Fitness and functional landscapes of the *E. coli* RNase III gene *rnc*

Ryan Weeks[1] and Marc Ostermeier[2]*

[1]Chemistry-Biology Interface Graduate Program

[2]Department of Chemical and Biomolecular Engineering, Johns Hopkins University, Baltimore, 21218, USA

* Corresponding author: Email: oster@jhu.edu

Abstract

How protein properties such as protein activity and protein essentiality affect the distribution of fitness effects (DFE) of mutations are important questions in protein evolution. Deep mutational scanning studies typically measure the effects of a comprehensive set of mutations on either protein activity or fitness. Our understanding of the underpinnings of the DFE would be enhanced by a comprehensive study of both for the same gene. Here, we compared the fitness effects and in vivo protein activity effects of ~4,500 missense mutations in the *E. coli rnc* gene. This gene encodes RNase III, a global regulator enzyme that cleaves diverse RNA substrates including precursor ribosomal RNA and various mRNAs including its own 5' untranslated region (5'UTR). We find that RNase III's ability to cleave dsRNA is the most important determinant of the fitness effects of *rnc* mutations. The DFE of RNase III was bimodal, with mutations centered around neutral and deleterious effects, consistent with previously reported DFE's of enzymes with a singular physiological role. Fitness was buffered to small effects on RNase III activity. The enzyme's RNase III domain (RIIID), which contains the RNase III signature motif and all active site residues, was more sensitive to mutation than its dsRNA binding domain (dsRBD), which is responsible for recognition and binding to dsRNA. Differential effects on fitness and functional scores for mutations at highly conserved residues G97, G99, and F188 suggest that these positions may be important for RNase III cleavage specificity.

Introduction

In the last ten years, comprehensive experiments on the effects of all single-codon missense mutations, called deep mutational scanning (DMS) (Fowler and Fields 2014), have been performed on a variety of proteins (e.g. Roscoe, et al. 2013; Firnberg, et al. 2014; Stiffler, et al. 2015; Mavor, et al. 2016; Wrenbeck, et al. 2017; Wu, et al. 2022). Many of these studies report on the mutational effects on protein properties. These studies often use non-native reporter assays, sometimes with the protein expressed in a non-native organism, to determine, for example, mutational effects on the affinity of an inhibitor against H1N1 influenza hemagglutinin (Whitehead, et al. 2012). These DMS studies show how changes to the amino acid sequence affect protein activity. Although these studies often are described as producing "fitness landscapes", these landscapes are not fitness landscapes in the traditional sense of the word "fitness" (i.e. the ability of an organism to survive and procreate). In contrast, traditional fitness landscapes describe the effect of mutations on the fitness of the organism (e.g. Lind, et al. 2010; Roscoe, et al. 2013; Firnberg, et al. 2014; Lind, et al. 2016; Mavor, et al. 2016; Lundin, et al. 2017; Wrenbeck, et al. 2017; Hartman and Tullman-Ercek 2019; Asgari, et al. 2020) and are thus more relevant for protein evolution. In this paper, we use the term "fitness landscape" in the traditional sense, referring to the mutational effects on the fitness of the organism. In our experiments, fitness is measured as the effect of

the mutation on the exponential growth rate of the bacteria. In contrast we use the term "functional landscape" when referring to mutational effects on protein activity. In our experiments, that protein activity is the level of protein activity in the cell, which depends on both specific protein activity and the abundance of active protein.

An important question in evolutionary biology is what causes the fitness effects of mutations. While it is expected that fitness will depend on the protein's properties and activity, the functional form of this dependence is expected to be nonlinear (in part because of buffering), has only been studied for a limited number of mutations (Bershtein, et al. 2012; Rodrigues, et al. 2016; Sarkisyan, et al. 2016; Thompson, et al. 2020), and will likely take different forms for different proteins and different environments. In addition, the collateral fitness effects of mutations (i.e. those effects that do not originate from changes in the ability of the gene to perform its physiological function) may play an important role (Mehlhoff, et al. 2020). DMS studies on fitness landscapes (e.g. Roscoe, et al. 2013; Firnberg, et al. 2014; Stiffler, et al. 2015; Mavor, et al. 2016; Mehlhoff, et al. 2020; Thompson, et al. 2020; Wu, et al. 2022) and functional landscapes (e.g. Bandaru, et al. 2017; Heredia, et al. 2018; Shah and Kuriyan 2019; Subramanian, et al. 2021) indicate that the distribution of fitness effects (DFE) of both types of landscapes is typically bimodal with peaks centered on fully deleterious and neutral effects. The fractions of mutations observed in each peak of the DFEs is a measure of the protein's mutational robustness, which has implications for a protein's evolution. For many proteins, most mutations are neutral (Roscoe, et al. 2013; Firnberg, et al. 2014; Stiffler, et al. 2015; Lind, et al. 2016; Mavor, et al. 2016; Lundin, et al. 2017), but for others most are deleterious (Lind, et al. 2016). Important factors causing such differences in DFEs include the level of gene expression and the environment in which fitness is measured (Stiffler, et al. 2015; Adams, et al. 2016; Mavor, et al. 2016; Lundin, et al. 2017; Noda-García, et al. 2019). When expression is sufficiently high, fitness effects can be buffered to mutational effects that do not decrease activity below an activity threshold, as observed with *S. enterica* HisA and *S. cerevisiae* Hsp90 (Hietpas, et al. 2011; Jiang, et al. 2013; Lundin, et al. 2017). Recent studies of *S. cerevisiae URA3* (Wu, et al. 2022) and *E. coli TEM-1* β-lactamase (Stiffler, et al. 2015; Mehlhoff, et al. 2020) indicate that a gene's essentiality in the growth environment also shapes the DFE, with mutations less deleterious to protein function being tolerated in environments where the gene is less essential for growth. Other properties, such as the role of the protein in the cell and the centrality of the protein to the cell's function may affect the DFE (Lind, et al. 2010; Melamed, et al. 2013; Roscoe, et al. 2013; Mavor, et al. 2016; Subramanian, et al. 2021).

We are not aware of any study reporting comprehensive fitness and functional landscapes for the same gene. Such a study would add to our overall understanding of how fitness depends on protein function. Here, we chose to determine and compare the fitness and functional landscapes of *E. coli*

ribonuclease III (RNase III). RNase III is a dsRNA-specific, Mg^{2+} dependent endonuclease that is involved in many processes in the cell including the maturing of rRNA and post-transcriptionally regulating numerous genes including its own (Dunn and Studier 1973; Gitelman and Apirion 1980; Sim, et al. 2010; Stead, et al. 2011; Court, et al. 2013; Sim, et al. 2014; Gordon, et al. 2017; Altuvia, et al. 2018). Members of the RNase III family are found in organisms from bacteria to humans and are characterized by the presence of a ten-residue signature motif and by their ability to cleave dsRNA. Although the functional roles of RNase III family members differ, sequence conservation remains high. Within the RNase III enzyme family, *E. coli* RNase III is one of the simplest enzymes. Despite its involvement in many processes, *E. coli* RNase III is not essential (Apirion and Watson 1974; Gegenheimer and Apirion 1975; Babitzke, et al. 1993; Hauk, et al. 2022), but cells lacking RNase III grow about 20-27% slower (Babitzke, et al. 1993). Monomers of RNase III homodimerize to create two identical active sites surrounding target dsRNA and catalyzes the two strand-cleavage reactions (Blaszczyk, et al. 2001; Gan, et al. 2006; Court, et al. 2013; Altuvia, et al. 2018). Each RNase III monomer is composed of an N-terminal RNase III domain (RIIID), which contains the RNase III signature motif and all active site residues, and a C-terminal dsRNA binding domain (dsRBD) which is responsible for recognition and binding to dsRNA (Court, et al. 2013). Acting on its most abundant substrate, RNase III regulates ribosome formation through maturation of rRNA transcribed from rRNA operons (Kaczanowska and Rydén-Aulin 2007; Deutscher 2009; Court, et al. 2013). RNase III also activates and represses gene expression through dsRNA cleavage of an array of different mRNA messages (Sim, et al. 2010; Stead, et al. 2011; Adkar, et al. 2012; Gordon, et al. 2017; Altuvia, et al. 2018). For activation, RNase III cleavage can remove barriers that block translation, such as freeing the ribosome binding site in *adhE* mRNA for complex with the ribosome (Aristarkhov, et al. 1996). Conversely, RNase III cleavage can make mRNA more susceptible to degradation, as is exemplified by the negative feedback that regulates RNase III activity in cells. When cellular RNase III levels are high, cleavage of its own 5' untranslated region (5'-UTR) leaves the *rnc* mRNA vulnerable to RNase E degradation thus downregulating RNase III levels (Bardwell, et al. 1989; Régnier and Grunberg-Manago 1989, 1990; Matsunaga, Dyer, et al. 1996; Matsunaga, Simons, et al. 1996; Nicholson 1999). In addition to its relevance for protein evolution, a comprehensive fitness and functional landscapes of RNase III would allow for a more complete picture of how the sequence of *E. coli* RNase III is related to its function, as no crystal structure exists of the *E. coli* enzyme.

Results and Discussion

Cells lacking a functional RNase III grow 25% slower. We previously demonstrated that when RNase III is placed under the pBAD promoter on a plasmid, RNase III is expressed even in the absence of

the promoter's inducer, arabinose (Hauk, et al. 2022). Under these leaky expression conditions, the cleavage pattern of the *rnc* 5'UTR (a native substrate) is similar to that when RNase III is chromosomally-expressed (Hauk, et al. 2022). Thus, we presume the RNase III expression levels were approximately physiological. In those experiments, the plasmid-encoded RNase III lacked its native 5'UTR. In our experiments here, we integrated the native *rnc* 5'-UTR in place of the pBAD 5'-UTR on pRnc-gGFP (Hauk, et al. 2022) to make pRnc2-gGFP (**fig. 1A**) to better reflect the native autoregulation of RNase III.

We first confirmed that expression of RNase III in this manner restored wildtype growth rates in cells lacking a functional *rnc* gene on the chromosome. We expressed wildtype RNase III, the nuclease-null RNase III E117K, and the empty vector (EV) in MG1693 and SK7622 (MG1693 *rnc*$^-$) cells. RNase III E117K lacks the ability to catalytically cleave dsRNA but maintains the ability to specifically bind with dsRNA (Dasgupta, et al. 1998; Sun and Nicholson 2001). We measured the growth rate at 37˚C of exponentially growing cells in liquid lysogeny broth (LB) media containing 0.2% glucose (a repressor of the pBAD promoter) and ampicillin (Amp) for plasmid maintenance (**supplementary fig. S1**, Supplementary Material online). The SK7622 cells expressing RNase III-E117K and EV grew approximately 25% slower than cells with chromosomally-expressed RNase III (P<0.0001, 0.0003 by unpaired Student's T-test for E117K and EV, respectfully). However, no significant difference in growth rate was observed between SK7622 cells expressing RNase III and MG1693 cells expressing any of the three constructs. As inducing higher expression levels of RNase III by adding typical amounts of arabinose results in irregular sized bacteria with a diminished growth rate (Hauk, et al. 2022), we interpret our no arabinose growth conditions as producing approximately physiological expression levels of RNase III. This is also consistent with the observation that *E. coli* growth rate is not affected by 0.1 to 10-times the endogenous levels of RNase III (Sim, et al. 2010). However, we did not quantify expression by measuring RNase III mRNA or protein levels.

Fitness landscape measured in growth competition experiments. We constructed a library of all single-codon substitutions in *rnc* on the pRnc2-gGFP plasmid in the *rnc*$^+$ cloning strain NEB5α. We constructed three sub-libraries corresponding to each third of the gene due to read-length limits of Illumina MiSeq DNA sequencing. All library experiments were performed in parallel using these sub-libraries. We used NEB5α as a library construction strain to avoid biasing the library by the fitness effects of RNase III mutations during library construction. We purified the library plasmid DNA, and deep sequencing revealed it consisted of nearly all single-codon amino acid mutations except for poor coverage at one position: I6. This plasmid library was transformed into SK7622 cells, which lack a functional RNase III gene. In two replica experiments, exponentially growing SK7622 cells containing the library were subject to a growth competition experiment in LB media for 10 generations (**fig. 1C**). We

determined the frequency of each *rnc* mutant by deep-sequencing at the beginning and end of the competition. From these frequencies, we determined relative fitness (*w*) values, corresponding to the mean growth rate of cells containing a mutant RNase III relative to the mean growth rate of cells containing alleles synonymous to wildtype (Mehlhoff and Ostermeier 2020). We utilize the frequency of wildtype synonymous alleles instead of the frequency of wildtype because wildtype synonyms occurred more frequently in the library and wildtype sequencing counts are more prone to being affected by the artifact of PCR template jumping (Pääbo, et al. 1990; Kebschull and Zador 2015) during the preparation of barcoded amplicons for deep-sequencing.

