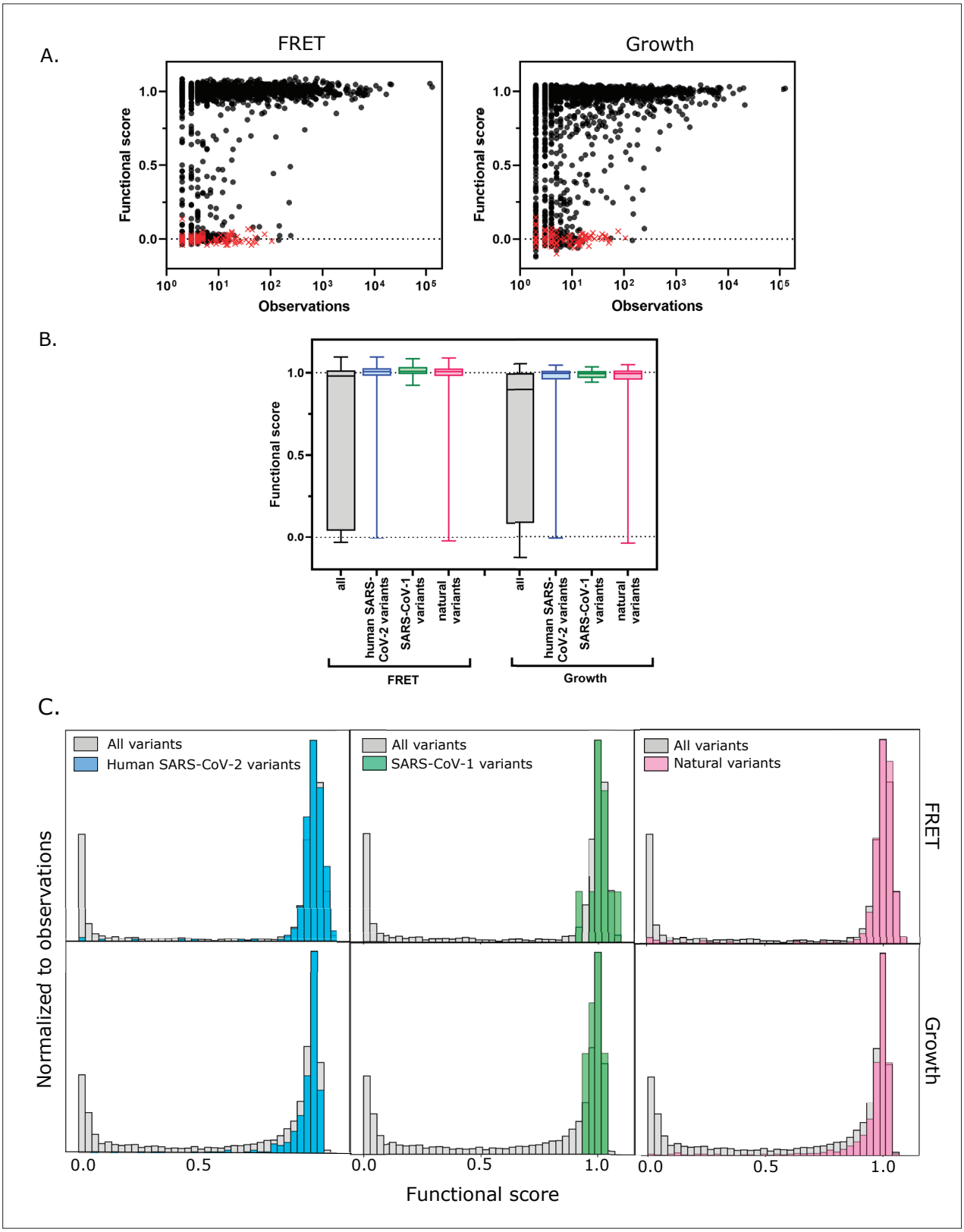


Figure 5. Functional scores indicate that natural amino acid variants of main protease (M^{pro}) are generally fit. (**A**) Comparison of functional scores in the FRET screen (left panel) and growth screen (right panel) to the number of observations among clinical samples. All missense mutations excluding stops are indicated with black circles and stop codons are indicated with red x's. (See ***Figure 5—source data 1***) (**B**) The distribution of functional scores of all variants in the FRET and growth screens compared to the observed clinically-relevant M^{pro} variants (human SARS-CoV-2 variants, blue), 12 amino

