# RNA-binding and HEPN nuclease activation are decoupled in CRISPR-Cas13a

Akshay Tambe[1,9], Alexandra East-Seletsky[1,2,9], Gavin J. Knott[1], Jennifer A. Doudna[1,3-6*], Mitchell R. O'Connell[1,7,8*]

[1]Department of Molecular and Cell Biology, University of California, Berkeley, California, 94720, USA.

[2]Present Address: Arrakis Therapeutics, 35 Gatehouse Drive, Waltham, Massachusetts, 02451, U.S.A.

[3]Department of Chemistry, University of California, Berkeley, California, 94720, USA

[4]Howard Hughes Medical Institute, University of California, Berkeley, California 94720, USA.

[5]Innovative Genomics Institute, University of California, Berkeley, California 94720, USA.

[6]MBIB Division, Lawrence Berkeley National Laboratory, Berkeley, California 94720, USA.

[7]Department of Biochemistry and Biophysics, School of Medicine and Dentistry, University of Rochester, Rochester, New York 14642, USA;

[8]Center for RNA Biology, University of Rochester, Rochester, New York 14642, USA.

[9]These authors contributed equally.

*Correspondence should be addressed to mitchell_oconnell@urmc.rochester.edu or doudna@berkeley.edu.

**ABSTRACT**

CRISPR-Cas13a is a CRISPR RNA (crRNA)-guided, RNA-activated ribonuclease (RNase) involved in bacterial adaptive immunity and shows promise as a useful tool for RNA detection, RNA imaging and RNA regulation. However, the relationship between Cas13a's crRNA-mediated RNA-activator binding and HEPN nuclease activation is not well understood. Using a comprehensive sequencing-based strategy coupled with *in vitro* biochemistry, we find that Cas13a's crRNA-guided RNA binding affinity and HEPN nuclease activity are differentially affected by the number and position of mismatches between the crRNA and activator-RNA. Unexpectedly, activator-RNAs with the tightest binding affinity are not perfectly complementary to the crRNA, but rather contain mismatches at specific positions. However, despite tight binding to Cas13a, these RNA-activators do not activate Cas13a's HEPN nuclease. Conversely, a subset of mismatched, weakly-bound activator-RNAs are able to maximally activate Cas13a's HEPN nuclease. These results demonstrate that RNA-activator binding and target cleavage by Cas13a are decoupled, highlighting a complex specificity landscape that allows for plasticity in the precise RNA sequence required by Cas13a for RNA binding and HEPN nuclease activation. Taken together, our findings underscore a need to consider the range of effects off-target recognition has on Cas13a's RNA binding and cleavage behavior in both bacterial immunity and RNA-targeting tool development.

## INTRODUCTION

Prokaryotic adaptive immune systems use CRISPRs (clustered regularly interspaced short palindromic repeats) and CRISPR associated (Cas) proteins for RNA-guided cleavage of foreign genetic elements[1,2]. Type VI CRISPR-Cas systems include a single protein, Cas13 (formerly C2c2), that when assembled with a CRISPR-RNA (crRNA) forms a crRNA-guided RNA-targeting effector complex[3,4]. Cas13 proteins studied to date possess two enzymatically distinct RNase activities that are required for optimal interference[3-5]. First, upon binding a precursor-crRNA (pre-crRNA), Cas13 cleaves within the crRNA direct repeat in a pre-crRNA array to form mature Cas13a-crRNA complexes[4-6]. Second, binding of an RNA target complementary to the crRNA (henceforth referred to as an activator-RNA) triggers Cas13 to cleave RNA non-specifically by activating the enzyme's two higher-eukaryotes-and-prokaryotes nucleotide-binding (HEPN) domains to form a single composite RNase active site[3-5,7,8].

In addition to its role during bacterial adaptive immunity, the RNA-activated RNA cleavage behavior of Cas13a provides a mechanism for RNA detection and possible diagnostic applications[4,9]. Versions of Cas13 where the HEPN domains have been catalytically inactivated[3-5] (deactivated Cas13; dCas13) may also be useful as programmable RNA binding proteins for imaging and regulating specific transcripts (for review, see ref. [10]). However, the relationship between crRNA: activator-RNA sequence complementarity and Cas13a binding or HEPN-nuclease activation have not been determined, and based on the behavior exhibited by other CRISPR-Cas systems (for review see[11]), it is likely that this relationship is complex.

To determine the interplay between activator-RNA recognition and HEPN-nuclease activation, we used a combination of biochemical experiments and a library-based high-throughput RNA-binding assay to interrogate the effects of mismatches on both activities. We find that both the number and distribution of mismatches between the crRNA and activator-RNA affect Lbu-Cas13a's binding affinity and HEPN-nuclease activity. Unexpectedly, mismatches within a discrete region of the crRNA: activator-RNA duplex promote higher Cas13a binding

affinity compared a fully complementary activator-RNA. Despite higher-affinity binding, these mismatched activator-RNAs fail to activate Cas13a's HEPN-nuclease activity. Conversely, a separate subset of mismatched, weakly-bound activator-RNAs were able to maximally activate Cas13a's ribonuclease activity. These results show that activator-RNA binding and HEPN-nuclease activation can be decoupled in Cas13a, revealing a proofreading mechanism that is not reflected by simple activator-RNA binding affinity but instead a complex specificity landscape that enables Cas13a to distinguish between complementary and mismatched RNA transcripts prior to crRNA-guided RNase activation, all while allowing for relaxed RNA sequence specificity. Together, we show that the accuracy of activator-RNA binding and RNA-activated RNA cleavage by Cas13a enzymes are different, which has implications for understanding the role of Cas13a in bacterial immunity, as well as the development of Cas13a as a high-fidelity RNA-targeting tool.

## RESULTS

### Base pair mismatches alter crRNA-guided Cas13a: activator-RNA binding affinity

To test the effects of base pair mismatches between a crRNA and a complementary activator-RNA sequence on Cas13a binding affinity, we used fluorescence anisotropy to measure RNA binding by HEPN-nuclease-deactivated Lbu-Cas13a (Lbu-dCas13a). We paired activator-RNAs containing sets of four contiguous nucleotide mismatches with an otherwise complementary crRNA (**Fig. 1A**), tiled across the 20-nt. guide sequence **(Fig. 1B).** For three out of five of these activator-RNAs (activator-RNAs #2, #3 & #6), the affinity of the crRNA-Cas13a interaction decreased by 6- to 8- fold compared to the interaction with a fully complementary RNA (activator-RNA #1; **Fig. 1B, C).** In contrast, mismatches in nucleotide positions 9-12 of the activator-RNA (activator-RNA #4) led to a 47-fold decrease in binding affinity compared to a fully complementary ssRNA, while mismatches in nucleotide positions 13-16 of the target RNA

(activator-RNA #5) led to a minimal two-fold decrease in binding affinity. Taken together, these data suggest that LbuCas13a's binding affinity to activator-RNA is predominantly influenced by complementarity within the middle of the crRNA: activator-RNA duplex.

Given that Cas13a's HEPN-nuclease activity is supported by a range crRNA guide sequence lengths[3-6,8,9], we carried out fluorescence anisotropy binding experiments using an analogous set of mismatched activator-RNAs and a crRNA containing a 24-nt. guide sequence **(Supplementary Fig. 1A, B).** Extending the crRNA: activator-RNA duplex to 24-nt did not increase the overall affinity of the Cas13a: activator-RNA interaction compared to a 20-nt crRNA **(Supplementary Fig. 1C).** This observation suggests that either the additional base-pairing interactions do not occur or are energetically neutral with respect to binding affinity between a crRNA and its perfectly complementary activator-RNA. We also found that the profile of mismatches that impact binding affinity was similar to that measured using the crRNA with a 20-nt guide sequence, although magnitudes of binding defects were smaller **(Supplementary Fig. 1C).** Since the crRNA with 24-nt guide sequence did not enhance binding affinity or specificity for Lbu-Cas13a recognition of RNA targets, we used crRNAs with 20-nt spacer sequences for all further binding experiments.

**Discrete mismatches between crRNA and activator-RNA enhance Cas13a binding affinity**

For both the 20-nt and 24-nt crRNA guide sequences, we noticed that the overall change in fluorescence anisotropy between unbound and fully crRNA-Cas13a-bound activator-RNA depended on the position of the mismatches within the crRNA:activator-RNA duplex. We speculated that this phenomenon was due to undesired interactions between the fluorescein fluorophore and either the Cas13a protein or the activator RNA, and might preclude accurate fluorescence anisotropy binding affinity measurements[12]. We therefore used a filter-binding assay to confirm that the tighter binding affinity detected for RNA #5, relative to other

mismatched activator-RNA sequences, was due to a *bona fide* Cas13a: target RNA binding interaction.

Using filter binding assays with $^{32}$P 5'-radiolabelled activator-RNAs and programmed Cas13a:crRNA complexes, we observed good concordance of binding affinities measured by fluorescence anisotropy and filter binding for Cas13a:crRNA and perfectly complementary activator-RNAs, as well as for activator-RNAs with mismatches to the crRNA guide sequence at nucleotides 9-12 (RNA#4) **(Fig. 2A, B).** In addition, the binding affinities were consistent across two different crRNA: target RNA sequences (λ3 and λ4; **Fig. 2A, B).** However, and to our surprise, we measured a binding affinity approximately ~10-100-fold tighter for RNA sequences bearing mismatches to the crRNA guide sequence at nucleotides 13-16 (RNA #5), compared to perfectly complementary activator-RNA sequences (**Fig. 2A, B).** This result demonstrated that fluorescein at the 3'-end of the activator-RNA #5 interacts with Cas13a leading to an underestimated binding affinity. Taken together, these data indicate that four contiguous base-pair mismatches with the crRNA at nucleotides 13-16 of the activator-RNA sequence promote tighter Cas13a-RNA binding relative to a perfectly complementary crRNA-guided Cas13a: activator-RNA interaction.

**High-throughput profiling of Lbu-Cas13a activator-RNA binding preferences**

To determine the effects of individual crRNA: activator-RNA mismatches on the binding interaction with Cas13a, we designed an *in vitro* library-based assay to assess binding mismatch tolerances and the effects of sequence variation outside the Cas13a:crRNA binding site. To distinguish between weak crRNA-dependent target binding and crRNA-independent nonspecific binding, our experimental design included two different sequence libraries that were mutagenized in the same manner (**Fig. 3A; see Methods)**. Only one crRNA was used in any given binding experiment, allowing us to use the non-complementary target as a reference point for nonspecific binding. The libraries were mixed at a 1:1 ratio and incubated separately with

two concentrations of biotinylated Lbu-dCas13a protein (**see Methods & Supplementary Fig. 2**) complexed with crRNA. After incubation, bound RNAs were eluted and subjected to Illumina-based sequencing (**Fig. 3B; Supplementary Data Table 1**). Controls were conducted to determine input library distribution and the effect of non-specific Cas13a background binding.

To analyze these sequencing data, we developed a computational pipeline (**see Methods**) that first removes PCR duplicates (by counting unique molecular identifier sequences), and then counts the number of observations for each library variant in a given sample. For each variant, we identified which input sequence it was derived from by Hamming distance, and recorded its relative abundance in a sample. After applying a threshold to filter out low-abundance library members, data from the remaining ~18,000 unique sequences were highly reproducible across replicates, with Pearson R >= 0.9 for all but one case (**Supplementary Data Table 1**)**.** We identified the crRNA-dependent signal from our data by computing the fold-change in abundance between pulldowns with crRNA-bound Cas13a and crRNA-unbound apo-Cas13a. A scatter plot of this fold-change for each of the different crRNA containing samples (λ3 and λ4) showed that target sequences that were strongly enriched by Cas13a: crRNA-λ3 were depleted by the Cas13a:crRNA-λ4, and vice versa (**Fig. 3D, E**).

Unsurprisingly, in each case, the enriched and depleted sequences corresponded to the on-target and off-target activator-RNA libraries, respectively. This clear signal of orthogonally indicated that the sequence enrichment and depletion behavior we observed is crRNA-dependent. Interestingly, we observed that this 'orthogonal signal' was significantly diminished when comparing sequence enrichment by Cas13a:crRNA relative to the input library, rather than relative to the crRNA-unbound apo-Cas13a sample. (**Supplementary Fig. 3**). This implies that apo-Cas13a protein has a non-negligible affinity for RNA that is sequence and/or structure dependent. Although we control for this in the analysis presented here, this property might need to be considered when using this Cas13a as an RNA-targeting tool.