Curiously, although fitness values correlated between the two replica experiments ($\rho = 0.62$, $p < 10^{-15}$, **supplementary fig. S2**, Supplementary Material online), the mean fitness in replica 1 (0.952) was higher than that of replica 2 (0.923) and over twice as many mutants in replica two were deemed significantly deleterious at $P<0.01$ (48% vs 21%). A comparison of the two DFEs shows that a subset of mutants that were near neutral in replica 1 became slightly deleterious in replica 2 and those mutations deleterious in replica 1 became, on average, slightly more deleterious in replica 2 (**supplementary fig. S2**, Supplementary Material online). For unknown reasons, it appears that the RNase III activity required for wildtype growth rates was more stringent during replica 2. In replica 2 we observed that a wildtype reference monoculture grown at the same time as the libraries grew 3% faster and the libraries grew 1.1-1.7% faster than in replica 1 (**supplementary fig. S3**, Supplementary Material online), indicating that something was different about the growth conditions between the two replica experiments; perhaps the temperature was slightly higher in replica 2. We chose to present our fitness values as the weighted mean of the two replicas.

We obtained fitness data for 99.58% (4501/4520) of single-codon mutation in RNase III (**fig. 2A**). A total of 44.8% (2006/4479) of mutations nominally had fitness values under 0.98, with 54.6% being near-neutral (fitness values between 0.98 and 1.02). A total of 19.0% (P <0.01, Z-test) and 13.6% (P<0.001) of the mutations were deleterious in both replica experiments. As expected, regions of predicted secondary structure, namely the α-helices in the RIID and dsRBD domains, were more prone to deleterious mutations than loop regions between secondary structural elements. Mutations to the most exchangeable amino acid (Yampolsky and Stoltzfus 2005) tended to have smaller fitness effects with a median fitness of 0.994 ± 0.054 (**supplementary fig. S4**, Supplementary Material online).

Surprisingly, the distribution of fitness effects (DFE) showed an apparent unimodal distribution centered at neutral effects and a long tail of deleterious mutations (**fig. 3A**). Typically, DFE's are bimodal (Lind, et al. 2010; Peris, et al. 2010; Roscoe, et al. 2013; Firnberg, et al. 2014; Lind, et al. 2016; Lundin, et al. 2017; Mehlhoff and Ostermeier 2020), with one peak centered at neutral effects and the other corresponding to fitness values for an inactive protein (Lind, et al. 2010). Similar unimodal distributions

have been observed in ribosomal proteins shown to be highly robust to mutations (Lind, et al. 2010; Lind, et al. 2016). However, the tail of deleterious mutations was very broad, and we noticed that mutations with low fitness measurements tended to have lower than average sequencing counts at the start of the growth competition. These low counts resulted in a decreased ability to precisely measure their fitness effect as reflected in the uncertainty and *p*-values, which are inversely proportional to the sequencing counts. Furthermore, many nonsense mutations had fitness values well-above or well-below the expected value 0.75, the fitness value of cells lacking functional RNase III. Instead, nonsense mutations had a wide range of fitness values and a median fitness of 0.839 ± 0.109 (**supplementary fig. S4**, Supplementary Material online). Fitness values of nonsense mutations typically had large uncertainties stemming from low counts at the start of the experiment. We hypothesized that cells with deleterious RNase III mutations might not survive the transformation step as frequently, might form smaller colonies on the transformation plate, and would grow slower or have a longer lag phase in the pre-exponential incubation period prior to the growth competition phase of the experiment. This imprecision in the measurement of deleterious effects would broaden the expected peak corresponding to inactivating mutations, leading to a false appearance of a unimodal distribution. We sought another method to more accurately measure the fitness effects of mutations of deleterious mutations.

Deleterious fitness effects observed through mutant-dependent differences in transformation efficiency. We deep sequenced the NEB5α and SK7622 libraries to assess whether the transformation step negatively affected the frequency of deleterious mutants in the library (**fig. 1B**). From the frequencies of each mutant in each library, we determined transformation scores (w_t) corresponding to the log of the ratio of the enrichment of cells containing a mutant relative to the enrichment of cells containing alleles synonymous for the wildtype (**fig. 2B**). Transformation fitness scores are presented in log form, such that deleterious mutants with a score of -1 are ten-fold depleted in the SK7622 library relative to wildtype synonyms.

The results indicated that transformation of the library into SK7622 altered the frequencies of certain types of mutations. If transformation effectiveness was unaffected by RNase III function, one would expect that transformation fitness scores would show a random distribution centered about a value of zero regardless of RNase III function. However, this was not the case as nonsense mutations (excluding the last three codons) had a weighted mean transformation score of -1.36, but wildtype synonyms had a weighted mean transformation fitness score of 0.00 (**supplementary fig. S5**, Supplementary Material online). Nonsense mutation frequency diminished from 4.0% in the NEB5α library (close to the expected frequency of 4.7% (3/64)) to nearly tenfold less, representing only 0.55% of mutations in the SK7622 library. Conversely, residues mutated to those that are considered the most exchangeable were enriched from 4.3% in in our NEB5α library to 5.6% in the SK7622 library.

The distribution of transformation fitness scores was bimodal, with a neutral peak near 0 and a deleterious peak near -1.25 (**fig. 3B**). The landscapes of fitness and transformations scores looked qualitatively similar (**fig. 2A,B**), and fitness correlated with transformation scores ($\rho = 0.76$, $p < 10^{-15}$, **fig. 3D**). Thus, transformation scores likely report on many of the same fitness-impacting mutational effects that occur when the cells grow exponentially in liquid media. Transformation scores represent the relative ability of *rnc*$^-$ cells to acquire the plasmid-encoded mutant RNase III allele and grow on solid media. Presumably, mutant and wildtype plasmids are equally efficient at entering an *E. coli* cell. The difference in transformation scores thus must reflect how the RNase III mutation affects the ability of the cell to recover from the stress of electroporation and the rate at which the cell replicates on the plate (which impacts the size of the colony and thus the frequency at which the mutation is observed). We do not know whether recovering from the stress of electroporation is mutation-dependent, but we know that cells lacking a functional RNase III grow slower and thus should form smaller colonies. Thus, we view transformation score as a measure of fitness, albeit one in a different environment (liquid vs. solid media) using cells with a different internal state (i.e. exponentially growing cells vs. recently frozen cells prepared for electroporation). We conclude that the apparent absence of an obvious peak for loss of function mutations in the RNase III fitness landscape is just an artifact of the inability of the fitness measurements to precisely measure deleterious fitness effects because deleterious mutations were depleted prior to the experiment. For these reasons we view transformation score as a fitness measure that is more precise at measuring deleterious effects.

Interestingly, a sizable subset of mutants had deleterious transformation scores, but near neutral fitness values for exponential growth (**fig. 3D**). This difference may have several, non-mutually exclusive reasons. First, this difference may arise because certain functions of RNase III become more important for fitness during the stress of transformation or for growth on solid media. Second, the stress of transformation may make the cells more sensitive to small changes in RNase III catalytic activity, changes that are too small to have fitness consequences during exponential growth in liquid media. Finally, transformation scores might be more sensitive to small deleterious effects of mutations on RNase III activity because transcription of the RNase III gene starts after transformation of the plasmid into the cell. Thus, there would be a lag before the RNase III protein reaches steady state levels and can ameliorate the negative fitness effects experienced by the *rnc*$^-$ cells prior to transformation. Regardless of the mechanism(s), transformation scores are an important added dimension for understanding the fitness consequences of RNase III mutations in addition to being a more precise measure of the fitness effects of deleterious mutants.

Functional landscape for RNase III cleavage of a native substrate in vivo. How fitness depends on protein properties is an important question for the study of fitness landscapes. One difficulty

in addressing this question is the lack of methods for high-throughput measurements of protein activity in vivo. We developed a high throughput system that reports the ability of RNase III to cleave a native RNase III substrate: the 5' UTR of its own mRNA message (Hauk, et al. 2022) (upper right inset of **fig. 1**). This system utilizes an *E. coli* strain in which chromosomally-integrated GFP is transcriptionally-repressed by nuclease-null dCas9 when a GFP-targeting guide RNA (gGFP) is co-expressed (Qi, et al. 2013). Instead of using gGFP, we engineered a hybrid gRNA in which a fragment of the *rnc* 5'-UTR (nucleotides 1-129) is fused to the 5' end of gGFP (gUTR129GFP) (Hauk, et al. 2022). In the absence of RNase III, the gUTR129GFP is compromised in its ability to function with dCas9 to repress GFP expression, presumably by sterically blocking complexation with dCas9. However, RNase III cleavage of the gUTR129GFP causes release of the gGFP, which results in dCas9-mediated GFP repression. We showed that this fluorescent-reporter system allowed us to sort cells using flow cytometry based on RNase III activity (Hauk, et al. 2022). Cells lacking RNase III activity could be enriched from an excess of cells containing RNase III activity and vice versa.

Here, we use this system to measure a functional score for all mutations. This functional score reports on the cleavage of just one of the many substrates upon which RNase III acts: its own 5'UTR. Thus, functional scores for mutations might have been different if we had chosen an alternate RNase III substrate to fuse to the gRNA (provided the alternate system functioned as a reporter for that activity, which is not a given). We chose the 5'UTR of RNase III because it is the substrate that RNase III uses to self-regulate the level of RNase III in the cell, and thus is a strong candidate for a generally representative substrate.

We transformed the NEB5α library into the fluorescent-reporter, *rnc*$^{-}$ strain PH0919, which co-expresses dCas9 from a separate, compatible plasmid. The library was sorted into two populations by FACS: cells with higher fluorescence and cells with lower fluorescence (**fig. 1D**). The gates for selecting the two populations were determined based on the forward scatter versus fluorescence plots of cells expressing wildtype RNase III and cells expressing the E117K inactive mutant, as previously demonstrated (Hauk, et al. 2022). The frequency of each mutant in each population was determined by deep-sequencing. The enrichment of each mutant was defined as the ratio of the frequency of each mutant in the low-fluorescence, active population to its frequency in the high-fluorescence, inactive population. From these enrichments, we determined the functional scores (w_f) corresponding to the log of the ratio of the enrichment of cells containing a mutant relative to the enrichment of cells containing alleles synonymous for the wildtype.

Functional scores were calculated in two replica experiments, but the range of values in the first replica was larger than in the second (**supplementary fig. S6A-C**, Supplementary Material online). For example, non-amber nonsense mutations within the RIIID had a weighted mean functional score of -1.85

in the first replica and -0.80 in the second. We believe the difference in ranges might have resulted from differences in the length of time between transformation and the sorting experiment (i.e. cells had been frozen at -70°C for a longer time in the second replica), differences in the precision in placing the sorting gates, and differences in the efficiency of sorting. Although the two replicas had different ranges, the landscapes were qualitatively similar (**supplementary fig. S6A,D**, Supplementary Material online), so we present the functional scores as the weighted mean of the two replicas (**fig. 2C**). As observed for fitness values, deleterious mutants tended to have lower counts, which negatively impacted the precision of the functional scores of deleterious mutants. However, the problem was not as severe as in the fitness measurements because flow cytometry was performed on cells from the colonies on the transformation plates (i.e. no additional growth period occurred, unlike in the growth competition experiment).

The functional score represents how effective a mutant is at cleaving a native substrate, albeit in a non-native context. As observed with transformation fitness scores, the distribution for functional scores was bimodal, with peaks centered around -1.5 and 0 (**fig. 3C**). The bimodal nature of the distribution is best seen in each replica individually (**supplementary fig. S6B,C**, Supplementary Material online) as the averaging of the two replicas further broadened the lower peak beyond the broadening caused by the depletion of deleterious mutants during transformation. Functional scores positively correlated with fitness values ($\rho = 0.73$, $p < 10^{-15}$, **fig. 3E**). Similarly, mutations with low functional scores tended to have low transformation scores ($\rho = 0.70$, $p < 10^{-15}$, **fig. 3F**).