Figure 5 continued on next page

Figure 5 continued

acid differences between SARS-CoV-2 and SARS-CoV-1 (green), and the different amino acids in a broad sample of M^{pro} SARS-CoV-2 homologs (natural variants, pink). Distributions are significantly different as measured by a two-sample Kolmogorov-Smirnov (KS) (All FRET vs. human SARS-CoV-2 variants: N = 6044, 289, p<0.0001, D = 0.3258; All FRET vs. SARS-CoV-1 variants: N=6044, 12, p=0.0398, D=0.4223; All FRET vs. natural variants: N = 6044, 1205, p<0.0001, D = 0.2984; All Growth vs. human SARS-CoV-2 variants: N = 6044, 289, p<0.0001, D = 0.3938; All growth vs. SARS-CoV-1 variants: N=6044, 12, p=0.0024, D=0.5533; All growth vs. natural variants: N=6044,1205, p<0.0001, D = 0.3462) (**C**) Histogram of functional scores of all variants (grey) compared to that of human SARS-CoV-2 variants (blue), SARS-CoV-1 variants (green), and natural variants (pink).

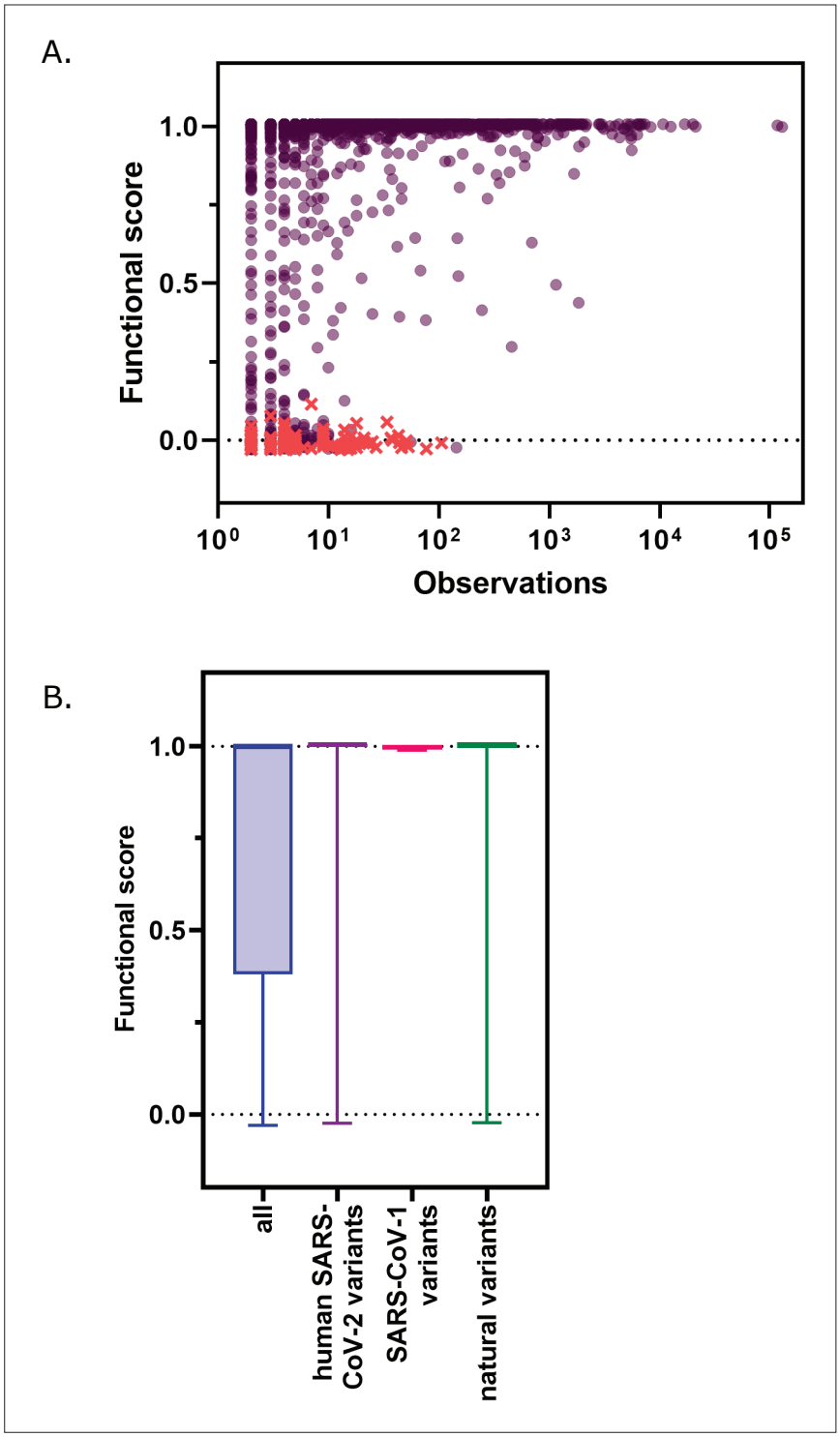


Figure 5—figure supplement 1. Functional scores indicate that natural amino acid variants of main protease (M^{pro}) are generally fit. (**A**) Comparison of functional scores in the transcription factor (TF) screen to the number of observations among clinical samples. All missense mutations excluding stops are indicated with black circles and stop codons are indicated with red x's. (See ***Figure 5—source data 1***) (**B**) The distribution of functional scores of all variants in the TF screen compared to the observed clinically relevant M^{pro} variants (human severe acute respiratory syndrome coronavirus-2 (SARS-CoV-2) variants, blue), 12 amino acid differences between SARS-CoV-2 and SARS-CoV-1 (green), and the different amino acids in a broad sample of M^{pro} SARS-CoV-2 homologs (natural variants, pink). Distributions are significantly different as measured by a two-sample Kolmogorov-Smirnov (KS) (all

Figure 5—figure supplement 1 continued on next page

Figure 5—figure supplement 1 continued

TF vs human SARS-CoV-2 variants: N=6038, 289, p<0.0001, D=0.2845; all TF vs SARS-CoV-1 variants: N=6038, 12, p=0.0196, D=0.4589; all TF vs natural variants: N=6038, 1205, p<0.0001, D=0.2608).

