Published in final edited form as:

Nature. 2021 March ; 591(7848): 152–156. doi:10.1038/s41586-021-03222-x.

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, Case Western Reserve University, Cleveland, OH 44106

[(2)]Department of Biochemistry, School of Medicine, Case Western Reserve University, Cleveland, OH 44106

[(3)]Department of Chemistry, Case Western Reserve University, Cleveland, OH 44106

[(4)]Department of Physics, Case Western Reserve University, Cleveland, OH 44106

[(5)]Case Comprehensive Cancer Center, School of Medicine, Case Western Reserve University, Cleveland, OH 44106

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 kinetic

[(6)]Corresponding authors: Donny D. Licatalosi: ddl33@case.edu; Eckhard Jankowsky: exj13@case.edu.

CONTRIBUTIONS

D.S., C.E.C.-H., D.D.L. and E.J. conceptualized the study. C.E.C.-H. and M.M.B. adapted the pulsed fs laser setup for time-resolved crosslinking experiments. D.S. and M.M.B. optimized and performed fs laser crosslinking experiments. D.S. and L.L.Z. optimized and performed NGS library preparations. L.L.Z and D.D.L. performed iCLIP, RNA-seq and ribosome profiling experiments. X.Y. optimized Dazl(RRM) overexpression, purified recombinant Dazl(RRM) and performed fluorescence anisotropy studies. D.S. and E.J. devised the framework for KIN-CLIP data analysis and the Dazl regulatory program. D.S. performed the data analysis. All authors contributed to the writing of the manuscript.

CODE AVAILABILITY

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

COMPETING INTERESTS

Deepak Sharma and Eckhard Jankowsky are founders of Bainom Inc. None of the other authors have conflicts to declare.

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]. RBP binding and dissociation kinetics have 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], but inaccessibility of binding and dissociation kinetics of RBPs in cells limits 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 and show how these kinetic parameters inform a quantitative understanding of the cellular function of Dazl.

Time-resolved laser crosslinking

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, calculation of rate constants requires a sufficient number of experimental constraints. These can be established by measuring crosslinking timecourses at different protein concentrations and different crosslinking efficiencies (Fig.1b), while ensuring that crosslinking rate constants are 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), which had been shown to efficiently photo-crosslink proteins to DNA through multi-photonic excitation [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). RNA degradation with the fs laser was reduced, 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 than for the steady-state UV source, fewer photons are absorbed by the RNA over a given time (Extended Data Fig.1c), because fs pulses are emitted only once per millisecond and the cross-section for multi-photonic absorption is smaller than for single-photonic absorption with a steady-state UV source [13].

Crosslinking of the 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). We determined binding, dissociation and crosslinking rate constants for RbFox(RRM)-RNA binding from crosslinking timecourses at different laser powers and 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 published values [14]. We next determined binding, dissociation and crosslinking rate constants for a mutated RbFox^{mut}(RRM) [15] and

for the RNA binding protein Dazl(RRM) [16] (Fig.1e, Extended Data Fig.1g,h). RNA affinities of these two proteins, calculated from rate constants, were also similar to affinities measured with fluorescence anisotropy (Fig.1e, Extended Data Fig.1j,k). The data with three RBPs 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). 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 bulk 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 (Extended Data Fig.2d–g, ref.[17]). These data show that fs laser crosslinking maintains the characteristics of crosslink sites seen with steady-state UV-crosslinking.

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 (Fig.2a, Supplementary Material Table S6). This normalized read coverage was used to calculate a dissociation rate constant ($k_{diss.}$), observed association rate constants at low and high Dazl concentration ($k_{on}^{(1xDazl)}$, $k_{on}^{(4.2xDazl)}$) and crosslinking rate constants for both laser powers ($k_{XL}^{(1\ mW)}$, $k_{XL}^{(2.6\ mW)}$) for each binding site. (Fig.2b, Extended Data Fig.3a–k). Obtained rate constants faithfully described the experimental data (Fig.2b, 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 1xDazl were lower than those at 4.2xDazl (Fig.2c). 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.2d). Association rate constants varied to a larger degree than dissociation rate constants (Fig.2d). 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 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.2d), 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 $\tau_B < 1s$ (Fig.2d). Binding events are infrequent and even at high Dazl concentrations occur rarely more than six times per minute (Fig.2d). Accordingly, the probability of Dazl to be bound at any time is less than 10% for many binding sites (Fig.2d), indicating 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 *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.2d, 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 (Extended Data Fig.4w). Association rate constants and binding probabilities were higher for binding sites in 3'UTRs than for sites in 5' UTRs, introns and ORFs, and higher in mRNAs, compared with other RNA classes (Extended Data Fig.4q,r,u,v). The maximal fractional occupancy of binding sites did not significantly vary in the different mRNA regions, but was higher in mRNA, compared with other RNA classes (Extended Data Fig.4t,x). 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 mRNA 3'UTRs 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 bound mRNAs. 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 patterns of moderate correlation (Extended Data Fig.5k). 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 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 binds in a 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 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 its function.

A Dazl regulatory program

To delineate the connection between Dazl binding kinetics and Dazl impact on mRNA function, 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 linked to 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 and provide a framework for obtaining quantitative, steady-state protein-RNA binding information from CLIP with conventional crosslinking sources. Combining time-resolved fs laser crosslinking and kinetic analysis might also 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 Dazl-free 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 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], 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. Finally, cellular kinetic data allow the decoding of a complex link between Dazl-RNA-binding patterns and Dazl function. Because our experimental and data analysis approaches are applicable to other RBPs, KIN-CLIP provides a blueprint for delineating regulatory programs for other RBPs.

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 5.2 (GE) software. The fraction degraded RNA (Frac D) at each time point was calculated according to:

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

(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]

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

(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 4.1.1. (Synergy, PA) [16].

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

(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:

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

(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 10.0.200513 (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 $\boldsymbol{k}_{on}$, $\boldsymbol{k}_{off}$ and $\boldsymbol{k}_{xl}$. 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 ($\boldsymbol{k}_{on}$, $\boldsymbol{k}_{off}$ and $\boldsymbol{k}_{xl}$) varied individually (1D fit space, Supplementary Material Fig.S2a–c) and for co-variations of $\boldsymbol{k}_{on}$ and $\boldsymbol{k}_{off}$ (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 (ATCC, cat # CRL-2053) 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]). Cells tested negative for mycoplasma contamination with the MycoAlert Mycoplasma Detection System (Lonza, cat # LT07-118) and the MycoAlert Assay Control Set (Lonza, cat # LT07-518) as a positive control. 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 1.8.0.). Lane background was used to normalize the band intensities.

KIN-CLIP read processing, refinement and mapping

Raw sequencing reads were assessed for quality (FastQC 0.11.9, 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 0.0.13 (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 0.0.13 (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 2.4.2 [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 2.8x [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 (1635.2) and intersected individually with KIN-CLIP binding site coordinates using Bedtools (2.29.1)

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 Compseq 6.0.0 [44], R package 'randomizeR 2.0.0' [45] and 'Random' [46] module in Python 3.9.0. To generate z-scores, shuffled control sets were generated for each dataset analyzed using random 3.9.0 available in Python 3.9.0 (Shuffle N = 10,000). Enriched motifs were visualized using Weblogo 2.8.2.

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 (for details on both approaches see Supplementary Information). Parameters from both methods were averaged for subsequent data analysis (Extended Data Fig. 3).

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:

$$P_{(4.2\text{xDazl})} = \frac{k_{\text{on}}^{(4.2\text{xDazl})}}{k_{\text{diss.}} + k_{\text{on}}^{(4.2\text{xDazl})}} \quad \text{(Eq.5)}$$

$$P_{(1\text{xDazl})} = \frac{k_{\text{on}}^{(1\text{xDazl})}}{k_{\text{diss.}} + k_{\text{on}}^{(1\text{xDazl})}} \quad \text{(Eq.6)}$$

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. $\Phi^{max} = 1$ 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:

$$\Phi^{max} = \alpha^{max} \cdot L^{-1} \quad \text{(Eq.7)}$$

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 $[Dazl]^{rel}$:

$$\alpha^{max} = \alpha^{(4.2xDazl)} \cdot (P_{(4.2xDazl)})^{-1} \quad \text{(Eq.8)}$$

$\alpha^{(4.2xDazl)}$: 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:

$$k_{on}^{(Dazl)} = k_{on}^{max} \times \frac{[Dazl]^{rel}}{[Dazl]^{rel} + K'} \quad \text{(Eq.9)}$$

($k_{on}^{(Dazl)}$: $k_{on}^{(1xDazl)}$, $k_{on}^{(4.2xDazl)}$, K': apparent relative binding constant)

The binding probability extrapolated to $[Dazl]^{rel}$ saturation (P_{max}) is:

$$P_{max} = \frac{k_{on}^{max}}{k_{diss.} + k_{on}^{max}} \quad \text{(Eq.10)}$$

and

$$\alpha_{max} = \alpha_{(4.2xDazl)} \cdot \frac{P_{max}}{P_{(4.2xDazl)}} \quad \text{(Eq.11)}$$

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 3.0.10 [47]. 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. Inter-group differences were significant ($p < 0.05$) when the F value exceeded the critical F value for the given degrees of freedom [48].

