Article

Pervasive transcriptome interactions of protein-targeted drugs

Linglan Fang [1], Willem A. Velema [1], Yujeong Lee[1], Lu Xiao [1], Michael G. Mohsen [1], Anna M. Kietrys[1] & Eric T. Kool [1,2] ✉

The off-target toxicity of drugs targeted to proteins imparts substantial health and economic costs. Proteome interaction studies can reveal off-target effects with unintended proteins; however, little attention has been paid to intracellular RNAs as potential off-targets that may contribute to toxicity. To begin to assess this, we developed a reactivity-based RNA profiling methodology and applied it to uncover transcriptome interactions of a set of Food and Drug Administration-approved small-molecule drugs in vivo. We show that these protein-targeted drugs pervasively interact with the human transcriptome and can exert unintended biological effects on RNA functions. In addition, we show that many off-target interactions occur at RNA loci associated with protein binding and structural changes, allowing us to generate hypotheses to infer the biological consequences of RNA off-target binding. The results suggest that rigorous characterization of drugs' transcriptome interactions may help assess target specificity and potentially avoid toxicity and clinical failures.

Dose-limiting toxicity is a routinely encountered problem in therapeutic drug development that is often discovered in clinical trials[1]. This is both detrimental to positive human health outcomes and costly in capital and human efforts. Even for approved drugs, dose-limiting toxicity results in adverse outcomes and post-marketing drug withdrawal[2]. These drug toxicities are commonly attributed to off-target binding to unintended cellular proteins[3]. Given RNA's broad bioregulatory roles in human physiology and the structural resemblance of protein-targeted drugs with known RNA-binding molecules[4,5] (Fig. 1a–c), we hypothesize that many protein-targeted Food and Drug Administration (FDA)-approved small-molecule drugs can interact with the human transcriptome (and RNA–protein interfaces) in vivo and may confer serious toxicity in patients as a result.

Studies provide support for this notion; for example, the investigational drug Dovitinib engages pre-miR-21 in triple-negative breast cancer cells, which guided the development of repurposed Dovitinib-RIBOTAC to target pre-miR-21 (ref. 6). A parallel microarray-based library-versus-library screening in vitro indicated that kinase inhibitors can bind RNA and topoisomerase inhibitors can engage RNAs such as pre-miR-21 (ref. 7). A computation-based

method (Inforna) also suggested that protein-targeted topoisomerase inhibitors can bind this same pre-miRNA[7]. These prior studies have led to the postulation that failed drug toxicity may be due in part to off-target RNA interactions[8]. Our analysis suggests that nearly 35% of all small-molecule FDA-approved drugs chemically resemble at least one known RNA-binding molecule (Tanimoto coefficient (Tc)[9], ≥0.7) (Fig. 1b and Supplementary Tables), and at least ten approved drugs are already documented as RNA binders in vitro (Supplementary Tables)[5]. Such chemical similarity with RNA-binding molecules also extends to drugs from preclinical studies and all phases in clinical trials (Fig. 1c).

Investigation of small-molecule–RNA interactions has been carried out recently via photocrosslinking (diazirine)[6,10–12], alkylation (chlorambucil)[13,14], in-line probing[15] and SHAPE[16], and these have emerged as useful tools to identify RNA–ligand interactions in vitro and recently in cultured cells[17]. Although elegant, application of these methods in vivo is potentially limited by transcriptional effects during prolonged probe treatment[18], cellular damage by ultraviolet irradiation[19] and nucleobase biases[20–22]. To address these issues and potentially complement prior methods that identify RNA–ligand interactions, we developed cell-permeable RNA 2′-hydroxyl (2′-OH) acylating

[1]Department of Chemistry, Stanford University, Stanford, CA, USA. [2]Sarafan ChEM-H Institute, Stanford University, Stanford, CA, USA. ✉e-mail: kool@stanford.edu
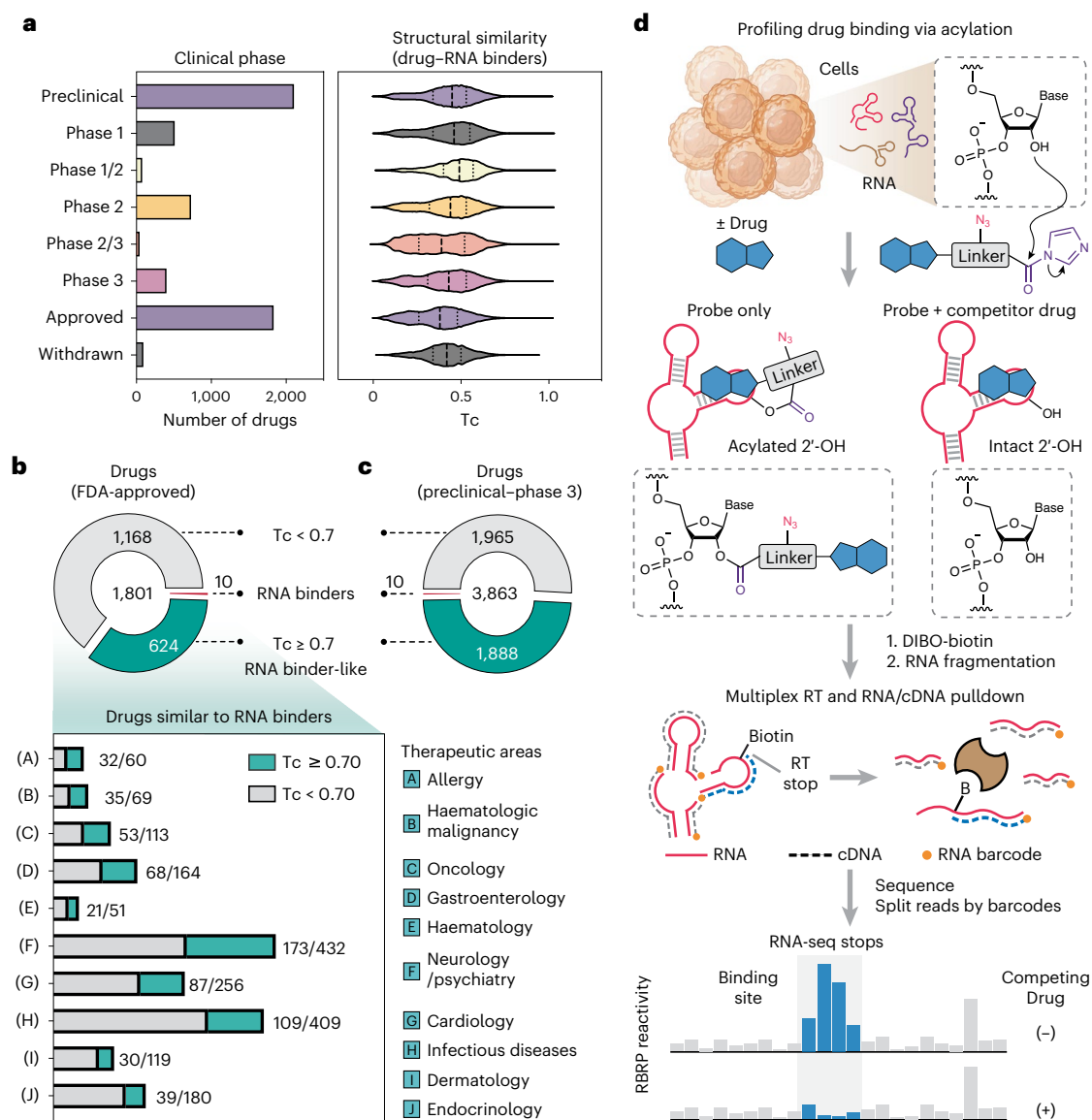
Fig. 1 | RBRP decodes transcriptome interaction of protein-targeted small-molecule drugs in vivo. a, Structure key-based 2D Tanimoto coefficient (Tc) characterizes pairwise structural similarity between 131 reported small-molecule RNA binders[5] and 6,043 drugs at various clinical phases. The number of small-molecule drugs at each clinical phase[58] is shown in the left panel. **b**, Pie chart showing the fractions of FDA-approved drugs that structurally resemble at least one known RNA binder in the top ten therapeutic areas (areas A–J). **c**, Fractions of drugs from preclinical to phase 3 stages that structurally resemble at least one known RNA binder as measured by Tc. See Supplementary Tables for data.

d, Schemes for comparative RBRP to identify transcriptome interactions of small-molecule drugs. An acylating analogue of the drug binds rapidly at specific RNA sites, promoting local acylation, which is analysed by RT stops via multiplexed RNA-seq. Pulldown of acylated RNAs greatly enhances the signal-to-noise ratio, and competition with unmodified drugs reveals specific sites. Binding sites exhibit statistically significant differences in RBRP reactivities. Blue, drug; red, azido group; purple, acylimidazole. Schematic elements in **d** were created with BioRender.com.

analogues of drugs that permit in vivo analysis of RNA–drug contacts in a sequence-independent manner. This reactivity-based RNA profiling (RBRP) methodology allows us to decode transcriptome interactions of drugs via deep sequencing and bioinformatics (Fig. 1d).

In this study, we employed RBRP to map in vivo transcriptome interactions of representative protein-targeted drugs in human cells. We provide evidence that RNA off-targets exist for FDA-approved protein-targeted drugs in cells. These transcriptome interactions are largely determined by the chemical structure of small-molecule drugs. We further demonstrate that drug engagement with RNA off-targets can exhibit target-associated cellular outcomes; for example, drug engagement with an RNA G-quadruplex (G4) in the YBX1 5′ untranslated region (UTR) inhibits the translation of genes encoded downstream of

the UTR. Further, drug engagement with small nucleolar RNA (snoRNA) sequences modulates downstream 2′-OH methylation of ribosomal RNAs. Moreover, RBRP enables mapping of the binding site across ~16,000 RNAs containing a poly(A) tail with nucleotide-level precision, and reveals intricate interplay of RNA–drug binding, RNA-binding proteins (RBPs) and RNA structural accessibility and dynamics.

Results and discussion
Decoding transcriptome interactions of protein-targeted drugs
Our profiling methodology involves the use of cell-permeable RNA acylation probes (employing acylimidazole-substituted linkers to react at RNA 2′-OH groups) to evaluate and quantify the tendency of

drugs to bind cellular RNAs (Fig. 1d). The binding of an acylimidazole conjugate of a drug to structured RNAs or protein–RNA interfaces should lead to enrichment of acylated 2′-OH near the drug-binding sites. We identify this binding-promoted acylation by modifying in vivo RNA mapping protocols[23], analysing messenger RNAs (mRNAs) and non-coding RNAs (ncRNAs) by poly(A) pulldown and deep sequencing the resulting libraries at high depth (>11 million reads per replicate). Specifically, RNA drug-binding sites are enriched with acylated 2′-OH groups, which cause reverse transcriptase (RT) to stop prematurely. We designed, optimized and validated a workflow to enrich these stops over random RNA breaks, random site acylation by acylimidazole warheads, non-specific binding events by biotin-mediated pulldown[23] and potential transcriptional changes upon treatment with the unmodified drug (Extended Data Fig. 1). This comparative workflow allows us to locate and quantify proximal binding sequences within the desired cellular RNA population; only sites that exhibit competition with the unmodified drug are scored as authentic drug-binding sites.

RBRP reveals the transcriptome interactions of Levofloxacin

We prototyped in vivo RBRP experiments with an acylimidazole conjugate of the small-molecule drug levofloxacin (Lev), containing an azido 'click' handle, in human embryonic kidney cells HEK293 (Fig. 2a–d). Lev is a member of a group of widely prescribed fluoroquinolone antibiotics known to cause neuropathy, fatigue and depression in patients[24]. Its structure suggests possible affinity for folded RNAs, given its fused aromatic rings and positive charge, and its close structural similarity with known RNA binders (Fig. 2b). To test this possibility, we treated HEK293 cells with the acylating analogue of Lev (Lev-AI, Fig. 2c) at 50 μM (near the maximum tissue concentration of Lev reported in patient; note that the exact cytosolic concentration of Lev is unknown)[25] for 30 min in the absence or presence of competing unmodified Lev in excess. We also performed RNA-sequencing (RNA-seq) experiments with HEK293 cells that were treated only with the competing drug or vehicle control (dimethylsulfoxide (DMSO)) (Extended Data Fig. 2a,b). After confirming that treatment with excess competing drug did not substantially alter the expression level of most cellular transcripts, we performed RBRP for acylating probe-treated polyadenylated transcripts. Deep sequencing results from RBRP showed a strong correlation of transcript expression value (RPKM) between two biological replicates (Pearson correlation $r = 1.00$) (Extended Data Fig. 3a–c). The concordance of RT-stop frequencies also remained high for most nucleotides with read depths higher than the optimized cut-off value (200) (Extended Data Fig. 3d–f). Thus, we performed bioinformatics analysis at the regions with read depths high enough to provide strong concordance among replicates. Specifically, we used a modified icSHAPE pipeline[23,26] (read depth, 200 as threshold) to quantify the yield of 2′-OH acylation at each nucleotide transcriptome-wide and generated 0.8 billion measurements in vivo (Extended Data Fig. 1). To remove random site acylation by the acylimidazole warheads and non-specific binding events by biotin-mediated pulldown, we subtracted the yield of 2′-OH acylation by the linker alone (Linker-AI) from that by Lev-AI. We also embedded RNA-seq experiments to account for potentially changed transcript abundance. Finally, RBRP reactivities at each site were calculated as the surplus yield of 2′-OH acylation for experiments performed in the absence over the presence of excess competing drug.

