The kinetic landscape of an RNA binding protein in cells

Deepak Sharma [1,2], Leah L. Zagore [1,2], Matthew M. Brister [3], Xuan Ye [1,2],
Carlos E. Crespo-Hernández [3], Donny D. Licatalosi [1,2,6] & Eckhard Jankowsky [1,2,4,5,6]

(1) Center for RNA Science and Therapeutics, School of Medicine
(2) Department of Biochemistry, School of Medicine
(3) Department of Chemistry
(4) Department of Physics
(5) Case Comprehensive Cancer Center, School of Medicine
Case Western Reserve University
Cleveland, OH 44106

(6) Corresponding authors

ABSTRACT

Gene expression in higher eukaryotic cells orchestrates interactions between thousands of RNA binding proteins (RBPs) and tens of thousands of RNAs [1]. The kinetics by which RBPs bind to and dissociate from their RNA sites are critical for the coordination of cellular RNA-protein interactions [2]. However, these kinetic parameters were experimentally inaccessible in cells. Here we show that time-resolved RNA-protein crosslinking with a pulsed femtosecond UV laser, followed by immunoprecipitation and high throughput sequencing allows the determination of binding and dissociation kinetics of the RBP Dazl for thousands of individual RNA binding sites in cells. This kinetic crosslinking and immunoprecipitation (KIN-CLIP) approach reveals that Dazl resides at individual binding sites only seconds or shorter, while the sites remain Dazl-free markedly longer. The data further indicate that Dazl binds to many RNAs in clusters of multiple proximal sites. The impact of Dazl on mRNA levels and ribosome association correlates with the cumulative probability of Dazl binding in these clusters. Integrating kinetic data with mRNA features quantitatively connects Dazl-RNA binding to Dazl function. Our results show how previously inaccessible, kinetic parameters for RNA-protein interactions in cells can be measured and how these data quantitatively link RBP-RNA binding to cellular RBP function.

The binding and dissociation of RBPs at their cognate RNA sites in cells are critical for the regulation of gene expression [2]. Yet, association and dissociation kinetics of RBPs at individual binding sites in cells have not been experimentally accessible. RBP binding and dissociation kinetics have only been measured *in vitro*, while in cells, only steady-state patterns of RNA-protein interactions have been determined [2-6]. For a small number of RBPs, equilibrium binding parameters measured *in vitro* correlate with steady-state binding patterns in cells [7,8]. Although these observations advanced understanding of RBP function, the inaccessibility of binding and dissociation kinetics of RBPs in cells limits or even precludes the establishment of quantitative connections between RBP-RNA interactions and cellular RBP function. Here, we measure binding and dissociation kinetics of the RBP Dazl at thousands of individual binding sites in cells. We then show how these kinetic parameters inform a quantitative understanding of the cellular function of Dazl.

Time-resolved fs laser crosslinking *in vitro*

To measure binding and dissociation kinetics of proteins at individual RNA sites in cells, we devised a time-resolved RNA-protein crosslinking approach (**Fig.1a**). Because kinetic parameters in cells must be determined from the steady-state between free and RNA-bound protein, a sufficient number of experimental constraints are required to calculate rate constants. These constraints can be established by measuring crosslinking timecourses at different protein concentrations and different crosslinking efficiencies (**Fig.1b**), while ensuring that crosslinking rate constants are roughly equal or larger than dissociation and apparent association rate constants. To achieve sufficiently fast protein-RNA crosslinking, we employed a pulsed femtosecond (fs) UV laser (**Fig.1c, Extended Data Fig.1a**). Pulsed UV lasers had been shown to efficiently photo-crosslink proteins to DNA through multi-photonic excitation of the crosslinking species [9-12].

To examine the utility of a pulsed fs UV laser for determining binding and dissociation rate constants of RNA-protein interactions, we performed time-resolved crosslinking reactions with purified proteins and RNAs (**Fig.1d,e**). UV-mediated RNA degradation was reduced upon irradiation with the fs laser, compared with a steady-state UV light source (**Extended Data Fig.1b**). Although the photon density during the laser pulse is orders of magnitude greater, compared with the steady-state UV light source, fewer photons are absorbed by the RNA over a given amount of time (**Extended Data Fig.1c**). This is because fs pulses are emitted once per millisecond and the cross-section for multi-photonic absorption is smaller than for single-photonic absorption with a steady-state UV light source [13].

Crosslinking of the purified RNA-binding protein RbFox(RRM) to its cognate RNA with the fs laser was markedly more efficient, compared with the steady-state UV source (**Extended Data Fig.1d-f**). Observed crosslinking rates increased with laser power and protein concentration, as expected (**Fig.1d**). We determined binding, dissociation and crosslinking rate constants for RbFox(RRM)-RNA binding from the crosslinking timecourses at two different laser powers and two different protein concentrations (**Fig.1b,d,e, Supplementary Material Fig.S2**). The apparent affinity ($K_{1/2}$) of RbFox(RRM) for its cognate RNA, calculated from association and dissociation rate constants, was similar to the affinity measured by fluorescence anisotropy (**Fig.1e, Extended Data Fig.1i**) and consistent with previously reported values [14]. We next determined binding, dissociation and crosslinking rate constants for a mutated $\text{RbFox}^{\text{mut}}$(RRM) [15] and for the RNA binding protein Dazl(RRM) [16], using fs laser crosslinking (**Fig.1e, Extended Data Fig.1g,h**). RNA affinities of these two proteins, calculated from the rate constants, were also similar to affinities measured with fluorescence anisotropy (**Fig.1e, Extended Data Fig.1j,k**). The data with three RBPs collectively indicate that binding and dissociation rate constants for RNA-protein interactions can be determined by time-resolved, fs laser crosslinking.

Laser crosslinking in cells

We adapted the time-resolved fs laser crosslinking approach to measure binding and dissociation rate constants of the RNA-binding protein Dazl to individual RNA sites in mouse GC-1 cells [17,18]. Dazl is essential for male and female gametogenesis [19-22]. The protein contains one RNA recognition motif (RRM), binds predominantly to 3'UTRs of mRNAs and regulates mRNA stability, translation, or both [23]. Dazl was expressed under the control of a doxycycline-inducible promotor [17]. Varying the doxycycline concentration allowed measurements at different Dazl concentrations in GC-1 cells (**Extended Data Fig.2a**). To perform time-resolved fs laser crosslinking experiments, cells were transferred to a quartz cuvette and under constant stirring placed in the laser beam. Crosslinking measurements were performed with GC-1 cells expressing two different Dazl concentrations and two different laser powers for 30, 180 and 680 s (**Extended Data Fig.2b**). We also measured the bulk degree of crosslinking at each time point (**Extended Data Fig.2c**) and determined transcript levels at each Dazl concentration by RNA-

Seq. Approximately 10% of cells showed signs of physical damage after crosslinking, which is comparable to cell damage by conventional steady-state UV-crosslinking (**Supplementary Material Table S4**).

We prepared and sequenced cDNA libraries for each timepoint sample and for controls without crosslinking (**Extended Data Fig.2b**, **Supplementary Material Table S5**, refs.[24,25]). Dazl crosslinking sites with the fs laser were virtually identical to sites identified by conventional steady-state UV-crosslinking with respect to RNA types, location in 3'UTRs and crosslinking site characteristics (**Fig.2a**, **Extended Data Fig.2d-f**, ref.[17]). These data show that fs laser crosslinking maintains the characteristics of crosslink sites seen with steady-state UV-crosslinking. Our kinetic crosslinking and immunoprecipitation approach (KIN-CLIP) thus faithfully maps Dazl binding sites.

To calculate association and dissociation rate constants for Dazl binding at individual binding sites, we normalized the sequencing reads for each CLIP library to the bulk amount of crosslinking, thereby converting sequencing reads into a concentration-equivalent of crosslinked RNA at a given binding site (**Fig.2b, Supplementary Material Table S6**). This normalized read coverage was used to calculate a dissociation rate constant ($k_{\mathrm{diss.}}$), observed association rate constants at low and high Dazl concentration ($k_{\mathrm{on}}^{(1\mathrm{xDazl})}$, $k_{\mathrm{on}}^{(4.2\mathrm{xDazl})}$) and crosslinking rate constants for both laser powers ($k_{\mathrm{XL}}^{(1\ \mathrm{mW})}$, $k_{\mathrm{XL}}^{(2.6\ \mathrm{mW})}$) for each binding site. (**Fig.2c**, **Extended Data Fig.3a-k**). Obtained rate constants faithfully described the experimental data (**Fig.2c, Extended Data Fig.3l,m, Supplementary Material Fig.S4**).

Dazl-RNA binding kinetics in cells

For most binding sites (89%), the observed association rate constants at the 1xDazl concentration were lower than those at the 4.2xDazl concentration (**Fig.2d**). These data indicate that only a small fraction of binding sites is saturated with Dazl at low protein concentration and implies a population of free Dazl in the cell, at least at the high Dazl concentration. Although 85% of Dazl crosslinking sites showed the consensus 5'-GUU motif (**Extended Data Fig.4a-d**), association and dissociation rate constants varied by several orders of magnitude (**Fig.2e**). Association rate constants varied to a larger degree than dissociation rate constants (**Fig.2e**). These observations suggest that Dazl binding and dissociation kinetics in cells depend not exclusively on the consensus motif. A_n, U_n and $(GU)_n$ stretches were overrepresented in the vicinity of binding sites with high association rate constants (**Extended Data Fig.4e-p**). No further sequence signatures in the vicinity of crosslinking sites correlated with other rate constants (**Extended Data Fig.4i-p**).

The dissociation rate constant for Dazl(RRM) *in vitro* (**Fig.1e**) is on the low end of the spectrum of cellular dissociation rate constants (**Fig.2e**), indicating that Dazl dissociates from most cellular binding sites more frequently than from its cognate RNA *in vitro*. Dazl resides at most cellular binding sites for less than $\tau_B < 1$s (**Fig.2e**). Binding events are infrequent and even at high Dazl concentrations occur rarely more than six times per minute (**Fig.2e**). Accordingly, the probability of Dazl to be bound at any time is less than 10% for many binding sites (**Fig.2e**). This observation indicates that Dazl operates at a sub-saturating regime with respect to its mRNA targets in GC-1 cells. This notion is consistent with kinetic parameters of Dazl measured *in vitro* (**Fig.1e**), and a cellular Dazl concentration roughly at or below its affinity *in vitro* [26]. We also determined a maximal fractional occupancy (Φ^{max}, **Fig.2e, Supplementary Material Fig.S3**), which describes the extent by which a given RNA site would be occupied at saturating Dazl concentrations. The data suggest that most binding sites are not fully accessible for Dazl binding during the course of the experiment.

Dissociation rate constants for binding sites did not vary significantly for different RNA classes (**Extended Data Fig.4s**) or between mRNA 3'UTRs, 5'UTRs, introns and open reading frames (**Fig.2f**). Association rate constants and binding probabilities, which depend on both, association and dissociation rate constants, were higher for binding sites in 3'UTRs than for sites in 5' UTRs, introns and ORFs (**Fig.2f**), and higher in mRNAs, compared with other RNA classes (**Extended Data Fig.4q,r**). The maximal fractional occupancy of binding sites did not significantly vary in the different mRNA regions (**Fig.2f**), but was higher in mRNA, compared with other RNA classes (**Extended Data Fig.4t**). Because Dazl function has been linked to binding in 3'UTRs [17], our data raised the possibility that association rate constants, binding probabilities, or both, influence cellular roles of Dazl more than its residence time at the binding sites. Collectively, the kinetic data revealed highly dynamic Dazl-RNA interactions with most Dazl binding events being rare and transient.

Dazl binds mRNAs in clusters

To understand how Dazl regulates mRNA function in this highly dynamic fashion, we examined the patterns of the kinetic parameters for all Dazl binding sites on each bound mRNA. The majority of Dazl binding sites are in 3'UTRs (**Fig.2a**), and frequently proximal to the polyadenylation site (PAS, **Extended Data Figs.2e, 5a**). Most Dazl-bound mRNAs contained multiple Dazl binding sites with an inter-site distance markedly smaller than expected by chance (**Fig.3a**), even when distant to the PAS (**Extended Data Figs.5b,c**). This observation suggested clustering of multiple Dazl binding sites on most 3'UTRs (**Extended Data Fig.5d-g**). The number of binding sites within a 3’UTR cluster increased with proximity to the PAS (**Fig.3b**). Dissociation rate constants and maximal fractional occupancies did not scale with the number of binding sites in a cluster (**Extended Data Fig.5i,j**). However, association rate constants for individual binding sites scaled with the number of binding sites in a cluster, regardless of the distance of the cluster to the PAS. (**Fig.3c**). Binding probabilities showed a similar pattern (**Extended Data Fig.5h**). These observations suggest cooperative association steps.

Kinetic parameters within clusters showed consistent patterns of moderate correlation (**Extended Data Fig.5k**). However, fractional occupancies for binding sites within a given cluster were closely correlated (**Fig.3d**, **Extended Data Fig.5k**), suggesting that binding site context, possibly including RNA structure or proximal binding of other proteins, play a prominent role in determining similar accessibility of binding sites within a cluster. This notion, together with the scaling of association rate constants with the number of binding sites (**Fig.3c**), raised the possibility that binding site clusters are important for Dazl function.

Clusters correlate with Dazl function

To test this hypothesis, we quantified Dazl binding in a given cluster by calculating a cumulative binding probability (ΣB) from the kinetic constants of the binding sites in the cluster. ΣB describes the probability that Dazl occupies at least one site in a given cluster at any given time (**Fig.4a**). ΣB increased with the number of binding sites in a cluster and with proximity to the PAS (**Extended Data Fig.6a,b**). We compared ΣB values in a given cluster to changes in ribosome association and transcript levels at low and high Dazl concentrations (**Fig.4b**). Dazl binding had been shown to increase transcript levels and ribosome association for many, but not all mRNAs [17]. At the high Dazl concentration, compared with the low Dazl concentration, we detected an overrepresentation of clusters with high ΣB in mRNAs that increased in transcript level, ribosome association, or both (**Fig.4c, Extended Data Fig.6c,d**). Clusters with low ΣB values were overrepresented in mRNAs that decreased in transcript levels and ribosome association at the high Dazl concentration (**Fig.4c**). We detected no comparable correlation between the Dazl impact on transcript levels or ribosome association and binding probabilities of individual binding sites, clusters with scrambled binding sites or with simultaneous occupancy of multiple binding sites in a given cluster (**Extended Data Fig.6e-k**). ΣB values thus instructively link binding kinetics to Dazl impact on mRNA function, further supporting the notion that Dazl clusters are critical for the function of this RBP.

A Dazl regulatory program.

To delineate the connection between Dazl binding kinetics and Dazl impact on mRNA function in more detail, we identified additional mRNA and Dazl cluster characteristics that correlated with Dazl function. Besides ΣB, we detected correlations for the number of binding sites in a cluster, the difference in cumulative binding probabilities at low and high Dazl concentrations (ΔΣB), number of clusters in a 3'UTR, length of the 3'UTR, and proximity of a cluster to the PAS (**Extended Data Fig.7**). Some of these characteristics correlate with each other ($R^2 \leq 0.6$), but each parameter contributes separately to the Dazl impact on mRNA function (**Extended Data Fig.8a-e**). Proximity of Dazl binding to the PAS had been previously noted to influence Dazl impact on mRNA function [17].

Principal component analysis and t-distributed stochastic neighbor embedding independently identified 21 mRNA groups with a distinct combination of kinetic, cluster and mRNA characteristics (**Extended Data Fig.8b-e**). Each of these 21 groups falls into a class of Dazl impact on transcript level and ribosome association (**Fig.4d**, **Extended Data Fig.8c-f, Extended Data Fig.9**). Translation efficiencies also vary for groups in mRNA classes where mRNA level and ribosome association do not scale proportionally (**Extended Data Fig.10a**). The mRNAs in each group belong to defined GO-terms (**Fig.4d**), and in many cases encode proximal proteins in a given pathway (**Extended Data Fig.8h**). mRNA groups with high values of ΣB or ΔΣB predominantly function in mRNA processing and transport, in DNA replication and in cell cycle regulation. mRNA groups with low ΣB or ΔΣB values are primarily associated with mRNA decay, membrane transport and stress response (**Fig.4d**). Collectively, the results indicate a link between the biological role of a given mRNA and Dazl binding kinetics, binding site clusters, their location on the 3'UTR and mRNA features (**Extended Data Figs.8h,9**). These characteristics represent a basic Dazl regulatory program that connects Dazl binding in 3'UTRs to its impact on mRNA function (**Fig.4d**).

To quantify this regulatory program, we employed a multiple linear regression model (**Fig.4e-h**; **Extended Data Fig.10b-e, Supplementary Material Figs.S5-S7.**). The model explains changes in ribosome association, mRNA levels (**Fig.4g,h**), translation efficiencies and changes in translation from luciferase reporters between low and high Dazl concentration (**Extended Data Fig.10f-h**). The largest contribution is seen for the cumulative binding probabilities, which derive from the kinetic parameters of Dazl binding, and for the numbers of Dazl clusters in the 3'UTR (**Fig.4e**,**f, Extended Data Fig.10f**). For mRNAs that increase in ribosome association, the distance of the Dazl clusters to the PAS also has an effect (**Fig.4e**), consistent with previously reported data [17]. Collectively, our data show that Dazl impacts bound mRNAs in a complex, yet tractable manner that depends prominently on kinetic parameters.

Discussion

We devised and applied a time-resolved crosslinking approach to measure cellular binding and dissociation kinetics of RNA-protein interactions at individual binding sites on a transcriptome-wide scale. Key to this KIN-CLIP approach is a pulsed fs UV laser, which increases crosslinking efficiencies without altering RNA-protein crosslinking patterns, compared with steady-state UV irradiation. KIN-CLIP should enable the biochemical characterization of other RNA-protein interactions in cells. Our approach also provides a framework for obtaining quantitative, steady-state protein-RNA binding information from CLIP with conventional crosslinking sources. Moreover, combining time-resolved fs laser crosslinking and kinetic analysis might allow quantitative, biochemical analysis of DNA-protein [12] and even of protein-protein interactions [27] in cells.

For Dazl, KIN-CLIP reveals highly dynamic RNA binding. Dazl resides at individual binding sites only seconds or shorter, while cognate sites remain free of Dazl for most of the time. These findings are consistent with kinetic data for Dazl-RNA binding *in vitro* and the notion that cellular Dazl concentrations are sub-saturating relative to its RNA targets [26]. Highly dynamic binding allows for rapid changes in RNA binding patterns, which might be critical for Dazl function. Since *in vitro* RNA binding kinetics of Dazl are similar to those of other RBPs [6], many of which might also operate in cells at sub-saturating concentrations relative to their RNA targets [7], our findings raise the possibility that other RBPs bind their cognate RNA sites also transiently and infrequently. If true for many RBPs, few regulatory RBPs and occasionally none might be bound to a given mRNA at a given time.

Access to cellular kinetic data allows the decoding of a complex link between Dazl-RNA-binding patterns and Dazl function. Dazl affects mRNA level and ribosome association according to a regulatory program that integrates the collective binding kinetics of Dazl at multiple cognate sites in a cluster, the number of binding sites in a cluster, location of clusters on the 3’UTR, proximity to the PAS, and 3’UTR length. Because our experimental and data analysis approaches are applicable to other RBPs, KIN-CLIP provides a blueprint for delineating regulatory programs for other RBPs.

ACKNOWLEDGEMENTS

We thank Dr. Gabriele Varani (UW, Seattle) for the gift of purified RbFox(RRM) and RbFox^{mut}(RRM), Dr. Anton Komar (Cleveland State University) for the design of the codon-optimized Dazl construct, and Dr. Wei Huang (CWRU) for assistance with the fluorescence polarization experiments. This work was supported by the NIH (GM118088 to E.J., GM107331 to D.D.L) and the NSF (CHE-1800052 to C.E.C.-H.)

DATA AVAILABILITY

Sequencing data are available at the NCBI Gene Expression Omnibus (Accession number: GSE150214).

CODE AVAILABILITY

Customized R and Python scripts are available at: https://github.com/deebratforlife/KIN-CLIP.

FIGURE CAPTIONS

Figure 1 | Time-resolved, fs laser RNA-protein crosslinking *in vitro*. **a**. Kinetic scheme for RNA-protein binding and crosslinking. **b**. Reaction scheme **c**. Schematic of pulsed fs UV laser crosslinking. **d**. RNA Crosslinking timecourses for RbFox(RRM) with fs laser at different laser power and protein concentrations. Lines show the fit to the data in panel e. **e**. Rate constants for association (k_{on}), dissociation (k_{off}) and crosslinking at both laser powers ($k_{XL}^{(1mW)}$, $k_{XL}^{(2.6mW)}$) determined with the fs laser for RbFox(RRM), a mutated RbFox^{mut}(RRM), and Dazl(RRM). Equilibrium dissociation constants ($K_{1/2}$) for fs laser are calculated from these rate constants and measured by fluorescence anisotropy (**Extended Data Fig.1h-j**). Errors mark one standard deviation.

Figure 2 | Kinetics of Dazl-RNA binding and dissociation in cells. **a**. Distribution of CLIP sequencing reads across RNA classes and mRNA regions for fs laser (4.2xDazl, 2.6 mW) and conventional crosslinking (Stratalinker; 4.2xDazl). **b**. Normalized sequencing reads for the 3'UTR of a representative transcript (Thbs1) at increasing crosslinking times (left side), different protein concentrations and different laser power (right side, scale: normalized coverage = 11 for all traces). Reads for conventional iCLIP are indicated below. **c**. Crosslinking timecourses for two binding sites (1,2, panel b). Datapoints show the normalized read coverage (Lines: best fit to the parameters in the table. Error bars: 95% confidence interval for normalized peak coverage value, determined by minimizing X^2. For crosslinking rate constants of all binding sites see **Suppl. Material Table S9**). Each binding site was fitted independently using two mutually exclusive methods. **d**. Association rate constants for 1xDazl and 4.2xDazl for all binding sites (N = 10,341). Arrows mark the confidence range for the rate constants. The diagonal line marks equal rate constants at both Dazl concentrations. **e**. Transcriptome-wide distributions of dissociation rate constants ($k_{diss.}$), association rate constants at high Dazl concentration ($k_{on}^{4.2xDazl}$), binding probability ($P^{4.2xDazl}$), and maximal fractional occupancy (Φ^{max}) for all Dazl binding sites. Select dwell times of Dazl bound (τ_b) and away from binding sites (τ_f) are marked (bin sizes for frequency distributions: $k_{diss.}$: 0.35s^{-1}, $k_{on}^{4.2xDazl}$: 0.015s^{-1}, $P^{4.2xDazl}$: 0.019, Φ^{max}: 0.02). **f**. Distributions of kinetic parameters for all binding sites in the indicated mRNA regions (p-values: one way ANOVA, n.s.: not significant; for boxplots: vertical line: median, box limits: interquartile range (IQR); whiskers 1.5x IQR)

Figure 3 | Clustering of Dazl binding sites in 3'UTRs. a. Distribution of Dazl binding sites in 3'UTRs as function of the distance between neighboring binding sites. The grey line shows the distribution if sites were randomly distributed across all 3'UTRs (*p* value: t-test). **b**. Proximity of clusters with varying number of binding sites to the PAS. **c**. Correlation between association rate constants and number of binding sites in clusters. (p-values: one way ANOVA; for boxplots: vertical line: median, box limits: interquartile range (IQR); whiskers 1.5x IQR). **d**. Heatmap depicting correlation of values for maximal fractional occupancy in clusters with 6 binding sites.

Figure 4 | Link between Dazl-RNA binding and Dazl impact on mRNA function. a. Distribution of cumulative binding probabilities (ΣB) for Dazl in all clusters (N = 1,690). **b**. Changes in transcript levels (ΔRNA) and ribosome association (ΔRPF) between low and high Dazl concentration for Dazl-bound mRNAs (N = 968). Data points represent averages from triplicate ribosome profiling and RNAseq experiments [17]. **c**. Correlation between cumulative binding probabilities and functional mRNA classes. Colors correspond to the enrichment (hypergeometric test, red: $p < 0.05$, shades of yellow: not enriched). **d**. Upper panel: Heatmap of the Dazl regulatory program, linking functional mRNA classes to kinetic parameters (ΣB, ΔΣB), cluster characteristics (number of binding sites in cluster, cluster distance from PAS) and 3'UTR features (numbers of clusters, on 3'UTR, 3'UTR length, transcript level), all shown in terciles (**Extended Data Fig.8f**). Numbers mark the groups with characteristic combinations of ΣB, ΔΣB, cluster and mRNA features. Lower panel: Link between Dazl-code and Gene ontology (GO) terms. **e,f**. Linear regression model linking the Dazl regulatory program to impact of Dazl binding on changes in transcript levels (ΔRNA) and ribosome association (ΔRPF) (panel b). Points represent the differential intercept (DI) linear coefficient (LC) (red: DILCs for transcript levels and ribosome association that increase at high Dazl concentration, green: black: DILCs for transcript levels and ribosome association that decrease at high Dazl concentration). **g,h**. Correlation between experimental values for ΔRNA and ΔRPF and values predicted with the linear regression model (R: adjusted linear correlation coefficient) for the test data set unseen by the model (N = 492).

REFERENCES

1 Gerstberger, S., Hafner, M. & Tuschl, T. A census of human RNA-binding proteins. *Nat Rev Genet* **15**, 829-845 (2014).

2 Licatalosi, D. D., Ye, X. & Jankowsky, E. Approaches for measuring the dynamics of RNA-protein interactions. *Wiley Interdiscip Rev RNA* **11**, e1565 (2020).

3 Corley, M., Burns, M. C. & Yeo, G. W. How RNA-Binding Proteins Interact with RNA: Molecules and Mechanisms. *Mol Cell* **78**, 9-29 (2020).

4 Ule, J., Hwang, H. W. & Darnell, R. B. The Future of Cross-Linking and Immunoprecipitation (CLIP). *Cold Spring Harb Perspect Biol* **10** (2018).

5 Van Nostrand, E. L. *et al.* Principles of RNA processing from analysis of enhanced CLIP maps for 150 RNA binding proteins. *Genome Biol* **21**, 90 (2020).

6 Gleitsman, K. R., Sengupta, R. N. & Herschlag, D. Slow molecular recognition by RNA. *RNA* **23**, 1745-1753 (2017).

7 Jarmoskaite, I. *et al.* A Quantitative and Predictive Model for RNA Binding by Human Pumilio Proteins. *Mol Cell* **74**, 966-981 (2019).

8 Sutandy, F. X. R. *et al.* In vitro iCLIP-based modeling uncovers how the splicing factor U2AF2 relies on regulation by cofactors. *Genome Res* **28**, 699-713 (2018).

9 Hockensmith, J. W., Kubasek, W. L., Vorachek, W. R. & von Hippel, P. H. Laser cross-linking of nucleic acids to proteins. Methodology and first applications to the phage T4 DNA replication system. *J Biol Chem* **261**, 3512-3518 (1986).

10 Pashev, I. G., Dimitrov, S. I. & Angelov, D. Crosslinking proteins to nucleic acids by ultraviolet laser irradiation. *Trends Biochem Sci* **16**, 323-326 (1991).

11 Russmann, C. *et al.* Crosslinking of progesterone receptor to DNA using tuneable nanosecond, picosecond and femtosecond UV laser pulses. *Nucleic Acids Res* **25**, 2478-2484 (1997).

12 Steube, A., Schenk, T., Tretyakov, A. & Saluz, H. P. High-intensity UV laser ChIP-seq for the study of protein-DNA interactions in living cells. *Nat Commun* **8**, 1303 (2017).

13 Budowsky, E. I., Axentyeva, M. S., Abdurashidova, G. G., Simukova, N. A. & Rubin, L. B. Induction of polynucleotide-protein cross-linkages by ultraviolet irradiation. Peculiarities of the high-intensity laser pulse irradiation. *Eur J Biochem* **159**, 95-101 (1986).

14 Auweter, S. D. *et al.* Molecular basis of RNA recognition by the human alternative splicing factor Fox-1. *EMBO J* **25**, 163-173 (2006).

15 Chen, Y. *et al.* Targeted inhibition of oncogenic miR-21 maturation with designed RNA-binding proteins. *Nat Chem Biol* **12**, 717-723 (2016).

16 Jenkins, H. T., Malkova, B. & Edwards, T. A. Kinked beta-strands mediate high-affinity recognition of mRNA targets by the germ-cell regulator DAZL. *Proc Natl Acad Sci U S A* **108**, 18266-18271 (2011).

17 Zagore, L. L. *et al.* DAZL Regulates Germ Cell Survival through a Network of PolyA-Proximal mRNA Interactions. *Cell Rep* **25**, 1225-1240 e1226 (2018).

18 Hofmann, M. C., Narisawa, S., Hess, R. A. & Millan, J. L. Immortalization of germ cells and somatic testicular cells using the SV40 large T antigen. *Exp Cell Res* **201**, 417-435 (1992).

19 Fu, X. F. *et al.* DAZ Family Proteins, Key Players for Germ Cell Development. *Int J Biol Sci* **11**, 1226-1235 (2015).

20 Lin, Y. & Page, D. C. Dazl deficiency leads to embryonic arrest of germ cell development in XY C57BL/6 mice. *Dev Biol* **288**, 309-316 (2005).

21 Ruggiu, M. *et al.* The mouse Dazla gene encodes a cytoplasmic protein essential for gametogenesis. *Nature* **389**, 73-77 (1997).

22 Saunders, P. T. *et al.* Absence of mDazl produces a final block on germ cell development at meiosis. *Reproduction* **126**, 589-597 (2003).

23 Yang, C. R. *et al.* The RNA-binding protein DAZL functions as repressor and activator of mRNA translation during oocyte maturation. *Nat Commun* **11**, 1399 (2020).

24 Haberman, N. *et al.* Insights into the design and interpretation of iCLIP experiments. *Genome Biol* **18**, 7 (2017).

25 Huppertz, I. *et al.* iCLIP: protein-RNA interactions at nucleotide resolution. *Methods* **65**, 274-287 (2014).

26 Reynolds, N. *et al.* Dazl binds in vivo to specific transcripts and can regulate the pre-meiotic translation of Mvh in germ cells. *Hum Mol Genet* **14**, 3899-3909 (2005).

27 Itri, F. *et al.* Femtosecond UV-laser pulses to unveil protein-protein interactions in living cells. *Cell Mol Life Sci* **73**, 637-648 (2016).

EXTENDED DATA, FIGURE CAPTIONS

Extended Data Figure 1 | Time-resolved RNA-protein crosslinking with fs laser *in vitro*. a. Schematics of fs laser setup. **b**. Degradation of RNA (38 nt) under steady-state and fs laser illumination. Data points represent averages of 3 independent measurements. Error bars mark one standard deviation. Lines show a linear trend. **c**. Dose absorbed over time for crosslinking with conventional UV (Stratalinker, 200 mJ/cm^2, λ = 254 nm) and fs laser (2.6 mW) **d**. Representative denaturing polyacrylamide gel electropherogram (PAGE) for a crosslinking reaction of 50 nM RbFox(RRM) (laser: 2.6 mW) (lanes 5 – 12) and control reactions with RNA only (lanes 1 – 3) and RbFox(RRM) only (lane 4), with (lanes 2-4) or without (lanes 1 and 5) crosslinking. **e**. Representative denaturing PAGE for a crosslinking reaction of 50 nM RbFox(RRM) with Stratalinker (200 mJ/cm^2, λ = 254 nm), lanes 4 - 8) and control reactions (lanes 1 - 3). **f**. Timecourse of crosslinking reaction of 50 nM RbFox(RRM) with Stratalinker (200 mJ/cm^2, λ = 254 nm) vs. fs laser (**Fig**.**1d**). Datapoints are averages from triplicate experiments (error bars: one standard deviation). **g**. RNA Crosslinking timecourses for Dazl(RRM) with fs laser at different laser power and protein concentrations. Data points represent averages of 3 independent measurements (error bars: one standard deviation). Lines show the fit to the data in **Fig.1e**. **h**. RNA Crosslinking timecourses for RbFox^{mut}(RRM) with fs laser at different laser power and protein concentrations. Data points represent averages of 3 independent measurements (error bars: one standard deviation). Lines show the fit to the data in **Fig.1e**. **i-k**. Binding isotherms for RbFox(RRM), RbFox^{mut}(RRM) and Dazl(RRM) to cognate RNAs measured by fluorescence anisotropy. Experiments were performed multiple times, all datapoints are shown. Apparent equilibrium binding constants ($K_{1/2}$, **Fig.1e**) were calculated with the quadratic binding equation.

Extended Data Figure 2 | Dazl-RNA crosslinking with fs laser in GC-1*spg* cells. a. Western Blot of Doxycyline dependent Dazl expression in GC-1 cells. **b**. Schematic of the time-resolved crosslinking approach in cells. Numbers mark the respective CLIP libraries. **c**. Representative PAGE for bulk Dazl-RNA crosslinking. The intensity of crosslinked RNA (marked) is used to convert NGS reads to a concentration-equivalent parameter (for bulk crosslinking intensities see **Supplementary Material, Table S6**) **d**. Dazl binding sites identified by fs laser (KIN-CLIP) and conventional UV crosslinking (iCLIP) on all RNAs and 3'UTRs. **e**. Metagene distribution of Dazl binding sites identified by KIN-CLIP and iCLIP on 3'UTRs proximal to stop codon and PAS. The dotted lines mark the background of a random distribution of binding sites on 3'UTRs. **f**. CITS (<u>C</u>rosslink <u>I</u>nduced <u>T</u>runcation <u>S</u>ite) analysis [28,29] of 6-mer and 4-mer enrichment at 5'-termini of sequencing reads for KIN-CLIP (upper panels) and iCLIP (lower panels). The data indicate a virtually identical sequence context of crosslinking sites for KIN-CLIP and iCLIP. Sequence enrichment reflects the statistical overrepresentation of 6-mer and 4-mer sequences with respect to randomized sequences (Z-score, 11 nucleotide region, ± 5 nt from the 5'-terminal nucleotide).

Extended Data Figure 3 | Determination of kinetic parameters from fs laser, time-resolved Dazl-RNA crosslinking in cells. a. Flowchart of the approach to calculate kinetic parameters for individual Dazl-RNA binding sites in cells (for details see Materials and Methods). Unless otherwise stated, rate constants averaged from both approaches are used in subsequent data analyses. **b**. Scaling of X^2 with the number of iterative fitting cycles for analytical and numerical approaches. **c,d**. Distribution of X^2 at first and last (642) fitting cycle for analytical (**c**) and numerical (**d**) approaches (COD: Coefficient Of Determination, R^2: linear correlation coefficient). **e-i**. Correlation of parameters calculated with analytical and numerical fitting procedures (R^2: linear correlation coefficient). **j**. Correlation between crosslinking rate constants for low and high laser power. Rate constants are averaged from parameters obtained with numerical and analytical approach. Crosslinking rate constants at higher laser power were larger than at lower for 92% of binding sites. **k**. Confidence range for dissociation rate constants (for details see Materials and Methods). **l**. Normalized read densities measured experimentally and calculated from the kinetic parameters for all Dazl binding sites. **m.** Distribution of X^2 for experimental values compared with values calculated with the kinetic parameters.

Extended Data Figure 4 | Kinetic parameters of Dazl binding sites and sequence context. a-d. Sequences surrounding Dazl binding sites, arranged according to decreasing values for $k_{on}^{(4.2xDazl)}$, $k_{diss.}$, $k_{XL}^{(2.6mW)}$, and Φ^{max}. Sequences are aligned at the peak nucleotide (most frequent crosslink site (± 11 nt peak nucleotide), **Extended Data Fig.2f**, position 0). **e-h**. Frequency of 6-mer sequences surrounding Dazl crosslink sites (± 111 nt peak nucleotide) in top and bottom 5% of sequences arranged according to the kinetic parameters in panels (**a-d**). **i-l**. Relative frequency of 6-mer sequences in top and bottom 5% of sequences (panels **e-h**), arranged according to the kinetic parameters in panels **a-d**. Sequences below the diagonal line correspond to enrichment of a 6-mer in the top 5% versus the bottom 5%. (R^2: linear correlation coefficient). A_6, U_6 and U_3GU_2 are most enriched in the vicinity of the binding sites with the fastest apparent association rate constants, compared to the binding sites with the slowest apparent association rate constants. No comparable enrichment is seen for other kinetic parameters. **m-p**. Relative frequency of 4-mers in top and bottom 5% of sequences arranged according to the kinetic parameters in panels (**a-d**). **q-t**. Distribution of association and dissociation rate constants, binding probabilities ($^{P(4.2xDazl)}$) and maximal fractional occupancy (Φ^{max}) for binding sites on different RNA classes. P values (one-way ANOVA, significant for $p < 0.05$) indicate inter-group differences. Φ^{max}, but not other parameters vary significantly for different RNA classes (boxplots: vertical line: median, box limits: interquartile range (IQR); whiskers 1.5x IQR).