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 3.6.2 package in R [49]) 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:

$$\sum B = \sum_{i=1}^{n} \left(\Phi^{\max(i)} \cdot \frac{k_{\mathrm{on}(i)}^{(4.2\mathrm{xDazl})}}{k_{\mathrm{on(i)}}^{(4.2\mathrm{xDazl})} + k_{\mathrm{diss.(i)}}} \right) = \sum_{i=1}^{n} (\Phi^{\max(i)} \cdot P_{(4.2\mathrm{xDazl})(\mathrm{i})}) \quad \text{(Eq.12)}$$

[*n*: number of binding sites in a given cluster; i: individual binding site, $\Phi^{\max(\mathrm{i})}$: 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:

$$\Delta \sum B = \sum_{i=1}^{n} \Phi^{\max(i)} \cdot \left(\frac{k_{\mathrm{on(i)}}^{(4.2\mathrm{xDazl})}}{k_{\mathrm{on(i)}}^{(4.2\mathrm{xDazl})} + k_{\mathrm{diss.(i)}}} - \frac{k_{\mathrm{on(i)}}^{(1\mathrm{xDazl})}}{k_{\mathrm{on(i)}}^{(1\mathrm{xDazl})} + k_{\mathrm{diss.(i)}}} \right) = \sum_{i=1}^{n} [\Phi^{\max(i)} \cdot (P_{(4.2\mathrm{xDazl})(\mathrm{i})} - P_{(1\mathrm{xDazl})(\mathrm{i})})] \quad \text{(Eq.13)}$$

[Variables as above, $k_{on(i)}^{(1xDazl)}$: association rate constant at 1xDazl for the binding site (i); $k_{diss.(i)}$, dissociation rate constant for binding site (i); $P_{(1xDazl)(i)}$: 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 ($RPKM_{4.2xDazl} < 6.0$) 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 [50] according to:

$$p = \mathrm{F}(x|M, K, N) = \sum_{i=0}^{x} \frac{\frac{(K)(M-K)}{(i)(N-i)}}{\frac{(M)}{(N)}} \quad \text{(Eq.14)}$$

(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 [50]) Hypergeometric tests were performed with *Scipy* hypergeom module 62 (ref.[51]) in Python 3.9.0.

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 [52]. 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 $K \leq n$ ($K = 2$ 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 [53]. PCA was conducted in R Project for Statistical Computing 4.3.0 using the prcomp 3.6.2 function. To visualize subgrouping within functional mRNA classes (Extended Data Fig.8d), the Barnes-Hut t-SNE implementation in R (rtSNE 0.13.0) [54] was used with the recommended parameters (perplexity 5 - 30, iterations 5 - 3000) as described [55].

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). Details of the multiple linear regression model are described in Supplementary Methods.

Decision Tree Classifier

We employed a Chi-squared Automatic Interaction Detection (CHAID) algorithm, which makes no assumption about underlying data [56,57], 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 [58]. The observed cell frequencies and the expected cell frequencies were then used to calculate Pearson chi-squared statistic, according to:

$$\chi^2 = \sum_{J=1}^{J} \sum_{I=1}^{I} \frac{(n_{IJ} - \acute{m}_{IJ})^2}{\acute{m}_{IJ}} \quad \text{(Eq.15)}$$

(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 [56,57].

The p value is:

$$p = \Pr(\chi_D^2 > \chi^2) \quad \text{(Eq.16)}$$

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 [58].

CHAID analysis was performed using CHAID 5.3.0 (ref.[59]) in Python 3.9.0.

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 [60]. 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 (ref.61) and plotted as a heatmap using ggplot2 (Fig.4d).

Pathway Analysis

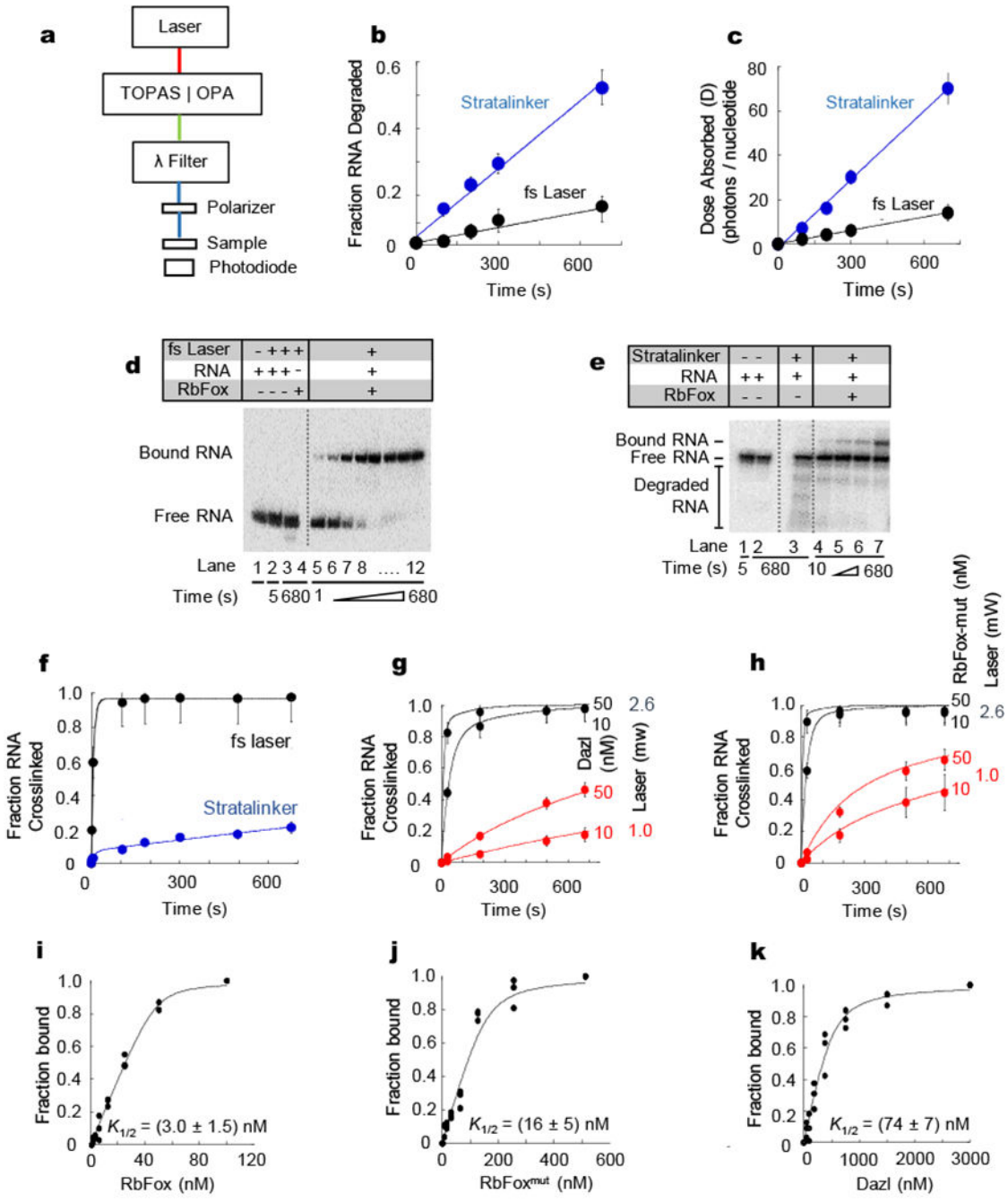
Pathways (Extended Data Fig.8h) were obtained from REACTOME [62,63]. mRNA classes were mapped on pathways with Cytoscape 3.4.0. [64].

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.

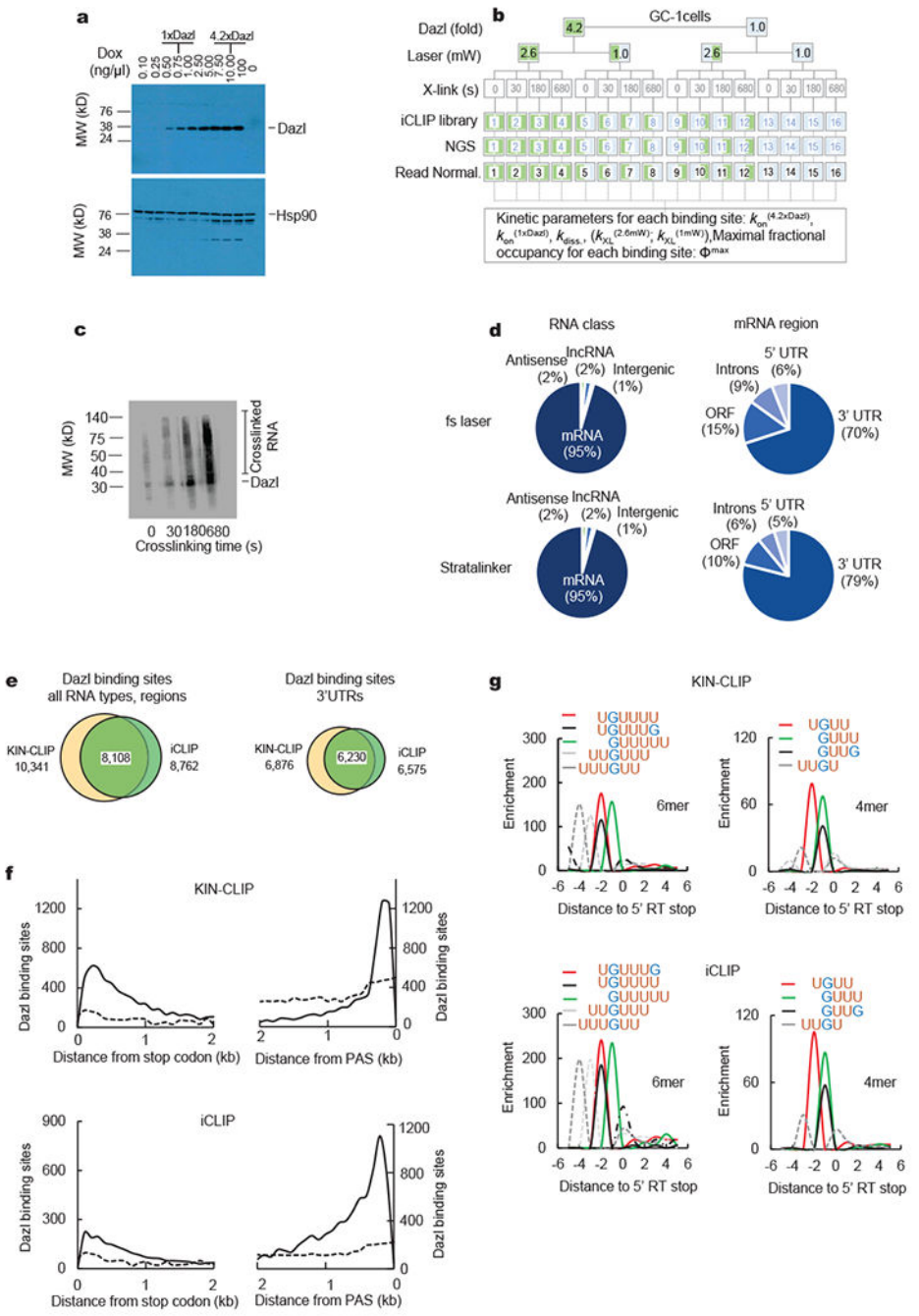
Extended Data



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). Error bars mark one standard deviation (N = 3 independent measurements). Lines

