

# FLASH: ultra-fast protocol to identify RNA–protein interactions in cells

Ibrahim Avsar Ilik<sup>1,2,\*</sup>, Tugce Aktas<sup>1,2</sup>, Daniel Maticzka<sup>3</sup>, Rolf Backofen<sup>3,4</sup> and Asifa Akhtar<sup>1,\*</sup>

<sup>1</sup>Max Planck Institute of Immunobiology and Epigenetics, 79108 Freiburg, Germany, <sup>2</sup>Otto Warburg Laboratories, Max Planck Institute for Molecular Genetics, 14195 Berlin, Germany, <sup>3</sup>Bioinformatics Group, Department of Computer Science, University of Freiburg, 79110 Freiburg, Germany and <sup>4</sup>Centre for Biological Signalling Studies (BOSS), University of Freiburg, 79104 Freiburg, Germany

Received December 12, 2018; Revised November 17, 2019; Editorial Decision November 20, 2019; Accepted December 03, 2019

## ABSTRACT

Determination of the *in vivo* binding sites of RNA-binding proteins (RBPs) is paramount to understanding their function and how they affect different aspects of gene regulation. With hundreds of RNA-binding proteins identified in human cells, a flexible, high-resolution, high-throughput, highly multiplexible and radioactivity-free method to determine their binding sites has not been described to date. Here we report FLASH (Fast Ligation of RNA after some sort of Affinity Purification for High-throughput Sequencing), which uses a special adapter design and an optimized protocol to determine protein–RNA interactions in living cells. The entire FLASH protocol, starting from cells on plates to a sequencing library, takes 1.5 days. We demonstrate the flexibility, speed and versatility of FLASH by using it to determine RNA targets of both tagged and endogenously expressed proteins under diverse conditions *in vivo*.

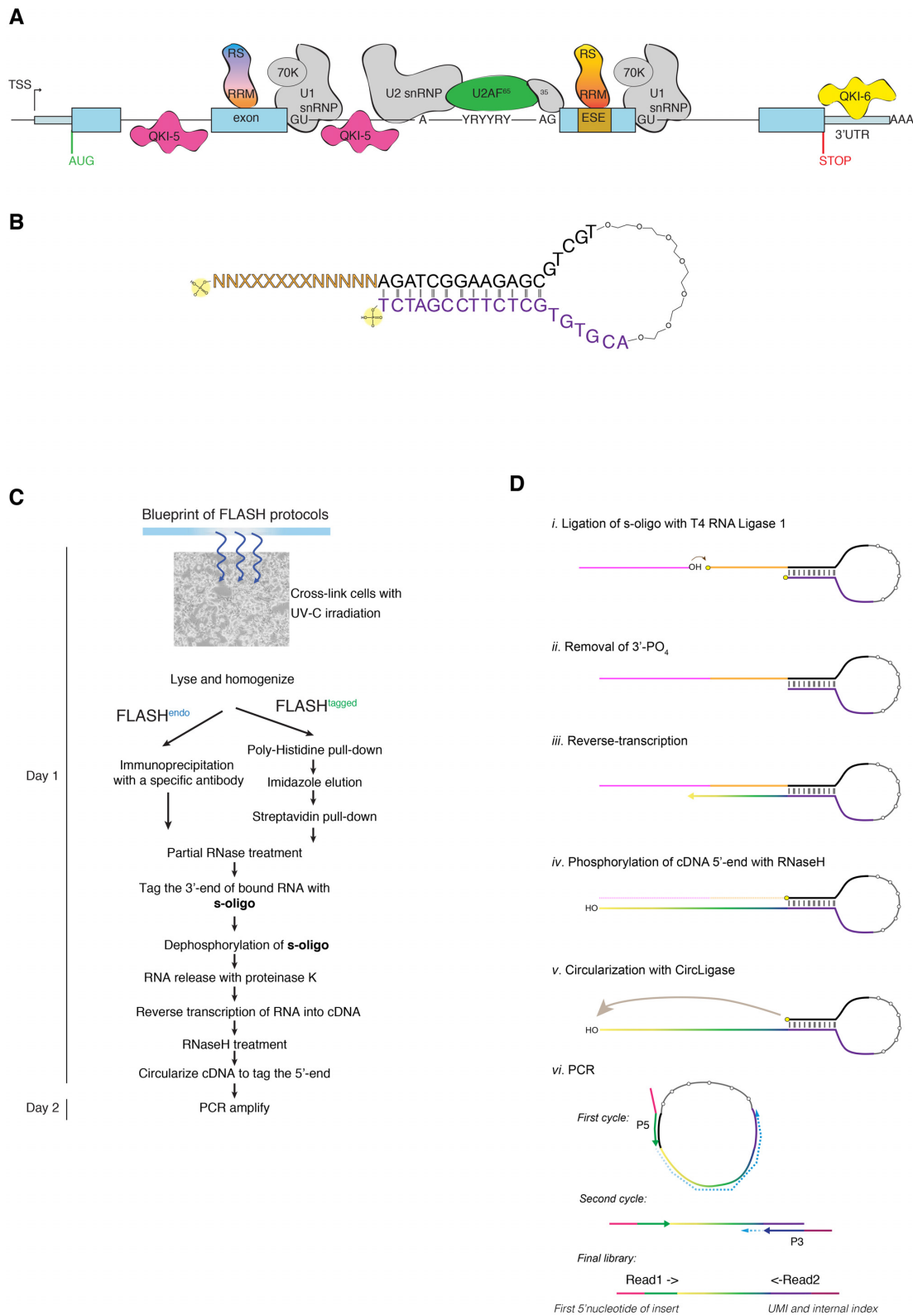
## INTRODUCTION

In eukaryotic organisms, RNA polymerase II is responsible for the transcription of all mRNAs that serve as templates for protein synthesis in the cytoplasm. Several highly regulated RNA-processing events are required to produce translation-competent mRNAs and guide their passage from the nucleus to the cytoplasm. These events are almost exclusively carried out by RNA-binding proteins (RBPs) that recognize short sequence motifs, specific RNA structures and/or RNA modifications (1,2). Understanding how the collective action of RBPs determines the fate of mRNAs requires the accurate and precise identification of their cellular targets. Current protocols that are up to this challenge are generally called CLIP-seq (crosslinking and immunoprecipitation) approaches, where either UV-

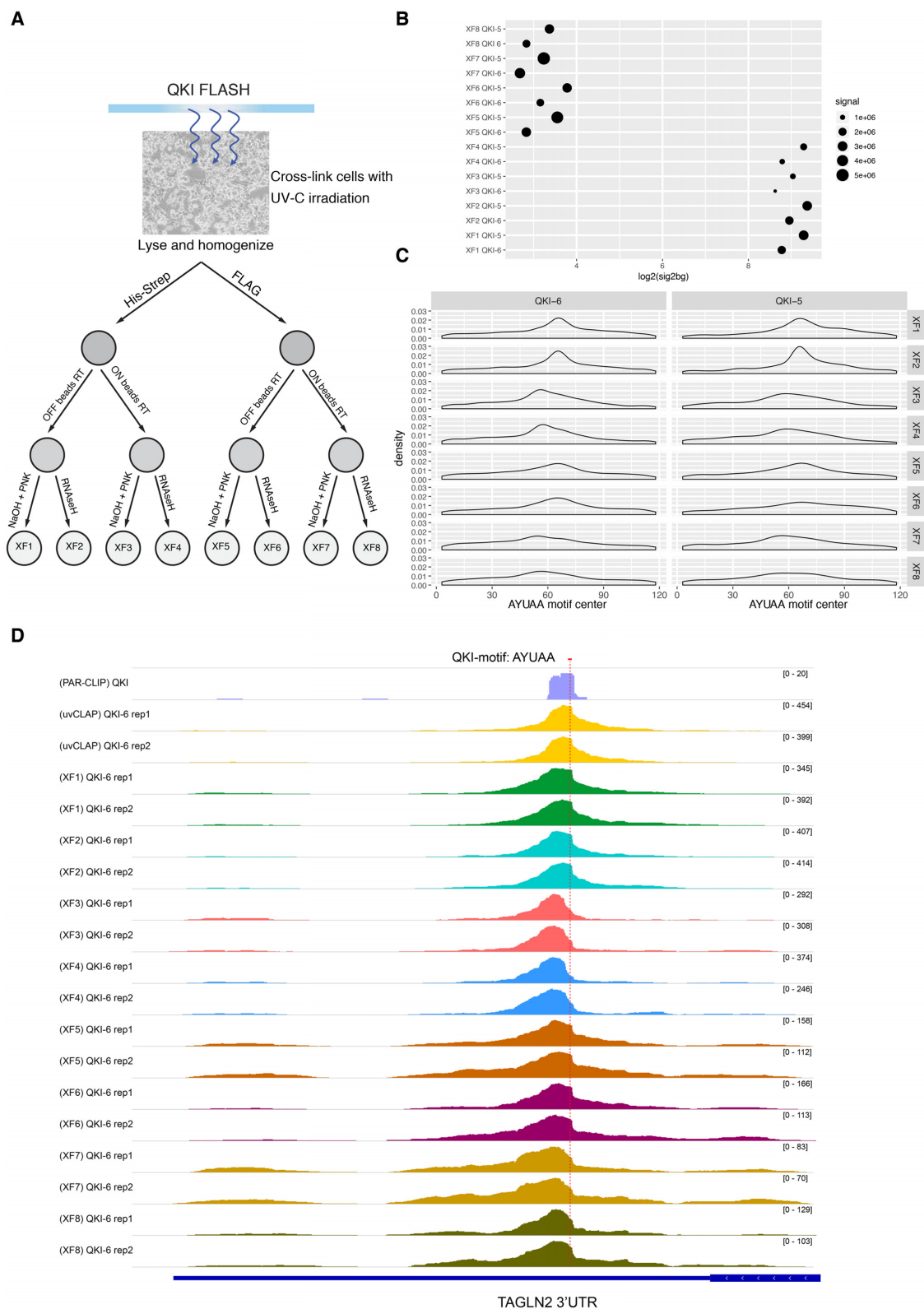
C light (wavelength ~254 nm) or UV-A light (wavelength ~365 nm, in combination with a photoactivatable ribonucleotide analogue incorporated into target RNAs) is used to directly induce protein–RNA crosslinks, then the resulting RNA–protein adducts are purified, proteins are typically removed by Proteinase K treatment and the RNA is cloned into a sequencing library, sequenced and finally analyzed. Many state-of-the-art CLIP-seq methods are, however, notoriously difficult as they require use of radioactive substances and isolation of minute amounts of RNA from nitrocellulose paper (3). Due to these design choices, and other technical challenges, these protocols take several days to complete, and are typically restricted to one or two proteins per iteration. Recently developed techniques address some of these problems by bypassing radioactive labeling and/or isolation of RNA from nitrocellulose paper or acrylamide gels (4–6). Here, we report a new protocol, FLASH, that surpasses these techniques in terms of flexibility, multiplexibility and speed, while retaining single-nucleotide resolution and specificity owing to a new adapter design.