Extended Data Figure 5 | Arrangement of 3'UTR Dazl binding sites in clusters.

a. Arrangement of Dazl binding sites in 3'UTRs. Binding sites are colored according to $k_{on}^{(4.2xDazl)}$ and $k_{diss.}$ as indicated in the key panel. Right panel: number of clusters in corresponding 3'UTR. Colors mark number of binding sites in a cluster, as indicated in legend bar (right) (N = 1,313 3'UTRs, 1,690 clusters, 6,085 binding sites) **b**. Distribution of Dazl binding sites in 3'UTRs closer than 500 nt to PAS, as function of the distance between neighboring binding sites. The grey line shows the distribution if sites were randomly distributed across all 3'UTRs (*p* value: t-test). **c**. Distribution of Dazl binding sites in 3'UTRs farther than 500 nt from PAS, as function of the distance between neighboring binding sites. The grey line shows the distribution if sites were randomly distributed across all 3'UTRs (*p* value: t-test). **d**. Large windows: genome browser traces of representative 3'UTRs with 5 clusters (Nucks1) and 2 clusters (D'Rik, D030056L22Rik). Bars show the normalized read coverage for 4.2xDazl, 2.6 mW laser and 680s crosslinking time. Numbers mark the distance between clusters. Small windows: zoom into cluster 1 of Nucks1 with 3 binding sites and in cluster 1 of D'Rik with 2 binding sites (numbers mark the distance between binding sites). **e.** Number of clusters in 3'UTRs with Dazl binding sites. Colors show the number of binding sites in a cluster as indicated in panel **a**. (red: 20; cornsilk: 1). **f**. Distances between clusters in 3'UTRs with 2 to 4 clusters. Number 1 represents the cluster most proximal to the PAS. **g**. Distribution of distances between neighboring binding sites in clusters (2-9 binding sites). Number 1 represents the 3' binding site (boxplots: vertical line: median, box limits: interquartile range (IQR); whiskers 1.5x IQR). **h-j**. Correlation between the number of binding sites for clusters proximal (blue: < 0.5 kb) and distant (red: ≥ 0.5 kb) to the PAS and ($P^{(4.2xDazl)}$, **h**), dissociation rate constants ($k_{diss.}$, **i**), and maximal fractional occupancy (Φ^{max}, **j**), for individual binding sites in a given cluster. P-values (one way ANOVA) indicate significant inter-group differences for $P^{(4.2xDazl)}$ and Φ^{max}, but not for k_{diss}. $P^{(4.2xDazl)}$ and Φ^{max} depend on $k_{on}^{(4.2xDazl)}$, which correlates with the number of binding sites in a cluster, (**Fig.3c**). **k**. Correlation between kinetic parameters of individual binding sites in clusters with 6, 5, 4, and 3 binding sites. The Pearson correlation coefficient is indicated in the legend bar. Binding site number 1 indicates the 3' binding site in a cluster.

Extended Data Figure 6 | Link between Dazl binding in 3'UTRs and impact on mRNA level and ribosome association. **a**. Correlation between cumulative binding probabilities (ΣB) and number of binding sites in a cluster (N = 1,313 3'UTRs, 6,085 binding sites, 1,690 clusters), R^2: linear correlation coefficient). **b**. Correlation between ΣB and distance of the cluster from the PAS, R^2: linear correlation coefficient). **c**. Correlation of ΣB terciles (H: high; M: medium; L: low, **Fig.4a**) and changes in ribosome association (ΔRPF, **Fig.4b**) for the corresponding transcripts (N = 968) between low (1xDazl) and high (4.2xDazl) concentration (P value: one-way ANOVA). For UTRs with multiple clusters, the cluster closest to the PAS was utilized (boxplots: vertical line: median, box limits: interquartile range (IQR); whiskers 1.5x IQR). **d**. Correlation of ΣB terciles (H: high; M: medium; L: low, **Fig.4a**) and changes in transcript levels (ΔRNA, **Fig.4b**) for the corresponding transcripts between low (1xDazl) and high (4.2xDazl) concentration (P value: one-way ANOVA). For UTRs with multiple clusters, the cluster closest to the PAS was utilized. **e**. Distribution of binding probabilities for individual Dazl binding sites in 3'UTRs for transcripts in THRH, THRM, TLRM, TLRL, TMRH, TMRL mRNA classes (**Fig.4b).** The dotted lines mark terciles (H: high; M: medium; L: low), (for details, see Materials and Methods). **f**. Correlation between binding probabilities for individual binding sites and functional mRNA classes (**Fig.4b**). Colors mark the enrichment (hypergeometric test, red: $p < 0.05$, shades of yellow: not enriched, see color bar). No significant enrichment is observed. **g**. Distribution of cumulative binding probabilities for Dazl clusters in 3'UTRs with scrambled binding sites. The dotted lines mark terciles (H: high; M: medium; L: low). **h**. Correlation between cumulative binding probabilities of Dazl clusters with binding sites scrambled between clusters (panel **g**) and functional mRNA classes (**Fig.4b**). Colors mark the enrichment (hypergeometric test, Red: $p < 0.05$, shades of yellow: not enriched, see color bar). No significant enrichment is observed. **i.** Correlation between additive binding probabilities of two Dazl sites in a cluster and functional mRNA classes (hypergeometric test, red: $p < 0.05$, shades of yellow: not enriched, see color bar). For clusters with > 2 binding sites, permutations of two sites were tested and sites with highest additive binding probability were selected. The model tests whether the additive binding probability of any two Dazl binding sites in a given cluster can explain the impact of Dazl on the transcript to the same extent as considering cumulative binding probabilities for the entire cluster (**Fig.4c**). The model is only able to explain the TLRL, TLRM mRNA classes, which frequently contain transcripts with clusters that have only few Dazl binding sites. **j**. Correlation between conditional binding probabilities of two Dazl sites in a cluster (terciles) and functional mRNA classes (hypergeometric test, Red: $p < 0.05$, shades of yellow: not enriched, see color bar). For clusters with > 2 binding sites, permutations of two sites were tested and combinations of sites with the highest multiplicative binding probability were selected. The model tests whether the conditional binding probability of any two Dazl binding sites (e.g. whether Dazl needs to bind simultanously to both sites) in a given cluster can explain the impact of Dazl on the transcript to the same extent as considering cumulative binding probabilities for the entire cluster (**Fig.4c**). The model explains only mRNA classes which frequently contain transcripts with Dazl clusters that have only few binding sites. For these clusters cumulative and conditional binding probabilities scale similarly. The data suggest that simultaneous binding of Dazl to two sites in a cluster is not required for general Dazl function. **k**. Correlation between conditional binding probabilities of three Dazl sites in a cluster (terciles) and functional mRNA classes (hypergeometric test, Red: $p < 0.05$, shades of yellow: not enriched, see color bar). For clusters with > 3 binding sites, permutations of three sites were tested and combinations of sites with the highest multiplicative binding probability were selected. The model tests whether the conditional binding probability of three Dazl binding sites (e.g. whether Dazl needs to bind simultaneously to three sites) in a given cluster can explain the impact of Dazl on the transcript to the same extent as considering cumulative binding probabilities for the entire cluster (**Fig.4c**). The model explains only mRNA classes which frequently contain transcripts with Dazl clusters that have only few binding sites. For these clusters cumulative and conditional binding probabilities scale similarly. The data suggest that simultaneous binding of Dazl to two or more sites in a cluster is not required for Dazl function.

Extended Data Figure 7 | Link between Dazl clusters in 3'UTRs and impact on mRNA level and ribosome association. a. Distribution of transcript levels at 4.2xDazl **b.** Distribution of 3'UTR lengths [17,30,31]. For UTRs with multiple lengths, coordinates for the longest 3'UTR were utilized. **c.** Distribution of distances of Dazl clusters from PAS. **d.** Distribution of differential cumulative binding probability (ΔΣB) for all Dazl clusters. The dotted lines mark terciles (H: high;

M: medium; L: low). Terciles were defined by obtained standard deviations from the mean for each feature described above. **e**. Link between Dazl impact on mRNA level and ribosome association and cluster features (upper graphs: number of Dazl clusters in 3'UTR: black line; ΣB: blue vertical lines, lower end marking ΣB at 1 x Dazl, upper end ΣB at 4.2 x Dazl; middle graphs: ΔΣB for each cluster and number of Dazl binding sites in each cluster; Heatmaps below the graphs: terciles of transcript features obtained from panels **a**-**c**. Each panel shows one functional mRNA class [defined in **Fig.4b**; first letter T: change in ribosome association, second third letter R: change in transcript level upon increase in Dazl concentration. H-high (increase at high Dazl concentration), M-medium (no change), L-low (decrease at high Dazl concentration)]. Functional classes not displayed contained too few or no transcripts (TLRH: 0, THRL: 2) or showed no change in ribosome association and transcript level (TMRM). Numbers represent the groups in the Dazl-code (**Fig.4d**). Clusters with ΣB > 1 (N = 4) are not shown.

Extended Data Figure 8 | The Dazl regulatory program. **a**. Pairwise correlation between Dazl cluster features. Colors correspond to Pearson's' correlation coefficient. Cluster features are marked as indicated on the right. **b**. Variance of data reflected in the eigenvalues of principal component axes (N = 7) obtained by PCA. Each eigenvalue corresponds to a principal component axis. Each axis reflects a linear combination of Dazl cluster features (N = 7), obtained from panel (**a**). The eigenvalues and the corresponding principal component axis are sorted according to the initial variance they represent. The first three principal component axes explain roughly 90% variance. **c**. Biplots of Dazl cluster features (arrows) projected on the first two principal components (PC1,2; panel **b**). Dots represent transcripts. Colors correspond to terciles of the distributions of values for ΔRPF (H = High, M: Medium, L: Low, **Fig.4b**), ΔRNA (H = High, M: Medium, L: Low, **Fig.4b**), Colors correspond to terciles of the distributions of values for ΔRPF (TH = High, TM: Medium, TL: Low, **Fig.4b**), ΔRNA (RH = High, RM: Medium, RL: Low, **Fig.4b**), and functional mRNA classes (THRH, THRM, TLRM, TLRL, TMRH, TMRL, **Fig.4b**). Each arrow represents a cluster feature (labels as in panel (**a**)). Proximity of arrows scales with correlation between the corresponding features. Arrows in the x-direction (positive or negative) contribute to PC1, arrows in the y-direction (positive or negative) contribute to PC2. Short arrows (transcript level, proximity to PAS) indicate that additional principal components (PC3-7) are required to explain the corresponding feature. **d.** T-distributed Stochastic Neighbor Embedding (t-SNE, Perplexity = 10, Iterations = 2,000) of cluster features (panel **a**). Identified groups are marked 1-21. Each point represents a transcript. **e.** Biplots of Dazl cluster features (arrows) projected on three principal components (PC1,2,3, panel **b**). Dots represent transcripts. Colors correspond to functional mRNA classes (THRH, THRM, TLRM, TLRL, TMRH, TMRL, **Fig.4b**). Separation of transcripts in 21 groups is marked as 1-21. **f**. Link of functional mRNA classes to kinetic parameters (ΣB, ΔΣB), cluster features (number of binding sites in cluster, proximity to PAS) and UTR features (numbers of clusters on UTR, UTR length, transcript level). Left panel: enrichment of terciles (H, M, L; **Fig.4a**, **Extended Data Fig.7a-d**) for ΣB, ΔΣB, number of binding sites in cluster, cluster distance from PAS, UTR length and transcript level in group 1. Numbers and color indicate the degree of enrichment. The row on the left marks the visualization of the Dazl code for group 1 that is used in **Fig.4d**. Right panel: enrichment of terciles for the features indicated in the left panel for all groups (1-21). Functional mRNA classes for the respective groups are shown on the bottom. **g**. Genome browser traces of representative transcripts of select groups. mRNA classes are indicated. The y-axis represents normalized coverage value. **h**. Mapping of transcripts from select groups on two biological networks. Groups are colored as indicated in the legend. Proximity of transcripts of a given group in the network indicates closely related biological functions.

Extended Data Figure 9 | Decision tree classification linking the Dazl code to functional impact of Dazl binding. a. Decision tree classifier (Chi-squared automatic interaction detection (CHAID) algorithm [32-34] of 7 features (ΣB, ΔΣB, distance to PAS, 3'UTR length, transcript level; Clust/UTR: number of clusters in a given 3'UTR, **Extended Data Fig.8**) in terciles (H: high, M: medium, L: low, **Extended Data Fig.7**). Nodes (◊) mark the given feature and corresponding partition (high, medium, low). Circles indicate the number of transcripts, donut graphs mark the functional mRNA classes, color coded as shown on the right. Circled numbers left to the heatmap with the Dazl code (identical to that in **Fig.4d**) indicate the number of transcripts in a given group. The decision tree was calculated by cross-tabulation of predictor variables (transcripts, N = 413) with target variables (functional mRNA classes THRH, THRM, TLRM, TLRL, TMRH, TMRL, **Fig.4b**) followed by partitioning of predictor variables into statistically significant subgroups (X^2 test, for independence with significance threshold: 0.05 (ref.[35], **Supplementary Material Table S10**). **b**. Confusion matrix corresponding to the decision tree. Validation 1 (N = 24 transcripts) and Validation 2 (N = 21 transcripts) are predictions for transcripts that were not included in the decision tree classification.

Extended Data Figure 10 | Linear regression models for linking the Dazl code to Dazl impact on changes in transcript levels, ribosome association and translation efficiency. **a:** Distribution of changes in translational efficiency values (ΔTE) between high and low Dazl concentration for transcripts in the 21 groups of the Dazl regulatory program, defined in **Fig 4d**. mRNA functional classes are defined in **Fig.4b**. The grey area in the plot center marks unchanged ΔTE (95% confidence interval). p-values were calculated by one-way ANOVA of inter-group variations for each mRNA functional class (boxplots: horizontal line: median, box limits: interquartile range (IQR); whiskers 1.5x IQR)**. b**. Linear Regression models tested. (yellow: dummy coding, using terciles of the variables, **Extended Data Fig.8**. Red: no dummy coding; use of continuous data. Grey: variable was omitted. **c**. adjusted R^2 for each model. **d**. Differential Intercept Linear Coefficients (DILC) for each model. Grey boxes mark models without the respective variable. **e**. Significance of each DILC for each model (White: $p < 0.05$ - significant, Black $p > 0.05$ – not significant, p-values: student t-test on each coefficient term). M1 is the only model with consistently significant DILCs. Models 24-27 include interaction terms corresponding to 7 independent variable terms and test impact of multicollinearity. Interaction terms for each of the models were as follows: M24: ΣB | ΔΣB and ΣB | # binding sites in a cluster. M25: ΣB | ΔΣB. M26: ΣB | ΔΣB and ΣB: Proximity from PAS. M27: ΣB | Proximity to PAS. Interaction terms are the cross product of encompassing independent variable terms and were selected based on pairwise correlation coefficients (**Extended Data Figure 8a**). **f:** Linear regression model linking the Dazl regulatory program to changes in translational efficiency values (ΔTE) (panel **a**). Points represent the differential intercept (DI) linear coefficient (LC) (red: DILCs for translational efficiencies that increase at high Dazl concentration, black: DILCs for translational efficiencies that decrease at high Dazl concentration). **g:** Correlation between experimental values for ΔTE and values predicted with the linear regression model (Adj. R: adjusted linear correlation coefficient) for test dataset. **h:** Correlation between predicted values for ΔRPF (N = 6) and changes in luciferase activity between high and low Dazl concentration for reporter RNA constructs. Reporters were generated by appending the 3'UTR of the respective transcripts to a luciferase ORF, and measurements were performed as described in ref.17. Naa40 and Ptma were part of model building data set (training data set). Calm2, Cxcl1, D'Rik and Spp1 were part of the test dataset. (R: linear correlation coefficient).

MATERIALS AND METHODS

Laser Setup

The cross-linking experiments were performed by using a Ti:Sapphire regenerative amplifier laser system (Libra-HE, Coherent, Inc.; λ = 800 nm (center wavelength, nominal), pulse width ≤100 fs (Full Width at Half Maximum), 4.0 W at 1 kHz, contrast ratio > 1000:1 pre-pulse; > 100:1 post-pulse; root mean square (8 h) energy stability under stable environmental conditions after system warmup < 0.5 %). The 800 nm fundamental beam was converted to the 270 nm excitation beam by second harmonic sum frequency generation with an optical parametric amplifier (TOPAS, Quantronix/Light Conversion)[36,37]. Contributions to the excitation beam from other wavelengths were removed by a set of dichroic mirrors (λ-filter) and a Glan-Taylor polarizer [37]. The excitation beam was collimated to a spot size of 6.0 mm. The photon flux at the sample was $1.25 \cdot 10^{16}$ $cm^{-2}s^{-1}$ (2.6 mW) and $4.81 \cdot 10^{15}$ $cm^{-2}s^{-1}$ (1 mW) at 270 nm with a pulse duration of 200 (± 50) fs, assuming a Gaussian-shaped pulse [38]. Stability of the laser output at λ = 270 nm was monitored with a silicon photodiode (S120VC, ThorLabs). The power of the excitation beam was attenuated with a neutral density filter for the crosslinking experiments with the average power of 2.6 mW and 1.0 mW. The crosslinking experiments were conducted in a 2 mm optical path length quartz cell with a maximum sample volume of 0.7 mL, placed orthogonal to the excitation beam. Homogeneity of the sample in the cuvette was maintained with a Teflon-coated magnetic stirring bar (Sterna Cells, Inc.) throughout the measurement. Temperature in the cuvette before and immediately after measurements was monitored with a thermo-coupling device.

RNA degradation measurements

Cy3 labelled RNA oligonucleotide was purchased from Dharmacon (Lafayette, Colorado). RNA degradation by fs laser was measured for 0.15 µM of 38 nt Cy3 labelled RNA substrate (V = 600 µL, 60 mM KCl, 6 mM HEPES-pH 7.5, 0.2 mM $MgCl_2$, 5'-GCU UUA CGG UGC UUA AAA CAA AAC AAA ACA AAA CAA AA-Cy3-3'), irradiated with the fs laser (2.6 mW) as described above for 0, 100, 200, 300 and 680 s. RNA degradation by steady-state UV irradiation was measured for 0.15 µM of the 38 nt Cy3 labelled RNA substrate (V = 50 µL ,60 mM KCl, 6 mM HEPES-pH 7.5, 0.2 mM $MgCl_2$) irradiated in a Stratalinker (Fisher Scientific, 200 mJ/cm^2) for same time points. Following irradiation, samples were subjected to denaturing PAGE (4-12% Novex NuPage Bis-Tris (Invitrogen), 60 min, 100 V). Samples on the gels were quantified using a

Phosphorimager (GE) in fluorescence detection mode. Intact and degraded RNA bands were quantified using the ImageQuantTL (GE) software. The fraction degraded RNA (Frac D) at each time point was calculated according to:

$$Frac\ D = I_D \cdot (I_{ND} + I_D)^{-1} \qquad \text{(Eq.1)}$$

(I_D: fluorescence intensity degraded RNA, I_{ND}: fluorescence intensity non-degraded RNA)

Photons absorbed over time (**Extended Data Fig.1b**) were calculated according to [11,13]

$$Dose\ absorbed = [I^0 \cdot t \cdot \sigma \cdot (1-10^{-A})] \cdot (2.3 \cdot A) \qquad \text{(Eq.2)}$$

(I^0 = intensity of incident light in photons $cm^{-2}\,s^{-1}$; t = duration of irradiation; A = absorbance of protein-RNA solution in Absorbance Units (AU), σ = mean cross section of absorption of nucleic acids). For the fs laser: $I^0 = 2 \cdot 10^{27}$ photons $cm^{-2}\,s^{-1}$ (refs. [9,13]), A_{270} = 0.99 AU (Absorbance Units of protein-RNA solution), σ = 2.7 x 10^{-17} cm^2 $molecule^{-1}$ (ref.[13]). For the steady-state UV irradiation (Stratalinker, 400 mJ /cm^2) I^0 = 2 .10^{15} photons $cm^{-2}\,s^{-1}$, A_{270} = 0.99 AU, σ = 2.7 x 10^{-17} cm^2 $molecule^{-1}$ (ref.[13]).

Protein expression and purification

Mus musculus Dazl(RRM) (amino acids 32 - 117) was codon-optimized (Dapcel, OH) for expression in *E.coli*. (**Supplementary Material Table S1**). The DNA construct was chemically synthesized (Genscript, NJ) and cloned into a pET-22b vector with an N-terminal His_6 - Sumo cleavable tag. Protein was expressed in *E.coli* (BL21) cells overnight at 19°C and purified through Ni^{2+} affinity column [16]. Samples were dialyzed (20 mM HEPES, pH7.5, 100 mM NaCl), the His_6-Sumo tag was removed with Sumo protease (Ulp1) at 4°C overnight. Dazl(RRM) protein was further purified by gel filtration chromatography (Superdex 75) equilibrated in 20 mM HEPES (pH 7.5), 100 mM NaCl, 5% (v/v) glycerol. Peak fractions were pooled and concentrated with Amicon ultra centrifugal filters. RbFox(RRM) (amino acids 109-208) and $RbFox^{mut}$(RRM) (amino acids 109-208, R118D, E147R, N151S, E152T mutations) proteins were prepared as described [15]. Protein concentrations were determined by UV absorbance at 280 nm and validated with Bradford assays.

RNA-protein affinity measurements by fluorescence polarization

Purified proteins RbFox(RRM), $RbFox^{mut}$(RRM), Dazl(RRM) at different concentrations and corresponding cognate 3'-Cy3 RNAs (20 nM, RbFox: 5′-UCCUGCAUGUUUA-Cy3-3', Dazl: 5′-UUGUUCUUU-Cy3-3', cognate motifs underlined; modified RNAs purchased from Dharmacon, Lafayette, Colorado) were incubated for 10 min (20 mM HEPES (pH 7.5), 100 mM NaCl and 0.01% (v/v) NP-40). Solutions were transferred to a 96-well plate (Greiner Bio-one), and fluorescence polarization was measured in a Tecan M1000-Pro microplate reader (Tecan, Switzerland). Plots of the fraction bound RNA vs. protein concentrations were fitted against the quadratic binding equation using KaleidaGraph (Synergy, PA) [16].

$$\text{Fraction Bound} = A \times \frac{(K_{1/2}+R_0+P_0) - \sqrt{\{(K_{1/2}+R_0+P_0)^2 - 4\times R_0\times P_0\}}}{2\times R_0} \qquad \text{(Eq.3)}$$

(A: reaction amplitude, $K_{1/2}$: apparent dissociation constant, R_0: RNA concentration, P_0: protein concentration)

fs laser RNA-protein crosslinking *in vitro*

Cy3 labelled RNA oligonucleotides corresponding to cognate sequences for RbFox(RRM) and Dazl(RRM) (described above, 5 nM, final concentration) and protein (10 nM, 50 nM, final concentration) were combined in a cuvette (V = 600 µL, 20 mM HEPES (pH 7.5), 100 mM NaCl, 5% (v/v) glycerol, 25°C) and incubated for 5 min. Longer incubation times did not change results, indicating that equilibrium was reached. The solution in the cuvette was constantly stirred during the reaction (200 rpm), using a magnetic stirbar. Laser power during the measurement was monitored with a photodiode, as described above. Temperature in the cuvette was measured before and after reactions. The RNA-protein mix was irradiated with the UV laser at two different powers (1.0 mW and 2.6 mW, 270 nm). Each timepoint was measured in a separate reaction, avoiding volume changes during the crosslinking experiment. Following crosslinking, samples were removed from the cuvette and stored on ice. Crosslinked and non-crosslinked RNA were separated on denaturing PAGE (4-12% Novex NuPage Bis-Tris gel, 200 V, 45 min). Fluorescence of crosslinked and non-crosslinked RNA in the gels was measured with a Phosphorimager (GE) and quantified with the ImageQuant TL Software (GE). The fraction cross-linked RNA (Frac XL) at each time point was calculated according to:

$$Frac\ XL = I_{XL} \cdot (I_{XL} + I_{NX})^{-1} \qquad \text{(Eq.4)}$$

(I_{XL}: fluorescence intensity crosslinked material, I_{NX}: fluorescence intensity non-crosslinked material).

Determination of kinetic parameters from RNA-protein crosslinking experiments *in vitro*.

Timecourses at different protein concentrations and laser intensities were globally fit to a two-step kinetic model (**Fig.1a**) using KinTek Global Kinetic Explorer (Kintek, Austin TX). Data fit started from a pre-equilibrated mixture of protein and RNA, mirroring the experiments. Initial conditions were identified from an array of different starting values for $\boldsymbol{k}_{\mathrm{on}}$, $\boldsymbol{k}_{\mathrm{off}}$ and $\boldsymbol{k}_{\mathrm{xl}}$. Multiple iterations were performed with various combinations of floating and fixed rate constants until the best fit to all data sets was achieved (**Fig.1e**). The quality of the global fit was assessed by computation of Chi-squared (X^2) values with each parameter ($\boldsymbol{k}_{\mathrm{on}}$, $\boldsymbol{k}_{\mathrm{off}}$ and $\boldsymbol{k}_{\mathrm{xl}}$) varied individually (1D fit space, **Supplementary Material Fig.S2a-c**) and for co-variations of $\boldsymbol{k}_{\mathrm{on}}$ and $\boldsymbol{k}_{\mathrm{off}}$ (2D fits pace, **Supplementary Material Fig.S2d**) Confidence intervals are given as upper and lower bounds at 95% of the relative X^2. To visually assess the quality of the fit, curves with calculated rate constants were overlaid on experimental values.

Cell culture

GC-1*spg* cells with inducible DAZL expression were maintained in DMEM high glucose medium (ThermoFisher) supplemented with 10% (v/v) Tet-system approved FBS (Clontech), 100 U/mL penicillin, 100 mg/mL streptomycin, 5 mg/mL blasticidin, and 300 mg/mL Zeocin (all from ThermoFisher) at 37°C, 5% (v/v) CO_2 (ref.[17]). Doxycycline induction of Dazl was performed and lysates for generation of cDNA libraries and quantification of Dazl levels were prepared as described [17]. Equal amounts of protein were run on a SDS-PAGE (10% NEXT Gel, Amresco) and transfered to a PVDF membrane. Western blotting was performed with anti-Dazl (Rabbit; 1:5000, US Biological) and anti-Hsp90 (Rabbit; 1:10,000; US Biological) antibodies. Chemiluminescence was quantified with the ImagequantTL software.

fs laser crosslinking of GC-1 cells

GC-1*spg* cells (with doxycyline induction of Dazl expression) were grown in 150 mm plates to 70% confluency. Cells were rinsed with 2 mL PBS (per plate), scraped, re-suspended in 600 µL PBS, transferred to the quartz cuvette and stirred with a magnetic stir bar (described above). Crosslinking of the cell suspension was performed as described above at two laser powers (1.0 mW, 2.6 mW) in separate experiments for 30, 180 and 680 s (25°C). Each crosslinking reaction contained a constant number of cells ($6 \cdot 10^5$). To generate sufficient material for timepoints with low crosslinking yield, multiple identical experiments were conducted and pooled. Temperature in the cuvette was measured before and after crosslinking (increase was less than 1°C after 680 s). Cell integrity after crosslinking was measured by Trypan-blue staining [39] and cell counting in a hemocytometer. After crosslinking, cell suspensions were pelleted at 1,000 g for 5 min (4°C). The pellet was suspended in PBS (3x dry volume). Cells were pelleted again (1,000 g for 5 min), the supernatant was removed, and pellets were frozen and stored at -80°C until further processing.

cDNA library preparation

Cell lysates for each sample were split into two aliquots (A1, A2). RQ1 DNase (PromegaM6101) and RNAse A (USB70194Y) were added at 1:100 (A1) and 1: 20,000 (A2). Over-digested sample (A1) confirmed the size of the Dazl-RNA radioactive band on SDS-PAGE gel. The under-digested cell supernatant from the under-digested sample (A2, equivalent to ~150 mg of cell lysate) was mixed with protein G Dynabeads (ThermoFisher 10009D) with anti Dazl antibody (Rabbit; 1:5000) in separate Eppendorf tubes for each sample (N = 16). Samples were treated with CIP (Roche712023). RNA linker ligation and PNK (NEBM0201S) treatment were performed as described [17]. The supernatants were loaded onto separate Novex NuPAGE 4-12% Bis-Tris gels, and crosslinked material was transferred to a nitrocellulose membrane. Samples were located on the membrane by autoradiography and RNA-Dazl complexes at 50 - 70 kDa (Dazl molecular weight; 37 kD) were cut. Nitrocellulose fragments were treated with proteinase K (Roche1373196). Dazl bound RNA was isolated, reverse transcribed (SuperScript III; Invitrogen18080051), circularized and amplified to obtain 16 cDNA libraries. The RT primers used contained iSP18 spacers and phosphorylated 5' end for circularization of first strand cDNA to generate PCR template without linearization [17]. Unique molecular identifiers (UMIs, randomized barcodes, 11 nt with 4 nt random nucleotides) were used to determine PCR amplification artifacts (primer sequences: **Supplementary Material Table S2**). cDNA diversity in each library was tested before next generation sequencing by cloning cDNA from each library into pBS plasmid, subsequent transformation in competent cells, colony PCR and DNA sequencing. Illumina Sequencing for all cDNA libraries was performed at the Case Western sequencing core facility.

Measurement of bulk crosslinking

For each KIN-CLIP library, cells were cross linked and cell lysate was prepared as described above. 200 µL aliquots (equivalent to 150 mg of cell lysate) for each KIN-CLIP sample were treated with RQ1 DNase and RNAse (at 1: 20,000) as described above. Treated lysates were centrifuged in a pre-chilled ultra-centrifuge, polycarbonate tubes, TLA 120.2 rotor at 30,000 rpm, 20 min, 75 µl of the supernatant were removed and RNA was 5'-radiolabeled with PNK. Samples were run on a SDS-PAGE gel and transferred to a nitrocellulose membrane. The radioactivity was measured by quantifying the intensity of the radioactive bands (using ImageJ software). Lane background was used to normalize the band intensities.

KIN-CLIP read processing, refinement and mapping

Raw sequencing reads were assessed for quality (FastQC, https://www.bioinformatics.babraham.ac.uk) and de-multiplexed. Low-quality reads were removed if ≤ 80% of sequenced bases in a read had a PHRED quality score of ≤ 25. De-multiplexing and read filtering was performed with the FASTX-Toolkit (http://hannonlab.cshl.edu/fastx_toolkit/) using standard commands [40]. Filtered reads were stored in FASTQ format. Barcode and UMI (randomized 4nt sequence) were kept appended to line 1 of the FASTQ for each read.

Read duplicates, as identified by UMIs were collapsed into a single read. Linkers and concatamers were removed with the FASTX-Toolkit (http://hannonlab.cshl.edu/fastx_toolkit/), using permutations (N = 25) of linker sequences as target. Reads with ≥ 15 nt were retained for subsequent analysis. Processed reads were aligned against the mouse genome (mm10) by using bowtie2 [41] with the following settings for a 50 bp sequencing run: Number of mismatches allowed in seed alignment during multi-seed alignment = *1*, length of the seed substrings to align during multi-seed alignment = *15*, set a function governing the interval between seed substrings to use during multi-seed alignment = *S,1,0.50,* function governing the maximum number of ambiguous characters (N's and/or '.'s) allowed in a read as a function of read length = *L,0,0.15*, disallow gaps within this many positions of the beginning or end of the read = *4*, set a function governing the minimum alignment score needed for an alignment to be considered `valid` = *L, -0.6, -0.6*, set the maximum (`MX`) and minimum (`MN`) mismatch penalties, both integers = *6,2,* sets penalty for positions where the read, reference, or both, contain an ambiguous character such as `*N*` = *1,* gap opening penalty = *5*, gap extension penalty = *3*, attempt that many consecutive seed extension attempts to `fail` before Bowtie 2 moves on, using the alignments found so far = *20*, set the maximum number of times Bowtie 2 will `re-seed` reads with repetitive seeds = *3*. End-to-end alignment mode was used. Only uniquely mapped reads were retained. To evaluate the stringency of filtering and sequence alignment, the fraction of uniquely mapped tags over all mapped reads was assessed [40] by employing different permutations of read mapping parameters described above. In total, 55 parameter permutations for mapping were tested. The setting yielding the largest number of uniquely mapped reads is shown above. The *BAM* index of mapped reads corresponding to the 16 KIN-CLIP libraries was then converted to *BED*/*bedgraph* using the standard command line version of –bedtools (V2.29.1) and –samtools (V1.10) [42]. Bedgraph files were visualized in the IGV [43].

Identification of KIN-CLIP peaks

Genomic coordinates of the 5'-terminal nucleotide (5'nt) of every mapped read were obtained. Adjacent 5'nt were summed at single nucleotide resolution level by creating a sliding window of 11nt (stride = 1, steps = 5nt on either side or until no new reads were detected), with the 5'nt position at the center. Crosslinking peaks were defined by plotting the distribution of the count of 5'nt reads in these windows for every location. The peak apex represents the coordinate for the crosslinking peak and the associated coverage value. Error ranges for coverage values corresponding to each crosslinking peak were defined as the 95% confidence interval from the apex of crosslinking peaks. Coordinates of crosslinking peaks present in all KIN-CLIP libraries, except at the zero timepoint were used to define Dazl binding sites for further analysis. For peaks with coverage at the zero timepoint (~0.2% of peaks), the peak value at $t = 0$ was subtracted from the KIN-CLIP peaks. Coverage values for each Dazl binding site were converted into a concentration equivalent by normalizing to the amount of bulk crosslinked RNA for each KIN-CLIP library (**Supplementary Material Table S6**). The normalized read coverage values were used for calculating kinetic parameters and other subsequent analyses.

Analysis of read distribution

To annotate KIN-CLIP Dazl binding sites, RefSeq coding regions, 5'UTRs, 3'UTRs, ORF, introns, and RNA types were obtained from the UCSC genome browser and intersected individually with KIN-CLIP binding site coordinates using *bedtools*.

CITS analysis and sequence enrichment

Crosslink Induced Truncation Site (CITS) analysis was performed as described [28,29]. Enrichment of motifs at and around CLIP regions was performed using the EMBOSS tool Compseq [44], R package 'randomizeR' [45] and 'Random' [46] module in Python. To generate z-scores, shuffled control sets were generated for each dataset analyzed using Random module available in Python (Shuffle N = 10,000).

Distribution of Dazl-RNA contacts in 3'UTRs

Metagene analysis of Dazl-3'UTR interactions was performed on 3'UTRs as defined by PolyA-Seq [17]. To define 3'UTR length, coordinates from Refseq and Ensembl [30,31] were matched with PolyA-Seq data [17]. For transcripts with multiple 3'UTR length annotations, coordinates for the longest 3'UTR were utilized. 3'UTRs that overlapped with intron sequences annotated in either RefSeq or Ensembl were omitted. To calculate distances of binding sites to PAS and stop codons, the distance between coordinates for each KIN-CLIP binding relative to the Stop codon and to the PAS (10 nt window) was measured. For each 3'UTR, the random distribution of binding sites was determined by scrambling all Dazl binding sites (1,000 times) in that 3'UTR into all probable 10 nt bins in that 3'UTR and obtaining the average.

Calculation of kinetic parameters

Kinetic parameters were calculated from normalized peak coverage values for each Dazl binding site (N = 10,341). A Dazl binding site was defined by the presence of more than 5 normalized sequencing reads in the library for the (4.2xDazl, 2.6 mW laser) 680 s timepoint, within 11 nucleotides of the peak apex for the binding site in all libraries. Sites without normalized reads for the 30s (1XDazl, 1.0 mW laser) timepoint were excluded, as it is not possible to calculate meaningful kinetic parameters from such sparse data. Kinetic parameters were calculated according to two different approaches: (i) a numerical and (ii) an analytical method. Parameters from both methods were averaged for subsequent data analysis (**Extended Data Fig. 3**).

Numerical approach

The numerical approach to calculate kinetic parameters is based on numerically fitting crosslinking timecourses to the differential equations describing the Dazl-RNA binding and crosslinking process (**Fig.1a**), according to:

$$\frac{d(DR)}{dt} = k_{\mathrm{on}}(D)(R) - k_{\mathrm{diss.}}(DR) - k_{\mathrm{XL}}(DR) \qquad \text{(Eq.5)}$$

$$\frac{d(DR^{*})}{dt} = k_{\mathrm{XL}}(DR) \qquad \text{(Eq.6)}$$

(DR: concentration of non-crosslinked Dazl-RNA complex (for each binding site), DR*: concentration of crosslinked Dazl-RNA complex (for each binding site), D: Dazl concentration, R: RNA concentration (binding site), k_{on}: association rate constant, $k_{\text{diss.}}$: dissociation rate constant, k_{XL}: crosslinking rate constant).

Because concentrations of free Dazl and RNA in the cell are experimentally inaccessible, the second order association process (k_{on}) was treated as pseudo-first order reaction at each of the two Dazl concentrations. Accordingly, we calculated a pseudo first order rate constant for each Dazl concentration ($k_{\text{on}}^{(1\text{xDazl})}$, $k_{\text{on}}^{(4.2\text{xDazl})}$), and $k_{\text{diss.}}$, $k_{\text{XL}}^{(1\text{mW})}$ and $k_{\text{XL}}^{(2.6\text{mW})}$ for each binding site. Numerical fitting of timecourses of normalized read coverage for each binding site (**Fig.2c**) was performed in R with packages deSolve (with ODE function) [47], ggplot2 [48], reshape2 [49] and rmarkdown [50].