To stringently eliminate the low-confidence transcriptome interactions, we validated all the putative drug-binding sites with low-throughput reverse transcription quantitative PCR (RT–qPCR) experiments (Extended Data Fig. 4). Deep sequencing results from RBRP showed a reasonably strong correlation with RT–qPCR (Pearson correlation $r = 0.90$) for most putative binding sites, where RNA off-targets are defined as nucleotides with RBRP score ≥0.12 (Fig. 2e). In addition, we computed the two-tailed P-value of Welch's t-test, comparing RBRP scores at each nucleotide transcriptome-wide in with/without competitor experiments to quantify the statistical significance of the RBRP score.

These comparative RBRP data allowed us to identify probe-promoted, Lev-competable 2′-OH acylation as imprints of selective RNA–ligand interactions. Approximately 15 selective targets were identified in ~16,000 RNAs of HEK293 cells, all of which were independently confirmed by RT–qPCR (Extended Data Fig. 4). Because RBRP scores generally reflect the level of competition by unmodified drug (Fig. 2e), we can approximately rank the binding affinity of targets based on their RBRP score (Fig. 2f). We found that RNA off-targets of Lev were enriched for exons of mature mRNA (Supplementary Tables). For example, Lev engaged with the open reading frames (ORFs) of an mRNA that encodes interleukin enhancer binding factor 2 (ILF2). RBRP also showed that Lev targets ncRNAs such as snoRNA C/D box 110 (SNORD110) and a U1 small nuclear 88 pseudogene RNA (RNU1-88P), which are polyadenylated within their annotated 3′ ends in HEK293 cells[27] (Supplementary Fig. 1). These observations provide evidence for off-target transcriptome interactions of Lev in living human cells.

In addition, RBRP allowed us to estimate the binding affinity of RNA interactions of small-molecule drugs. For example, RBRP indicated that Lev targets the 5′ UTR of an mRNA that encodes the Y-box binding protein 1 (YBX1) (chr1:42,682,457-42,682,458) (Fig. 2g). With RT–qPCR, we showed dose-dependent 2′-OH acylation within the YBX1 5′ UTR in HEK293 cells (Fig. 2h). Pretreatment of HEK293 cells with increasing concentrations of unmodified Lev followed by treatment with Lev-AI demonstrated a dose-dependent reduction of 2′-OH acylation at the YBX1 5′ UTR site, with a clinically relevant IC_{50} value of $1.9 ± 0.8$ μM (Fig. 2i).

To gather more evidence for transcriptome interactions of Lev and further evaluate the performance of RBRP, we directly compared RBRP to an orthogonal method that utilizes diazirine-conjugated probes to identify RNA interactions[6,12,28,29]. We constructed a photocrosslinking analogue of Lev (Lev-diazirine) which structurally resembles that of Lev-AI (Extended Data Fig. 5a). We conducted profiling experiments with Lev-diazirine in HEK293 cells with or without excess competitor drug (Extended Data Fig. 5b), and photo-crosslinked RNAs were enriched following a reported protocol[6]. We then analysed these crosslinked RNAs with deep sequencing and identified putative RNA hits with the enrichment of RNA transcripts in the absence or presence of excess competing drug. Profiling with Lev-diazirine identified 14 putative RNA interactions (Extended Data Fig. 5c,d). Five 'hits' discovered by RBRP were orthogonally confirmed by diazirine-based profiling experiments, with another four hits by RBRP less confidently validated by Lev-diazirine (Extended Data Fig. 5e). Both probes also identified additional sets of unique off-target interactions, suggesting complementarity between the two methods. Notably, the nucleotide composition in the RBRP sequencing libraries was found to be distributed almost evenly across all four nucleobases (Extended Data Figs. 2j and 5f). This is in contrast with diazirine-based profiling methods that display pronounced preference toward guanine nucleobase[20–22]. Taken together, the data highlight the viability of RBRP for profiling ligand–RNA interactions.

Drug structure determines transcriptome interactions

Existing protein-targeted drugs differ greatly in their physicochemical and molecular recognition features, covering a broad multidimensional chemical space (Fig. 3a and Extended Data Fig. 6). To explore whether transcriptome interaction of protein-targeted drugs extends beyond Lev, we next investigated two additional compounds that are structurally distinct from Lev and have strongly divergent protein targets.

We first characterized the in vivo transcriptome interaction of hydroxychloroquine (HCQ), which was originally approved for the treatment of malaria and recently studied for treatment of COVID-19 infections, and causes retinopathy and cardiomyopathy of unknown
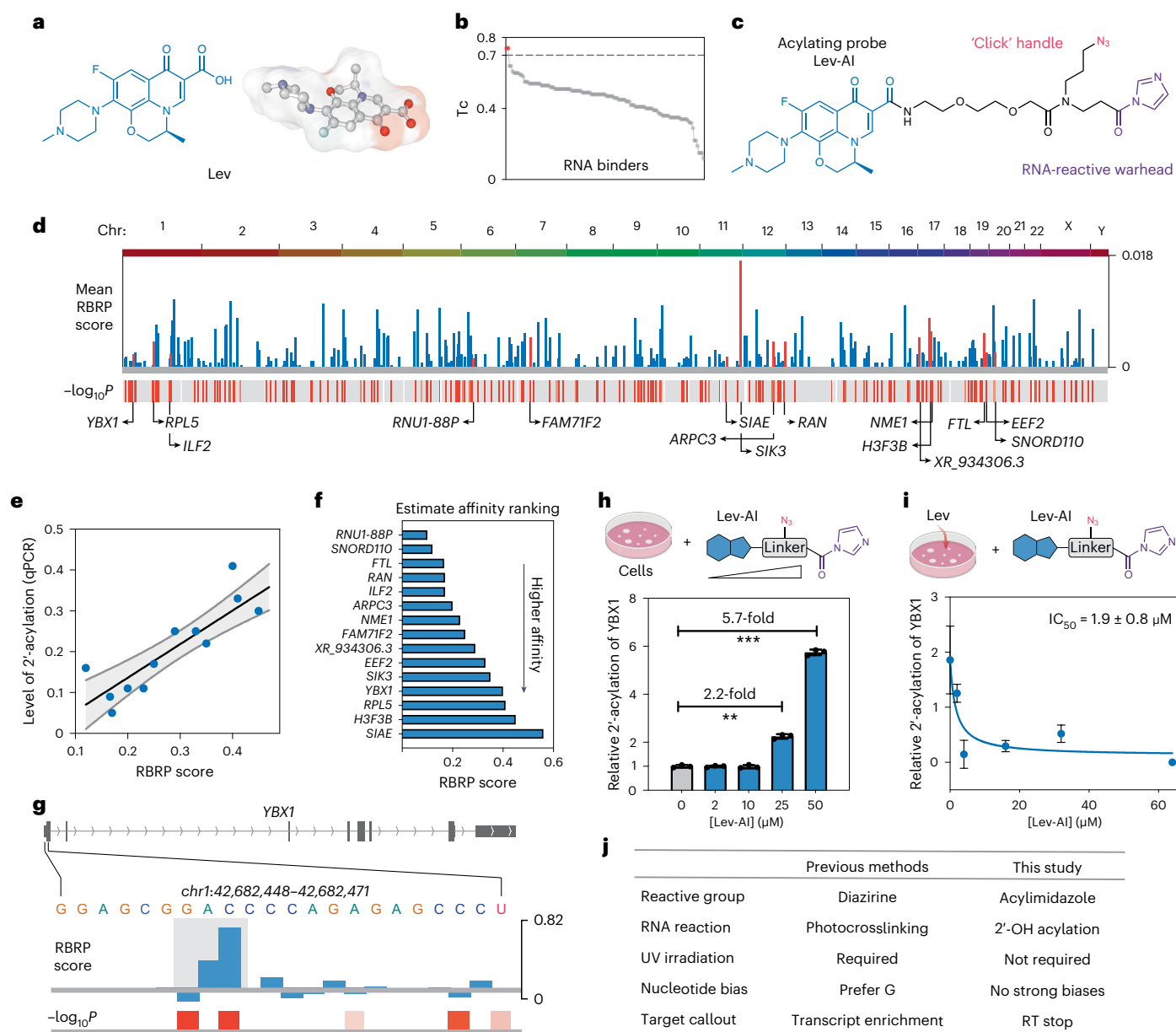
Fig. 2 | RBRP reveals transcriptome interaction of Lev. a, Left: chemical structure of Lev; right: 3D structure and Connolly surface of Lev. **b**, Dot plot showing the pairwise 2D Tc between Lev and known RNA binders. RNA binders with Tc ≥ 0.7 are red; other RNA binders are grey. The horizontal dashed line indicates Tc = 0.7. **c**, A modular design of acylating probe (Lev-AI) to identify the transcriptome interaction of Lev. Blue, drug part; black, linker; red, 'click' handle; purple, RNA-reactive warhead. **d**, UCSC tracks showing the mean RBRP scores (y axis) in live HEK293 cells along the human chromosomes (top) and the $-\log_2 P$ value of the Welch t-test (bottom) of the Lev-AI probe. Mean RBRP reactivity scores ranged from 0 to 0.018. **e**, Linear regression analysis showing high concordance between RBRP score and the level of 2′-acylation determined by independent qPCR for most targeted sites. Pearson's correlation $r = 0.90$. **f**, RBRP scores enables estimated ranking of binding affinity of RNA interactions

of Lev. **g**, UCSC tracks showing RBRP score at a locus in YBX1 mRNA (top) and the $-\log_2 P$-value of the Welch t-test (bottom), suggesting binding of Lev. Statistical significance was calculated with two-tailed Welch's t-tests. **h**, Lev-AI increases 2′-OH acylation in a dose-dependent manner at a binding site in YBX1 mRNA. Data shown are mean ± s.e.m., $n = 3$ independent biological replicates. Statistical significance was calculated with two-tailed paired Student's t-tests: **$P < 0.01$, ***$P < 0.001$. P values are 0.0019 (25 μM Lev) and 0.0001 (50 μM Lev), respectively. **i**, Unmodified Lev dose-dependently reduces 2′-OH acylation at the YBX1 binding site, with an IC_{50} of $1.9 ± 0.8$ μM. Data shown are mean ± s.e.m., $n = 3$ independent biological replicates. IC_{50} is mean ± s.e.m., $n = 3$ independent measurements. **j**, Comparing RBRP with previously reported diazirine-based methods for detecting RNA interactions of small molecules. Statistical significance was calculated with two-tailed Student's t-tests: $P < 0.01$, $P < 0.001$, $P < 0.0001$.

origins[30,31] (Fig. 3b). HCQ structurally resembles at least eight known RNA binders (Tc ≥ 0.7) and contains fused aromatic rings and a positive charge under the physiological conditions that may add affinity for folded RNAs (Fig. 3c). After confirming that treatment with excess HCQ did not substantially alter the abundance of most transcripts (Extended Data Fig. 2c), we conducted RBRP experiments with an acylating analogue of HCQ (HCQ-AI) in HEK293 at 50 μM (below the