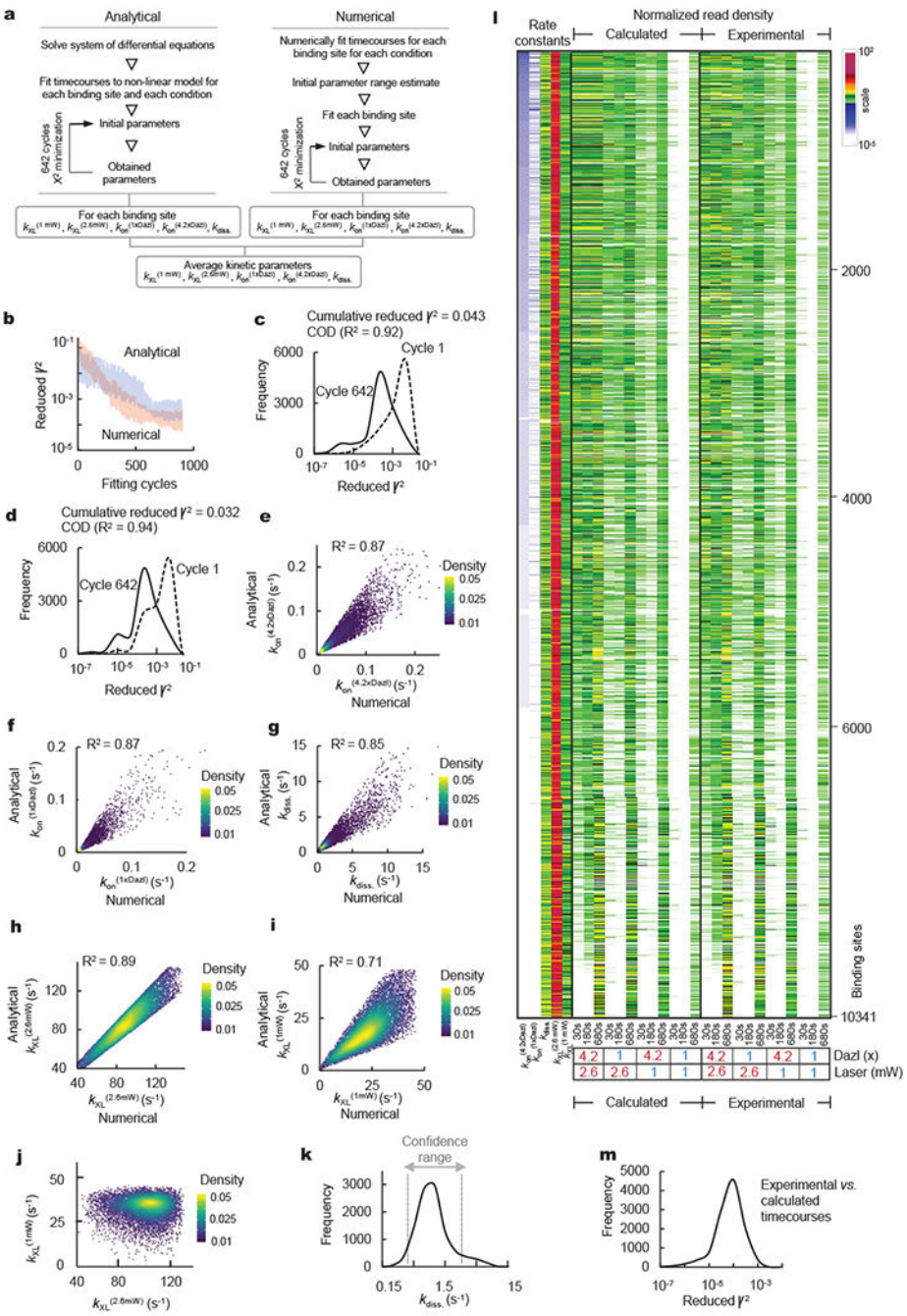
show a linear trend. **d**. Representative denaturing polyacrylamide gel electropherogram (PAGE) for a crosslinking reaction of 50 nM RbFox(RRM) (laser: 2.6 mW) (lanes 5 – 12) and control reactions with RNA only (lanes 1 – 3) and RbFox(RRM) only (lane 4), with (lanes 2-4) or without (lanes 1 and 5) crosslinking. Three independent measurements provided similar results. **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). Three independent measurements provided similar results. **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. Four independent experiments provided similar results. **b**. Schematic of the time-resolved crosslinking approach in cells. Numbers mark the respective CLIP libraries. **c**. Representative PAGE for bulk Dazl-RNA crosslinking. 3 independent experiments provided similar results. The intensity of crosslinked RNA (marked) is used to convert NGS reads to a concentration-equivalent parameter (for bulk crosslinking intensities and associated standard errors see Supplementary Material, Table S6) **d**. Distribution of CLIP sequencing reads across RNA

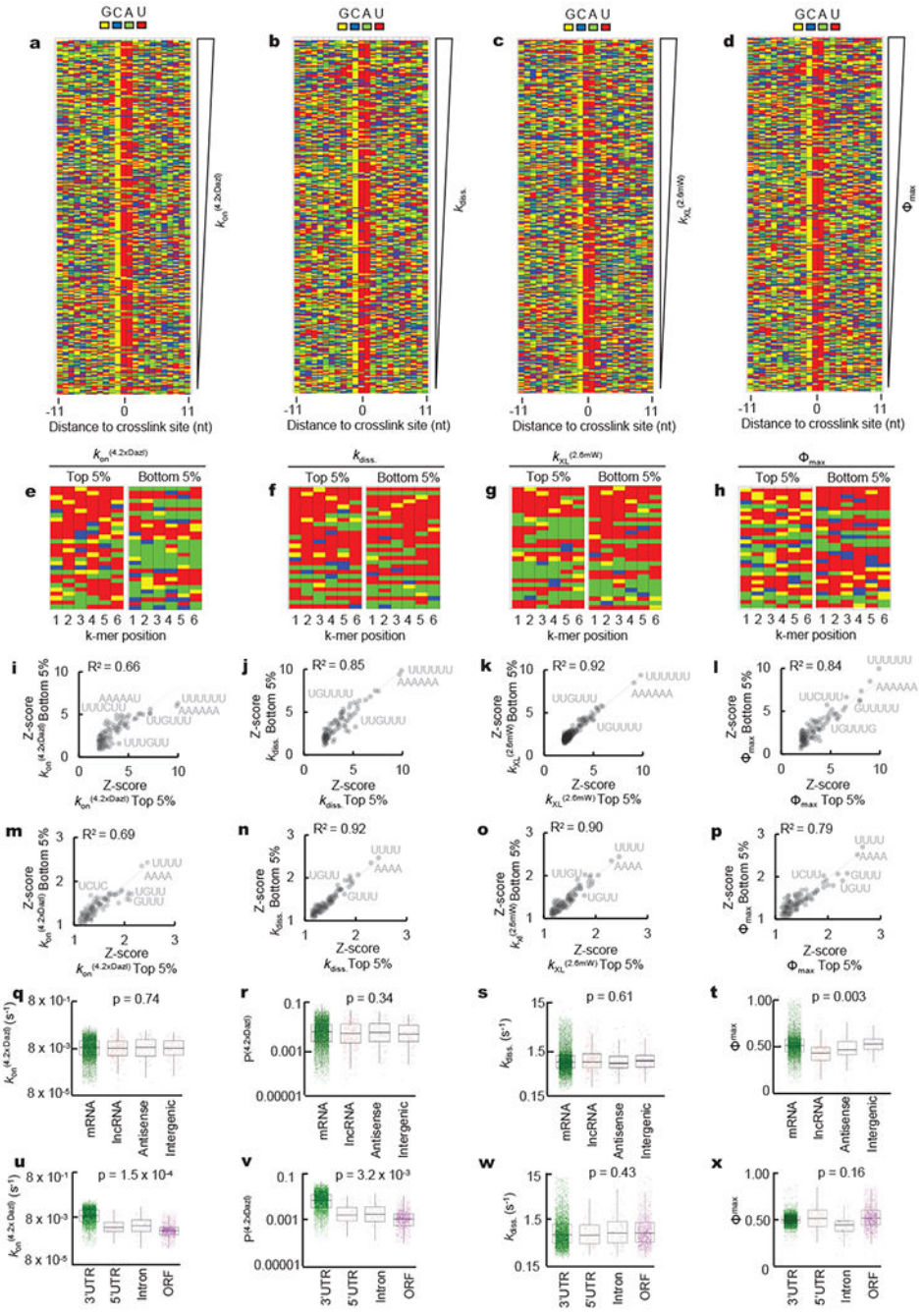
classes and mRNA regions for fs laser (4.2xDazl, 2.6 mW) and conventional crosslinking (Stratalinker; 4.2xDazl). Distributions for laser crosslinking experiments were calculated for binding sites with sequencing reads for all 12 measurements. Distribution for iCLIP experiments were calculated from three independent measurements [17]. **e**. Dazl binding sites identified by fs laser (KIN-CLIP) and conventional UV crosslinking (iCLIP) on all RNAs and 3'UTRs. **f**. 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. **g**. CITS (Crosslink Induced Truncation Site) 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**.