Since it is not immediately obvious how a change in adapter design can lead to dramatic improvements in the protocol, we first detail the design elements of this new adapter, followed by the description of how these elements come together to solve several technical challenges in CLIP or CLIP-like experiments that require ligation-mediated cloning of RNA. In order to test the performance and applicability of FLASH in various experimental scenarios, we report FLASH profiles of several important RBPs with different cellular roles (Figure 1A), beginning with the optimization of the FLASH protocol using two isoforms of the KH-type RBP QKI, QKI-5 and QKI-6 (Figure 2). QKI is a well-studied RBP with roles in splicing, circular RNA formation, RNA stability and translation (7). QKI is able to influence both nuclear and cytoplasmic events since the endogenous gene produces protein variants through alternative splicing that localize to the nucleus, cytoplasm or both compart-

\*To whom correspondence should be addressed. Tel: +49 0 7615108565; Fax: +49 0 7615108566; Email: akhtar@ie-freiburg.mpg.de  
Correspondence may also be addressed to Ibrahim Avsar Ilik. Email: ilik@molgen.mpg.de



**Figure 1.** (A) A typical multi-exonic mammalian transcript depicted together with a plethora of RNA binding proteins. (B) Schematic description of the s-oligo. See text for details. (C) A ‘Blueprint’ of FLASH protocols, depicting various experimental alternatives. (D) Step-wise explanation of the enzymatic steps of the FLASH protocol, starting with the ligation of the s-oligo to the target and ending with PCR amplification.



**Figure 2.** (A) Schematic representation of all the conditions that were tested with QKI-5 and QKI-6. For a more complete description see Supplementary Figures S1 and S2. (B) Signal-to-noise ratios of the XF libraries, higher ratios are indicative of lower background. (C) Frequency distribution of crosslinked nucleotides in the vicinity of the QKI motif AYUAA in XF libraries. (D) An Integrative Genomics Viewer (IGV) screenshot of QKI FLASH data compared to PAR-CLIP data (10) and uvCLAP data (6).

ments (8). Furthermore, cellular RNA targets of QKI have been described before, including its preferred sequence motif, and preferred target classes for both isoforms, making it possible to directly assess the performance of different protocol variations (6,9–10). With the optimal conditions established, we then report FLASH profiles of U2AF65 (Figure 3), an RRM-type RBP that recognizes polypyrimidine tracts and which together with U2AF35 is instrumental in defining splice-acceptor sites. Another important class of RBPs that are critical for splicing are the serine/arginine (SR)-rich proteins, which play important roles in exon definition as well as nuclear export of mRNAs (11). We report FLASH profiles of nine SR proteins and a clinically important point mutant of SRSF2 (Figure 4). Finally, to exploit and demonstrate the speed and flexibility of FLASH, we profile transiently transfected QKI constructs in a 6-well plate format, and assess the practicality and utility of such a setup (Figure 5).

## MATERIALS AND METHODS

### Cell culture and generation of stable cell lines

Flp-In T-REx 293 (Thermo Fisher Scientific, R78007) cells were maintained with DMEM-Glutamax supplemented with sodium pyruvate, glucose and 10% FBS. In addition to those supplements, Flp-In T-REx 293 cells were maintained in zeocin (Thermo Fisher Scientific, R25001)- and blasticidin (Thermo Fisher Scientific, A1113903)-containing medium according to manufacturer's protocol and the zeocin selection is exchanged with hygromycin (Thermo Fisher Scientific, 10687010) upon transgene transfection.

All the transgenes were cloned into pCDNA5-FRT/TO (Thermo Fisher Scientific, V6520-20) with a C-terminal 3× Flag-HBH tag by PCR amplification of the coding sequence from cDNA prepared from Flp-In T-REx 293 cells with the oligos listed in Supplementary Table S1. U2AF65 was ordered from the BIOSS Toolbox, University of Freiburg, and originally is from human ORFeome V5.1 collection (Open Biosystems) with ID:4551. Transgenes were co-transfected with pOG44 (Thermo Fisher Scientific, V600520) plasmid with a 1:9 DNA concentration ratio for the generation of stable cell lines. Cells were re-plated in different dilutions (in order to assure the impact of cell density on cell survival after antibiotics selection) 24 h after transfection and 150 µg per ml hygromycin selection was initiated 48 h after transfection. Cell lines were maintained with blasticidin and hygromycin at all times and the transgenes were induced over-night with 0.1 µg per ml doxycycline.

### Transfection and induction for transient FLASH

A total of 800 000 Flp-In T-REx 293 cells were transfected with 1 µg of plasmid DNA (pCDNA5-FRT/To plasmid carrying the tagged gene) in a 6-well plate. The medium was exchanged 6 h after the transfection with a medium carrying 0.1 µg per ml doxycycline.

### s-oligo

The adapters used in FLASH protocols, the s-oligo were ordered from Integrated DNA Technologies (IDT). The com-