maximum predicted cytosolic concentration of HCQ by a physiologically based pharmacokinetic model[32]; note that the clinically relevant cytosolic concentration of HCQ has not been experimentally determined) (Fig. 3d). The comparative RBRP workflow enabled us to identify probe-promoted, HCQ-competable 2′-OH acylation as imprints of selective RNA–ligand interactions. Approximately four selective targets were identified in ~16,000 RNAs of HEK293 cells (Fig. 3e),
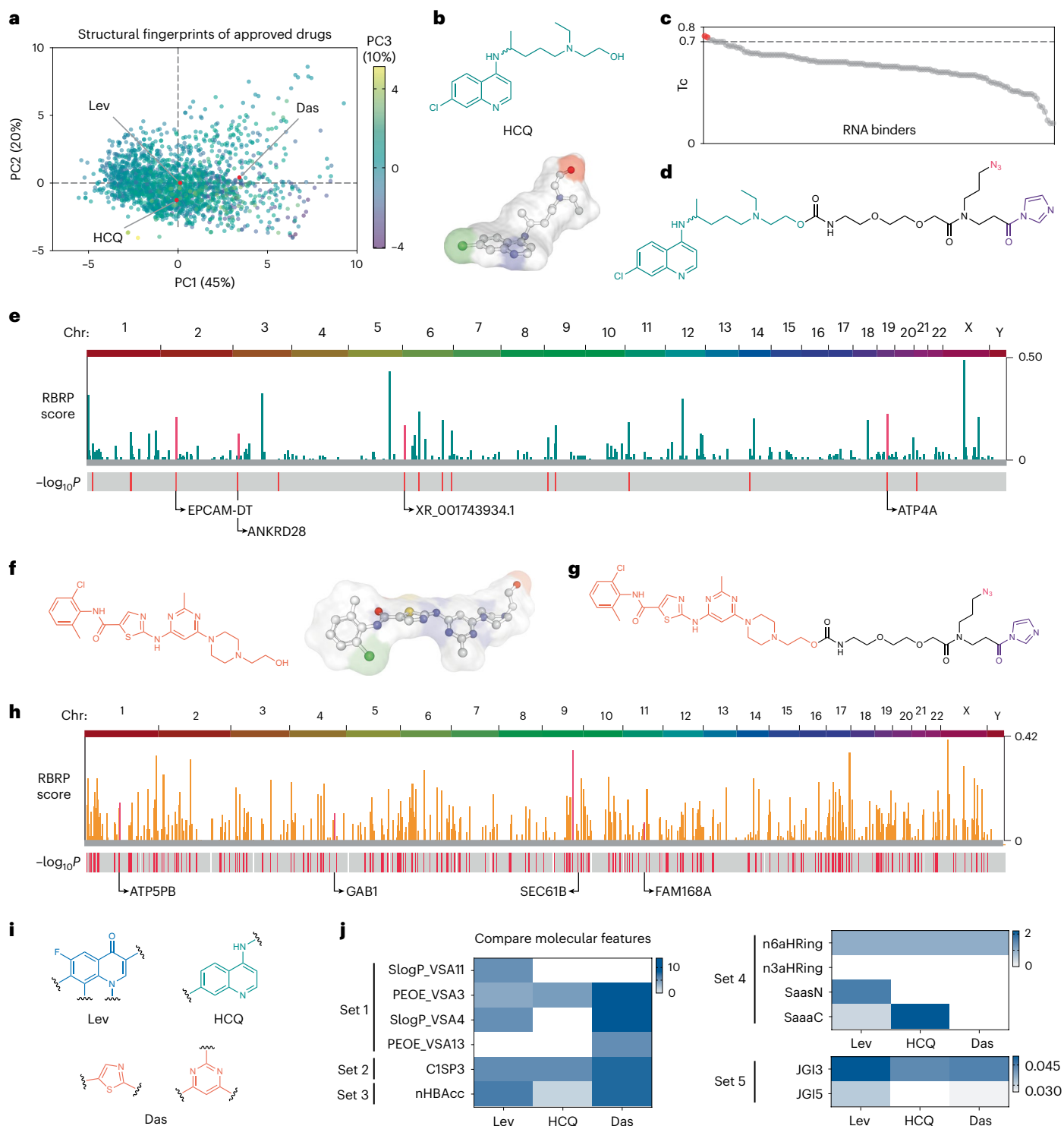
Fig. 3 | Structural fingerprints of small-molecule drugs determine their transcriptome interactions in vivo. a, Principal component analysis describing the structural variance within the selected FDA-approved drug library. Multivariate plot showing PC1, PC2 and PC3. ChemMine Tools calculate the Open Babel descriptors as structural fingerprints[59]. The contributions of each parameter and the loading plots for the first three PCs are shown in Supplementary Tables 1–3. **b**, Chemical structures of HCQ. **c**, Dot plot showing the pairwise 2D Tc between Lev and known RNA binders. RNA binders with Tc ≥ 0.7 are red; other RNA binders are grey. **d**, Chemical structure of acylating probe (HCQ-AI) to identify the transcriptome interaction of HCQ. Dark green, drug part; black, linker; red, 'click' handle; purple, RNA-reactive warhead. **e**, UCSC tracks showing the RBRP scores (*y* axis) in live HEK293 cells along the

human chromosomes (top) and the −log$_2$ *P* value of the two-tailed Welch's *t*-tests (bottom) of the HCQ-AI probe. **f,g**, Chemical structures of Das (**f**) and acylating probe (Das-AI) (**g**). **h**, UCSC tracks showing the RBRP scores (*y* axis) in live HEK293 cells along the human chromosomes (top) and the −log$_2$ *P* value of the two-tailed Welch's *t*-tests (bottom) of the Das-AI probe. **i**, Lev, HCQ and Das contain different nitrogen heterocycles including fluoroquinolone, aminoquinoline, pyrimidine and thiazole, which may affect RNA selectivity. **j**, Heatmap analysis showing that the three drugs have distinct structural features related to van der Waals surface area (set 1), *sp*^{3} character (set 2), hydrogen-bond acceptor (set 3), aromaticity and nitrogen rings (set 4) and topological charge (set 5), which are predictive of RNA–ligand interactions[36].

most of which were independently confirmed by RT–qPCR (Extended Data Fig. 7a,b).

We also profiled the transcriptome interactions of dasatinib (Das), a multitargeted kinase inhibitor known to cause neutropenia and myelosuppression, heart failure and serious pulmonary arterial hypertension[33] (Fig. 3f). The cationic Das contains heterocyclic flat structures, and amide linkages, reminiscent of groove-binding ligands for nucleic acids[34,35]. Base stacking and partial intercalation by Das are also probably feasible. RBRP with an acylating analogue of Das (Das-AI) revealed three distinct off-target interactions in cells (Fig. 3g,h). We identified and validated several RNA loci that were targeted by Das in HEK293 cells (Extended Data Fig. 7c). This is consistent with a previous microarray-based drug screening that showed RNA interaction of FDA-approved kinase inhibitors[7].

Profiling of these additional compounds via RBRP revealed distinct off-target interactions of drugs with the human transcriptome in cells, consistent with three drugs that differ greatly in their overall structures. The three drugs examined here contain different nitrogen heterocycles, including fluoroquinolone, aminoquinoline, pyrimidine and thiazole (Fig. 3i), that may contribute to their differential RNA selectivity[34,36]. The drugs also differ greatly in other features including van der Waals surface area, sp^3 character, aromaticity and nitrogen rings, and topological charge, which also probably influence their RNA recognition (Fig. 3j)[36,37].

Analyse RBP and RNA structure to infer off-target effects

Despite the existence of pervasive transcriptome interactions, whether these off-target drug binding events may cause unintended biological effects remained unclear. To infer potential effects, we sought to identify the binding loci that engage with RBPs, which extensively associate with RNAs to exert biological functions[38] (Fig. 4a). Taking the off-target snoRNA (SNORD110) as an example, SNORD110 binds to Lev and contains an oligo(A) tail at its 3′ end that allowed detection by RBRP[27] (Fig. 4b and Supplementary Fig. 1). Crosslinking immunoprecipitation followed by sequencing (CLIP-seq) experiments[39] showed that Lev engages with an RBP-binding sequence within SNORD110 (Fig. 4b). At this locus, Gene Ontology analyses demonstrated that the underlying RBPs possess numerous RNA-processing functions (for example, rRNA modification) (Fig. 4c), suggesting potential perturbation of these functions by Lev. Two RBPs, nucleolar protein 58 (Nop58) and fibrillarin (FBL), form the small nucleolar ribonucleoprotein complex (SnoRNP) with SNORD110, which mediates site-specific 2′-OH methylation of nascent pre-rRNA at 18S rRNA:U1288 and impacts the translational functions of the mature ribosome (Fig. 4d)[40,41]. qPCR and deep sequencing validated the selective SNORD110 interaction of Lev (Extended Data Fig. 4 and Supplementary Fig. 2). Titration of unmodified Lev in HEK293 cells revealed dose-dependent enhancement of 2′-OH methylation at 18S rRNA:U1288 at clinically relevant doses, with an EC_{50} of 10 ± 2.1 µM (Fig. 4e). Thus, comparative analysis of RBRP (drug binding) and existing iCLIP (RBP binding) data may allow us to generate testable hypotheses to infer possible biological outcomes of RNA off-target binding.

RNA structure also plays a central role in many cellular processes, including RNA transcription, translation and degradation[42–44], and is subject to extensive regulation in vivo by RBP interactions that can modulate intermolecular accessibility to structural probing reagents. Thus, we also searched for functional binding events that alter folded structure at RNA loci. SHAPE analysis of the YBX1 5′ UTR in cell lysates demonstrated that Lev enhances the structural accessibility of G11 in the YBX1 5′ UTR at an RBP-binding sequence (Fig. 4f). The underlying RBPs at this YBX1 locus are involved in numerous cellular functions including negative regulation of translation (Fig. 4g), leading us to wonder whether Lev may not only alter RNA structure at this locus, but also modulate the translation of YBX1 in cells in an off-target way.

Levofloxacin targets an RNA G4 to inhibit YBX1 translation

To test whether Lev modulates YBX1 translation, we treated HEK293 cells with increasing concentrations of Lev and quantified the cellular expression level of YBX1 with Western-blot analysis. We observed an ~30% reduction in YBX1 expression level in cells treated with Lev at clinically relevant doses (Fig. 5a).

To probe the underlying mechanism of action, we first investigated Lev's minimal binding sequence in the YBX1 5′ UTR. RBRP provides information about proximal drug-binding sites at near single-nucleotide level, allowing the identification of the drug's minimal binding sequences. The 5′ UTR of the Y-box binding protein 1 (YBX1) mRNA contains a G4 motif (5′-GGTAGCGGGAGCGGAGAGCGG-3′)[45, 46], which is located 18 nucleotides downstream of the 5′ cap and 109 nucleotides upstream of the start codon (Fig. 5b). The Lev-AI probe acylates the YBX1 5′ UTR at a dinucleotide sequence that is immediately adjacent 3′ to the G4 sequence. Notably, an analogue of Lev, CX-5461 (pidnarulex), has been implicated as a G4 binder[47]. We have also shown tight engagement of Lev with YBX1 5′ UTR in cells at clinically relevant doses (Fig. 2i). Thus, we speculated that Lev might target the G4 sequence within the YBX1 5′ UTR.

To confirm the existence of a stable RNA quadruplex structure within the YBX1 5′ UTR, we biophysically characterized a synthetic RNA corresponding to the isolated sequence from the YBX1 5′ UTR with circular dichroism (CD) spectroscopy (Fig. 5c,d), confirming the formation of a parallel G4 structure within the isolated RNA sequence (Fig. 5c,d)[48].

To isolate the YBX1 5′ UTR as the drug-binding site of Lev, we biochemically quantified the impacts of Lev on the translation of genes encoded downstream of the YBX1 5′ UTR with a cell-free translation system coupled to a reporter gene assay (Fig. 5e,f). Strong dose-dependent inhibition by Lev of d2GFP translation was seen, with an $EC_{50} = 1.6 \pm 1.0$ µM, which is consistent with the binding affinity measured in cells (Fig. 5g), whereas no dose-dependent effect was seen with a scrambled UTR unable to form G4 structure. The results are consistent with previous observations that small molecules that target RNA G4 in 5′ UTRs can modulate the translation of mRNA[49,50].

Mutations can disrupt favourable target interactions of drugs, and can be applied to identify the mechanism of drug actions[51]. Thus, we engineered two mutants to disrupt G4 formation (Fig. 5e)[52]. G4-disrupting mutations (GG–AA) resulted in an increase in translation efficiency and loss of Lev response relative to the native sequence (wild-type (WT)) (Fig. 5h,i). These data also indicate that the natural YBX1 RNA G4 in its 5′ UTR has an inhibitory effect on translation. In addition, deletion of the G4 motif (YBX1-delG4) resulted in nearly complete resistance to Lev-mediated suppression of mRNA translation, which further documents this G4 as a Lev off-target binding site.

Finally, we showed that Lev targets the G4 at the YBX1 5′ UTR in living cells. YBX1 UTR reporter mRNAs were lipofected into HEK293 cells with or without Lev at a clinically relevant dose (10 µM). The translation kinetics of YBX1 reporter were monitored by green fluorescence signal in living HEK293 cells (Fig. 5j). We found that treatment with Lev markedly reduced the translation efficiency of G4-containing YBX1-WT reporter (Fig. 5k, left panel). In contrast, Lev did not affect the cellular translation efficiency of YBX1-mutG2A, further supporting that Lev targets YBX1 G4 in cells (Fig. 5k, right panel). Thus, Lev-dependent reduction of YBX1 protein expression is at least partially caused by Lev targeting the translation-suppressive G4 within the YBX1 5′ UTR in living cells.