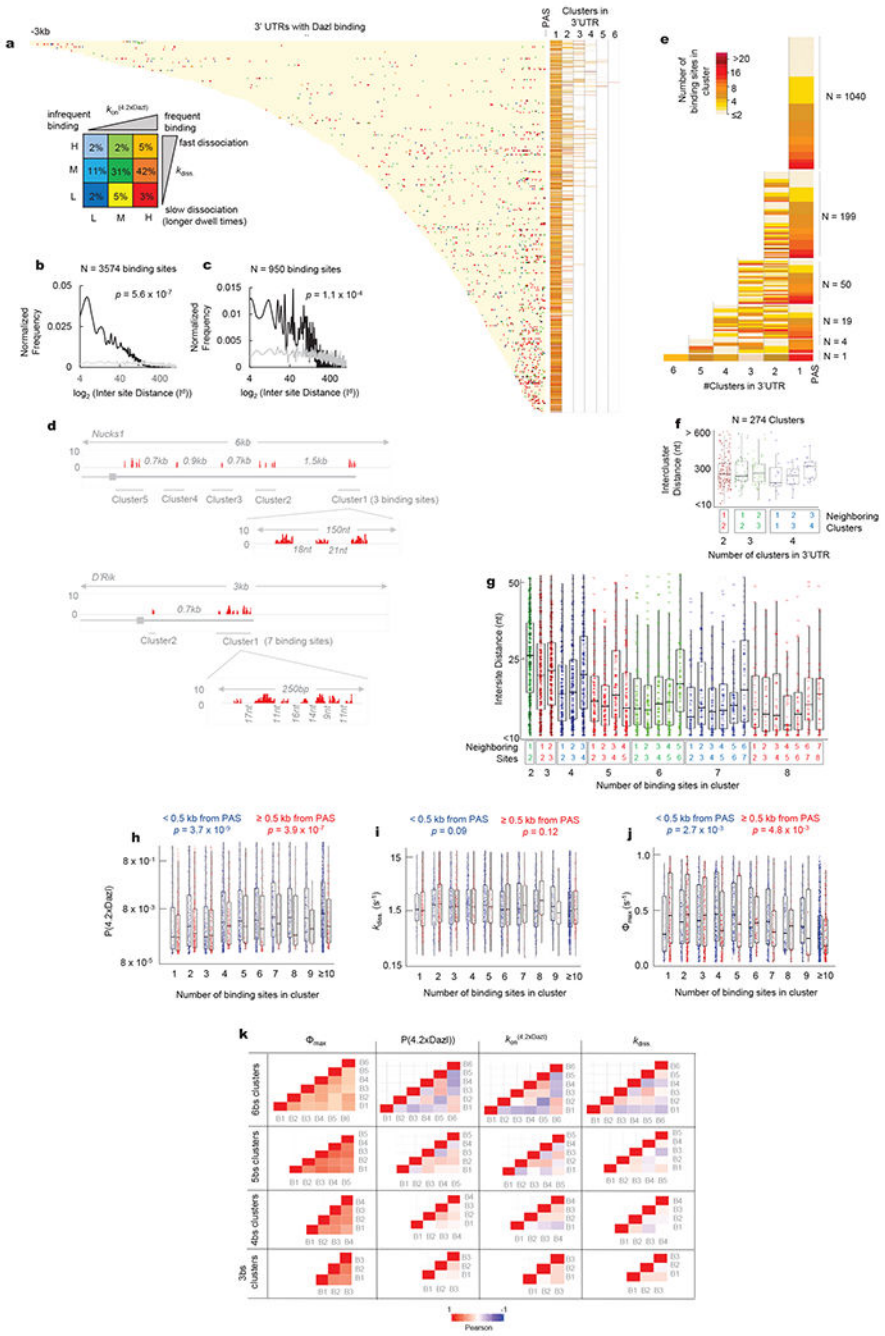
Correlation of parameters calculated with analytical and numerical fitting procedures (R^2: linear correlation coefficient). **j**. Correlation between crosslinking rate constants for low and high laser power. Rate constants are averaged from parameters obtained with numerical and analytical approach. 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**.

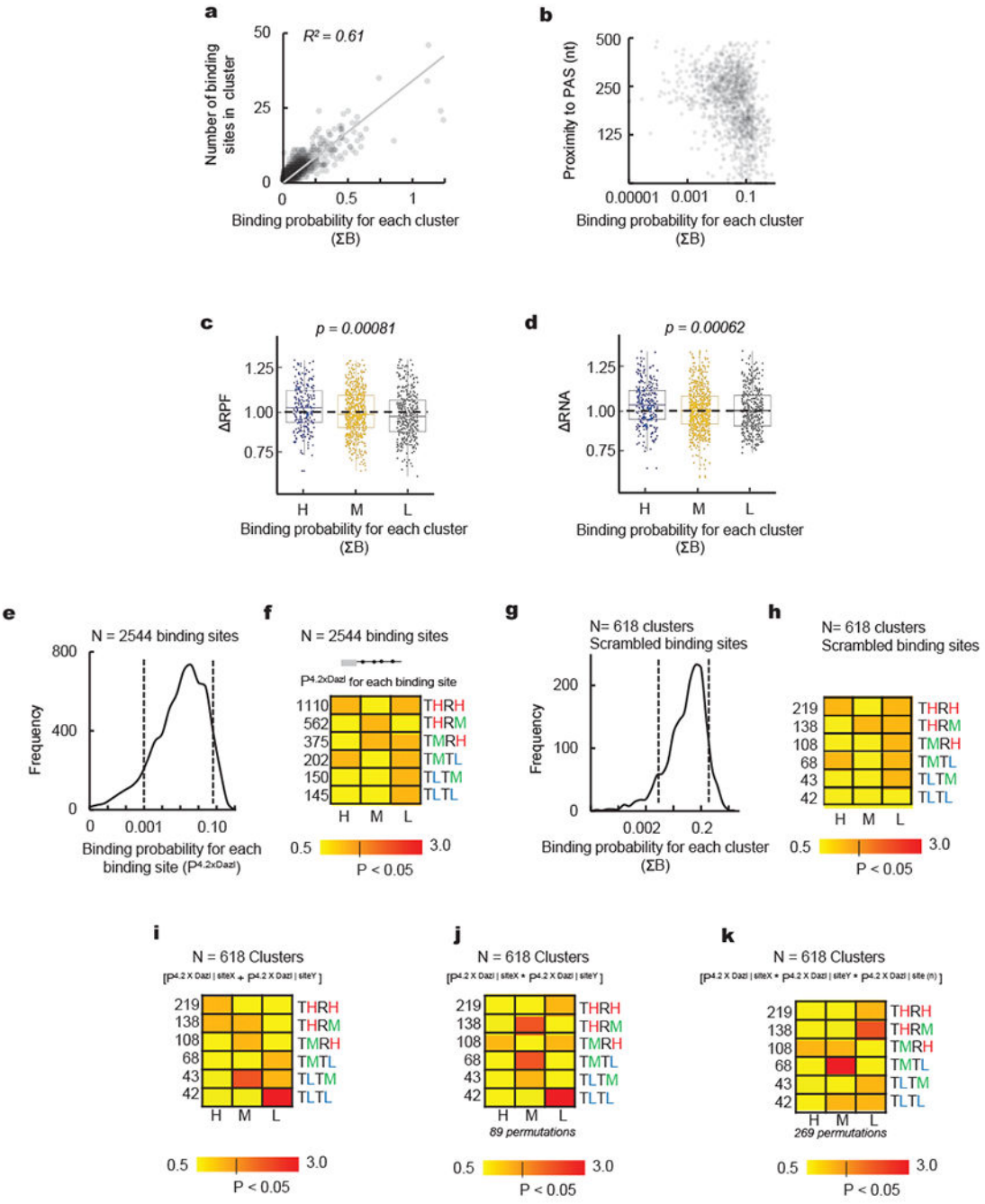
Sequences below the diagonal line correspond to enrichment of a 6-mer in the top 5% versus the bottom 5%. (R^2: linear correlation coefficient). A_6, U_6 and U_3GU_2 are most enriched in the vicinity of the binding sites with the fastest apparent association rate constants, compared to the binding sites 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 (N = 8,696, binding sites with associated values for fractional occupancy) 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). **u-x**. Distributions of kinetic parameters for all binding sites (N = 8,212, binding sites with associated values for fractional occupancy) in the indicated mRNA regions (p-values: one-way ANOVA; boxplots: vertical line: median, box limits: interquartile range (IQR); whiskers 1.5x IQR). $k_{on}^{(4.2xDazl)}$ and $P^{(4.2xDazl)}$, but not the other parameters vary significantly for different mRNA regions.



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: one sided 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: one sided 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 (N = 2,888) 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 (N = 6,546) 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). Boxplots: vertical line: median, box limits: interquartile range (IQR); whiskers 1.5x IQR. **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.