Fitness is buffered to small deleterious effects on cellular RNase III activity. We constructed ten mutants of RNase III to accurately quantify their fitness and functional effects. The ten mutations include six to active site residues: E38A, E38V, E38Q, E65P, D114G, and D114R (Blaszczyk, et al. 2001; Sun, et al. 2004; Zhang, et al. 2004; Gan, et al. 2006; Xiao, et al. 2009). Other variants include D155E at the end of the flexible linker, expected to sterically interfere with RNase III binding to dsRNA (Inada and Nakamura 1995; Gan, et al. 2006), as well as F188D and C192D within β-strands in the dsRBD. We replaced the start codon with TCT, which cannot function as an alternative start (Hecht, et al. 2017), for a negative "no RNase III" control. We subjected these ten mutants to more in-depth experiments performed at the individual variant level (**Table 1**).

We measured the growth rates of SK7622 cells expressing each mutant under conditions identical to those in which we measured the fitness values by DMS. We calculated the relative fitness of these cells by dividing their growth rate by the growth rate of cells expressing the unmutated RNase III. A comparison of these fitness values to their transformation scores further illustrated that transformation score is a reasonable proxy for fitness in liquid media (**fig. 4A**).

We also performed two RNase III functional assays on cultures of our exponentially-growing fluorescent reporter PH0919 cells expressing the mutants: 1) we measured the fluorescence of the cells,

and 2) we used a primer extension assay to determine the gUTR129GFP cleavage efficiency in vivo (**supplementary fig. S7**, Supplementary Material online). A comparison of the cleavage assay to the functional score (**fig. 4B**) and to the fluorescent assay (**fig. 4C**) indicated that these fluorescence-based functional assays do not detect reductions in gUTR129GFP cleavage until it is reduced below a certain threshold. Below this threshold, functional scores decrease rapidly, and fluorescent values increase rapidly. Fitness was also buffered against deleterious effects on the ability of RNase III to cleave its own 5'UTR (**fig. 4F**). This buffering is expected from the known feedback control mechanism through which RNase III regulates its own expression by cleaving its own mRNA message (Bardwell, et al. 1989; Matsunaga, Dyer, et al. 1996; Matsunaga, Simons, et al. 1996). Such buffering could help ensure that the cell has an excess of RNase III activity above that necessary for wildtype growth rate. While we do not expect that our expression level is precisely physiological, we do observe sufficient expression to fully rescue the growth deficiency caused by loss of RNase III (**supplementary fig. S1**, Supplementary Material online) but do not have the excess expression that would decrease growth rate (Sim, et al. 2010) and cause abnormal cell morphology (Hauk, et al. 2022). However, if our expression level is much higher than the native level, excess buffering of fitness to small functional effects may result.

Mutations at F188 differentially effect fitness and cleavage of the *rnc* 5'UTR. RNase III has many substrates in the cell including ribosomal RNA and those that directly or indirectly affect the mRNA levels of dozens of genes, and thus presumably affect their cognate protein expression levels (Sim, et al. 2010; Stead, et al. 2011; Gordon, et al. 2017). Mutations in RNase III thus have the potential to affect fitness through a variety of mechanisms. We developed our fluorescence-based, high-throughput functional score to interrogate the relationship between RNase III function and fitness. However, our functional score reports on the cleavage of just one RNase III substrate, the 5' UTR of its own mRNA. Despite using a single substrate, our functional score and fluorescence assay can explain most of the variation in fitness among our selected mutants (**fig. 4D,E**), and a positive correlation between functional score and fitness among all mutations is apparent (**fig. 3E**). This suggests that the 5'UTR of *rnc* is a good, representative RNase III substrate. Mutational changes in the rate of cleavage of the 5'UTR of *rnc* reflect changes in the rate of cleavage of the RNase III substrate(s) that impact fitness. Such an outcome with our chosen substrate was not guaranteed, as one can imagine specificity-altering mutations that differentially affect the cleavage of the 5'UTR of *rnc* and RNase III substrates important for fitness.

Visual comparison of the fitness, transformation, and functional landscapes (**fig. 2**) does not suggest that specificity-altering mutations are common, though limitations in the resolution of our measurements make it difficult to discover single substitutions that alter specificity or positions with small effects on RNase III specificity. However, our data suggests F188 is important for RNase III specificity. The fitness and transformation score effects of F188D were much more deleterious than

expected based on all three functional measures (**fig. 4D-F, H**). F188 is highly conserved and lies in a β-strand right after RBM2 (residues 181-186) only ~3.4 Å from the dsRNA in our structural model. Interestingly, no amino acid substitution at F188 had a significantly deleterious functional score, yet several substitutions at F188 are deleterious for both fitness and the transformation score, most notably mutations to Asp, Glu, Pro, and Gly (**fig. 4G,H**), which are four of the five least common amino acids in β-strands (Fujiwara, et al. 2012). This difference suggests that these F188 mutations, while not being very deleterious for cleavage of the 5'UTR of *rnc*, are more deleterious for cleavage of at least one other RNase III substrate, specifically one that is consequential for fitness.

RNase III's tolerance to mutation. A common way to quantify the sequence conservation of a protein family is through calculation of the Shannon entropy at each residue using a multiple sequence alignment of the family members (Shenkin, et al. 1991). The calculation makes use of the probability of finding each amino acid at that position, values which vary from 0 to 1. From this entropy each position's k^* value can be calculated, which is considered the average number of amino acids that are tolerated at a position. A k^* of 20 means that all 20 amino acids are equally probable, while a k^* of 1 means that only one amino acid is found at that position. These k^* values are a prediction of the mutational tolerance by position for the protein family based on what is observed naturally.

We have proposed an analogous calculation for position-dependent mutational tolerance of a single protein using the set of fitness values from DMS experiments, provided the fitness values range from 0 to 1 (Firnberg, et al. 2014, 2016). These k^* values are a measure of the actual mutational tolerance at each position for that protein. For RNase III, fitness values do not have a value of near zero for non-functional mutants – a necessity for calculating meaningful k^* values. However relative enrichment values used to calculate the transformation and functional scores span from zero to around 1. Thus, using the relative enrichment values we calculated k^* values for transformation and functional scores.

We compared these values to each other and to k^* values we determined from a multiple sequence alignment of 355 unique UniProt-curated RNase III family sequences (**fig. 5**). The distribution of k^* values indicate that many positions in *E. coli* RNase III have a high degree of mutational tolerance, with more than 50% tolerating 15 or more different amino acids (**fig. 5A,B**). We find that if a position is relatively variable in the RNase III family, the position is almost always highly tolerant to mutation in the *E. coli* protein (**fig. 5C**). However, the inverse is not necessarily true. While key active site residues have low k^* values in all calculations, several positions that are largely invariant in nature can tolerate many mutations in the *E. coli* protein. Several factors are likely contributing to this difference. Our k^* values report tolerance specifically for E. coli RNase III, not the protein family collectively. Natural diversity is a product of the evolution of genetically diverse populations over much longer time-scales and is subject to stochasticity (Boucher, et al. 2016). In addition, our measure of fitness is for a very narrowly defined

set of conditions, whereas fitness in nature is a complicated function of many factors (e.g. growth rate, survival, etc.) under many environments and internal states of the organism (i.e. perhaps mutations appearing tolerant in our data have greater fitness relevance under different environmental conditions). Finally, mutations with deleterious fitness effects smaller than can be measured in our experiments will be purged in natural evolution (Boucher, et al. 2016).

We find that k^* values calculated from transformation and functional scores correlate linearly (R^2 = 0.82), further illustrating how the cellular level of catalytic activity is the major determinant of fitness (**fig. 5D**). Interestingly, positions G97 and G99 in RBM4 were strong outliers of this trend. Except for mutations to small residues, which are near neutral for function and fitness at this position, most mutations at G97 and G99 cause a much smaller reduction in transformation scores than would be predicted from their large deleterious effect on functional scores (**fig. 5D**). This result suggests that although cleavage of the 5'UTR of *rnc* strongly requires small amino acids at G97 and G99, cleavage of RNase III substrates that are important for fitness are not nearly so negatively impacted by mutations to large residues at these positions despite both G97 and G99 being highly conserved in the RNase III family (k^* values 1.9 and 2.1, respectively). This result predicts that mutations to non-small amino acids at G97 and G99 substantially impact the sequence specificity of RNase III.

RNase III fitness and functional landscapes in the context RNase III's known sequence-structure-function relationship. The RNase III monomer can be divided into three regions: the RIIID, dsRBD, and the flexible linker that connects these two regions. RNase III's positional tolerance to mutation, as measured by k^* values for transformation score, are broadly consistent with the structural model of RNase III (**fig. 5E**). Positions less tolerant to mutation tend to fall at the protein-RNA and dimer interfaces or in the interior of the protein, with surfaces residues tending to be more tolerant. Within the RIIID are all the active site residues responsible for cleavage of dsRNA as well as all the points of contact between the two monomers. Each RNase III active site is composed of six negatively charged residues that contribute to positioning and cleavage of the dsRNA: E38, E41, D45, E65, D114, and E117 (Blaszczyk, et al. 2001; Gan, et al. 2006; Gan, et al. 2008; Court, et al. 2013). We find very low tolerance for mutations at these positions except for D114 (**fig. 5F** and **supplementary fig. S9**, Supplementary Material online). At positions E41, D45, and E117, no mutation is tolerated confirming previous reports that these residues are essential for RNase III activity (Inada and Nakamura 1995; Dasgupta, et al. 1998; Blaszczyk, et al. 2001; Sun and Nicholson 2001; Sun, et al. 2004; Zhang, et al. 2004). Three of these residues are in RNase III's signature motif of ten residues (residues 38-47). The motif is highly conserved across the RNase III family of proteins (Mian 1997; MacRae and Doudna 2007; Court, et al. 2013; Nicholson 2014) and generally intolerant to mutations in this study (**fig. 5F** and **supplementary fig. S9**, Supplementary Material online).

Assisting the six negatively charged active site residues are two RNA binding motifs (RBM) in the RIIID that help position the scissile bond at the correct positions for cleavage (Court, et al. 2013). RBM3 shows little tolerance to mutations except to like residues and those found in RNase III family enzymes in other organisms. Binding occurs between RBM3 and dsRNA when S69 forms a hydrogen bond with the RNA at the cleavage site and the side chain of E65 forms a hydrogen bond within the dsRNA major groove (Gan, et al. 2006). Most mutations at E65 are very deleterious, but quite surprisingly E65P is neutral for fitness and among the least deleterious for the functional score. RBM4 was much more tolerant to mutations than RBM3 (**fig. 5F**; median values 0.99±0.01 versus 0.90±0.11, $p<$ 0.0001, Wilcoxon rank-sum test). RMB3's greater sensitivity to mutation is consistent with its active site proximity and greater number of contacts with the dsRNA. The most severe mutations in RBM4 were those to G97 and G99, but only for the functional score as described previously.

The dsRBD was more robust to mutations than the RIIID in each of the landscapes. The dsRBD is composed of two α-helices and a β-sheet made up of three antiparallel strands. Most of the deleterious mutations in the dsRBD were within the two α-helices, whereas the β-sheets were less affected by mutation. The two RNA binding motifs within the dsRBD are found in the first α-helix (RBM1) of the domain and the loop connecting the β1 and β2 strands (RBM2). RBM1 was much less tolerant to mutation than RBM2 (**fig. 5F**), as might be expected (Gan, et al. 2006) due to its cleavage site proximity and larger number of dsRNA interactions. Residues in RBM2 interact with two bases within the minor groove of the RNA and the side chain of H184 makes one hydrogen bond with an O2' hydroxyl (Gan, et al. 2006) but these interactions must be of minor importance for function and fitness based on RBM2's high mutational tolerance.

The dsRBD is essential for activity in vivo (Nicholson 1999; Sun, et al. 2001), although RNase III lacking the dsRBD is capable of dsRNA-specific cleavage in vitro in non-physiological salt concentrations (Sun, et al. 2001). Our DMS data indicated that the dsRBD is essential as the region was intolerant to nonsense mutations, except for the final three residues. To confirm this result, we created a ΔdsRBD mutant through triple mutation of Q153*/K154*/D155* to test the effects of removing the dsRBD (**Table 1**). We found that deleting the dsRBD reduced fitness to 0.75 ± 0.01, equivalent to cells lacking RNase III. The fluorescent assay and cleavage assay indicated Q153*/K154*/D155* was severely impaired in cleaving the *rnc* 5'UTR, thus confirming that the dsRBD is required for in vivo cleavage activity.

The flexible linker region connecting the RIIID and dsRBD (residues 149 – 155) was fairly tolerant to mutation except for D155, the most highly conserved linker positions (**fig. 5F**). Within the linker, mutations to proline, especially at G150, were the most deleterious. Based on structural models of *T. maritima* RNase III with and without dsRNA, free rotation is observed around the flexible linker until

both the RIIID and dsRBD have complexed with the dsRNA (Gan, et al. 2005, 2006). Introduction of a rigid proline with the linker, particularly at G150, likely interferes with the linker adopting the necessary conformation to orient the two domains correctly.