Taken together, our data strongly suggest that FDA-approved protein-targeted drugs can cause unintended biological outcomes by engaging with off-target cellular RNAs. Stringent transcriptome-wide validation of drug-binding sites may provide more comprehensive insights into the potential mechanisms of dose-limiting drug side effects.
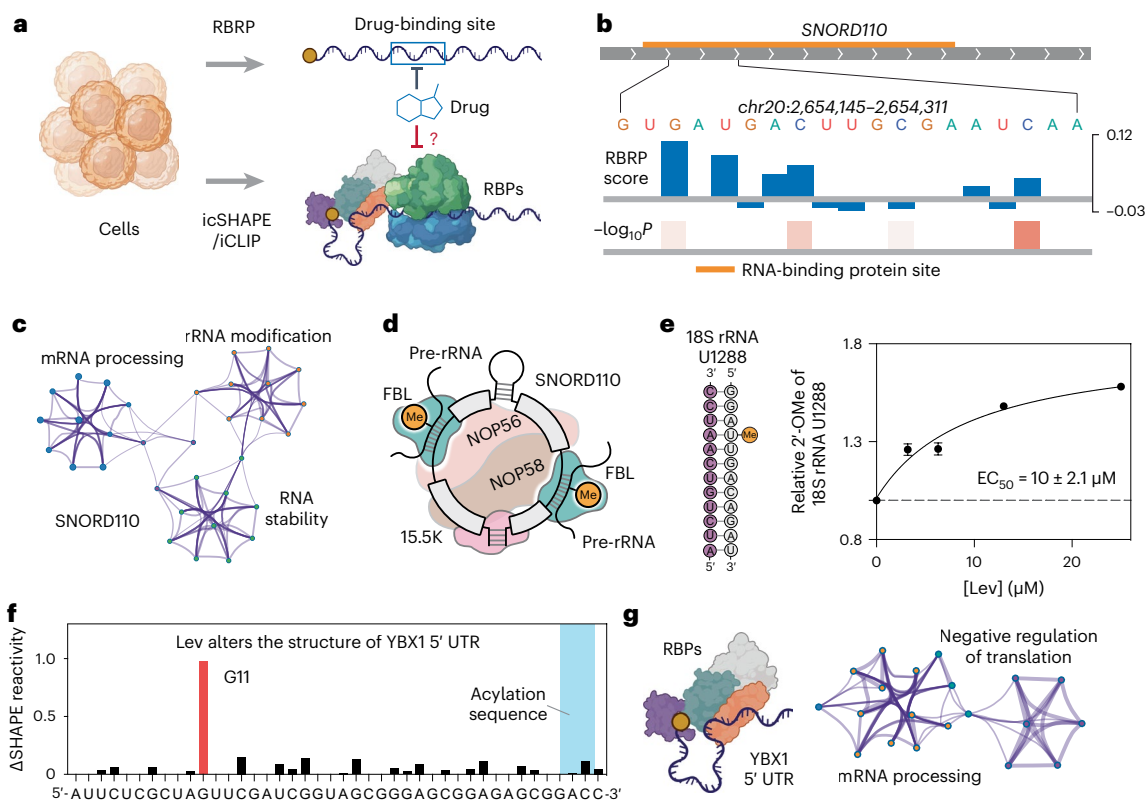
Fig. 4 | Altered RNA structure and RBP binding can associate with drug engagement. a, Meta-analysis reveals association among RBRP reactivity, RNA structural accessibility and RBP binding to infer biological outcomes of off-target RNA interactions. **b**, UCSC tracks showing RBRP score at an RNA locus in SNORD110-expressing sequence (top) and the $-\log_2 P$-value of two-tailed Welch's t-tests (bottom), suggesting binding of Lev. Orange, RBP-binding sequence. **c**, Metascape enrichment network showing the top non-redundant enrichment clusters of SNORD110-binding proteins proximal to the Lev-AI binding site. **d**, snoRNP complex guides 2′-OH methylation of the nascent pre-rRNA. Boxes indicate the conserved sequences of SNORD110. Fibrillarin (FBL), NOP56, NOP58 and 15.5K are small nucleolar proteins. Me indicates the ribosomal 2′-OH to be methylated. **e**, Lev dose-dependently enhances 2′-O-methylation of 18S rRNA U1288, with an $EC_{50} = 10 \pm 2.1$ μM. Data shown are mean ± s.e.m., $n = 3$ independent biological replicates. **f**, Acylation probing (SHAPE) showing Lev-enhanced structural accessibility of 2′-OH at G11 in the YBX1 5′ UTR, a locus proximal to the acylation sequence by Lev-AI. Values shown are the differential SHAPE reactivities at probe-binding sites in the presence versus the absence of 10 μM of Lev, determined by DNA fragment analysis. ΔSHAPE reactivity = SHAPE reactivity (Lev-treated) − SHAPE reactivity (DMSO). **g**, Illustration of RBPs binding to the YBX1 5′ UTR (left panel) and Metascape enrichment network showing the top non-redundant enrichment clusters of the YBX1 5′ UTR-binding proteins proximal to the Lev-AI binding site. Schematic elements in **a** and **g** were created with BioRender.com.

Conclusions

It is generally understood that small-molecule drugs can exhibit toxicity by off-target interactions within the human proteome[53]. Our data suggest that, in addition to proteome interactions, off-target engagement with specific RNAs in the human transcriptome may also be a mechanism by which small molecules cause unintended biological consequences. Each ostensibly protein-targeted drug tested in this study was found to target numerous transcripts in cells. This is consistent with previous observations that known drugs, including FDA-approved drugs, can bind RNAs (for example, mitoxantrone targets pre-miR-21), which was discovered with an in vitro microarray-based screening and later validated in cells[7]. Considering the high chemical similarity of many existing protein-targeted drugs with known RNA-binding molecules (Tanimoto coefficient, ≥0.7), this phenomenon is probably common among many FDA-approved drugs and those in clinical trials and preclinical studies. Given that ~90% of drugs fail to receive FDA approval and hundreds of previously approved drugs have been withdrawn[53,54], RNA off-targets probably contribute to failure rates. Thus, stringent characterization of transcriptome interactions for small-molecule drugs may be helpful in shedding light on these problems.

This study highlights the power of RBRP for target validation of small-molecule drugs, which requires the use of cell-permeable

RNA acylation probes (reacting at 2′-OH groups) and a comparative workflow to scout off-target interactions of drugs in the human transcriptome. Direct comparison with an orthogonal diazirine-based RNA-ligand profiling method validated and highlighted the power of RBRP to identify transcriptome interactions of drugs, complementing existing diazirine-based methodology. RBRP bridges a gap in RNA-sequencing technologies that currently lack the ability to impartially identify drug-binding sites at all four nucleotides (Extended Data Fig. 5f), without potential interference from extended incubation or ultraviolet light. The ability to quantify off-target interactions at each nucleotide in more than ~16,000 poly(A)+ RNAs enables unprecedented identification of altered RNA structural accessibility and RBP underlying drug engagement, and allows the users to infer the cellular consequences of off-target RNA binding. Notably, we quantified and observed minimal correlation between the RBRP and icSHAPE reactivities transcriptome-wide (Supplementary Fig. 3a,b), confirming that RBRP signals are drug-promoted rather than simply reflecting structure-dependent stochastic acylation. This observation is also confirmed by the fact that RBRP signals were minimally correlated among three different drugs (Supplementary Fig. 3c).

Our results indicate that many existing drugs may exert unintended biological effects through off-target interactions with the
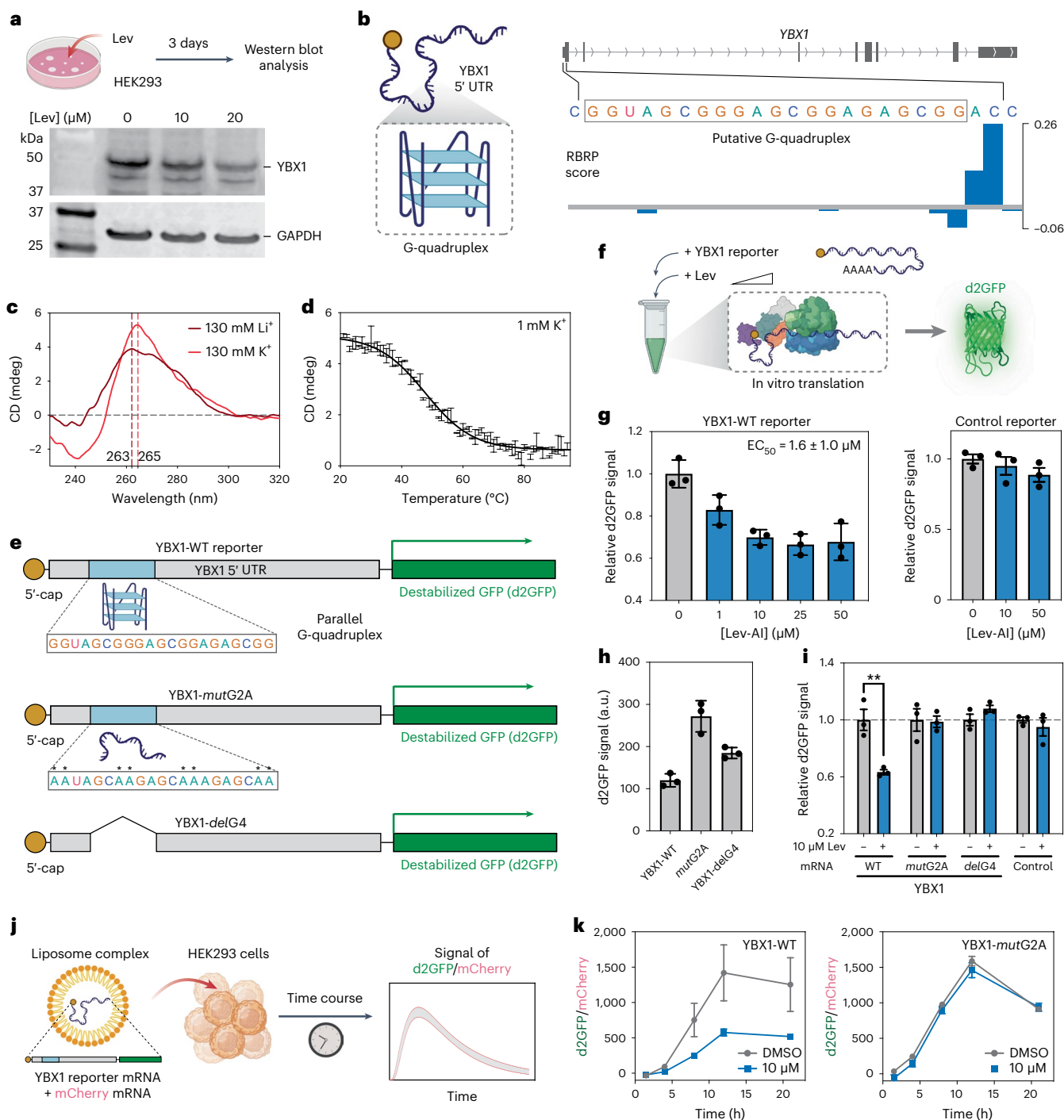
Fig. 5 | Lev targets a G4 structure in the YBX1 5′ UTR and inhibits translation.
a, Western blot analysis showing dose-dependent suppression of YBX1 expression level in HEK293 cells at clinically relevant doses. The experiment was independently conducted twice ($n = 2$) with similar results. **b**, Schematic representation of an RNA G4 within the YBX1 5′ UTR (left panel) and a UCSC track showing that Lev-AI acylates 2′-OH groups immediately 3′ to the RNA G4 in the YBX1 5′ UTR (right panel). **c**, CD spectrum of YBX1 G4 at a 10 μM strand concentration in 10 mM Tris, pH 7.0 and 130 mM of KCl or LiCl at 25 °C. The vertical dashed line indicates the wavelength of the positive peak in the CD spectrum. **d**, Melting CD curve of YBX1 G4 at a 10 μM strand concentration in 10 mM Tris, pH 7.0 and 1 mM KCl. Data represent mean ± s.e.m., $n = 2$ biologically independent experiments. **e**, Schematic representation of chimeric d2GFP reporter mRNA constructs. YBX1-WT (top), full-length wild-type YBX1 5′ UTR; YBX1-*mut*G2A (middle), GG to AA mutation within G4 sequence to disrupt quadruplex formation; YBX1-*del*G4 (bottom), G4 sequence deleted from the YBX1 5′ UTR. **f**, Workflow determining translation efficiency with cell-free in vitro

translation systems. **g**, Lev dose-dependently suppressed translation of YBX1-WT reporter, with an $EC_{50} = 1.6 ± 1.0$ μM (left). Lev did not substantially alter the translation of a control d2GFP reporter with a scrambled 5′ UTR (right).
h, Translation efficiency of three chimeric YBX1 reporter mRNAs as quantified by d2GFP signal. **i**, Relative translation efficiency of three chimeric YBX1 reporter mRNAs and a control mRNA with a scrambled 5′ UTR in the presence or absence of 10 μM Lev. Statistical significance was calculated with two-tailed Student's *t*-tests: **P < 0.01. *P* value is 0.0085. Data in **g**–**i** are mean ± s.e.m., $n = 3$ independent biological replicates. **j**, Workflow to determine translation efficiency of YBX1 reporter mRNA in HEK293 cells. Chimeric YBX1 reporters are co-transfected with mCherry mRNA. A fluorescence-based microplate reader monitors the translation of d2GFP and mCherry. **k**, Lev inhibits the translation of chimeric YBX1-WT reporter mRNA in HEK293 cells (left), while showing no effects on the translation of YBX1-*mut*G2A at a clinically relevant dose (10 μM) (right). Data represent mean ± s.e.m., $n = 3$ biologically independent experiments. Schematic elements in **a**, **b**, **e**, **f** and **j** were created with BioRender.com.

human transcriptome in vivo. For example, we find that off-target transcriptome interaction may unexpectedly affect 2'-OH methylation levels of ribosomal RNAs. Our data also show that the antibiotic Lev targets an RNA G4 in the YBX1 5' UTR and causes collateral inhibition of YBX1 translation in cells at clinically relevant doses. Not surprisingly, our findings provide evidence that transcriptome interaction of drugs depends greatly on their chemical structures[34,37,55]. Indeed, structural features such as positive charge and flat aromatic structures for groove binding and intercalation can add affinity for structured RNAs in vitro with low to moderate selectivity[34,35,56]. Consistently, our work shows that intercalator-like Lev extensively interacts with highly structured RNAs in cells, whereas Das is more structurally complex than a simple intercalator and has fewer RNA off-targets. Although our work provides strong evidence that transcriptome interaction can alter cellular outcomes in specific cases, it is likely that some off-target interactions within the human transcriptome will not cause unintended cellular consequences if the binding does not alter RNA conformation or RPB interactions. In addition, it remains possible that some drug targets will prove to be cell-type-dependent, which is not studied here and will require further investigation.