The fitting strategy encompassed two steps: (i) estimation of parameter ranges following a sequential parameter estimation procedure [51] and (ii) fitting the timecourses using estimated parameter ranges as input (**Supplementary Material, Scheme 1**). Estimation of parameter ranges was also performed in two steps, (i,a) *initial* parameter range estimation for $k_{\text{on}}^{(1\text{xDazl})}$, $k_{\text{on}}^{(4.2\text{xDazl})}$, $k_{\text{diss.}}$, $k_{\text{XL}}^{(1\text{mW})}$ and $k_{\text{XL}}^{(2.6\text{mW})}$, and (i,b) refinement of *initial* parameter range estimates to obtain *final* parameter range estimates (**Supplementary Material, Scheme 1**). To estimate *initial* parameter ranges, timecourses from reactions with 4.2xDazl at high laser power (2.6 mW) and low laser power (1mW) were fit separately. Starting values were based on the kinetic parameters measured *in vitro* (**Fig.1**; $k_{\text{on}}^{(1\text{xDazl})} = 0.0001$ s^{-1}, $k_{\text{on}}^{(4.2\text{xDazl})} = 0.0001$ s^{-1}, $k_{diss.} = 1$ s^{-1}, $k_{\text{XL}}^{(1\text{mW})} = 1$ s^{-1} and $k_{\text{XL}}^{(2.6\text{mW})} = 10$ s^{-1}. Use of significantly different starting values did not yield acceptable fits for the majority of binding sites). This step provided average initial values for $k_{\text{on}}^{(4.2\text{xDazl})}$ and $k_{\text{diss.}}$ as well as initial values for $k_{\text{XL}}^{(1\text{mW})}$ and $k_{\text{XL}}^{(2.6\text{mW})}$. Next, timecourses at 1xDazl at high laser power (2.6mW) and low laser power (1mW) were fit separately, yielding average initial values for $k_{\text{on}}^{(1\text{xDazl})}$ and $k_{diss.}$ and initial values for $k_{\text{XL}}^{(1\text{mW})}$ and $k_{\text{XL}}^{(2.6\text{mW})}$. This process was performed for each binding site until the X^2 was minimized (no change in X^2 for 4 consecutive cycles) or 1,000 fitting cycles were completed. The process provided 10,341 x 5 parameter values, which were plotted as distribution (10,341 values for each parameter). The *initial* parameter range estimate represents the 95% confidence interval from the mean of the distribution for $k_{\text{on}}^{(1\text{xDazl})}$, $k_{\text{on}}^{(4.2\text{xDazl})}$, $k_{\text{diss.}}$, $k_{\text{XL}}^{(1\text{mW})}$ and $k_{\text{XL}}^{(2.6\text{mW})}$.

To refine parameter range estimates to obtain *final* parameter range estimates, the initial parameter range estimates were used as input to fit multiple, random subsets of 2,000 randomly selected binding sites. 10,000 iterations, each with a unique random subset of 2,000 binding were performed. Each iteration yielded a distribution. All 10,000 distributions were superimposed and the median apex of all distributions was identified. The *final* parameter range estimates represent the 95% confidence interval from the median apex of the averaged distributions. The *final* parameter range estimate was about 35% smaller than the *initial* parameter range estimate.

The estimated parameter ranges were used as input for fitting of the timecourses (**Supplementary Material, Scheme 1**). We fitted timecourses for reactions at 4.2xDazl at the different laser powers (1 mW, 2.6 mW), varying linked $k_{\text{on}}^{(4.2\text{xDazl})}$ and $k_{\text{diss.}}$ (which do not scale with laser power), and differing $k_{\text{XL}}^{(1\text{ mW})}$ and $k_{\text{XL}}^{(2.6\text{ mW})}$. We then fit timecourses at 1xDazl at both laser powers, varying linked $k_{\text{on}}^{(1\text{xDazl})}$ and $k_{\text{diss.}}$, and differing $k_{\text{XL}}^{(1\text{ mW})}$ and $k_{\text{XL}}^{(2.6\text{ mW})}$. Utilizing parameters obtained from these two steps we fit all 4 timecourses linking $k_{\text{on}}^{(4.2\text{xDazl})}$ and $k_{\text{on}}^{(1\text{xDazl})}$ for differing laser powers, linking $k_{\text{XL}}^{(2.6\text{ mW})}$, $k_{\text{XL}}^{(1\text{ mW})}$ for differing Dazl concentrations and linking $k_{\text{diss.}}$ for all conditions. The process of fitting all 4 timecourses for each binding site was repeated 642 times, after which χ^2 did not show significant fluctuation (< 5% for 4 consecutive cycles). Obtained rate constants were used as final kinetic parameters for the numerical approach (**Extended Data Fig.3b-d**).

Fitting quality was assessed by calculating chi-squared (χ^2) for each binding site, the overall cumulative reduced chi-squared (χ_v^2) and the coefficient of determination/R^2 (COD) according to:

$$\chi^2 = \sum_i \frac{(O_i - C_i)^2}{\sigma_i^2} \qquad \text{(Eq.7)}$$

(*O*: observed value, *C*: calculated value for each binding site (i). σ_i^2 is the squared variance between data points *O, C*);

$$\chi_v^2 = \frac{\chi^2}{v} \qquad \text{(Eq.8)}$$

[*v*: degree of free; equals *(n – m)*, with *n*: number of observations (*n = 16*), *m*: number of fitted parameters (*m* = *5*)].

The coefficient of determination/R^2 (COD) was calculated using the standard method as described [52]. The COD describes correlation between calculated and observed timecourses. For the last fitting cycle, COD = 0.92, X_v^2 = 0.043 (**Extended Data Fig.3c**).

Analytical approach

The analytical approach to calculate kinetic parameters is based on fitting of crosslinking timecourses to explicit solutions of the system of differential equations (Eqs.5,6) for the kinetic scheme (**Fig.1a)**. To solve the system of differential equations, we considered that at any given time *(t)* during crosslinking, the accessible fraction of a given Dazl binding site is either free *(R)*, occupied *(DR)* or crosslinked *(DR*)*:

$$(R)_t + (DR)_t + (DR)_t^* = 1 \qquad \text{(Eq.9)}$$

In addition, at $t \rightarrow \infty$, 100% of the accessible fraction of a given Dazl binding site is crosslinked. As described for the numerical approach, the second order association process (k_{on}) was treated as pseudo-first order process at each Dazl concentration.

Treating second order association process (k_{on}) as pseudo-first order process, considering Eq.9 and rearranging Eq.5 yields:

$$\frac{d(DR)}{dt} = k_{\mathrm{on}}\left[1 - [DR]_{(t)} - DR_{(t)}^{*}\ \right] - k_{\mathrm{diss.}}[DR]_{(t)} - k_{\mathrm{XL}}[DR]_{(t)} \qquad \text{(Eq.10)}$$

Before crosslinking (t = 0), at steady-state of the binding reaction,

$$\frac{\mathrm{d}(DR^*)}{\mathrm{dt}} = 0 \qquad \text{(Eq.11)}$$

because

$$k_{\mathrm{XL}} = 0 \qquad \text{(Eq.12)}$$

From Eq.5, we thus obtain:

$$0 = k_{\mathrm{on}}[R] - k_{\mathrm{diss.}}[DR] \qquad \text{(Eq.13)}$$

which yields, after rearranging,

$$[DR]_{(t)} = \frac{k_{\mathrm{on}}}{k_{\mathrm{on}} + k_{\mathrm{diss.}}} \qquad \text{(Eq.14)}$$

At t = ∞, crosslinking is complete, and thus

$$\frac{d(DR)}{dt} = 0 \qquad \text{(Eq.15)}$$

$$\frac{\mathrm{d}(DR^*)}{\mathrm{dt}} = 0 \qquad \text{(Eq.16)}$$

The boundary limits are:

$$\lim \cong {DR^*}_0 \cong 0 \qquad \text{(Eq.17)}$$

$$\lim \cong {DR^*}_\infty \cong 0 \qquad \text{(Eq.18)}$$

Equations 11-18 define the boundary conditions.

Crosslinking timecourses represent amount of crosslinked material at a given time (t), expressed as normalized coverage value for each binding site $[DR^*]_{(t)}$. $[DR^*]_{(t)}$ depends on amount of Dazl-RNA complex [DR] at the time (t) (Eq.6) and thus on k_{on}, $k_{\mathrm{diss.}}$ and k_{XL}. Absolute concentrations of [D], [R] and [DR] are not known in our system. To extract k_{on}, $k_{\mathrm{diss.}}$ and k_{XL} for each binding site from the crosslinking timecourses we integrate Eq.6 after appropriate substitution of [DR]. To accomplish this, we take a second differential of Eq.10, considering the boundary conditions (Eq.11-18). We obtain the general solution of the second order differential equation:

$$\frac{d^2(DR)}{dt^2} = k_{\mathrm{on}}\frac{\mathrm{d(DR)}}{\mathrm{dt}} - k_{\mathrm{on}}\frac{\mathrm{d(DR^*)}}{\mathrm{dt}} - k_{\mathrm{diss.}}\frac{\mathrm{d(DR)}}{\mathrm{dt}} - k_{\mathrm{XL}}\frac{\mathrm{d(DR)}}{\mathrm{dt}} \qquad \text{(Eq.19)}$$

$$\frac{d^2(DR)}{dt^2} = -(k_{\mathrm{on}} + k_{\mathrm{diss.}} + k_{\mathrm{XL}})\frac{\mathrm{d(DR)}}{\mathrm{dt}} + (k_{\mathrm{XL}}k_{\mathrm{on}})[DR](t) \qquad \text{(Eq.20)}$$

Equation 20 is a constant coefficient, homogenous, linear, second order differential equation with two independent solutions (*y1*, *y2*) [53]:

$$y(t) = c1y1(t) + c2y2(t) \qquad \text{(Eq.21)}$$

The coefficients *c1* and *c2* (by the principle of superposition) [54] are obtained after providing the boundary conditions from equations 11-18. We identify a function *y* where a constant multiplied by its second derivative *y"* plus another constant times *y'* plus a third constant multiplied by *y* equals zero [54].

The exponential function

$$y = e^{rx} \text{ (}r\text{: constant).} \qquad \text{(Eq.22)}$$

has the property that its derivative is a constant multiple of itself:

$$y' = re^{rx} \qquad \text{(Eq.23)}$$

Furthermore,

$$y'' = r^2e^{rx} \qquad \text{(Eq.24)}$$

Substituting these expressions into (Eq.20), we obtain:

$$ar^2 + br + c = 0 \qquad \text{(Eq.25)}$$

Equation 25 is the auxiliary (characteristic) equation of the differential equation 20 (ref. [55]). The equation is transformed into an algebraic equation by replacing

$$\frac{d^2(DR)}{dt^2} = r^2, \qquad \text{(Eq.26)}$$

$$\frac{\mathrm{d(DR)}}{\mathrm{dt}} = r \qquad \text{(Eq.27)}$$

and $[DR]$ by 1.

The roots of Eq.25 are found by factoring [55]:

$$r_1 = \frac{(k_{\mathrm{on}}+k_{\mathrm{diss.}}+k_{\mathrm{XL}}) + \sqrt{(k_{\mathrm{on}}+k_{\mathrm{diss.}}+k_{\mathrm{XL}})^2-4(k_{\mathrm{XL}}k_{\mathrm{on}}[P])}}{2} \qquad \text{(Eq.28)}$$

$$r_2 = \frac{(k_{\mathrm{on}}+k_{\mathrm{diss.}}+k_{\mathrm{XL}}) - \sqrt{(k_{\mathrm{on}}+k_{\mathrm{diss.}}+k_{\mathrm{XL}})^2-4(k_{\mathrm{XL}}k_{\mathrm{on}}[P])}}{2} \qquad \text{(Eq.29)}$$

With Eq.21-29, the general solution of Eq.20 is [56]:

$$[DR]_t = c_1e^{r_1t} + c_2e^{r_2t} \qquad \text{(Eq.30)}$$

To obtain our observable *$[DR^*]_{(t)}$*, we integrate Eq.6 under consideration of the boundary conditions (Eqs.11-18):

$$[DR^*]_{(t)} - [DR^*]_{(0)} = k_{\mathrm{XL}}\int_0^t [DR]_{(t)}\,dt \qquad \text{(Eq.31)}$$

Substituting *$[DR]_t$* from Eq.30 yields

$$[DR^*]_{(t)} - [DR^*]_{(0)} = k_{\mathrm{XL}}[r_1c_1(1-e^{r_1t}) + r_2c_2(1-e^{r_2t}) \qquad \text{(Eq.32)}$$

Substituting *c1* and *c2* by providing the boundary conditions (Eqs.11-18) and considering (Eqs.21-29), we obtain:

$$[DR^*]_{(t)} = k_{\mathrm{XL}}[\frac{1}{k_{\mathrm{XL}}} - r_1(1-\frac{k_{\mathrm{on}}}{k_{\mathrm{on}}+k_{\mathrm{diss.}}})(1-e^{r_1t}) + r_2(1-\frac{k_{\mathrm{on}}}{k_{\mathrm{on}}+k_{\mathrm{diss.}}})(1-e^{r_2t}) \qquad \text{(Eq.33)}$$

Equation 33 is an explicit nonlinear equation of the form:

$$Y = f(t,\beta) + \varepsilon \qquad \text{(Eq.34)}$$

$t = (t_1, t_2, \ldots \ldots . t_n)$ are the independent variables (the normalized read coverage values at different timepoints), *$\beta = (\beta_1, \beta_2, \ldots \ldots . \beta_n)'$* are the parameters ($k_{\mathrm{on}} = k_{\mathrm{on}}^{4.2x\,\#i}$, $k_{\mathrm{on}}^{1x\,\#i}$, $k_{\mathrm{XL}} = k_{\mathrm{XL}}^{2.6mW\,\#i}$, $k_{\mathrm{XL}}^{1mW\,\#i}$ and $k_{\mathrm{diss.}} = k_{\mathrm{diss.}}^{\#i}$, where *#i* represents the crosslinking conditions. *ε* is the fitting error between observed and expected timecourses. *f(t, β)* represents the functional relationship between t, β and Y.

Equation 33, adapted to the different Dazl concentrations and different laser powers was used to fit the crosslinking timecourses for each binding site. The resulting equations represent the non-linear model:

For 4.2xDazl, 2.6 mW laser:

$$[DR^*]'_{(t)} = k_{\mathrm{XL}}^{2.6\mathrm{mW}\,\#1}[\frac{1}{k_{\mathrm{XL}}^{2.6\mathrm{mW}\,\#1}} - r_1(1-\frac{k_{\mathrm{XL}}^{2.6\mathrm{mW}\,\#1}}{k_{\mathrm{XL}}^{2.6\mathrm{mW}\,\#1}+k_{\mathrm{diss.}}^{\#1}})(1-e^{r_1 t})+ r_2(1-\frac{k_{\mathrm{on}}^{4.2\mathrm{x}\,\#1}}{k_{\mathrm{on}}^{4.2\mathrm{x}\,\#1}+k_{\mathrm{diss.}}^{\#1}})(1-e^{r_2 t})$$

(Eq.35)

For 4.2xDazl, 1 mW laser:

$$[DR^*]'_{(t)} = k_{\mathrm{XL}}^{1\mathrm{mW}\,\#2}[\frac{1}{k_{\mathrm{XL}}^{1\mathrm{mW}\,\#2}} - r_1(1-\frac{k_{\mathrm{XL}}^{1\mathrm{mW}\,\#2}}{k_{\mathrm{XL}}^{1\mathrm{mW}\,\#2}+k_{\mathrm{diss.}}^{\#2}})(1-e^{r_1 t})+ r_2(1-\frac{k_{\mathrm{on}}^{4.2\mathrm{x}\,\#2}}{k_{\mathrm{on}}^{4.2\mathrm{x}\,\#2}+k_{\mathrm{diss.}}^{\#2}})(1-e^{r_2 t})$$

(Eq.36)

For 1xDazl, 2.6 mW laser:

$$[DR^*]'_{(t)} = k_{\mathrm{XL}}^{2.6\mathrm{mW}\,\#3}[\frac{1}{k_{\mathrm{XL}}^{2.6\mathrm{mW}\,\#3}} - r_1(1-\frac{k_{\mathrm{XL}}^{2.6\mathrm{mW}\,\#3}}{k_{\mathrm{XL}}^{2.6\mathrm{mW}\,\#3}+k_{\mathrm{diss.}}^{\#3}})(1-e^{r_1 t})+ r_2(1-\frac{k_{\mathrm{on}}^{1\mathrm{x}\,\#3}}{k_{\mathrm{on}}^{1\mathrm{x}\,\#3}+k_{\mathrm{diss.}}^{\#3}})(1-e^{r_2 t})$$

(Eq.37)

For 1xDazl, 1 mW laser:

$$[DR^*]'_{(t)} = k_{\mathrm{XL}}^{1\mathrm{mW}\,\#4}[\frac{1}{k_{\mathrm{XL}}^{1\mathrm{mW}\,\#4}} - r_1(1-\frac{k_{\mathrm{XL}}^{1\mathrm{mW}\,\#4}}{k_{\mathrm{XL}}^{1\mathrm{mW}\,\#4}+k_{\mathrm{diss.}}^{\#4}})(1-e^{r_1 t})+ r_2(1-\frac{k_{\mathrm{on}}^{1\mathrm{x}\,\#4}}{k_{\mathrm{on}}^{1\mathrm{x}\,\#4}+k_{\mathrm{diss.}}^{\#4}})(1-e^{r_2 t})$$

(Eq.38)

r1 and r2 are:

$$r_1 = \frac{\left(k_{\mathrm{on}}^{h}+k_{\mathrm{diss.}}^{i}+k_{\mathrm{XL}}^{j}\right) + \sqrt{\left(k_{\mathrm{on}}^{h}+k_{\mathrm{diss.}}^{i}+k_{\mathrm{XL}}^{j}\right)^2 - 4\left(k_{\mathrm{XL}}^{j}k_{\mathrm{on}}^{h}\right)}}{2}$$

(Eq.39)

$$r_2 = \frac{\left(k_{\mathrm{on}}^{h}+k_{\mathrm{diss.}}^{i}+k_{\mathrm{XL}}^{j}\right) - \sqrt{\left(k_{\mathrm{on}}^{h}+k_{\mathrm{diss.}}^{i}+k_{\mathrm{XL}}^{j}\right)^2 - 4\left(k_{\mathrm{XL}}^{j}k_{\mathrm{on}}^{h}\right)}}{2}$$

(Eq.40)

h represents 4.2xDazl #1 (Eq. 35), 4.2xDazl #2 (Eq.36), 1xDazl #3 (Eq.37) and 1xDazl #4 (Eq.38). *i* represents #1 (Eq.35), #2 (Eq.36), #3 (Eq.37) and #4 (Eq.38). *j* represents 2.6 mW #1 (Eq.35), 1 mW #2 (Eq.36), 2.6 mW #3 (Eq.37) and 1 mW #4 (Eq.38).

Timecourses for 4.2xDazl at high laser (2.6 mW), 4.2xDazl at low laser (1mW), 1xDazl at high laser power (2.6 mW) and 1xDazl at low laser power (1mW) were separately fit to the non-linear model (**Supplementary Material Scheme 2**).

A matrix of initial parameters was obtained,

4.2xDazl: 2.6 mW laser	$k_{\text{on}}^{(4.2\text{xDazl})\ \#1}$	$k_{\text{diss.}}^{\#1}$	$k_{\text{XL}}^{(2.6\text{mW})\ \#1}$
4.2xDazl: 1 mW laser	$k_{\text{on}}^{(4.2\text{xDazl})\ \#2}$	$k_{\text{diss.}}^{\#2}$	$k_{\text{XL}}^{(1\text{mW})\ \#2}$
1xDazl: 2.6 mW laser	$k_{\text{on}}^{(1\text{xDazl})\ \#3}$	$k_{\text{diss.}}^{\#3}$	$k_{\text{XL}}^{(2.6\text{mW})\ \#3}$
1xDazl: 1 mW laser	$k_{\text{on}}^{(1\text{xDazl})\ \#4}$	$k_{\text{diss.}}^{\#4}$	$k_{\text{XL}}^{(1\text{mW})\ \#4}$

Next, a global datafit for all four timecourses (#1-4) for an individual binding site was performed. Initial parameters were iteratively adjusted, considering the following criteria:

$k_{\text{on}}^{(4.2\text{xDazl})\ \#1} \cong k_{\text{on}}^{(4.2\text{xDazl})\ \#2}$ (at different laser powers)

$k_{\text{on}}^{(1\text{xDazl})\ \#3} \cong k_{\text{on}}^{(1\text{xDazl})\ \#4}$ (at different laser powers)

$k_{\text{diss.}}^{\#1} \cong k_{\text{diss.}}^{\#2} \cong k_{\text{diss.}}^{\#3} \cong k_{\text{diss.}}^{\#4}$

$k_{\text{XL}}^{(2.6\text{mW})\ \#1} \cong k_{\text{XL}}^{(2.6\text{mW})\ \#3}$ (at 2.6 mW laser power)

$k_{\text{XL}}^{(1\text{mW})\ \#2} \cong k_{\text{XL}}^{(1\text{mW})\ \#4}$ (at 2.6 mW laser power)

Fits were repeated until the best fit was reached (no change in X^2 for 4 successive fittings), as measured by Chi-squared X^2 minimization, according to:

$$\chi^2 = \sum_{i=1}^{n}[\frac{Y_i - f(x_i', \beta)}{\sigma_i}]^2 \qquad \text{(Eq.41)}$$

x_i' is the row vector for the *ith* (*i* = 1, 2, ... , n; n = 10,341) observation. *β* is the parameter under consideration. Y_i is the estimated parameter value for the i_{th} (i = 1, 2, ... , n; n = 10,341) observation. σ_i is the variance between observed and estimated parameter values. $f(x_i', \beta)$ represents the function for which x_i' and *β* are measured.

Obtained parameters were further refined by additional rounds of fitting using the analytical, Levenberg-Marquardt (L-M) least squares algorithm, which combines the Gauss-Newton and the steepest descent method [57]. Utilizing the values obtained above, parameters for timecourses at 4.2xDazl at high laser power (2.6 mW) and low laser power (1mW) were adjusted together. $k_{\text{on}}^{(4.2\text{xDazl})\ \#2}$ was increased or decreased (depending on initial values for a given binding site) in small increments (∂b) in order to move $k_{\text{on}}^{(4.2\text{xDazl})\ \#2}$ closer to $k_{\text{on}}^{(4.2\text{xDazl})\ \#1}$. ∂b was set as 5% of $k_{\text{on}}^{(4.2\text{xDazl})\,\#2}$ for a given binding site. Following each increment, the timecourse was fitted to the non-linear model and X^2 calculated. $k_{\text{diss.}}^{\#2}$ and $k_{\text{XL}}^{(1\text{mW})\,\#2}$ were floated during the fitting. If $X^2\,(b + \partial b) \geq X^2\,(b)$ for >3 consecutive fitting cycles, $k_{\text{on}}^{(4.2\text{xDazl})\,\#1}$ was increased or decreased (depending on initial values) in small increments to improve fitting. This fitting procedure was repeated for N = 642 cycles.

Next, the parameters for timecourses at 1xDazl at high (2.6 mW) and low laser power (1mW) were adjusted, providing $k_{\text{on}}^{(4.2\text{xDazl})\,\#1}$, $k_{\text{on}}^{(4.2\text{xDazl})\,\#2}$, $k_{\text{on}}^{(1\text{xDazl})\,\#3}$ and $k_{\text{on}}^{(1\text{xDazl})\,\#4}$. Keeping the adjusted k_{on} constant (floating k_{XL}), were subsequently adjusted $k_{\text{diss.}}^{\#1}$, $k_{\text{diss.}}^{\#2}$, $k_{\text{diss.}}^{\#3}$ and $k_{\textit{diss.}}^{\#4}$ (within 25% range of each other). Finally, $k_{\text{XL}}^{(2.6\text{mW})\,\#1}$ and $k_{\text{XL}}^{(2.6\text{mW})\,\#3}$ were adjusted by increasing or decreasing $k_{\text{on}}^{(4.2\text{xDazl})\,\#1}$ and $k_{\text{on}}^{(4.2\text{xDazl})\,\#3}$ in small increments ($\partial b \leq 5\%$ of parameter values) while maintaining $k_{\text{on}}^{(4.2\text{xDazl})\,\#1} > k_{\text{on}}^{(4.2\text{xDazl})\,\#3}$. Additionally, $k_{\text{diss.}}^{\#1}$ and $k_{\text{diss.}}^{\#3}$ were increased or decreased in increments of $\partial b \leq 1\%$. The same process was performed for adjusting $k_{\text{on}}^{(4.2\text{xDazl})\,\#2}$ and $k_{\text{on}}^{(4.2\text{xDazl})\,\#4}$. Every parameter adjustment cycle was repeated 642 times after which X^2 values computed in 4 successive iterations showed fluctuations of less than 5% for > 95% of binding sites.

Calculation of binding probabilities.

The binding probability (P) describes the probability by which the accessible fraction of a given binding site is bound by Dazl. P for each Dazl concentration was calculated according to:

$$P_{(4.2\text{xDazl})} = \frac{k_{\text{on}}^{(4.2\text{xDazl})}}{k_{\text{diss.}} + k_{\text{on}}^{(4.2\text{xDazl})}} \qquad \text{(Eq.42)}$$

$$P_{(1\text{xDazl})} = \frac{k_{\text{on}}^{(1\text{xDazl})}}{k_{\text{diss.}} + k_{\text{on}}^{(1\text{xDazl})}} \qquad \text{(Eq.43)}$$

Calculation of fractional occupancy.

The fractional occupancy (Φ^{max}) describes the fraction of a given binding site that is occupied by Dazl extrapolated to saturating concentrations. Φ^{max} is a measure of binding site accessibility during the course of the experiment. $\Phi^{\text{max}} = 1$ indicates complete accessibility, decreasing values indicate decreasing accessibility. Φ^{max} was calculated by plotting the maximal amplitude (α^{max}: probability of Dazl bound to the fraction of a given binding site that is accessible during the course of the experiment, extrapolated to saturating concentrations of Dazl) vs. level of the corresponding transcript (L, in RPKM) (**Supplementary Material Figure S3**). Φ^{max} corresponds to the slope of the plots, and was calculated according to:

$$\Phi^{max} = \alpha^{max} \cdot L^{-1} \qquad \text{(Eq.44)}$$

Reported Φ^{max} values were normalized to a scale of zero to 1. To define α^{max}, apparent association rate constants at both Dazl concentrations $k_{on}^{(4.2xDazl)}$, $k_{on}^{(1xDazl)}$ were plotted against the relative cellular Dazl concentrations ($[Dazl]^{rel}$, **Supplementary Material Figure S3**).

For binding sites where $k_{on}^{(4.2xDazl)}$, $k_{on}^{(1xDazl)}$ increased linearly with $[Dazl]^{rel}$:

$$\alpha^{max} = \alpha^{(4.2xDazl)} \cdot (P_{(4.2xDazl)})^{-1} \qquad \text{(Eq.45)}$$

$\alpha^{(4.2xDazl)}$: normalized read density at the 30s time point for the timecourse with 4.2xDazl and 2.6 mW laser power for a given binding site, $P_{(4.2xDazl)}$: binding probability at 4.2xDazl (Eq.42).
For binding sites where $k_{on}^{(4.2xDazl)}$, $k_{on}^{(1xDazl)}$ increased with $[Dazl]^{rel}$ in a hyperbolic fashion, we determined the maximal apparent binding rate constant k_{on}^{max} by fitting the plot of $k_{on}^{(4.2xDazl)}$, $k_{on}^{(1xDazl)}$ vs. $[Dazl]^{rel}$ to:

$$k_{\text{on}}^{(\text{Dazl})} = k_{\text{on}}^{\text{max}} \times \frac{[Dazl]^{\text{rel}}}{[Dazl]^{\text{rel}}+K'} \qquad \text{(Eq.46)}$$

($k_{on}^{(Dazl)}$: $k_{on}^{(1xDazl)}$, $k_{on}^{(4.2xDazl)}$, K': apparent relative binding constant)
The binding probability extrapolated to $[Dazl]^{rel}$ saturation (P_{max}) is:

$$P_{\text{max}} = \frac{k_{\text{on}}^{\text{max}}}{k_{\text{diss.}}+k_{\text{on}}^{\text{max}}} \qquad \text{(Eq.47)}$$

and

$$\alpha_{max} = \alpha_{(4.2\text{xDazl})} \cdot \frac{P_{\text{max}}}{P_{(4.2\text{xDazl})}} \qquad \text{(Eq.48)}$$

A plot was defined as hyperbolic if $k_{on}^{max} < 4 \cdot k_{on}^{(4.2xDazl)}$.

Analysis of Variance (ANOVA):

One-way ANOVA was calculated in R using libraries – car [58]. Mean square differences between and within groups were calculated. Obtained F values were compared with the critical value in the F table to obtain p values [58]. Inter-group differences were significant ($p < 0.05$) when the F value exceeded the critical F value for the given degrees of freedom [59].

Determination of distances between neighboring binding sites.

Distances between neighboring binding sites (genomic coordinates: mm10) were calculated between first and last read coordinates of adjacent peaks recorded with a sliding window, (start: l = 0 (chr1), length = 2 nt, stride = 1 nt) for each transcript. The number of inter-site distances for a given value was divided by the overall number of distances to yield the normalized frequency (**Fig.3a**). The random distribution of inter-site distances was obtained by Monte Carlo simulations (**Fig.3a**). A random binding site was defined as a genomic coordinate encompassing a non-overlapping 5 nt long sequence (in the entire mouse transcriptome, **Fig.3a**) within 500 nt of PAS, or excluding 500 nt proximal to PAS, (**Extended Data Fig.5**). 10,341 binding sites were randomly distributed over these windows, their distribution was recorded and plotted as described above. Monte Carlo simulations (Vignette package in R [60] were carried out 1,000 times. Obtained distributions were averaged and plotted (**Fig.3a**).

Dazl cluster definition and distribution

A cluster of Dazl binding sites was defined by an inter-binding site distance of < 40 nt and absence of additional binding sites < 120 nt around the cluster. The distribution of clusters in 3'UTRs (**Fig.3b**) was calculated by dividing the 3'UTRs in 100 nt bins, starting at the PAS. The number of clusters in each bin was counted and the cumulative frequency of clusters with different numbers of binding sites was plotted against the 3'UTR bins.

Calculation of cumulative and differential binding probabilities.

Cumulative binding probabilities (ΣB) for each cluster of Dazl binding sites were calculated according to:

$$\Sigma B = \sum_{i=1}^{n}\left(\Phi^{\max(i)} \cdot \frac{k_{\mathrm{on}(i)}^{(4.2\mathrm{xDazl})}}{k_{\mathrm{on(i)}}^{(4.2\mathrm{xDazl})}+k_{\mathrm{diss.(i)}}}\right) = \sum_{i=1}^{n}\left(\Phi^{\max(i)} \cdot P_{(4.2\mathrm{xDazl})(\mathrm{i})}\right) \qquad \text{(Eq.49)}$$

[*n*: number of binding sites in a given cluster; i: individual binding site, $\Phi^{\max(i)}$: fractional occupancy for the binding site (i); $k_{\mathrm{on(i)}}^{(4.2\mathrm{xDazl})}$: association rate constant at 4.2xDazl for the binding site (i); $k_{\mathrm{diss.(i)}}$, dissociation rate constant for the binding site (i); $P_{(4.2\mathrm{xDazl})(\mathrm{i})}$: binding probability at 4.2xDazl) for the binding site (i)].

The differential cumulative binding probabilities (ΔΣB) for each cluster of Dazl binding sites were:

$$\Delta\Sigma B = \sum_{i=1}^{n} \Phi^{\max(i)} \cdot \left(\frac{k_{\mathrm{on(i)}}^{(4.2\mathrm{xDazl})}}{k_{\mathrm{on(i)}}^{(4.2\mathrm{xDazl})} + k_{\mathrm{diss.(i)}}} - \frac{k_{\mathrm{on(i)}}^{(1\mathrm{xDazl})}}{k_{\mathrm{on(i)}}^{(1\mathrm{xDazl})} + k_{\mathrm{diss.(i)}}} \right)$$

$$= \sum_{i=1}^{n} [\Phi^{\max(i)} \cdot \left(P_{(4.2\mathrm{xDazl})(\mathrm{i})} - P_{(1\mathrm{xDazl})(\mathrm{i})} \right)]$$

(Eq.50)

[Variables as above, $k_{\mathrm{on(i)}}^{(1\mathrm{xDazl})}$: association rate constant at 1xDazl for the binding site (i); $k_{\mathrm{diss.(i)}}$, dissociation rate constant for binding site (i); $P_{(1\mathrm{xDazl})(\mathrm{i})}$: binding probability at 1xDazl for binding site (i)].

Ribosome Profiling and RNA-seq

Ribosome profiling and RNA–seq, performed in biological triplicates at both Dazl concentrations was described [17]. Deposited sequencing data (GEO: GSE108997) were analyzed as described [17]. Averages from the triplicate datasets were used for subsequent data analysis.

Definition of functional mRNA classes

Changes in ribosome protected fragments (ΔRPF) from 4.2xDazl to 1xDazl (RPKM) and changes in transcript levels (ΔRNA) from 4.2xDazl to 1xDazl (RPKM) for each transcript with a Dazl binding site, represented in all ribosome profiling and RNA-seq datasets were plotted (**Fig.4b**). Low abundance transcripts ($\mathrm{RPKM}_{4.2\mathrm{xDazl}} < 6.0$) were removed. ΔRPF and ΔRNA distributions for Dazl bound transcripts were divided into terciles, based on testing the significance ($p < 0.05$) of the deviation from the mean (H = High; ΔRPF = 1.063, ΔRNA = 1.088, M = Medium; 1.063 ≤ ΔRPF ≤ 0.913, ΔRNA = 1.088 ≤ ΔRPF ≤ 0.974, L = Low; ΔRPF = 0.913, ΔRNA = 0.974). Terciles for ΔRPF and ΔRNA yield nine functional mRNA classes (**Fig.4b**). The HL and LH classes contained too few transcripts (< 10) for meaningful examination and were therefore not considered in subsequent analyses. The MM class was not further considered because neither ribosome occupancy nor transcript level changed significantly upon changes in Dazl concentration.

Enrichment Analysis

Statistical enrichment of clusters with high, medium and low cumulative binding probabilities (ΣB, **Fig.4a**) in transcripts belonging to each of the functional mRNA classesTHRH, THRM, TMRH, TMRL, TLRM and TLRL (**Fig.4c**), was calculated with the cumulative distribution function (CDF) of a hypergeometric distribution [61] according to:

$$p = \mathrm{F}(x|M,K,N) = \sum_{i=0}^{x} \frac{\frac{(K)(M-K)}{(i)\ (N-i)}}{\frac{(M)}{(N)}} \qquad \text{(Eq.51)}$$

(M: number of total clusters in Dazl bound transcripts, K: number of clusters in each functional mRNA class (THRH, THRM, TMRH, TMRL, TLRM and TLRL), N: number of clusters in a given ΣB tercile (H, M, L), i: number of clusters with a ΣB tercile in a given functional mRNA class (for example, number of clusters with high ΣB in THRH functional mRNA class). x represents a cluster and F (x|M,K,N) is enrichment of x given M, K and N (by Fishers' t-test represented as F). p is theLL hypergeometric p value of enrichment, based on the F-test [61]) Hypergeometric tests were performed with *Scipy* hypergeom module [62] in Python 3.6.5.

PCA and t-SNE*.*

A data matrix (X) with the seven features of Dazl clusters and of transcripts with Dazl binding sites in 3'UTR (number of clusters in 3'UTR, ΣB, ΔΣB, number of binding sites in a cluster, UTR length, proximity to PAS, transcript level), corresponding to each transcript, was generated. In transcripts with multiple clusters in the 3'UTR, ΣB, ΔΣB and number of binding sites in a cluster represent values of the cluster closest to the PAS. Proximity to PAS in transcripts of multiple clusters represents the median pattern for the clusters (for example, in a UTR with 5 clusters, 4 of which distant to the PAS, the median was considered distant to the PAS). The empirical mean for each column of the data matrix was calculated (sample mean of each column, shifted to zero to center data). Data were centered and scaled and a covariance matrix for the seven features was calculated (**Extended Fig. 8a**). This covariance matrix was used to calculate eigenvectors and eigenvalues, as described [63]. Eigenvalues were sorted in descending order and *K* largest eigenvalues were selected. *K* is the desired number of dimensions (Principal Components) of a new feature subspace Y with K ≤ n (*K* = 2 for **Extended Fig.8c** and *K* = 3 for **Extended Fig.8e**). A projection matrix (W) was created from the selected (*K*) eigenvalues through orthogonal transformation of the original dataset (*X*) in order to obtain a *K*-dimensional feature subspace Y. Proportion of variance, cumulative variance, factor loadings and eigenvalues explained by each component were recorded (**Supplementary Material Table S11**). Functional mRNA classes (**Extended Fig.8c**) and Dazl code groups (1 - 21, **Extended Fig.8e**) were identified and mapped onto the feature space (*Y*) by k-means clustering [64]. PCA was conducted in R using the *prcomp()* function. To visualize subgrouping within functional mRNA classes (**Extended Data Fig.8d**), the Barnes-Hut t-SNE implementation in R [65] was used with the recommended parameters (perplexity 5 - 30, iterations 5 - 3000) as described [66].