Conclusions. Our study represents the first comprehensive report on both the fitness and in vivo functional effects of missense mutants to the same gene. We developed three landscapes each reporting on different aspects of the effect of RNase III mutations. Our fitness landscape shows how a mutation affects the growth rate of cells during exponential growth in liquid media, whereas the transformation score fitness landscape shows how each mutation affects the fitness of cells during plasmid transformation and subsequent growth on solid media. The functional landscape shows how each mutation affects the cellular level of RNase III catalytic activity on its own 5'UTR but does not directly measure effects on the many other RNase III activities in the cell. Despite this limitation, both fitness and transformation scores strongly correlate with functional scores, indicating that the ability of RNase III to properly cleave dsRNA is the major fitness determinant for this gene.

The distribution of fitness effects (as measured by the transformation score) and functional effects were both bimodal, with the largest fraction of mutations near neutral, but a sizable fraction of mutations causing a fitness effect approximately equivalent to that of loss of RNase III function. We observed that key residues including the signature motif, active sites residues, RBM1, and RBM3 were largely intolerant to mutation in each of our landscapes. A comparison of the RNase III family's sequence diversity and *E. coli* RNase III's mutational tolerance (**fig. 5**) showed that positions in *E. coli* RNase III that were intolerant to mutation lacked sequence diversity in the RNase III family. Conversely, however, some regions that lacked diversity in nature were highly tolerant to mutation in *E. coli* RNase III, most notably RBM4 and RBM2 and the β-strands in the dsRBD. Several factors may contribute to this difference (Boucher, et al. 2016). We identified three residues (G97, G99, and F188) at which mutations have significantly different effects on functional score and fitness. We speculate that these mutations having different effects on the cleavage of the 5'UTR of *rnc* and the cleavage of RNase III substrate(s) important for fitness (i.e. the mutations impact RNase III substrate specificity), but this hypothesis needs to be tested experimentally

Materials and Methods

***E. coli* strains and growth conditions.** Strain NEB5α (*fhu2 (argF-lacZ)U169 phoA glnV44 80 (lacZ)M15 gyrA96 recA1 relA1 endA1 thi-1 hsdR17)*(C2987) was purchased from New England Biolabs. Strain MG1693 (*E. coli* MG1655: K12 F^- lambda$^-$ *ilvG*$^-$ *rfb*-50 *rph*-1 *thyA715)* and SK7622 (MG1693 Δ*rnc-38* Km^R) were a gift from Sidney Kushner. The PH0919 strain was a gift from Pricila Hauk.

Except where noted, all lysogeny broth (LB) media (10 g/L tryptone, 5 g/L yeast extract, 10 g/L NaCl) was supplemented with 50 µg/mL thymine, 100 µg/mL ampicillin to maintain pRnc2-gRNA, and 50 µg/mL chloramphenicol to maintain pdCas9. Overnight cultures were grown in test tubes containing 10 mL LB-media supplemented with appropriate antibiotics shaking at 250 rpm at 37˚C for 16 h. Overnight cultures of *rnc*⁻ strains were grown overnight supplemented with 25 µg/mL kanamycin and transferred to media without kanamycin for experimentation. Growth experiments were performed in 500 mL baffled shake flasks containing 100 mL LB-media supplemented as above and with 0.2% w/v glucose shaking at 250 rpm at 37˚C.

Vectors. pRnc2-gRNA (#Addgene #189563) was made by inserting the Phusion High Fidelity polymerase (New England Biolabs) amplified *rnc* 5'-UTR (nucleotides 1-160) in place of the pBAD 5'-UTR on pRnc-gGFP129 (Hauk, et al. 2022) by blunt ligation. pdCas9 (#44249) was purchased from Addgene.

Monoculture growth assay for *rnc*⁻ growth rate. MG1693 and SK7622 (*rnc* ⁻*)* cells harboring pRnc2-gGFP, pRnc2-GFP with mutant *rnc* alleles, or the empty vector (EV) were grown overnight and diluted to an (optical density) OD_{600} of 0.05 in unbaffled shaker flasks. OD_{600} was monitored until cells reached the exponential growth phase (~OD_{600} of 0.5) and then diluted to an OD_{600} of 0.02 in baffled shaker flasks containing pre-warmed to 37˚C LB-media. Cells were grown for approximately 10 generations; OD_{600} was measured for every 30 min for approximately 6 h with a 3.1 ml/96.9 mL dilution in new pre-warmed to 37˚C LB-media at about 3 h.

Fitness was determined as the ratio of the mean growth rate of cells expressing the *rnc* mutant to the mean growth rate of cells expressing the wildtype *rnc*. The mean growth rate was calculated from the OD_{600} data as a function of time between the first time point (30 min) and final timepoint (after ~ 10 generations of growth). The growth rate was the slope of the linear fit of $\ln(OD_{600})$ vs. time.

Library creation. Libraries of all possible single-codon substitutions in *rnc* were constructed essentially as described (Mehlhoff, et al. 2020); the protocol is described in detail here. All single-codon substitutions were generated by inverse PCR with Phusion High Fidelity polymerase in HF Buffer (New England Biolabs) using degenerative codon NNN targeted to each codon of the *rnc* gene. Amplification at each codon was verified by gel electrophoresis on a 1% TAE-agarose gel. Inverse PCR and gel verification was repeated for any codons lacking prominent bands corresponding to the size of pRNc2-gRNA (4.8 kb). Successfully amplified vectors were pooled together based on the relative intensity of bands. Plasmids were pooled in three sublibraries – one for each third of the gene (region 1, residues 1-75; region 2, residues 76-150,; region 3, residues 151-226). Sublibraries were used due to read constraints of Illumina MiSeq. Pooled DNA was separated on a 1% TAE-agarose gel and desired band extracted using PureLink Quick Gel Extraction Kit (Invitrogen). The extracted DNA was cleaned and concentrated

using a DNA Clean & Concentrator Kit (Zymo) and phosphorylated using T4 polynucleotide Kinase (New England Biolabs). The phosphorylated DNA was cleaned and concentrated again as above and then ligated using T4 DNA Ligase (New England Biolabs). The ligation products were transformed into NEB5α electrocompetent cells and plated on LB-agar bioassay dishes supplemented with 50 μg/mL ampicillin and 2% w/v glucose. We estimated that a target of 22,000 transformants was sufficient to constitute a library that was about 99% complete (Bosley and Ostermeier 2005) and repeated the transformation step until the number of transformants exceeded 22,000 in total. Successful library creation was verified by Sanger sequencing of random transformants (Genewiz). All colonies were collected from bioassay dishes in 15 mL of LB-media containing 15% glycerol, centrifuged at 3000 xg, and the pellet was resuspended in a small volume of supernatant then stored at -70˚C.

An aliquot of each region of the NEB5α *rnc* library was thawed on ice and DNA purified using a QIAprep Spin Miniprep Kit (Qiagen). DNA concentration was estimated by NanoDrop spectrophotometer and verified by gel electrophoresis on a 1% TAE-agarose gel. To avoid double mutant transformants (Goldsmith, et al. 2007), 8.3 ng of each region was transformed into electrocompetent SK7622 cells for the fitness landscape experiment and electrocompetent PH0919 cells harboring the pdCas9 plasmid for the functional landscape experiment. Colonies were collected from the transformation plates as before, centrifuged, resuspended in a small volume of supernatant, and stored at -70˚C.

Growth competition experiment of the SK7622 library. Growth competition was performed essentially as described (Mehlhoff, et al. 2020). SK7622 frozen library stocks for each of the three regions were diluted to OD_{600} of 0.01 in 100 mL of LB-media in unbaffled shaker flasks and incubated until the OD_{600} reached approximately 0.5. Ten mL of this culture was placed on ice and centrifuged at 3000 xg for 10 min at 4˚C. The cultured cells at OD_{600} of 0.5 were diluted to OD_{600} of 0.02 in 100 mL of prewarmed to 37˚C LB-media in baffled shake flasks. Flasks were incubated at 37˚C with shaking until OD_{600} was approximately 0.640 (5 generations of growth). The culture was diluted at a ratio of 3.1 mL culture to 96.9 mL of new pre-warmed to 37˚C LB-media in a separate baffled shaker flask and incubated at 37˚C until OD_{600} reached approximately 0.64 (5 additional generations of growth). Ten mL of this culture was placed on ice and centrifuged at 3000 xg for 10 min at 4˚C. Plasmid from the initial and final time point was extracted and purified from pelleted cells using a QIAprep Spin Miniprep Kit (Qiagen).

Transformation score experiment. Frozen NEB5α and SK7622 library stocks were thawed on ice then centrifuged at 3000 xg for 10 min at 4˚C. Plasmid DNA was extracted from pelleted cells. Extracted DNA was stored then prepared for Illumina MiSeq as described below.

Functional score experiment. Electrocompetent PH0919 cells containing pdCas9 were transformed with pRnc2-gGFP with wildtype RNase III or the E117K mutant, incubated overnight on bioassay dishes, and colonies collected then frozen as described above. Aliquots of the frozen PH0919

library, wildtype, and E117K mutant were thawed on ice. Twenty µL of frozen cells were diluted into 1980 µL of cold phosphate- buffered saline (PBS) and the OD_{600} was measured. Cells were then diluted to an OD_{600} of 0.1 in a final volume of 2 mL; region 2 samples were spiked with 1% E117K sample based upon OD_{600}. Fluorescence-activated cell sorting (FACS) of each library region was performed based on GFP-fluorescence and forward scatter gating placed around wildtype RNase III and E117K samples using a Becton Coulter MoFlo Legacy cell sorter. Cells were sorted until greater than 1 x 10^6 high fluorescent events were collected, during which time approximately 1.5-2.0 x 10^7 low fluorescent events were collected. Sorted cells were centrifuged at 3000xg and DNA extracted, stored, and then prepared for Illumina MiSeq as described below.

Illumina MiSeq sample preparation. Plasmids collected during the growth competition experiment, the transformation score experiment and the functional score experiment were digested with XhoI (New England Biolabs) and purified using a DNA Clean & Concentrator Kit (Zymo). Custom Illumina adapter sequences (IDT) containing unique barcode to identify which region and time point DNA corresponded to were added to the linearized DNA by PCR amplification with Phusion High Fidelity polymerase (New England Biolabs) and HF buffer (New England Biolabs). For replica 1 of the growth competition experiment, PCR amplification was performed with ~50 ng of DNA template and following a protocol of 98˚C for 30 sec, 25 cycles of 98˚C for 15 sec, 54˚C for 15 sec, and 72˚C for 1 min, followed by 72˚C for 5 min. For replica 2 of the growth competition experiment and the transformation experiment, PCR amplification was performed as replica 1 but with ~25 ng of DNA template and following a protocol with 15 cycles instead of 25. For replica 1 and 2 of the functional experiment, PCR amplification was performed as above with ~25 ng of DNA and following a protocol with 20 cycles instead of 25. Proper DNA size was verified by gel electrophoresis on a 2% TAE-agarose gel. Samples were pooled and submitted for Illumina MiSeq (2 x 300 bp reads) at the Transcriptomics and Deep Sequencing Core facility at Johns Hopkins University.

Deep sequencing analysis. Illumina MiSeq reads were analyzed essentially as described (Mehlhoff, et al. 2020). Paired-end reads were merged using PEAR (Zhang, et al. 2013) set to a minimum assembly length of 200 base pairs. Illumina adapters and base pairs outside the desired region were cropped from the merged reads using Trimmomatic (Bolger, et al. 2014) (Region 1- Headcrop: 39, crop 225; Region 2- Headcrop: 33 , crop: 225; Region 3 – Headcrop: 35 , crop: 228). The resulting trimmed read were input to Enrich2 (Rubin, et al. 2017), which counted the variants for use in calculating enrichment to determine fitness, transformation score, functional score, and corresponding variances. Reads containing bases with a quality score below 20, bases marked as N, or mutations at more than one codon were filtered out.

Fitness, transformation score, and functional score calculation. Fitness values (*w*) were determined as described (Mehlhoff, et al. 2020) using Equation 1, except the rate of wildtype synonyms was used in place of the growth rate for wildtype

$$w_i = \frac{\ln(r\varepsilon_i)}{\ln(r\varepsilon_{wt,syn})} (1)$$

where *r* is the fold increase in total number of cells and ε_i is the enrichment of non-synonymous mutation *i*. Fitness was calculated on the amino acid level by pooling the sequencing counts for all synonymous codons.