plete list of s-oligos used in this study can be found in Supplementary Table S1. We use one example to explain the chemical makeup of the adapters:

```

/5Phos/rNrNrCrArCrUrGrNrYrYrNrNAGATC
GGAAGAGCGTCGT/iSp18/ACGTGTGCTCTCCGA
TCT/3Phos/
/5Phos/: 5'-phosphate group
rN: ribonucleotide (random in this case)
/iSp18/: 8-atom hexa-ethyleneglycol spacer
/3Phos/: 3'phosphate group
A, G, T or C: DNA moieties

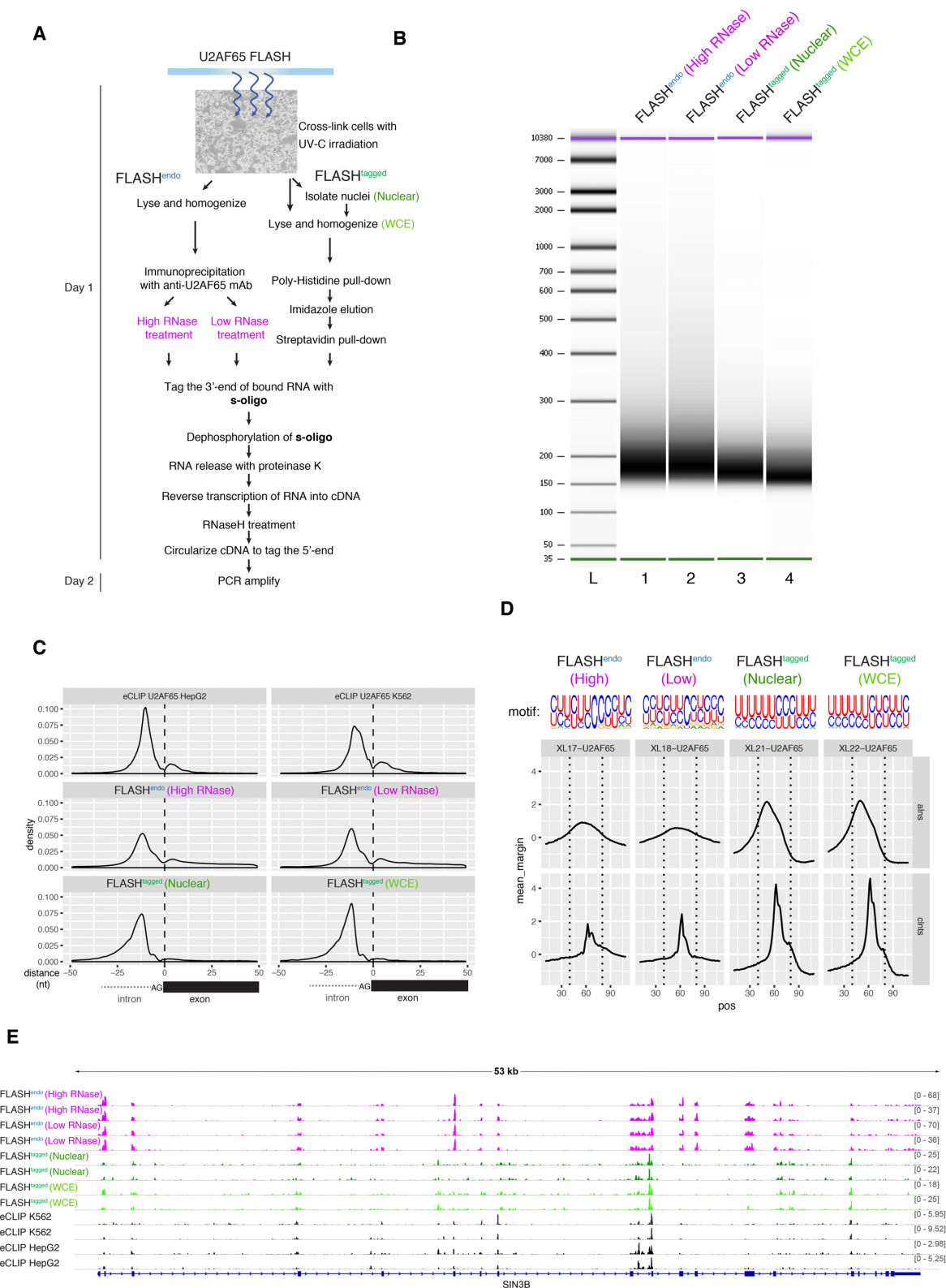
```

### FLASH protocols

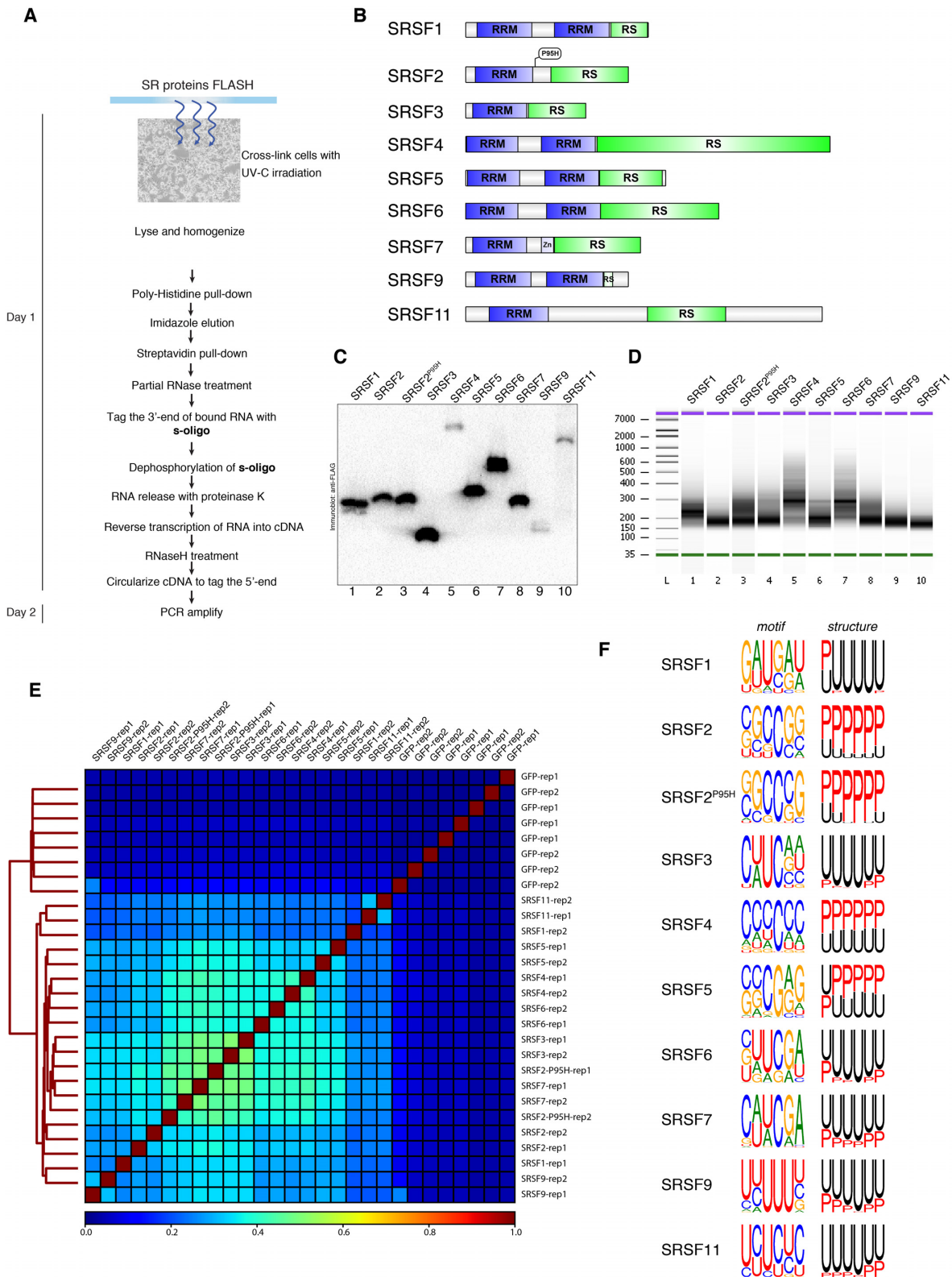
The XF2 flavor of the protocol, described below, is deposited to protocols.io as a step-by-step protocol: [https://www.protocols.io/view/flash-zv9f696?version\\_warning=no](https://www.protocols.io/view/flash-zv9f696?version_warning=no)

For most FLASH protocols, cells grown in 15-cm dishes and are induced with 0.1 µg/ml doxycycline for ~16 h. Medium is removed, cells are washed once with 6 ml of ice-cold PBS, then 6 ml of fresh ice cold PBS is carefully layered on top of the cells and cells are crosslinked on an ice-water tray with 0.15 mJ/cm<sup>2</sup> UV-C light. The cells are collected into a 15-ml Falcon tube with the help of a cell scraper, spun down in a cold centrifuge at 500g, washed once with 1 ml ice-cold PBS, re-pelleted at 500g and snap-frozen in liquid nitrogen until use. On the day of the experiment, the cells are allowed to thaw on ice for ~2 min and resuspended with 550 µl of NLB (1× PBS, 0.3 M NaCl, 1% Triton-X, 0.1% Tween-20). The lysate is then sonicated with a Bioruptor sonifier (Diagenode) for 5 mins (30 s ON, 30 s OFF, LOW, 5 cycles at 4°C). Insoluble material is removed by centrifugation at 20 000g for 10 min at 4°C.