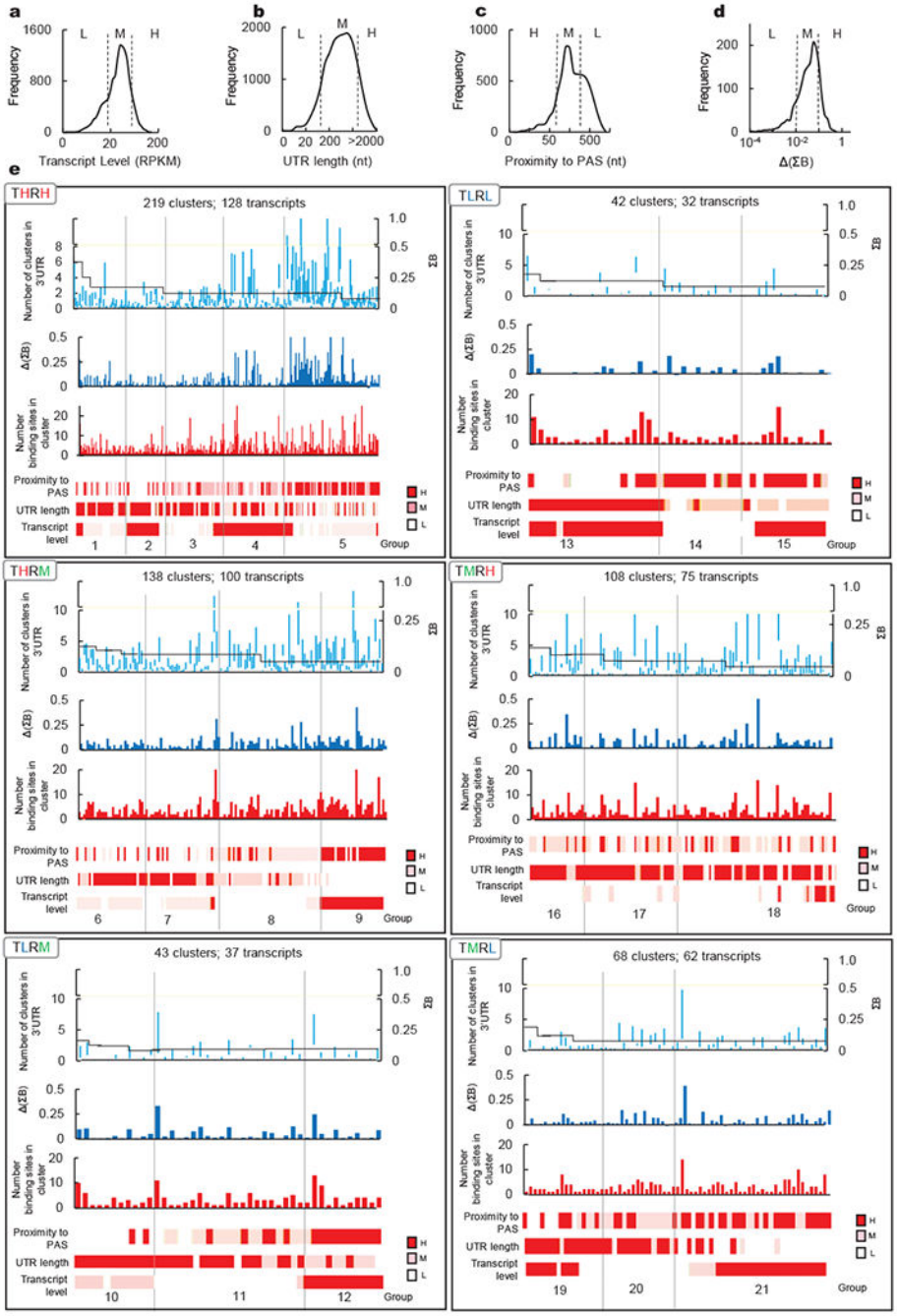


Extended Data Figure 6 |. Link between Dazl binding in 3'UTRs and impact on mRNA level and ribosome association.

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 in transcripts with associated values for ΔRNA and ΔRPF; 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. (Boxplots: vertical line: median, box limits: interquartile range (IQR); whiskers 1.5x IQR). **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 of a given ΣB tercile compared to a random distribution (hypergeometric test, one-sided, red: $p < 0.0005$ to 0.05, shades of yellow: $p > 0.05$ to 0.5, not enriched). 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 of a given ΣB tercile compared to a random distribution (hypergeometric test, one-sided, red: $p < 0.0005$ to 0.05, shades of yellow: $p > 0.05$ to 0.5, not enriched). No significant enrichment is observed. **i**. Correlation between additive binding probabilities of two Dazl sites in a cluster and functional mRNA classes. Colors mark the enrichment of a given ΣB tercile compared to a random distribution (hypergeometric test, one-sided, red: $p < 0.0005$ to 0.05, shades of yellow: $p > 0.05$ to 0.5, not enriched). 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. Colors mark the enrichment of a given ΣB tercile compared to a random distribution (hypergeometric test, one-sided, red: $p < 0.0005$ to 0.05, shades of yellow: $p > 0.05$ to 0.5, not enriched). 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. Colors mark the enrichment of a given ΣB tercile compared to a random distribution

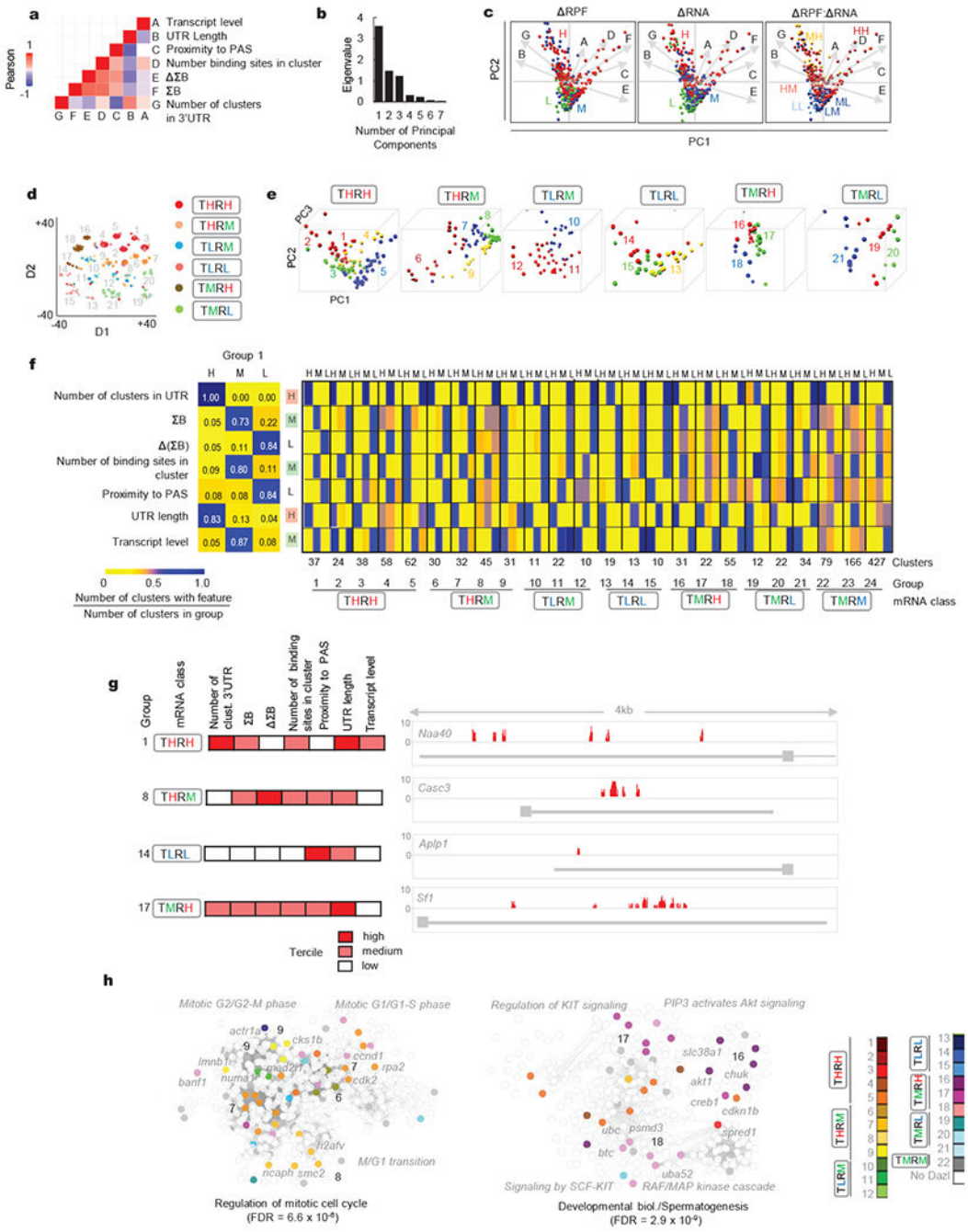
(hypergeometric test, one-sided, red: $p < 0.0005$ to 0.05, shades of yellow: $p > 0.05$ to 0.5, not enriched). For clusters with > 3 binding sites, permutations of three sites were tested and combinations of sites with the highest multiplicative binding probability were selected. The model tests whether 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