**High-throughput profiling identifies activator-RNA mismatch sensitivity hotspots**

We calculated the relative enrichment of target RNA library members with exactly one mismatch to each crRNA-guided Cas13a sample, revealing several interesting observations **(Fig. 4A-D)**. First, as expected, the magnitude of enrichment for each single-mismatch member within each targeted library was highly dependent on the position of the single-mismatch within the crRNA-targeting region. Second, when targeting the λ4 RNA library with Cas13a:crRNA-λ4, we observed that single-mismatch members were least enriched when their mismatches occurred within a contiguous stretch of nucleotides in positions 7-12 compared to all other positions within the crRNA: target RNA duplex (**Fig. 4A, C**), with the effect most pronounced at high Cas13a: crRNA concentration (**Fig. 4C**). This suggests that single base pair mismatches in this region can have a significant negative effect on Cas13a's affinity for activator-RNA, in agreement with the binding data for quadruple mismatches in this region (**Fig.s 1, 2**). Conversely, library members with single base pair mismatches outside this central region (i.e. in activator-RNA sequence positions 2-6 & 13-20) are enriched as well as or even greater than the perfectly complementary target RNA, indicating that mismatches in these positions in the target RNA: crRNA duplex have little effect on the overall binding affinity relative to a perfectly complementary target. In some cases, they may even promote tighter binding than observed for perfectly complementary activator-RNA (**Fig. 4A, C**).

Although less obvious, a similar effect is seen when the λ3 RNA library is targeted by Cas13a:crRNA-λ3, where single-mismatches in positions 11-12 of the crRNA:activator-RNA duplex are also less tolerated **(Fig. 4B, D**). The difference observed in the magnitude of mismatch sensitivity when comparing λ3 and λ4 target libraries could be due to crRNA and target RNA sequence composition, secondary structure propensity of the crRNA and/or target RNA, and the overall efficiency of the sample recovery.

To determine the effects of additional base pair mismatches on Cas13a's binding affinity for target RNA, we calculated the relative enrichment of activator-RNA library members with

exactly two adjacent mismatches (as a sliding window). This analysis revealed a similar pattern of mismatch sensitivities across the length of the crRNA:activator-RNA duplex for both λ3 and λ4 target libraries (**Fig. 4E, F**). Notably, however, we found little evidence that sequence elements outside the crRNA: target RNA duplex (termed the protospacer-flanking motif or PFS[3,7]) contribute to crRNA-guided Cas13a binding to target sequences. Thus, at least for Lbu-Cas13a, a PFS isn't required for optimal activator-RNA binding.

**Off-target RNAs differentially impact Cas13a HEPN-nuclease activation**

We wondered whether mismatched activator-RNAs might still activate Cas13a for trans-RNA cleavage directed by the HEPN-nuclease, and particularly whether we would observe mismatch position-dependent effects. To measure Cas13a HEPN-nuclease activation, we used a fluorescent RNA cleavage assay[4] that assesses HEPN-mediated RNA cleavage upon RNA-mediated Cas13 activation (**Fig. 5A**). The multiple turnover conditions of this assay coupled with the robust catalytic turnover of Lbu-Cas13a allows us to detect HEPN nuclease activity from as little as ~10pM of activator-RNA bound Cas13a:crRNA complex[4], enabling the rapid measurement of a range of parameters that alter the HEPN nuclease activity of Cas13a.

Using this assay, we determined relative apparent rates of fluorescent RNA reporter cleavage (normalized to the completely complementary RNA-activator #1) from each of the resulting time-courses (**Fig. 5B**). The data revealed a range in the ability of each mismatched activator-RNA to activate Cas13a trans-RNA cleavage, and similar effects were observed using a crRNA with a 24-nt guide sequence and a corresponding set of activator-RNAs (**Supplementary Fig. 4**).

Despite exhibiting weaker binding to the Cas13a: crRNA complex, RNAs #3 and #6 are still able to promote robust Cas13a HEPN-nuclease activity, with cleavage rates greater or similar to the catalytic rates promoted by a perfectly complementary activator-RNA (RNA #1), respectively (**Fig. 5B**). This observation suggests that under the equilibrium conditions of this

assay, any decrease in the amount of active activator-RNA-bound Cas13a complex (due to a reduced affinity) is offset by increased catalytic turnover of the HEPN nuclease when mismatches are present in these regions. On the other hand, some quadruple-mismatch activator-RNAs produced large reductions in Cas13a HEPN-nuclease activation (RNAs #2, #4 & #5). The defects in Cas13a catalytic activation exhibited by RNAs #2 and #4 could be either attributed to their weak binding affinities for Cas13a:crRNA (and the lack of a sufficient amount of active activator-RNA-bound Cas13a complex to create a fluorescent signal) or a specific requirement for a base-paired RNA duplex in these regions for HEPN nuclease activation.

Most surprisingly, we found that activator-RNAs with mismatches at positions 13-16 (i.e. RNA #5) are unable to trigger the Cas13a HEPN-nuclease activity (**Fig. 5B**), even though they are able to bind 10- to 100-fold more tightly to Cas13a:crRNA than a perfectly-complementary activator-RNA (**Fig. 2**). We confirmed that this phenomenon occurs across a number of different crRNA: activator-RNA sequences, and that HEPN-nuclease activation occurs upon restoration of base-pairing interactions between the crRNA and the activator-RNA in this region (**Supplementary Fig. 5**). This striking result indicates that stable activator-RNA binding and HEPN nuclease activation can be decoupled. Furthermore, the conformational rearrangement of the HEPN domains required to form a catalytically competent nuclease active site requires a base-paired RNA duplex at positions 13-16 of the activator-RNA. Together, these results reveal a complex interplay between crRNA-spacer: activator-RNA base-pairing requirements for stable Cas13a binding and for HEPN nuclease activation.

**RNA base-pairing is an allosteric activator of Cas13a HEPN-nuclease activity**

To determine how crRNA: activator-RNA base pairing triggers Cas13a HEPN-nuclease catalytic activity, we tested whether activator-RNAs containing single nucleotide mismatches to the crRNA guide sequence at positions 9-16 can activate Cas13a for non-specific RNA cleavage (**Fig. 6A & Supplementary Fig. 6**). Overall, Cas13a nuclease activation was more

affected by single-mismatches in the 13-16 region compared to the 9-12 region (**Fig. 6A**). The largest defect in activation was caused by a mismatch in position 15, which resulted in a five-fold reduction in cleavage rate compared to a perfectly complementary target RNA. This indicates that a single base-pair mismatch within the crRNA: target RNA duplex is sufficient to impair Cas13a HEPN nuclease activation, despite having minimal effect on Cas13a binding affinity (**Fig. 4B, C**).

Conversely, three out of four mismatches within the 9-12 region resulted in more robust Cas13a HEPN-nuclease activation relative to perfectly complementary RNA target sequences, indicating that single mismatches in this region do not preclude HEPN activation. In fact, these target RNAs promote higher catalytic activity (**Fig. 6A**), similar to our observations for an activator-RNA containing quadruple mismatches in positions 5-8. Comparing this result to the high-throughput mismatch profiling data (**Fig. 4**) where mismatches in this region had the greatest negative effect on binding affinity, reiterates that base-pair mismatches in some cases may act to improve HEPN nuclease catalytic activity.

Together, these two observations prompted us to revisit the high-throughput mismatch profiling data to see whether implementing a ranking-based approach would detect enrichment or depletion of base-pairing interactions regardless of whether the crRNA: target RNA duplex contained single or multiple mismatches. Specifically, we ranked all λ4-library members from the 100 nM Cas13a:crRNA-λ4 sample (**Fig. 3**) based on their fold-change enrichment / depletion compared to the apo-Cas13a sample. We then separately computed the nucleotide frequency by position for the top 500 members and the bottom 500 members of the list and used this to compute the frequency difference between the two subsets (**Fig. 6B**). To deal with any possible heterogeneity in the background nucleotide distribution, we performed permutation tests using shuffled lists of sequences as a null hypothesis (see **Methods**). This analysis identified several positions within the crRNA: activator-RNA duplex that exhibit interesting nucleotide preferences. Most striking was the substantial depletion at positions 15 and 16 of guanine nucleotides that

are complementary to the crRNA, and a slight enrichment of the remaining three nucleotide identities at these positions. This observation indicates that Cas13a:crRNA binds more tightly on average, to members of the RNA library that allow for mismatches in positions 15 and/or 16 between crRNA and the target RNA, even in the presence of other mismatches, in agreement with the filter-binding data in Fig. 2.

## DISCUSSION

Our investigation of specificity requirements for CRISPR-Cas13a RNA-activated RNA cleavage found that mismatches between crRNA and activator-RNA affect RNA binding affinity and HEPN-nuclease activation differently depending on mismatch position. These results reveal a proofreading mechanism following initial crRNA: activator-RNA base pairing that serves as a specificity checkpoint controlling HEPN-nuclease activation (**Fig. 6c).** Our data support a model for Cas13a target search that involves activator-RNA base-pairing with a 'mismatch hyper-sensitive' region (activator-RNA nucleotides 7-12) and a 'HEPN-nuclease switch' region (activator-RNA nucleotides 13-16). Target search by crRNA-loaded LbuCas13a requires activator-RNA base-pairing in the mismatch hyper-sensitive region. While complementarity at these positions is absolutely necessary, additional contributions from supplementary base-pairing at positions 1-6 and 17-20 are required for stable binding. After the activator is bound, HEPN-nuclease activation is gated by LbuCas13a detecting base-pairing between the crRNA and the HEPN-nuclease switch region of the activator-RNA, triggering HEPN domain conformational rearrangement and formation of a catalytically competent active-site[7,8]. Mismatches within the HEPN-nuclease switch region may block conformational change by trapping Cas13a in a non-catalytic yet more tightly-bound activator-RNA-associated state. We propose that mismatches in activator nucleotides 13-16 bypass an energetically unfavorable rearrangement that occurs when a fully complementary RNA activator binds Cas13a. Conversely, it appears that even though mismatches in regions 5-12 and 17-20 are

energetically unfavorable with respect to binding affinity, they are still able to promote robust HEPN-nuclease activity. This model is supported by data for both Lsh-Cas13a[3] and Lwa-Cas13[9] where mismatches in regions of the crRNA equivalent to activator-RNA nucleotides 7-12 and 13-16/15-16 compromise HEPN-nuclease activity. Furthermore, our data and model rationalize the mismatch scheme developed to maximize SNP sensitivity within the Lwa-Cas13a RNA detection platform[9]. Taken together, these similarities suggest broad conservation of activator-RNA binding and HEPN activation mechanisms (and their sensitivity to mismatches) across the U-cleaving Cas13a family. Subtle variation in the magnitude or position of these effects may exist between homologs, given the broad range of activator sensitivities and rates of *trans*-ssRNA cleavage observed across this family[3,4,6].

A number of nucleic-acid guided proteins employ accessible, pre-ordered 'seed' regions within their guides for target search (reviewed in [13,11]). The binary structures of Lsh-Cas13a[7], Lbu-Cas13a[8], and Lba-Cas13a[14], show the repeat proximal spacer nucleotides adopting a distorted 'U-turn' conformation within the nuclease lobe, while the repeat distal spacer nucleotides are solvent-exposed and disordered. Examining the crRNA-loaded structure of Lbu-Cas13a[8] shows that the 'mismatch sensitive' region we identify in this study is solvent-exposed in a near A-form conformation, but engaged in mismatched base-pairing with the 3'-end of the spacer. While the apparent pre-ordering and artefactual mismatch binding suggests this could be a seed for activator RNA binding, our data definitively identifies these nucleotides as the seed for Lbu-Cas13a.