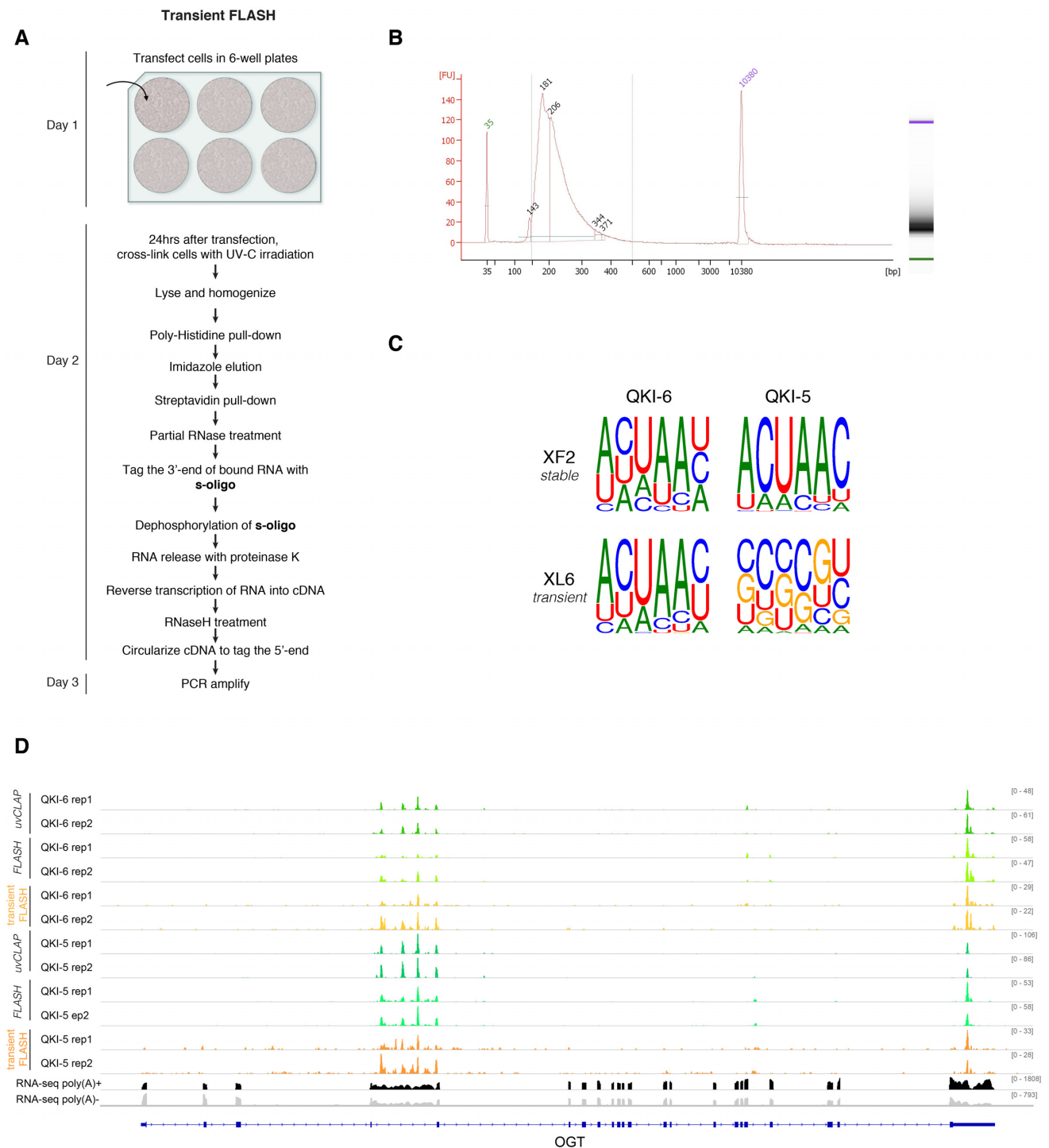
*QKI protocols (XF1, XF2, XF3 and XF4).* The clarified lysate in NLB is incubated for 5 min (maximum 10 min) with 25 µl of Dynabeads™ His-Tag Isolation and Pulldown beads (catalogue number: 10103D, Thermo Fisher Scientific), which are washed once with NLB and resuspended in 500 µl NLB. After the incubation, the beads are collected with a magnet, supernatant is removed, and the beads are washed with 800 µl of NLB. Elution is carried out with NLB supplemented with 250 mM imidazole, for 10 min on ice. The eluate is then incubated with 25 µl of Dynabeads™ MyOne™ Streptavidin C1 beads (catalogue number: 65002, Thermo Fisher Scientific), which are washed once with NLB and resuspended in 500 µl NLB and incubated in the cold-room (~4°C) for 1 h. The supernatant is removed, and the beads are washed with LDS buffer (20 mM Tris-Cl pH 7.4, 0.5 M LiCl, 1 mM EDTA, 0.5% LiDS), PLB (20 mM Tris-Cl pH 7.4, 0.5 M LiCl, 1 mM EDTA, 1% SDS), HSB (50 mM Tris-Cl pH 7.4, 1 M NaCl, 1% IGEPAL CA-630, 0.1% SDS, 1 mM EDTA) and NDB (50 mM Tris-Cl pH 7.4, 100 mM NaCl, 0.1% Tween-20). The beads are then resuspended with 1 ml of NDB, to which 2 µl of TURBO DNase (AM2238, Thermo Fisher Scientific) and 10 µl of diluted RNaseI (1:2000 dilution in NDB, AM2294, Thermo Fisher Scientific) is added and incubated at 37°C for 3 min. The lysates are cooled on ice for 2 min, before removal of the supernatant. Beads are then washed once with HSB and once with NDB. The dephosphoryla-



**Figure 3.** (A) The experimental design to compare FLASH<sup>endo</sup> (i.e. with antibody) and FLASH<sup>tagged</sup> protocols for U2AF65. (B) Bioanalyzer traces of the sequencing libraries generated using antibodies against endogenously expressed U2AF65 (lanes 2 and 3) and using a cell line that expresses tagged U2AF65 (lanes 4 and 5). (C) Positioning of U2AF65 binding events with respect to the intron-exon boundaries where U2AF65 typically binds to the polypyrimidine tract. (D) Distribution of the U2AF65 sequence motif as calculated by GraphProt on PEAKachu peaks determined from alignments ('alts', top lane) and crosslinked nucleotides ('clnts', bottom lane). WCE: Whole cell extract. (E) An IGV screenshot of U2AF65 FLASH data in comparison to eCLIP data generated in human cell lines.



**Figure 4.** (A) The common FLASH protocol that was used to generate SR-FLASH data. (B) Schematic representation of the domain architecture of SR proteins used in this work. (C) An immunoblot showing the levels of purified proteins used for FLASH. (D) Bioanalyzer quantification of SR sequencing libraries. (E) A correlation plot for all the SR libraries and the internal GFP controls. (F) Extracted motifs and their structure for all the SR proteins. See Supplementary Figure S9 for longer motifs.



**Figure 5.** (A) Description of the transient FLASH protocol used for QKI-5 and QKI-6, with GFP as a negative control. (B) Bioanalyzer quantification of QKI libraries. (C) Motifs extracted from QKI libraries using GraphProt. (D) An IGV screenshot of transient FLASH and its comparison to uvCLAP and XF2 libraries.

tion of the 3'-cyclic phosphate is carried out at 37°C for 20 min in a 20 µl reaction that contains 10 µl of 2× PNK-MES buffer (50 mM MES pH 6.0, 100 mM NaCl, 10 mM MgCl<sub>2</sub>, 0.1% Tween-20), 0.5 µl RNasin (N2511, Promega), 1 µl of β-mercaptoethanol (0.1 M), 1 µl of T4 PNK (10 U/µl, M0201, NEB) and 7.5 µl of water. After the PNK-reaction, the beads are washed once with HSB and twice with NDB. Each sample is then ligated with a unique s-oligo at 25°C for 1 h, in a reaction mixture that contains 2 µl of 10× T4 RNA Ligase Buffer, 4 µl of PEG8000, 1 µl of s-oligo (10 µM), 2 µl of ATP (1 mM), 0.5 µl of RNasin Plus (40 U/µl), 1 µl T4 RNA Ligase 1 (M0204L, NEB) and 9.5 µl of water. The beads are then washed once with HSB, once with NDB. At this stage relevant samples are mixed as they are now uniquely tagged. The 3'-phosphate group of the s-oligo is removed with T4 PNK, with the reaction setup described above, after which the beads are washed once with HSB and once with NDB.

**For XF1 and XF2.** The beads are resuspended with 100 µl of Proteinase K mix (100 mM Tris-Cl pH 7.4, 50 mM NaCl, 0.1% Tween-20, 10 mM EDTA, 0.1% SDS, 10 µl Proteinase K [20 mg/ml 25530049, Thermo Fisher Scientific]), and incubated at 37°C to digest all proteins and release the RNA into solution. The RNA is then purified using the Oligo Clean & Concentrator kit (Zymo Research, D4060) with 200 µl of binding buffer and 400 µl of ethanol for binding to the column and 9.5 µl of water for elution. The eluted RNA is reverse-transcribed at 42°C for 10 min, 50°C for 10 min and 55°C for another 10 min with SuperScript III in a reaction mixture that contains 2 µl of 10× Buffer, 1 µl of 10mM dNTPs, 4 µl of 25 mM MgCl<sub>2</sub>, 2 µl of 0.1 M DTT, 1 µl of RNaseOUT, 1 µl of SuperScript III and 9 µl of the RNA eluate.