Derivation of the Dazl regulatory program.

Seven features of Dazl clusters and of transcripts with Dazl binding sites in 3'UTR (number of clusters in 3'UTR, ΣB, ΔΣB, number of binding sites in a cluster, UTR length, proximity to PAS, transcript level) were utilized to further group transcripts in each functional mRNA class (**Fig.4d**). In transcripts with multiple clusters in the 3'UTR, ΣB, ΔΣB and number of binding sites in a cluster represent values of the cluster closest to the PAS. Proximity to PAS in transcripts of multiple clusters represents the median pattern for the clusters (for example, in a UTR with 5 clusters, 4 of which distant to the PAS, the median was considered distant to the PAS). PCA and t-SNE independently identified 21 groups (1-21) in the 6 functional mRNA classes (**Extended Data Figs.7,8**). To create the Dazl code from identified groups 1-21, we first defined terciles (High, Median, Low) for each of the 7 features of Dazl binding patterns (number of clusters in 3'UTR, ΣB, ΔΣB, number of binding sites in a cluster, UTR length, proximity to PAS, transcript level) on the basis of significance testing ($p < 0.05$) for the deviation from the mean. The number of clusters of each tercile type (H, M or L) for each of the 7 features was then counted in each group. This yielded a data matrix with count of feature tercile (example: [group 1; ΣB]; H = 2, M = 27, L = 8, Total = 37 Clusters, **Extended Data Fig.8f**). The tercile count per feature (per group) was then normalized to total number of clusters in the group to obtain fraction of each feature tercile in a group (example: [group 1; ΣB]; H = 0.05, M = 0.73, L = 0.22, Total = 37 Clusters). For every group, the tercile for a feature that encompassed >50% of the clusters was utilized as the code for that group **(Extended Data Fig.8f)**.

Multiple Linear Regression Analysis

Multiple linear regression (MLR) analysis was performed with "dummy coding", e.g. transformation of categorical independent variables into dichotomous variables [67]. The dependent variables, ΔRPF and ΔRNA, were used as continuous data, either separately or merged (**Extended Data Fig.10**). 45 models were formulated describing Dazl binding and corresponding mRNA characteristics for various combinations of “dummy coded” independent variables, “continuous” independent variables, “continuous” dependent variables (separate ΔRPF and ΔRNA) and “merged” dependent variables (**Extended Data Fig.10**). Models were progressively shortlisted and the best performing model (M1) was selected after 4 steps.

Step1. We utilized the best subsets regression procedure (Ref1) to identify all possible model permutations of parameters (N = 45) that satisfied the following criterion:

1. Models contain n ≥ 3 independent variables
2. Models account for Dazl kinetics and binding pattern along with RNA features.
3. Selected independent variables do not show multi collinearity (assessed by pairwise correlation).

The data was randomly divided into training (70%, N = 699) and test set (30%, N = 492). The training set was utilized to evaluate, estimate and identify the optimal models and cross-validation was performed using the test set. Each model was regressed on associated independent variables and adjusted R^2 and root mean standard errors (RMSE) were calculated according to:

$$Adjusted\ R^2 = 1 - \left(\frac{n-1}{n-(k+1)}\right)(1-R^2) \qquad \text{(Eq.52)}$$

(n = 699, number of observations; k=7: number of independent variable terms). The root mean standard error (i.e. estimated standard deviation; σ^2 of the error term u) was obtained as:

$$RMSE = \sqrt{\frac{SSE}{n-(k+1)}} \qquad \text{(Eq.53)}$$

(n = 699, number of observations; k=7: number of independent variable terms; SSE: sum of squares error, difference between observed and predicted value).

As expected, the adjusted R^2 showed inverse correlation with RMSE.

We selected the models with the highest adjusted R^2 (≥ 0.5) and lowest root mean standard errors (RMSE; top 50%). We also examined models with R^2 ≥ 0.5 despite low adjusted R^2 , high RMSE according to:

$$R^2 = \frac{SSR}{SSTO} = 1 - \frac{SSE}{SSTO} \qquad \text{(Eq.54)}$$

SSR (sum of squares due to regression; the sum of the differences between the predicted value and the mean of the dependent variable, measures unexplained variance) is equivalent to the distance from each point to the regression line. SSR was calculated according to:

$$SSR = \sum_i (y_i - y')^2$$

(Eq.55)

(y_i = predicted value; y' = mean)

SSTO (sample variance) was calculated according to:

$$SSTO = \sum_i (x_i - y')^2$$

(Eq.56)

(y_i = observed value; y' = mean)

With this approach, we shortlisted 24 models with according to adjusted R^2, RMSE and R^2 values (**Supplementary Material Fig.S4**).

We next determined information criterion statistics (ICS) for these models. ICS combines the SSE, number of parameters in the model, and sample size. We utilized three established information criterion parameters [68]: Akaike's Information Criterion (AIC), the Bayesian Information Criterion (BIC) and Amemiya's Prediction Criterion (APC), which were calculated according to:

$$AIC_k = n\ln(SSE) - n\ln(n) + 2(k+1) \qquad \text{(Eq.57)}$$

$$BIC_k = n\ln(SSE) - n\ln(n) + (k+1)\ln(n) \qquad \text{(Eq.58)}$$

$$APC_k = \frac{(n+k+1)}{n(n-k-1)} SSE \qquad \text{(Eq.59)}$$

(n: sample size, k: number of predictor terms, e. g. k+1 = number of regression parameters in the model, including the intercept). We compared all 24 models and ranked the models according values for AIC, BIC and APC (lowest value – highest rank). At this stage, no model was removed.

Step 2. Further shortlisting was performed by comparing information criteria with model fitness parameters. To determine the fitness of the shortlisted models, two different hypothesis tests for slopes were conducted. We first tested the hypothesis that at least one slope parameter is 0:

$$H_0: \beta_1 = \beta_2 = \beta_{(n\ldots)} = 0$$

$$H_\alpha: At\ least\ one\ \beta_i \neq 0\ (for\ i = 1, 2, n \ldots.)\ where\ \alpha = 0.05 \qquad \text{(Eq.60)}$$

using the general linear F test (ANOVA F statistic) by obtaining error sum of squares (the squared distances between the observed and predicted responses) for full (with all independent variables) and reduced models (with intercept only). p values were computed.

We next tested the hypothesis that only one of the slope parameters is 0:

$$H_0: \beta_1 = 0$$

$$H_\alpha: \beta_1 \neq 0\ where\ \alpha = 0.05 \qquad \text{(Eq.61)}$$

using t-test statistics for each independent variable in the model. p values were computed.

Next, we compared information criterion parameters (AIC, BIC and APC), general linear F statistic and t-test statistic values for all 24 models. We shortlisted the models with the lowest AIC, BIC and APC values, most significant general linear F statistic and significant t-test statistic for all associated independent variables were shortlisted (**Supplementary Material Fig.S5**). All models satisfied the general linear F statistic condition, indicating that addition of selected independent variables (i.e. features) increased the explanatory power of the models. 13 out of 24 models had significantly lower information criterion parameters (**Supplementary Material Fig.S5)**. We further assessed these 13 models according to obtained coefficients, standard errors, t-statistic, p-value and confidence intervals for all the independent variables. 6 out of 13 models showed significant t-statistics (p-values) for all coefficient terms and the smallest confidence interval ranges (**Supplementary Material Fig.S5**).

Step 3. To estimate the quality of the remaining 6 models, we tested 4 multiple linear regression conditions (LINE conditions):

1. The mean of the response, $E(Y_i)$, at each set of values of predictors, $(x_{1i}, x_{2i}, x_{(n)i})$ is a Linear function of the predictors.
2. The errors, ε_i, are Independent.
3. The errors, ε_i, at each set of values of the predictors are Normally distributed.
4. The errors, ε_i, at each set of values of predictors have Equal variance (σ^2).

To visually validate the LINE conditions (assessment of the distribution of errors), we recorded residuals vs. predicted values, and plotted a histogram of residuals for each model (**Supplementary Material Fig.S6)**. We also performed the Kolmogorov-Smirnov Test (K-S test) for all 6 models [69]. Three models, M1, M19 and M24 showed normal distribution of error residuals, absence of outliers and equal variance and hence were selected for cross-validation (**Supplementary Material Fig.S6**).

Step 4. These three models were validated using the test dataset (N = 492) and model M1 was identified as the optimal model on the basis of smallest Mean Squared Prediction Error value (MSPE) (**Extended Data Figure 10e, Supplementary Material Fig.S7**). This model (M1) consisted of seven independent variables: number of clusters in 3'UTR, ΣB, ΔΣB, number of binding sites in a cluster, UTR length, proximity to PAS, transcript level all expressed as dummy coded variables in terciles of their respective distributions.

Multiple regression on a training data set of N = 699 was performed according to:

$$Y_i' = b_0 + b_1X_{1i} + b_2X_{2i} + b_3X_{3i} + b_4X_{4i} + b_5X_{5i} + b_6X_{6i} + +b_7X_{7i} + u \qquad \text{(Eq.62)}$$

(Y': predicted dependent, continuous variable (ΔRPF and ΔRNA) or predicted dependent, merged continuous variable, $b_{(i=0...7)}$: differential intercept linear coefficients, X($_{n)i}$: independent variables, u: error term). The differential intercept linear coefficients (DILC) associated with each dummy coded/continuous independent variable terms are the expected difference in the mean of the outcome for that variable, compared to the reference group (TMRM class), with all other predictors constant [67,69]. The "b_n" values represent regression weights that were computed by minimization of the sum of squared deviations:

$$\sum_{i=1}^{n}(Y_{i-} Y_i')^2 \qquad \text{(Eq.63)}$$

(n = 699, sample size of training data set, Y_i : observed value for the dependent variable ΔRPF and ΔRNA). The optimal regression model was:

$$\Delta\text{RPF} = 1.01 + (cluster)\ \genfrac{}{}{0pt}{}{+0.02^{Hi}}{-0.03_{Lo}} + (\text{bind. prob.})\ \genfrac{}{}{0pt}{}{+0.03^{Hi}}{-0.02_{\ Lo}} + (\mathit{\Delta}\text{bind. prob.})\ \genfrac{}{}{0pt}{}{+0.11^{Hi}}{+0.05_{Lo}} +$$

$$(\text{\#bind. sites})\ \genfrac{}{}{0pt}{}{+0.03^{Hi}}{-0.15_{\ Lo}} + (\text{dist. PAS})\ \genfrac{}{}{0pt}{}{-0.005^{Hi}}{+0.01_{\ Lo}} + (UTR\ len)^{+0.06}_{+0.03} + (\text{RPKM})(\text{-}0.00004) + 0.07$$

$$\Delta\text{RNA} = 1.01 + (cluster)\ \genfrac{}{}{0pt}{}{+0.03^{Hi}}{-0.003_{Lo}} + (\text{bind. prob.})\ \genfrac{}{}{0pt}{}{+0.05^{Hi}}{-0.02_{\ Lo}} + (\mathit{\Delta}\text{bind. prob.})\ \genfrac{}{}{0pt}{}{+0.03^{Hi}}{+0.01_{\ Lo}} +$$

$$(\text{\#bind. sites})\ \genfrac{}{}{0pt}{}{+0.04^{Hi}}{-0.11_{\ Lo}} + (\text{dist. PAS})\ \genfrac{}{}{0pt}{}{+0.01^{Hi}}{+0.007_{\ Lo}} + (UTR\ len)^{+0.07}_{-0.03} + (\text{RPKM})(\text{-}0.000007) + 0.06$$

(Eq.64)

$\left(\mathit{\Delta RPF} = \frac{RPF\ at\ high\ Dazl}{RPF\ at\ low\ Dazl};\ \mathit{\Delta RNA} = \frac{RNA\ at\ high\ Dazl}{RNA\ at\ low\ Dazl}\right).$

The model was evaluated on a test data set (N = 492, 30% of the data; **Fig.4**). Regression analysis was performed using Scikit-learn [71] and Statsmodels [72] modules in Python 3.6.5.

Decision Tree Classifier

We employed a Chi-squared Automatic Interaction Detection (CHAID) algorithm, which makes no assumption about underlying data [73,74], in order to determine how categorical independent variables (seven transcript and cluster features, above) best combine to predict the functional mRNA classes. A data matrix was formed using classes of Y (transcript and cluster features) as columns and categories of the predictor X (functional mRNA classes) as rows. The expected cell frequencies under the null hypothesis were estimated as described [73]. The observed cell frequencies and the expected cell frequencies were then used to calculate Pearson chi-squared statistic, according to:

$$\chi^2 = \sum_{J=1}^{J}\sum_{I=1}^{I}\frac{(n_{IJ}-\acute{m}_{IJ})^2}{\acute{m}_{IJ}} \qquad \text{(Eq.65)}$$

(n_{IJ} is the observed cell frequency for cell $(x_n = I \mid y_n = j)$. m_{IJ} is the estimated expected cell frequency for cell $(x_n = I \mid y_n = j)$ from independence model [73,74].

The p value is:

$$p = \Pr(\chi_D^2 > \chi^2) \qquad \text{(Eq.66)}$$

X_D^2 follows a Chi-squared distribution with degrees of freedom $d = (J - 1)\ (I - 1)$

Pr: probability. The adjusted p-value is calculated as Bonferroni multiplier [75].

CHAID analysis was performed using CHAID 5.3.0 (ref.[76]) in Python 3.6.5.

Gene Ontology Analysis

GO term analyses for transcripts in groups 1-21 (**Fig.4d**) was performed with REACTOME (refs. [77,78]) using a hypergeometric statistical test and Benjamini and Hochberg FDR correction (significance level of 0.05) to identify enriched terms after multiple testing correction [79]. Redundant GO terms were merged to create a parent term. Transcripts for each Dazl group (1-21) were clustered using Ward's minimum variance method in R [80] and plotted as a heatmap using ggplot2 [48] (**Fig.4d**).

Pathway Analysis

Pathways (**Extended Data Fig.8h**) were obtained from REACTOME [77,78]. mRNA classes were mapped on pathways with Cytoscape [81].

Luciferase Reporter Measurements

Luciferase reporters were generated as previously described [17]. Briefly, DAZL target 3'UTRs with at least 100 nt of downstream sequence were cloned into the pRL-TK vector (Promega), replacing the SV40 late poly(A) region. Transfections and luciferase assays were also performed as previously described [17]. GC-1 spg cells were induced with doxycycline as described above. After 24 hours, pRL-TK 3′UTR reporters and pGL4.54[luc2/TK] (Promega) firefly luciferase control plasmids were transfected into GC-1 spg cells using Lipofectamine 2000 (Thermofisher). The media was replaced after 4-6 hours and cells were harvested after 24 hours. Dual luciferase assays were performed using the Dual-Luciferase Reporter Assay System (Promega) according to manufacturer's instructions. Renilla luciferase levels were normalized to firefly luciferase activity.

ADDITIONAL REFERENCES

28 Weyn-Vanhentenryck, S. M. *et al.* HITS-CLIP and integrative modeling define the Rbfox splicing-regulatory network linked to brain development and autism. *Cell Rep* **6**, 1139-1152 (2014).

29 Zhang, C. & Darnell, R. B. Mapping in vivo protein-RNA interactions at single-nucleotide resolution from HITS-CLIP data. *Nat Biotechnol* **29**, 607-614 (2011).

30 Aken, B. L. *et al.* Ensembl 2017. *Nucleic Acids Res* **45**, D635-D642 (2017).

31 O'Leary, N. A. *et al.* Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. *Nucleic Acids Res* **44**, D733-745 (2016).

32 Magdison, J. Common Pitfalls in Causal Analysis of Categorical Data. *Journal of Marketing Research* **19**, 461 - 471 (1982).

33 Breiman, L., Friedman, J. H., Olshen, R. A. & Stone, C. J. *Classification and Regression Trees.* (Chapman & Hall/CRC, 1984).

34 Blake, C. L., Keogh, E. & Merz, C. J. UCI repository of machine learning databases. . (1998). <http://www.ics.uci.edu/~mlearn/MLRepository.html>.

35 Kass, G. V. An exploratory technique for investigating large quantities for categorical data. *Applied Statistics* **20**, 119 - 127 (1980).

36 Brister, M. M. & Crespo-Hernandez, C. E. Direct Observation of Triplet-State Population Dynamics in the RNA Uracil Derivative 1-Cyclohexyluracil. *J Phys Chem Lett* **6**, 4404-4409 (2015).

37 Brister, M. M. & Crespo-Hernandez, C. E. Excited-State Dynamics in the RNA Nucleotide Uridine 5'-Monophosphate Investigated Using Femtosecond Broadband Transient Absorption Spectroscopy. *J Phys Chem Lett* **10**, 2156-2161 (2019).

38 Paschotta, R. *Encyclopedia of Laser Physics and Technology* (Wiley-VCH, 2008).

39 Strober, W. Trypan blue exclusion test of cell viability. *Curr Protoc Immunol* **Appendix 3**, Appendix 3B, doi:10.1002/0471142735.ima03bs21 (2001).

40 Moore, M. J. *et al.* Mapping Argonaute and conventional RNA-binding protein interactions with RNA at single-nucleotide resolution using HITS-CLIP and CIMS analysis. *Nat Protoc* **9**, 263-293 (2014).

41 Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9**, 357-359 (2012).

42 Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841-842 (2010).

43 Robinson, J. T. *et al.* Integrative genomics viewer. *Nat Biotechnol* **29**, 24-26 (2011).

44 <http://emboss.bioinformatics.nl/cgi-bin/emboss/help/compseq >

45 Schindler, D. *Randomize R*, <https://cran.rproject.org/web/packages/randomizeR/randomizeR.pdf> (2019).

46 <https://docs.python.org/3/library/random.html >

47 Soetaert, K. *deSolve: Solvers for Initial Value Problems of Differential Equations* <https://cran.r-project.org/web/packages/deSolve/index.html>

48 <https://www.r-graph-gallery.com/79-levelplot-with-ggplot2.html>

49 <https://www.rdocumentation.org/packages/reshape2/versions/1.4.3>

50 <https://rmarkdown.rstudio.com/>

51 Schön, T. B., Wills, A. & Ninness, B. System identification of nonlinear state-space models. *Automatica* **47**, 39-49 (2011).

52 Ross, S. M. *Introductory Statistics.* (Elsevier Inc., 2010).

53 Boyce, W. E. & DiPrima, R. C. *Elementary differential equations and boundary value problems.* (Wiley, 2012).

54 Polyanin, A. D. & Zaitsev, V. F. *Handbook of Ordinary Differential Equations: Exact Solutions, Methods, and Problems.* (CRC Press, 2017).

55 Edwards, C. H. & Penney, D. E. *Differential Equations: Computing and Modeling.*, 156–170 (Pearson Education).

56 Kreyszig, E. *Advanced Engineering Mathematics* (Wiley, 1972).

57 Gill, P. E. & Murray, W. Algorithms for the solution of the nonlinear least-squares problem. *SIAM Journal on Numerical Analysis* **15**, 977–979 (1978).

58 Fox, J. *Car: Companion to Applied Regression* <https://cran.r-project.org/web/packages/car/index.html> (2020).

59 Thompson, H. W., Mera, R. & Prasad, C. The Analysis of Variance (ANOVA). *Nutr Neurosci* **2**, 43-55 (1999).

60 <https://cran.rproject.org/web/packages/MonteCarlo/vignettes/MonteCarlo-Vignette.html>

61 Cao, J. & Zhang, S. A Bayesian extension of the hypergeometric test for functional enrichment analysis. *Biometrics* **70**, 84-94 (2014).

62 <https://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.hypergeom.html>

63 Jolliffe, I. T. & Cadima, J. Principal component analysis: a review and recent developments. *Philos Trans A Math Phys Eng Sci* **374**, 20150202 (2016).

64 Kerr, G., Ruskin, H. J., Crane, M. & Doolan, P. Techniques for clustering gene expression data. *Comput Biol Med* **38**, 283-293 (2008).

65 Krijthe, J. <https://cran.r-project.org/web/packages/Rtsne/index.html> (2018).

66 van der Maaten, L. Visualizing Data using t-SNE. *J Mach Learn Res* **9**, 2579-2605 (2008).

67 Draper, N. R. & Smith, H. *Applied Regression Analysis*. (Wiley & Sons, Inc., 1998).

68 Zhang Z. (2016). Variable selection with stepwise and best subset approaches. *Ann Transl Med 4*(7), 136

69 Mishra P, Pandey CM, Singh U, Gupta A, Sahu C, Keshri A. Descriptive statistics and normality tests for statistical data. *Ann Card Anaesth* 22, 67-72 (2019)

70 Krzywinski, M. & Altman, N. Multiple linear regression. *Nat Methods* **12**, 1103-1104 (2015).

71 <https://scikit-learn.org/stable/>

72 <https://www.statsmodels.org/stable/index.html>

73 Bigss, D., Ville, B. & Suen, E. A Method of Choosing Multiway Partitions for Classification and Decision Trees. *J Appl Stat* **18**, 49-62 (1991).

74 Goodman, L. A. Simple Models for the Analysis of Association in CrossClassifications Having Ordered Categories. *J Am Stat Assoc* **74**, 537-552 (1979).

75 Armstrong, R. A. When to use the Bonferroni correction. *Ophthalmic Physiol Opt* **34**, 502-508, doi:10.1111/opo.12131 (2014).

76 <https://pypi.org/project/CHAID/>

77 Fabregat, A. *et al.* Reactome pathway analysis: a high-performance in-memory approach. *BMC Bioinformatics* **18**, 142 (2017).

78 Jassal, B. *et al.* The reactome pathway knowledgebase. *Nucleic Acids Res* **48**, D498-D503 (2020).

79 Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J Royal Stat Soc, Series B* **57**, 289–300 (1995).

80 *Ward's minimum variance method* <https://uc-r.github.io/hc_clustering>

81 Shannon, P. *et al.* Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res* **13**, 2498-2504 (2003).

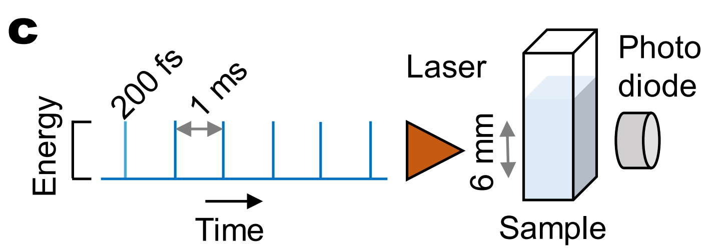
a

$$\text{Protein + RNA} \underset{k_{off}}{\overset{k_{on}}{\rightleftarrows}} [\text{Protein-RNA}] \xrightarrow{k_{XL}} \underset{\textit{crosslinked}}{[\text{Protein-RNA}]^*}$$

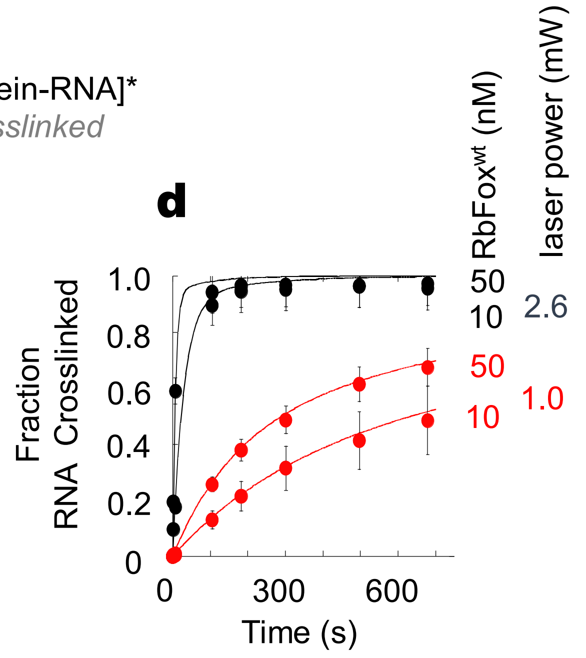
b

	Timecourses			
	1	2	3	4
Protein concentration	High	High	Low	Low
Crosslinking Efficiency	High	Low	High	Low

c



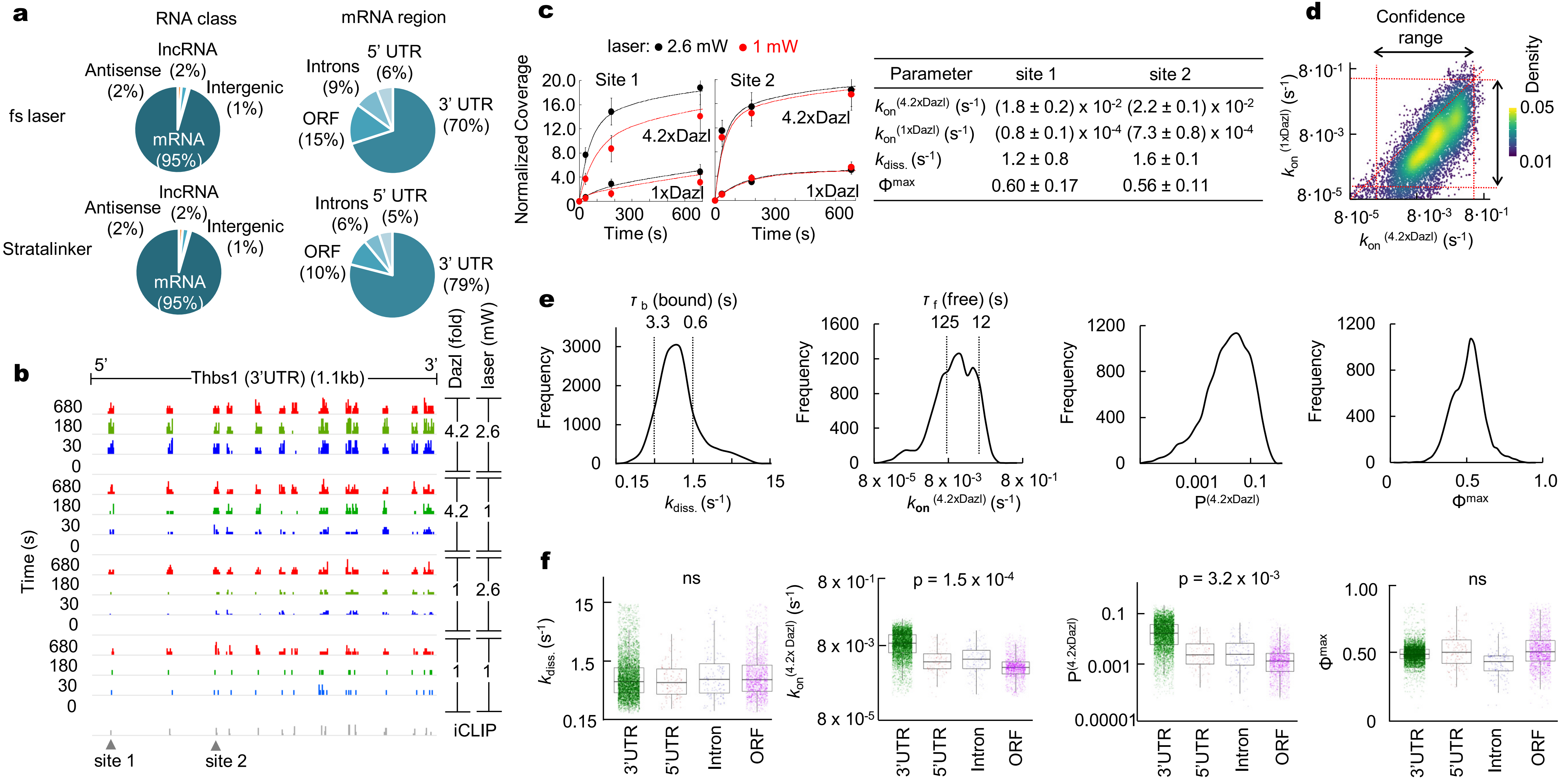
d



e

	$k_{XL}^{(2.6mW)}$	$k_{XL}^{(1mW)}$	k_{on} (10^6 $M^{-1}s^{-1}$)	k_{off} (s^{-1})	$K_{1/2}$ (nM)	
	fs Laser				Anisotropy	fs Laser
RbFox^{wt}(RRM)	2.4 ± 0.4	0.06 ± 0.02	8.2 ± 1.8	0.07 ± 0.02	3.0 ± 1.5	8.5 ± 3.1
RbFox^{mut}(RRM)	2.6 ± 0.5	0.05 ± 0.01	4.6 ± 2.0	0.11 ± 0.05	16 ± 5	23 ± 6
Dazl(RRM)	2.5 ± 0.3	0.06 ± 0.01	3.8 ± 1.2	0.49 ± 0.12	74 ± 7	128 ± 32

FIGURE 1



Parameter	site 1	site 2
$k_{on}^{(4.2xDazl)}$ (s^{-1})	(1.8 ± 0.2) x 10^{-2}	(2.2 ± 0.1) x 10^{-2}
$k_{on}^{(1xDazl)}$ (s^{-1})	(0.8 ± 0.1) x 10^{-4}	(7.3 ± 0.8) x 10^{-4}
$k_{diss.}$ (s^{-1})	1.2 ± 0.8	1.6 ± 0.1
Φ^{max}	0.60 ± 0.17	0.56 ± 0.11

FIGURE 2

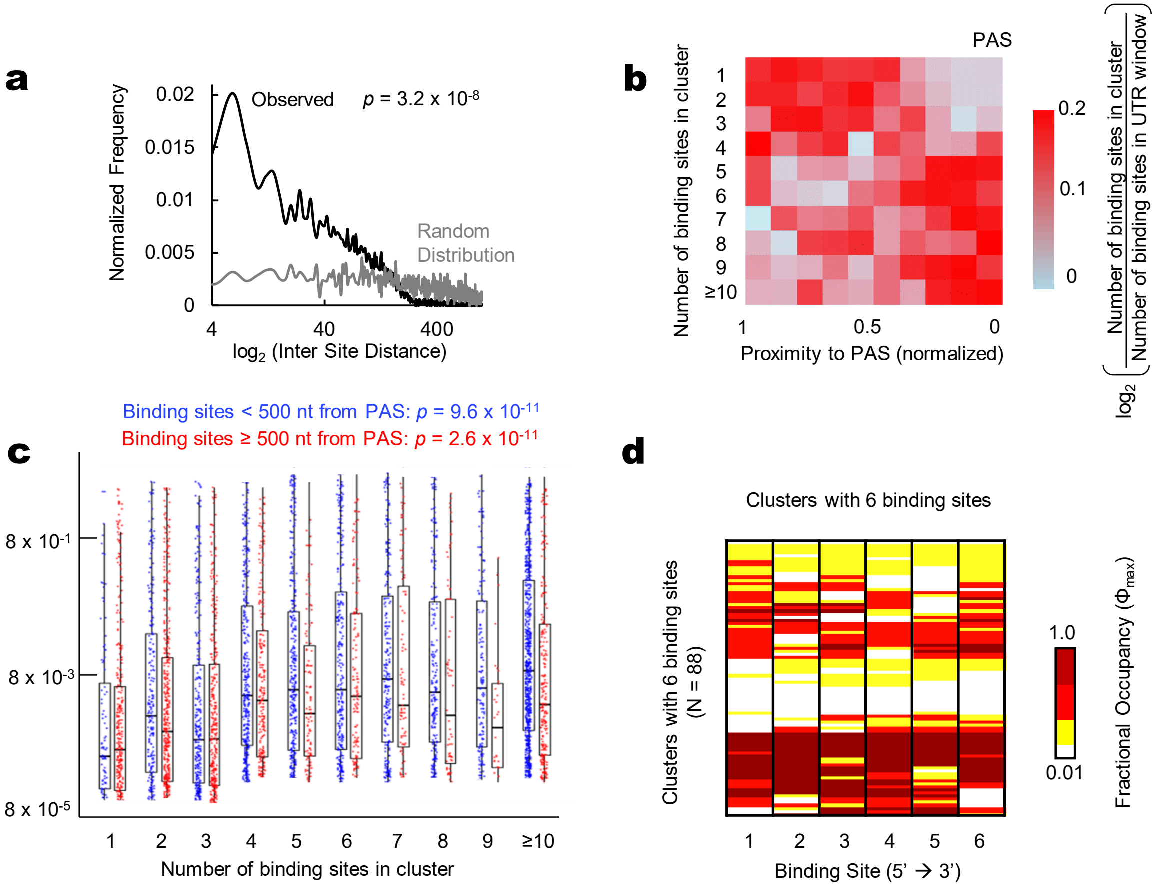

a
Normalized Frequency
0.02
0.015
0.01
0.005
0
Observed
p = 3.2 x 10^{-8}
Random Distribution
4
40
400
log$_2$ (Inter Site Distance)
b
PAS
Number of binding sites in cluster
1
2
3
4
5
6
7
8
9
≥10
1
0.5
0
Proximity to PAS (normalized)
0.2
0.1
0
log$_2$ (Number of binding sites in cluster / Number of binding sites in UTR window)
Binding sites < 500 nt from PAS: p = 9.6 x 10^{-11}
Binding sites ≥ 500 nt from PAS: p = 2.6 x 10^{-11}
c
$k_{on}^{(4.2xDazl)}$ (s^{-1})
8 x 10^{-1}
8 x 10^{-3}
8 x 10^{-5}
1
2
3
4
5
6
7
8
9
≥10
Number of binding sites in cluster
d
Clusters with 6 binding sites
Clusters with 6 binding sites (N = 88)
1.0
0.01
Fractional Occupancy (Φ_{max})
1
2
3
4
5
6
Binding Site (5' → 3')