The structure of Lbu-Cas13a:crRNA bound to activator-RNA[8] combined with our mismatch profiling, together provide the basis to rationalize how the HEPN-nuclease switch region might allosterically regulate the HEPN-nuclease. Both the 3' flank of the crRNA-repeat and the HEPN-1 domain undergo a concerted conformational change that drives the formation of a catalytically competent active-site[8]. Of note is S786, positioned within an alpha helix of HEPN-1 that is closely packed across the phosphate backbone of the activator-RNA HEPN-

nuclease switch region (nucleotides 13-16). S786 interacts through a bidentate hydrogen bonding interaction with a non-bridging oxygen from the phosphate of activator-RNA nucleotide 15, and C(-1) within the 3' flank of the crRNA-repeat. This interdependent sensing of activator-RNA nucleotide 15 by the crRNA 3' flank C(-1) via the HEPN-1 domain may drive conformational activation of the HEPN nuclease. Consistent with this, mismatches in this region (**Fig. 6**) would shift the positions of the activator RNA phosphates, disrupting the geometry of the crRNA: activator duplex leading to conformational inhibition of the HEPN-nuclease. This network of interactions might have implications for the mechanism by which the PFS[3] is able to gate HEPN nuclease activity, given that C(-1) has been implicated in sensing the PFS[7,8]. In addition to this interaction, a non-bridging oxygen from the phosphate of crRNA nucleotide 7 (Fig. 1 numbering), which base-pairs with nucleotide 14 in our activator-RNA, interacts with K558 from the Helical-2 domain, and mutation of this residue to alanine resulted in a significant defect in HEPN-nuclease activity[8]. Based on our data, one can imagine that the role of K558 and S786 is to sense simultaneously the positions of crRNA and activator-RNA phosphates, respectively, relative to one another only when they exist in an RNA duplex. It is worth noting that even single mismatches in an RNA:RNA duplex can have large effects on the local geometry of the duplex including positions of the phosphate groups[15]. Further studies will be required to understand how the dynamics of Cas13a:crRNA complexes allow them to overcome the energetic barriers imposed by crRNA organization to drive activator-RNA binding, and how sensing of the RNA duplex in this region by the helical-II and HEPN1 domains as well as the crRNA is able to elicit HEPN-nuclease activation.

Despite a fundamental difference in guide RNA organization, there are some striking similarities between Cas13a and Argonaute (Ago) proteins with respect to target ssRNA recognition and catalytic activation. First, relative to perfectly complementary targets, mismatches in guide nucleotides 9-12 improve mouse Ago2's affinity for target ssRNA[16] much like what we observed with mismatches in crRNA region 5-8 (activator-RNA nucleotides 13-16)

of Lbu-Cas13a. Second, base-pairing in this region is required for Ago2 proteins to achieve a catalytically competent 'slicer' conformation active for nucleic acid cleavage[17-19], similar to HEPN nuclease activation within Cas13a. And finally, while Ago2 proteins in animals have retained their ability to cleave perfectly-complementary target RNAs (i.e. through the siRNA pathway), they primarily use mismatched guides (as describe above) to bind target RNA and recruit host factors to carry out translational repression of target mRNAs (i.e. the miRNA pathway)[20]. It is tantalizing to speculate that this function could be used to modulate the expression of viral or self mRNAs upon viral infection, and/or during normal function, or modulate to virulence, much like the CRISPR-Cas9 system from *Francisella novicida*[21,22]. Alternatively, similar to other CRISPR systems, there might be a role for mismatched crRNA-spacer: ssRNA targets in the acquisition phase of CRISPR immunity, akin to primed spacer acquisition in type I CRISPR systems[23,24]. Additionally, our observation that mismatches in some cases seemed to be tolerated, if not preferred with respect to maximal HEPN nuclease activity might be indicative of an evolved mismatch tolerance required to maintain immunity when faced with the rapidly evolving nature of phage genomes. Indeed, recent data suggests that RNA-targeting by type III CRISPR-Cas systems exhibit relaxed interference specificity relative to type I and II DNA targeting CRISPR-Cas systems[25]. Nevertheless, further work is required to understand whether there is a role for both of these behaviors of within the native hosts of type VI CRISPR-Cas systems.

Understanding the specificity of Cas13a activator-RNA binding and subsequent HEPN nuclease activation is crucial for all current[4,6,9] and potential applications of Cas13a's programmable RNA-binding and/or cleavage capabilities. Our data enable rational design of crRNAs with optimal specificity and activity for each type of application, and underscore a need to consider the range of effects that Cas13a off-target recognition might have on Cas13a's ssRNA-binding and/or cleavage behavior, and ultimately the fate of the RNA pool present in each experiment.

## METHODS

### Lbu-dCas13a-Avitag overexpression and purification

HEPN nuclease deactivated Lbu-Cas13a (dCas13a: R472A, H477A, R1048A, H1053; Addgene plasmid #: 83485) was subcloned into pDuBir (a generous gift from Laura Lavery) in order to express a Lbu-dCas13a with a *N*-terminal $His_6$-MBP-TEV cleavage site tag, a C-terminal BirA biotinylation signal (Avitag[26]) as well as a separate BirA polypeptide, which biotinylates the Avitag. Expression of both polypeptides is driven by a T7 promoter. pDuBir-Lbu-dCas13a-avitag (Addgene plasmid #: 100817) was transformed into Rosetta2 *E. coli* cells grown in Terrific broth at 37 °C. *E. coli* cells were induced at $OD_{600}$ ~1 with 0.5 M ITPG and 50 μM D-biotin (in 10 mM bicine buffer pH8.3) was added as substrate for the biotinylation reaction. The temperature was then reduced to 16 °C for overnight expression of His-MBP-Cas13a-Avitag. Cells were subsequently harvested, resuspended in lysis buffer (50 mM Tris-HCl pH 7.0, 500 mM NaCl, 5% glycerol, 1 mM TCEP, 0.5mM PMSF, 50 μM D-biotin, and EDTA-free protease inhibitor (Roche)), lysed by sonication, and the lysates were clarified by centrifugation. Soluble His-MBP-Cas13a-Avitag was isolated by Ni-NTA affinity chromatography. While concurrently dialyzing into ion exchange buffer (50 mM Tris-HCl pH 7.0, 250 mM KCl, 5% glycerol, 1 mM TCEP), the His-MBP-dCas13a-Avitag containing eluate was incubated with TEV protease at 4 °C overnight in order to cleave off the $His_6$-MBP tag. dCas13a-Avitag was loaded onto a HiTrap SP column and eluted over a linear 0.25-1M KCl gradient. Cation exchange chromatography fractions containing dCas13a-Avitag were pooled and further purified via size-exclusion chromatography on an Superdex S200 column and stored in gel filtration buffer (20 mM Tris-HCl pH 7.0, 200 mM KCl, 5% glycerol, 1 mM TCEP) at -80C until required. Biotinylation efficiency was determined using a streptavidin-magnetic bead pulldown experiment, as described in ref. [27]. Wildtype Lbu-Cas13a (Addgene plasmid #: #83482) was expressed and purified as previously described[4], essentially as above except D-biotin was added during overexpression and purification.

**Preparation of RNA**

All 3'-prime fluorescein (6-FAM) labelled activator-RNAs were purchased from Integrated DNA technologies (IDT). All crRNAs and unlabeled activator-RNAs were transcribed *in vitro* from either PCR-amplified double-stranded DNA templates (crRNAs) or single-stranded DNA templates with an annealed dsDNA T7 promoter region (IDT). All templates contained a sequence encoding a hepatitis delta virus (HDV) ribozyme 3′ to the desired RNA product to ensure 3' end homogeneity. Transcriptions were carried out as described previously[28] with the following modifications: RNA were transcribed at 37°C for 3 h, followed by treatment with RQ1 RNase-free DNase (Promega) and 5 mM $MgCl_2$ (37°C, 1 hr) to remove DNA templates and drive ribozyme cleavage, respectively. Transcription reactions were purified by 15% denaturing polyacrylamide gel electrophoresis (PAGE), activator-RNAs were resuspended in cleavage buffer (20 mM HEPES pH 6.8, 50 mM KCl, 5 mM $MgCl_2$, and 5% glycerol), while crRNAs were first resuspended in RNase-free water, then prior complexing with Lba-Cas13a, crRNAs were annealed in 1X annealing buffer (10 mM Na-HEPES pH 7.4, 30 mM KCl, 1.5 mM $MgCl_2$) by heating at 75°C for 5 min followed by rapid cooling on ice. For filter-binding experiments, 5' triphosphates were removed from activator-RNAs by calf intestinal phosphate (New England Biolabs) prior to radiolabeling and activator-RNAs were then 5'-end labeled using T4 polynucleotide kinase (New England Biolabs) and [$\gamma$-$^{32}$P]-ATP (Perkin Elmer) as described previously[28].All oligonucleotide sequences used in this study can be found in Supplementary Data Table 2.

**Fluorescence anisotropy binding assays**

Lbu-dCas13a-avitag:crRNA complexes were preassembled by incubating 1μM of Lbu-dCas13a with 1.1 μM of crRNA in modified RNA processing buffer (20 mM HEPES pH 6.8, 50 mM KCl, 5 mM $MgCl_2$, 10 μg $ml^{-1}$ BSA, 100 μg $ml^{-1}$ yeast tRNA, 0.01% Igepal CA-630 and 5%

glycerol) for 2 min at 37°C, followed by 8 min at 25°C. These complexes were then serially diluted multiple times in modified RNA processing buffer to evenly cover between 0 - 1 μM Lbu-dCas13a:crRNA. These complexes were then further diluted two-fold in modified RNA processing buffer in the presence of 40 nM (final concentration) of 3'-labelled 6-FAM activator-RNA. These reactions were incubated for one hour at 37°C before fluorescence anisotropy measurements were taken ($\lambda_{ex}$: 485 nm; $\lambda_{em}$: 535 nm). Fluorescence anisotropy values (mA) were fit using a solution of the quadratic binding equation using GraphPad Prism, to avoid any effects of protein depletion when [RNA] > $K_D$ (see equation (6) in ref. [12]). For each curve fit, the maximum anisotropy value was floated to allow for activator-RNA specific effects on magnitude of maximum anisotropy value. All experiments were carried out in triplicate. Dissociation constants and associated errors are reported in the Fig.s and/or Fig. legends.

**Filter-binding assays**

Lbu-dCas13a-avitag:crRNA complexes were preassembled as described above. These complexes were then serially diluted multiple times in modified RNA processing buffer to evenly cover between 0 - 1 μM Lbu-dCas13a:crRNA. These complexes were then further diluted two-fold in modified RNA processing buffer in the presence of ~0.5 nM (final concentration) of 5'-$^{32}$P-labelled activator-RNA. These reactions were incubated for one hour at 37°C prior to loading samples onto the dot-blot apparatus. Tufryn, Protran and Hybond-N+ were assembled onto a dot-blot apparatus in the order listed above. Prior to sample application, each membrane was washed twice with 50μL Equilibration Buffer (20 mM HEPES pH 6.8, 50 mM KCl, 5 mM $MgCl_2$ and 5% glycerol). 25 μL of each sample was loaded per well, followed by a single 50 μL Equilibration Buffer wash. Membranes were then briefly dried and visualized by phosphorimaging. Fraction-bound values were fit using a solution of the quadratic binding equation using GraphPad Prism, to avoid any effects of protein depletion when [RNA] > $K_D$ (see

equation (6) in ref. [12]). For each curve fit, the max fraction bound value was constrained be greater than 0.6 to ensure (a) that weakly bound activator-RNAs affinities where not overestimated, and (b) affinities were not underestimated for activator-RNA interactions where the fraction bound values plateaued at less than 1.0. All experiments were carried out in triplicate. Dissociation constants and associated errors are reported in the Fig.s and/or Fig. legends.

**Fluorescent ssRNA reporter nuclease assays**

LbuC2c2:crRNA complexes were preassembled by incubating 1μM of WT Lbu-Cas13a with 500 nM crRNA in modified RNA processing buffer for 2 min at 37°C, followed by 8 min at 25°C.. These complexes were then diluted to 100nM LbuC2c2: 50 nM crRNA in modified RNA processing buffer in the presence of 150 nM of RNAase-Alert substrate (Thermo-Fisher), 100 ng of HeLa total RNA and 100 pM of activator-RNA. These reactions were incubated in a fluorescence plate reader for up to 120 min at 37°C with fluorescence measurements taken every 5 min ($\lambda_{ex}$: 485 nm; $\lambda_{em}$: 535 nm). Background-corrected fluorescence values were obtained by subtracting fluorescence values obtained from reactions carried out in the absence of target ssRNA. Background normalized apparent cleavage rates were calculated using a single-exponential decay using GraphPad Prism and associated errors are reported in each Fig. as standard deviation error bars from three independent experiments.