**For XF1.** About 2.5 µl of 1 M NaOH is added to the reverse-transcription reaction which is then incubated at 70°C for 10 min. The reaction is then neutralized with the addition of 2.5 µl of 1 M HCl. To the neutralized and cooled reaction mixture, 47.5 µl of 10 mM Tris-Cl (pH 7.4), 0.75 µl of ATP (0.1 M), 0.75 µl of β-mercaptoethanol (0.1 M) and 1 µl of T4 PNK is added and the tubes are incubated at 37°C for 15 min to phosphorylate the 5'-ends of cDNA molecules, which is necessary for circularization. The phosphorylated cDNA is then purified with Oligo Clean & Concentrator kit with 150 µl of binding buffer and 300 µl of ethanol for binding to the column. About 6.5 µl of water was used for the elution.

**For XF2.** About 1 µl of RNaseH is added to the reverse-transcription reaction and incubate at 37°C for 20 min. About 54 µl of water is added to the reaction to bring the volume to 75 µl, after which the phosphorylated cDNA is purified with Oligo Clean & Concentrator kit with 150 µl of binding buffer and 300 µl of ethanol for binding to the column. About 6.5 µl of water was used for elution.

**For XF3 and XF4.** Directly after ligation of the s-oligo and appropriate mixing of samples, the beads are resuspended with the following reverse-transcription mixture: 4 µl of 5× Buffer, 1 µl MonsterScript, About 15 µl of water, and

incubated at 42°C for 5 min, then at 60°C for 15 min, at 70°C for 5 min and finally at 75°C for 5 min.

**For XF3.** NaOH and T4 PNK treatment, same as XF1.

**For XF4.** RNaseH treatment, same as XF2.

5'-phosphorylated cDNA from XF1, XF2, XF3 and XF4 are circularised using CircLigaseII (Lucigen, CL9021K) in a 10 µl reaction consisting of 1 µl of 10× Buffer, 0.5 µl of 50 mM MnCl<sub>2</sub>, 2 µl of 5 M Betaine, 0.5 µl of CircLigaseII and 6 µl of eluted 5'-phosphorylated cDNA. The reactions are carried out at 60°C in a hybridization oven (air-incubator) for ~16 h. The circularization reaction is then used, without purification, in a PCR reaction: 20 µl of 2× NEBNext Q5 Master Mix, 1 µl of P5 primer (10 µM, universal), 1 µl of P3 primer (10 µM, barcoded), 9 µl of cDNA, 8 µl of water. The cycle number is either guessed based on experience or determined using a qPCR reaction where 1 µl of circularized cDNA is used and the C<sub>t</sub> value is used as the final cycling number with the NEBNext enzyme (NEB, M0541S). Final PCR reaction is cleaned up with 1.5× AMPure beads (Agencourt, A63881), twice to completely remove leftover primers and eluted with 10mM Tris.Cl (pH 8) supplemented with 0.1% Tween-20. Sequencing libraries are quantified with Qubit DNA HS assay (Thermo Fisher Scientific, Q32851) and with a Bioanalyzer High Sensitivity DNA chip (Agilent, 5067–4626) and submitted for high-throughput sequencing to the Deep Sequencing Facility at the Max Planck Institute of Immunobiology and Epigenetics.

**XF5, XF6, XF7 and XF8.** The clarified lysate is incubated for 60 min with 25 µl Dynabeads™ Protein G beads (10003D, Thermo Fisher Scientific), coupled to 1 µl of FLAG-M2 mAb (Sigma, F1804) then washed once with NLB. The beads are then resuspended with 1 ml of NDB and the protocols described for XF1-4 above are followed exactly, in that XF5 ~ XF1, XF6 ~ XF2, XF7 ~ XF3 and XF8 ~ XF4. For an overview of these protocols see Figure 2, Supplementary Figures S1 and S2.

**U2AF65 protocols (XL17, XL18, XL21, XL22).** For XL17 and XL18 the XF6 protocol was followed with the following modifications: the FLAG mAb was replaced with a monoclonal antibody against U2AF65 (Sigma U4758, ~1 mg/ml, 5 µl of mAb coupled to 25 µl of Dynabeads™ Protein G beads), for XL17 the RNaseI was diluted 1:500 in NDB, for XL18, RNaseI dilution was kept at 1:2000.

For XL21 and XL22, the XF2 protocol was followed with an RNaseI dilution of 1:1000, using a stable cells that expresses U2AF65<sup>3x<sup>F</sup>BH</sup>. Additionally, for the XL21 library, nuclei was isolated first by incubating the cells on ice for 10 min with HLB (10 mM HEPES.Cl pH 7.9, 1.5 mM MgCl<sub>2</sub>, 10 mM KCl), after which the IGEPAL CA-630 was added to 0.5% (final) and the cells were left on ice for 5 min. The nuclei are then pelleted by centrifugation at 500g for 2 min, washed with HLB once, and re-suspended with 500 µl NLB before proceeding with sonication as with the other protocols described above.



**SR protocols.** XF2 protocol was followed with an RNaseI dilution of 1:2000 using cell lines that express SR proteins tagged with the 3xFHBH tag.

Also see Supplementary Table S1 for FLASH s-oligo sequences used for libraries reported in this study.

### Antibodies

U2AF65 (monoclonal mouse antibody Sigma U4758), Flag-HRP (Sigma A8592), FLAG-M2 mAb (Sigma, F1804).

### Bioinformatics

FLASH read processing and mapping was performed using the Galaxy platform (12). Adapters were trimmed using Flexbar (v2.5) (13). Libraries were demultiplexed using bctools (<https://github.com/dmaticzka/bctools>, v0.2.0) and Flexbar (v2.5). Custom FLASH adapters contained two barcodes and random nucleotides adjacent to the 3'-adapters according to the pattern NNB<sub>1</sub>B<sub>2</sub>NT<sub>1</sub>T<sub>2</sub>T<sub>3</sub>T<sub>4</sub>T<sub>5</sub>T<sub>6</sub>NN (N = random tag nucleotide; T = tag nucleotide; B = RY-space tag nucleotide). Random tags were used to merge PCR duplicates, regular tags were used to specify the pull-down condition. The semi-random RY-space tags were used to distinguish the biological replicates of libraries XF1-XF8. The remaining libraries employ regular tags to distinguish both pull-down condition and biological replicates. Possible readthroughs into the barcoded regions were removed by clipping 13 nt from the 3' ends of first mate reads. Reads were mapped to reference genome hg19 using bowtie2 (v2.2.6) (14) with parameters: `-very-sensitive -end-to-end -no-mixed -no-discordant -maxins 500`. We excluded all reads for which bowtie2 could identify multiple distinct alignments as indicated by the `XS:i` flag and used the alignments of the remaining uniquely mapped reads to determine crosslinking events as previously described (15).

Peaks used for motif detection were called using PureCLIP (v1.0.4) (16).

Binding motifs (Figures 4F and 5C, Supplementary Figures S1 and S9) were created using GraphProt (v1.1.7) (17). Unbound sequences used for training were selected by randomly placing peaks within genes with at least one binding site and at least 100 nucleotides apart from any bound site. GraphProt sequence models were trained based on the 60 nucleotides surrounding peak centers. GraphProt sequence-and-structure models additionally used 150 nucleotides up- and downstream to calculate RNA secondary structures.

GraphProt parameters were optimized using 200 bound and unbound sites. Motifs were generated based on the 20% highest-scoring sequences among the bound training instances.

Peaks used for the analysis shown in Figure 3D were called using PEAKachu (version 0.1.0, parameters: `- pairwise_replicates -m 0 -n manual -size_factors 1 1 0.75 0.75`) (18), using the two replicates of the respective pull-down condition as foreground and the two replicates of the corresponding control pull-down condition (specifically, IgG pull-downs for XL17-18 using cross-linked cells) as background. Since library XL22 was not paired with a

control, its peaks were called using the control of library XL21, which was generated using a UV-crosslinked cell-line expressing GFP<sup>3xFHBH</sup>. For XF1-8 libraries, a Flp-In T-REx 293 cell line with the parental tagging plasmid without an insert was constructed and used for background determination after UV-crosslinking (Supplementary Figure S2). For SR- and FLASH libraries with transiently expressed QKI proteins (Figures 4 and 5), a GFP<sup>3xFHBH</sup> cell line and a GFP<sup>3xFHBH</sup> expression plasmid were used respectively after UV-crosslinking. For the analysis shown in Figure 3D we created two sets of peaks using a) crosslinked nucleotides extended by 5 nucleotides up- and downstream and b) whole alignments of the corresponding reads.

Heatmaps showing pairwise Spearman correlations (Figure 4E and Supplementary Figure S6) and the principal component analysis (Supplementary Figure S2B) were created using deepTools (v3.1.1) (19) and are based on crosslinking event counts of non-overlapping 200 nucleotide genomic bins.

Bar plots showing overlap between U2AF2 data (FLASH and eCLIP) and transcript features and motif analysis of U2AF2 FLASH data in Supplementary Figure S8 were generated by RCAS (20) using crosslinking sites identified with PureCLIP (16).