More broadly, we suggest that rigorous characterization of transcriptome interaction should be essential not only for existing drugs, but also future drugs. Animal toxicity testing fails to predict side effects and toxicity in ~50% of drugs between Phase I human trials and early post-market withdrawals[57], calling for comprehensive target validation during preclinical studies. Others in the field have also postulated that off-target RNA interactions probably fuel the failure of drug safety assessment[8]. Given that dose-limiting toxicity is a chief driver for clinical trial failures and drug withdrawal, stringent validation of drug targets in both the human proteome and transcriptome should not only reduce costs, but also allow the redistribution of resources towards lead compounds that are more likely to pass clinical trials. We envision that ultraviolet irradiation-free, minimally nucleobase-biased RBRP experiments can facilitate our understanding of RNA-mediated drug effects, indicate potential toxicity biomarkers and characterize toxicity mechanisms.

Online content

Any methods, additional references, Nature Portfolio reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41557-023-01309-8.

References

1. Lin, A. et al. Off-target toxicity is a common mechanism of action of cancer drugs undergoing clinical trials. *Sci. Transl. Med.* **11**, eaaw8412 (2019).
2. Onakpoya, I. J., Heneghan, C. J. & Aronson, J. K. Post-marketing withdrawal of 462 medicinal products because of adverse drug reactions: a systematic review of the world literature. *BMC Med* **14**, 10 (2016).
3. Harrison, R. K. Phase II and phase III failures: 2013–2015. *Nat Rev Drug Discov* **15**, 817–818 (2016).
4. Morgan, B. S. et al. R-BIND: an interactive database for exploring and developing RNA-targeted chemical probes. *ACS Chem. Biol.* **14**, 2691–2700 (2019).
5. Donlic, A. et al. R-BIND 2.0: an updated database of bioactive RNA-targeting small molecules and associated RNA secondary structures. *ACS Chem. Biol.* **17**, 1556–1566 (2022).
6. Zhang, P. et al. Reprogramming of protein-targeted small-molecule medicines to RNA by ribonuclease recruitment. *J. Am. Chem. Soc.* **143**, 13044–13055 (2021).
7. Velagapudi, S. P. et al. Approved anti-cancer drugs target oncogenic non-coding RNAs. *Cell Chem. Biol.* **25**, 1086–1094 (2018).
8. Disney, M. D. Targeting RNA with small molecules to capture opportunities at the intersection of chemistry, biology, and medicine. *J. Am. Chem. Soc.* **141**, 6776–6790 (2019).
9. Kim, S. Exploring chemical information in PubChem. *Curr Protoc* **1**, e217 (2021).
10. Suresh, B. M. et al. A general fragment-based approach to identify and optimize bioactive ligands targeting RNA. *Proc. Natl Acad. Sci. USA* **117**, 33197–33203 (2020).
11. Balaratnam, S. et al. A chemical probe based on the PreQ1 metabolite enables transcriptome-wide mapping of binding sites. *Nat. Commun.* **12**, 5856 (2021).
12. Mukherjee, H. et al. PEARL-seq: a photoaffinity platform for the analysis of small molecule–RNA interactions. *ACS Chem. Biol.* **15**, 2374–2381 (2020).
13. Velagapudi, S. P., Li, Y. & Disney, M. D. A cross-linking approach to map small molecule-RNA binding sites in cells. *Bioorg. Med. Chem. Lett.* **29**, 1532–1536 (2019).
14. Rzuczek, S. G. et al. Precise small-molecule recognition of a toxic CUG RNA repeat expansion. *Nat. Chem. Biol.* **13**, 188–193 (2017).
15. Regulski, E. E. & Breaker, R. R. In-line probing analysis of riboswitches. *Methods Mol. Biol.* **419**, 53–67 (2008).
16. Zeller, M. J. et al. SHAPE-enabled fragment-based ligand discovery for RNA. *Proc. Natl Acad. Sci. USA* **119**, e2122660119 (2022).
17. Tong, Y. et al. Transcriptome-wide mapping of small-molecule RNA-binding sites in cells informs an isoform-specific degrader of QSOX1 mRNA. *J. Am. Chem. Soc.* **144**, 11620–11625 (2022).
18. Chang, M. T. et al. Identifying transcriptional programs underlying cancer drug response with TraCe-seq. *Nat. Biotechnol.* **40**, 86–93 (2022).
19. Wurtmann, E. J. & Wolin, S. L. RNA under attack: cellular handling of RNA damage. *Crit. Rev. Biochem. Mol. Biol.* **44**, 34–49 (2009).
20. Qiu, Z., Lu, L., Jian, X. & He, C. A diazirine-based nucleoside analogue for efficient DNA interstrand photocross-linking. *J. Am. Chem. Soc.* **130**, 14398–14399 (2008).
21. Nakamoto, K. & Ueno, Y. Diazirine-containing RNA photo-cross-linking probes for capturing microRNA targets. *J. Org. Chem.* **79**, 2463–2472 (2014).
22. Di Antonio, M., McLuckie, K. I. & Balasubramanian, S. Reprogramming the mechanism of action of chlorambucil by coupling to a G-quadruplex ligand. *J. Am. Chem. Soc.* **136**, 5860–5863 (2014).
23. Spitale, R. C. et al. Structural imprints in vivo decode RNA regulatory mechanisms. *Nature* **519**, 486–490 (2015).
24. Marchant, J. When antibiotics turn toxic. *Nature* **555**, 431–433 (2018).
25. von Baum, H., Bottcher, S., Abel, R., Gerner, H. J. & Sonntag, H. G. Tissue and serum concentrations of levofloxacin in orthopaedic patients. *Int. J. Antimicrob. Agents* **18**, 335–340 (2001).
26. Flynn, R. A. et al. Transcriptome-wide interrogation of RNA secondary structure in living cells with icSHAPE. *Nat. Protoc.* **11**, 273–290 (2016).
27. Herrmann, C. J. et al. PolyASite 2.0: a consolidated atlas of polyadenylation sites from 3' end sequencing. *Nucleic Acids Res.* **48**, D174–D179 (2020).
28. Balaratnam, S. et al. A chemical probe based on the PreQ(1) metabolite enables transcriptome-wide mapping of binding sites. *Nat. Commun.* **12**, 5856 (2021).
29. Wang, J., Schultz, P. G. & Johnson, K. A. Mechanistic studies of a small-molecule modulator of SMN2 splicing. *Proc. Natl Acad. Sci. USA* **115**, E4604–E4612 (2018).
30. Shahsavarinia, K. et al. An umbrella review of clinical efficacy and adverse cardiac events associated with hydroxychloroquine or chloroquine with or without azithromycin in patients with COVID-19. *Anesth Pain Med* **11**, e115827 (2021).

31. Begley, C. G. et al. Drug repurposing: misconceptions, challenges, and opportunities for academic researchers. *Sci. Transl. Med.* **13**, eabd5524 (2021).

32. Kolli, A. R., Calvino-Martin, F. & Hoeng, J. Translational modeling of chloroquine and hydroxychloroquine dosimetry in human airways for treating viral respiratory infections. *Pharm. Res.* **39**, 57–73 (2022).

33. Force, T. & Kolaja, K. L. Cardiotoxicity of kinase inhibitors: the prediction and translation of preclinical models to clinical outcomes. *Nat. Rev. Drug Discov.* **10**, 111–126 (2011).

34. Padroni, G., Patwardhan, N. N., Schapira, M. & Hargrove, A. E. Systematic analysis of the interactions driving small molecule-RNA recognition. *RSC Med. Chem.* **11**, 802–813 (2020).

35. Guo, P. et al. Compound shape effects in minor groove binding affinity and specificity for mixed sequence DNA. *J. Am. Chem. Soc.* **140**, 14761–14769 (2018).

36. Yazdani, K. et al. Machine learning informs RNA-binding chemical space. *Angew. Chem. Int. Ed.* **62**, e202211358 (2023).

37. Cai, Z., Zafferani, M., Akande, O. M. & Hargrove, A. E. Quantitative structure–activity relationship (QSAR) study predicts small-molecule binding to RNA structure. *J. Med. Chem.* **65**, 7262–7277 (2022).

38. Van Nostrand, E. L. et al. A large-scale binding and functional map of human RNA-binding proteins. *Nature* **583**, 711–719 (2020).

39. Yang, Y. C. et al. CLIPdb: a CLIP-seq database for protein–RNA interactions. *BMC Genomics* **16**, 51 (2015).

40. Khoshnevis, S., Dreggors-Walker, R. E., Marchand, V., Motorin, Y. & Ghalei, H. Ribosomal RNA 2'-O-methylations regulate translation by impacting ribosome dynamics. *Proc. Natl Acad. Sci. USA* **119**, e2117334119 (2022).

41. Jansson, M. D. et al. Regulation of translation by site-specific ribosomal RNA methylation. *Nat. Struct. Mol. Biol.* **28**, 889–899 (2021).

42. Sun, L. et al. RNA structure maps across mammalian cellular compartments. *Nat. Struct. Mol. Biol.* **26**, 322–330 (2019).

43. Mustoe, A. M. et al. Pervasive regulatory functions of mRNA structure revealed by high-resolution SHAPE probing. *Cell* **173**, 181–195 (2018).

44. Mortimer, S. A., Kidwell, M. A. & Doudna, J. A. Insights into RNA structure and function from genome-wide studies. *Nat. Rev. Genet.* **15**, 469–479 (2014).

45. Yeung, P. Y. et al. Systematic evaluation and optimization of the experimental steps in RNA G-quadruplex structure sequencing. *Sci Rep.* **9**, 8091 (2019).

46. Yu, H., Qi, Y., Yang, B., Yang, X. & Ding, Y. G4Atlas: a comprehensive transcriptome-wide G-quadruplex database. *Nucleic Acids Res.* **51**, D126–D134 (2023).

47. Xu, H. & Hurley, L. H. A first-in-class clinical G-quadruplex-targeting drug. The bench-to-bedside translation of the fluoroquinolone QQ58 to CX-5461 (Pidnarulex). *Bioorg. Med. Chem. Lett.* **77**, 129016 (2022).

48. Tang, C. F. & Shafer, R. H. Engineering the quadruplex fold: nucleoside conformation determines both folding topology and molecularity in guanine quadruplexes. *J. Am. Chem. Soc.* **128**, 5966–5973 (2006).

49. Balaratnam, S. et al. Investigating the NRAS 5' UTR as a target for small molecules. *Cell Chem. Biol.* **30**, 643–657 (2023).

50. Fay, M. M., Lyons, S. M. & Ivanov, P. RNA G-quadruplexes in biology: principles and molecular mechanisms. *J. Mol. Biol.* **429**, 2127–2147 (2017).

51. Wacker, S. A., Houghtaling, B. R., Elemento, O. & Kapoor, T. M. Using transcriptome sequencing to identify mechanisms of drug action and resistance. *Nat. Chem. Biol.* **8**, 235–237 (2012).

52. Kumari, S., Bugaut, A., Huppert, J. L. & Balasubramanian, S. An RNA G-quadruplex in the 5' UTR of the NRAS proto-oncogene modulates translation. *Nat. Chem. Biol.* **3**, 218–221 (2007).

53. Pottel, J. et al. The activities of drug inactive ingredients on biological targets. *Science* **369**, 403–413 (2020).

54. Mullard, A. Parsing clinical success rates. *Nat. Rev. Drug Discov.* **15**, 447 (2016).

55. Morgan, B. S., Forte, J. E., Culver, R. N., Zhang, Y. & Hargrove, A. E. Discovery of key physicochemical, structural, and spatial properties of RNA-targeted bioactive ligands. *Angew. Chem. Int. Ed.* **56**, 13498–13502 (2017).