**High-throughput mismatch profiling library design rationale and preparation**

Two distinct ssDNA 100-nt libraries containing a 34-nt 'target-region' comprising a partially-randomized library (76% frequency of on-target nt. and 8% frequency of each other nt. in each position) of either λ3 or λ4 sequences, flanked by illumina sequencing adaptor complementary sequences, a random 6-nt. 'Cluster ID' region for efficient cluster calling, and a

5' T7 promoter for *in vitro* transcription (**Fig. 3A**). The 34-nt target should theoretically cover all single- and pairwise- error combinations, as well as a substantial portion of triple- mismatch combinations (see Simulate_fragments.ipynb script github.com/OConnell-lab-URMC/High-throughput-RNA-mismatch-profiling to simulate mismatch distributions). Each library is barcoded twice: (1) during final PCR, a 'P7' index-containing adaptor was appended to allow for sample multiplexing within each sequencing experiment/run, and (2) during reverse transcription, a modified 'P5' adaptor containing a randomized 13-n.t. 'unique molecule identifier' was appended to correct for any library PCR bias. Each library was produced by in vitro transcription as described above from single-stranded DNA template (IDT) annealed to a short ssDNA to create a dsDNA T7 promoter region.

**High-throughput mismatch profiling ssRNA library dCas13a-avitag pull-down**

Lbu-dCas13a-avitag:crRNA complexes were preassembled as described above (excepted for Apo-Lbu-dCas13a-avitag samples, where the crRNA addition was omitted). λ3 and λ4 activator-RNA libraries (final concentration 100 nM of each) were added to modified RNA processing buffer with 40 U RNasin (Promega), and Lbu-dCas13a-avitag:crRNA complexes (10 or 100 nM Lbu-dCas13a-avitag) in a total volume of 100 µL and incubated at 37 °C for 1 h. This mixture was then added to 12 µL magnetic streptavidin beads (Dynabeads MyOne Streptavidin C1; Life Technologies) pre-equilibrated in modified RNA processing buffer and agitated at 4 °C for 1 h. Beads were then washed once with 1 mL wash buffer (20 mM Tris-HCl, pH 7.5, 50 mM NaCl, 5mM $MgCl_2$, 0.1% Triton X-100, 5% glycerol, 1mM DTT). Bound RNAs were eluted from dCas13a-avitag by DEPC-water/trizol extraction, and purified by phenol-chloroform extraction and ethanol precipitation.

**High-throughput mismatch profiling Illumina sequencing library construction**

RNA pellets from the library pulldown were resuspended with 15μL RNase-free water. 33% of the sample (or 0.5 pmol λ3 and λ4 activator-RNA library as input control) was annealed with 2pmol RT primer (see Supplementary Data Table 2 for sequence) and reverse transcribed using Superscript III reverse transcriptase (Invitrogen) following the manufacturer's protocol. The RT primer contains a 13-nt randomized sequence (instead of a standard P7 index) which serves as a unique molecular identifier (UMI) to enable accurate molecule counting. Excess primers were degraded with Exonuclease I (New England Biolabs), and products were purified by ethanol precipitation. cDNAs were subjected to limited PCR amplification (12 cycles) with Phusion polymerase (New England Biolabs), using the NEBNext library barcoding kit to append Illumina sequencing handles and sample barcode identifiers to the library fragments. The amplified library was then purified and size-selected using AmpureXP beads (Beckman / Agencourt) with a 1:1 ratio of beads to sample. Library quality was verified by Bioanalyzer prior to high-throughput Illumina sequencing.

**High-throughput mismatch profiling Illumina sequencing**

High-throughput RNA mismatch profiling library samples were pooled and sequenced on an Illumina HiSeq 4000 using RapidRun configuration, with 50bp single-end reads, standard P7 index cycles, and custom P5 index cycles to sequence the complete 13-mer UMI. Sequencing was performed at the Vincent J. Coates Genomics Sequencing Laboratory at UC Berkeley. All sequencing reads are available through the NCBI SRA at BioProject ID: PRJNA399726.

**High-throughput mismatch profiling sequencing analysis/bioinformatics**

Sequencing data was analyzed using custom python scripts. All scripts and analysis presented here is publically available at github.com/OConnell-lab-URMC/High-throughput-RNA-

mismatch-profiling. Reads were de-duplicated by counting instances of the 13 nucleotide unique molecular identifier, and de-duplicated reads were then assigned to one of the two (λ3 and λ4 activator-RNA) target sequences based on Hamming distance. The number of unique observations of each sequence variant was computed, and entries with low counts (fewer than 20 unique counts in each of 3 replicates for any given condition) were filtered out. Counts were normalized to the total number of unique reads in a given sample. Fold change between conditions was calculated as the ratio of normalized counts in a pulldown condition vs the apo-Cas13a condition, and uncertainty in the fold-change was calculated by error propagation / partial derivatives, using the standard deviation in counts (across 3 replicates). These fold-change values were used to compute Fig.s 4a-1f.

We identified highly enriched or depleted nucleotide / position combinations by ranking sequences by fold change (relative to the apo-Cas13a control) and computing the frequencies of each nucleotide at each position for the top 500 or bottom 500 sequence entries in this ranked list. To verify that the observed frequencies are significant, they were compared against a null hypothesis generated by permuting the order of this ranked list (i.e. shuffling the fold-changes), using 10,000 permutations. We then used a two-tailed t-test and a p-value cut-off of $p <= 0.001$ to identify frequency / position combinations that were significantly enriched / depleted in the fold change ranking.

## REFERENCES


1 Mohanraju, P., Makarova, K. S., Zetsche, B., Zhang, F., Koonin, E. V. & van der Oost, J. Diverse evolutionary roots and mechanistic variations of the CRISPR-Cas systems. *Science* **353**, aad5147, doi:10.1126/science.aad5147 (2016).

2 Wright, A. V., Nunez, J. K. & Doudna, J. A. Biology and Applications of CRISPR Systems: Harnessing Nature's Toolbox for Genome Engineering. *Cell* **164**, 29-44, doi:10.1016/j.cell.2015.12.035 (2016).

3 Abudayyeh, O. O., Gootenberg, J. S., Konermann, S., Joung, J., Slaymaker, I. M., Cox, D. B., Shmakov, S., Makarova, K. S., Semenova, E., Minakhin, L., Severinov, K., Regev, A., Lander, E. S., Koonin, E. V. & Zhang, F. C2c2 is a single-component programmable RNA-guided RNA-targeting CRISPR effector. *Science* **353**, aaf5573, doi:10.1126/science.aaf5573 (2016).

4 East-Seletsky, A., O'Connell, M. R., Knight, S. C., Burstein, D., Cate, J. H., Tjian, R. & Doudna, J. A. Two distinct RNase activities of CRISPR-C2c2 enable guide-RNA processing and RNA detection. *Nature* **538**, 270-273, doi:10.1038/nature19802 (2016).
5 Smargon, A. A., Cox, D. B., Pyzocha, N. K., Zheng, K., Slaymaker, I. M., Gootenberg, J. S., Abudayyeh, O. A., Essletzbichler, P., Shmakov, S., Makarova, K. S., Koonin, E. V. & Zhang, F. Cas13b Is a Type VI-B CRISPR-Associated RNA-Guided RNase Differentially Regulated by Accessory Proteins Csx27 and Csx28. *Mol. Cell* **65**, 618-630 e617, doi:10.1016/j.molcel.2016.12.023 (2017).
6 East-Seletsky, A., O'Connell, M. R., Burstein, D., Knott, G. J. & Doudna, J. A. RNA Targeting by Functionally Orthogonal Type VI-A CRISPR-Cas Enzymes. *Mol. Cell* **66**, 373-383 e373, doi:10.1016/j.molcel.2017.04.008 (2017).
7 Liu, L., Li, X., Wang, J., Wang, M., Chen, P., Yin, M., Li, J., Sheng, G. & Wang, Y. Two Distant Catalytic Sites Are Responsible for C2c2 RNase Activities. *Cell* **168**, 121-134 e112, doi:10.1016/j.cell.2016.12.031 (2017).
8 Liu, L., Li, X., Ma, J., Li, Z., You, L., Wang, J., Wang, M., Zhang, X. & Wang, Y. The Molecular Architecture for RNA-Guided RNA Cleavage by Cas13a. *Cell* **170**, 714-726 e710, doi:10.1016/j.cell.2017.06.050 (2017).
9 Gootenberg, J. S., Abudayyeh, O. O., Lee, J. W., Essletzbichler, P., Dy, A. J., Joung, J., Verdine, V., Donghia, N., Daringer, N. M., Freije, C. A., Myhrvold, C., Bhattacharyya, R. P., Livny, J., Regev, A., Koonin, E. V., Hung, D. T., Sabeti, P. C., Collins, J. J. & Zhang, F. Nucleic acid detection with CRISPR-Cas13a/C2c2. *Science* **356**, 438-442, doi:10.1126/science.aam9321 (2017).
10 Mackay, J. P., Font, J. & Segal, D. J. The prospects for designer single-stranded RNA-binding proteins. *Nat. Struct. Mol. Biol.* **18**, 256-261, doi:10.1038/nsmb.2005 (2011).
11 Jackson, R. N., van Erp, P. B., Sternberg, S. H. & Wiedenheft, B. Conformational regulation of CRISPR-associated nucleases. *Curr. Opin. Microbiol.* **37**, 110-119, doi:10.1016/j.mib.2017.05.010 (2017).
12 Roehrl, M. H., Wang, J. Y. & Wagner, G. A general framework for development and data analysis of competitive high-throughput screens for small-molecule inhibitors of protein-protein interactions by fluorescence polarization. *Biochemistry* **43**, 16056-16066, doi:10.1021/bi048233g (2004).
13 Nakanishi, K. Anatomy of RISC: how do small RNAs and chaperones activate Argonaute proteins? *Wiley Interdiscip Rev RNA* **7**, 637-660, doi:10.1002/wrna.1356 (2016).
14 Knott, G. J., East-Seletsky, A., Cofsky, J. C., Holton, J. M., Charles, E., O'Connell, M. R. & Doudna, J. A. Guide-bound structures of an RNA-targeting A-cleaving CRISPR-Cas13a enzyme. *Nat. Struct. Mol. Biol.* **advance online publication**, doi:10.1038/nsmb.3466
http://www.nature.com/nsmb/journal/vaop/ncurrent/abs/nsmb.3466.html#supplementary-information (2017).
15 Pan, B., Mitra, S. N. & Sundaralingam, M. Structure of a 16-mer RNA duplex r(GCAGACUUAAAUCUGC)2 with wobble C.A+ mismatches. *J. Mol. Biol.* **283**, 977-984, doi:10.1006/jmbi.1998.2140 (1998).
16 Wee, L. M., Flores-Jasso, C. F., Salomon, W. E. & Zamore, P. D. Argonaute divides its RNA guide into domains with distinct functions and RNA-binding properties. *Cell* **151**, 1055-1067, doi:10.1016/j.cell.2012.10.036 (2012).
17 Martinez, J. & Tuschl, T. RISC is a 5' phosphomonoester-producing RNA endonuclease. *Genes Dev.* **18**, 975-980, doi:10.1101/gad.1187904 (2004).
18 Haley, B. & Zamore, P. D. Kinetic analysis of the RNAi enzyme complex. *Nat. Struct. Mol. Biol.* **11**, 599-606, doi:10.1038/nsmb780 (2004).
19 Schwarz, D. S., Ding, H., Kennington, L., Moore, J. T., Schelter, J., Burchard, J., Linsley, P. S., Aronin, N., Xu, Z. & Zamore, P. D. Designing siRNA that distinguish between