Also see Supplementary Table S2 for mapping statistics for the libraries reported in this study.

## RESULTS

### The s-oligo enables rapid RNA-protein interaction mapping

We designed a new adapter (the s-oligo) for CLIP-seq experiments with the following considerations: **(a)** It should be ligated to RNA molecules with a 3'-OH end, **(b)** it should contain random nucleotides (also referred to as UMI: Unique Molecular Identifier) that can be used to remove PCR duplicates, **(c)** it should contain unique, pre-determined sequences for indexing and multiplexing, **(d)** it should contain sequence elements that should minimize and/or eliminate gel purifications and **(e)** it should contain chemical groups that suppress various ligation and PCR artifacts. Several different flavors of this adapter design can be used for different types of experimental requirements, however we will describe the use of the type depicted in Figure 1B.

The s-oligo is an RNA/DNA chimera, with a 5'-dangling single-stranded RNA that is made up of 6 non-random index ribonucleotides (XXXXXX) that are sandwiched between 7 random ribonucleotides that constitute the Unique Molecular Identifier (UMI, NNXXXXXXXXNNNN). Splitting the UMI into two and sandwiching the internal index between them serves two purposes: to reduce ligation bias, as the first two nucleotides ligated by the T4 RNA ligase (or other suitable ligase) on the adapter side are random, and to facilitate cluster identification in Illumina sequencers (21). The single-stranded RNA part is followed by a DNA duplex that is linked by a non-nucleic acid C18 spacer. Both the 5'- and the 3'-ends of the adapter are phosphorylated for reasons described below.

Ligation of the s-oligo to RNA is typically carried out by T4 RNA Ligase 1 (Figure 1D), which ligates the dangling, 5'-phosphorylated ssRNA of the adapter to the tar-

get RNA. The phosphate group at the 3'-end blocks self-ligation of the adapter to itself, which is essential to prevent generating 'empty insert' amplicons at the end of the protocol. After ligation, the 3'-phosphate group is removed with a phosphatase. The very 3'-end of the s-oligo base-pairs with the DNA segment that immediately follows the dangling RNA part (yellow nucleotides in Figure 1B), therefore this molecule is ready for reverse transcription (RT) immediately after dephosphorylation since the duplex-DNA part serves as a RT primer (purple nucleotides in Figure 1B). Reverse transcription is then carried out by adding an RNA-dependent DNA polymerase and dNTPs to the ligation products. After RT is completed, RNaseH is added to the reaction. This step is generally carried out to remove RNA and increase the efficiency of PCR (22). In FLASH, we exploit the fact that RNaseH requires, and leaves, a 5'-phosphate group as it degrades RNA/DNA hybrids progressively. By doing so, we not only remove all RNA but also effectively phosphorylate the very first deoxyribonucleotide moiety that follows the dangling ssRNA segment (Figure 1D). This creates a cDNA molecule that has a 5'-phosphate at its 5'-terminus and a 3'-OH at its 3'-terminus, which can be circularized using CircLigase. After circularization, cDNA is amplified directly by PCR without the linearization of the cDNA or further purification (Figure 1B-D).

How and why does this cloning procedure with the s-oligo dramatically speed up CLIP protocols? There are several reasons. One of the most persistent problems in ligation-mediated cloning of RNA (which is the general path taken by almost all CLIP protocols (3)) is the production of so-called 'adapter-dimers', which in most cases can be thought of as 'zero-length inserts'. These are products of either self-ligation of adapters or reverse-transcription primers which can eventually become PCR templates. We undercut these artifacts by blocking self-ligation by a 3'-phosphate group, which we later remove for reverse transcription (Figure 1D). Additionally, since we do not use additional reverse-transcription primers, the problems associated with using excess RT primer also do not exist in our protocol. Finally, we use a C18 spacer group, both to give flexibility to our adapter so that smaller inserts are circularized properly, and to function as a polymerase-block, enabling us to directly use circularization products for PCR. Electrophoresis of RNA-protein adducts on a polyacrylamide gel and isolation of distinct regions that are thought to contain RNAs that are bound to the protein-of-interest is a typical part of many CLIP protocols (3); however cutting a membrane or a gel is an intrinsically imprecise process and will inevitably be somewhat different for each protein, antibody, lane of gel and for each replicate; in addition to being labor-intensive, time-consuming and inefficient procedure. For this reason and as a result of the design choices detailed above, we do not use gel purifications at any stage of the protocol, and thus, starting from cells on plates and ending up with a dsDNA library ready for high-throughput sequencing, the protocol can be completed within ~10 h. However, we split the protocol into two days for convenience and to increase the efficiency of the cDNA circularization step (Figure 1C).

### Optimization of FLASH protocol with QKI

Flexibility of the FLASH protocol allows users to explore a large number of parameters in order to find the best possible experimental setup for a given purpose. The *Blueprint* depicted in Figure 1C, shows the general architecture of FLASH experiments. The first important technical fork in FLASH experiments is the choice of the purification method to enrich for the protein-of-interest. One can either use specific antibodies against a target (FLASH<sup>endo</sup>) or use a stringent affinity-purification scheme to pull down a tagged protein (FLASH<sup>tagged</sup>, see Figure 1C). We will describe the use of specific antibodies in the next section. However, since every antibody is unique in its biophysical properties, we optimized our method with tagged proteins first.

We used two cell lines: one expressing QKI-5<sup>3x<sub>F</sub>H<sub>B</sub>H</sup>, a nuclear isoform of QKI and another that expresses QKI-6<sup>3x<sub>F</sub>H<sub>B</sub>H</sup>, an isoform that has a significant cytoplasmic pool. These two isoforms and their *in vivo* targets were characterized in detail recently (6). We tested three forks in the FLASH protocol using these two tagged isoforms of QKI with defined binding behaviors (Figure 2A; Supplementary Figures S1 and S2A for the complete setup). First, we tested the impact of the purification scheme on the ability of FLASH to precisely locate binding and the quality of the final data, which is rarely, if at all, experimentally tested in other CLIP protocols. For this purpose, we used either single-step FLAG immunoprecipitation or polyhistidine (HIS) pull-down followed by streptavidin (STREP) purification. Second, we tested whether a FLASH protocol variation that does not have an RNA purification step would be possible. For researchers accustomed to published CLIP-seq protocols, this can be quite an unusual variation, in which the target RNA, crosslinked to the protein-of-interest on beads, is directly reverse-transcribed by simply resuspending the beads with a reaction mix containing a reverse-transcriptase enzyme and dNTPs, bypassing not only gel purifications and/or RNA isolation from nitrocellulose paper but also proteinase K treatment, phenol-chloroform or Trizol extractions and/or column purifications. This is possible in FLASH due to the geometry of the s-oligo, which comes with its reverse-transcription primer covalently linked to it (Figure 1B). Finally, we tested two different ways to phosphorylate the 5'-end of the cDNA after reverse transcription, which is necessary for the circularization reaction that follows it. Here, we used either base-catalyzed degradation of RNA with 0.1 M sodium hydroxide (NaOH) and subsequent phosphorylation of the 5'-ends with T4 Polynucleotide kinase (PNK) and Adenosine Triphosphate (ATP) or we simply added RNaseH to our samples immediately after the reverse-transcription reaction. RNaseH is generally used to remove RNA-DNA hybrids after reverse transcription to improve PCR performance, but due to the geometry of the s-oligo, it ends up leaving a phosphate group at the 5'-end of the cDNA molecule (Figure 1D). We processed a total of 48 samples distributed over 8 sequencing libraries for this experiment alone (Figure 2A and Supplementary Figure S2A). A PCA plot demonstrates that QKI libraries XF1-4 group with each other and close to uvCLAP libraries that were generated using the same cell lines, whereas XF5-8 libraries

cluster close together but further from uvCLAP libraries (6) (Supplementary Figure S2B).

In summary, we observed that HIS-STREP purifications (XF1-4, for RNA target distributions, see Supplementary Figure S3) performed better than FLAG (XF4-8, for RNA target distributions, see Supplementary Figure S4) purifications without any compromises on protein purity as judged by silver-stained gels (see (6)). Second, we observed no significant differences between RNaseH-mediated phosphorylation of cDNA ends and T4 PNK and ATP-mediated phosphorylation of cDNA ends after hydrolysis of RNA with 0.1M NaOH (XF1 versus XF2, XF3 versus XF4, XF5 versus XF6 and XF7 versus XF8, Figure 2B–D). Finally, skipping RNA purification and carrying out reverse transcription on beads did have an interesting effect on the data. As can be seen in Figure 2D, the libraries prepared with proteinase K treatment and RNA-purification steps (XF1-2 and XF5-6) the peak of RBD binding shows abrupt end at a QKI motif (AYUAA), with a smaller tail following it, presumably showing the cross-linking site and read-through events, while on-beads reverse-transcribed libraries (XF3-4 and XF7-8) do not show an abrupt end at QKI motif, and rather show an accumulation of events mostly before the motif and the tail is no longer seen. A plot that shows the distribution of the QKI motif among cross-linked fragments (Figure 2C) verifies these observations genome-wide, in that the motif peak shifts ~10 nucleotides to the left, similar to the example shown in Figure 2D. Presumably, this observation indicates that, without proteinase K treatment and RNA isolation, the reverse transcriptase is unable to reach the crosslinking site due to steric hindrance, and cannot travel beyond it, providing ‘RNA toeprints’ (23) rather than cross-linking sites. When all is taken into account, we recommend the XF2 protocol to be the canonical FLASH<sup>tagged</sup> protocol for most applications.