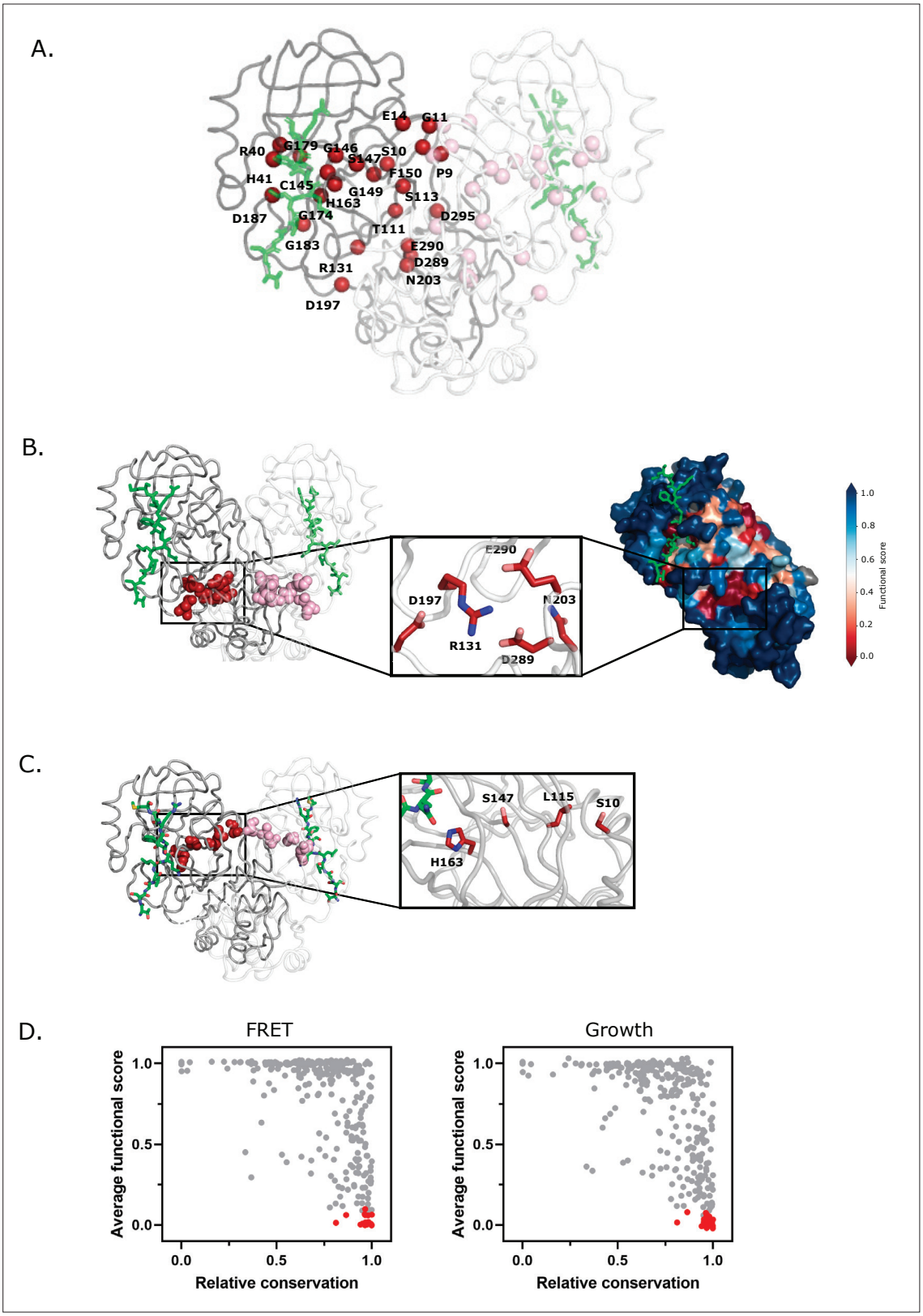


Figure 6. Structural distribution of main protease (M^{pro}) positions that are intolerant to mutation. (**A**) M^{pro} positions that are intolerant of mutations with 17 or more substitutions having null-like function are represented by red spheres on chain A (shown in gray) and pink spheres on chain B (shown in white). The Nsp4/5 substrate peptide is shown in green (PDB 7T70). (**B**) Representation of a cluster of the mutation-intolerant positions (red spheres) at a site distal to the active site. (**C**) A cluster of mutation-intolerant residues (red spheres) appear to be part of a communication network between the active

Figure 6 continued on next page

Figure 6 continued

site and the dimerization interface. (**D**) Comparison of the average functional score of each position to conservation observed in a broad sample of severe acute respiratory syndrome coronavirus-2 (SARS-CoV-2) M^{pro} homologs. The 24 mutation-intolerant positions shown as red spheres in part A are highlighted in red. Positions exhibiting the strongest evolutionary conservation exhibit a broad range of experimental sensitivity to mutation while the most evolutionary variable positions are experimentally tolerant to mutations.