genes that differ by a single nucleotide. *PLoS Genet.* **2**, e140, doi:10.1371/journal.pgen.0020140 (2006).
20 Bartel, D. P. MicroRNAs: target recognition and regulatory functions. *Cell* **136**, 215-233, doi:10.1016/j.cell.2009.01.002 (2009).
21 Sampson, T. R., Napier, B. A., Schroeder, M. R., Louwen, R., Zhao, J., Chin, C. Y., Ratner, H. K., Llewellyn, A. C., Jones, C. L., Laroui, H., Merlin, D., Zhou, P., Endtz, H. P. & Weiss, D. S. A CRISPR-Cas system enhances envelope integrity mediating antibiotic resistance and inflammasome evasion. *Proc Natl Acad Sci U S A* **111**, 11163-11168, doi:10.1073/pnas.1323025111 (2014).
22 Sampson, T. R., Saroj, S. D., Llewellyn, A. C., Tzeng, Y. L. & Weiss, D. S. A CRISPR/Cas system mediates bacterial innate immune evasion and virulence. *Nature* **497**, 254-257, doi:10.1038/nature12048 (2013).
23 Datsenko, K. A., Pougach, K., Tikhonov, A., Wanner, B. L., Severinov, K. & Semenova, E. Molecular memory of prior infections activates the CRISPR/Cas adaptive bacterial immunity system. *Nat Commun* **3**, 945, doi:10.1038/ncomms1937 (2012).
24 Fineran, P. C., Gerritzen, M. J., Suarez-Diez, M., Kunne, T., Boekhorst, J., van Hijum, S. A., Staals, R. H. & Brouns, S. J. Degenerate target sites mediate rapid primed CRISPR adaptation. *Proc Natl Acad Sci U S A* **111**, E1629-1638, doi:10.1073/pnas.1400071111 (2014).
25 Pyenson, N. C., Gayvert, K., Varble, A., Elemento, O. & Marraffini, L. A. Broad Targeting Specificity during Bacterial Type III CRISPR-Cas Immunity Constrains Viral Escape. *Cell Host Microbe*, doi:10.1016/j.chom.2017.07.016 (2017).
26 Beckett, D., Kovaleva, E. & Schatz, P. J. A minimal peptide substrate in biotin holoenzyme synthetase-catalyzed biotinylation. *Protein Sci.* **8**, 921-929, doi:10.1110/ps.8.4.921 (1999).
27 O'Connell, M. R., Oakes, B. L., Sternberg, S. H., East-Seletsky, A., Kaplan, M. & Doudna, J. A. Programmable RNA recognition and cleavage by CRISPR/Cas9. *Nature* **516**, 263-266, doi:10.1038/nature13769 (2014).
28 Sternberg, S. H., Haurwitz, R. E. & Doudna, J. A. Mechanism of substrate selection by a highly specific CRISPR endoribonuclease. *RNA* **18**, 661-672, doi:10.1261/rna.030882.111 (2012).

## FIGURE LEGENDS

**Fig. 1. Position of crRNA: activator-RNA mismatches differentially impacts Lbu-Cas13a activator-RNA binding affinity.** (**A**) Schematic of the Lbu-Cas13a crRNA: ssRNA activator interaction highlighting the different activator-RNAs tested in this study. (**B**) Fluorescence anisotropy binding curves for the interaction between Lbu-dCas13a:crRNA-λ4 and λ4 activator-RNAs depicted in (A). Binding data fits (see Methods) shown as solid lines. Error-bars represent the s.d of the anisotropy from three independent experiments. Dissociation constants ($K_D$) and their associated standard errors (SE) from three independent experiments are shown in the Fig.

(**C**) Data from (B) normalized as a percentage binding affinity relative to the affinity to a perfectly complementary activator-RNAs (RNA #1).

**Fig. 2. Activator-RNAs with mismatches in nucleotide positions 13-16 are bound more tightly by Lbu-Cas13a then perfectly complementary activator-RNA.** (**A**) filter binding curves for curves for the interaction between Lbu-dCas13a:crRNA-λ3 (or λ4) and λ3 (or λ4) activator-RNAs, respectively. Position of mismatches in RNA #1, RNA #4, and RNA #5 are depicted in Fig. 1A. Binding data fits (see Methods) shown as solid lines. A binding curve could not be fit λ3 RNA #4. Error-bars represent the s.d of the fraction bound from three independent experiments. Dissociation constants ($K_D$) and their associated standard errors (SE) from three independent experiments are shown in the Fig. (**B**) Data from (A) normalized as a percentage binding affinity relative to the affinity to a perfectly complementary λ4 activator-RNA (λ4 RNA#1) with error bars representing the normalized s.d from three independent experiments.

**Fig. 3. Development of High-throughput RNA mismatch profiling to determine activator-RNA binding preferences for Lbu-Cas13a.** (**A**) Schematic of the high-throughput RNA mismatch profiling library design. T7 promotor: for producing RNA via in vitro transcription. This sequence lost upon *in vitro* transcription of the DNA template (see Methods). P7 Seq Primer: binding site for the P7 Illumina sequencing primer that includes an index for multiplexing library samples. 34-mer Partially Randomized Library: either λ3 or λ4 target sequence with 76% frequency of on-target nucleotides and 8% frequency of each other nt. at each position. 20nt-spacer crRNA binding site positon is shown. Cluster ID: 6-mer randomized region to ensure robust Illumina sequence cluster calling. P5 Seq Primer + PCR UMI: a modified P5 Illumina sequencing primer binding site that incorporates a 13-mer randomized 'unique molecular identifier' (UMI) within the index region to filter out PCR duplicates downstream. (**B**) Schematic of the high-throughput RNA mismatch profiling pulldown workflow. (**C**) mismatch distribution

profile for the input λ3- and λ4- activator-RNA libraries. (**D**) A scatter plot displaying fold-change in abundance of each λ3- and λ4- libraries member between samples with 10 nM crRNA-bound Cas13a (vs. crRNA-unbound apo-Cas13a) programmed to target λ3-or λ4- ssRNA sequences. (**E**) A scatter plot displaying fold-change in abundance of each λ3- and λ4- libraries member between samples with 100 nM crRNA-bound Cas13a (vs. crRNA-unbound apo-Cas13a) programmed to target λ3- or λ4- ssRNA sequences.

**Fig. 4. dCas13a high-throughput RNA mismatch profiling confirms binding mismatch sensitivity hotspots for single- and adjacent double- mismatched activator-RNAs.** Fold-change enrichment analysis (relative to Apo-dCas13a) for activator-RNA members of both libraries that contain only a single mismatch across the shown spacer-targeting and flanking regions when (**A**) targeting λ4 activator-RNA library with 10 nM dCas13a:crRNA-λ4, (**B**) targeting λ3 activator-RNA library with 10 nM dCas13a:crRNA-λ3, (**C**) targeting λ4 activator-RNA library with 100 nM dCas13a:crRNA-λ4, (**D**) targeting λ3 activator-RNA library with 100 nM dCas13a:crRNA-λ3. Fold-change enrichment/depletion analysis (relative to Apo-dCas13a) for activator-RNA members of both libraries that contain only adjacent double mismatches as a sliding window across the shown spacer-targeting and flanking regions when **(E)** targeting λ4 activator-RNA library with 100 nM dCas13a:crRNA-λ4, and **(F)** targeting λ3 activator-RNA library with 100 nM dCas13a:crRNA-λ3. In each plot, each individual point in a position represents each unique mismatch combination possible in that position, the solid line represents as an average fold-enrichment (of all mismatches combinations) in that position, and the error bars represent the s.d. from three independent experiments. The fold-change enrichment of the perfectly complementary targets in each experiment are show on the right of each plot

**Fig. 5. Position of crRNA: activator-RNA mismatches differentially impact Lbu-dCas13a HEPN ssRNA nuclease activation** (**A**) representative time course of background corrected fluorescence measurements generated by Lbu- Cas13a:crRNA- λ4 activation by the addition of 100 pM λ4 activator-RNAs with either zero (RNA #1) or four consecutive mismatches across the crRNA-spacer targeting region. See Fig. 1A for more information. Exponential fits are shown as solid lines. (**B**) Data from (A) normalized as a percent cleavage rate relative to a perfectly complementary λ4 activator-RNA (λ4 RNA#1) with error bars representing the normalized s.d from three independent experiments.

**Fig. 6. Lbu-Cas13a HEPN nuclease activation is gated by 'unfavorable' base-pairing within the context of the Cas13a crRNA: activator-RNA interaction.** (**A**) Background corrected apparent cleavage rates generated by Lbu- Cas13a:crRNA- λ4 activation by the addition of 100 pM λ4 activator-RNA with either zero or single mismatches in positions 9-16 across the crRNA-spacer targeting region. Apparent cleavage rates are plotted as a normalized percentage cleavage rate relative to a perfectly complementary λ4 activator-RNA (λ4 RNA#1) with error bars representing the normalized s.d from three independent experiments. (**B**) Change in nucleotide frequency in each position shown between the top 500 enriched members and the bottom 500 enriched members (as determined by fold-change relative to Apo Lbu-dCas13a) for 100 nM dCas13a: crRNA-λ4 targeting the λ4 activator-RNA library. The λ4 activator-RNA target sequence is shown above. Values with $p \geq 0.001$ in the permutation test are not shown. (**C**) A Revised model for Lbu-Cas13a HEPN nuclease activation based on the key findings in this study.

**SUPPLEMENTAL FIGURES/DATA**

**Supplementary Fig. 1. Additional crRNA: activator-RNA mismatch binding affinity experiments with a 24-nt. crRNA-spacer.** (**A**) Schematic of the Lbu-Cas13a crRNA-24-nt.-spacer: activator-RNA interaction highlighting the different activator-RNAs tested in this study. (**B**) Fluorescence anisotropy binding curves for the interaction between Lbu-dCas13a:crRNA-λ4 (with a 24-nt. spacer) and λ4 activator-RNAs depicted in (A). Binding data fits (see Methods) shown as solid lines. Error-bars represent the s.d of the anisotropy from three independent experiments. Dissociation constants ($K_D$) and their associated standard errors (SE) from three independent experiments are shown in the Fig. (**C**) Data from (B) normalized as a percentage binding affinity relative to the affinity to a perfectly complementary activator-RNA (RNA #1).

**Supplementary Fig. 2. Streptavidin pulldown experiment to confirm Lbu-dCas13a-avitag biotinylation.** (**A**) Streptavidin bead binding assay with biotinylated Lbu-dCas13a-avitag or non-biotinylated (non-biot.) MBP-Lbu-Cas13a and streptavidin magnetic beads. 8.5 pmol of with biotinylated Lbu-dCas13a-avitag or non-biotinylated (non-biot.). Each were separately mixed with 12 µL streptavidin magnetic beads and incubated in modified processing buffer at 37C for 1 h, followed by three washes with modified processing buffer and elution in boiling SDS-PAGE loading buffer. Input, incubation flow-through & elution for each sample were analyzed using SDS-PAGE. Biotinylated Lbu-dCas13a-avitag only remains specifically bound to the beads after biotin labeling.

**Supplementary Fig. 3.** A scatter plot displaying fold-change in abundance of each λ3- and λ4-libraries member between samples with (**A**) 10 nM crRNA-bound Cas13a (vs. input library) or (**B**) 10 nM crRNA-bound Cas13a (vs. input library) programmed to target λ3- or λ4- ssRNA sequences.

**Supplementary Fig. 4. Additional crRNA: activator-RNA mismatch HEPN-nuclease activation experiments with a 24-nt. crRNA-spacer.** (**A**) representative time course of background corrected fluorescence measurements generated by Lbu- Cas13a:crRNA- λ4 (with a 24-nt. spacer) activation by the addition of 100 pM λ4 activator-RNAs with either zero (RNA #1) or four consecutive mismatches across the crRNA-spacer targeting region. See Fig. 1A for more information. Exponential fits are shown as solid lines. (**B**) Data from (A) normalized as a percent cleavage rate relative to a perfectly complementary λ4 activator-RNA (λ4 RNA#1) with error bars representing the normalized s.d from three independent experiments.