FIGURE 3

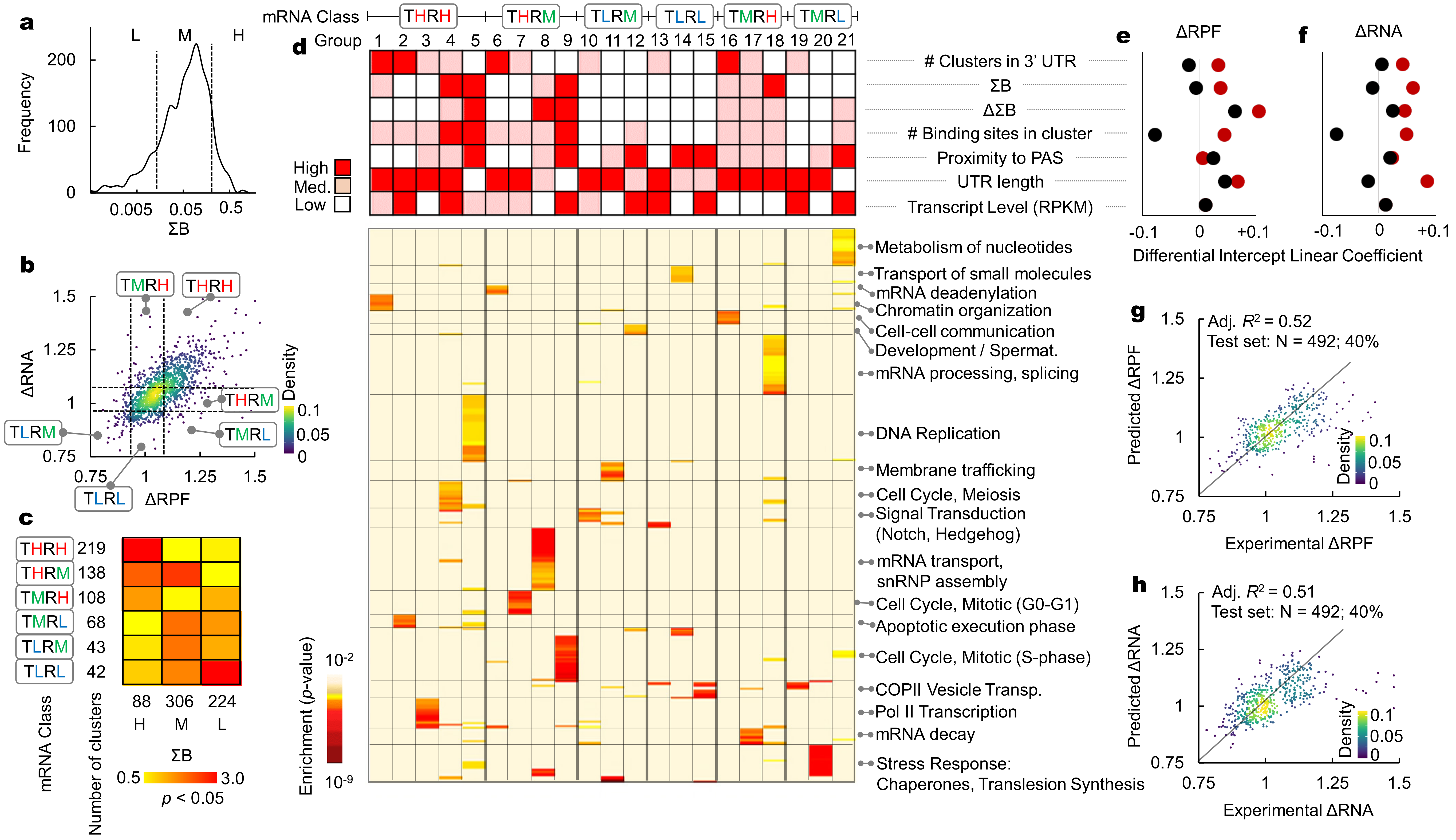
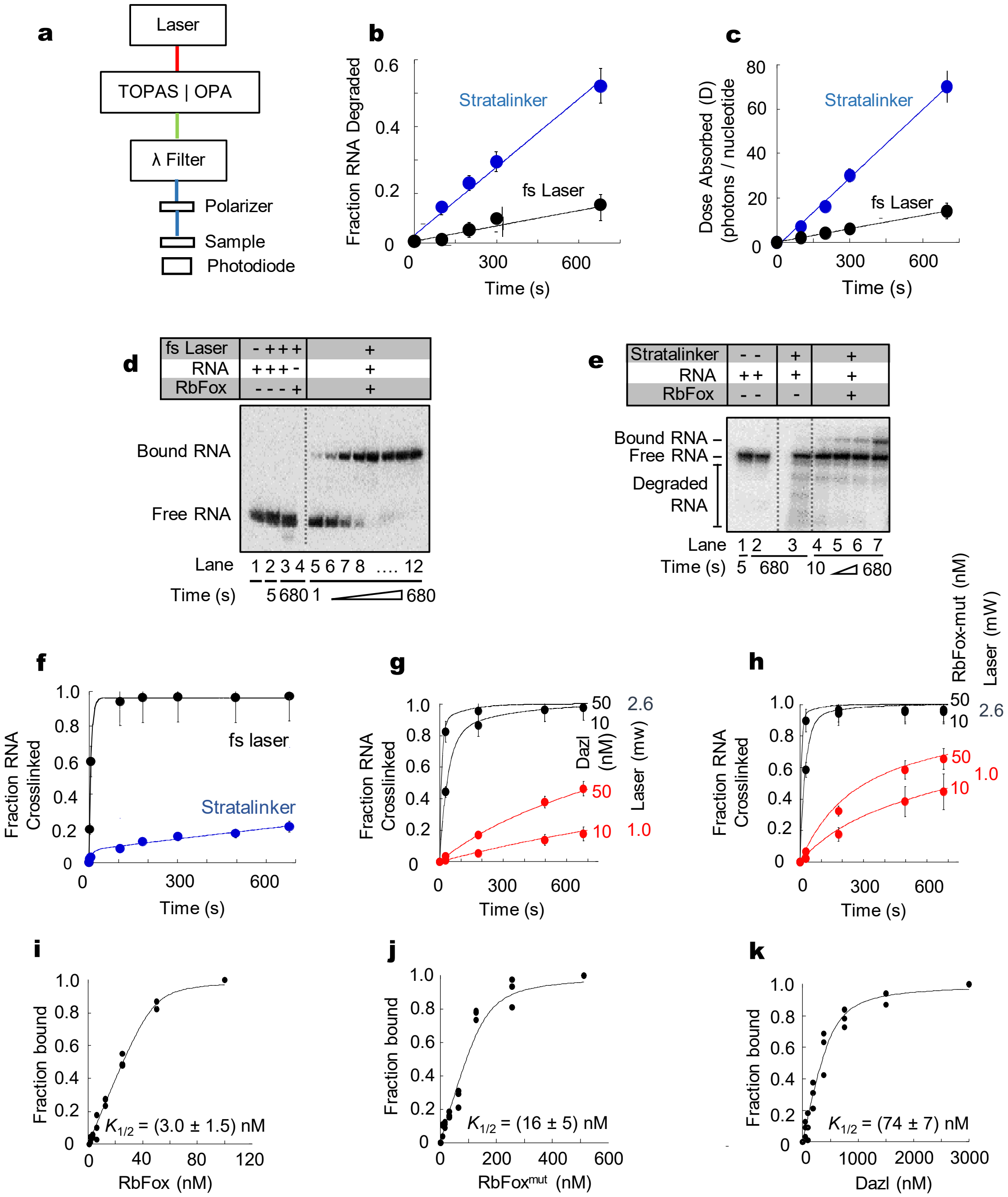
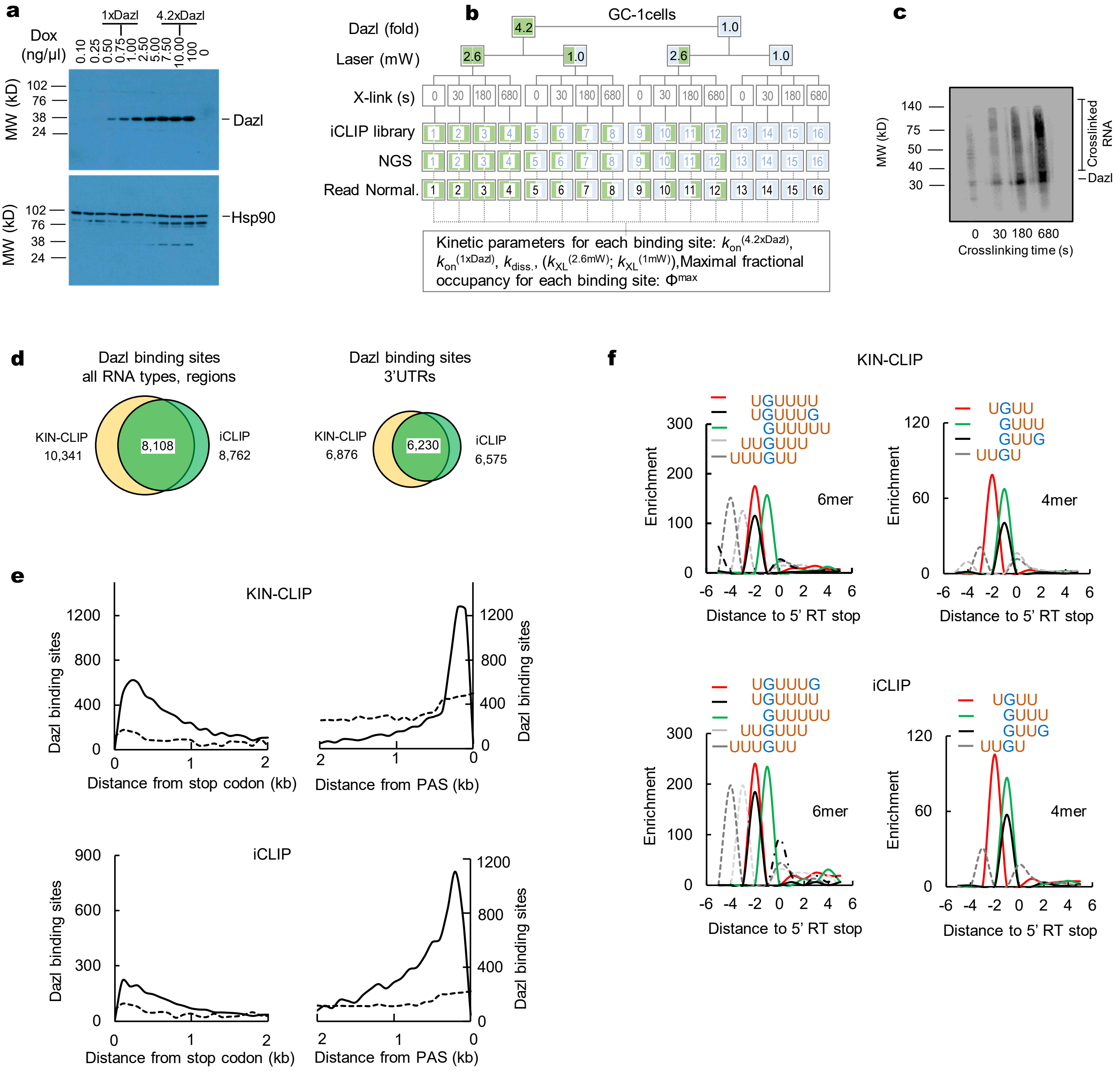

a
L
M
H
Frequency
200
100
0
0.005
0.05
0.5
ΣB
b
TMRH
THRH
1.5
1.25
1
0.75
ΔRNA
THRM
TLRM
TMRL
Density
0.1
0.05
0
0.75
1
1.25
1.5
TLRL
ΔRPF
c
THRH 219
THRM 138
TMRH 108
TMRL 68
TLRM 43
TLRL 42
88 306 224
H M L
ΣB
0.5
3.0
$p < 0.05$
mRNA Class
Number of clusters
mRNA Class
THRH
THRM
TLRM
TLRL
TMRH
TMRL
d
Group
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21
High
Med.
Low
# Clusters in 3' UTR
ΣB
ΔΣB
# Binding sites in cluster
Proximity to PAS
UTR length
Transcript Level (RPKM)
Enrichment (p-value)
10^{-2}
10^{-9}
Metabolism of nucleotides
Transport of small molecules
mRNA deadenylation
Chromatin organization
Cell-cell communication
Development / Spermat.
mRNA processing, splicing
DNA Replication
Membrane trafficking
Cell Cycle, Meiosis
Signal Transduction
(Notch, Hedgehog)
mRNA transport,
snRNP assembly
Cell Cycle, Mitotic (G0-G1)
Apoptotic execution phase
Cell Cycle, Mitotic (S-phase)
COPII Vesicle Transp.
Pol II Transcription
mRNA decay
Stress Response:
Chaperones, Translesion Synthesis
e
ΔRPF
f
ΔRNA
-0.1
0
+0.1
-0.1
0
+0.1
Differential Intercept Linear Coefficient
g
Adj. R^2 = 0.52
Test set: N = 492; 40%
1.5
1.25
1
0.75
Predicted ΔRPF
Density
0.1
0.05
0
0.75
1
1.25
1.5
Experimental ΔRPF
h
Adj. R^2 = 0.51
Test set: N = 492; 40%
1.5
1.25
1
0.75
Predicted ΔRNA
Density
0.1
0.05
0
0.75
1
1.25
1.5
Experimental ΔRNA

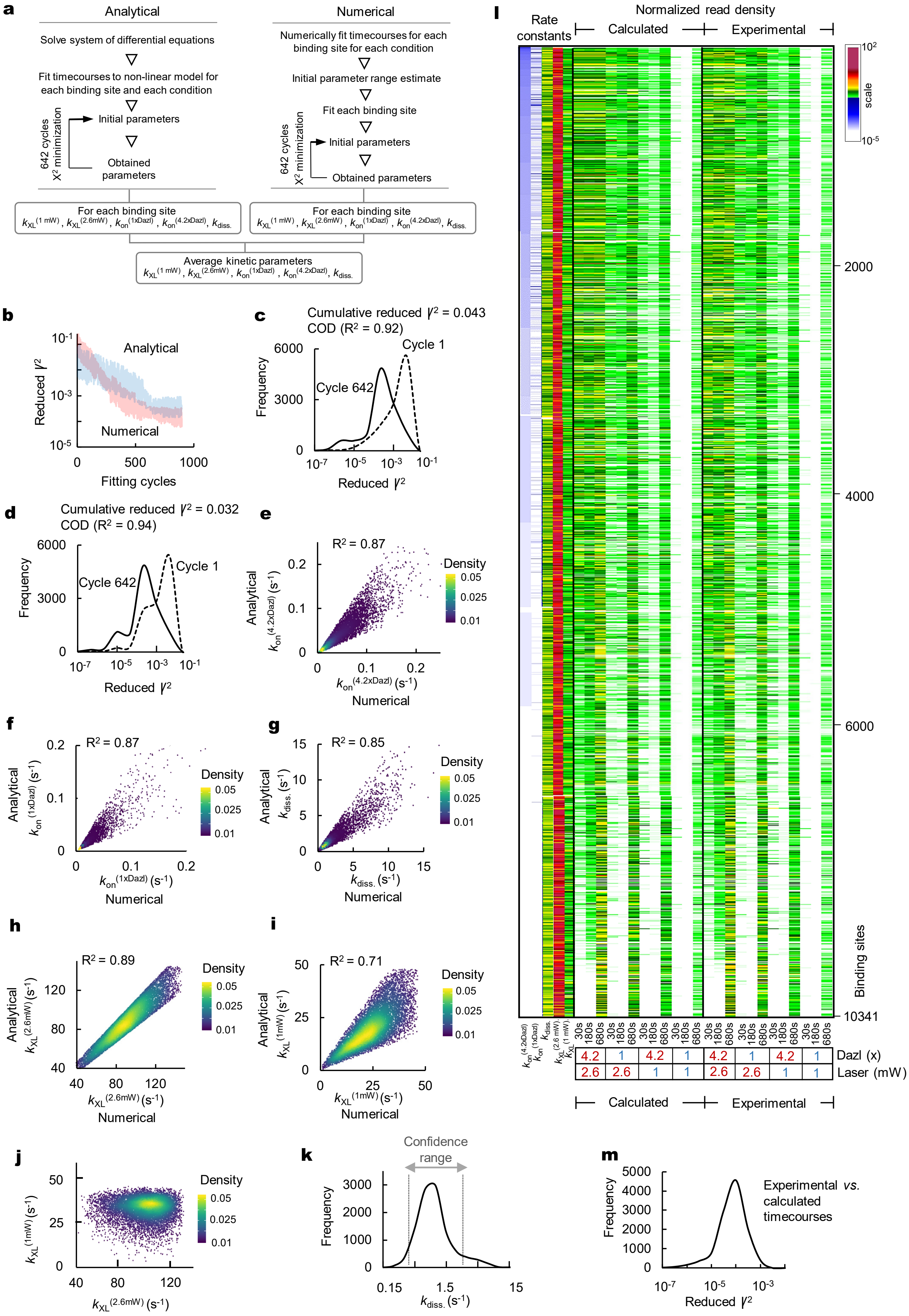
FIGURE 4



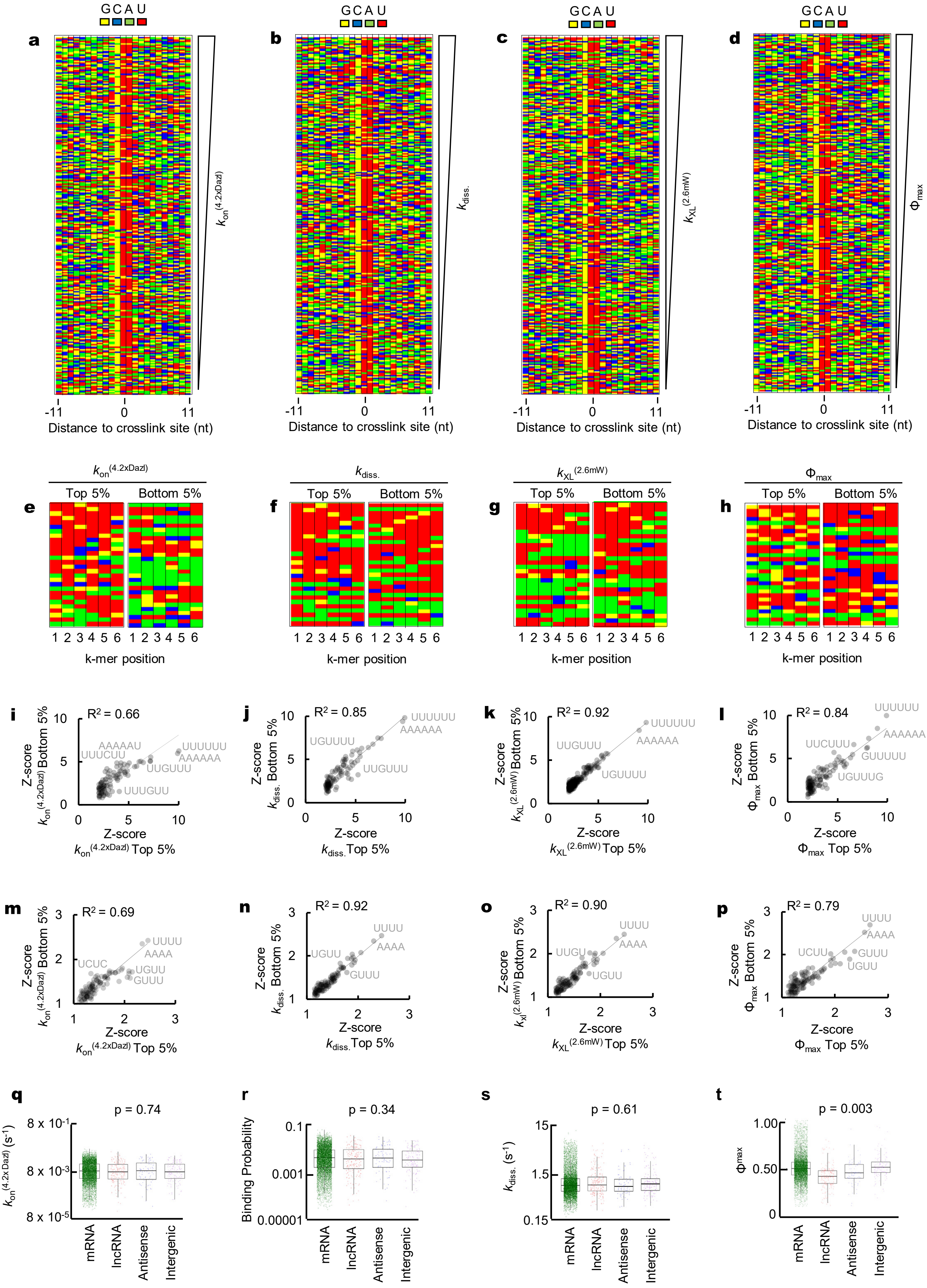
Extended Data Figure 1



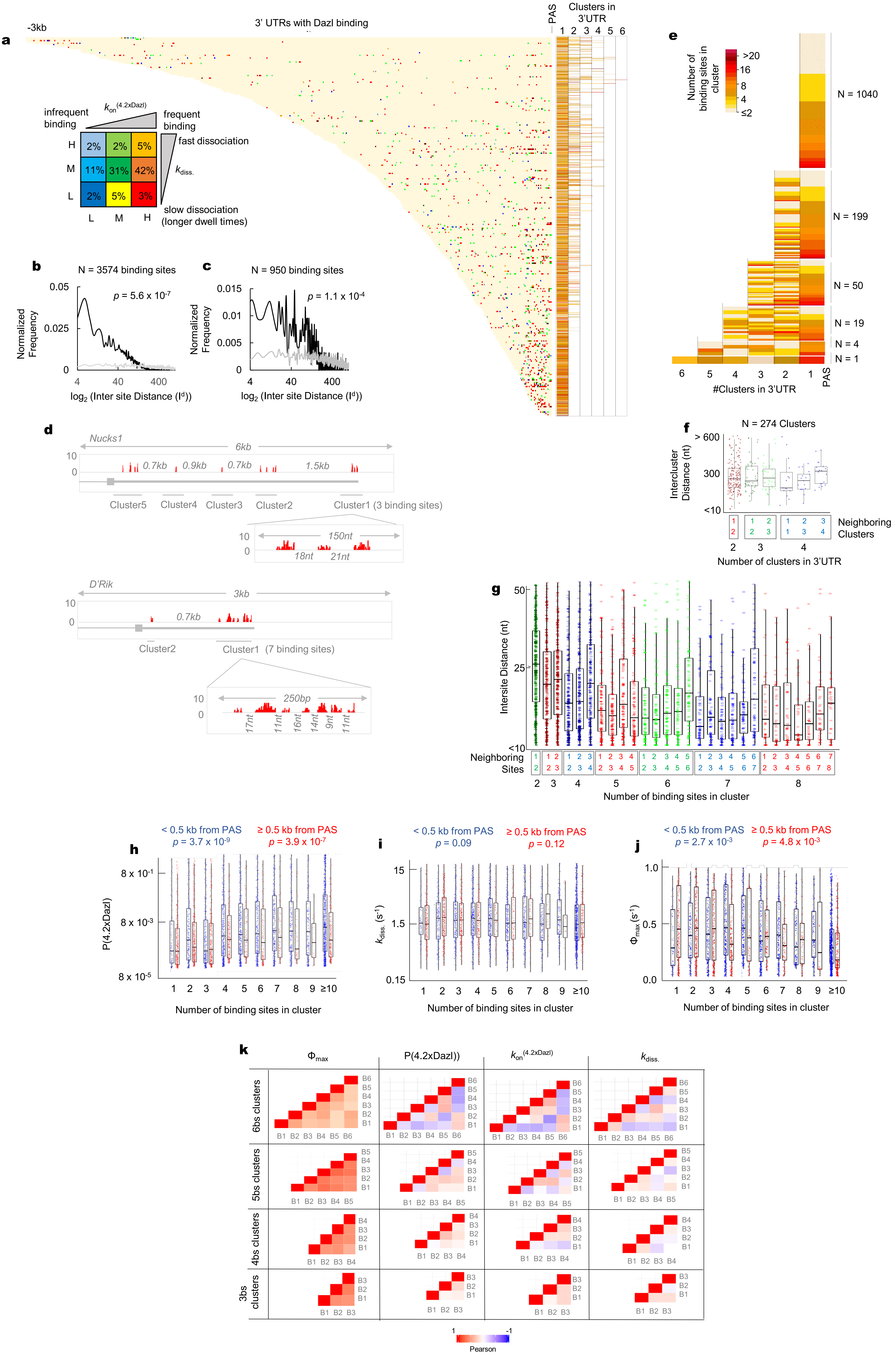
Extended Data Figure 2



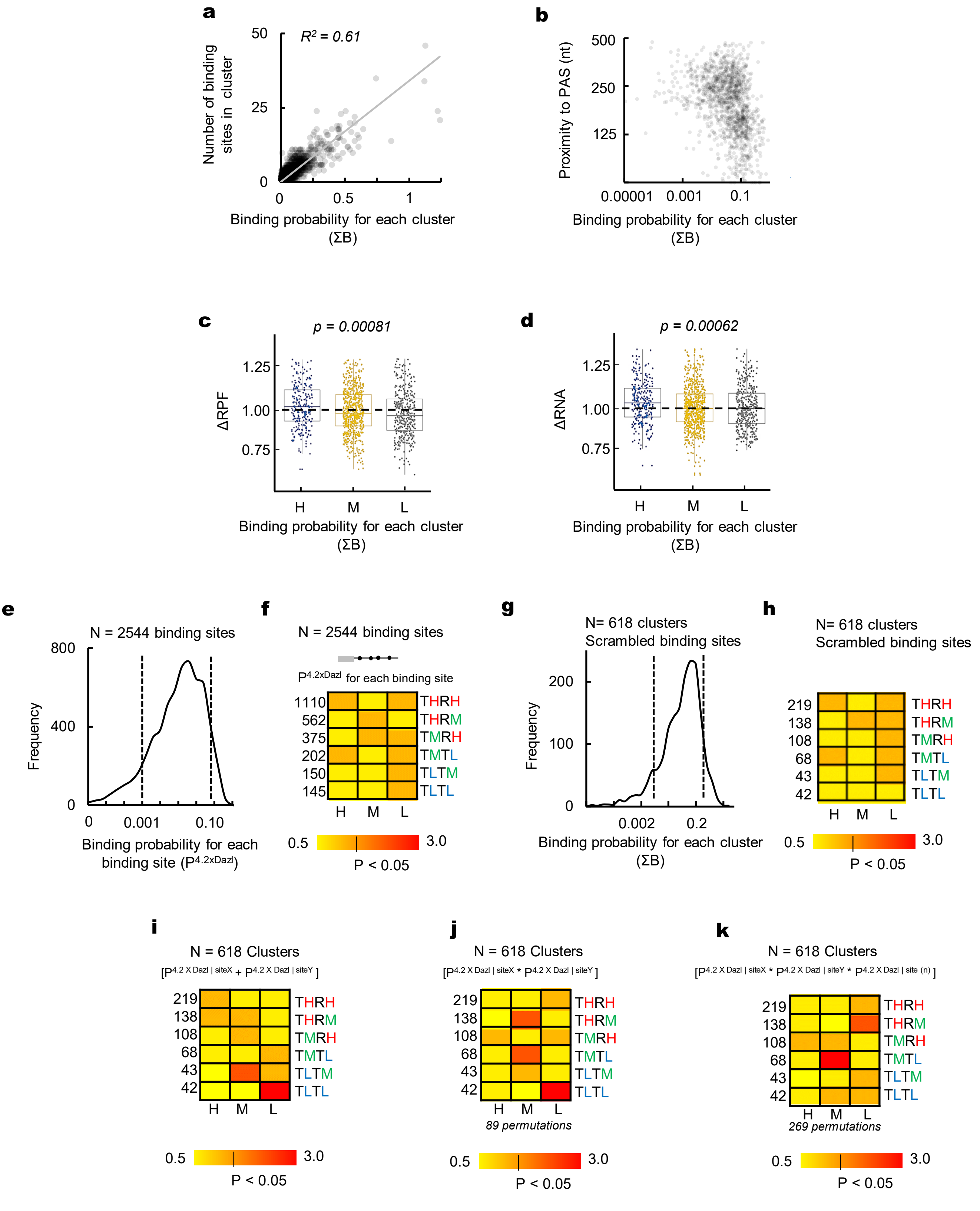
Extended Data Figure 3



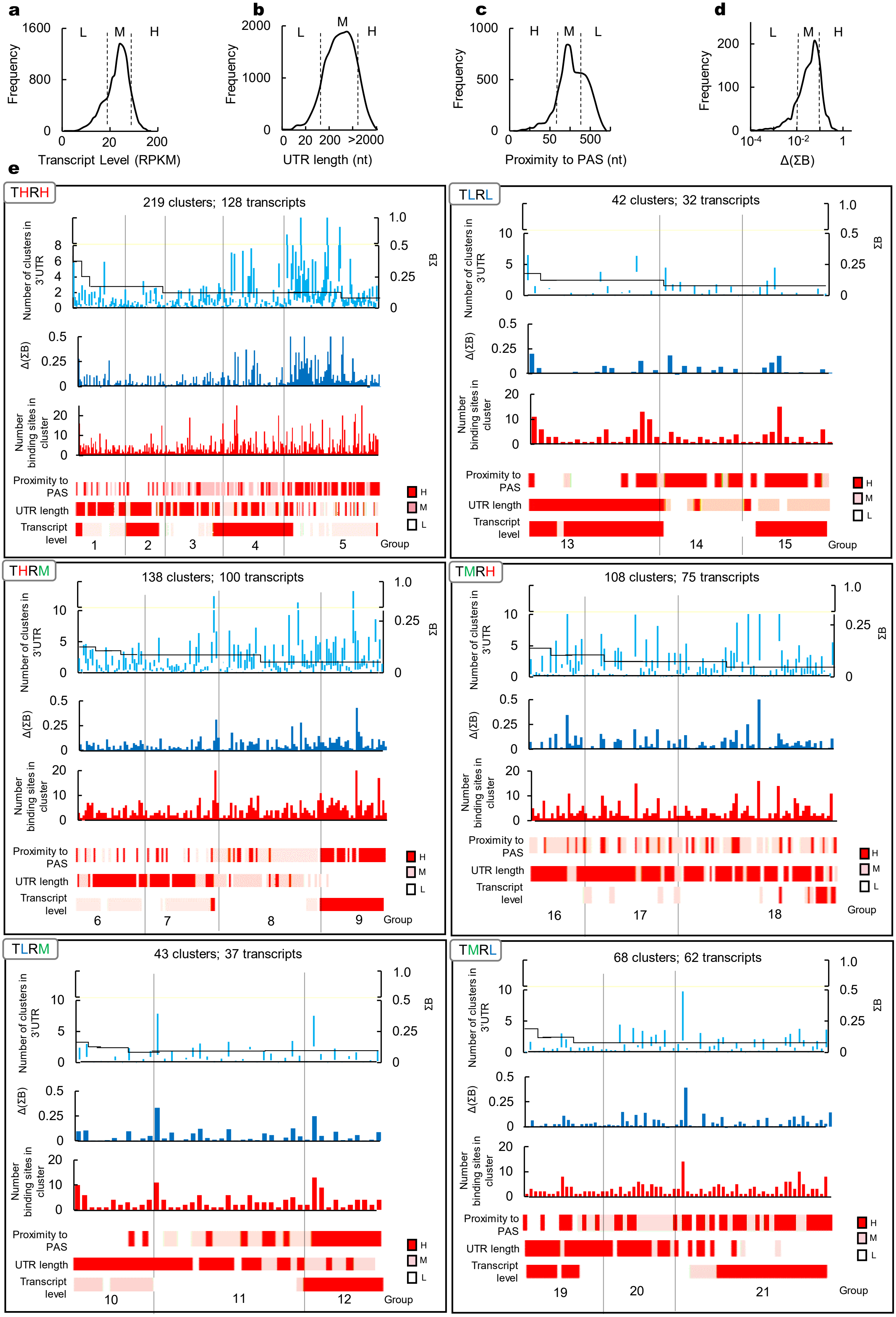
Extended Data Figure 4



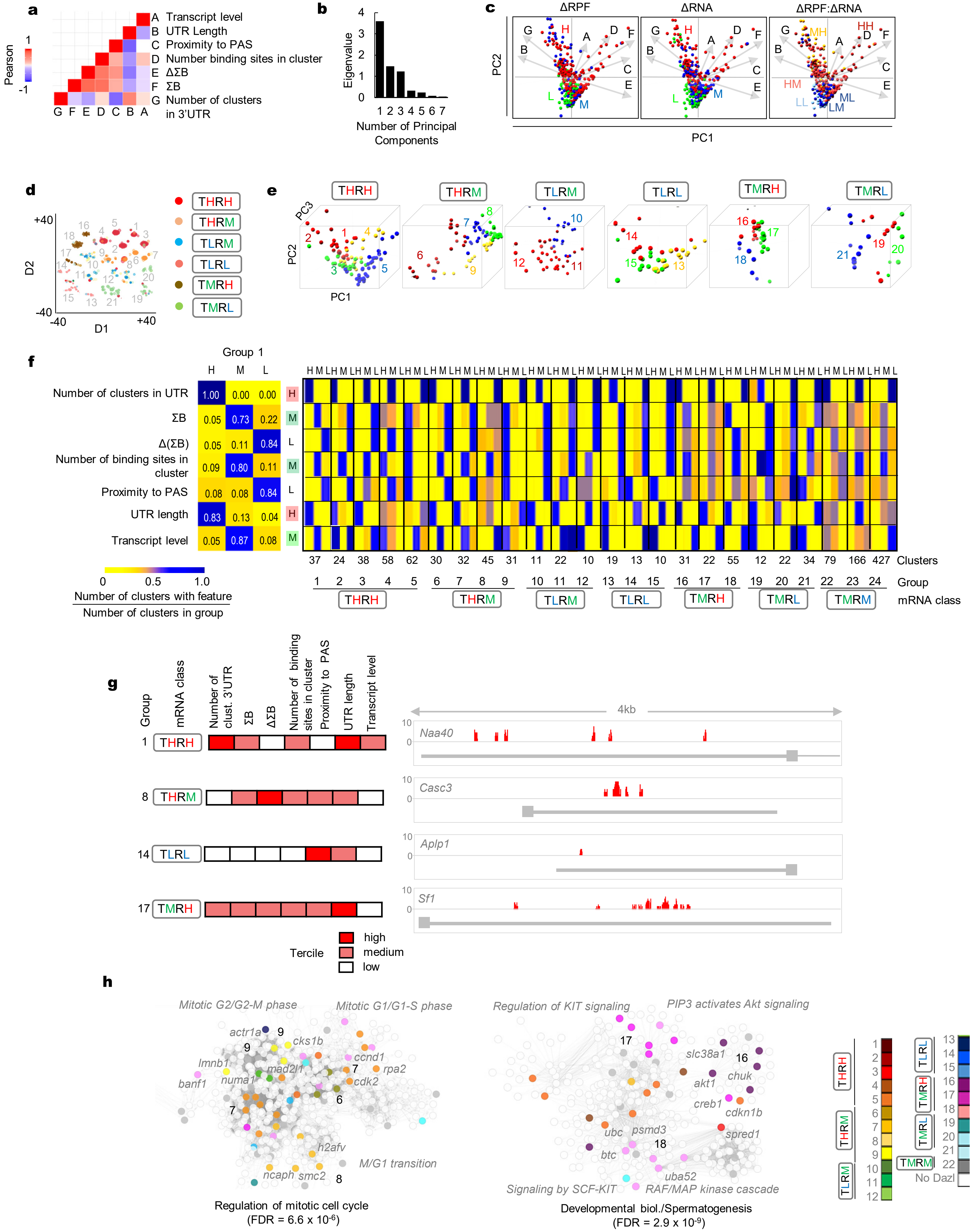
Extended Data Fig.5



Extended Data Figure 6

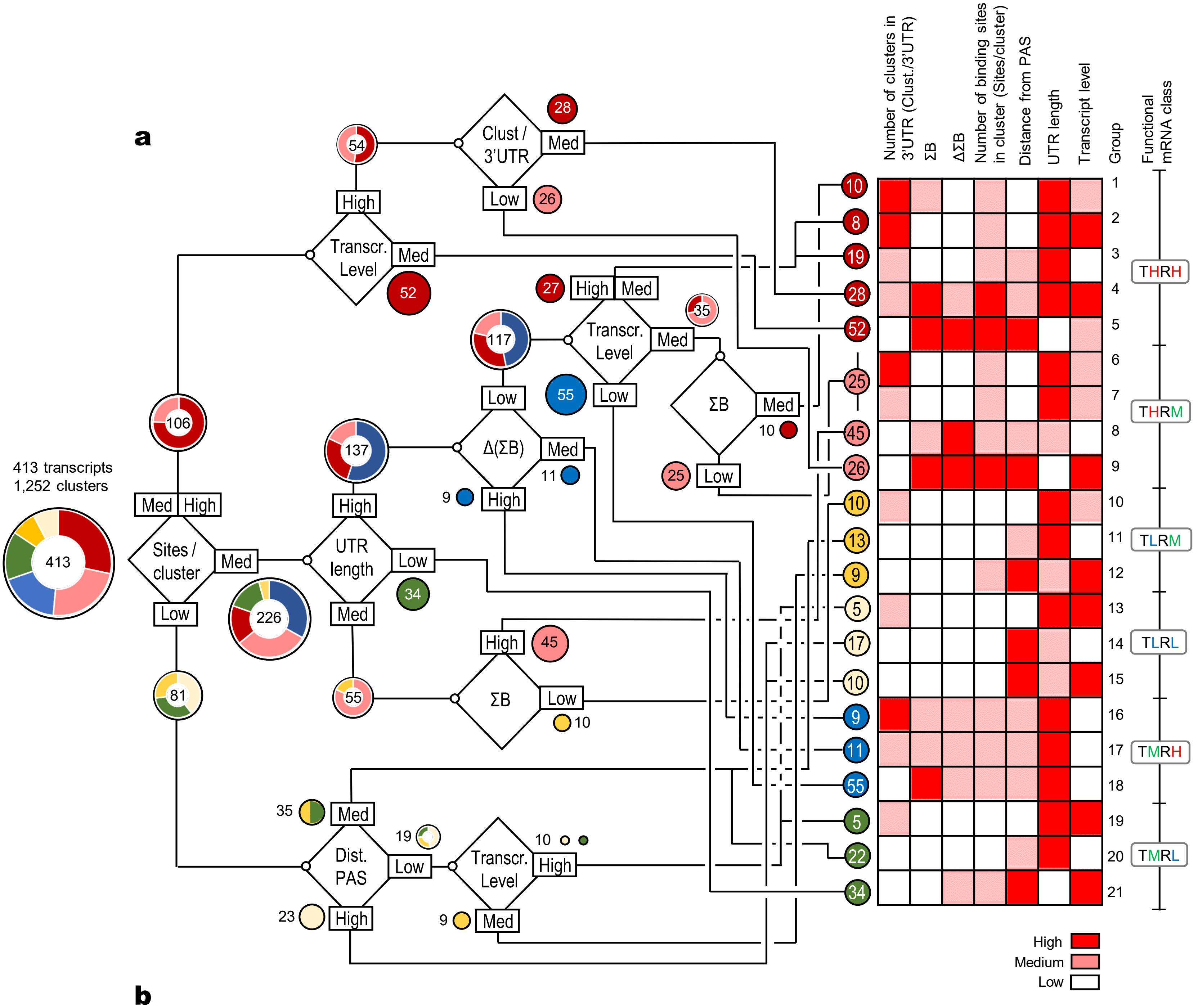


Extended Data Figure 7



Extended Data Fig.8

a



Group	Clust./3'UTR	ΣB	ΔΣB	Sites/cluster	Distance from PAS	UTR length	Transcript level
1	🟥	▨	□	▨	□	🟥	▨
2	🟥	□	□	▨	□	🟥	🟥
3	▨	□	□	▨	▨	🟥	□
4	▨	🟥	▨	🟥	▨	🟥	🟥
5	□	🟥	🟥	🟥	🟥	□	▨
6	🟥	□	□	▨	□	🟥	▨
7	▨	□	□	▨	□	🟥	▨
8	□	▨	🟥	▨	▨	▨	□
9	□	🟥	🟥	🟥	🟥	□	🟥
10	▨	□	□	□	□	🟥	▨
11	□	□	□	□	▨	🟥	□
12	□	□	□	▨	🟥	▨	🟥
13	▨	□	□	□	□	🟥	🟥
14	□	□	□	□	🟥	▨	□
15	□	□	□	□	🟥	▨	🟥
16	🟥	▨	▨	▨	▨	🟥	□
17	▨	▨	▨	▨	▨	🟥	□
18	□	🟥	▨	▨	▨	🟥	□
19	▨	□	□	□	□	🟥	🟥
20	□	□	□	□	▨	🟥	□
21	□	□	▨	▨	🟥	□	🟥