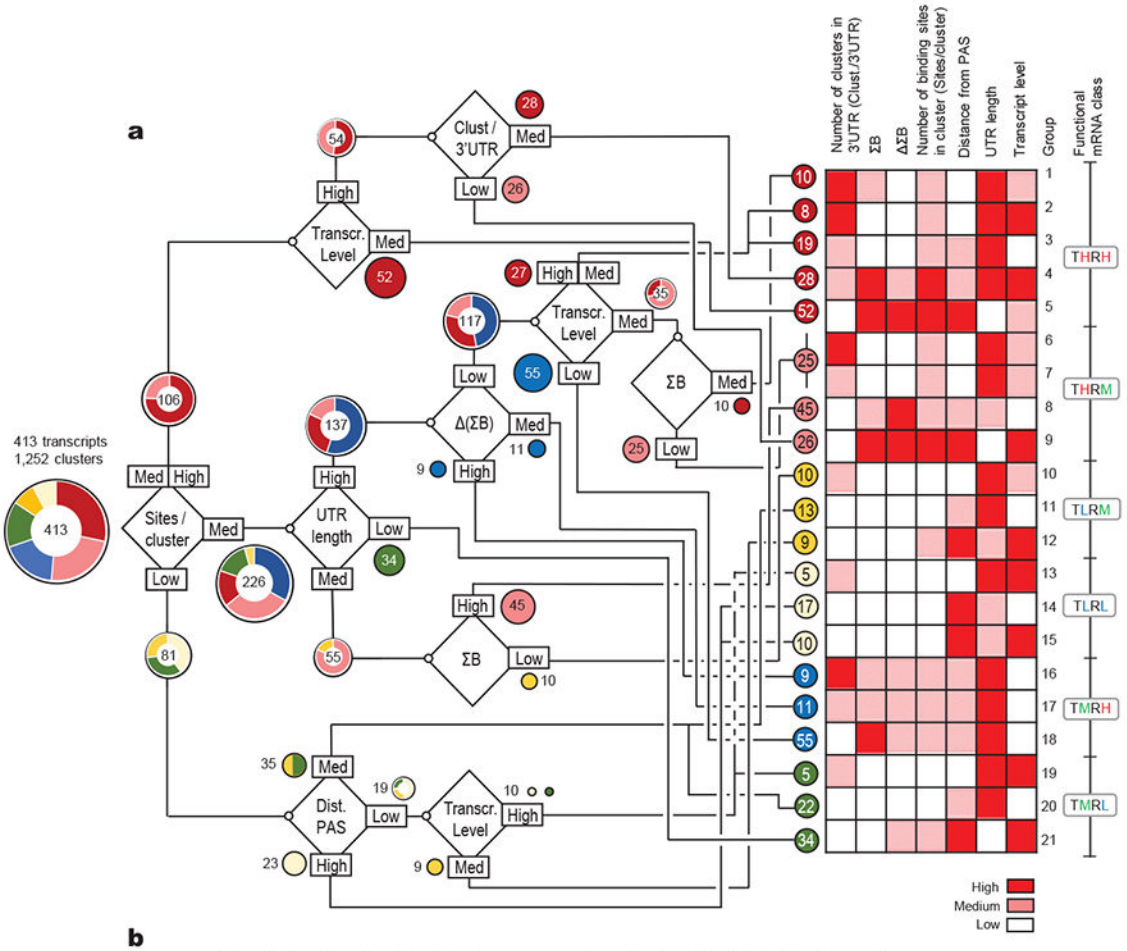
association and cluster features (upper graphs: number of Dazl clusters in 3'UTR: black line; ΣB: blue vertical lines, lower end marking ΣB at 1 x Dazl, upper end ΣB at 4.2 x Dazl; middle graphs: ΔΣB for each cluster and number of Dazl binding sites in each cluster; Heatmaps below the graphs: terciles 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 the 7 principal component axes obtained by PCA. Each eigenvalue corresponds to a principal component axis. Each axis reflects a linear combination of the 7 characteristics of a Dazl cluster, 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.



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

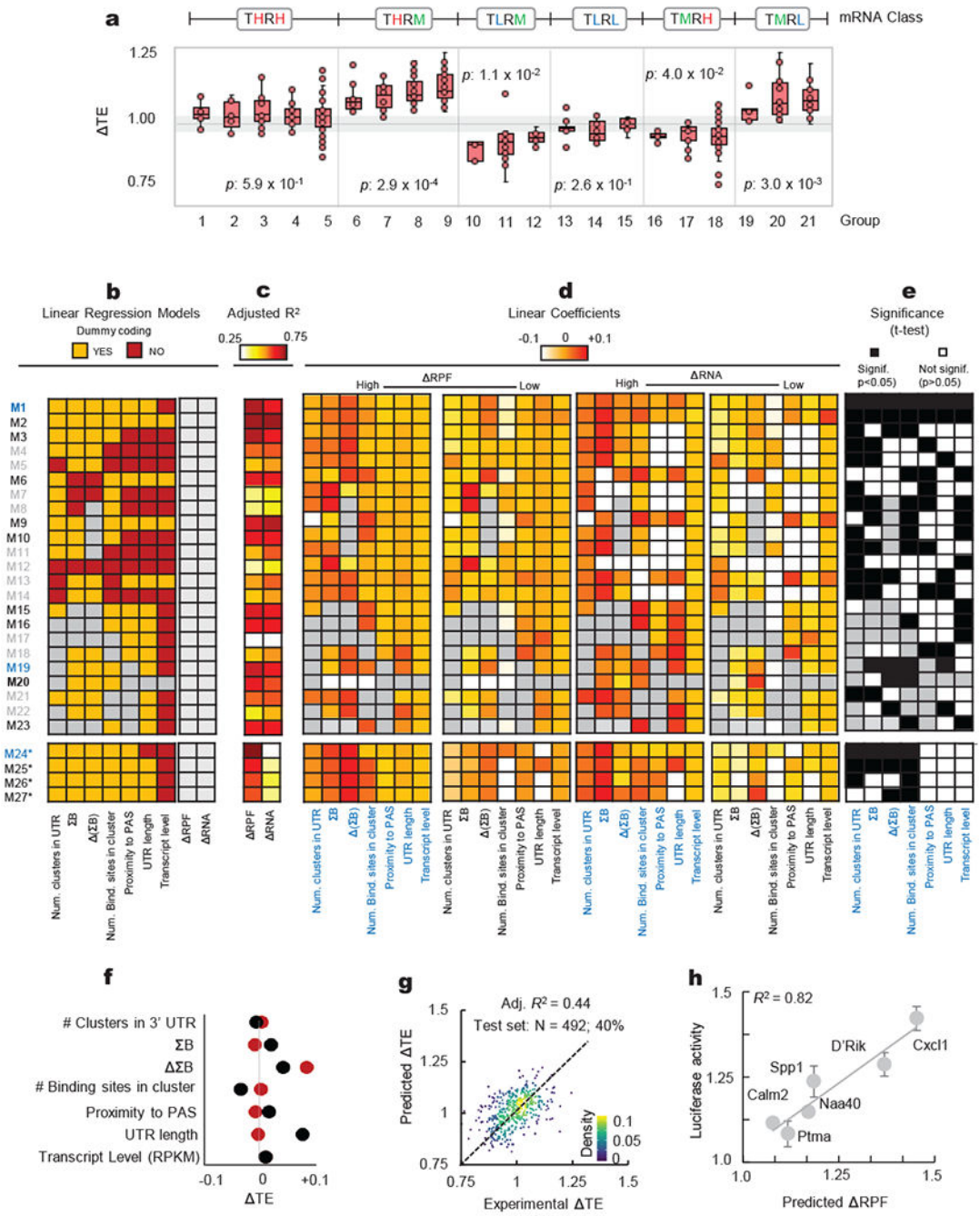
from \ to	HH	HM	LL	LM	MH	ML	Total	% correct
HH	8	1	0	0	0	0	8	90.000
HM	0	5	0	0	0	0	5	100.000
LL	0	0	2	1	0	0	3	67.000
LM	0	0	0	1	0	1	2	50.000
MH	0	0	0	0	4	0	4	100.000
ML	0	0	1	0	0	1	2	50.000

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 |. 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, significant: $p < 0.005$ to 0.05, black, not significant $p > 0.05$, p-values: one-sided student t-test on each coefficient). Only M1 shows consistently significant DILCs. Models 24-27 include interaction terms corresponding to 7 independent variable terms and test impact of multi-collinearity. 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 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. Error bars represent the standard error from the mean for each data point, corresponding to 5 independent experiments. 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).

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

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).

REFERENCES

1. Gerstberger S, Hafner M & Tuschl T A census of human RNA-binding proteins. Nat Rev Genet 15, 829–845 (2014). [PubMed: 25365966]
2. Licatalosi DD, Ye X & Jankowsky E Approaches for measuring the dynamics of RNA-protein interactions. Wiley Interdiscip Rev RNA 11, e1565 (2020). [PubMed: 31429211]