Enrichment of non-synonymous mutations was calculated as described (Mehlhoff, et al. 2020) using Equation 2 where c_i is the count of non-synonymous mutation *i* in the designated population

$$\varepsilon_i = \frac{\frac{c_{i,f}}{c_{T,f}}}{\frac{c_{i,o}}{c_{T,o}}} = \frac{(c_{i,f})(c_{T,o})}{(c_{i,o})(c_{T,f})} \approx \frac{(c_{i,f}+0.5)(c_{T,o})}{(c_{i,o}+0.5)(c_{T,f})} \quad (2)$$

and C_T is the total count of the designated population. C_i was calculated by summing the counts of all synonymous mutations for each amino acid substitution. We added the 0.5 to all counts of mutants to assist with mutations that have counts that are zero (i.e. so that a fitness value can still be calculated).

Since there was not cell growth in the transformation and functional experiments, we defined the transformation and functional scores as:

$$w_{t_i} = \log \frac{\varepsilon_i}{\varepsilon_{wt\ syn}} (3)$$

$$w_{f_i} = \log \frac{\varepsilon_i}{\varepsilon_{wt\ syn}} \quad (4)$$

The enrichments were determined using Equation 2. For the transformation score, the NEB5α library was population *o*, while the SK7622 library was population *f*. For the functional score, the highly fluorescent (i.e. most similar to E117K) population was population *o* while the dim GFP population (most similar to wildtype) was population *f*.

Statistical treatment of DMS measurements. The statistical analysis for fitness values was performed as described (Mehlhoff, et al. 2020). Statistical calculations for transformation and functional scores were calculated as below. For transformation score, the NEB5α library was population *o*, while the SK7622 library was population *f*. For the functional score, the highly fluorescent (i.e. most like E117K) population was population *o* while the dim GFP population (most like wildtype) was population *f*. The pre-log variance of transformation score and functional score was defined as the variance σ^2, which depends on the proportion *p* and the number of samples from the population:

$$\sigma^2 = \frac{p\,(1-p)}{n} \quad (5)$$

For our calculations $p = \frac{c_i}{c_T}$ and $n = c_T$.

In propagating uncertainty, the variance of a proportion of proportions was defined as:

$$\sigma_F^2 = F^2\left[\left(\frac{\sigma_A}{A}\right)^2 + \left(\frac{\sigma_B}{B}\right)^2 - 2\frac{\sigma_{AB}}{AB}\right],\ F = \frac{A}{B} \quad (6)$$

Here the $\frac{\sigma_{AB}}{AB}$ term was equal to 0 since there is no covariance between proportions A and B. To calculate the variance for transformation score and functional score, there were four proportions to consider: the frequencies of the mutant in the *o* and *f* populations, and the frequency of WT synonyms in *o* and *f* populations. Thus, the variance of the functional fitness was defined as:

$$\sigma_F^2 = F^2\left[\left(\frac{\sigma_A}{A}\right)^2 + \left(\frac{\sigma_B}{B}\right)^2 + \left(\frac{\sigma_C}{C}\right)^2 + \left(\frac{\sigma_D}{D}\right)^2\right],\ F = \frac{\frac{A}{B}}{\frac{C}{D}},\ \mathrm{A} = p_{i,f},\ \mathrm{B} = p_{i,o},\ \mathrm{C} = p_{wt\,syn,f},\ \mathrm{D} = p_{wt\,syn,o.} \quad (7)$$

We defined the range of values for the transformation and functional scores as the log of the score ± variance from equation 7.

Confidence intervals and *p*-values for transformation and functional scores were calculated as described (Mehlhoff, et al. 2020) using the variance.

Calculation of *k values.** The *k** values were calculated from the relative enrichments using Equation 8-10 in a manner that is analogous to calculations of Shannon entropy (Firnberg, et al. 2014, 2016).

$$p_i = \frac{10^{w_i}}{\sum_{j=1}^{n} 10^{w_j}} \quad (8)$$

$$S = -\sum_{i=1}^{n} p_i \log_2 p_i \quad (9)$$

$$k^* = \left(\frac{20}{n}\right) 2^S \quad (10)$$

The term w_i is one of the n functional or transformation scores at a position (see Equations 3 and 4) and the prefactor ($20/n$) is a way of estimating the value of k^* when a position has less than a full complement of 20 scores at that position.

RNase III fluorescence activity assay. Overnight inoculums were diluted to a final OD_{600} of 0.00625 in 100 mL of LB-media in baffled shake flasks. Cultures were incubated at 37˚C with shaking at 250 rpm for approximately 4 hours. Triplicates of 100 µL from each culture were added to a clear/black bottom 96-well microplate. OD_{600} and GFP fluorescence at 485 nm excitation/525 nm emission were measured using a SpectraMax M3 microplate reader.

In vitro primer extension. The primer extension assay was performed essentially as described (Sharma and Woodson 2020); the protocol is described in detail here (Hauk, et al. 2022). After 6 hr of growth in 100 mL of LB-media, 1 mL of 1x 10^8 cells/mL were harvested and pelleted at 3000x g. The cell pellet was incubated in 200 µL of TE buffer (Gentrox) containing 1 mg/mL lysozyme at room

temperature for 5 min, vortexing for 10 s every 90 s. A total of 680 µL of RLT (Qiagen) buffer with 1% 2-mercaptoethanol was added to the resuspended pellet followed by addition of 490 µL of 100% ethanol. This lysate was used for RNA extraction with an RNeasy Mini Kit (Qiagen) according to the manufacturer's standard protocol for bacterial RNA purification.

The purified RNA was used to amplify cDNA from the gUTRGFP template with the 5' Cy-5 labelled pWP252fluor primer (5' ATAACGGACTAGCCTTA 3') using the standard Superscript III First-Strand Synthesis System (Thermo Fisher) with 3 µg of total RNA and elongating at 52.5˚C for 60 min. Next, undyed urea loading buffer (8 M urea, TE buffer) was added to cDNA and the DNA was denatured at 95˚C for 3 min prior to loading in Novex 6% TBE-urea gel (Thermo Fisher). Samples were separated for 45 min at 180 V, 19 mA in TBE buffer (Quality Biological). The gel was visualized for Cy-5 fluorescence using a Typhoon 9410. DNA standards (ssDNA standards) were amplified from pRnc-gRNA DNA using Phusion High Fidelity DNA Polymerase (New England BioLabs) with the pWP252fluor reverse primer and forward primers annealing to the expected 5' termini of the cleaved and uncleaved 5'UTR that would yield the desired size products. PCR products were separated on a 4% agarose gel at 110V for 40 min and proper-sized products extracted using QIAquick gel extraction kit (Qiagen). Following denaturation in undyed loading buffer at 95˚C for 3 min, DNA standards and cDNA both migrate as ssDNA in denaturing gels.

Cleavage scores were quantified using ImageJ software (Schneider, et al. 2012). Cleavage percentages are determined as the ratio of the intensity of the sum of upstream and downstream cleavage bands to the sum of the uncleaved, upstream and downstream cleaved bands. Cleavage scores represent the mean of three biological replicates each with two technical duplicates.

Acknowledgments

We thank Hao Zhang for his support with FACS and Jacob Mehlhoff for his assistance with deep sequencing analysis. This research was supported by National Science Foundation grants MCB-1817646 and MCB-2113019 to M.O.

Data availability

DMS sequencing data can be found at BioProject (https://www.ncbi.nlm.nih.gov/bioproject) using accession no. PRJNA887445. Sequencing counts for the DMS studies can be found in **supplementary data S1-9**, Supplementary Material online.

References

Adams RM, Mora T, Walczak AM, Kinney JB. 2016. Measuring the sequence-affinity landscape of antibodies with massively parallel titration curves. *eLife* 5:e23156.

Adkar BV, Tripathi A, Sahoo A, Bajaj K, Goswami D, Chakrabarti P, Swarnkar MK, Gokhale RS, Varadarajan R. 2012. Protein model discrimination using mutational sensitivity derived from deep sequencing. *Structure* 20:371-381.

Altuvia Y, Bar A, Reiss N, Karavani E, Argaman L, Margalit H. 2018. In vivo cleavage rules and target repertoire of RNase III in Escherichia coli. *Nucleic Acids Res.* 46:10380-10394.

Apirion D, Watson N. 1974. Analysis of an Escherichia coli strain carrying physiologically compensating mutations one of which causes an altered ribonuclease 3. *Mol. Gen. Genet.* 132:89-104.

Aristarkhov A, Mikulskis A, Belasco JG, Lin EC. 1996. Translation of the adhE transcript to produce ethanol dehydrogenase requires RNase III cleavage in Escherichia coli. *J. Bacteriol.* 178:4327-4332.

Asgari M, Ilbeigikhamsehnejad M, Rismani E, Dinparast Djadid N, Raz A. 2020. Molecular characterization of RNase III protein of Asaia sp. for developing a robust RNAi-based paratransgensis tool to affect the sexual life-cycle of Plasmodium or Anopheles fitness. *Parasites & Vectors* 13:42.

Babitzke P, Granger L, Olszewski J, Kushner SR. 1993. Analysis of mRNA decay and rRNA processing in Escherichia coli multiple mutants carrying a deletion in RNase III. *J. Bacteriol.* 175:229-239.

Bandaru P, Shah NH, Bhattacharyya M, Barton JP, Kondo Y, Cofsky JC, Gee CL, Chakraborty AK, Kortemme T, Ranganathan R, et al. 2017. Deconstruction of the Ras switching cycle through saturation mutagenesis. *eLife* 6:e27810.

Bardwell JC, Régnier P, Chen SM, Nakamura Y, Grunberg-Manago M, Court DL. 1989. Autoregulation of RNase III operon by mRNA processing. *EMBO J.* 8:3401-3407.

Bershtein S, Mu W, Shakhnovich EI. 2012. Soluble oligomerization provides a beneficial fitness effect on destabilizing mutations. *Proc. Natl. Acad. Sci. USA* 109:4857-4862.

Blaszczyk J, Tropea JE, Bubunenko M, Routzahn KM, Waugh DS, Court DL, Ji X. 2001. Crystallographic and modeling studies of RNase III suggest a mechanism for double-stranded RNA cleavage. *Structure* 9:1225-1236.

Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30:2114-2120.

Bosley AD, Ostermeier M. 2005. Mathematical expressions useful in the construction, description and evaluation of protein libraries. *Biomol. Eng.* 22:57-61.

Boucher JI, Bolon DN, Tawfik DS. 2016. Quantifying and understanding the fitness effects of protein mutations: Laboratory versus nature. *Protein Sci.* 25:1219-1226.

Court DL, Gan J, Liang YH, Shaw GX, Tropea JE, Costantino N, Waugh DS, Ji X. 2013. RNase III: Genetics and function; structure and mechanism. *Annu. Rev. Genet.* 47:405-431.

Dasgupta S, Fernandez L, Kameyama L, Inada T, Nakamura Y, Pappas A, Court DL. 1998. Genetic uncoupling of the dsRNA-binding and RNA cleavage activities of the Escherichia coli endoribonuclease RNase III--the effect of dsRNA binding on gene expression. *Mol. Microbiol.* 28:629-640.

Deutscher MP. 2009. Chapter 9 Maturation and Degradation of Ribosomal RNA in Bacteria. In: Condon C, editor. Molecular Biology of RNA Processing and Decay in Prokaryotes. p. 369-391.

Dunn JJ, Studier FW. 1973. T7 early RNAs and Escherichia coli ribosomal RNAs are cut from large precursor RNAs in vivo by ribonuclease 3. *Proc. Natl. Acad. Sci. USA* 70:3296-3300.

Firnberg E, Labonte JW, Gray JJ, Ostermeier M. 2016. A Comprehensive, High-Resolution Map of a Gene's Fitness Landscape. *Mol. Biol. Evol.* 33:1378.

Firnberg E, Labonte JW, Gray JJ, Ostermeier M. 2014. A comprehensive, high-resolution map of a gene's fitness landscape. *Mol. Biol. Evol.* 31:1581-1592.

Fowler DM, Fields S. 2014. Deep mutational scanning: a new style of protein science. *Nat. Methods* 11:801-807.

Fujiwara K, Toda H, Ikeguchi M. 2012. Dependence of α-helical and β-sheet amino acid propensities on the overall protein fold type. *BMC Struct. Biol.* 12:18.

Gan J, Shaw G, Tropea JE, Waugh DS, Court DL, Ji X. 2008. A stepwise model for double-stranded RNA processing by ribonuclease III. *Mol. Microbiol.* 67:143-154.