High 🟥
Medium ▨
Low □

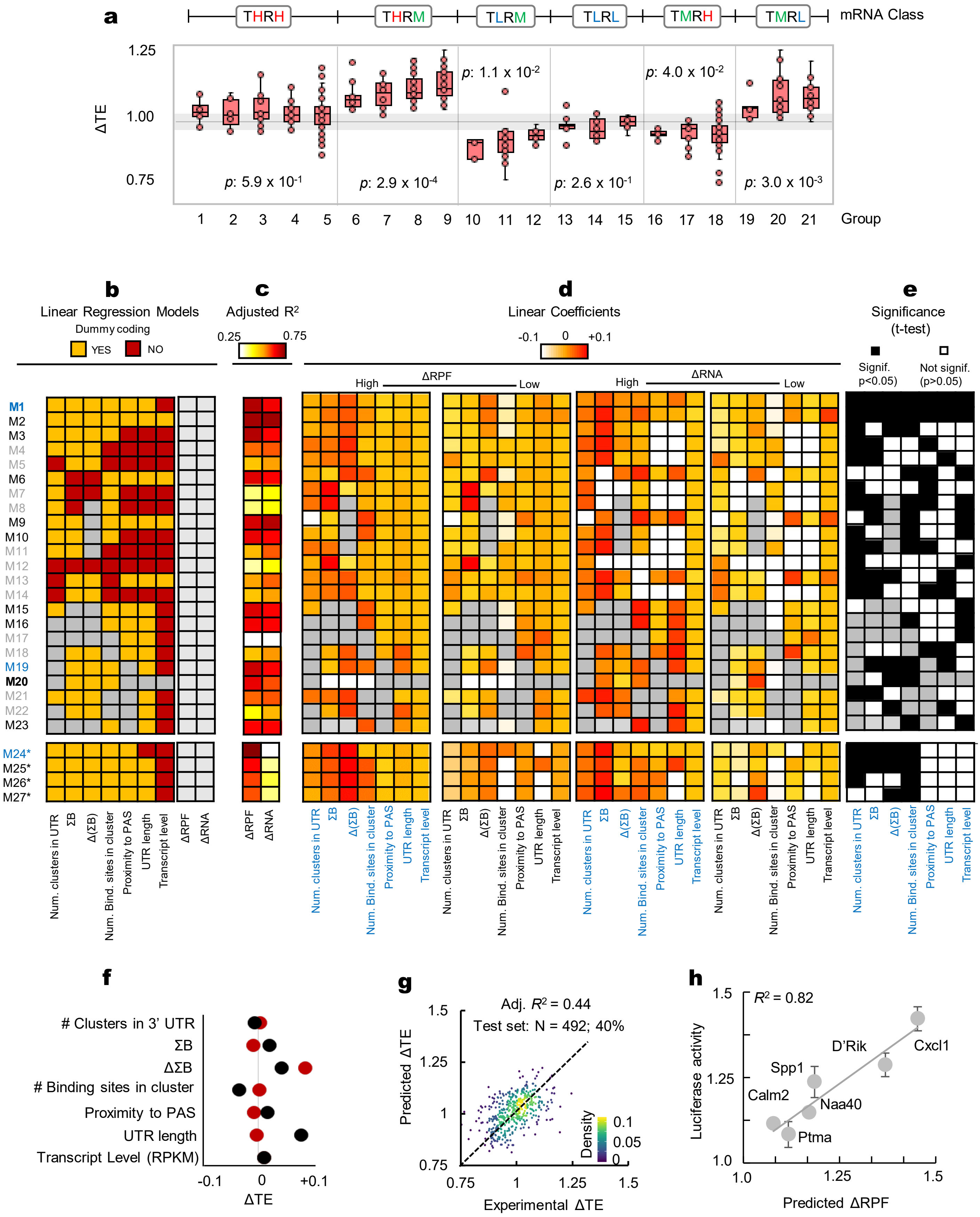
b

from \ to	HH	HM	LL	LM	MH	ML	Total	% correct
HH	117	0	0	0	0	0	117	**100.000**
HM	0	96	0	0	0	0	96	**100.000**
LL	0	0	32	0	0	0	32	**100.000**
LM	0	0	0	19	0	13	32	**59.375**
MH	0	0	0	0	75	0	75	**100.000**
ML	0	0	5	0	0	56	61	**91.803**

from \ to	HH	HM	LL	LM	MH	ML	Total	% correct
HH	8	1	0	0	0	0	8	**90.000**
HM	0	5	0	0	0	0	5	**100.000**
LL	0	0	2	1	0	0	3	**67.000**
LM	0	0	0	1	0	1	2	**50.000**
MH	0	0	0	0	4	0	4	**100.000**
ML	0	0	1	0	0	1	2	**50.000**

from \ to	HH	HM	LL	LM	MH	ML	Total	% correct
HH	4	1	0	0	0	0	5	**90.000**
HM	0	4	0	0	1	0	5	**100.000**
LL	0	0	4	0	0	1	5	**67.000**
LM	0	0	0	1	0	1	2	**50.000**
MH	0	0	0	0	2	0	2	**100.000**
ML	0	0	1	0	0	1	2	**50.000**

Extended Data Figure 9



Extended Data Figure 10

SUPPLEMENTARY MATERIAL

The kinetic landscape of an RNA binding protein in cells

Deepak Sharma, Leah L. Zagore, Matthew M. Brister, Xuan Ye,

Carlos E. Crespo-Hernández, Donny D. Licatalosi & Eckhard Jankowsky

Supplementary Table S1 | Codon optimized *Mus musculus* Dazl (RRM) DNA construct (amino acids 32 -117) **and primers for cloning.**

Dazl (RRM) DNA construct

SacI and XhoI restriction sites are underlined. Complete DNA construct was purchased from Genscript.

```
GGAAATATAGAGCTCTTGCCGGAAGGCAAGATCATGCCGAACACCGTATTCGTAGGAGGAATAG
ACGTACGCATGGACGAAACCGAAATCCGCTCTTTTTTCGCACGCTACGGCTCTGTAAAGGAGGT
TAAAATAATCACGGACAGAACGGGGGTTTCGAAAGGCTACGGATTCGTCTCTTTCTACAACGAT
GTTGACGTTCAGAAAATAGTAGAGTCTCAGATAAACTTTCATGGGAAGAAACTGAAGCTGGGCC
CGGCTATCCGCAAACAATAATGACCTCGAGGGCTGCAA
```

Primers for cloning

SacI and XhoI restriction sites are underlined.

Dazl Forward

```
5’-GGAAATATAGAGCTCTTGCCGGAAGGCAAGATCATGC
```

Dazl Reverse

```
5’-TTGCAGCCCTCGAGGTCATTATTGTTTGCGGATA
```

Supplementary Table S2 | Sequencing adapters and primers.

RNA linkers (Dharmacon)

RL5: 5’-OH AGG GAG GAC GAU GCG G 3’-OH

RL5D: 5’-OH AGG GAG GAC GAU GCG Gr(N)r(N) r(N)r(N)G 3’-OH

RL3: 5’-P GUG UCA GUC ACU UCC AGC GG 3’-puromycin

DNA primers (Operon)

DP5: 5’-AGG GAG GAC GAT GCG G-3’

DP3: 5’-CCG CTG GAA GTG ACT GAC AC-3’

Solexa Fusion Primers (Operon)

SSP1: 5'-CTA TGG ATA CTT AGT CAG GGA GGA CGA TGC GG-3'

Circularization RT primer (Dharmacon)

5’Phos/(GGTTA)(CCGCATCGTCCTCCCT)(CCCTATAGTGAGTCGTATTA)/iSp18/CACTCA/iSp18/(CCGCTGGAAGTGACTGACAC)3’

Antisense DP5 Antisense T7 Promoter DP3

1) 5’Phos-GNNNN CGTGAT CCGCATCGTCCTCCCTC CCTATAGTGAGTCGTATTA - iSp18 - CACTCA -iSp18 – CCGCTGGAAGTGACTGACAC

2) 5’Phos-GNNNN ACATCG CCGCATCGTCCTCCCTC CCTATAGTGAGTCGTATTA - iSp18 - CACTCA -iSp18 – CCGCTGGAAGTGACTGACAC

3) 5’Phos-GNNNN GCCCTA CCGCATCGTCCTCCCTC CCTATAGTGAGTCGTATTA - iSp18 - CACTCA -iSp18 – CCGCTGGAAGTGACTGACAC

4) 5’Phos-GNNNN TGGTCA CCGCATCGTCCTCCCTC CCTATAGTGAGTCGTATTA - iSp18 - CACTCA -iSp18 – CCGCTGGAAGTGACTGACAC

5) 5’Phos-GNNNN CACAGT CCGCATCGTCCTCCCTC CCTATAGTGAGTCGTATTA - iSp18 - CACTCA -iSp18 – CCGCTGGAAGTGACTGACAC

6) 5’Phos-GNNNN ATTGGC CCGCATCGTCCTCCCTC CCTATAGTGAGTCGTATTA - iSp18 - CACTCA -iSp18 – CCGCTGGAAGTGACTGACAC

Complementary barcode sequence

1) ATCACGNNNNG……………

2) CGATGTNNNNG……………

3) TAGGGCNNNNG…………..

4) TGACCANNNNG…………..

5) ACTGTGNNNNG…………..

6) GCCAATNNNNG…………..

Time (s)	Dazl: 4.2x L: 2.6 mW	Dazl: 4.2x L: 1 mW	Dazl: 1x L: 2.6 mW	Dazl: 1x L: 1 mW	Stratalinker
0	$5 \cdot 10^6$	$6 \cdot 10^6$	$4 \cdot 10^6$	$3 \cdot 10^6$	$5 \cdot 10^6$
30	$3 \cdot 10^6$	$3.6 \cdot 10^6$	$4 \cdot 10^6$	$8 \cdot 10^6$	$5 \cdot 10^6$
180	$1.9 \cdot 10^6$	$2.4 \cdot 10^6$	$4 \cdot 10^6$	$5 \cdot 10^6$	$5 \cdot 10^6$
680	$0.6 \cdot 10^6$	$1.2 \cdot 10^6$	$2 \cdot 10^6$	$3 \cdot 10^6$	$5 \cdot 10^6$

Supplementary Table S3 | Number of cells used in each crosslinking experiment
(L: laser power)

Time (s)	Dazl: 4.2x L: 2.6 mW	Dazl: 4.2x L: 1 mW	Dazl: 1x L: 2.6 mW	Dazl: 1x L: 1 mW	Stratalinker
30	88%	98%	80%	91%	91%
180	79%	92%	82%	87%	84%
680	87%	81%	93%	91%	83%

Supplementary Table S4 | Cell Viability after each crosslinking experiment (L: laser power). Cell viability was measured by Trypan-blue staining and cell counting in a hemocytometer (Materials and Methods).

Conditions	680 s	180 s	30 s	0	680 s	180 s	30 s	0	680 s	180 s	30 s	0	680 s	180 s	30 s	0
	Dazl: 4.2x				Dazl: 1x				Dazl: 4.2x				Dazl: 1x			
	Laser: 2.6 mW				Laser: 2.6 mW				Laser: 1 mW				Laser: 1 mW			
Post processed reads [a]	3,372,238	466,053	357,206	13,800	545,542	283,506	150,313	12,720	249,005	364,176	141,804	15,650	394,016	227,026	175,420	8,730
Mapped Reads [b]	1,140,415	341,785	214,324	828	256,405	172,939	111,232	865	186,754	185,730	90,755	1,001	165,487	154,378	112,269	567
% Reads Mapped	33.81	73.33	60.00	6.0	47.00	61.00	74.00	6.8	75.00	51.00	64.00	6.4	42.00	68.00	64.00	6.5
Correction factor [c]	0.89	2.28	2.56	1	2.88	2.22	3.11	1	1.84	2.17	2.2	1	1.88	2.54	3	1
Reads - Peak Intersection [d]	252,932	185,659	173,943	0	204,474	86,071	92,228	0	153,860	48,334	74,552	0	79,527	11,271	14,910	0

Supplementary Table S5 | Sequencing and read processing statistics.

[a] Post processed reads: Reads remaining after de-multiplexing, adapter removal and PCR duplicate collapsing.

[b] Mapped reads: Reads mapped to mouse genome (mm10).

[c] Correction factor: Intensity per read obtained by normalizing number of reads per condition with total crosslinked RNA.

[d] Reads-Peak intersection: Number of reads corresponding to Dazl binding site peaks common to all KIN-CLIP conditions.

Conditions	680 s	180 s	30 s	0	680 s	180 s	30 s	0	680 s	180 s	30 s	0	680 s	180 s	30 s	0
	Dazl: 4.2x				Dazl: 1x				Dazl: 4.2x				Dazl: 1x			
	Laser: 2.6 mW				Laser: 2.6 mW				Laser: 1 mW				Laser: 1 mW			
Bulk Crosslinking Intensity (10^6)	1.012 ± 0.25	0.775 ± 0.22	0.537 ± 0.07	10^{-5}	0.722 ± 0.19	0.384 ± 0.11	0.346 ± 0.07	10^{-5}	0.343 ± 0.10	0.403 ± 0.07	0.199 ± 0.07	10^{-5}	0.311 ± 0.11	0.392 ± 0.07	0.336 ± 0.06	10^{-5}

Supplementary Table S6 | Bulk crosslinking intensity for each crosslinking condition.

Bulk crosslinking (AU; pixel density as described in Image J) was measured as described in Materials and Methods. The errors associated with intensity represent deviation in bulk cross linking as obtained by measuring bulk cross linking for at least three replicates for each time point.

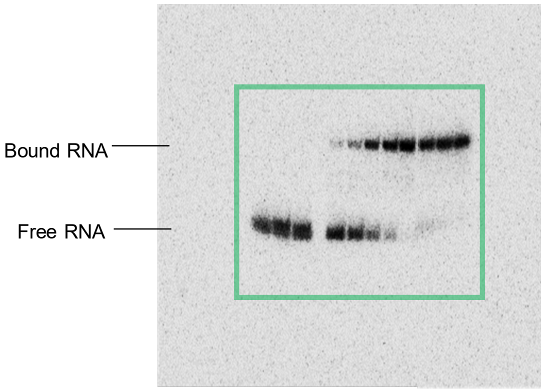
	GC1 Replicate 1	GC1 Replicate 2	GC1 Replicate 3
Post processed Reads[(a)]	1,351,295	910,651	996,650
Mapped Reads [(b)]	123,851	59,674	71,288

Supplementary Table S7 | Sequencing and read processing statistics for iCLIP experiments.

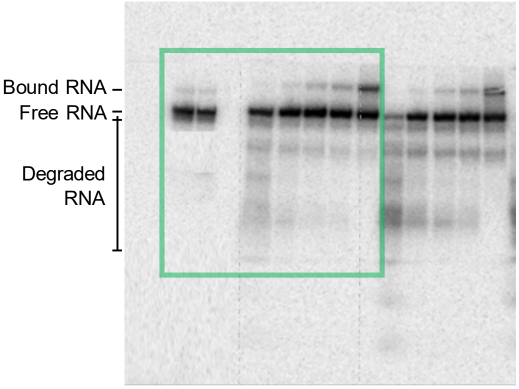
[(a)] Reads remaining after adapter removal and PCR duplicate collapsing.

[(b)] Reads mapped to mouse genome (mm9).

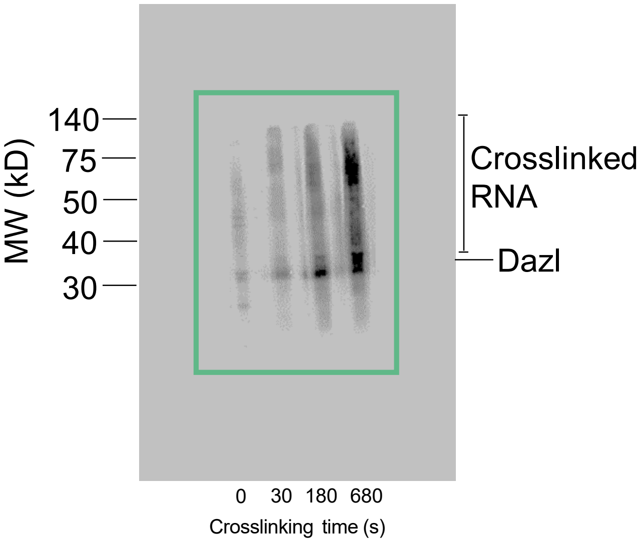
Extended Data Figure 1d



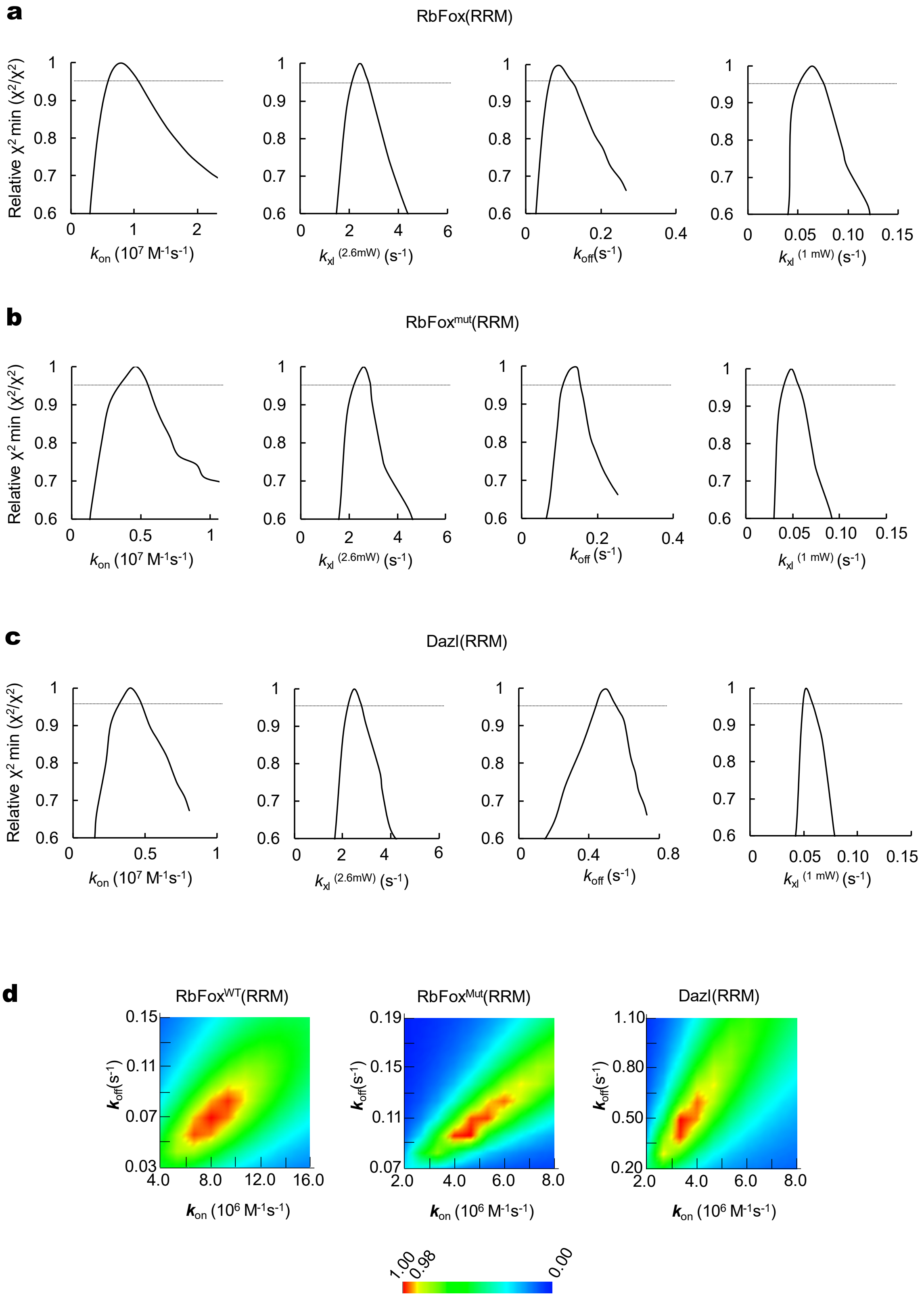
Extended Data Figure 1e



Extended Data Figure 2c

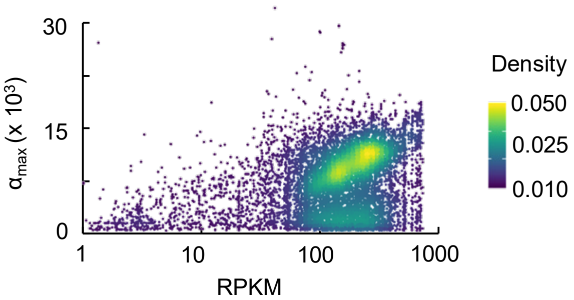


Supplementary Figure S1 | Gel Source data for images shown in Extended Data Figures 1d,e and Extended Data Fig.2d.
Green solid rectangle: gel regions shown in the Extended Data Figures.


a
RbFox(RRM)
Relative χ2 min (χ2/χ2)
1
0.9
0.8
0.7
0.6
0 1 2
kon (107 M-1s-1)
0 2 4 6
kxl (2.6mW) (s-1)
0 0.2 0.4
koff(s-1)
0 0.05 0.10 0.15
kxl (1 mW) (s-1)
b
RbFoxmut(RRM)
Relative χ2 min (χ2/χ2)
0 0.5 1
kon (107 M-1s-1)
0 2 4 6
kxl (2.6mW) (s-1)
0 0.2 0.4
koff (s-1)
0 0.05 0.10 0.15
kxl (1 mW) (s-1)
c
Dazl(RRM)
Relative χ2 min (χ2/χ2)
0 0.5 1
kon (107 M-1s-1)
0 2 4 6
kxl (2.6mW) (s-1)
0 0.4 0.8
koff (s-1)
0 0.05 0.10 0.15
kxl (1 mW) (s-1)
d
RbFoxWT(RRM)
koff(s-1)
0.15
0.11
0.07
0.03
4.0 8.0 12.0 16.0
kon (106 M-1s-1)
RbFoxMut(RRM)
koff(s-1)
0.19
0.15
0.11
0.07
2.0 4.0 6.0 8.0
kon (106 M-1s-1)
Dazl(RRM)
koff(s-1)
1.10
0.80
0.50
0.20
2.0 4.0 6.0 8.0
kon (106 M-1s-1)
1.00
0.98
0.00

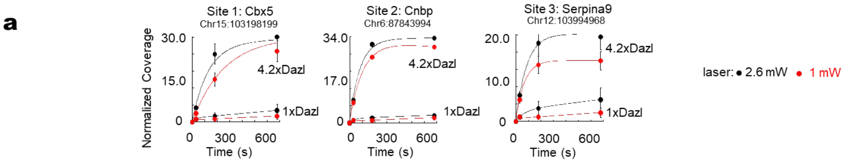
Supplementary Figure S2 | fs laser crosslinking fit space parameters.

1D Fit space analysis (KINTEK) for obtained kinetic parameters (k_{on}, $k_{xl}^{2.6mW}$, k_{off} and k_{xl}^{1mW}) for **(a).** RbFox(RRM), **(b)** RbFox$^{\mathrm{Mut}}$(RRM) and **(c)** Dazl(RRM). (**Fig.1e**). The relative X^2 represents the smallest (optimal) X^2 divided by the X^2 obtained for the entire thermodynamic model. For the optimal parameter value, the relative X^2 = 1. Horizontal lines mark the 95% confidence interval. **d.** 2D Fit space analysis of the relative X^2 of co-varying k_{on} and k_{off}. Both rate constants are constrained for all 3 proteins with a well-defined local minimum (red).

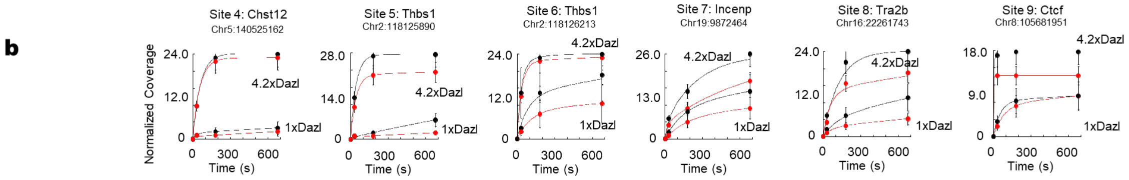


Supplementary Figure S3 | Determination of fractional occupancy (Φ^{max})

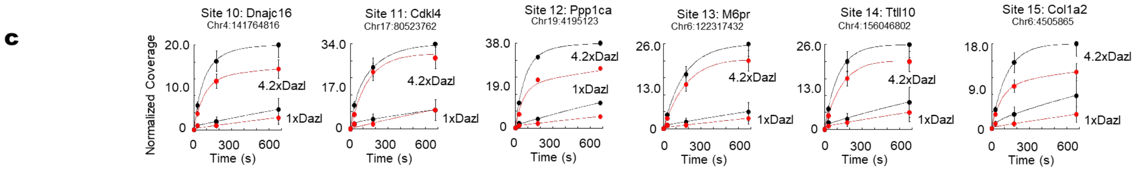
Maximal amplitude (α^{max}: probability of Dazl bound to the fraction of a given binding site that is accessible during the course of the experiment, extrapolated to saturating concentrations of Dazl) plotted vs. level of the corresponding transcript (RPKM). Eq.44 (Materials and Methods) is used to calculate the maximal fractional occupancy (Φ^{max}).



Parameter	Site 1: Cbx5	Site 2: Cnbp	Site 3: Serpina9
$k_{on}^{(4.2xDazl)}$ (s^{-1})	(1.1 ± 0.3) x 10^{-2}	(1.4 ± 0.4) x 10^{-2}	(1.4 ± 0.3) x 10^{-2}
$k_{on}^{(1xDazl)}$ (s^{-1})	(0.4 ± 0.1) x 10^{-3}	(0.3 ± 0.1) x 10^{-3}	(0.8 ± 0.3) x 10^{-3}
k_{off} (s^{-1})	15.3 ± 4.8	1.0 ± 0.4	0.28 ± 0.1
$k_{xl}^{(2.6mW)}$ (s^{-1})	94.3 ± 21.4	73.2 ± 18.2	84.4 ± 22.2
$k_{xl}^{(1.0mW)}$ (s^{-1})	20.5 ± 6.3	26.6 ± 8.3	21.6 ± 8.2



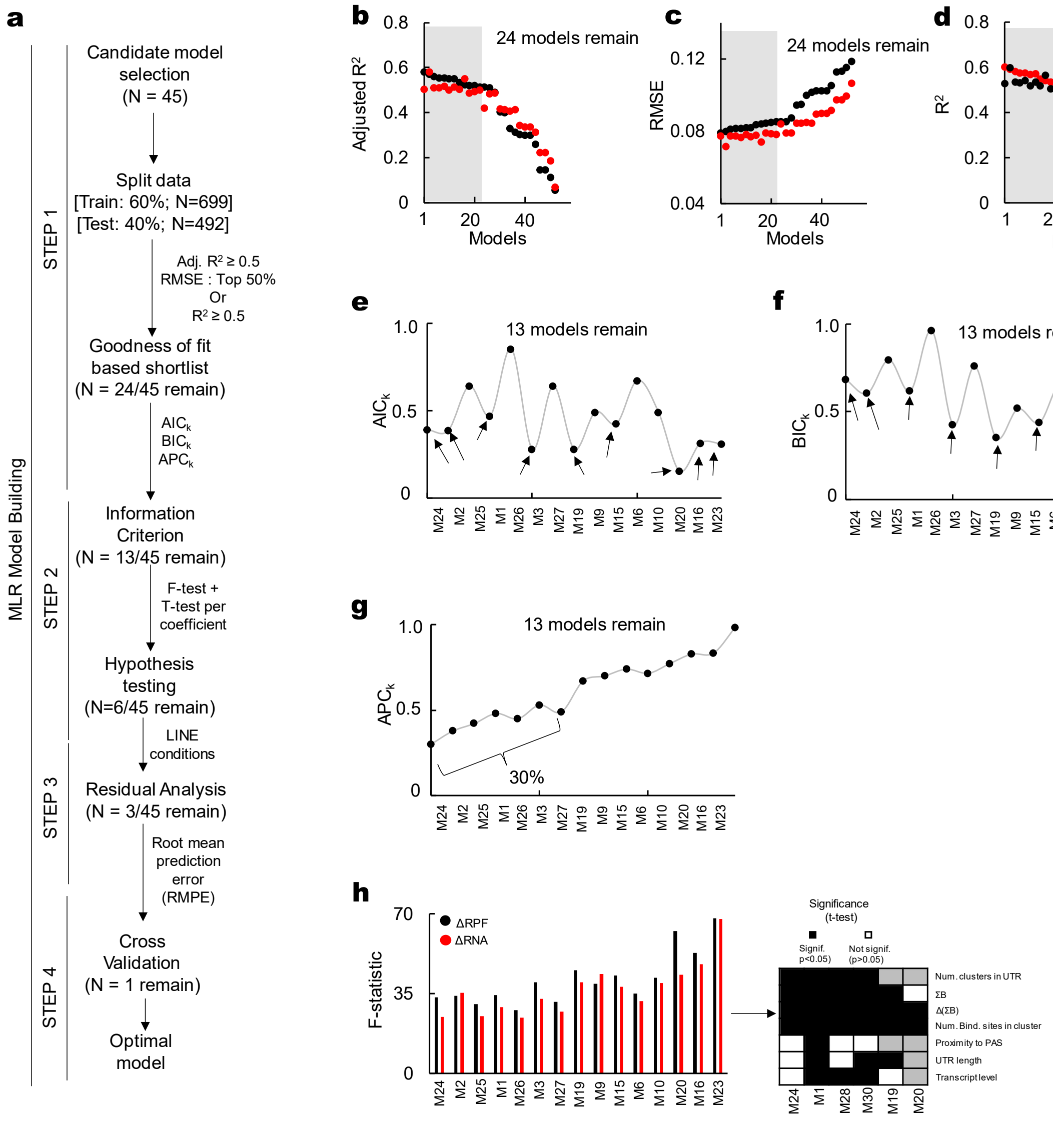
Parameter	Site 4: Chst12	Site 5: Thbs1	Site 6: Thbs1	Site 7: Incenp	Site 8: Tra2b	Site 9: Ctcf
$k_{on}^{(4.2xDazl)}$ (s^{-1})	(1.8 ± 0.5) x 10^{-2}	(1.8 ± 0.4) x 10^{-2}	(2.1 ± 0.7) x 10^{-2}	(0.4 ± 0.1) x 10^{-2}	(1.0 ± 0.4) x 10^{-2}	(5.4 ± 1.2) x 10^{-2}
$k_{on}^{(1xDazl)}$ (s^{-1})	(0.4 ± 0.1) x 10^{-3}	(0.8 ± 0.2) x 10^{-3}	(7.0 ± 1.3) x 10^{-3}	(0.4 ± 0.1) x 10^{-3}	(0.4 ± 0.2) x 10^{-3}	(0.8 ± 0.3) x 10^{-3}
k_{off} (s^{-1})	1.85 ± 0.5	1.20 ± 0.5	1.55 ± 0.4	1.95 ± 0.6	1.28 ± 0.5	1.4 ± 0.4
$k_{xl}^{(2.6mW)}$ (s^{-1})	96.3 ± 31.4	90.9 ± 15.2	77.4 ± 25.2	44.7 ± 11.4	92.3 ± 15.2	90.5 ± 24.4
$k_{xl}^{(1.0mW)}$ (s^{-1})	28.8 ± 8.2	27.4 ± 6.6	20.7 ± 8.2	30.7 ± 7.3	44.7 ± 9.3	30.7 ± 8.1



Parameter	Site 10: Dnajc16	Site 11: Cdkl4	Site 12: Ppp1ca	Site 13: M6pr	Site 14:Ttll10	Site 15: Col1a2
$k_{on}^{(4.2xDazl)}$ (s^{-1})	(0.9 ± 0.2) x 10^{-2}	(0.7 ± 0.3) x 10^{-2}	(0.9 ± 0.2) x 10^{-2}	(0.5 ± 0.1) x 10^{-2}	(0.9 ± 0.4) x 10^{-2}	(0.8 ± 0.3) x 10^{-2}
$k_{on}^{(1xDazl)}$ (s^{-1})	(0.6 ± 0.1) x 10^{-3}	(0.7 ± 0.2) x 10^{-3}	(1.0 ± 0.3) x 10^{-3}	(0.5 ± 0.1) x 10^{-3}	(0.1 ± 0.04) x 10^{-3}	(0.1 ± 0.03) x 10^{-3}
$k_{diss.}$ (s^{-1})	1.64 ± 0.5	1.42 ± 0.3	1.03 ± 0.4	1.00 ± 0.3	1.82 ± 0.5	1.92 ± 0.4
$k_{xl}^{(2.6mW)}$ (s^{-1})	107.2 ± 41.4	60.2 ± 18.2	40.1 ± 12.2	82.56 ± 31.4	83.74 ± 38.2	84.12 ± 42.2
$k_{xl}^{(1.0mW)}$ (s^{-1})	21.2 ± 6.3	28.1 ± 8.3	22.7 ± 8.2	46.12 ± 6.3	20.72 ± 8.3	4.32 ± 1.2

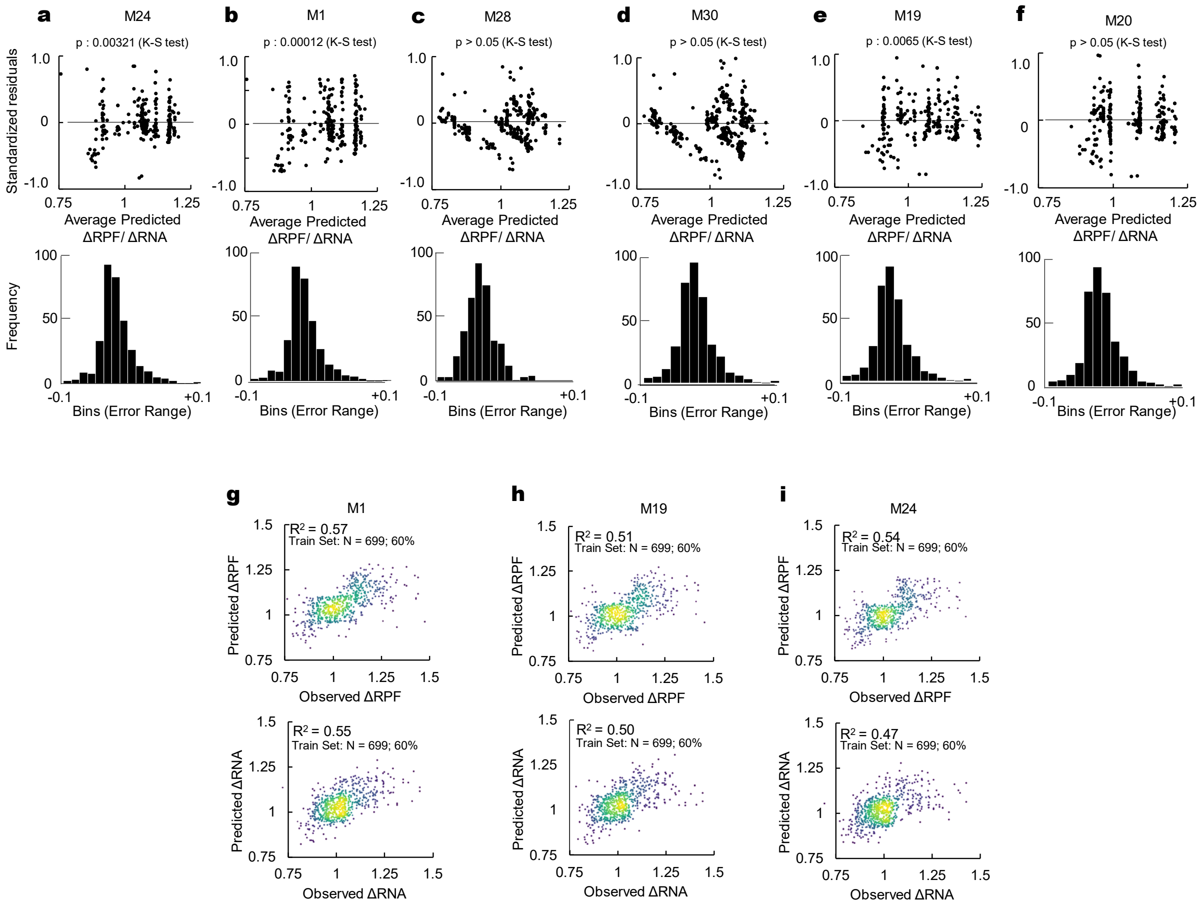
Supplementary Figure S4 | Impact of rate constant variation on crosslinking time courses.

(a) Time courses for Dazl binding sites with differing k_{off} values (highlighted; high, medium and low range of the distribution of k_{off} values) and similar values for other rate constants. **(b)** Time courses for Dazl binding sites with differing values for $k_{on}^{(1xDazl)}$ (s^{-1}) (left) and $k_{on}^{(4.2xDazl)}$ (s^{-1}) (right), and similar values for other rate constants. **(c)** Time courses for Dazl binding sites with differing values for $k_{xl}^{(2.6mW)}$ (s^{-1}) (left) and $k_{xl}^{(1.0mW)}$ (s^{-1}) (right) and similar values for other rate constants. Points mark the experimental normalized peak coverage value (error bars: 95% confidence interval for normalized peak coverage value, determined by minimizing X^2), lines show the curves with calculated rate constants.