3. Corley M, Burns MC & Yeo GW How RNA-Binding Proteins Interact with RNA: Molecules and Mechanisms. Mol Cell 78, 9–29 (2020). [PubMed: 32243832]
4. Ule J, Hwang HW & Darnell RB The Future of Cross-Linking and Immunoprecipitation (CLIP). Cold Spring Harb Perspect Biol 10 (2018).
5. Van Nostrand EL et al. Principles of RNA processing from analysis of enhanced CLIP maps for 150 RNA binding proteins. Genome Biol 21, 90 (2020). [PubMed: 32252787]
6. Gleitsman KR, Sengupta RN & Herschlag D Slow molecular recognition by RNA. RNA 23, 1745–1753 (2017). [PubMed: 28971853]
7. Jarmoskaite I et al. A Quantitative and Predictive Model for RNA Binding by Human Pumilio Proteins. Mol Cell 74, 966–981 (2019). [PubMed: 31078383]
8. Sutandy FXR et al. In vitro iCLIP-based modeling uncovers how the splicing factor U2AF2 relies on regulation by cofactors. Genome Res 28, 699–713 (2018). [PubMed: 29643205]
9. Hockensmith JW, Kubasek WL, Vorachek WR & von Hippel PH 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). [PubMed: 3949776]
10. Pashev IG, Dimitrov SI & Angelov D Crosslinking proteins to nucleic acids by ultraviolet laser irradiation. Trends Biochem Sci 16, 323–326 (1991). [PubMed: 1835191]
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). [PubMed: 9171102]
12. Steube A, Schenk T, Tretyakov A & Saluz HP High-intensity UV laser ChIP-seq for the study of protein-DNA interactions in living cells. Nat Commun 8, 1303 (2017). [PubMed: 29101361]
13. Budowsky EI, Axentyeva MS, Abdurashidova GG, Simukova NA & Rubin LB Induction of polynucleotide-protein cross-linkages by ultraviolet irradiation. Peculiarities of the high-intensity laser pulse irradiation. Eur J Biochem 159, 95–101 (1986). [PubMed: 2427338]
14. Auweter SD et al. Molecular basis of RNA recognition by the human alternative splicing factor Fox-1. EMBO J 25, 163–173 (2006). [PubMed: 16362037]
15. Chen Y et al. Targeted inhibition of oncogenic miR-21 maturation with designed RNA-binding proteins. Nat Chem Biol 12, 717–723 (2016). [PubMed: 27428511]
16. Jenkins HT, Malkova B & Edwards TA 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). [PubMed: 22021443]
17. Zagore LL et al. DAZL Regulates Germ Cell Survival through a Network of PolyA-Proximal mRNA Interactions. Cell Rep 25, 1225–1240 e1226 (2018). [PubMed: 30380414]
18. Hofmann MC, Narisawa S, Hess RA & Millan JL Immortalization of germ cells and somatic testicular cells using the SV40 large T antigen. Exp Cell Res 201, 417–435 (1992). [PubMed: 1322317]
19. Fu XF et al. DAZ Family Proteins, Key Players for Germ Cell Development. Int J Biol Sci 11, 1226–1235 (2015). [PubMed: 26327816]
20. Lin Y & Page DC Dazl deficiency leads to embryonic arrest of germ cell development in XY C57BL/6 mice. Dev Biol 288, 309–316 (2005). [PubMed: 16310179]
21. Ruggiu M et al. The mouse Dazla gene encodes a cytoplasmic protein essential for gametogenesis. Nature 389, 73–77 (1997). [PubMed: 9288969]
22. Saunders PT et al. Absence of mDazl produces a final block on germ cell development at meiosis. Reproduction 126, 589–597 (2003). [PubMed: 14611631]
23. Yang CR et al. The RNA-binding protein DAZL functions as repressor and activator of mRNA translation during oocyte maturation. Nat Commun 11, 1399 (2020). [PubMed: 32170089]
24. Haberman N et al.. Insights into the design and interpretation of iCLIP experiments. Genome Biol 18, 7 (2017). [PubMed: 28093074]
25. Huppertz I et al.. iCLIP: protein-RNA interactions at nucleotide resolution. Methods 65, 274–287 (2014). [PubMed: 24184352]
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). [PubMed: 16278232]

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). [PubMed: 26265182]

ADDITIONAL REFERENCES

28. Weyn-Vanhentenryck SM 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). [PubMed: 24613350]

29. Zhang C & Darnell RB Mapping in vivo protein-RNA interactions at single-nucleotide resolution from HITS-CLIP data. Nat Biotechnol 29, 607–614 (2011). [PubMed: 21633356]

30. Aken BL et al. Ensembl 2017. Nucleic Acids Res 45, D635–D642 (2017). [PubMed: 27899575]

31. O'Leary NA et al. Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. Nucleic Acids Res 44, D733–745 (2016). [PubMed: 26553804]

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

33. Breiman L, Friedman JH, Olshen RA & Stone CJ Classification and Regression Trees. (Chapman & Hall/CRC, 1984).

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

35. Kass GV An exploratory technique for investigating large quantities for categorical data. Applied Statistics 20, 119–127 (1980).

36. Brister MM & Crespo-Hernandez CE Direct Observation of Triplet-State Population Dynamics in the RNA Uracil Derivative 1-Cyclohexyluracil. J Phys Chem Lett 6, 4404–4409 (2015). [PubMed: 26538051]

37. Brister MM & Crespo-Hernandez CE 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). [PubMed: 30995048]

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 MJ 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). [PubMed: 24407355]

41. Langmead B & Salzberg SL Fast gapped-read alignment with Bowtie 2. Nat Methods 9, 357–359 (2012). [PubMed: 22388286]

42. Quinlan AR & Hall IM BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics 26, 841–842 (2010). [PubMed: 20110278]

43. Robinson JT et al. Integrative genomics viewer. Nat Biotechnol 29, 24–26 (2011). [PubMed: 21221095]

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

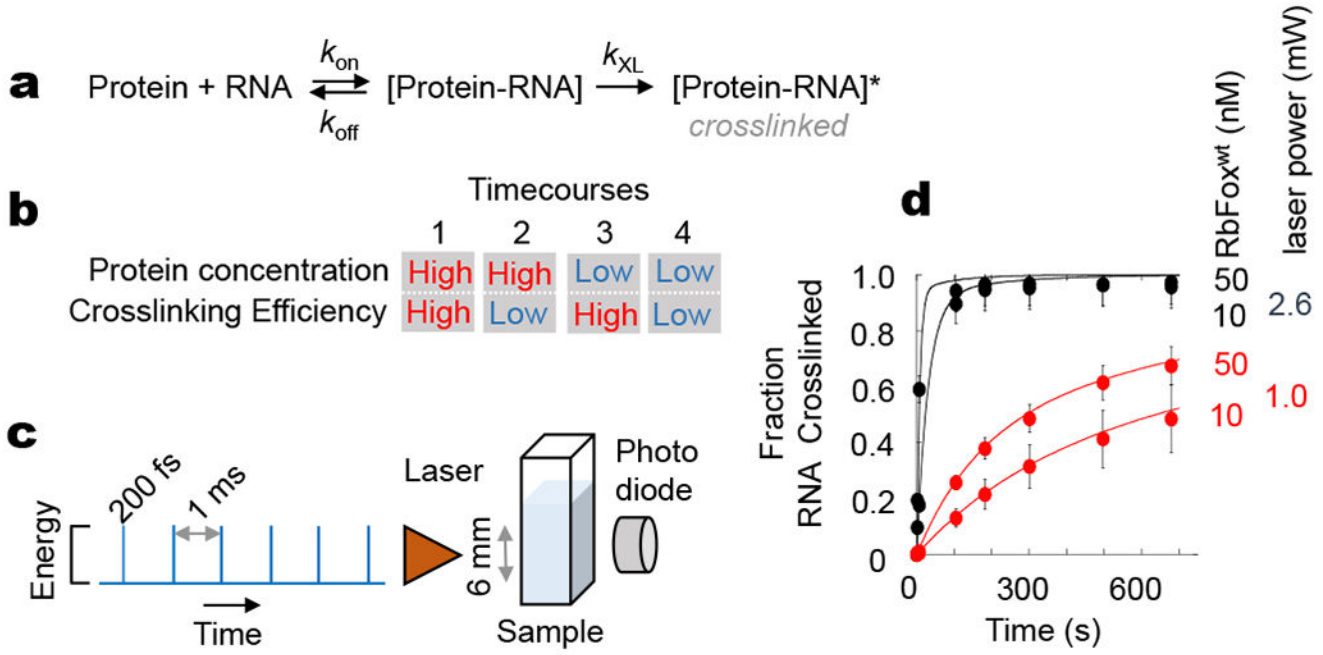
48. Thompson HW, Mera R & Prasad C The Analysis of Variance (ANOVA). Nutr Neurosci 2, 43–55 (1999). [PubMed: 27406694]

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

50. Cao J & Zhang S A Bayesian extension of the hypergeometric test for functional enrichment analysis. Biometrics 70, 84–94 (2014). [PubMed: 24320951]

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

52. Jolliffe IT & Cadima J Principal component analysis: a review and recent developments. Philos Trans A Math Phys Eng Sci 374, 20150202 (2016). [PubMed: 26953178]
53. Kerr G, Ruskin HJ, Crane M & Doolan P Techniques for clustering gene expression data. Comput Biol Med 38, 283–293 (2008). [PubMed: 18061589]
54. Krijthe J <https://cran.r-project.org/web/packages/Rtsne/index.html> (2018).
55. van der Maaten L Visualizing Data using t-SNE. J Mach Learn Res 9, 2579–2605 (2008).
56. Bigss D, Ville B & Suen E A Method of Choosing Multiway Partitions for Classification and Decision Trees. J Appl Stat 18, 49–62 (1991).
57. Goodman LA Simple Models for the Analysis of Association in CrossClassifications Having Ordered Categories. J Am Stat Assoc 74, 537–552 (1979).
58. Armstrong RA When to use the Bonferroni correction. Ophthalmic Physiol Opt 34, 502–508, doi:10.1111/opo.12131 (2014). [PubMed: 24697967]
59. <https://pypi.org/project/CHAID/>
60. 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).
61. Ward's minimum variance method <https://uc-r.github.io/hc_clustering>
62. Fabregat A et al. Reactome pathway analysis: a high-performance in-memory approach. BMC Bioinformatics 18, 142 (2017). [PubMed: 28249561]
63. Jassal B et al. The reactome pathway knowledgebase. Nucleic Acids Res 48, D498–D503 (2020). [PubMed: 31691815]
64. Shannon P et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res 13, 2498–2504 (2003). [PubMed: 14597658]



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 |. 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. Error bars mark one standard deviation from the mean (N = 3 independent measurements). **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.