Gan J, Tropea JE, Austin BP, Court DL, Waugh DS, Ji X. 2005. Intermediate states of ribonuclease III in complex with double-stranded RNA. *Structure* 13:1435-1442.

Gan J, Tropea JE, Austin BP, Court DL, Waugh DS, Ji X. 2006. Structural insight into the mechanism of double-stranded RNA processing by ribonuclease III. *Cell* 124:355-366.

Gegenheimer P, Apirion D. 1975. Escherichia coli ribosomal ribonucleic acids are not cut from an intact precursor molecule. *J. Biol. Chem.* 250:2407-2409.

Gitelman DR, Apirion D. 1980. The synthesis of some proteins is affected in RNA processing mutants of Escherichia coli. *Biochem. Biophys. Res. Commun.* 96:1063-1070.

Goldsmith M, Kiss C, Bradbury AR, Tawfik DS. 2007. Avoiding and controlling double transformation artifacts. *Protein Eng. Des. Sel.* 20:315-318.

Gordon GC, Cameron JC, Pfleger BF. 2017. RNA Sequencing Identifies New RNase III Cleavage Sites in Escherichia coli and Reveals Increased Regulation of mRNA. *mBio* 8:e00128-00117.

Hartman EC, Tullman-Ercek D. 2019. Learning from protein fitness landscapes: a review of mutability, epistasis, and evolution. *Curr. Opin. Sys. Biol.* 14:25-31.

Hauk P, Weeks R, Ostermeier M. 2022. A CRISPR-dCas9 System for Assaying and Selecting for RNase III Activity In Vivo in Escherichia coli. *CRISPR J.*

Hecht A, Glasgow J, Jaschke PR, Bawazer LA, Munson MS, Cochran JR, Endy D, Salit M. 2017. Measurements of translation initiation from all 64 codons in E. coli. *Nucleic Acids Res.* 45:3615-3626.

Heredia JD, Park J, Brubaker RJ, Szymanski SK, Gill KS, Procko E. 2018. Mapping Interaction Sites on Human Chemokine Receptors by Deep Mutational Scanning. *Journal of Immunology* 200:3825-3839.

Hietpas RT, Jensen JD, Bolon DNA. 2011. Experimental illumination of a fitness landscape. *Proc. Natl. Acad. Sci. USA* 108:7896-7901.

Inada T, Nakamura Y. 1995. Lethal double-stranded RNA processing activity of ribonuclease III in the absence of suhB protein of Escherichia coli. *Biochimie* 77:294-302.

Jiang L, Mishra P, Hietpas RT, Zeldovich KB, Bolon DNA. 2013. Latent Effects of Hsp90 Mutants Revealed at Reduced Expression Levels. *PLoS Genet.* 9:e1003600.

Kaczanowska M, Rydén-Aulin M. 2007. Ribosome biogenesis and the translation process in Escherichia coli. *Microbiol. Mol. Biol. Rev.* 71:477-494.

Kebschull JM, Zador AM. 2015. Sources of PCR-induced distortions in high-throughput sequencing data sets. *Nucleic Acids Res.* 43:e143.

Lind PA, Arvidsson L, Berg OG, Andersson DI. 2016. Variation in Mutational Robustness between Different Proteins and the Predictability of Fitness Effects. *Mol. Biol. Evol.* 34:408-418.

Lind PA, Berg OG, Andersson DI. 2010. Mutational Robustness of Ribosomal Protein Genes. *Science* 330:825-827.

Lundin E, Tang P-C, Guy L, Näsvall J, Andersson DI. 2017. Experimental Determination and Prediction of the Fitness Effects of Random Point Mutations in the Biosynthetic Enzyme HisA. *Mol. Biol. Evol.* 35:704-718.

MacRae IJ, Doudna JA. 2007. Ribonuclease revisited: structural insights into ribonuclease III family enzymes. *Curr. Opin. Struct. Biol.* 17:138-145.

Matsunaga J, Dyer M, Simons EL, Simons RW. 1996. Expression and regulation of the rnc and pdxJ operons of Escherichia coli. *Mol. Microbiol.* 22:977-989.

Matsunaga J, Simons EL, Simons RW. 1996. RNase III autoregulation: structure and function of rncO, the posttranscriptional "operator". *RNA* 2:1228-1240.

Mavor D, Barlow K, Thompson S, Barad BA, Bonny AR, Cario CL, Gaskins G, Liu Z, Deming L, Axen SD, et al. 2016. Determination of ubiquitin fitness landscapes under different chemical stresses in a classroom setting. *eLife* 5:e15802.

Mehlhoff JD, Ostermeier M. 2020. Biological fitness landscapes by deep mutational scanning. *Methods Enzymol.* 643:203-224.

Mehlhoff JD, Stearns FW, Rohm D, Wang B, Tsou E-Y, Dutta N, Hsiao M-H, Gonzalez CE, Rubin AF, Ostermeier M. 2020. Collateral fitness effects of mutations. *Proc. Natl. Acad. Sci. USA* 117:11597-11607.

Melamed D, Young DL, Gamble CE, Miller CR, Fields S. 2013. Deep mutational scanning of an RRM domain of the Saccharomyces cerevisiae poly(A)-binding protein. *RNA* 19:1537-1551.

Mian IS. 1997. Comparative sequence analysis of ribonucleases HII, III, II PH and D. *Nucleic Acids Res.* 25:3187-3195.

Nicholson AW. 1999. Function, mechanism and regulation of bacterial ribonucleases. *FEMS Microbiol. Rev.* 23:371-390.

Nicholson AW. 2014. Ribonuclease III mechanisms of double-stranded RNA cleavage. *Wiley Interdiscip Rev RNA* 5:31-48.

Noda-García L, Davidi D, Korenblum E, Elazar A, Putintseva E, Aharoni A, Tawfik DS. 2019. Chance and pleiotropy dominate genetic diversity in complex bacterial environments. *Nat. Microbiol.* 4:1221-1230.

Pääbo S, Irwin DM, Wilson AC. 1990. DNA damage promotes jumping between templates during enzymatic amplification. *J. Biol. Chem.* 265:4718-4721.

Peris JB, Davis P, Cuevas JM, Nebot MR, Sanjuán R. 2010. Distribution of fitness effects caused by single-nucleotide substitutions in bacteriophage f1. *Genetics* 185:603-609.

Qi LS, Larson MH, Gilbert LA, Doudna JA, Weissman JS, Arkin AP, Lim WA. 2013. Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. *Cell* 152:1173-1183.

Régnier P, Grunberg-Manago M. 1989. Cleavage by RNase III in the transcripts of the met Y-nus-A-infB operon of Escherichia coli releases the tRNA and initiates the decay of the downstream mRNA. *J. Mol. Biol.* 210:293-302.

Régnier P, Grunberg-Manago M. 1990. RNase III cleavages in non-coding leaders of Escherichia coli transcripts control mRNA stability and genetic expression. *Biochimie* 72:825-834.

Rodrigues JV, Bershtein S, Li A, Lozovsky ER, Hartl DL, Shakhnovich EI. 2016. Biophysical principles predict fitness landscapes of drug resistance. *Proc. Natl. Acad. Sci. USA* 113:E1470-E1478.

Roscoe BP, Thayer KM, Zeldovich KB, Fushman D, Bolon DNA. 2013. Analyses of the Effects of All Ubiquitin Point Mutants on Yeast Growth Rate. *J. Mol. Biol.* 425:1363-1377.

Rubin AF, Gelman H, Lucas N, Bajjalieh SM, Papenfuss AT, Speed TP, Fowler DM. 2017. A statistical framework for analyzing deep mutational scanning data. *Genome Biol.* 18:150.

Sarkisyan KS, Bolotin DA, Meer MV, Usmanova DR, Mishin AS, Sharonov GV, Ivankov DN, Bozhanova NG, Baranov MS, Soylemez O, et al. 2016. Local fitness landscape of the green fluorescent protein. *Nature* 533:397-401.

Schneider CA, Rasband WS, Eliceiri KW. 2012. NIH Image to ImageJ: 25 years of image analysis. *Nat. Methods* 9:671-675.

Shah NH, Kuriyan J. 2019. Understanding molecular mechanisms in cell signaling through natural and artificial sequence variation. *Nat. Struct. Mol. Biol.* 26:25-34.

Sharma IM, Woodson SA. 2020. RbfA and IF3 couple ribosome biogenesis and translation initiation to increase stress tolerance. *Nucleic Acids Res.* 48:359-372.

Shenkin PS, Erman B, Mastrandrea LD. 1991. Information-theoretical entropy as a measure of sequence variability. *Proteins* 11:297-313.

Sim M, Lim B, Sim S-H, Kim D, Jung E, Lee Y, Lee K. 2014. Two Tandem RNase III Cleavage Sites Determine betT mRNA Stability in Response to Osmotic Stress in Escherichia coli. *PLOS One* 9:e100520.

Sim SH, Yeom JH, Shin C, Song WS, Shin E, Kim HM, Cha CJ, Han SH, Ha NC, Kim SW, et al. 2010. Escherichia coli ribonuclease III activity is downregulated by osmotic stress: consequences for the degradation of bdm mRNA in biofilm formation. *Mol. Microbiol.* 75:413-425.

Stead MB, Marshburn S, Mohanty BK, Mitra J, Pena Castillo L, Ray D, van Bakel H, Hughes TR, Kushner SR. 2011. Analysis of Escherichia coli RNase E and RNase III activity in vivo using tiling microarrays. *Nucleic Acids Res.* 39:3188-3203.

Stiffler MA, Hekstra DR, Ranganathan R. 2015. Evolvability as a function of purifying selection in TEM-1 β-lactamase. *Cell* 160:882-892.

Subramanian S, Gorday K, Marcus K, Orellana MR, Ren P, Luo XR, O'Donnell ME, Kuriyan J. 2021. Allosteric communication in DNA polymerase clamp loaders relies on a critical hydrogen-bonded junction. *eLife* 10:e66181.

Sun W, Jun E, Nicholson AW. 2001. Intrinsic double-stranded-RNA processing activity of Escherichia coli ribonuclease III lacking the dsRNA-binding domain. *Biochemistry* 40:14976-14984.

Sun W, Li G, Nicholson AW. 2004. Mutational analysis of the nuclease domain of Escherichia coli ribonuclease III. Identification of conserved acidic residues that are important for catalytic function in vitro. *Biochemistry* 43:13054-13062.

Sun W, Nicholson AW. 2001. Mechanism of action of Escherichia coli ribonuclease III. Stringent chemical requirement for the glutamic acid 117 side chain and Mn2+ rescue of the Glu117Asp mutant. *Biochemistry* 40:5102-5110.

Thompson S, Zhang Y, Ingle C, Reynolds KA, Kortemme T. 2020. Altered expression of a quality control protease in E. coli reshapes the in vivo mutational landscape of a model enzyme. *eLife* 9:e53476.

Whitehead TA, Chevalier A, Song Y, Dreyfus C, Fleishman SJ, De Mattos C, Myers CA, Kamisetty H, Blair P, Wilson IA, et al. 2012. Optimization of affinity, specificity and function of designed influenza inhibitors using deep sequencing. *Nat. Biotechnol.* 30:543-548.

Wrenbeck EE, Azouz LR, Whitehead TA. 2017. Single-mutation fitness landscapes for an enzyme on multiple substrates reveal specificity is globally encoded. *Nat. Commun.* 8:15695.

Wu Z, Cai X, Zhang X, Liu Y, Tian G-b, Yang J-R, Chen X. 2022. Expression level is a major modifier of the fitness landscape of a protein coding gene. *Nat. Ecol. Evol.* 6:103-115.

Xiao J, Feehery CE, Tzertzinis G, Maina CV. 2009. E. coli RNase III(E38A) generates discrete-sized products from long dsRNA. *RNA* 15:984-991.

Yampolsky LY, Stoltzfus A. 2005. The exchangeability of amino acids in proteins. *Genetics* 170:1459-1472.

Zhang H, Kolb FA, Jaskiewicz L, Westhof E, Filipowicz W. 2004. Single processing center models for human Dicer and bacterial RNase III. *Cell* 118:57-68.

Zhang J, Kobert K, Flouri T, Stamatakis A. 2013. PEAR: a fast and accurate Illumina Paired-End reAd mergeR. *Bioinformatics* 30:614-620.

Table 1: Properties of individually studied mutants of *E. coli* RNase III.