### Comparison of FLASH<sup>tagged</sup> and FLASH<sup>endo</sup> protocols with U2AF65

Using tagged proteins has many advantages in CLIP experiments, especially when the tag or the tag combination allows for stringent purifications that can efficiently eliminate co-purifying proteins, as in FLASH (Supplementary Figure S5 also see (6)). Such purity is virtually impossible to obtain with protocols that use specific antibodies, even when coupled to gel electrophoresis and nitrocellulose transfers, as co-purifying proteins with similar molecular weight to the protein-of-interest will co-migrate on the polyacrylamide gel as well. Moreover, tagged proteins make it easy to study derivatives of the protein of interest; be it deletions, point mutations or domain swaps, together with matched negative controls. In addition, the timing and amount of tagged protein can be modulated with inducible systems for expression or degradation of the protein of interest. When endogenous expression is preferable over ectopic expression, tags can be inserted into the endogenous locus using CRISPR-Cas9 with relative ease (24). However, not all proteins tolerate affinity tags, and even with the most stringent phenotypic tests claiming otherwise, affinity tags will usually affect some aspects of the target protein. For example, a tagged protein will definitely diffuse slower than

the native untagged protein, which may or may not affect the function or RNA targets of the protein. Furthermore, tagging target proteins may simply not be an option when working with challenging samples such as tissue biopsies or post-mortem brain samples (25). We thus generated a stable cell line expressing U2AF65<sup>3x<sup>F</sup>HBH</sup>, generated sequencing libraries with the FLASH<sup>tagged</sup> protocol, and compared the results to a FLASH<sup>endo</sup> protocol, where we used a monoclonal antibody against U2AF65 (MC3) which is frequently used in CLIP experiments (26). The FLASH<sup>endo</sup> protocol is almost identical to the FLASH<sup>tagged</sup> protocol, the only differences being the use of protein G-coupled paramagnetic beads, a single-step purification using the specific antibody and the skipping of stringent wash buffers that contain 0.5% LiDS or 1% SDS to preserve antibody-antigen interactions (Figure 3A). As with FLASH<sup>tagged</sup> protocols, no gels were used in this protocol variant. With all libraries (Figure 3B) (both endo and tagged) we could recover the known binding preference of U2AF65, both in terms of sequence identity (polypyrimidine tracts) and the positioning of its targets (upstream of exon-intron boundaries) (Figure 3C). The recovered motifs were sharply centered around cross-linking sites in both endo and tagged libraries, pointing to single-nucleotide resolution at binding sites (Figure 3D). As expected, the FLASH<sup>endo</sup> libraries contained higher coverage on exonic sequences compared to FLASH<sup>tagged</sup> libraries, possibly due to the inability to wash away some of the binding partners such as U2AF35 or SR proteins due to reduced stringency of washing conditions, a compromise made to preserve antibody-antigen interactions (also see Supplementary Figures S7 and S8 for RNA target distributions). Interestingly, a similar shoulder of exonic enrichment is also observed in libraries generated using eCLIP, a protocol which uses a polyclonal antibody against U2AF65, and goes through both PAA-electrophoresis, transfer to nitrocellulose paper and isolation of protein-RNA adducts from nitrocellulose. Finally, both FLASH<sup>tagged</sup> and FLASH<sup>endo</sup> profiles agree well with published eCLIP profiles both at canonical and non-canonical U2AF65 binding sites (Figure 3E) (4). In summary, FLASH<sup>endo</sup> protocol produces high-quality data when used with U2AF65 monoclonal antibodies and can be completed within 1.5 days. Other antibodies should be tested on a case-by-case basis, and where possible, should be compared to or supplemented with FLASH<sup>tagged</sup> data.

### Profiling ten human SR proteins with FLASH

SR proteins are an important group of RNA-binding proteins that are involved in splicing and export of mRNAs in all metazoans (27). They typically contain RRM(s) (RNA recognition motif) as RNA-binding domains and a repetitive region composed of serines and arginines, in which serine phosphorylation can affect their localization and/or modulate their RNA-binding properties. A recent study has generated iCLIP profiles of seven SR proteins in mice (SRSF1-7) using GFP-tagged transgenes (28). We decided to generate profiles for nine human SR proteins (SRSF1-7, SRSF9 and SRSF11) together with a point mutant of SRSF2 that is clinically important (SRSF2<sup>P95H</sup>) (29) as a resource (Figure 4A,B). We prepared stable cell lines ex-

pressing 3xFLAG-HBH SR proteins (Figure 4C) and followed the FLASH<sup>tagged</sup> protocol to generate sequencing libraries (Figure 4D). As expected, SR proteins showed exonic binding (Supplementary Figures S10 and S11) with a variety of sequence motifs summarized in Figure 4F. Interestingly, SRSF2<sup>P95H</sup> bound a motif similar to SRSF2, but with a cytosine replacing a guanine at the center of the motif (Figure 4F), which is similar to a recently published study which compared SRSF2 and SRSF2<sup>P95H</sup> using HITS-CLIP (30). This change is also visible in the correlation plot (Figure 4E), where SRSF2<sup>P95H</sup> moves closer to SRSF7 and SRSF3, both of which have a central cytosine in their recognition motif determined by our FLASH data. Using GraphProt (17), we also found that SR protein motifs are generally found in unstructured regions of RNA, with the exception of SRSF2 and its mutant SRSF2<sup>P95H</sup>, which are predicted to bind sequences that may form structures (Figure 4F, right).

### FLASH enables rapid *in vivo* RNA-profiling upon transient expression of RBPs

In order to be able to carry out a FLASH<sup>tagged</sup> protocol, it is necessary to either generate a cell line that expresses the protein-of-interest in an inducible manner, or the endogenous locus of the target should be modified so that the protein is expressed with the affinity tag necessary for FLASH experiments. These protocols are generally straightforward, but since the FLASH experiment itself takes only 1.5 days, generating these cell lines now becomes the bottleneck in high-throughput experiments. We thus attempted FLASH experiments on cells that are transiently transfected with plasmids encoding RNA-binding proteins (Figure 5). Once again, we used the QKI proteins as a reference, since we have a good understanding of their target sequence and target distribution, and transfected QKI-5<sup>3xFLAG-HBH</sup>, QKI-6<sup>3xFLAG-HBH</sup> and GFP<sup>3xFLAG-HBH</sup> expression plasmids into Flp-In T-REx 293 cells. We transfected cells in a 6-well format, and used one well (~1 million cells) as a replicate for each sample. Twenty-four hours post-transfection, cells were lysed and processed using the FLASH protocol optimized for tagged proteins (Figures 2A and 5A), ending with a sequence library similar to other FLASH libraries (Figure 5B). Both the duplications levels (avg. reads per event 5.38–5.73) and signal-to-noise ratios (6.71–6.83) were between libraries XF5-8 and XF1-4, which appear to be noisier than ideal (Supplementary Table S2). Nevertheless, it was possible to reproducibly recover the QKI motif with the QKI-6 sample (Figure 5C), and to recapitulate the *in vivo* binding pattern of both QKI-5 and QKI-6 (Figure 5D).

### DISCUSSION

FLASH solves several important challenges that are intrinsic to many CLIP-seq protocols and opens the way to solve other application-specific problems through rapid iteration cycles. This means that the experimenter can test, rather than assume, the effect of different parameters on the outcome through pilot experiments. For instance, instead of choosing one RNase concentration, a user can test a dozen concentrations and pool all samples for sequencing after

ligation of the s-oligo. The experimenter can test the effect of using different types of RNases, reverse-transcriptases, RNA ligases, or any other enzyme used in the protocol. The s-oligo can also be introduced into other CLIP-seq protocols in order to shorten the protocol by several days. Finally, since all of the purifications described in the protocol can also be carried out using paramagnetic bead-based purification strategies, FLASH has the unique potential to be completely automated using a liquid-handling system. This will enable truly high-throughput CLIP experiments to study and understand the role of hundreds of RNA-binding proteins and how they collectively regulate the fate of mRNAs as they emerge from RNA polymerase II.