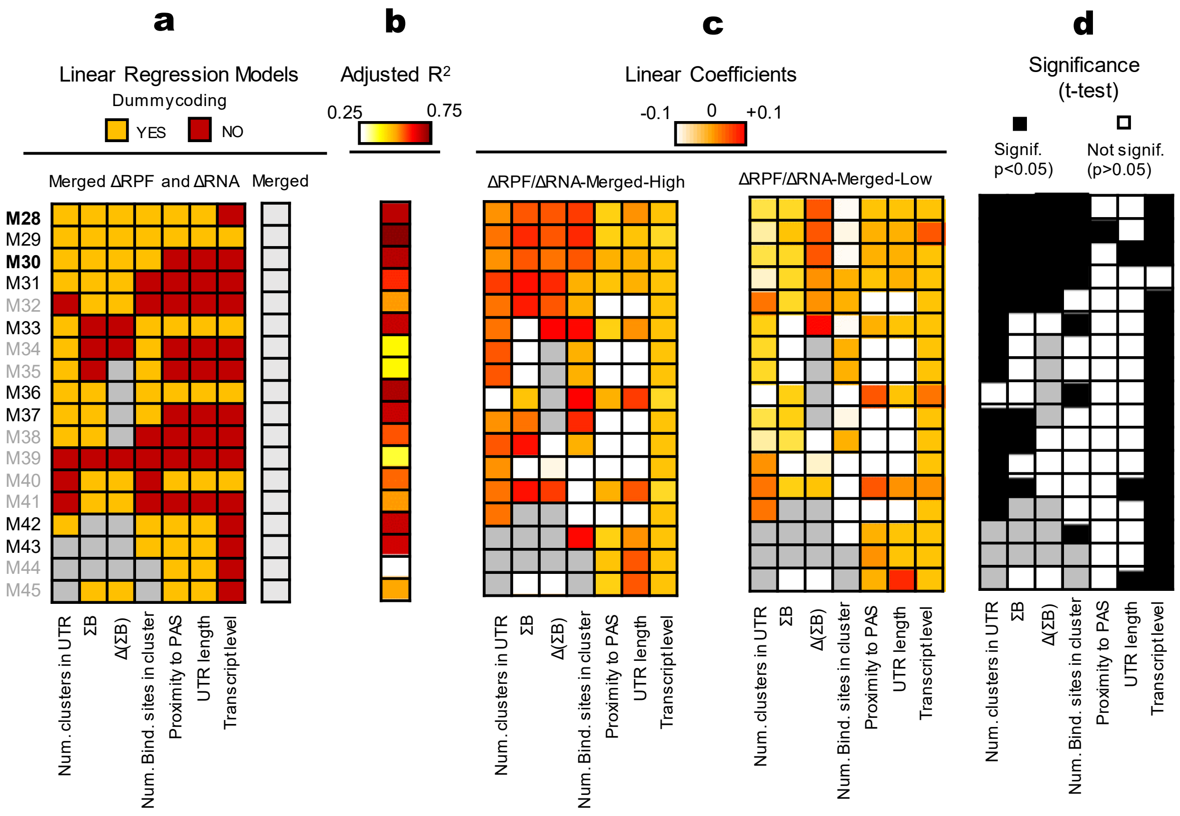


Supplementary Figure S5 | Generation of the Multiple Linear Regression Models.

(a) Flowchart for the development of the multiple linear regression (MLR) models. **(b)** Adjusted R^2 values for all selected candidate models (N = 45). **(c)** Root Mean Squared Error (RMSE) values for all selected candidate models (N = 45). **(d)** R^2 values for all selected candidate models (N = 45). Models with adjusted $R^2 \geq 0.5$, lowest 50% RMSE and/or $R^2 > 0.5$ were shortlisted (N = 24, grey area). Red dots: ΔRNA; Black dots: ΔRPF. **(e-g)** Information criterion statistics (ICS) for models with separate ΔRNA and ΔRPF terms (N = 15; **Extended Data Figure 10b-e**). ICS for models with merged ΔRNA and ΔRPF (N = 8; not shown) was carried out in the similar manner. **(e)** Models with lowest Akaike's Information Criterion (AIC) and **(f)** Bayesian Information Criterion (BIC) are marked (N = 9; arrows). **(g)** Models with lowest Amemiya's Prediction Criterion (APC) are selected (lowest 30%). 13 models remain after ICS criterion (9 with separate ΔRNA and ΔRPF and 4 with merged ΔRNA and ΔRPF, not shown). **(h)** F-statistic for models with separate ΔRNA and ΔRPF. All models satisfied general linear F-statistic condition (F-statistic > 15). Heatmap on the right show shortlisted models (N = 6) with significant t-tests for majority of independent variable terms (at least 60%, $p < 0.05$, black), lowest ICS and significant F-test statistic (N = 6).

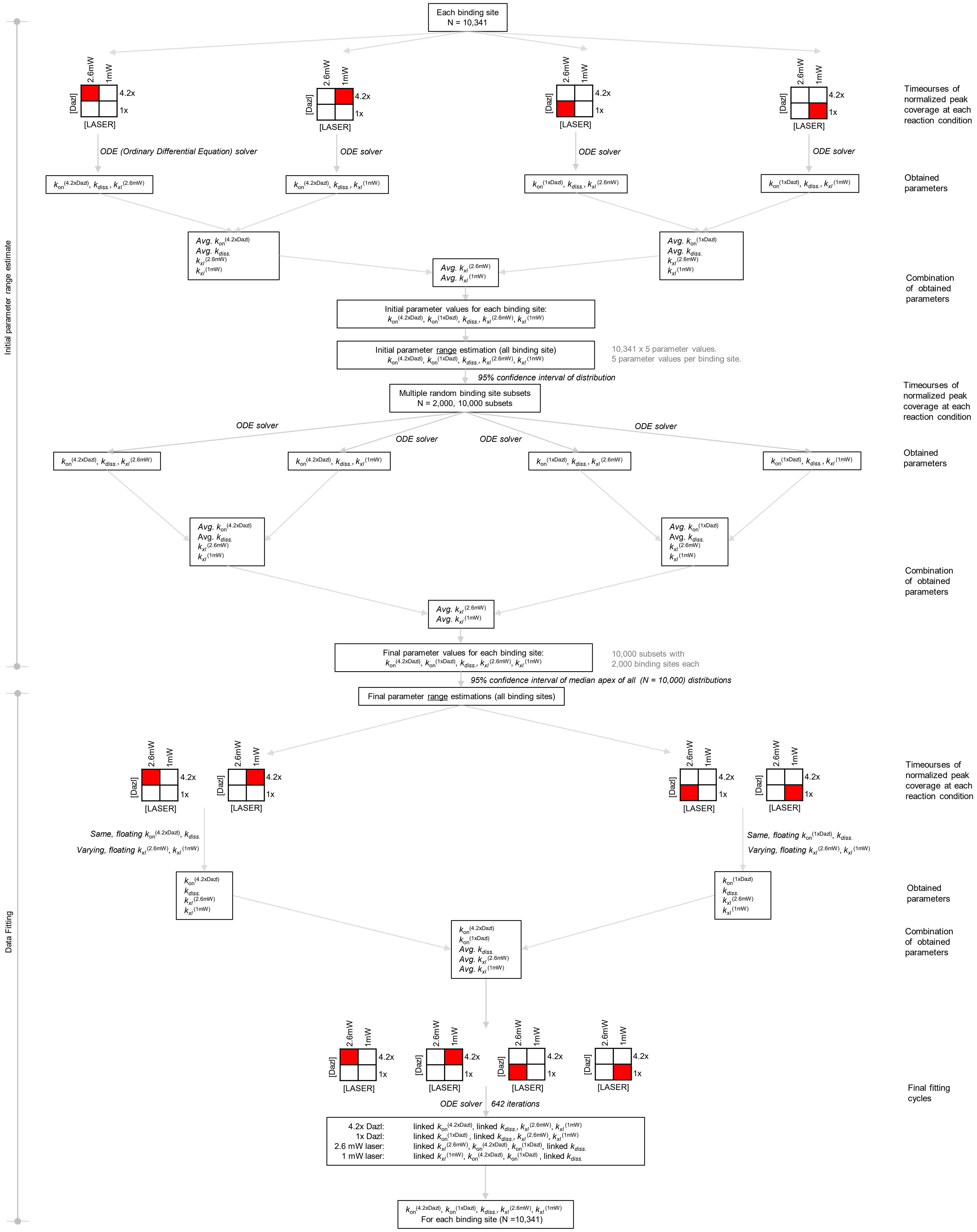


Supplementary Figure S6 | Shortlisting of Multiple Linear Regression Models. (a-f) Upper panels: Standardized residuals versus average predicted ΔRNA and ΔRPF values for models remaining after significance testing (N = 6, **Supplementary Materials Fig.S4**). p-value: Kolmogorov-Smirnov Test (K-S test) for error normality. $p < 0.05$ indicates normal distribution of error residuals (Models M1, M19, M24). Lower panels: Histogram of error residuals for models remaining after significance testing (N = 6, **Supplementary Materials Fig.S4**). **(g)** Correlation between experimental values for ΔRPF (top panel) and ΔRNA (bottom panel) (training data set, N = 699; 60%) and values calculated with the linear regression model (R: adjusted linear correlation coefficient) for models shortlisted in panels **a-f**.



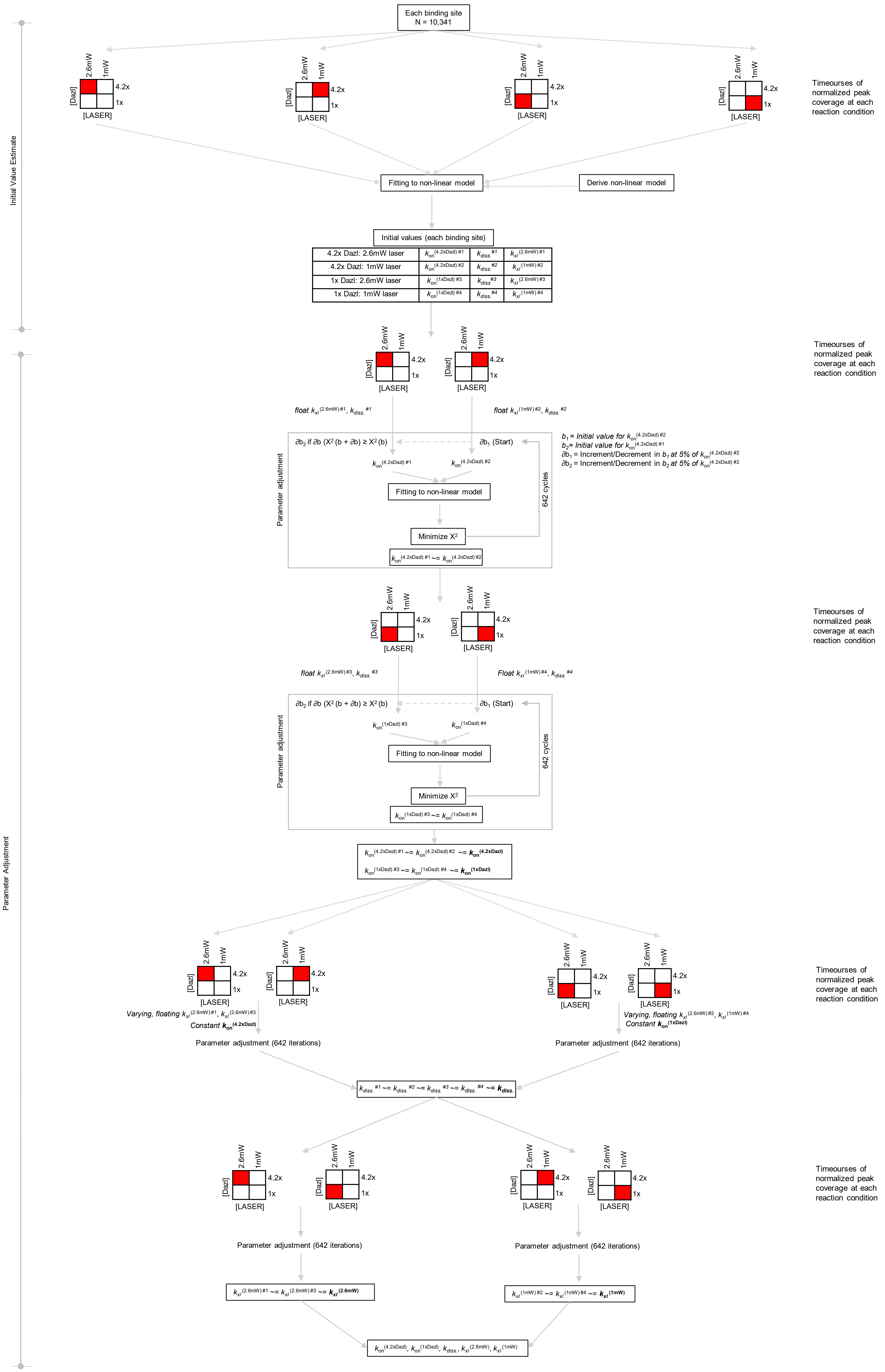
Supplementary Figure S7 | MLR models for merged ΔRPF and ΔRNA terms

(a) Linear Regression models tested (M28 – M45). (Yellow: dummy coding, using terciles of the variables, **Extended Data Fig.8**. Red: no dummy coding; use of continuous data. Grey: variable was omitted. ΔRPF and ΔRNA were merged by normalizing both, ΔRPF and ΔRNA to a scale of 0-1 and then multiplying [ΔRPF X (ΔRNA – 0.01]. The merged terms are distinct from the translation efficiency (ΔTE). **(b)** Adjusted R^2 for each model. **(c)** Differential Intercept Linear Coefficients (DILC) for each model. **(d)** p-values of t-test for each independent variable (N = 7) for all models. Black: p < 0.05; white: p > 0.05.


Initial parameter range estimate
Data Fitting
Each binding site
N = 10,341
2.6mW
1mW
4.2x
1x
[Dazl]
[LASER]
ODE (Ordinary Differential Equation) solver
ODE solver
Timeourses of normalized peak coverage at each reaction condition
$k_{on}^{(4.2xDazl)}$, $k_{diss.}$, $k_{xl}^{(2.6mW)}$
$k_{on}^{(4.2xDazl)}$, $k_{diss.}$, $k_{xl}^{(1mW)}$
$k_{on}^{(1xDazl)}$, $k_{diss.}$, $k_{xl}^{(2.6mW)}$
$k_{on}^{(1xDazl)}$, $k_{diss.}$, $k_{xl}^{(1mW)}$
Obtained parameters
Avg. $k_{on}^{(4.2xDazl)}$
Avg. $k_{diss.}$
$k_{xl}^{(2.6mW)}$
$k_{xl}^{(1mW)}$
Avg. $k_{on}^{(1xDazl)}$
Avg. $k_{diss.}$
$k_{xl}^{(2.6mW)}$
$k_{xl}^{(1mW)}$
Avg. $k_{xl}^{(2.6mW)}$
Avg. $k_{xl}^{(1mW)}$
Combination of obtained parameters
Initial parameter values for each binding site:
$k_{on}^{(4.2xDazl)}$, $k_{on}^{(1xDazl)}$, $k_{diss.}$, $k_{xl}^{(2.6mW)}$, $k_{xl}^{(1mW)}$
Initial parameter range estimation (all binding site)
$k_{on}^{(4.2xDazl)}$, $k_{on}^{(1xDazl)}$, $k_{diss.}$, $k_{xl}^{(2.6mW)}$, $k_{xl}^{(1mW)}$
10,341 x 5 parameter values.
5 parameter values per binding site.
95% confidence interval of distribution
Multiple random binding site subsets
N = 2,000, 10,000 subsets
Timeourses of normalized peak coverage at each reaction condition
ODE solver
Obtained parameters
Combination of obtained parameters
Final parameter values for each binding site:
$k_{on}^{(4.2xDazl)}$, $k_{on}^{(1xDazl)}$, $k_{diss.}$, $k_{xl}^{(2.6mW)}$, $k_{xl}^{(1mW)}$
10,000 subsets with
2,000 binding sites each
95% confidence interval of median apex of all (N = 10,000) distributions
Final parameter range estimations (all binding sites)
Timeourses of normalized peak coverage at each reaction condition
Same, floating $k_{on}^{(4.2xDazl)}$, $k_{diss.}$
Varying, floating $k_{xl}^{(2.6mW)}$, $k_{xl}^{(1mW)}$
Same, floating $k_{on}^{(1xDazl)}$, $k_{diss.}$
Varying, floating $k_{xl}^{(2.6mW)}$, $k_{xl}^{(1mW)}$
$k_{on}^{(4.2xDazl)}$
$k_{diss.}$
$k_{xl}^{(2.6mW)}$
$k_{xl}^{(1mW)}$
$k_{on}^{(1xDazl)}$
$k_{diss.}$
$k_{xl}^{(2.6mW)}$
$k_{xl}^{(1mW)}$
Obtained parameters
$k_{on}^{(4.2xDazl)}$
$k_{on}^{(1xDazl)}$
Avg. $k_{diss.}$
Avg. $k_{xl}^{(2.6mW)}$
Avg. $k_{xl}^{(1mW)}$
Combination of obtained parameters
Final fitting cycles
ODE solver
642 iterations
4.2x Dazl: linked $k_{on}^{(4.2xDazl)}$, linked $k_{diss.}$, $k_{xl}^{(2.6mW)}$, $k_{xl}^{(1mW)}$
1x Dazl: linked $k_{on}^{(1xDazl)}$, linked $k_{diss.}$, $k_{xl}^{(2.6mW)}$, $k_{xl}^{(1mW)}$
2.6 mW laser: linked $k_{xl}^{(2.6mW)}$, $k_{on}^{(4.2xDazl)}$, $k_{on}^{(1xDazl)}$, linked $k_{diss.}$
1 mW laser: linked $k_{xl}^{(1mW)}$, $k_{on}^{(4.2xDazl)}$, $k_{on}^{(1xDazl)}$, linked $k_{diss.}$
$k_{on}^{(4.2xDazl)}$, $k_{on}^{(1xDazl)}$, $k_{diss.}$, $k_{xl}^{(2.6mW)}$, $k_{xl}^{(1mW)}$
For each binding site (N =10,341)

Supplementary Scheme S1 | Numerical data fitting process

Steps for the numerical fitting of crosslinking timecourses to calculate kinetic parameters. Square boxes represent KIN-CLIP conditions (red).


Initial Value Estimate
Parameter Adjustment
Each binding site
N = 10,341
2.6mW
1mW
[Dazl]
4.2x
1x
[LASER]
Timeourses of normalized peak coverage at each reaction condition
Fitting to non-linear model
Derive non-linear model
Initial values (each binding site)
float $k_{xl}^{(2.6\mathrm{mW})\#1}$, $k_{diss.}^{\#1}$
float $k_{xl}^{(1\mathrm{mW})\#2}$, $k_{diss.}^{\#2}$
Parameter adjustment
∂b_2 if ∂b ($X^2(b + \partial b) \geq X^2(b)$)
∂b_1 (Start)
$k_{on}^{(4.2\mathrm{xDazl})\#1}$
$k_{on}^{(4.2\mathrm{xDazl})\#2}$
Fitting to non-linear model
Minimize X^2
642 cycles
$k_{on}^{(4.2\mathrm{xDazl})\#1} \sim= k_{on}^{(4.2\mathrm{xDazl})\#2}$
b_1 = *Initial value for* $k_{on}^{(4.2\mathrm{xDazl})\#2}$
b_2= *Initial value for* $k_{on}^{(4.2\mathrm{xDazl})\#1}$
∂b_1 = Increment/Decrement in b_1 *at 5% of* $k_{on}^{(4.2\mathrm{xDazl})\#2}$
∂b_2 = Increment/Decrement in b_2 *at 5% of* $k_{on}^{(4.2\mathrm{xDazl})\#2}$
float $k_{xl}^{(2.6\mathrm{mW})\#3}$, $k_{diss.}^{\#3}$
Float $k_{xl}^{(1\mathrm{mW})\#4}$, $k_{diss.}^{\#4}$
Parameter adjustment
∂b_2 if ∂b ($X^2(b + \partial b) \geq X^2(b)$)
∂b_1 (Start)
$k_{on}^{(1\mathrm{xDazl})\#3}$
$k_{on}^{(1\mathrm{xDazl})\#4}$
Fitting to non-linear model
Minimize X^2
642 cycles
$k_{on}^{(1\mathrm{xDazl})\#3} \sim= k_{on}^{(1\mathrm{xDazl})\#4}$
$k_{on}^{(4.2\mathrm{xDazl})\#1} \sim= k_{on}^{(4.2\mathrm{xDazl})\#2} \sim= \boldsymbol{k_{on}^{(4.2\mathrm{xDazl})}}$
$k_{on}^{(1\mathrm{xDazl})\#3} \sim= k_{on}^{(1\mathrm{xDazl})\#4} \sim= \boldsymbol{k_{on}^{(1\mathrm{xDazl})}}$
Varying, floating $k_{xl}^{(2.6\mathrm{mW})\#1}$, $k_{xl}^{(2.6\mathrm{mW})\#3}$
Constant $\boldsymbol{k_{on}^{(4.2\mathrm{xDazl})}}$
Parameter adjustment (642 iterations)
Varying, floating $k_{xl}^{(2.6\mathrm{mW})\#2}$, $k_{xl}^{(1\mathrm{mW})\#4}$
Constant $\boldsymbol{k_{on}^{(1\mathrm{xDazl})}}$
Parameter adjustment (642 iterations)
$k_{diss.}^{\#1} \sim= k_{diss.}^{\#2} \sim= k_{diss.}^{\#3} \sim= k_{diss.}^{\#4} \sim= \boldsymbol{k_{diss.}}$
Parameter adjustment (642 iterations)
Parameter adjustment (642 iterations)
$k_{xl}^{(2.6\mathrm{mW})\#1} \sim= k_{xl}^{(2.6\mathrm{mW})\#3} \sim= \boldsymbol{k_{xl}^{(2.6\mathrm{mW})}}$
$k_{xl}^{(1\mathrm{mW})\#2} \sim= k_{xl}^{(1\mathrm{mW})\#4} \sim= \boldsymbol{k_{xl}^{(1\mathrm{mW})}}$
$k_{on}^{(4.2\mathrm{xDazl})}$, $k_{on}^{(1\mathrm{xDazl})}$, $k_{diss.}$, $k_{xl}^{(2.6\mathrm{mW})}$, $k_{xl}^{(1\mathrm{mW})}$

4.2x Dazl: 2.6mW laser	$k_{on}^{(4.2\mathrm{xDazl})\#1}$	$k_{diss.}^{\#1}$	$k_{xl}^{(2.6\mathrm{mW})\#1}$
4.2x Dazl: 1mW laser	$k_{on}^{(4.2\mathrm{xDazl})\#2}$	$k_{diss.}^{\#2}$	$k_{xl}^{(1\mathrm{mW})\#2}$
1x Dazl: 2.6mW laser	$k_{on}^{(1\mathrm{xDazl})\#3}$	$k_{diss}^{\#3}$	$k_{xl}^{(2.6\mathrm{mW})\#3}$
1x Dazl: 1mW laser	$k_{on}^{(1\mathrm{xDazl})\#4}$	$k_{diss.}^{\#4}$	$k_{xl}^{(1\mathrm{mW})\#4}$

Supplementary Scheme S2 | Analytical data fitting process

Steps for the numerical fitting of crosslinking timecourses to calculate kinetic parameters. Square boxes represent KIN-CLIP conditions (red).

Point-by-point Response to Referee Comments.

We thank the referees for their encouraging and constructive comments, which have helped to make the manuscript clearer and stronger.

Referee Comments are black, our responses are in red and captions for figures are in purple.

Referee #1:

Summary

In their manuscript, Sharma and colleagues describe the development of KIN-CLIP, which allows to determine kinetics of protein-RNA interactions in living cells. To this end, they use time-resolved UV crosslinking of proteins to RNA. They first validate that time-resolved UV crosslinking enables kinetic studies on protein-RNA interaction in vitro. They then move on to study binding kinetics of the RNA-binding protein Dazl in mouse cells. They use the kinetic data to study properties of Dazl binding clusters and predict Dazl-mediated RNA and ribosome regulation based on KIN-CLIP parameters and additional features.

General appraisal

In my opinion, the authors made an outstanding effort to pull this project and to develop a technology that allows to determine kinetics of protein-RNA interactions in living cells. The application in form of the Dazl code is less convincing for me (see comments below). I think overall the manuscript is a great scientific contribution and I recommend publication in Nature. However, I would ask to fully address the following comments before publication.

Major comments

1) In vitro crosslinking: In the current format it is difficult to grasp the main concept of how to deduce kinetics from the UV time course experiments. I think it would be important to briefly explain in the main text how to deduce constants from the data. A little schematic on this in Fig. 1d would also be helpful. It would be good to explain why different time points and different protein concentrations are needed. (In the current version of the manuscript the focus lies on explaining why the laser is less harmful for the RNA, which is less relevant for the rest of the manuscript.)

We have now included a brief explanation of why the calculation of the kinetic parameters requires the measurement of crosslinking timecourses under different reaction conditions. We have also included a small scheme in Figure 1, as suggested. A detailed description of how the kinetic data are calculated from the timecourses is provided in the Materials and Methods.

We also considered the note on the RNA degradation data and we agree with the reviewer – the data is technical in nature, although the documentation of limited RNA degradation is essential. We have therefore moved these data to Extended Figure 1, thereby focusing Figure 1 better on the main narrative of the manuscript.

2) Kinetics in cells:

A) I was surprised by the extremely low errors estimated for the different constants (often <10%). This is surprising since measurements used as input such as total crosslinking signal on the membrane might not be very precise. In this context it would be very important to provide information on the reproducibility of the obtained data from replicate experiments. Also, the authors should double check the error models.

The errors for individual rate constants cover a range from less than 5% to sometimes 50% of the reported value. In our estimate this data range does not indicate very low errors.

The crosslinking measurements, which were used to determine the overall crosslinking efficiency for each respective CLIP library, were performed in several replicates. We have now provided the information on the replicates in **Supplementary Material Table S6.**

We have also verified the error models. The standard errors for rate constants indicate the bounds of the respective fitting quality, as assessed by minimized X^2. We have noted this in the caption for **Fig.2c.**

The error models are described in the Materials and Methods section, equation 7 for the numerical fit:

$$\chi^2 = \sum_i \frac{(O_i - C_i)^2}{\sigma_i^2} \qquad \text{(Eq.7)}$$

and by equation 41 for the analytical fit:

$$\chi^2 = \sum_{i=1}^{n} [\frac{Y_i - f(x_i', \beta)}{\sigma_i}]^2 \qquad \text{(Eq.41)}$$

The errors mark lower and upper bounds for the rate constants at a 95% confidence interval (CI). In other words, these errors describe how well the models fit the given data for a binding site and how much fluctuation in rate constants (95% CI) will still yield the same fit. A low range for 95% CI for the errors for a given rate constant indicates a constrained fit, a larger range indicates a poorer fit. Error distributions for both numerical and analytical fits are also shown in the form of reduced/minimized X^2 in **Extended Data Figure 2 c, d**.

B) It is convincing to see that about 85% of Dazl binding sites contain a GUU motif. However, it is difficult to understand that there is only very weak difference e.g. in the comparison of motif enrichment comparing binding sites with top and bottom scoring rate constants. In this context it might be good to validate that the constants obtained in vivo to show correlation with affinities obtained in vitro. Would there be an RNA Bind-N-Seq dataset (or similar) available for comparison?

We agree with the reviewer - the small difference in motif enrichment comparing binding sites with top and bottom scoring rate constants is notable. This result suggests that Dazl displays high selectivity for its cognate GUU motif in cells.

We have measured affinity, association and dissociation rate constants of Dazl(RRM) *in vitro* for an RNA with the cognate GUU motif (**Fig.1**, **Extended Data Fig.1**). These *in vitro* kinetic parameters are within the range of parameters measured for the Dazl binding sites in the cell (**Fig.2e**), and we note this in the text.

Unfortunately, no RNA Bind-n-seq dataset is available for Dazl. However, there is a RNA Bind-n-Seq dataset in the Encode Database for a human Dazl ortholog Daz3, which is 92% homologous to Dazl (**Fig.R1**). The data for Daz3 reveal a clear GUU consensus motif – essentially all RNA variants that were identified in RNA Bind-n-Seq experiments to bind Daz3 contain the GUU core consensus. These data are consistent with very high inherent selectivity of Daz3 towards the GUU core motif.

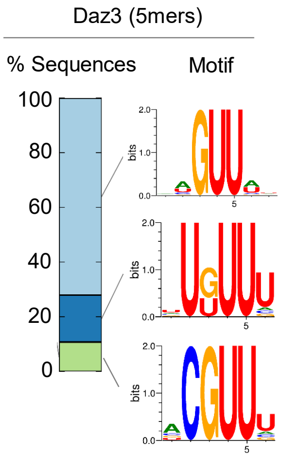


Figure R1 | Bind-n-Seq data for Daz3 from the Endcode project. The bar shows the distribution of sequence variants with the motifs indicated on the right (Figure adapted from: https://www.encodeproject.org/experiments/ENCSR449VKY/)

These data clearly raise the possibility that Dazl also shows high inherent selectivity for the GUU core motif, but definitive conclusions based on data with a Dazl ortholog might be premature. We have spent considerable effort to establish a Bind-n-Seq or a related approach for Dazl but have not been successful. This is because Dazl(RRM) has proven uncooperative in approaches based on gel separation or filter binding to separate bound vs. unbound substrate species. Although these experimental challenges are consistent with very high inherent selectivity (i.e. only few of the randomized substrate variants are bound, which is difficult to measure reliably), meaningful conclusions about the inherent binding selectivity of Dazl (*in vitro*) are premature until a reliable Bind-n-Seq approach for Dazl can be established.

While it remains to be shown what feature(s) determine(s) a given rate constant at a given binding site, our results suggest that the number of surrounding Dazl binding sites (i.e., Dazl clusters) impact association rate constants. We show this data in **Figure 3c** and **Extended Data Figure 5k**.

C) I was surprised to see that the distribution of binding constants is relatively narrow. This could be due to the initial selection of binding sites. Hence weaker binding sites might be underrepresented in the dataset. Including such weaker sites might also help to observe features that differ for high-affinity binding sites.

Our KIN-CLIP analysis includes > 90% of sites with corresponding sites in CLIP data (**Extended Data Figure 2d**). We had to exclude sites with low coverage values (< 5 reads at high laser power and high Dazl concentration at late timepoints, and sites without reads at low Dazl, low laser power, early timepoints time points), because it is not possible to calculate meaningful kinetic parameters from such sparse data.

Since we considered the vast majority of binding sites, it is unlikely that our analysis is markedly affected by binding site selection. We are confident that the resulting distribution of binding rate constants, which ranges over three orders of magnitude (**Fig.2e**) provides an unbiased view of the kinetic landscape. We nevertheless appreciate the reviewer’s comment and have now specifically noted in the manuscript that our KIN-CLIP analysis includes > 90% of sites with CLIP reads and is thus unlikely to be biased by binding site selection criteria.

3) Comparison of Dazl binding to the different transcript regions:

A) Introns are present mostly in the nucleus, whereas the other regions are dominant in the cytoplasm. Also, it is likely that Dazl concentration is very different for nuclear and cytoplasm. Will this affect calculation of the different constants?

Intracellular differences in Dazl concentration (e.g. nucleus vs. cytoplasm) is reflected in association rate constants (and parameters that include association rate constants, e.g. P, B and ΣB) and does not affect the calculation of rate constants.

B) The authors use Bowtie2, a mapper that is not splicing-aware. Hence will reads be systematically lost in the region of the ORF compared to the 3'UTR?

We used Bowtie2 to ensure back-compatibility of our data with previous published CLIP data for Dazl. Exon-exon spanning reads represent less than 0.1% of the total CLIP reads. The data are thus not unduly biased by using Bowtie2.

4) Dazl binds mRNA in clusters. It is interesting that clusters with more binding sites have higher k_{on} rates, but k_{off} rates are not affected. It would be nice to present some interpretation.

The scaling of association rate constants with the number of binding sites in clusters (**Fig.3c**) reflects a cooperative association process of Dazl protomers within a given cluster. That is, a bound Dazl protomer increases the binding of other protomers in the cluster. However, multiple bound protomers appear not to slow dissociation of Dazl from an individual binding site.

5) Clusters correlate with Dazl function. It is great to see that the cumulative Dazl binding probability correlates with the different regulatory classes (RNA and RPF). In my understanding the cumulative binding probability could maybe also be approximated with normal CLIP-seq data. Have the authors correlated cumulative binding probability with different scores for iCLIP signal normalized for expression (e.g. PureCLIP score, etc.)? This would give a better idea of the benefit of KIN-CLIP over normal CLIP. Or alternatively the authors could use the KIN-CLIP data to suggest how to use normal CLIP data in the best way. I think that will be very useful for the RNA community, where the majority of groups will only be able to perform normal CLIP experiments.

This is great point. It was our initial motivation for the presented work to obtain quantitative information from "conventional" CLIP approaches. However, the nature of the crosslinking process precludes a straightforward solution. This is because the extent of crosslinking for a protein at a given binding site depends on crosslinking time, crosslinking efficiency, association and dissociation rate constants and the accessibility of the binding for the protein over the course of the crosslinking experiment. These parameters cannot be deconvoluted in a single point measurement, neither absolutely nor relatively (e.g. comparing relative binding parameters for binding sites).

To illustrate this point, we have prepared a series of plots of KIN-CLIP (cumulate binding probability) and iCLIP parameters, normalized for expression (**Figure R2**). No significant correlation is seen between the iCLIP density, normalized to RNA expression levels, and binding probability (**Fig.R2a, b**), which is not unexpected. PURECLIP scores also show no correlation (not shown).

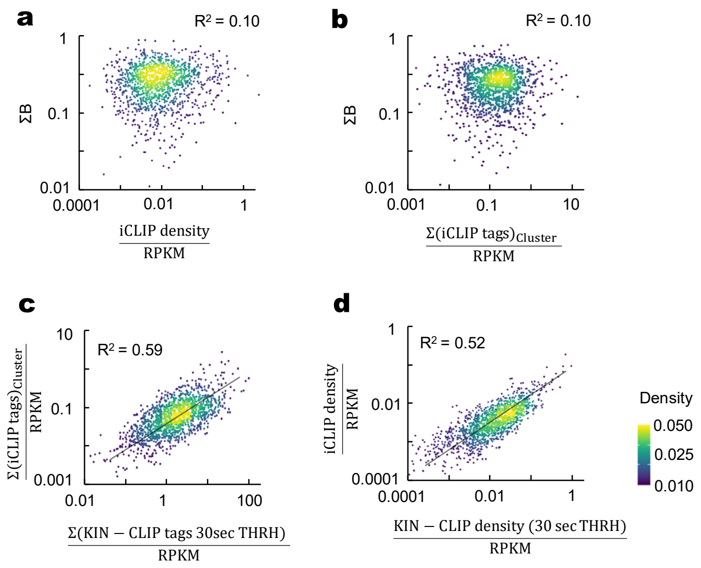


Figure R2 | Link between KIN-CLIP and iCLIP data. **a.** iCLIP read density normalized to mRNA expression level (X-axis) plotted versus binding probability calculated from KIN-CLIP data (Y-axis). No apparent correlation is detected. **b.** iCLIP read density as tags per Dazl-binding cluster, normalized to mRNA expression level (X-axis) plotted versus binding probability per Dazl cluster, calculated from KIN-CLIP data (Y-axis). No apparent correlation is detected **c.** Correlation between iCLIP read density as tags per Dazl-binding cluster, normalized to mRNA expression level, and KIN-CLIP read density as tags per Dazl-binding cluster, normalized to mRNA expression level for the 30s timepoint, at high Dazl concentration. **d.** Correlation between iCLIP read density per Dazl-binding site, normalized to mRNA expression level, and KIN-CLIP read density per Dazl-binding site, normalized to mRNA expression level for the 30s timepoint, at high Dazl concentration.

There is a correlation ($R^2 = 0.52$) between iCLIP read density and KIN-CLIP read density at high protein concentration and high laser power at the shortest timepoint (**Fig.R1c, d**). This KIN-CLIP density is linked but does not equal to the fraction of bound protein at steady state. At high crosslinking efficiency, protein bound to a given site is rapidly crosslinked. However, the KIN-CLIP read density is also influenced by binding probability and binding site accessibility, and therefore, the KIN-CLIP read density at high laser power cannot be directly interpreted in a mechanistically meaningful manner.

In addition to binding probability and binding site accessibility, iCLIP read density is influenced by the crosslinking rate constant. As we have shown, this rate constant varies for individual binding sites, but the range of this variation is comparably small (**Extended Data Figure 3i**). For this reason, we see the noted correlation between the iCLIP and the KIN-CLIP read density (**Fig.R2**). If the range of crosslinking rate constants for individual bindings sites would be larger, which is a possibility for other proteins, the correlation would diminish. For proteins other than Dazl, crosslinking rate constants for individual binding sites in cells are not known, and it is thus not possible to estimate to which extent iCLIP read densities of single point measurements are even residually reflective of binding probabilities.