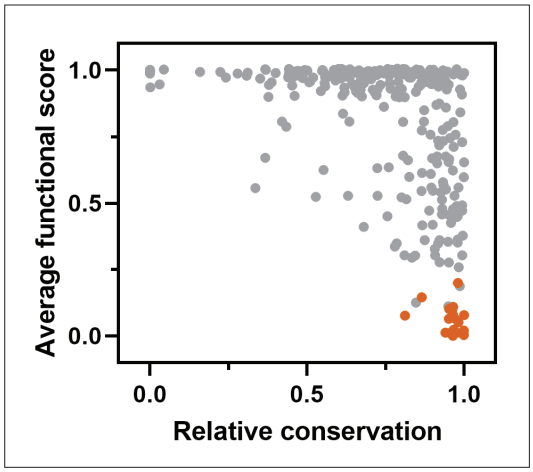
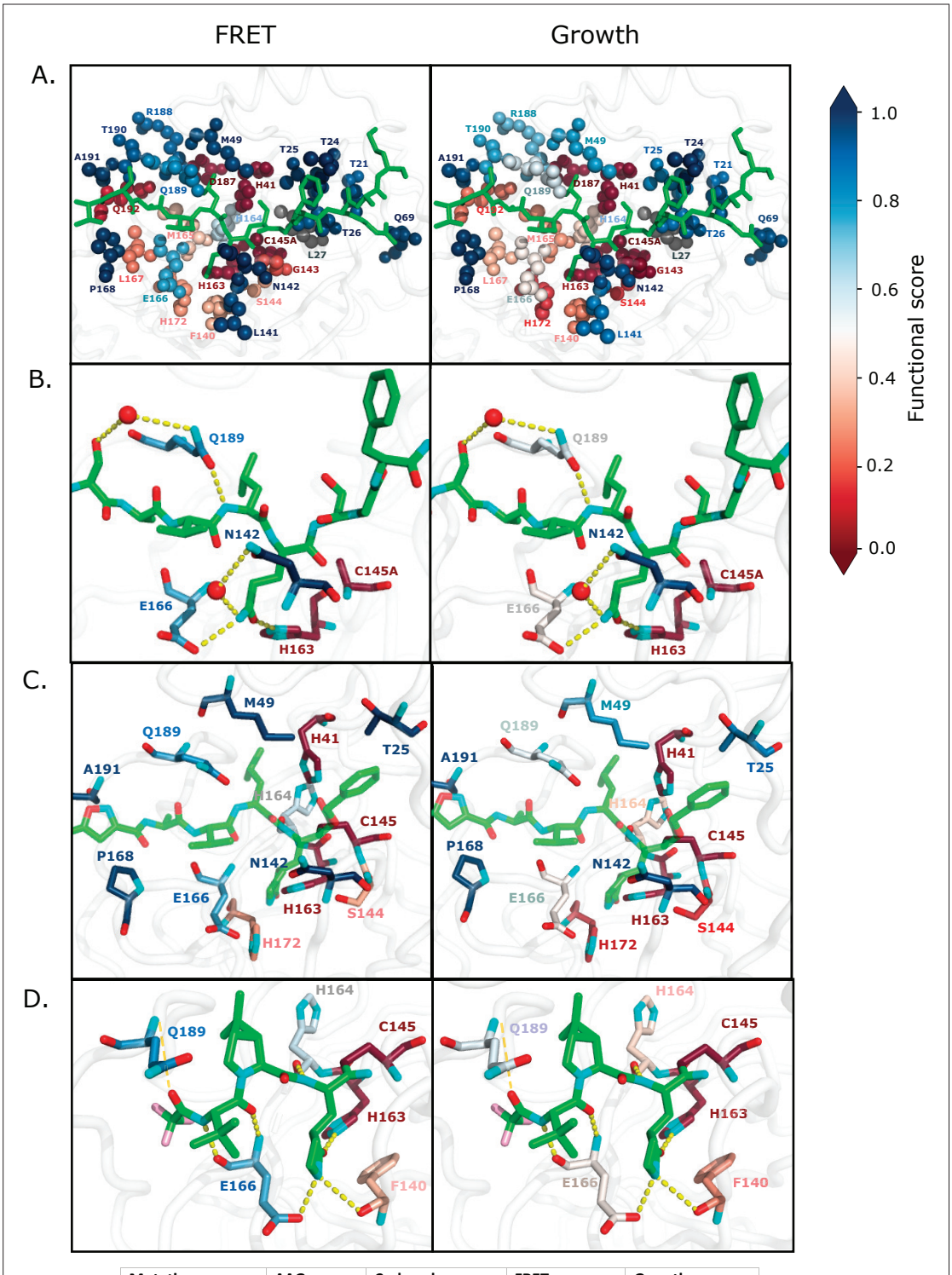


Figure 6—figure supplement 1. Comparison of the average transcription factor (TF) functional score of each position to conservation observed in a broad sample of severe acute respiratory syndrome coronavirus-2 (SARS-CoV-2) main protease (M^{pro}) homologs.