56. Vicens, Q. & Westhof, E. RNA as a drug target: the case of aminoglycosides. *ChemBioChem* **4**, 1018–1023 (2003).

57. Van Norman, G. A. Limitations of animal studies for predicting toxicity in clinical trials: is it time to rethink our current approach? *JACC Basic Transl. Sci.* **4**, 845–854 (2019).

58. Corsello, S. M. et al. The drug repurposing hub: a next-generation drug library and information resource. *Nat. Med.* **23**, 405–408 (2017).

59. Backman, T. W., Cao, Y. & Girke, T. ChemMine tools: an online service for analyzing and clustering small molecules. *Nucleic Acids Res.* **39**, W486–W491 (2011).

Methods

Materials

DNA and RNA were purchased from IDT or Stanford PAN facility unless otherwise stated (see Supplementary Table 4 for details). All chemicals purchased from commercial suppliers were used without further purification (see Supplementary Table 5 for details). All enzymes, kits, bioreagents and software were obtained from sources listed in Supplementary Table 5. Human cell line HEK293 (CRL-1573) was purchased from the American Type Culture Collection (ATCC) and maintained in DMEM supplemented with 10% fetal bovine serum at 37 °C in a humidified incubator containing 5% CO_2.

Synthesis of acylimidazole probes

HCQ-AI, Lev-AI and Das-AI were synthesized as described in the synthesis section of Supplementary Information file and activated immediately before use. In general, the carboxylic acid precursors (1.0 equiv.) were dissolved in dry DMSO. To it was added a suspension containing 1.4 equiv. of 1,1'-carbonyldiimidazole at room temperature. The solution was stirred at room temperature overnight under argon. After the reaction, the final solution can be used as a 50 mM acylimidazole stock without further purification. VNMRJ was used to acquire ^{1}H and ^{13}C NMR.

Curation of small-molecule drugs database

Lists of FDA-approved small-molecule drugs were acquired from a previous study[5]. Briefly, FDA-approved drugs were downloaded from DrugBank v.5.1.6 (released on 22 April 2020) on 15 June 2020. The drug list was further filtered to remove molecules according to the reported criteria[5]. Molecules with molecular weights larger than 140 Da and smaller than 706 Da were used for statistical and principal component analysis (PCA).

Drugs that are in preclinical studies, clinical trials and drugs withdrawn from the market were acquired from Drug Repurposing Hub[58] on 5 June 2022. The molecules were annotated based on their disease areas and stages in clinical studies. Molecules with molecular weights larger than 140 Da and smaller than 706 Da were used for statistical analysis.

The small-molecule (SM) RNA binders were downloaded from R-bind 2.0 (ref. 5) on 24 May 2022. Molecules (131) with known PubChem Compound ID number (CID) were further used for structural similarity analysis using the Tanimoto coefficient (Tc).

Quantification of structural similarity with the Tanimoto coefficient

Substructure key-based 2D Tanimoto coefficients were calculated using the PubChem Score Matrix Service. Lists of CIDs for 131 known RNA binders and molecules of interest were used as input to calculate the pairwise Tanimoto coefficients with each of 131 known RNA binders. Tanimoto coefficients of all drug–RNA binder pairs are listed as the matrix in Supplementary Tables.

Chemoinformatic calculation of chemical fingerprints

Fourteen Open Babel descriptors of the selected FDA-approved drugs were calculated using ChemMine Tools[59]. See Supplementary Tables for the detailed list.

PCA of chemical fingerprints of small-molecule drugs

Fourteen Open Babel descriptors were used as chemical fingerprints in the standardized PCA. PCA was performed using Graphpad Prism 9. The principal components (PC1 and PC2) were selected based on parallel analysis, with eigenvalues greater than the 95% percentile of Monte Carlo simulations (number of simulations, 1,000) on 'random' data of equal dimension to the input data. See Supplementary Tables 1 and 2 for details.

Treatment of cells with acylimidazole probes, isolation of total cellular RNA and enrichment of poly(A)+ transcripts

First, ~2×10^6 cells were pretreated with DMSO or excess competing drugs for 30 min at 37 °C. The cells were then immediately scraped into a 15 ml Falcon tube and resuspended in 1 ml of 1 × PBS, pH 7.4 at 37 °C containing 50 µM of acylimidazole probe, in the absence or presence of excess competing drugs. After incubation at 37 °C for 30 min, the reaction was stopped by centrifugation. To isolate total cellular RNA, cells were lysed with 6 ml of Trizol LS reagent. RNAs were then purified with a ZymoQuick-RNA Midiprep kit according to the manufacturer's protocol. To enrich poly(A)+ transcripts, a 250 µg aliquot of total cellular RNA was diluted to 600 µg ml^{-1} with RNA storage buffer and then purified with Poly(A)Purist MAG kit according to the manufacturer's protocol. The captured RNAs were further purified with RNeasy Mini columns according to the manufacturer's protocol, and stored as 500 ng aliquots at −80 °C.

Preparation of sequencing libraries

Sequencing libraries were prepared according to the following protocol based on previous reports[23,26] with modifications. First, the probe-labelled RNAs were modified with biotin by 'click' reaction. Then, 500 ng of poly(A)+ RNAs was incubated with 2 µl of 1.85 mM dibenzocyclooctyne-biotin and 1 µl of RiboLock at 37 °C for 2 h. After the reaction, modified RNAs were purified with a Zymo RNA Clean & Concentrator Column-5 according to the manufacturer's protocol.

RNA fragmentation, 3'-end repair and 3'-end ligation. RNAs were fragmented using RNA fragmentation reagent, which rendered RNA fragments with a medium length of ~100 nt. After purification with a Zymo RNA Clean & Concentrator Column-5, the 3' end of RNA fragments were repaired by FastAP and T4 PNK for 1 h at 37 °C according to the previously reported protocol[23,26]. Next, 3' ligation was performed to install an 3' blocked RNA ligand to the 3' end of RNA fragments. Briefly, RNAs were incubated with T4 RNA ligase 1 and preadenylylated RNA links (probe-treated sample: 3'-ddC blocked RNA linker /5rApp/ AGA TCG GAA GAG CGG TTC AG/3ddC/ was used; for DMSO-treated samples, 3'-biotin blocked RNA linker /5rApp/AGA TCG GAA GAG CGG TTC AG/3Biotin/ was used) for 4 h at 25 °C. After the 3'-end ligation reaction, RNA was purified with a Zymo RNA Clean & Concentrator Column-5 and then resolved on a 6% (w/v) UreaGel denaturing PAGE. RNAs with a length above ~50 nt were extracted with RNase-free water and further purified using a 10K Amicon filter according to the manufacturer's protocol.

cDNA synthesis, enrichment of biotin-modified RNAs and cDNA purification. The following steps were performed according to the previously reported protocol with slight modifications[23,26]. The 3'-ligated RNA was dissolved in 11 µl of RNase-free water and 1 µl of 1 µM barcoded RT primer (/5phos/DDD NN<u>A ACC</u> NNN NAG ATC GGA AGA GCG TCG TGG A/iSp18/GGATCC/iSp18/TACTGAACCGC, where D = A/G/T and N = A/T/G/C; the underlined four nucleotides are barcodes) (see Supplementary Table 4 for assignments of RT primers). The reaction was heated at 70 °C for 5 min and then cooled to 25 °C at 20 °C per 60 s. Next, 8 µl of RT reaction mixture (4 µl of 5× first-strand buffer, 0.75 µl of RiboLock, 1 µl of 100 mM dithiothreitol (DTT), 1 µl of 10 mM deoxyribonucleoside triphosphate (dNTP) and 1.25 µl of SuperScript III) was added to the above reaction. The resulting solution was heated at 25 °C for 3 min, 42 °C for 5 min and 52 °C for 30 min. Next, the cDNA–RNA heteroduplexes were enriched with MyOne C1 magnetic streptavidin beads and the cDNA was purified with a 6% (w/v) denaturing PAGE gel.

cDNA circularization and library amplification. The purified cDNA was circularized with CircLigase II for 3 h at 60 °C and purified with Zymo DNA Clean & Concentrator Column-5 according to the manufacturers' protocols. To amplify the library, 20 µl of 2× Phusion HF PCR master mix, 0.4 µl of SYBR Green I (25×), 1.0 µl of 10 µM P5-universal PCR primer (5'-AAT GAT ACG GCG ACC ACC GAG ATC TAC ACT CTT TCC CTA CAC GAC GCT CTT CCG ATC T-3') and 10 µM P3-universal PCR primer (5'-CAA GCA GAA GAC GGC ATA CGA GAT CGG TCT CGG

CAT TCC TGC TGA ACC GCT CTT CCG ATC T-3'). PCR amplification was monitored in real time to avoid overamplification. The PCR reaction was first denatured at 98 °C for 45 s, followed by cycles of denaturation (98 °C, 15 s), annealing (65 °C, 20 s) and extension (72 °C, 1 min). The PCR products were purified with a 3% low-melting-point agarose gel to remove primer dimers (~130 nt) and extracted with a QIAGEN MiniElute Gel extraction kit. The purified DNA was analysed by BioAnalyzer for quality control and quantified with Qubit.

Next-generation sequencing
Ten sublibraries were combined as a sequence library. The sequence library was sequenced on an Illumina-based HiSeq X-ten with a 23% spike-in PhiX sequencing control. Paired-end 150 bp sequencing was performed. The raw data were transformed into single-end reads for the following bioinformatics analysis with the icSHAPE pipeline. Note: the DMSO-treated sequence sublibraries are amplified for fewer cycle numbers and thus have higher complexity than the probe-treated sequence sublibraries. Thus, more sequencing depth should be allocated to the probe-treated sequence sublibraries.

Bioinformatics and RBRP data visualization
We use a bioinformatics pipeline modified based on icSHAPE[23,26] to calculate the RBRP score at each nucleotide transcriptome-wide. See the 'Bioinformatics and command lines' section in Supplementary Information for detailed protocols and command lines. Bigwig files were processed with wiggletools and can be visualized using alignment visualization tools such as Integrative Genomics Viewer (IGV) or UCSC Genome Browser.

Identification of off-target RNA binding sites
RBRP scores at each site were calculated as the surplus yield of 2'-OH acylation for experiments performed in the absence versus the presence of excess competing drug. We defined potential RNA-binding sites as nucleotides with RBRP scores ≥0.12 and read depths ≥200 (see Supplementary Tables for all competed acylation sites).

Target validation with RT–qPCR
HEK293 cells were treated with the acylation probe in the absence or presence of excess competing drugs. Total cellular RNA was extracted with the Trizol LS reagent and purified according to the protocol described above. First, 1 µg (2.5 µl, 400 ng µl^{-1}) of total RNA was mixed with 1 µl of 10 mM dNTPs, 1 µl of 2 µM RT primer and 8.5 µl of water. The reaction mixture was heated at 65 °C for 5 min and then cooled on ice for >1 min. Next, 4 µl of 5× first-strand buffer, 1 µl of 0.1 M DTT, 1 µl of RiboLock and 1 µl of SuperScript III were added to the reaction. The reaction solution was then incubated at 25 °C for 5 min and at 55 °C for 30 min. After inactivation at 70 °C for 15 min, 0.5 µl of RNase H was added and incubated at 37 °C for 20 min.

qPCR. First, 1 µl of the above cDNA product was added to a solution containing 10 µl of Luna University qPCR Master Mix, 0.5 µl of 10 µM forward primer, 0.5 µl of 10 µM reverse primer and water added to make 20 µl. The resulting mixture was amplified using a StepOnePlus real-time qPCR system, starting with an initial denaturation step at 95 °C for 60 s, followed by 40 cycles of denaturation (95 °C, 15 s), annealing and extension (60 °C, 30 s). The ΔCt value was calculated from the equation: ΔCt = Ct(U) − Ct(D). qPCR amplified regions upstream (U) and downstream (D) of the acylation sites. An increased ΔCt value indicates acylation at the tested sites.

CD experiments
CD experiments were performed using a Jasco J-810 spectropolarimeter. The RNA solution in 10 mM Tris (pH 7.0) and the desired concentration of KCl was heated at 95 °C for 3 min, then cooled to 25 °C over 1 h. Then, 250 µl of folded RNA was placed in a quartz cuvette with an optical path length of 1 mm. Three CD scans, over the wavelength range of 220–320 nm, were performed at 50 nm min^{-1} with a 1 nm bandwidth. The average CD curve was used for further analysis. In addition, the CD spectrum of the buffer was determined and subtracted from the spectrum obtained for the RNA-containing solution. Melting curve measurement: folded RNA was heated from 20 °C to 95 °C at a 1 °C min^{-1} ramping rate. CD ellipticity at 265 nm was measured with data integration time = 2 s.

Cloning
We PCR-amplified the open reading frame and the 3' UTR sequence of a d2GFP-encoding plasmid according to a previous report[60]. Next, 5 ng of the above PCR product (8.4 fmol) is further PCR-amplified for ten cycles using Q5 High-Fidelity 2× Master Mix (NEB) with sequence-specific forward primers to install the T7 promoter sequence, variants of the YBX1 5' UTR and the poly(T)$_{120}$ tail (Supplementary Tables). Finally, 1 µl of the second PCR product is further amplified with two universal primers (T7-PCR-fwd and Tail-PCR-rev) using Q5 High-Fidelity 2× Master Mix. This PCR product is purified with 1% agarose gel.