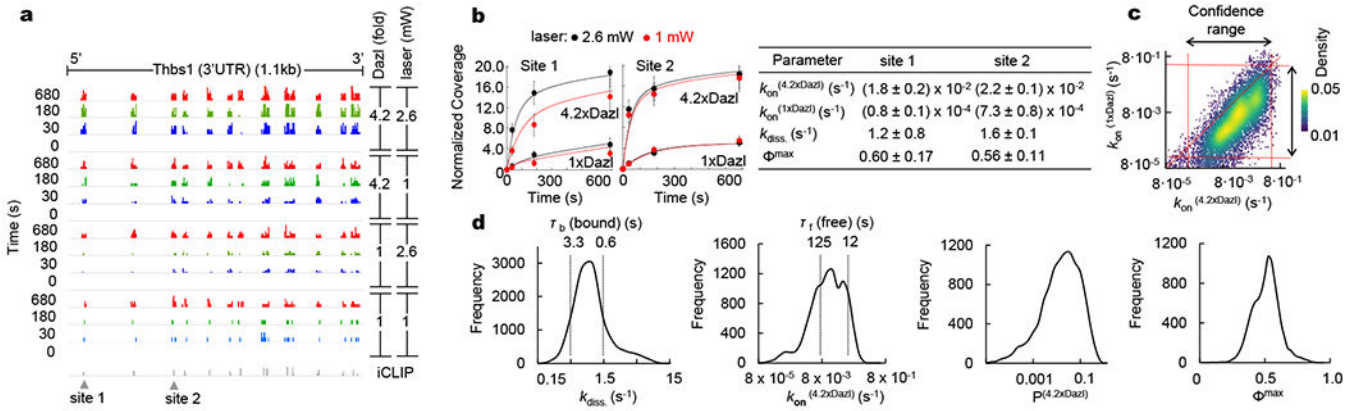


Figure 2 |. Kinetics of Dazl-RNA binding and dissociation in cells.

a. 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. **b**. 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. **c**. 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. **d**. 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 (T_b) and away from binding sites (T_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).

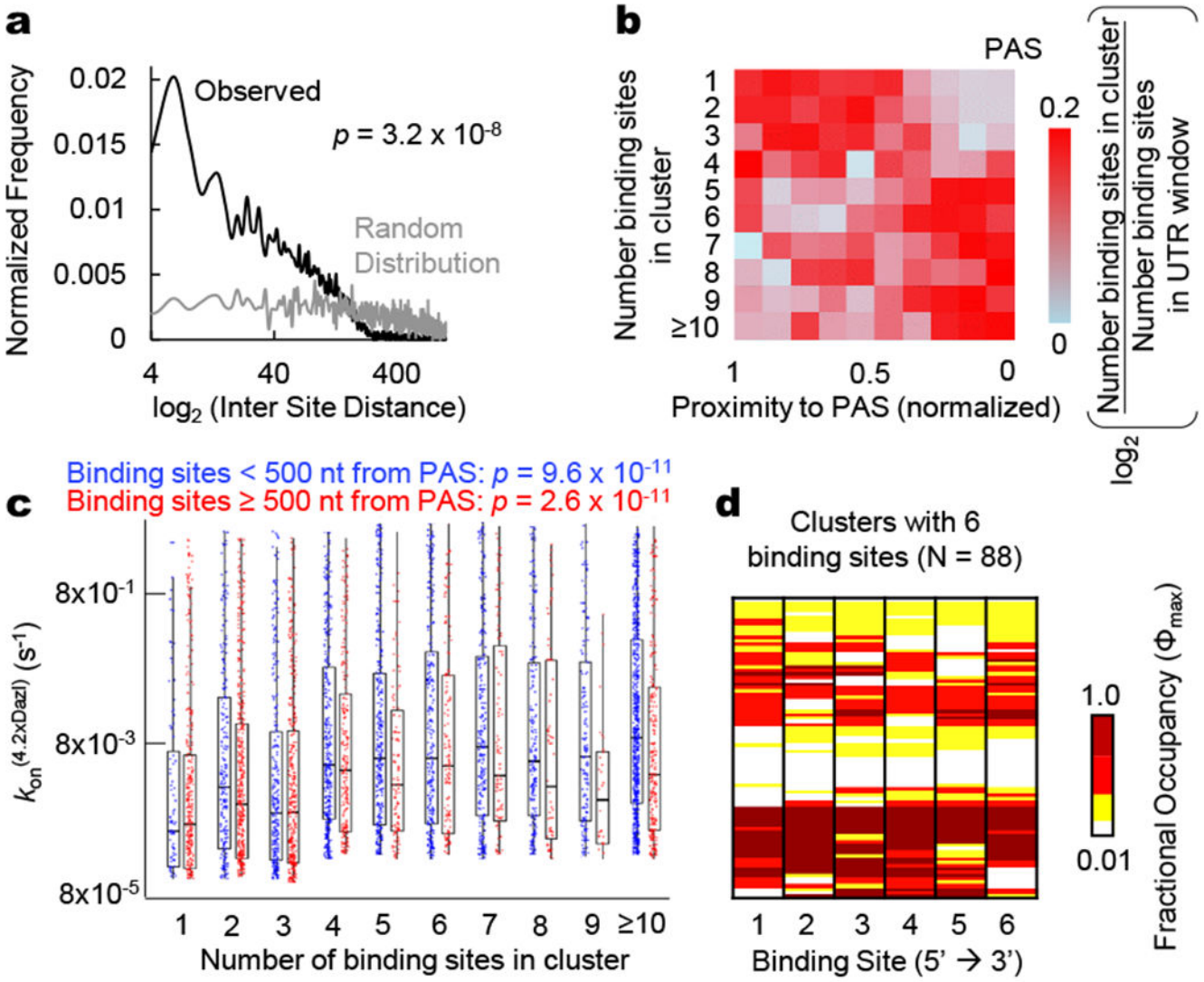


Figure 3 |. Clustering of Dazl binding sites in 3'UTRs.

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: one sided 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. (N = 6,546; 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.

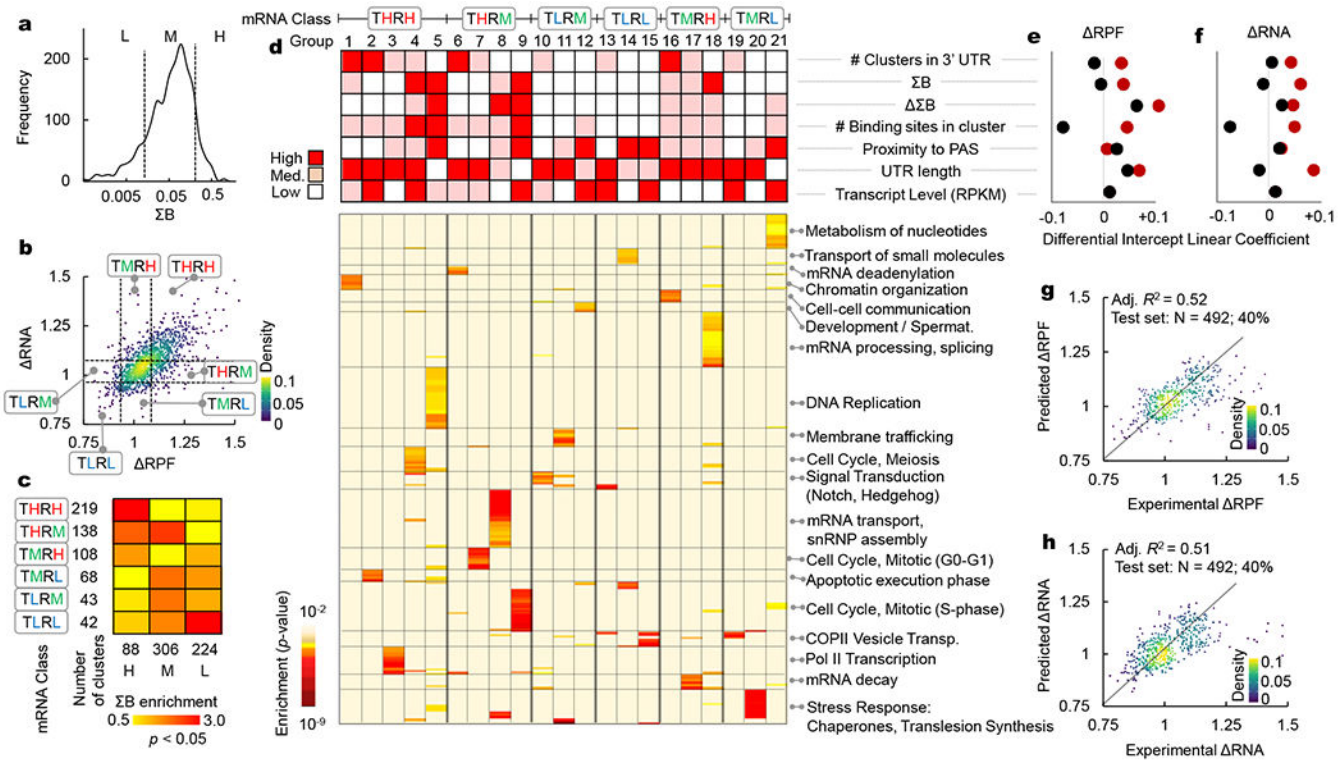


Figure 4 |. Link between Dazl-RNA binding and Dazl impact on mRNA function.

a. Distribution of cumulative binding probabilities (ΣB) for Dazl in all clusters in 3'UTRs (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 reflect enrichment of a given ΣB tercile compared to a random distribution. (hypergeometric test, one-sided, red: $p < 0.0005$ to 0.05, yellow: $p > 0.05$ to 0.5, 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, 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: transcripts).