Mutation	ΔΔG	Codon change	FRET score	Growth score
Q189E	1.11	CAA -> GAA	1.04	1.0
E166A	1.0	GAA -> GCA	1.0	0.88
E166Q	1.0	GAA -> CAA	1.0	0.96

Figure 7. Substrate and inhibitor binding sites are variably sensitive to mutation. (**A**) All main protease (M^{pro}) positions that contact the Nsp4/5 substrate peptide are represented in spheres and colored by their average fluorescence resonance energy transfer (FRET) functional score (left panel) and growth functional score (right panel; PDB 7T70). The Nsp4/5 peptide is shown in green. (**B**) M^{pro} positions that form hydrogen bonds with the Nsp4/5 substrate are shown in sticks and colored by their average FRET functional score (left panel) and growth functional score (right panel; PDB 7T70). Oxygens are

Figure 7 continued on next page

Figure 7 continued

shown in red and nitrogens in cyan. Water molecules are represented as red spheres and hydrogen bonds as yellow dashed lines. (**C**) M^{pro} positions shown to contact over 185 inhibitors in crystal structures (***Cho et al., 2021***) are shown in sticks and are colored by their average FRET functional score (left panel) and average growth functional score (right panel). Shown is a representative structure of M^{pro} bound to the N3 inhibitor (PDB 6LU7) (***Jin et al., 2020***). The N3 inhibitor is shown in green, oxygens in red, and nitrogens in cyan. (**D**) M^{pro} positions that form hydrogen bonds with the Pfizer inhibitor, PF-07321332, are represented by sticks and colored by their average FRET functional score (left panel) or growth functional score (right panel; PDB 7VH8) (***Owen et al., 2021***; ***Zhao et al., 2021***). PF-07321332 is shown in green, oxygens in red, nitrogens in cyan, fluorines in pink. Hydrogen bonds less than 4 Å are represented with thick yellow dashed lines and greater than 4 Å with a thin yellow dashed line. The table below lists the mutations with highest potential for being resistant against PF-07321332.