**Supplementary Data Fig. 5. Additional data showing that four mismatches in positions 13-16 of crRNA**: **activator-RNA duplex abolish Lbu-dCas13a HEPN ssRNA nuclease activation.** (**A**) representative time courses of background corrected fluorescence measurements generated by Lbu-Cas13a activation by the addition of 100 pM activator-RNAs with either zero (RNA #1) or four consecutive mismatches in positions 13-16 of the crRNA-spacer targeting region (RNA #5). Three different RNA sequences (λ3, λ4 & λ5) and their corresponding crRNAs were tested. Exponential fits are shown as solid lines, except no fit is shown for λ3 RNA#1 as this reaction when to completion before accurate measurements could be taken. (**B**) Schematic of the Lbu-Cas13a crRNA: ssRNA activator interaction highlighting the the position of the mismatches between RNAs #1 and #5 with standard crRNA λ4 (std) and variant crRNA λ4 (var.: with mismatches in positions 5-8 of the crRNA). (**C**) representative time courses of background corrected fluorescence measurements generated by crRNA λ4 std- or crRNA λ4 var.- loaded Lbu-Cas13a activated by the addition of 100 pM activator- RNA #1 or RNA #5. Exponential fits are shown as solid lines.

**Supplementary Data Fig. 6. Single mismatch HEPN-nuclease activation time course data.** (**A**) representative time course of background corrected fluorescence measurements generated

by Lbu-Cas13a: crRNA- λ4 activation by the addition of 100 pM λ4 activator-RNAs with either zero or single mismatches within positions 9-16 of crRNA-spacer targeting region. See Fig. 1A for more information. Exponential fits are shown as solid lines.

**Supplementary Data Table 1. Illumina sequencing statistics for each high-throughput RNA mismatch profiling sample.** See attached excel spreadsheet.

**Supplementary Data Table 2. Oligonucleotide sequences used in this study.** See attached excel spreadsheet.

**ONLINE DATA**

All sequencing reads are available through the NCBI SRA BioProject ID: PRJNA399726, and all scripts and analysis presented here is publically available at github.com/OConnell-lab-URMC/High-throughput-RNA-mismatch-profiling.

**ACKNOWLEDGEMENTS**

We thank Nan Ma and Kaihong Zhou for technical assistance. We thank Dmitri Ermolenko, and members of the O'Connell and Doudna laboratories for helpful discussions. We thank Shana McDevitt and Dylan Smith for high throughput sequencing assistance and helpful advice. We thank Laura Lavery and Callison Alcott for the pDuBir plasmid and tips on in vivo biotinylation. This work used the Vincent J. Coates Genomics Sequencing Laboratory at UC Berkeley, supported by the NIH S10 OD018174 Instrumentation Grant. A.T. was supported by the HARC Center National Institutes of Health Grant P50GM82250 (N. Krogan, PI; J.A.D., co-PI). G.J.K. acknowledges support from the Howard Hughes Medical Institute. M.R.O. was supported by a C.J Martin Overseas Early Career Fellowship from the National Health and Medical Research Council of Australia. This work was supported in part by a Frontiers Science award from the

Paul Allen Institute to J.A.D, the National Science Foundation (MCB-1244557 to J.A.D.) and the National Institutes of Health Center for RNA Systems Biology Grant P50-GM102706 (J. Cate, P.I.; J.A.D., co-PI). J.A.D is an Investigator of the Howard Hughes Medical Institute. J.A.D. is a co-founder of Caribou Biosciences, Inc., Editas Medicine, and Intellia Therapeutics.

**AUTHOR CONTRIBUTIONS**

M.R.O, A.T, A.E-S designed the project with input from G.J.K and J.A.D.

M.R.O, A.T, A-E.S carried out experiments. A.T. developed the bioinformatics pipeline, and analyzed all the sequencing data with assistance from M.R.O. M.R.O drafted the manuscript with assistance from A.T., and all authors reviewed and edited the manuscript.

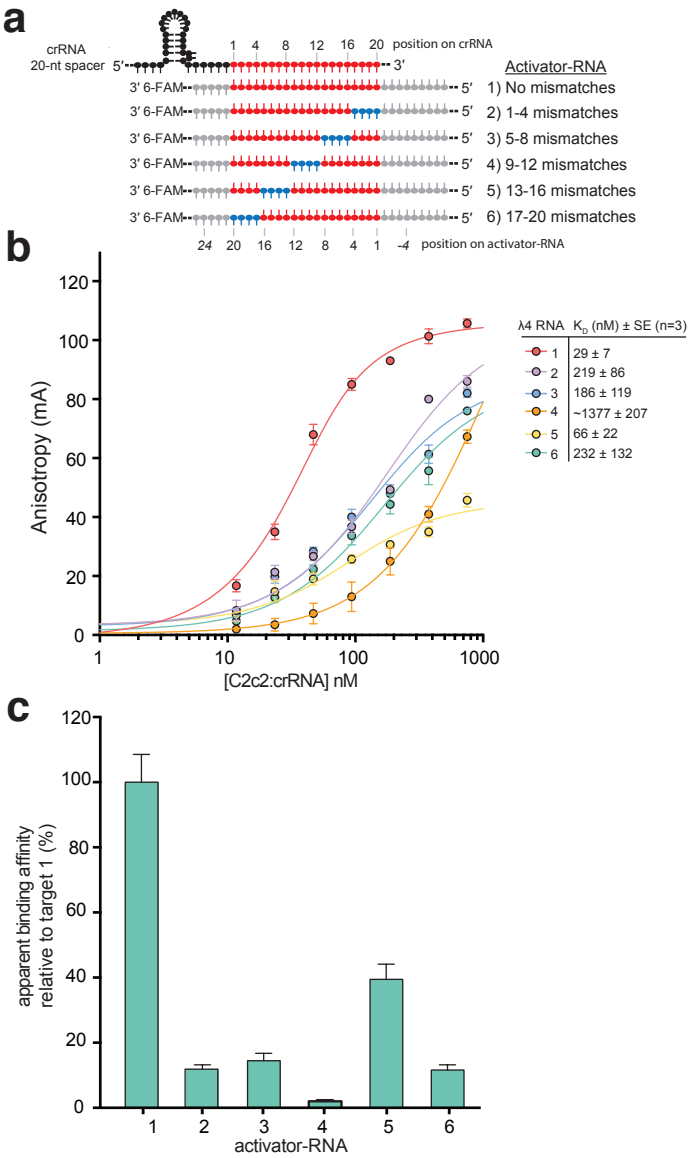
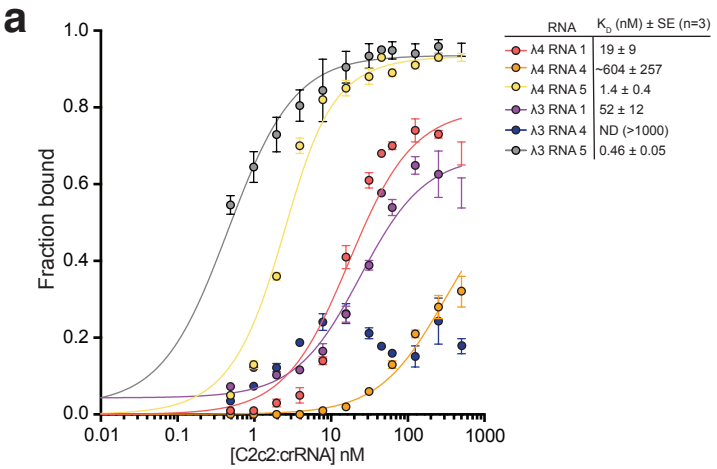

a
crRNA
20-nt spacer
position on crRNA
Activator-RNA
1) No mismatches
2) 1-4 mismatches
3) 5-8 mismatches
4) 9-12 mismatches
5) 13-16 mismatches
6) 17-20 mismatches
position on activator-RNA
b
Anisotropy (mA)
[C2c2:crRNA] nM
λ4 RNA
K_D (nM) ± SE (n=3)
1 29 ± 7
2 219 ± 86
3 186 ± 119
4 ~1377 ± 207
5 66 ± 22
6 232 ± 132
c
apparent binding affinity relative to target 1 (%)
activator-RNA

Figure 1


a
Fraction bound
[C2c2:crRNA] nM
RNA
$K_D$ (nM) ± SE (n=3)
λ4 RNA 1
19 ± 9
λ4 RNA 4
~604 ± 257
λ4 RNA 5
1.4 ± 0.4
λ3 RNA 1
52 ± 12
λ3 RNA 4
ND (>1000)
λ3 RNA 5
0.46 ± 0.05

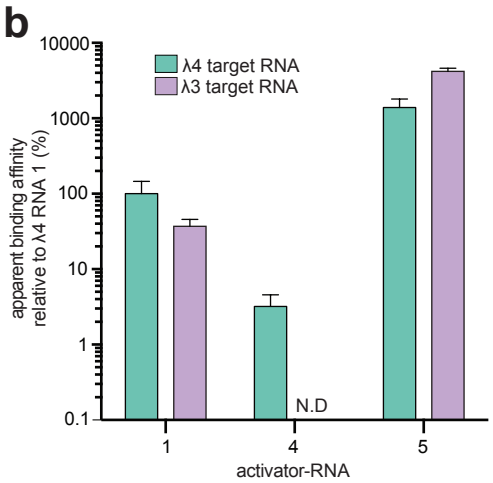

b
λ4 target RNA
λ3 target RNA
apparent binding affinity
relative to λ4 RNA 1 (%)
N.D
activator-RNA