Unless drastic simplifications of the crosslinking process are stipulated, such as equal or similar crosslinking efficiency for all binding sites and equal binding site accessibility, we do not see a possibility to use single point CLIP measurements to even semi-quantitatively estimate binding probabilities. Yet, as the correlation in **Fig.R2** shows, there can be quantitative information in conventional CLIP data, although one would have to present this with the appropriate caveats and stipulations.

Notwithstanding these considerations, it is in principle possible to obtain rigorous quantitative information from traditional CLIP approaches. Experiments would need to be conducted analogously to the KIN-CLIP approach – at multiple timepoints at different protein concentration and at different crosslinking efficiencies. Since the crosslinking rate constant would be slower than association and dissociation rate constants, those two parameters cannot be deconvoluted, but a quantitative binding probability (related to affinity) can be determined. This would be an extremely useful quantitative parameter for RNA-protein binding in cells. We have now noted this in the discussion.

Although space constraints preclude a detailed discussion of how our study can guide experiments aimed at obtaining quantitative information from conventional CLIP approaches, we agree with the reviewer that it will be very useful for the RNA community to outline principles of how to make conventional CLIP approaches quantitative. However, we feel this is best accomplished in a separate manuscript dedicated to this topic.

6) The Dazl code:

(A) I am not sure if in the current setup it can be really called cracking the code. I would tone this down.

We appreciate the reviewer's point. "Code" conveys certainty about an outcome based on a set of rules. Although we are confident that the set of rules we have delineated are a major step towards a "code" for Dazl, we have toned down the term accordingly. We now use "regulatory program", which does not suggest certainty, but still reflects the high degree of explanatory power of our model. We have changed the corresponding passages in the manuscript.

(B) The authors observe a nice correlation between the predicted and the experimental RPF/RNA changes. Did the authors control for overfitting? It would be important to leave out part of the data and use this for testing after fitting.

Yes, we have carefully controlled for overfitting. We kept 30% (N = 492) of data set for model cross-validation to assess overfitting, as well as over parametrization and to evaluate model quality. We realize that we did not explain the model part in sufficient detail, as also noted by reviewer 3.

We have now improved this section of the ms. We have updated the **Materials and Methods** section to describe model building, selection and validation (including controls for overfitting) in more detail, also in response to a comment by reviewer 3. We also have updated the plots in **Fig 4g, h** where we now show the test datasets, rather than the training sets and we provide Root Mean Prediction Error (RMPE) and adjusted R^2 values for test data (**Extended Data Figure 10, Fig 4g, h**). Finally, and also in response to the comment by reviewer 3, we now describe the steps of model building, refinement and selection in more detail in the **Materials and Methods** section, in **Extended Figure 10** and in additional schemes in the **Supplementary Materials (Suppl. Materials Figures S5-S7)**.

(C) I am confused about the data points shown in Fig. 4g,h. Shouldn't they be the same points as in Fig. 4b? However, the distribution is clearly shifted, with the majority of data points having values around 1.1 instead of 1.

We thank the reviewer for catching this. This was on error. The plots showed the training data set earlier. As noted above, we have now updated the plots and show the predictive power of our model using appropriate test data set unseen to the model.

(D) In the t-SNE analysis shown in Ext. Data Fig. 8d, the points from the different classes (HH, HM,…) completely separate, and there is not a single RNA going into the wrong cluster. I find this very unlikely, since the different categories (HH, HM,…) are not separated at all in Fig. 4b, and the boundaries between them are set more or less arbitrarily.

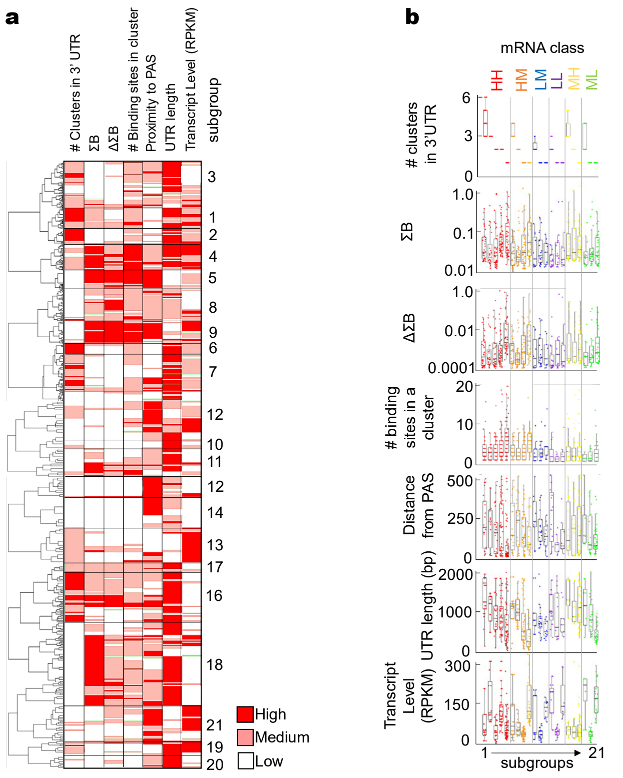


Figure R3 | Hierarchical clustering of mRNA features and KIN-CLIP parameters (a) and correlation of these parameters with the functional mRNA classes (b).

For the visual depiction shown in **Extended Data Figure 8** for t-SNE analysis, we first performed hierarchical clustering of the 7 characteristics that constitute the Dazl regulatory program and identified 21 subgroups (**Fig.R3**). We utilized a smoothened version of this hierarchical clustering, where we defined a subgroup based on the majority of a given characteristics in each subgroup. Accordingly, the RNA subgroups are completely separated in the t-SNE plot.

The reviewer's comment prompted us to re-think the smoothing approach and we have replaced the smoothened t-SNE plot in **Extended Data Figure 8d** with a t-SNE plot of un-smoothened data. This does not alter the clear emergence of 21 subgroups in the t-SNE analysis, while more directly reflecting underlying data.

7) To make the manuscript and the data accessible to the community, it will be important to make the code more accessible and well documented. Currently, it is hard to use. Also, it would be important to make the model for the fitting of the kinetic parameters available.

We annotated the code in more detail. We have also uploaded the code for fitting of the kinetic parameters available on Github. In addition, we are happy to provide more specific information upon request.

Minor

The authors could check, as a control, that the calculated rates do not correlate with RNA expression levels, since the crosslinking signal of individual transcripts correlates with RNA expression. Kinetic rates should be independent of transcript levels?

The rate constants do not correlate with RNA expression levels (**Fig.R4**). This is expected.

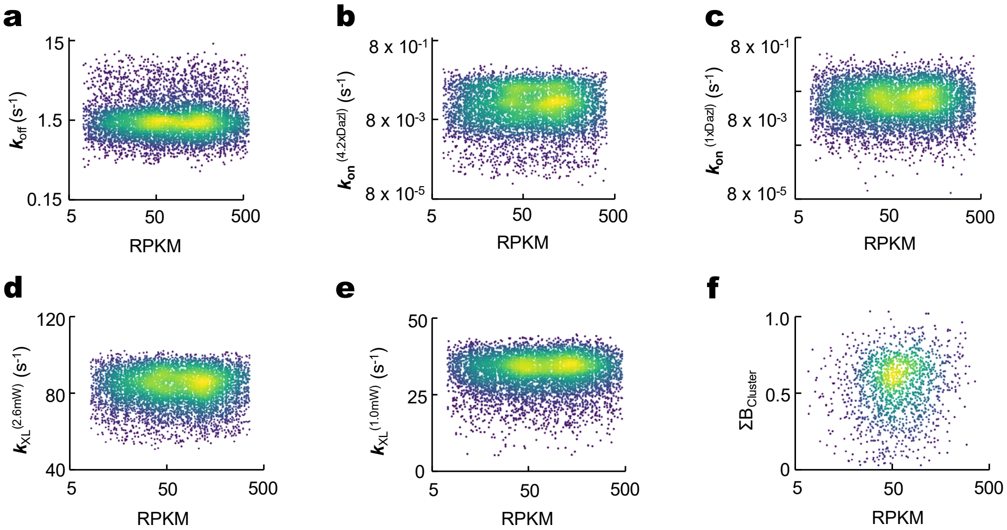


Figure R4 | Plots of rate constants (**a-e**) and binding probability (**f**) vs. mRNA level.

Fig. 2e: Are those really frequencies or rather densities? If it is frequencies, it would be good to indicate the bin size.

The curves show frequencies. We have added the bin size definitions in the Figure caption.

The colors in some of the figures are not suitable for color-blind people.

We thank the reviewer for pointing this out. We have updated the figures accordingly.

Consideration for normalization to crosslink signal: I think it is important that for this normalization the majority of reads map to the genome.

The majority of reads did map to the genome, although the actual numbers varied for the different libraries (**Suppl. Material Table S6**). However, it is not critical for the normalization that the majority of the reads map to the genome. This is because, as per our analysis, unmapped reads are predominantly adapter concatamers and other "artificial" reads that are generated during the cDNA library generation process. These reads are therefore not part of the crosslinked material. Accordingly, our normalization is for mapped reads, as those represent the physically crosslinked RNA. We note that each KIN-CLIP experiment (e.g. each crosslinking timepoint and condition) is normalized to its own bulk-crosslinking parameter, which were determined in multiple replicates.

Also, I wondered if for the higher protein concentration all material is pulled down. Otherwise, this normalization might be distorted.

Complete pulldown of crosslinked material and in fact of all Dazl, crosslinked and non-crosslinked was carefully optimized. The exact conditions are given in the ms.

Is there a difference in library sizes between the datasets? What are the overall read numbers? In this context, it would be good to have scales for the y-axes in the genome browser shots in Fig. 2b. Also, there is not information of the iCLIP only libraries generated, with details in the number of reads obtained.

There are differences in the library sizes. Overall read numbers are given in Supplementary Materials (**Supplementary Material Table S5**). **Fig.2b** shows identical scales for the normalized reads. We added the scale information in the figure caption (Normalized coverage = 11 for all traces). We have added the sequencing statistics information for the iCLIP libraries as **Supplementary Material Table S7**.

Is there any characteristic feature in the sites only recovered by iCLIP/KIN-CLIP? It would be interesting to see whether the sites that appear exclusively in iCLIP are lost in the KIN-CLIP due to a refinement of the crosslinking or just experimental variance.

We appreciate this question. We had not looked at potential differences in the features of sites unique in either the iCLIP or the KIN-CLIP data sets, but have now interrogated the respective datasets (**Fig.R5**).

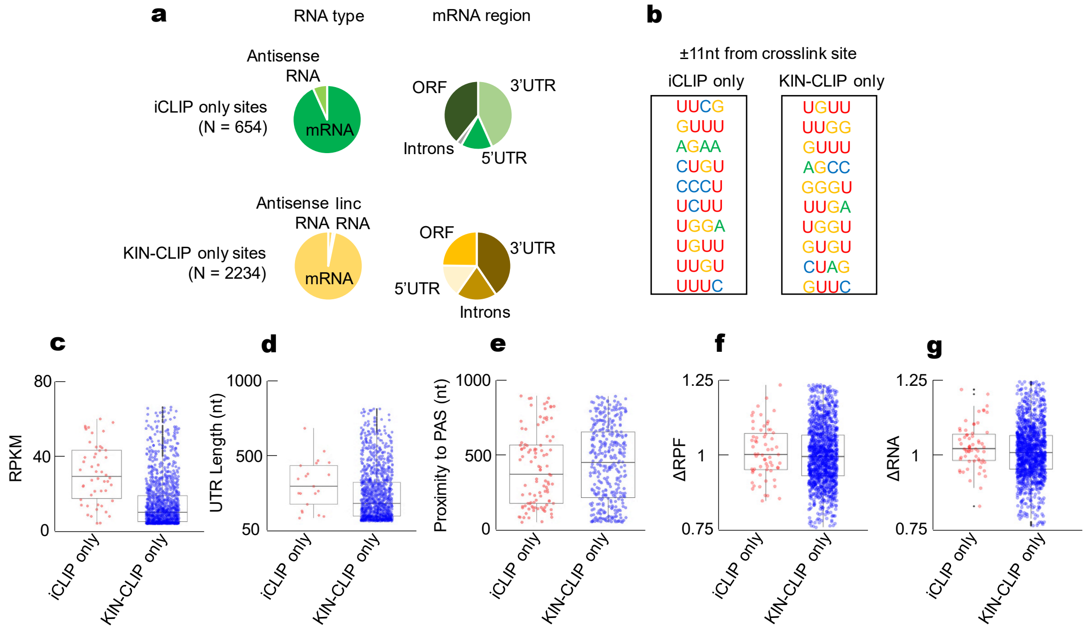


Figure R5 | Characteristics of Dazl binding sites for site detected only in iCLIP or in KIN-CLIP datasets. a. RNA classes and mRNA regions for iCLIP only and KIN-CLIP only binding sites. **b.** Most frequent sequence motifs at iCLIP only and KIN-CLIP only binding sites. **c-g.** Site features for iCLIP only and KIN-CLIP only binding sites. All plots were generated as described in the Materials and Methods Section.

The majority of reads for both datasets is in mRNAs, a smaller percentage in non-coding RNAs (**Fig.R4a**). In mRNAs most reads are in 3'UTRs, and the most notable difference is a larger fraction of reads in introns in the KIN-CLIP only dataset (**Fig.R4a**). A higher fraction of binding in introns is consistent with higher crosslinking efficiency, which enables the capture of binding sites in RNA regions that do not accumulate at high levels and might escape detection by conventional crosslinking. We also analyzed sequence characteristics of both datasets at the crosslinking site and detected no significant differences (**Fig.R4b**). We finally interrogated mRNA features for the datasets (**Fig.R4c-g**). Here we find that mRNAs in the KIN-CLIP only dataset have a markedly lower expression level, compared to the RNAs in the iCLIP only dataset **(Fig.R4c**). This observation is perhaps expected, given the higher crosslinking efficiency of the fs-laser crosslinking approach, which allows capture of crosslink for less abundant RNAs. Other RNA features do not significantly differ between the datasets (**Fig.R4d,e**). We also do not see differences in the datasets with respect to impact of Dazl on mRNA expression levels (ΔRNA) and translation state (ΔRPF) (**Fig.R4f**, **g**). In sum, the analysis of the iCLIP only and KIN-CLIP only datasets reveals no notable differences, aside from the perhaps expected greater sensitivity of the fs laser crosslinking approach. We can therefore conclude that sites absent in either KIN-CLIP or the iCLIP dataset are not lost due to refinement of the crosslinking but rather due to experimental variance.

The nucleotide color code in Ext. Data Fig. 4 does not match with the standard (A-Green, C-Blue, G- Yellow, T- Red).

We have changed the color code accordingly.

There is no information of the iCLIP-only libraries generated, with details in the number of reads obtained.

We have included this information for the iCLIP libraries as **Supplementary Material Table S7**.

Referee #2:

This manuscript introduces and applies what could be a tremendously important new method for monitoring the kinetics and equilibria (or steady state) of intermolecular interactions in cells. The method, deceptively simple, is to monitor the time dependence of UV crosslinking from a femtosecond pulsed laser, with the increase in crosslinked product reflecting a combination of the rate constants for association and dissociation as well as the rate of crosslinking from the bound species. After introducing the method in vitro using the RNA-binding proteins RbFox(RRM) and Dazl, the authors use an antibody pull-down and next-gen sequencing to monitor the kinetics of RNA binding by Dazl on a genome-wide scale. From the in vivo assays, the authors report a number of new discoveries. There is substantial variability in the binding rate constants for different Dazl binding sites, exceeding the variability in dissociation rate constants. Further, the binding rate constants do not seem to track with sequences but instead track together with neighboring sites, suggesting that there are major differences in RNA accessibility that impact binding frequencies. The majority of sites are not saturated with protein, and individual binding events are relatively short-lived (~1 s), indicating tremendous potential for Dazl binding to respond rapidly at global or local levels to changes in conditions, RNA accessibility, or Dazl expression level. Dazl binding sites appear to be organized in clusters, and the authors use their binding data to develop a model in which the frequency of Dazl binding to at least one site within a cluster is tightly linked to the regulatory effects of Dazl on translation and/or RNA decay.

The work builds in interesting ways on previous work of others that mapped Dazl binding sites genome-wide and probed the effects of Dazl binding. The overall presentation of the work, in clarity, organization, and economy, is outstanding.

Although the biological insights are significant, the most enduring value of the work likely lies in the introduction of the time-resolved UV crosslinking method. The method can be applied in a straightforward way to any RNA-binding protein and presumably to DNA-binding proteins as well. In my opinion, it is likely to be a game changer for quantitative research in protein-nucleic acid interactions in cells. This is probably the most important paper I have read this year.

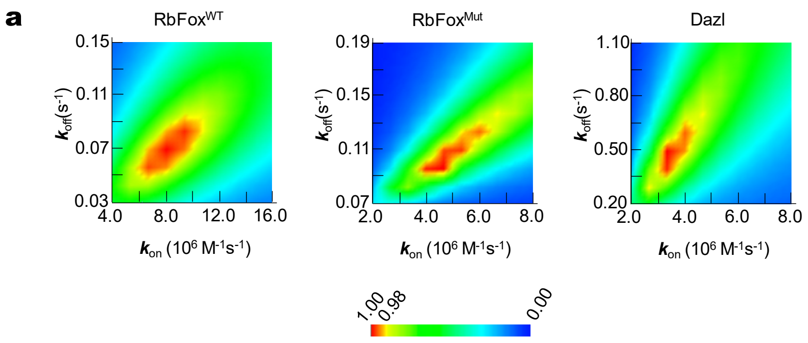
Overall, the authors should be commended for this terrific idea and for bringing it to fruition while also going most of the way toward establishing and benchmarking a robust infrastructure that will serve as a blueprint for future applications of the method. Importantly, the authors varied both the Dazl expression level and the crosslinking power, in addition to the exposure time, to establish constraints on the three rate constants that govern the overall behavior. They provide considerable detail on the methods of the analysis and the outcomes. The strategy and all of this information will be tremendously valuable for other who seek to apply the method to their own systems.

Nevertheless, there are some points about the method that are not completely clear. Although the method is conceptually straightforward, it is not at all trivial to perform the experiments in a way that defines the binding rate constant and particularly for the dissociation rate constant, even in the simple experiments with just one target in vitro. If the crosslinking rate constant (kxl) is much smaller than the dissociation rate constant (kdiss), the measurement will report on the binding equilibrium but not the rate constants. If kxl is much larger than kdiss, the measurement will report on the pseudo-first order binding rate constant (kon) or kxl, whichever is smaller, or some combination of the two, but it will not give direct information on kdiss, as dissociation does not happen to a significant extent. Perhaps information on kdiss is provided indirectly, because a fast phase of crosslinking could reflect protein that was already bound at time zero and therefore define the binding equilibrium, while the slower phase could define the binding rate constant of additional protein. With both the equilibrium and the kon values defined, the measurements would constrain kdiss. It is also possible that measurements under conditions such that kxl is similar in magnitude to koff would provide information that would define koff. As described further below, in some cases it is not clear from the manuscript how the measurements constrain the rate constants, through the scenarios described above (or perhaps in other ways?).

We appreciate the reviewer’s comment. Subsets of experiments provide only compound parameters, as the reviewer notes. However, the rate constants are linked to each other in a predictable manner, as outlined in the Materials and Methods section. Through variation of reaction conditions, the individual rate constants can be calculated. Key for this analysis is the variation of protein concentration and crosslinking power in a manner that each condition results in a timecourse that is sufficiently different from the timecourses under the other conditions. For example, if the crosslinking rate constant is too large at all laser powers, no differences in the timecourses would be detected and it would not be possible to calculate sufficiently constrained rate constants. In our case, timecourses at all conditions *in vitro* show sufficient differences for all conditions for the tested proteins, and well constrained rate constant can thus be calculated, as outlined in more detail in the responses to the following comments.

1. In vitro measurements: There is currently insufficient detail about how (and how well) the kinetic parameters are constrained in the in vitro experiments. Although the equilibrium values are benchmarked against another method, the values of the rate constants are not. While the addition of an alternative method to benchmark the kinetics measurements would be one way to strengthen the work, in my opinion it is not essential. However, in the absence of another method, it is especially critical to show how the measurements provide reliable values of the rate constants. The fit space profiles are included in Fig. S1, but these are 1D explorations of space and do not rule out the possibility that the measurements constrain the binding equilibrium but not the rate constants, which might vary together over a large range without impacting the fit. One possibility would be to evaluate the 2D fit space and include the results of co-varying kon and kdiss.

We see the reviewer's point and appreciate the suggestion. We have evaluated the 2D fit space (**Fig.R6**) and find that rate constants are sufficiently constrained in an independent manner.



b

Protein	Parameter	Value	Lower bound	Upper bound
RbFox1^{wt}	k_{on} (10^6 M^{-1}s^{-1})	8.2 ± 1.8	6.4	10.0
	k_{off} (s^{-1})	0.07 ± 0.02	0.05	0.09
RbFox1^{mut}	k_{on} (10^6 M^{-1}s^{-1})	4.6 ± 2.0	3.7	6.3
	k_{off} (s^{-1})	0.11 ± 0.05	0.08	0.14
Dazl	k_{on} (10^6 M^{-1}s^{-1})	3.8 ± 1.2	2.6	5.0
	k_{off} (s^{-1})	0.49 ± 0.12	0.35	0.61

Figure R6 | 2D fit space analysis. a. Relative X^2 of co-varying k_{on} and k_{off}. Both rate constants are constrained for all 3 proteins with a well-defined local minimum (red). **b**. Values for lower and upper bounds obtained with the analysis.

The results rule out a scenario where both rate constants vary together over a large range without impacting the fit. We have now included these data in the **Supplementary Material Figure S2**, together with the 1D Fitspace analysis.

We have also measured the dissociation rate constant for RbFox^{WT}(RRM), which is possible manually through pulse chase experiments monitored by PAGE (**Fig.R7**). The measured dissociation rate constant is similar to the rate constant obtained by crosslinking, providing additional evidence that rate constants obtained by time-resolved crosslinking provide reliable parameters. It is unfortunately not possible to perform similar measurements for Dazl with available means, since it has not been possible to establish conditions for gel shifts for the protein.

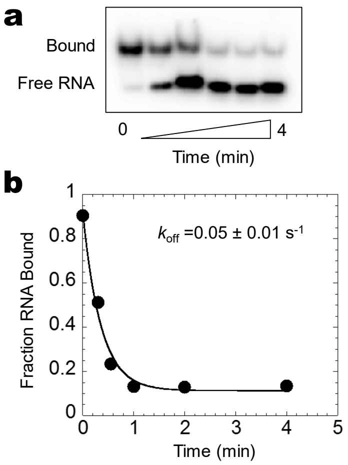


Figure R7 | Direct measurement of RbFox^{WT}(RRM) dissociation rate constant. a. PAGE for pulse chase reaction of and RbFox^{WT}(RRM) dissociation timecourse. Bound RNA (radiolabeled) was identical to that used in **Fig.1**, chase RNA was identical to the bound RNA, but not radiolabeled. Substrate RNA (final concentration 1 nM) was incubated with RbFox^{WT}(RRM) (final concentration 20 nM) for 30 minutes. Chase RNA (final concentration 1 μM) was added, and aliquots were removed at the indicated times, stored on ice and subsequently loaded on 8% non-denaturing PAGE **b**. Timecourse of the reaction in panel **a**. The solid line marks a fit against the integrated first order rate constant, the error range indicates the fitting error.

It would also be very helpful to include as a supporting figure all of the time traces and the accompanying fits by the simulation. They are currently shown in Fig. 1 for the wild-type RbFox(RRM) but not for the other two proteins.

We show these time traces and the corresponding fits in **Extended Data Figure 1g, h.**

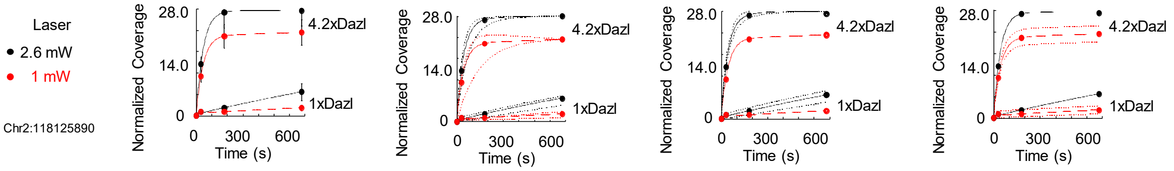
In addition, there is not much detail on how the global fitting by simulation was done. Does the simulated crosslinking begin from a pre-equilibrated mixture of protein and RNA (as in the experiment)? As noted above, under some conditions this might give clear fast and slow phases, which could be quite informative.

We apologize for the lack of detail on this topic. We have now included a more detailed description of the fitting procedure in the Materials and Methods Section. The global datafit was performed as in the experiment, starting from a pre-equilibrated mixture.

2. In vivo measurements: For these measurements, there is a lengthy and thorough section describing the analysis, and the methods and use of statistics and uncertainties seems to be appropriate. Still, it would be helpful to see representative examples of the data for a few individual sites presented in the same format as the in vitro binding measurements (as in Fig. 1d) and to get a more intuitive sense of how the measurements constrain the rate constants. These plots could be shown with a simulation curve overlaid using the determined rate constants. For many of the sites, it seems that the crosslinking rates are much greater than the dissociation rates, perhaps even at the lower laser power, and it is not clear to me how the measurements define kdiss.

We agree with the reviewer. Additional representative examples for individual binding sites are beneficial. Space limitations prevent us from showing these data in the extended Data Figures and we have therefore added a Supplementary Materials Figure (**Supplementary Materials Figure S4**) that displays timecourses (with the associated data fits) for 15 additional examples. We picked the examples for a low, medium and high value of each of the respective parameters, to emphasize how these parameters are reflected in the experimental data. We feel these examples are an instructive and nicely intuitive way to assess the link between experimental data and kinetic parameters, and we appreciate the suggestion to include these data.

In addition, we illustrate the impact of varying parameters on a specific example (**Fig.R8**). The data show variation of the timecourses for a range of dissociation and crosslinking rate constants, illustrating that the measured timecourses allow the determination of the dissociation and crosslinking rate constants.



Parameter	Site : Thbs1	Simulation 1	Simulation 2	Simulation 3
$k_{on}^{(4.2xDazl)}$ (s^{-1})	(1.8 ± 0.4) x 10^{-2}	(1.8 ± 0.4) x 10^{-2}	(1.8 ± 0.4) x 10^{-2}	(1.8 ± 0.4) x 10^{-2}
$k_{on}^{(1xDazl)}$ (s^{-1})	(0.8 ± 0.2) x 10^{-3}	(0.8 ± 0.2) x 10^{-3}	(0.8 ± 0.2) x 10^{-3}	(0.8 ± 0.2) x 10^{-3}
$k_{diss.}$ (s^{-1})	1.2 ± 0.5	*0.15 -15.0*	1.2 ± 0.5	1.2 ± 0.5
$k_{xl}^{(2.6mW)}$ (s^{-1})	90.2 ± 18.2	90.2 ± 18.2	*40 – 120*	90.2 ± 18.2
$k_{xl}^{(1.0mW)}$ (s^{-1})	27.4 ± 8.3	27.4 ± 8.3	27.4 ± 8.3	*4 - 40*

Figure R8 | Impact of rate constant variation on timecourses. The left panel shows the experimental data. The other panels show simulated timecourses (dotted lines) with changed dissociation and crosslinking rate constants, as indicated in the table.

We note that a certain number of rate constants are outside the range where they can be properly constraint. We have marked the corresponding confidence ranges in **Fig.2**, and **Extended Data Fig.3**, and described in the Materials and Methods how we calculated the confidence ranges.

3. Laser power and crosslinking rate: For the in vitro experiments, when the laser power is increased 2.6-fold, the measured crosslinking rates increase by more than 40-fold to >2 s(-1). Is that expected? For the in vivo experiments, the crosslinking rates are much larger than those in vitro, with the peak at the higher laser power centered at ~100 s(-1), 40-fold(ish) greater than the same laser power in vitro. Is that expected? I am also a bit confused about how the crosslinking rate can really be constrained at 100 s(-1) when the first time point is taken at 30 s. But perhaps this comes from information about the binding equilibrium, which together with the crosslinking rate may define the observed crosslinking time dependence under some conditions.

The non-linear increase of the crosslinking rate constant with the laser power is expected because the fs-laser crosslinking is a multi-photon process, which scales non-linearly with power. Re. constraints of the crosslinking rate constants at the chosen timepoints: the reviewer is correct; the corresponding information comes from the binding equilibrium. The newly included **Supplementary Material Figure S4**, which shows more examples of crosslinking timetraces and **Fig.R8** (above) further clarify this point.

Minor points

1. Line 105, “Association and dissociation rate constants varied by several orders of magnitude.": Based on Fig. 2e and elsewhere, it seems like the binding rate constants vary quite a bit more than the dissociation rate constants, which is quite interesting, and it may be worth re-wording this sentence to include the difference.

We have re-worded the passage accordingly.

2. Lines 151-153: The statement that the fractional occupancy of sites within a cluster often trended together is quite interesting because it suggests that the context of the sequence – i.e. which cluster it is part of – is as important or more important for binding than the RNA sequence, at least the sequence beyond the GUU motif. I wonder if this point can/should be expanded in a sentence or two? It certainly seems like an important topic for future work.

We have re-worded the passage accordingly.

3. I am confused about Ext. Data Fig. 6b. How does the cumulative binding probability exceed 1? I must be misunderstanding something, but I don’t know what. Also, is there a normalization for cluster size in this analysis? If not, it would seem that the correlation of binding probability with proximity to the polyadenylation site may arise because those closer to this site tend to have more binding sites per cluster.

The cumulative binding probability, i.e., the addition of the binding probabilities for each binding can exceed 1. Values greater than 1 indicate that more than 1 Dazl is bound at all times in a given cluster. Theoretically, there is no upper limit, and multiple Dazl could be bound at any given time. For a single binding site, however, the binding probability cannot exceed 1.

In the analysis in Fig.6, no normalization for cluster size was included. The reviewer is correct that clusters more proximal to the PAS contain on average more binding sites. We directly show this correlation in **Extended Data Fig.6a.** However, the number of binding sites in a cluster and the proximity to the cluster to the PAS are not redundant. Not all clusters that are proximal to PAS have a high number of binding sites and *vice versa*. In addition, cumulative binding probability can be high for clusters with few binding sites, and low for clusters with many binding sites. The non-redundant impact of number of binding sites and PAS proximity for Dazl function is also apparent in the regression models for Dazl function (**Extended Data Fig.10b**) where we show how systematic removal of parameters diminishes the predictive power of the model. These observations mean that both, number of binding sites and proximity to the PAS contribute independently to Dazl function, despite their correlation.

4. P. 31, halfway down, “kxl(1 mW) and kxl(2.6 mW) were then averaged.”: Maybe this is a typo? The crosslinking rates for the two laser powers should definitely be different from each other, and indeed are reported as being quite different from each other, so it does not seem appropriate or useful to average them.

Thanks for pointing this out. We have corrected the passage, which now reads: “Timecourses at 1xDazl and high laser power (2.6mW) and low laser power (1mW) were fit separately, yielding average initial values for kon(1xDazl) and kdiss and initial values for kXL(1mW) and kXL(2.6mW).”

Also on p. 31, a few lines up, it seems to indicate that the starting estimates of the crosslinking rates were 1 s(-1) for 2.6 mW and 10 s(-1) for 1 mW. I am guessing that it was actually the reverse and that this is a typo.

Thanks for noticing. This was a typo. We have corrected it.

Referee #3:

This manuscript from Sharma et al. is an important piece of work that will be of broad interest to the RNA community. Here, the authors developed a second-generation CLIP method called "KIN-CLIP" that makes use of a powerful laser to rapidly crosslink protein and RNA. By using this laser, the authors can infer kinetic binding values from CLIP libraries, allowing them to gain access to biochemical coefficients of protein–RNA interactions in living cells. Similar strategies have recently been applied to the microRNA field, but this is the first such approach to investigate RNA-binding proteins and to define biochemical coefficients purely from in vivo measurements. The authors apply this method to the RNA-binding protein DAZL, which is important for gametogenesis and mediates post-transcriptional regulation. Using their method, the authors find that DAZL dissociates quickly, sites spend most of their time unbound by DAZL, and features of efficacious sites. The authors combine features into a linear regression model to predict impact of binding on gene expression. This method will be very important to many in the community and opens up new areas of RNA biology. Nonetheless, I have some concerns about the paper (especially surrounding figure 4) that need to be addressed before I can support publication:

Major comments:

1. Does the scaling of association rate constants with the number of binding sites occur in vitro? More generally, how much of the differences in binding between different clusters can be recapitulated in vitro with just the RNA and DAZL (as opposed to binding being influenced by other proteins and the cellular milieu)?

We have not examined whether the scaling of association rate constants with the number of binding sites occurs *in vitro*, because this is not currently feasible, as it has not been possible to purify recombinant full-length Dazl. The protein, like many other RBPs, contains low complexity, most likely unstructured C and N-termini that cause irreversible aggregation in attempts to purify the protein, which we have encountered. The recombinant Dazl RRM, which we did test *in vitro* (**Fig.1**), is unlikely to show the cooperativity in the association process seen in the cell.

2. In Fig. 4, the analyses should be repeated, but this time looking at directly ΔTE. As I'm sure the authors are aware, RPF values are mix of translational changes and RNA changes (as evidenced by the strong correlation between ΔRNA and ΔRPF in Fig. 4b); this makes analysis with RPF values hard to interpret. For instance, how much of the changes in RPF values with high and low Dazl are explained by changes in RNA levels? Are there any other features that explain changes in TE? To put it another way, currently there is limited evidence that consider RPF changes in addition to RNA changes meaningfully impacts their predictive models.

We appreciate the reviewer's point and are grateful for the suggestion. We have repeated the analysis with the ΔTE values, which isolates and thus emphasizes the role of Dazl in translation regulation. We have included the new data in **Extended Data Fig.10**. The analysis with the ΔTE values reveals statistically significant differences in the ΔTE values for different mRNA groups of the Dazl regulatory program within a given mRNA class. These differences are not automatically expected within each mRNA functional class. The ΔTE analysis thus provides further, independent functional validation of the Dazl regulatory program. We have also established the regression model based on ΔTE values, which we show as well in **Extended Data Figure 10**, (due to space constraints in the main figures). The correlation coefficient is somewhat lower than for the ΔRNA and ΔRPF values, but this is expected, given that ΔTE values are compound parameters calculated from both, RNA and RPF values.

We would like to retain our analysis using ΔRNA and ΔRPF values to emphasize the impact of Dazl on both, translation and RNA level. As the reviewer notes, these two effects are difficult to disentangle for certain scenarios (e.g. increase or decrease in both, RPF and RNA), although they are clear for other cases (e.g. no change in RNA but changes in RPF, no change in RNA – but changes in RPF). Using just ΔTE foregoes the impact on RNA levels.

We have amended the text, highlighting the ΔTE analysis.

3. In considering poly(A) sites, how were genes with multiple 3' UTR isoforms dealt with?

For transcripts with multiple 3’UTR length annotations, preference was given to annotation obtained from polyA-Seq (Ref.17). In cases when polyA-Seq annotation for a given 3’UTR was absent, coordinates with the longest 3’UTR annotation were utilized. We have noted this information now in the Materials and Methods Section.

4. Much of the model building in Fig. 4 is hard to understand. For example, how much data was held back to test the regression model? Similarly, how was the goodness of the models measured? Ext. Data Fig. 10 suggests that M1 was chosen because all seven features were significant, but with the underlying methods missing, this is very hard to interpret.

We appreciate the reviewer’s point and agree that the information on the model building and selection was insufficient. We have markedly expanded the description of the model building in the Materials and Methods Section and have included two additional Supplementary Figures (**Supplementary Material Figures S5, S6**) outlining the model building and evaluation process along with model quality control data.

5. A major weakness of the paper is that the authors do not test their predictions of DAZL binding and effects with reporters or alternative measurements.

We have now included data for six luciferase reporter constructs (**Extended Data Fig.10h**). Each of these reporter constructs contains the 3’UTR of a different mRNA that vary in their scores for the of Dazl effect on ΔRPF, which is a proxy for change in protein production. The change in luciferase activity measured with the reporters correlates very well with the predicted change in ΔRPF (R^2 = 0.8, **Extended Data Fig.10h**). These data provide an independent test of our model through the measure of luciferase activity with non-endogenous RNAs. We note that the reporters were used and measurements were performed in a previous study (Ref.17).

Minor comments:

1. The use of several abbreviations (like HH) is not intuitive for the reader and make Fig. 4 challenging to interpret.

We appreciate the reviewer’s comment. We see that H, M, L can be confused with the same designation for the group characteristics. We have changed the mRNA class labels to a four letter “code”: T for ΔRPF and R for ΔRNA and H – high, M- medium, L – low, for the corresponding changes (e.g. THRM). We have also color coded the labels. We trust this change makes interpretation easier.

2. The color choice in the top panel of Fig. 4d is not intuitive (e.g., a graded color scheme would be much better), and additionally this scheme, and others, does not use an accessible palette.

We thank the reviewer for pointing this out. We have changed the color schemes in **Figure 4** and others accordingly.