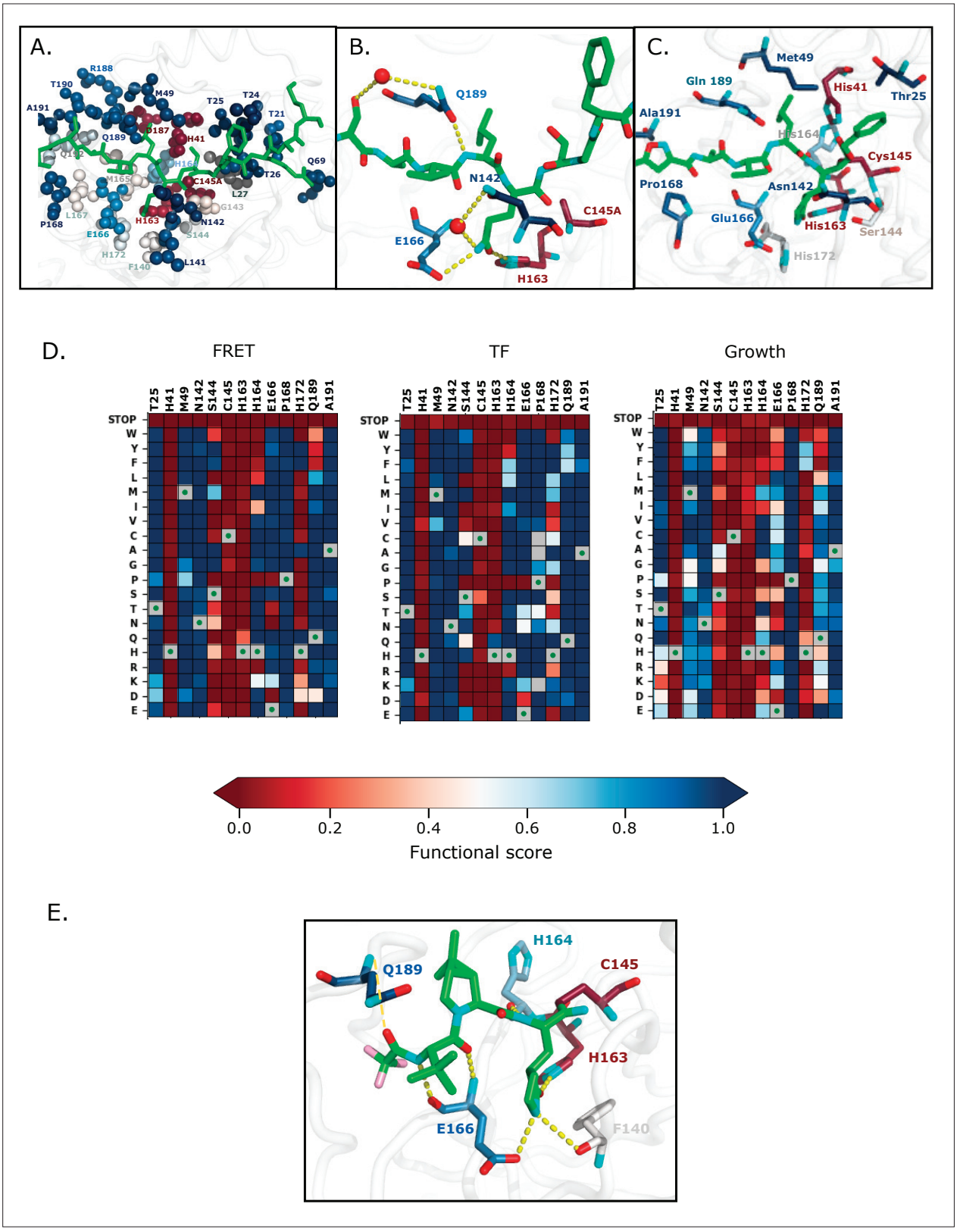


Figure 7—figure supplement 1. Substrate and inhibitor binding sites are variably sensitive to mutation. (**A**) All main protease (M^{pro}) positions that contact the Nsp4/5 substrate peptide are represented in spheres and colored by their average transcription factor (TF) functional score (PDB 7T70). The Nsp4/5 peptide is shown in green. (**B**) M^{pro} positions that form hydrogen bonds with the Nsp4/5 substrate are shown in sticks and colored by their average TF functional score (PDB 7T70). Oxygens are shown in red and nitrogens in cyan. Water molecules are represented as red spheres and

Figure 7—figure supplement 1 continued on next page

Figure 7—figure supplement 1 continued

hydrogen bonds as yellow dashed lines. (**C**) M^{pro} positions shown to contact over 185 inhibitors in crystal structures (***Cho et al., 2021***) are shown in sticks and are colored by their average TF functional score. Shown is a representative structure of M^{pro} bound to the N3 inhibitor (PDB 6LU7) (***Jin et al., 2020***). The N3 inhibitor is shown in green, oxygens in red, and nitrogens in cyan. (**D**) Heatmap representation of functional scores for the fluorescence resonance energy transfer (FRET) screen (left panel), TF screen (middle panel) and the growth screen (right panel) at key inhibitor-contact positions as illustrated in ***Figure 7c***. (**E**) M^{pro} positions that form hydrogen bonds with the Pfizer inhibitor, PF-07321332, are represented by sticks and colored by their average TF functional score (PDB 7VH8) (***Owen et al., 2021***; ***Zhao et al., 2021***). PF-07321332 is shown in green, oxygens in red, nitrogens in cyan. Hydrogen bonds less than 4 Å are represented with thick yellow dashed lines and greater than 4Å with a thin yellow dashed line.