Variant	Fitness[a,b]	Transformation score[c]	Functional score			Fluor./OD[e,b]	Gel cleavage (%)[b]
			Mean[d]	Replica 1[c]	Replica 2[c]		
Wildtype	1.00 ± 0.01	0.00	0.00	0.00	0.00	458 ± 78	69 ± 2
E65P	1.00 ± 0.01	-0.26 +0.02/-0.02	-0.09	0.04 +0.02/-0.02	-0.13 +0.01/-0.01	401 ± 34	37 ± 7
D114G	0.99 ± 0.01	-0.13 +0.03/-0.03	-0.04	0.07 +0.03/-0.04	-0.07 +0.01/-0.01	383 ± 5	48 ± 5
E38Q	0.99 ± 0.01	-0.19 +0.02/-0.02	-0.11	-0.14 +0.02/-0.02	-0.05 +0.01/-0.01	340 ± 6	47 ± 8
D114R	0.97 ± 0.01	-0.57 +0.03/-0.03	-0.02	0.01 +0.05/-0.05	-0.03 +0.01/-0.01	349 ± 37	17 ± 4
D155E	0.90 ± 0.01	-1.10 +0.08/-0.09	-0.36	-0.09 +0.04/-0.05	-0.45 +0.05/-0.06	371 ± 22	13 ± 5
E38A	0.88 ± 0.01	-1.05 +0.03/-0.03	-1.47	-1.51 +0.03/-0.03	-0.46 +0.04/-0.04	543 ± 51	15 ± 4
F188D	0.83 ± 0.01	-1.26 +0.08/-0.10	-0.13	-0.17 +0.07/-0.08	0.01 +0.04/-0.04	303 ± 8	27 ± 6
E38V	0.80 ± 0.01	-1.03 +0.03/-0.03	-1.55	-1.58 +0.02/-0.02	-0.65 +0.05/-0.06	747 ± 25	6 ± 4
C192C	0.80 ± 0.01	-1.66 +0.18/-0.31	-0.91	-0.96 +0.06/-0.07	-0.63 +0.16/-0.25	841 ± 145	13 ± 5
TCT[f]	0.76 ± 0.01	-0.71 +0.07/-0.09	-2.33	-2.40 +0.15/-0.23	-0.88 +0.20/-0.38	938 ± 54	7 ± 4
ΔdsRBD[g]	0.75 ± 0.01	n.d.[h]	n.d.	n.d.	n.d.	786 ± 95	2 ± 2
E117K	0.71 ± 0.01	-1.28 +0.15/-0.23	-2.18	-2.20 +0.01/-0.01	-0.84 +0.01/-0.02	1074 ± 36	3 ± 3

[a] Fitness determined as the growth rate relative to wildtype. Variants are ordered based on this value.
[b] Error is the standard deviation (*n*=3)
[c] Error is the standard deviation determined from the DMS sequencing counts
[d] Mean of replica 1 and replica 2 is weighted based on each replica's variance
[e] Fluorescence/OD-600nm of bacterial culture
[f] The M1 ATG start codon is replaced by the codon TCT, which does not function for translation initiation
[g] Q153*/K154*/D155*
[h] n.d. = not determined

Figure Legends

Figure 1: Experimental design for measuring the fitness, transformation, and functional landscapes of RNase III. (**A**) A comprehensive site-saturation library of the coding region of the RNase III gene was created on plasmid pRnc2-gGFP in *rnc*$^{+}$ NEB5α cells. (**B**) Transformation scores were determined from the change in mutation frequency in the library before and after transformation into *rnc*$^{-}$ SK7622 cells. (**C**) Fitness values were determined from the change in mutation frequency in SK7622 cells after ten generations of exponential growth in LB media. To measure (**D**) functional scores, the library was transformed into *rnc*$^{-}$ PH0919 cells that contain a fluorescent reporter for in vivo RNase III cleavage activity (Hauk, et al. 2022). These cells were sorted by FACS into high and low fluorescent populations. The functional score was determined by the difference in mutation frequency between the two populations relative to wild type.

Figure 2: Fitness, transformation scores, and functional scores are qualitatively alike. Heat maps show the (**A**) Fitness, (**B**) transformation score, and (**C**) functional score landscapes of *E. coli* RNase III. Red is deleterious, white is neutral, blue is beneficial, and black indicates insufficient data due to low sequencing counts (an average of ≤10 sequencing counts in the initial timepoint for fitness and transformation score and ≤ 24 total counts for functional score). The shading for each landscape was set such that full red represented the average value corresponding to complete loss of RNase III activity. The amino acid substitutions are shown on the left (* = nonsense mutation). Fitness and functional scores are reported as the weighted mean of two replica experiments. Transformation score is a single experiment. At the bottom are secondary structure elements (α-helices in blue and β-strands in green); functional motifs in orange including the signature motif, the four RNA binding motifs (RBM), and the six negatively charged active site residues (orange *); and the RIIID and dsRBD domains joined by the linker. The values and errors of all fitness, transformation scores, and functional scores can be found in **supplementary data S1-S3**, Supplementary Material online.

Figure 3: The distribution of mutational effects for transformation and functional scores are bimodal and the three fitness metrics correlate with each other. Histograms of (**A**) the weighted mean fitness, (**B**) the transformation score, and (**C**) the weighted mean functional score for missense (blue) and nonsense (red) mutations. The distribution of nonsense mutations is too broad to observe in A and C. The distribution of nonsense mutation functional scores can be seen in **supplementary fig. S6B,C**, Supplementary Material online. (**D-F**) Relationship between fitness, transformation score, and functional score. The values for wildtype (blue) and nonfunctional RNase III (red) are shown for reference. The value for nonfunctional RNase III fitness is the growth rate of cells with the M1S mutation (i.e. start

codon replaced with TCT) relative to that of wildtype. The values for nonfunctional RNase III transformation and functional scores are the weighted mean of all mutants with a non-functional start codon.

Figure 4: Relationship between function and fitness metrics for the select mutants of Table 1. (**A**) Fitness correlates with transformation score. (**B**, **C**) The functional score and the fluorescence activity assay are buffered against about fourfold reductions in RNase III catalytic activity**.** (**D**, **E**) Fitness correlates with functional score and the fluorescence activity assay. (**F**) Fitness is buffered against small changes to RNase III catalytic activity. (**G, H**) Effects of all mutations at F188. Several mutations at F188 with deleterious effects on fitness and transformation score are identified. Fitness was determined by measuring the growth rate relative to wildtype (error bars are the standard deviation *n*=3) except for panel G where the fitness values for all mutants except F188D are the weighted mean values from the DMS experiment. Primer extension assays (**supplementary fig. S7**, Supplementary Material online) were used to determine the % cleavage of the 5'UTR stem loop of *rnc* in gUTR129GFP (error bars are the standard deviation *n*=3). The fluorescence/OD of bacterial cultures reports on RNase III cleavage of the 5'UTR stem loop of *rnc* in gUTR129GFP (error bars are the standard deviation *n*=3). Dashed lines are linear fits except for panels B, C, and F, in which case they are guides for the eye. Analogous graphs for each replica of the functional score (as opposed to the weighted mean used in panels B, D, G, and H) are provided as **supplementary fig. S8**, Supplementary Material online.

Figure 5. RNase III mutational tolerance. Mutational tolerance as show by the distribution of *k** values for (**A**) functional score and (**B**) transformation score values. (**C**) Correlation of transformation score *k** values with those from a sequence alignment of the RNase III family. (**D**) The *k** values for functional score and transformation score correlate linearly. (**E**) Transformation score *k** values mapped onto a structural model of *E. coli* RNase III bound to an dsRNA substrate. RNase III model was generated from homology modeling to the *T. maritima* (1O0W) and *A. aeolicus* (2NUG) RNase III structures using Swiss Prot. The linker between the RIIID and the dsRBD is not included as it was poorly resolved in the 1O0W and 2NUG structures. (**F**) A comparison of *k** values mapped on the primary and secondary structure of *E. coli* RNase III. The range of *k** is 1 (white) to (20) red for functional and transformation scores and 1 (white) to 15 (red) for the sequence alignment. All *k** values are tabulated as **supplementary data S4**, Supplementary Material online.

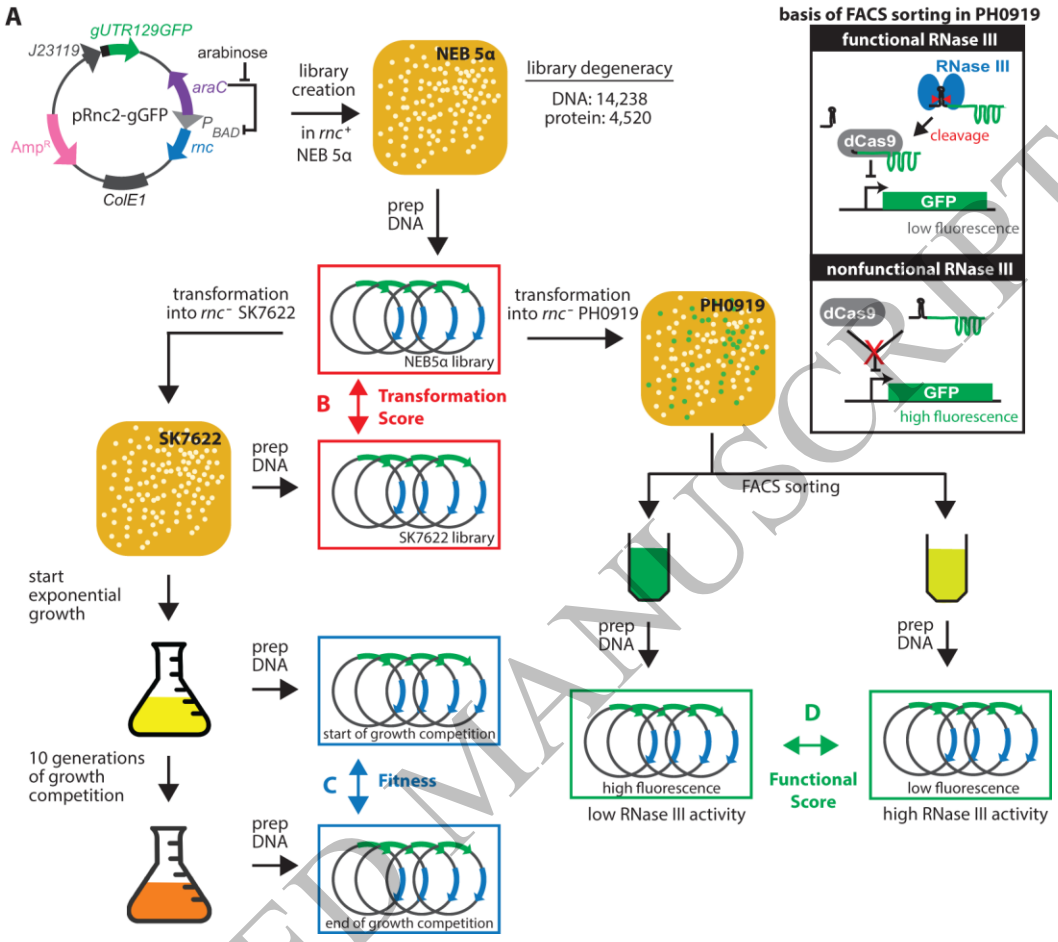

Figure 1
145x131 mm (.97 x DPI)