Figure 2

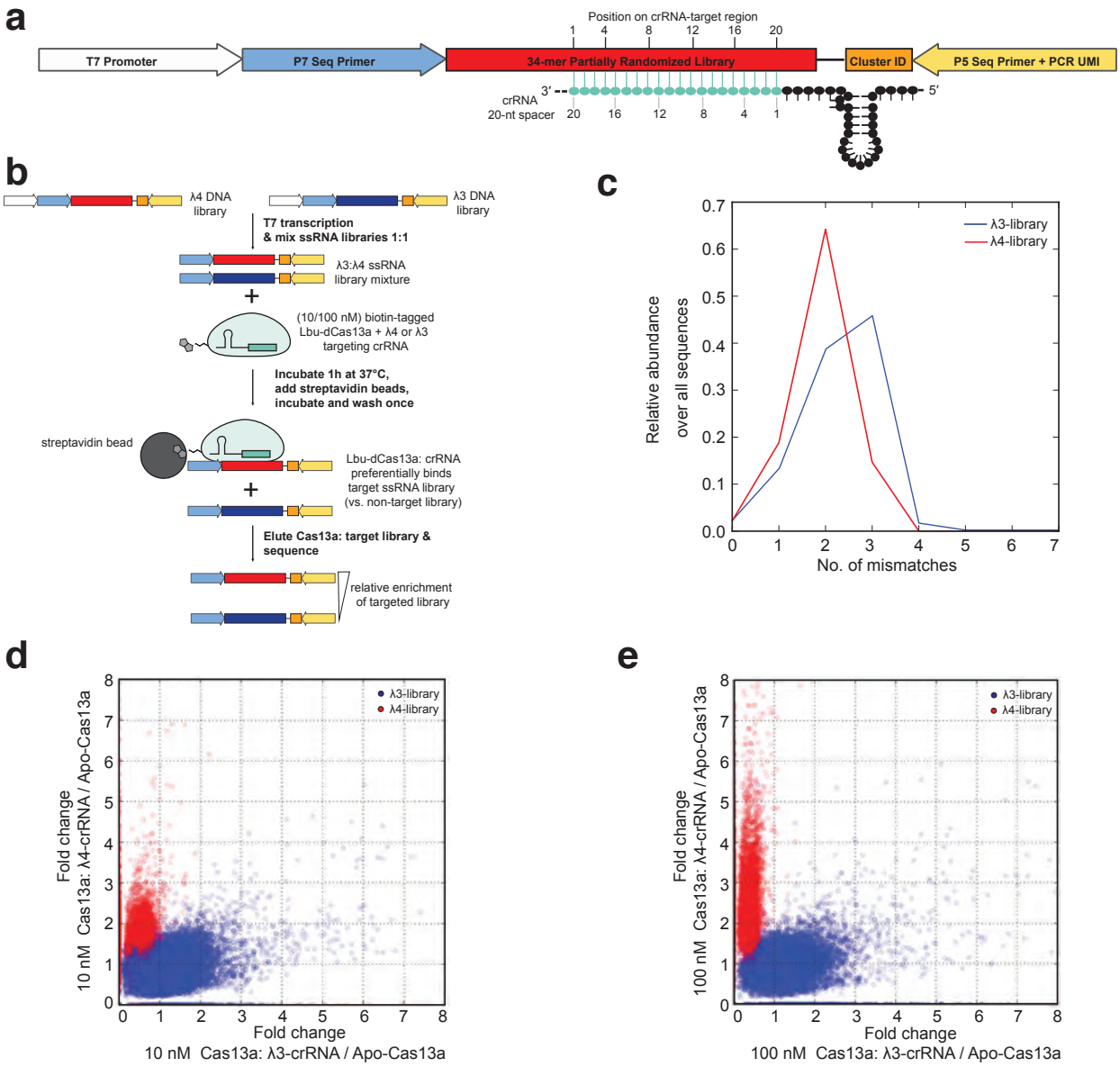

a
Position on crRNA-target region
1 4 8 12 16 20
T7 Promoter
P7 Seq Primer
34-mer Partially Randomized Library
Cluster ID
P5 Seq Primer + PCR UMI
crRNA
20-nt spacer
3′
5′
20 16 12 8 4 1
b
λ4 DNA library
λ3 DNA library
T7 transcription & mix ssRNA libraries 1:1
λ3:λ4 ssRNA library mixture
(10/100 nM) biotin-tagged Lbu-dCas13a + λ4 or λ3 targeting crRNA
Incubate 1h at 37°C, add streptavidin beads, incubate and wash once
streptavidin bead
Lbu-dCas13a: crRNA preferentially binds target ssRNA library (vs. non-target library)
Elute Cas13a: target library & sequence
relative enrichment of targeted library
c
λ3-library
λ4-library
Relative abundance over all sequences
No. of mismatches
d
λ3-library
λ4-library
Fold change
10 nM Cas13a: λ4-crRNA / Apo-Cas13a
Fold change
10 nM Cas13a: λ3-crRNA / Apo-Cas13a
e
λ3-library
λ4-library
Fold change
100 nM Cas13a: λ4-crRNA / Apo-Cas13a
Fold change
100 nM Cas13a: λ3-crRNA / Apo-Cas13a

Figure 3

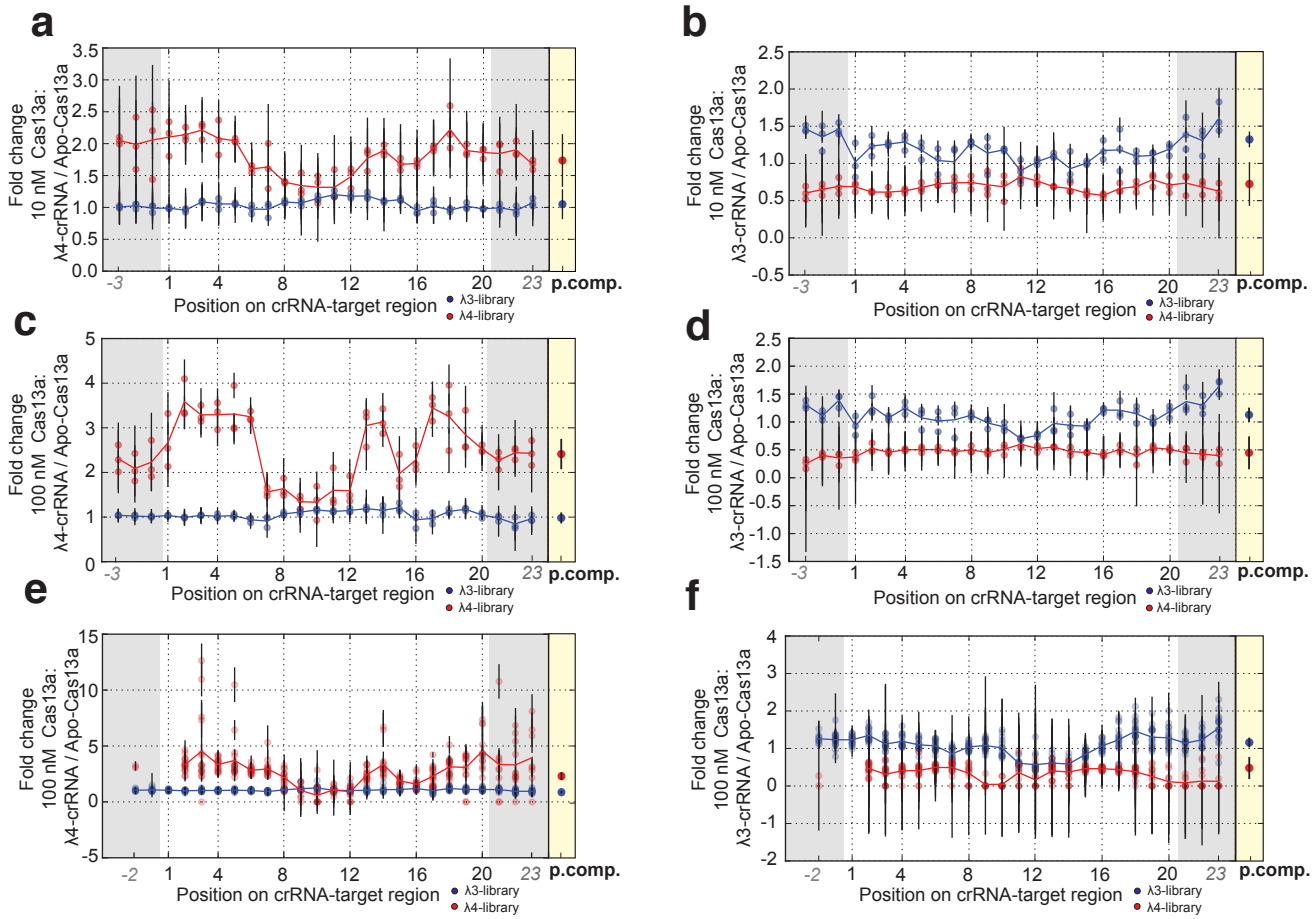

a
Fold change
10 nM Cas13a:
λ4-crRNA / Apo-Cas13a
Position on crRNA-target region
λ3-library
λ4-library
p.comp.
b
Fold change
10 nM Cas13a:
λ3-crRNA / Apo-Cas13a
Position on crRNA-target region
λ3-library
λ4-library
p.comp.
c
Fold change
100 nM Cas13a:
λ4-crRNA / Apo-Cas13a
Position on crRNA-target region
λ3-library
λ4-library
p.comp.
d
Fold change
100 nM Cas13a:
λ3-crRNA / Apo-Cas13a
Position on crRNA-target region
λ3-library
λ4-library
p.comp.
e
Fold change
100 nM Cas13a:
λ4-crRNA / Apo-Cas13a
Position on crRNA-target region
λ3-library
λ4-library
p.comp.
f
Fold change
100 nM Cas13a:
λ3-crRNA / Apo-Cas13a
Position on crRNA-target region
λ3-library
λ4-library
p.comp.

Figure 4

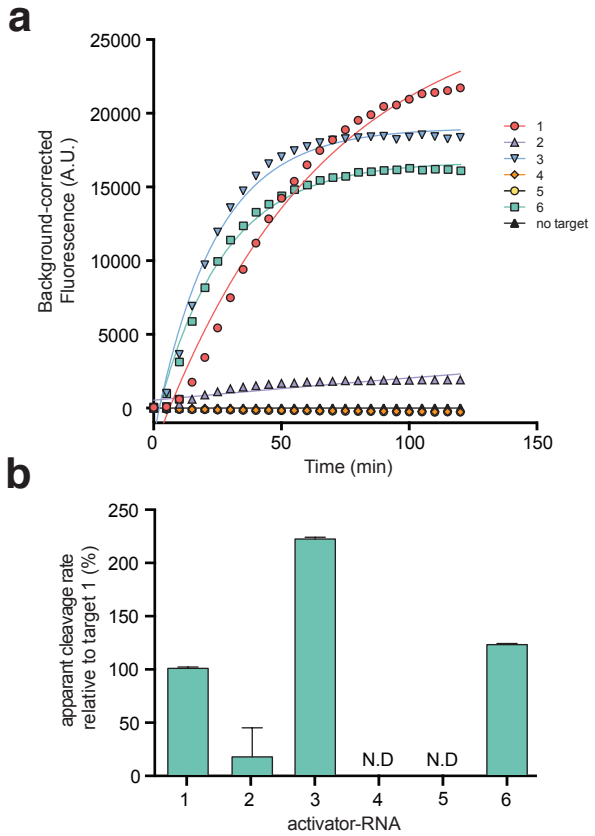

a
Background-corrected Fluorescence (A.U.)
25000
20000
15000
10000
5000
0
0
50
100
150
Time (min)
1
2
3
4
5
6
no target
b
apparant cleavage rate relative to target 1 (%)
250
200
150
100
50
0
N.D
N.D
1
2
3
4
5
6
activator-RNA