Synthesis of reporter mRNA
The dsDNA templates of YBX1-WT, YBX1-mutG2A and YBX1-delG4 were transcribed using HiScribe T7 Quick High Yield RNA Synthesis Kit (NEB) and 5'-capped with a Vaccinia Capping System according to manufacturer's protocol.

In vitro translation
First, 600 ng of mRNA (YBX1-WT, YBX1-mutG2A, YBX1-delG4) was diluted into 1× rG4 folding buffer (10 mM Tris, pH 7.0, 0.1 mM KCl) to 17.82 µl. mRNAs were folded by heating at 90 °C for 3 min and then cooling down to 25 °C at 0.1 °C s^{-1}. Next, 0.18 µl of 100× Lev solution at the desired concentration or DMSO was added and incubated at 25 °C for 30 min. Then, 6 µl of the above mRNA solution was added to a cell-free translation system containing 17.5 µl of rabbit reticulocyte lysate, 0.5 µl of RNasin Ribonuclease Inhibitor, 1 µl of 1 mM complete amino acid mixtures and 0.19 µl of the corresponding 100× Lev solution or DMSO. The resulting mixture was incubated at 30 °C for 90 min. After incubation, the green fluorescence signal was determined by a BioTek Synergy HT microplate reader (excitation/emission, 485 nm/520 nm).

SHAPE experiment
First, 1 µg of YBX1-WT mRNA was added to 1× rG4 folding buffer (10 mM Tris, pH 7.0, 0.1 mM KCl) to 11.88 µl. Next, 0.12 µl of 100× Lev solution or DMSO was added. After incubation at 25 °C for 30 min, 12 µl of the above mRNA solution was added to a cell-free system containing 35 µl of rabbit reticulocyte lysate, 1 µl of RNasin Ribonuclease Inhibitor, 1 µl of 1 mM complete amino acid mixtures, and 0.38 µl of 100× Lev solution (1–5 mM) or DMSO. After incubation at 30 °C for 30 min, 5.5 µl of 2.0 M NAI-N$_3$ reagent was added and incubated at 25 °C for 15 min. The reaction mixture was purified by Zymo RNA clean-up and Concentrator-5 column following manufacturer's protocol. The purified mRNA was mixed with 0.2 µl of 10 µM FAM-labelled RT primer, 0.5 µl of 10 mM dNTP and water to 6.5 µl. The mixture was heated at 75 °C for 5 min and chilled on ice for 2 min. Next, reverse transcription was performed by adding 2 µl of 5× first-strand buffer, 1 µl of 0.1 M DTT, 0.25 µL of 40 U µl^{-1} RiboLock RNase inhibitor and 0.25 µl of 200 U µl^{-1} SuperScript III. The reaction was heated at 25 °C for 10 min, at 50 °C for 50 min and at 55 °C for 50 min. After reaction, 1 µl of 1 N NaOH was added and incubated at 95 °C for 3 min, followed by the addition of 12 µl of 8 M urea loading buffer and heating at 95 °C for an additional 3 min. The resulting samples were resolved on an ABI 3130xl Genetic Analyzer. The data were analysed with Peak Scanner Software 2.

Quantification of the 2'-O-methylation level of human rRNA
To quantify 2'-O-methylation of rRNA, we performed reverse transcription at low dNTP concentrations followed by PCR (RTL-P)[61]. HEK293

cells (American Type Culture Collection, #CRL-1573) were seeded on a six-well plate in 2 ml DMEM containing 10% FBS. Then, 20 μl of DMSO or concentrated DMSO solution of drug was added to the cells to reach the desired drug concentrations. The cells were incubated at 37 °C in a humidified incubator containing 5% CO_2 for 2 days. After 2 days, cells were washed with 2 ml of DPBS and then lysed with 600 μl of Trizol LS reagent, followed by RNA purification with Quick RNA-miniprep kits and on-column digestion of genomic DNA.

For RTL-P, reverse transcription was performed with 1 μg extracted RNA (5 μl), 1 μl of dNTP at low concentration (10 μM) and 1 μl of 2 μM RT primer. The solution was heated at 65 °C for 5 min and then incubated on ice for >1 min. Next, 4 μl of 5× first-strand buffer, 1 μl of 0.1 M DTT, 1 μl of RiboLock and 1 μl of SuperScript III were added to the reaction. The reaction mixture was then incubated at 55 °C for 50 min, followed by incubation at 70 °C for 15 min. Then, 1 μl of the above cDNA product was added to a solution containing 10 μl of Luna University qPCR Master Mix, 0.5 μl of 10 μM forward primer, 0.5 μl of 10 μM reverse primer and water added to make 20 μl. The resulting mixture was amplified using a StepOnePlus real-time qPCR system, starting with an initial denaturation step at 95 °C for 60 s, followed by 40 cycles of denaturation (95 °C, 15 s), annealing and extension (60 °C, 30 s). The ΔCt value was calculated from the equation: $\Delta Ct = Ct(U) - Ct(D)$. qPCR amplified regions upstream (U) and downstream (D) of the 2′-OMe sites. An increased ΔCt value indicates acylation at the tested sites. An increased ΔCt value indicates enhanced 2′-O-methylation at the targeted sites.

Statistical analysis

For all statistical tests (unless otherwise noted), a one-tailed or two-tailed Student's t-test was used to compare means between two samples. Significance is denoted as follows: $*P < 0.05$, $**P < 0.01$, $***P < 0.001$, $****P < 0.0001$. Student's t-tests were performed in Graph-Pad Prism 9 software. Quantifications shown are mean ± s.e.m. unless otherwise stated.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

All sequencing data are available through the Gene Expression Omnibus (GEO) under accession GSE229331. Data supporting the findings of this study are available in the Article, and its Supplementary Information and Supplementary Tables. Source data and bedgraphs of RBRP scores are also freely available at figshare at https://doi.org/10.6084/m9.figshare.20326824. Source data are provided with this paper.

Code availability

The RBRP scripts used for bioinformatics analysis are freely available at https://github.com/linglanfang/RBRP.

References

60. Warren, L. et al. Highly efficient reprogramming to pluripotency and directed differentiation of human cells with synthetic modified mRNA. *Cell Stem Cell* **7**, 618–630 (2010).
61. Dong, Z. W. et al. RTL-P: a sensitive approach for detecting sites of 2′-O-methylation in RNA molecules. *Nucleic Acids Res.* **40**, e157 (2012).

Author contributions

E.T.K. conceived the project and supervised the work. E.T.K. and L.F. designed experiments. L.F. performed experiments and data analysis. L.F. developed codes and the bioinformatics pipeline. Y.L. synthesized the acylating probe of dasatinib. L.X. performed the in vitro translation and SHAPE experiments. W.A.V., M.G.M. and A.M.K. conducted the groundwork. L.F. and E.T.K. wrote the manuscript, with input from all authors.

Competing interests

The authors declare no competing interests.

Additional information

Extended data is available for this paper at https://doi.org/10.1038/s41557-023-01309-8.

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41557-023-01309-8.

**Correspondence and requests for materials** should be addressed to Eric T. Kool.

Peer review information *Nature Chemistry* thanks the anonymous reviewers for their contribution to the peer review of this work.