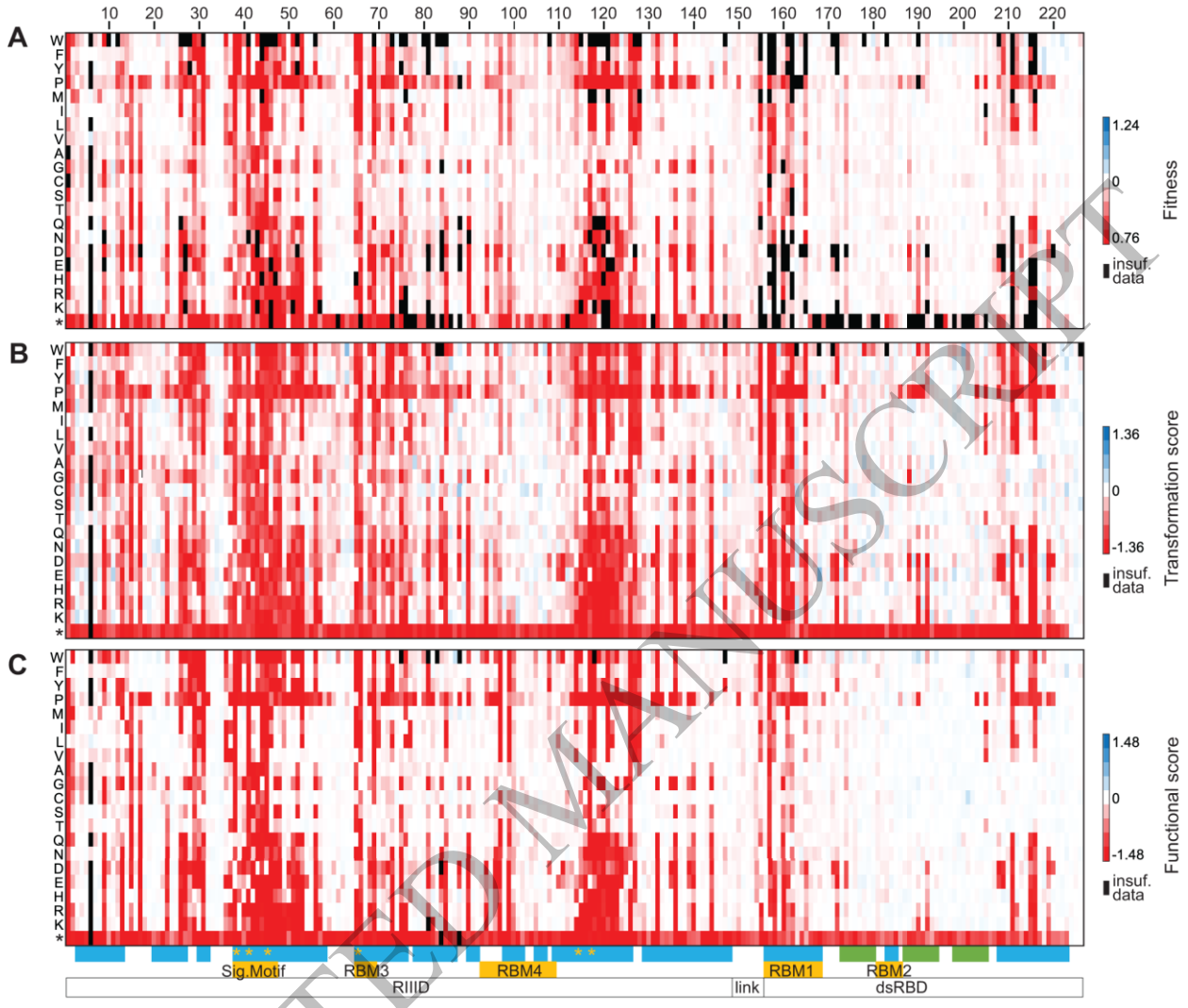

Figure 2
165x140 mm (.97 x DPI)

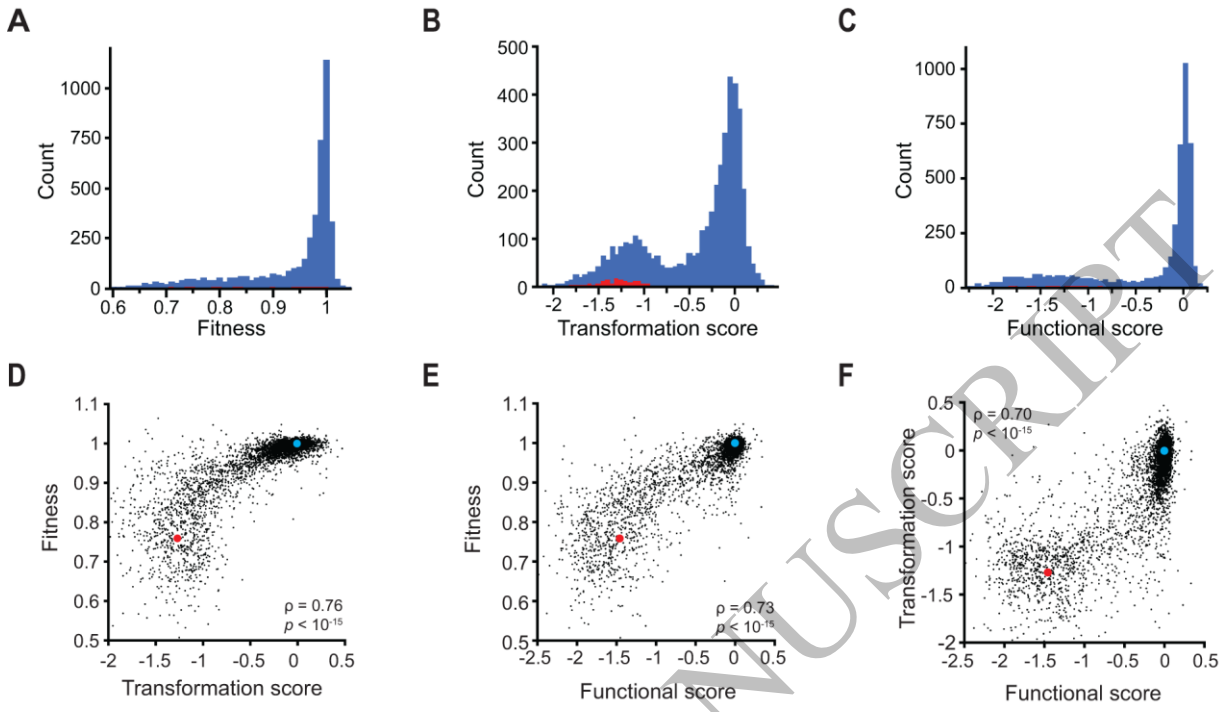

Figure 3
161x93 mm (.97 x DPI)

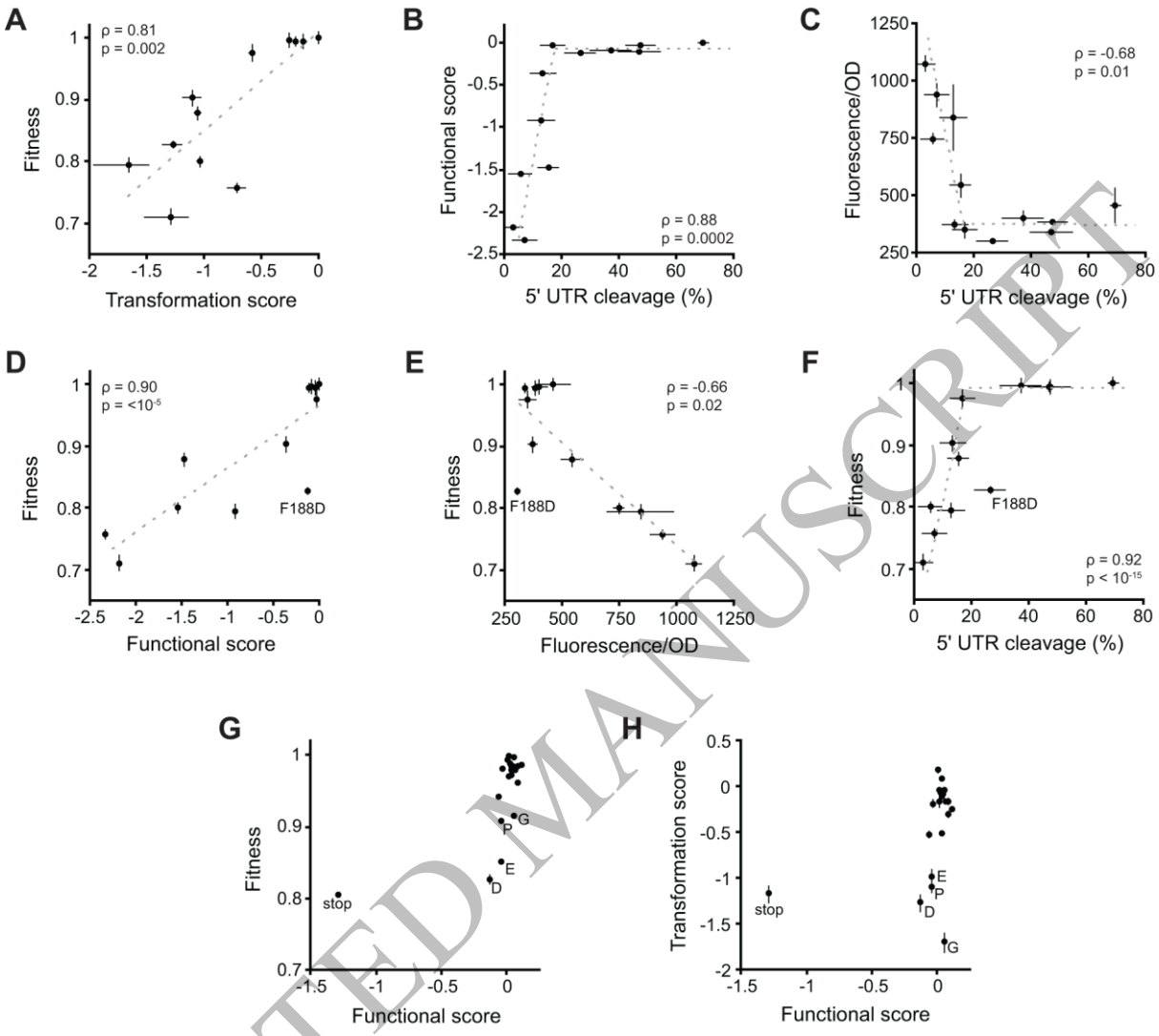

Figure 4
159x142 mm (.97 x DPI)

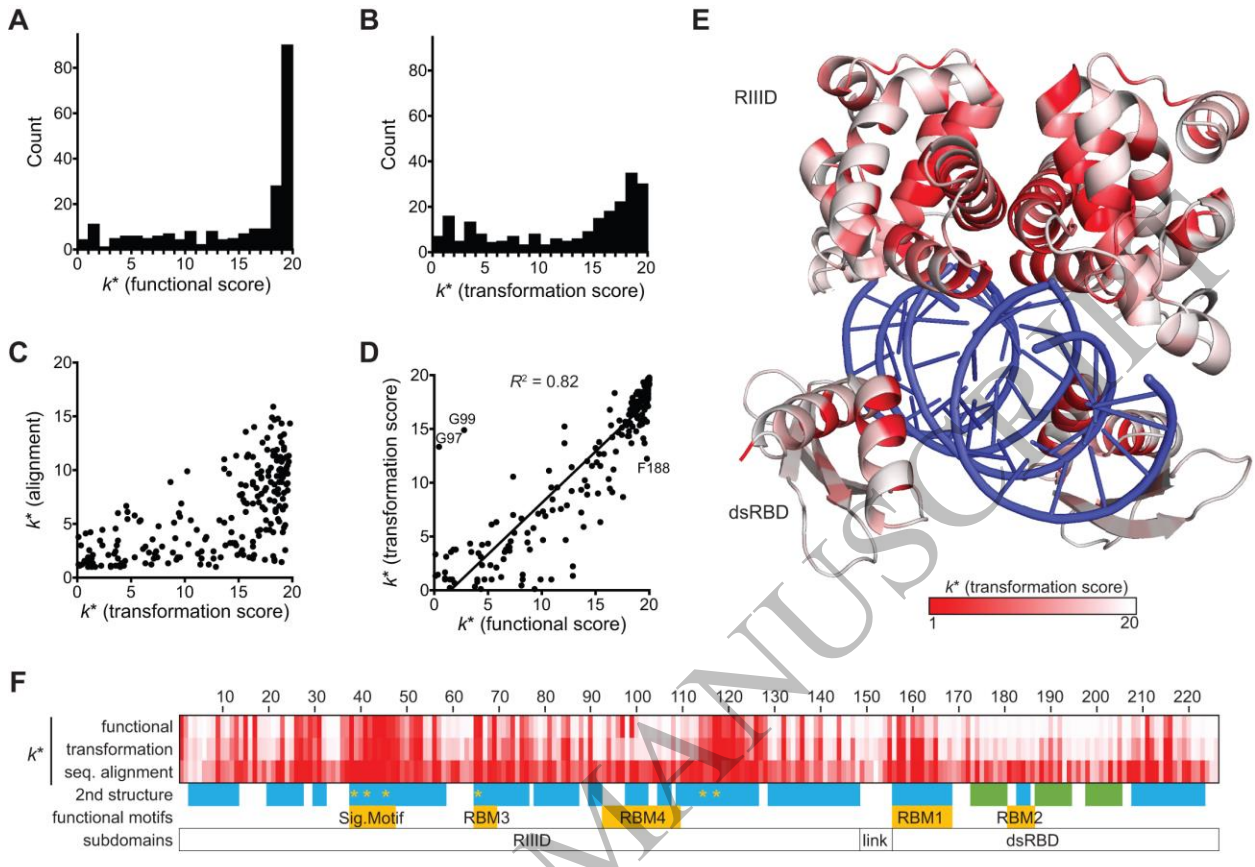

Figure 5

165x112 mm (.97 x DPI)