Figure 5

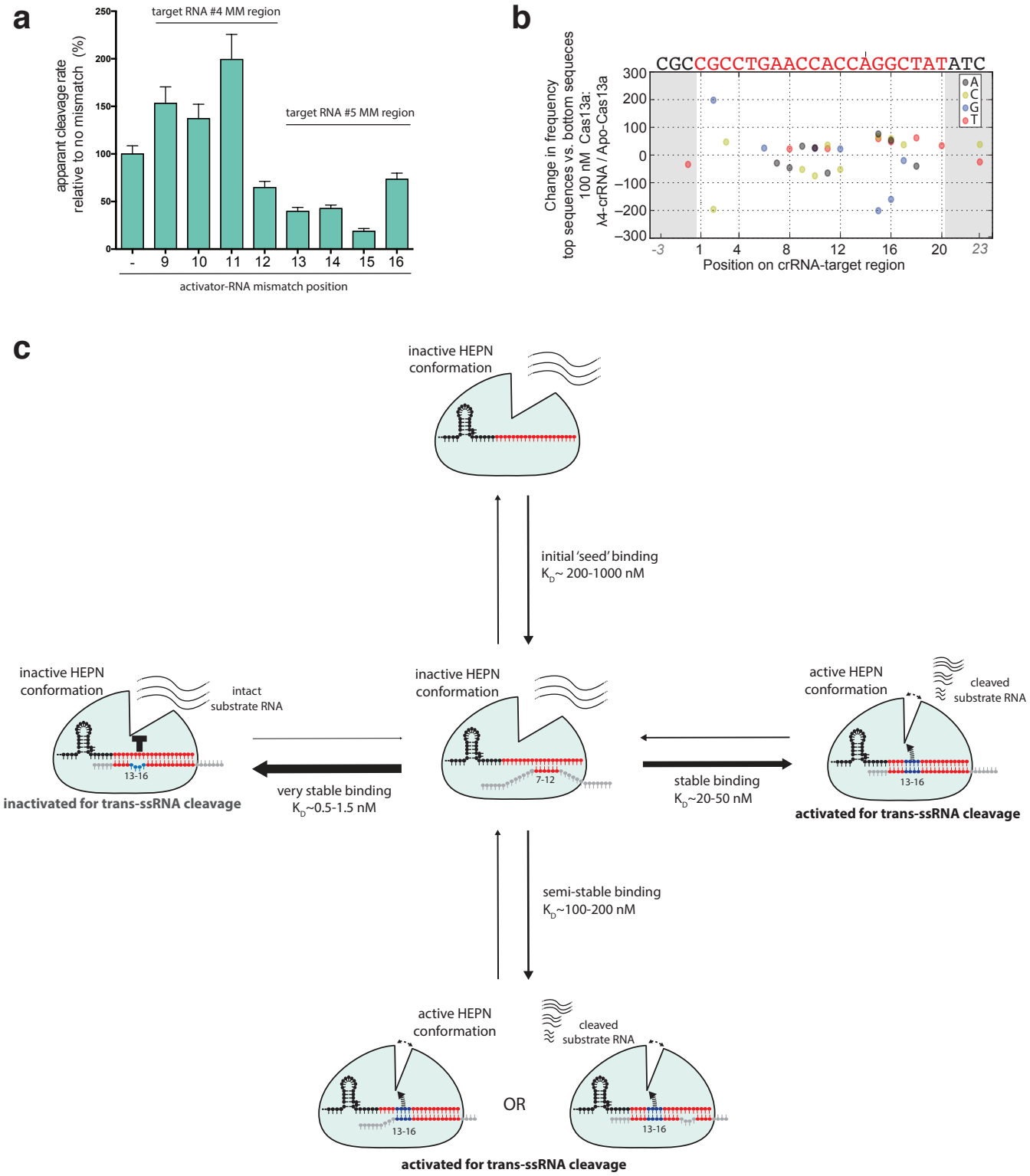
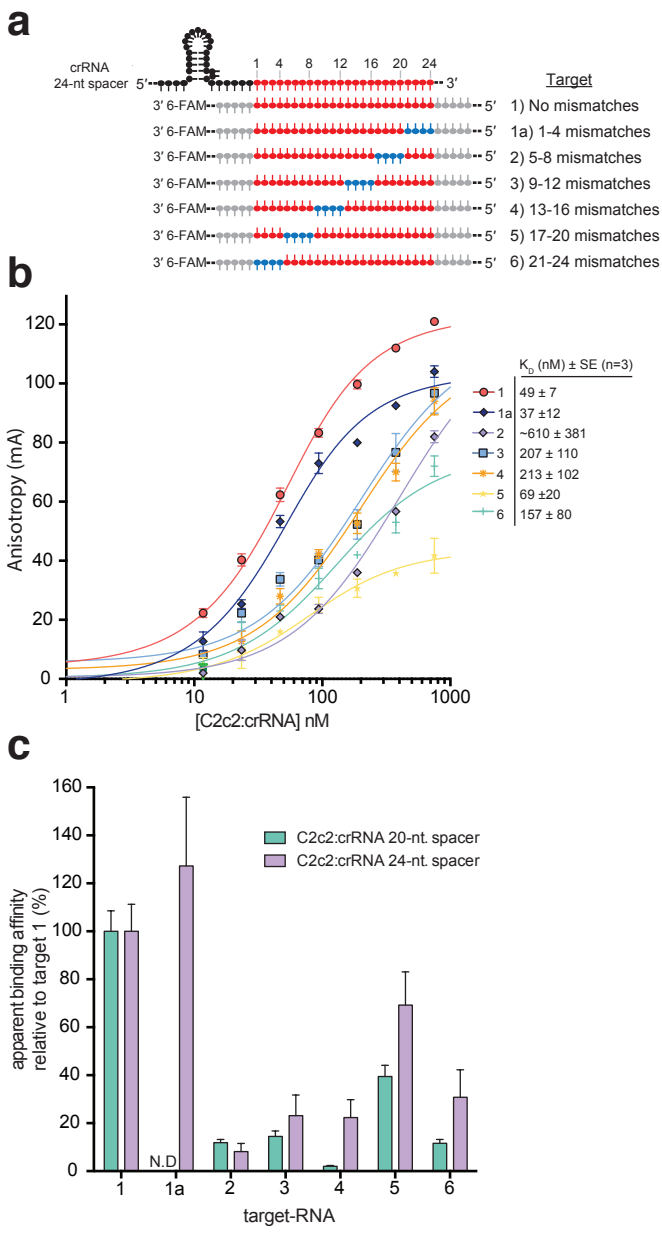
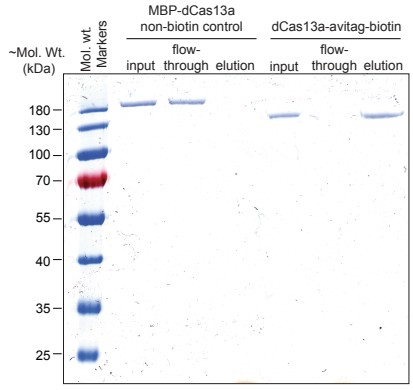

a
target RNA #4 MM region
target RNA #5 MM region
apparant cleavage rate relative to no mismatch (%)
250
200
150
100
50
0
-
9
10
11
12
13
14
15
16
activator-RNA mismatch position
b
CGCCGCCTGAACCACCAGGCTATATC
Change in frequency top sequences vs. bottom sequeces 100 nM Cas13a: λ4-crRNA / Apo-Cas13a
300
200
100
0
−100
−200
−300
-3
1
4
8
12
16
20
23
Position on crRNA-target region
A
C
G
T
c
inactive HEPN conformation
initial 'seed' binding
$K_D$~ 200-1000 nM
inactive HEPN conformation
intact substrate RNA
13-16
inactivated for trans-ssRNA cleavage
very stable binding
$K_D$~0.5-1.5 nM
inactive HEPN conformation
7-12
stable binding
$K_D$~20-50 nM
active HEPN conformation
cleaved substrate RNA
13-16
activated for trans-ssRNA cleavage
semi-stable binding
$K_D$~100-200 nM
active HEPN conformation
13-16
OR
cleaved substrate RNA
13-16
activated for trans-ssRNA cleavage

Figure 6

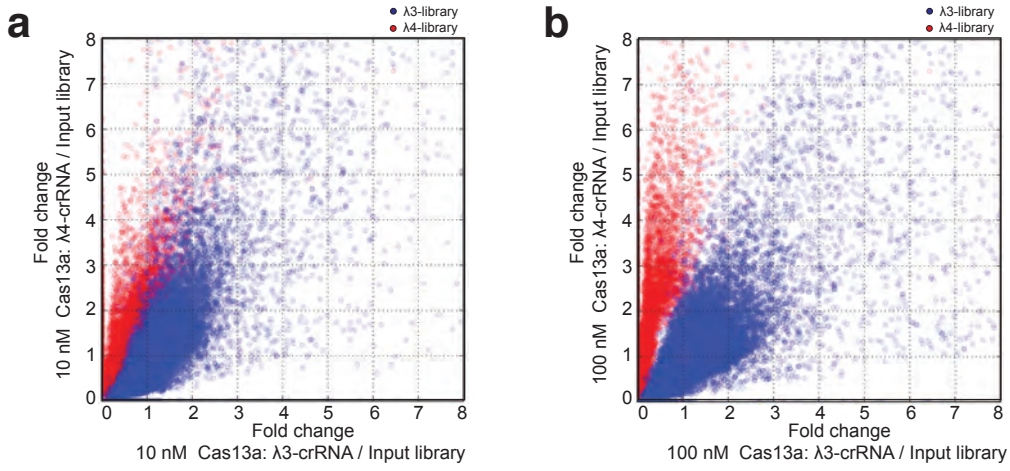



Supplementary Data Figure 1

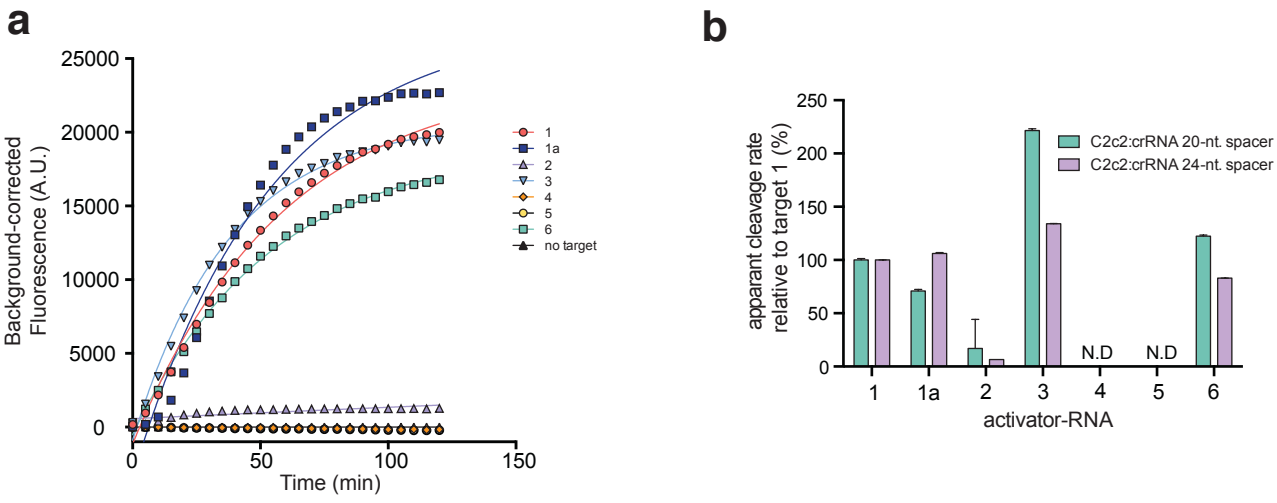
**a**



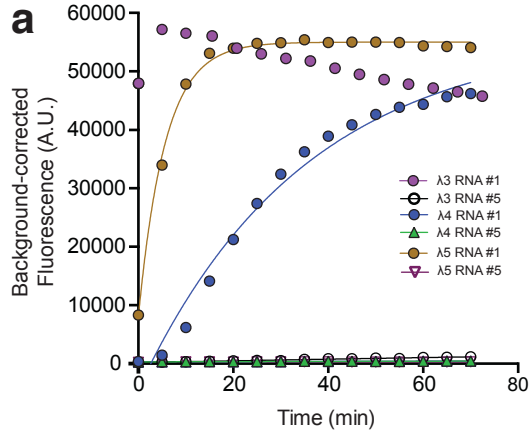

Supplementary Data Figure 2

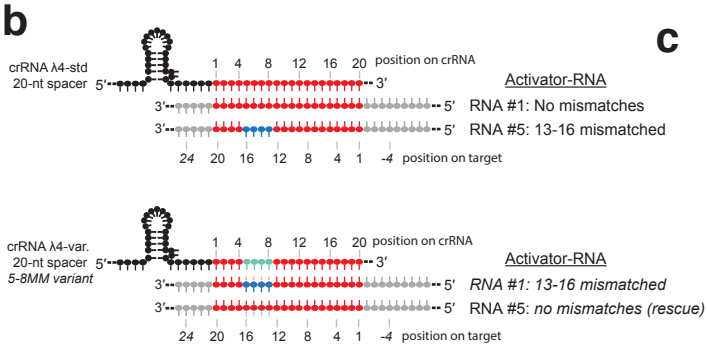


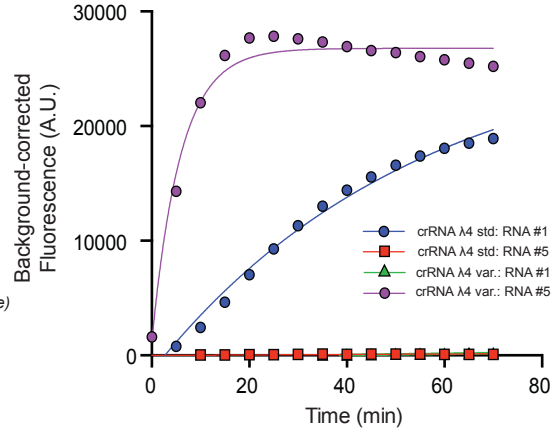

Supplementary Data Figure 3




Supplementary Data Figure 4

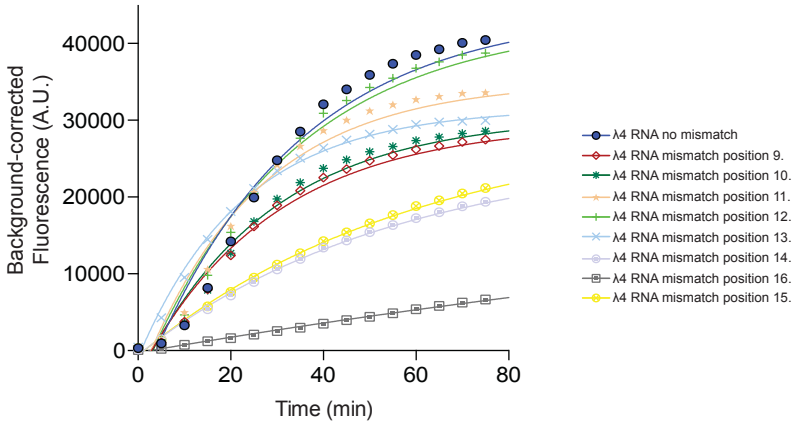



Supplementary Data Figure 5

**a**




Supplementary Data Figure 6