### DATA AVAILABILITY

FLASH: GSE118265 (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE118265>). FLASH with transiently expressed vectors: GSE121115 (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE121115>). Peaks are available at <http://doi.org/10.5281/zenodo.1458136>.

### SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

### ACKNOWLEDGEMENTS

We would like to thank the Akhtar group members who have already tested and adapted the FLASH protocol, thus gave us useful feedback, especially to Giuseppe Semplicio and Claudia Keller Valsecchi. We are also thankful for the support provided by the Bioinformatics Core Facility and the Deep Sequencing Facility (especially Ulrike Bönisch) of the Max Planck Institute of Immunobiology and Epigenetics.

### FUNDING

German Science Foundation [CRC992 to A.A., R.B.; CRC1381 (Project-ID 403222702) to A.A.]; Germany's Excellence Strategy (CIBSS – EXC-2189 – Project ID 390939984) to A.A.; Max Planck Research Group Leader Program to T.A. Funding for open access charge: German Science Foundation [CRC1381].

*Conflict of interest statement.* A patent entitled, 'Cloning of single-stranded nucleic acid' (WO2017013005) has been filed.

### REFERENCES

- Lewis, C.J.T., Pan, T. and Kalsotra, A. (2017) RNA modifications and structures cooperate to guide RNA–protein interactions. *Nat. Rev. Mol. Cell Biol.*, **18**, 202.
- Hentze, M.W., Castello, A., Schwarzl, T. and Preiss, T. (2018) A brave new world of RNA-binding proteins. *Nat. Rev. Mol. Cell Biol.*, **19**, 327–341.
- Lee, F.C.Y. and Ule, J. (2018) Advances in CLIP technologies for studies of Protein-RNA interactions. *Mol. Cell*, **69**, 354–369.
- Van Nostrand, E.L., Pratt, G.A., Shishkin, A.A., Gelboin-Burkhart, C., Fang, M.Y., Sundararaman, B., Blue, S.M., Nguyen, T.B., Surka, C., Elkins, K. *et al.* (2016) Robust transcriptome-wide discovery of RNA-binding protein binding sites with enhanced CLIP (eCLIP). *Nat. Methods*, **13**, 508–514.

5. Zarnegar,B.J., Flynn,R.A., Shen,Y., Do,B.T., Chang,H.Y. and Khavari,P.A. (2016) irCLIP platform for efficient characterization of protein-RNA interactions. *Nat. Methods*, **13**, 489–492.
6. Maticzka,D., Ilik,I.A., Aktas,T., Backofen,R. and Akhtar,A. (2018) uvCLAP is a fast and non-radioactive method to identify in vivo targets of RNA-binding proteins. *Nat. Commun.*, **9**, 1142.
7. Darbelli,L. and Richard,S. (2016) Emerging functions of the Quaking RNA-binding proteins and link to human diseases. *Wiley Interdiscip. Rev. RNA*, **7**, 399–412.
8. Pilotte,J., Larocque,D. and Richard,S. (2001) Nuclear translocation controlled by alternatively spliced isoforms inactivates the QUAKEING apoptotic inducer. *Genes Dev.*, **15**, 845–858.
9. Hayakawa-Yano,Y., Suyama,S., Nogami,M., Yugami,M., Koya,I., Furukawa,T., Zhou,L., Abe,M., Sakimura,K., Takebayashi,H. *et al.* (2017) An RNA-binding protein, Qki5, regulates embryonic neural stem cells through pre-mRNA processing in cell adhesion signaling. *Genes Dev.*, **31**, 1910–1925.
10. Hafner,M., Landthaler,M., Burger,L., Khorshid,M., Hausser,J., Berninger,P., Rothballer,A., Ascano,M Jr, Jungkamp,A.-C., Munschauer,M. *et al.* (2010) Transcriptome-wide identification of RNA-binding protein and microRNA target sites by PAR-CLIP. *Cell*, **141**, 129–141.
11. Howard,J.M. and Sanford,J.R. (2015) The RNAissance family: SR proteins as multifaceted regulators of gene expression. *Wiley Interdiscip. Rev. RNA*, **6**, 93–110.
12. Afgan,E., Baker,D., Batut,B., van den Beek,M., Bouvier,D., Cech,M., Chilton,J., Clements,D., Coraor,N., Grünig,B.A. *et al.* (2018) The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2018 update. *Nucleic Acids Res.*, **46**, W537–W544.
13. Dodt,M., Roehr,J.T., Ahmed,R. and Dieterich,C. (2012) FLEXBAR-Flexible barcode and adapter processing for Next-Generation sequencing platforms. *Biology*, **1**, 895–905.
14. Langmead,B. and Salzberg,S.L. (2012) Fast gapped-read alignment with Bowtie 2. *Nat. Methods*, **9**, 357–359.
15. Ilik,I.A., Quinn,J.J., Georgiev,P., Tavares-Cadete,F., Maticzka,D., Toscano,S., Wan,Y., Spitale,R.C., Luscombe,N., Backofen,R. *et al.* (2013) Tandem Stem-Loops in roX RNAs act together to mediate X chromosome dosage compensation in drosophila. *Mol. Cell*, **51**, 156–173.
16. Krakau,S., Richard,H. and Marsico,A. (2017) PureCLIP: capturing target-specific protein-RNA interaction footprints from single-nucleotide CLIP-seq data. *Genome Biol.*, **18**, 240.
17. Maticzka,D., Lange,S.J., Costa,F. and Backofen,R. (2014) GraphProt: modeling binding preferences of RNA-binding proteins. *Genome Biol.*, **15**, R17.
18. Holmqvist,E., Wright,P.R., Li,L., Bischler,T., Barquist,L., Reinhardt,R., Backofen,R. and Vogel,J. (2016) Global RNA recognition patterns of post-transcriptional regulators Hfq and CsrA revealed by UV crosslinking in vivo. *EMBO J.*, **35**, 991–1011.
19. Ramírez,F., Ryan,D.P., Grünig,B., Bhardwaj,V., Kilpert,F., Richter,A.S., Heyne,S., Dündar,F. and Manke,T. (2016) deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res.*, **44**, W160–W165.
20. Uyar,B., Yusuf,D., Wurmus,R., Rajewsky,N., Ohler,U. and Akalin,A. (2017) RCAS: an RNA centric annotation system for transcriptome-wide regions of interest. *Nucleic Acids Res.*, **45**, e91.
21. Kircher,M., Heyn,P. and Kelso,J. (2011) Addressing challenges in the production and analysis of illumina sequencing data. *BMC Genomics*, **12**, 382.
22. Kitabayashi,M. and Esaka,M. (2003) Improvement of reverse transcription PCR by RNase H. *Biosci. Biotechnol. Biochem.*, **67**, 2474–2476.
23. Nilsen,T.W. (2013) Toeprinting. *Cold Spring Harb. Protoc.*, **2013**, 896–899.
24. Aktaş,T., Aşvar Ilik,İ., Maticzka,D., Bhardwaj,V., Pessoa Rodrigues,C., Mittler,G., Manke,T., Backofen,R. and Akhtar,A. (2017) DHX9 suppresses RNA processing defects originating from the Alu invasion of the human genome. *Nature*, **544**, 115–119.
25. Modic,M., Ule,J. and Sibley,C.R. (2013) CLIPing the brain: studies of protein-RNA interactions important for neurodegenerative disorders. *Mol. Cell Neurosci.*, **56**, 429–435.
26. Zarnack,K., König,J., Tajnik,M., Martincorena,I., Eustermann,S., Stévant,I., Reyes,A., Anders,S., Luscombe,N.M. and Ule,J. (2013) Direct competition between hnRNP C and U2AF65 protects the transcriptome from the exonization of Alu elements. *Cell*, **152**, 453–466.
27. Shepard,P.J. and Hertel,K.J. (2009) The SR protein family. *Genome Biol.*, **10**, 242.
28. Müller-McNicoll,M., Botti,V., de Jesus Domingues,A.M., Brandl,H., Schwich,O.D., Steiner,M.C., Turk,T., Poser,I., Zarnack,K. and Neugebauer,K.M. (2016) SR proteins are NXF1 adaptors that link alternative RNA processing to mRNA export. *Genes Dev.*, **30**, 553–566.
29. Yoshida,K., Sanada,M., Shiraishi,Y., Nowak,D., Nagata,Y., Yamamoto,R., Sato,Y., Sato-Otsubo,A., Kon,A., Nagasaki,M. *et al.* (2011) Frequent pathway mutations of splicing machinery in myelodysplasia. *Nature*, **478**, 64–69.
30. Liang,Y., Tebaldi,T., Rejeski,K., Joshi,P., Stefani,G., Taylor,A., Song,Y., Vasic,R., Maziarz,J., Balasubramanian,K. *et al.* (2018) SRSF2 mutations drive oncogenesis by activating a global program of aberrant alternative splicing in hematopoietic cells. *Leukemia*, **32**, 2659–2671.