**Reprints and permissions information** is available at www.nature.com/reprints.
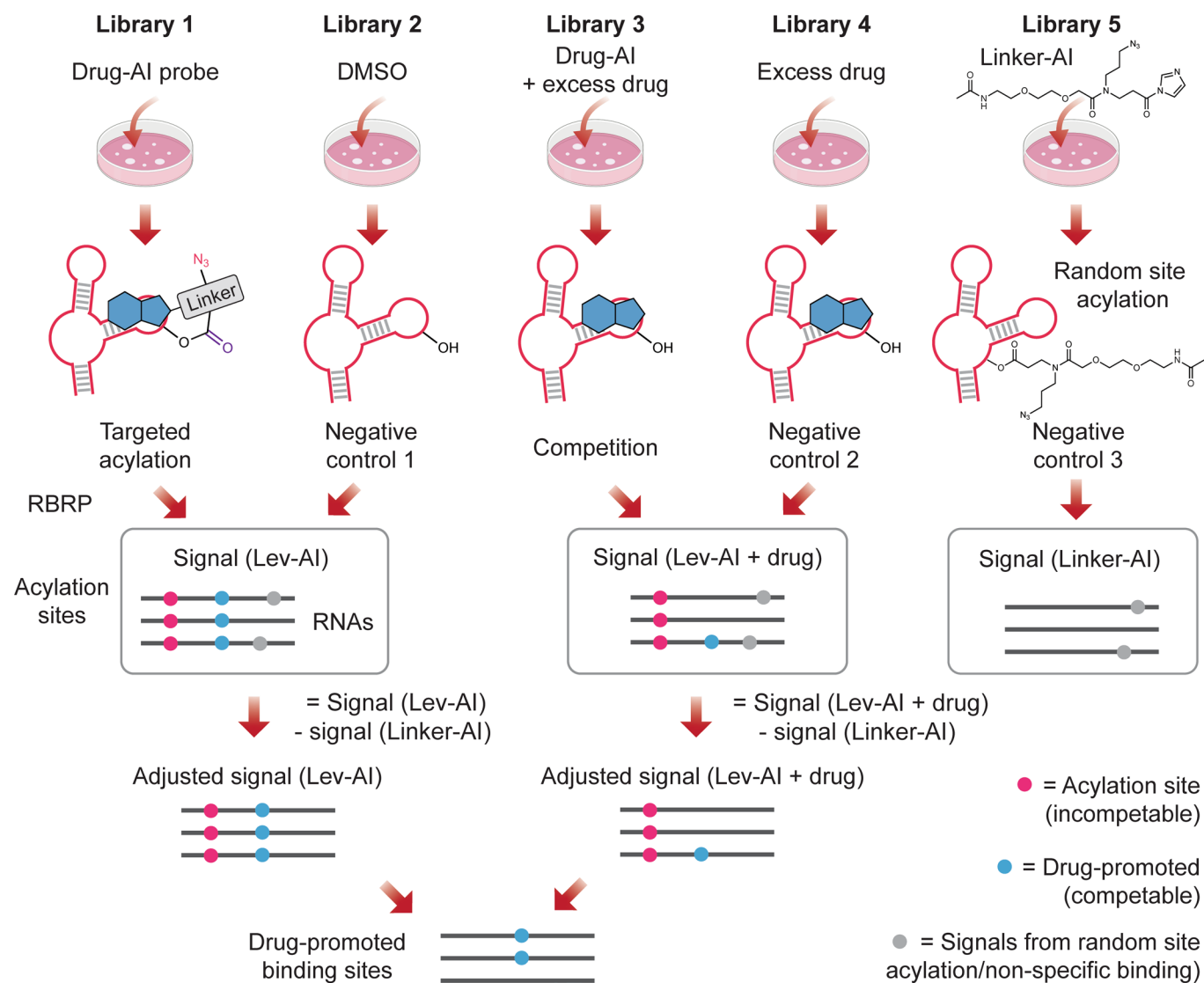
Extended Data Fig. 1 | RBRP workflow and experimental setup. Library 1: Cells were treated with the drug-conjugated acylating probe. RBRP workflow identifies RBRP signals from drug-promoted acylation, random site acylation, and non-specific binding events during biotin-mediated pulldown. **Library 2:** Cells were treated with DMSO. RBRP workflow identifies RBRP signals from non-specific binding events. **Library 3:** Cells were treated with drug-conjugated acylating probe and the unmodified drug. RBRP workflow identifies RBRP signals from random site acylation, non-specific binding events during biotin-mediated pulldown, and changes in transcript abundance. **Library 4:** Cells were treated with excess unmodified drugs. RBRP workflow identifies RBRP signals from non-specific binding events during biotin-mediated pulldown and changes in transcript abundance. **Library 5:** Cells were treated with Linker-AI. RBRP workflow identifies RBRP signals from random site acylation and non-specific binding events during biotin-mediated pulldown.
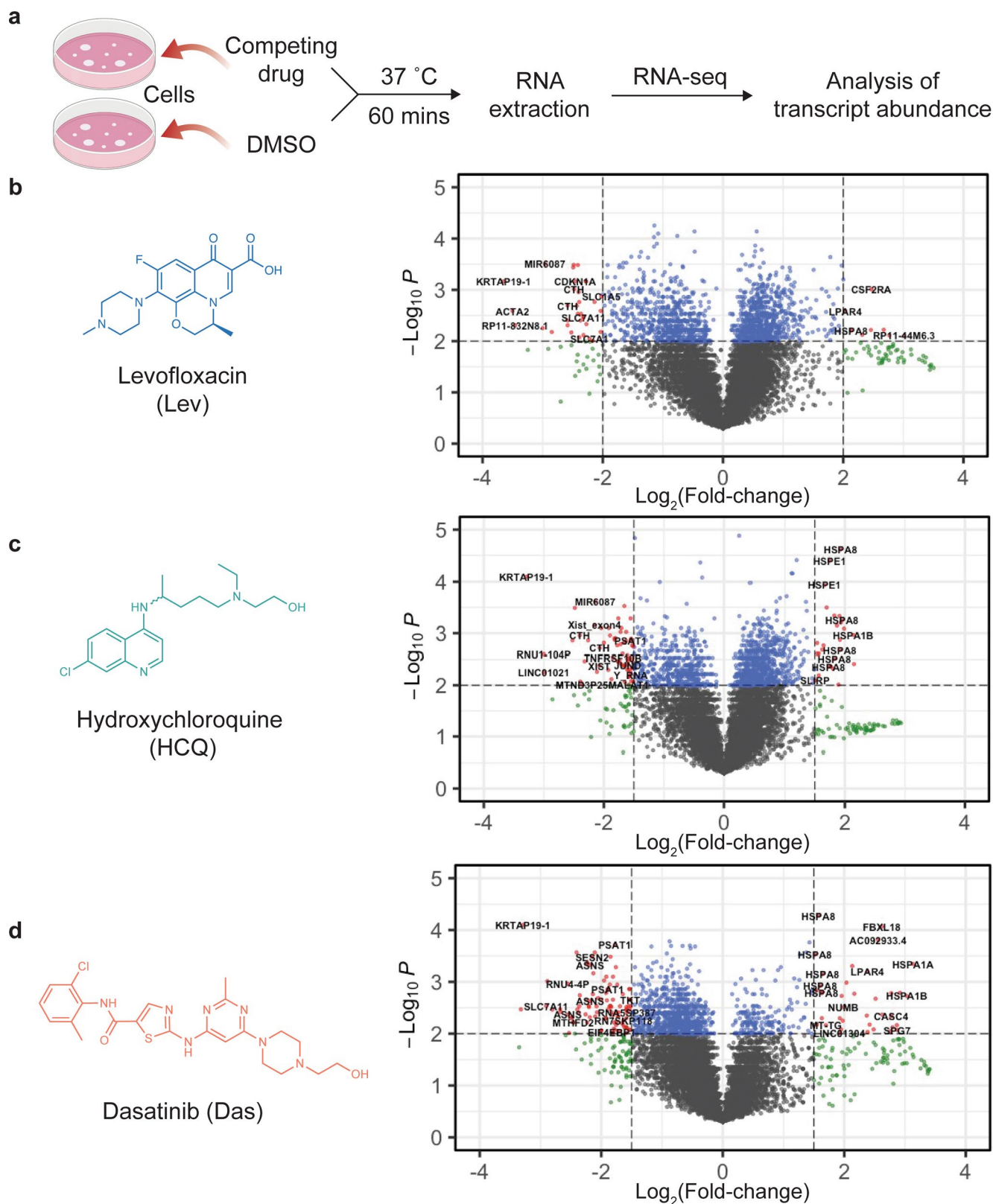
Extended Data Fig. 2 | RNA-seq determines how excess unmodified drug influences transcripts abundance in HEK293 cells. a, Workflow of RNA-seq with HEK293 cells treated with unmodified drugs or DMSO. **b-d**, Volcano plot showing effects of unmodified Lev (**b**), HCQ (**c**), and Das (**d**) on transcript abundance. X-axis: log-transformed ratio of transcript abundance in the presence over the absence of unmodified drug. Y-axis: the negative value of log-transformed P-value calculated with DESeq2 using a two-tailed Wald test. Fold-change=transcript abudance (drug-treated)/transcript abudance (DMSO).
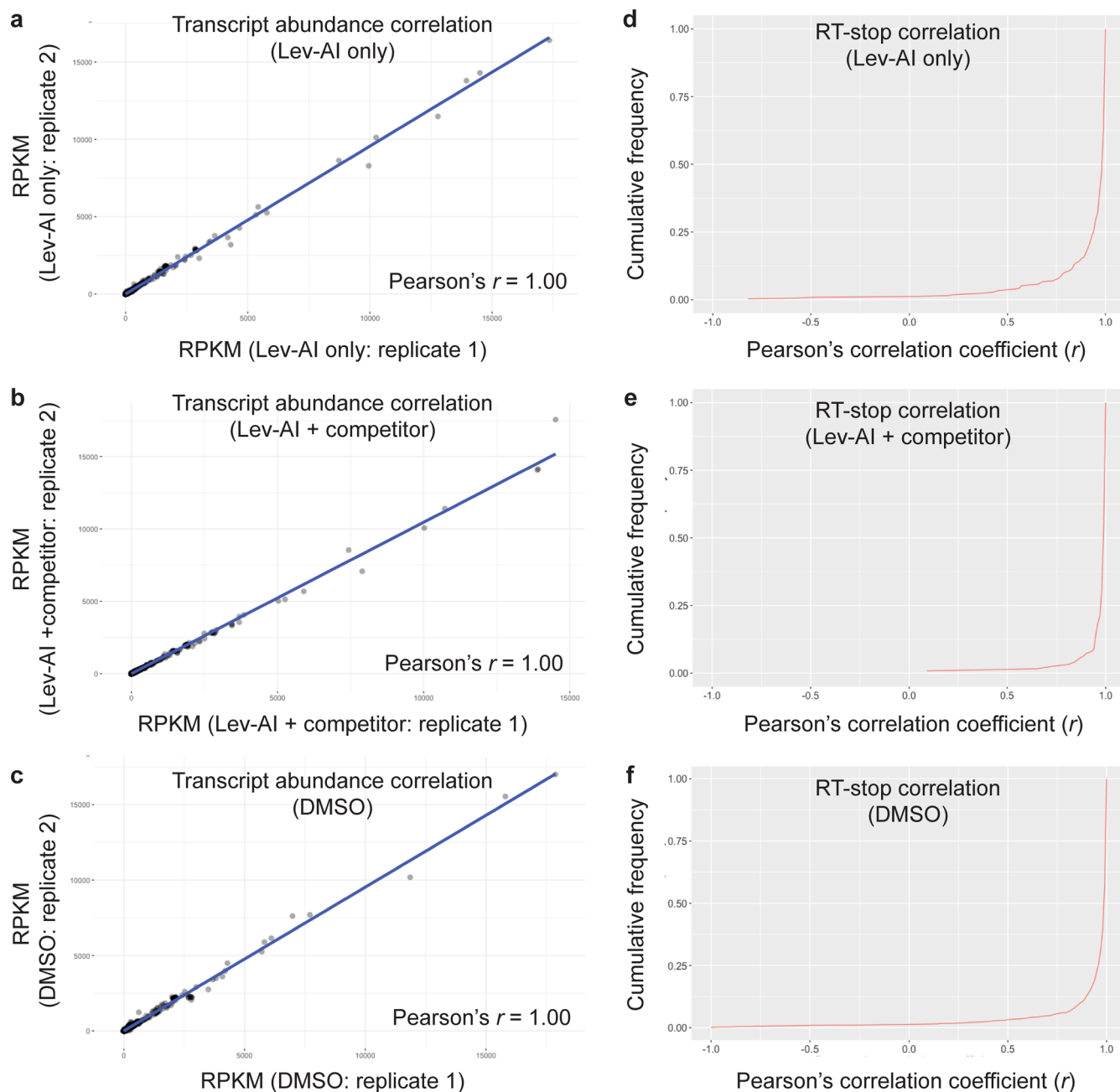
Extended Data Fig. 3 | Transcript abundance and RT-stop frequencies are strongly concordant between RBRP sequencing libraries from two biological replicates. a-c, Scatter plot showing very strong correlation of transcript expression value (RPKM) between two biological replicates (Pearson correlation r = 1.00) in HEK293 cells treated with Lev-AI only (**a**), Lev-AI and excess unmodified Lev (**b**), and DMSO (**c**). **d-f**, The concordance of RT-stop frequencies is high for most transcripts of read depth higher than the optimized cut-off value (200) in sequencing libraries of Lev-AI only (**d**), Lev-AI and excess unmodified Lev (**e**), and DMSO (**f**).
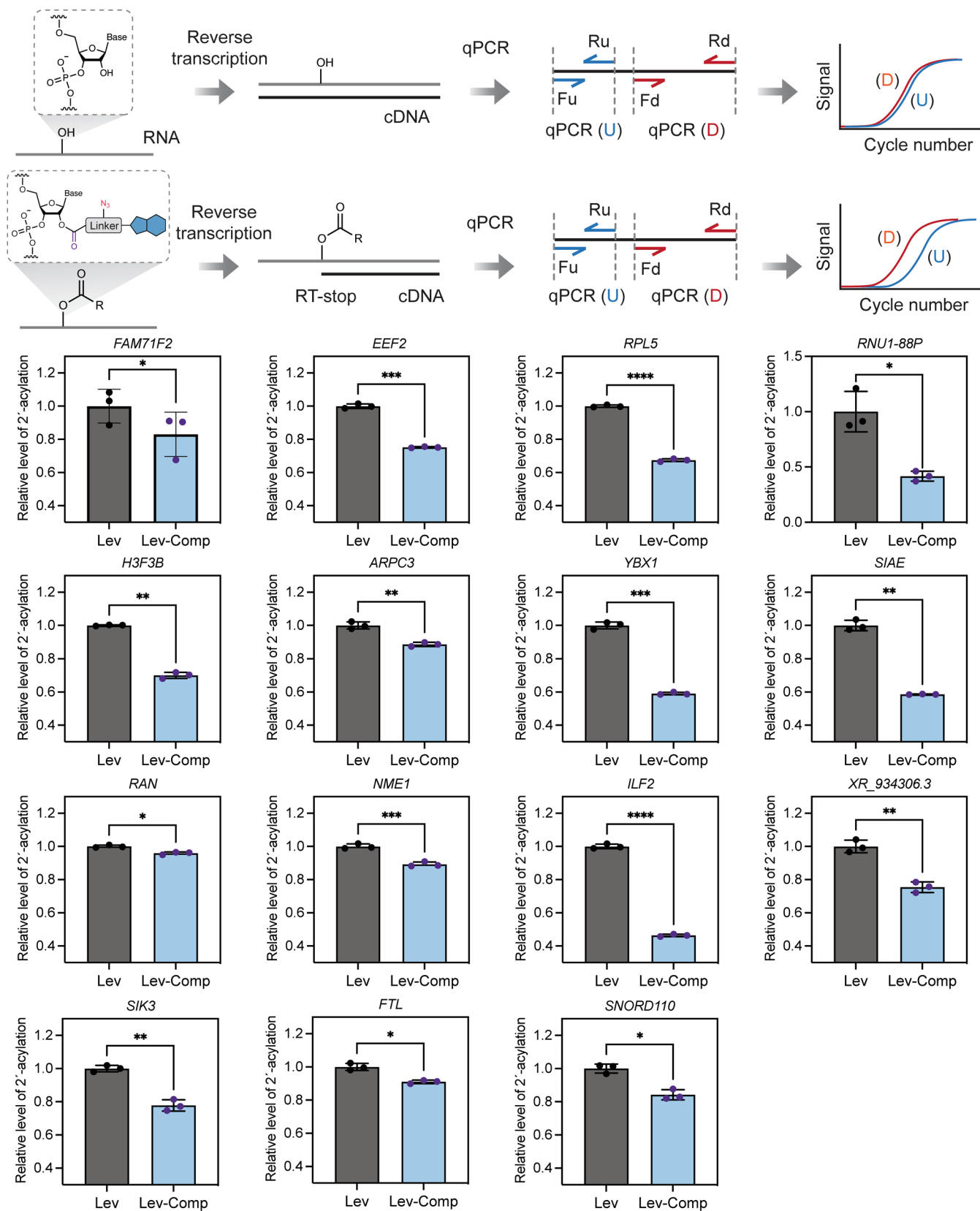
Extended Data Fig. 4 | qPCR independently validates the transcriptome interactions of Levofloxacin (Lev) at 15 transcriptome binding sites in HEK293 cells. Workflow showing the strategies of validating competable 2′-OH acylation sites with qPCR (Top panel). qPCR validated and quantified the relative level of 2′-OH acylation at the drug-binding loci. Data represent mean ± s.e.m., n = 3 biologically independent experiments. Statistical significance was calculated with two-tailed unpaired Student's t-tests: *P < 0.05, **P < 0.01, ***P < 0.001, ****P < 0.0001. P values are 0.0227 (*FAM71F2*), 0.0007 (*EEF2*), <0.0001 (*RPL5*), 0.0346 (*RNU1-88P*), 0.0015 (*H3F3B*), 0.0023 (*ARPC3*), 0.0009 (*YBX1*), 0.0019 (*SIAE*), 0.0478 (*RAN*), 0.0006 (*NME1*), <0.0001 (*ILF2*), 0.0055 (*XR_934306.3*), 0.0051 (*SIK3*), 0.0140 (*FTL*), and 0.0103 (*SNORD110*).
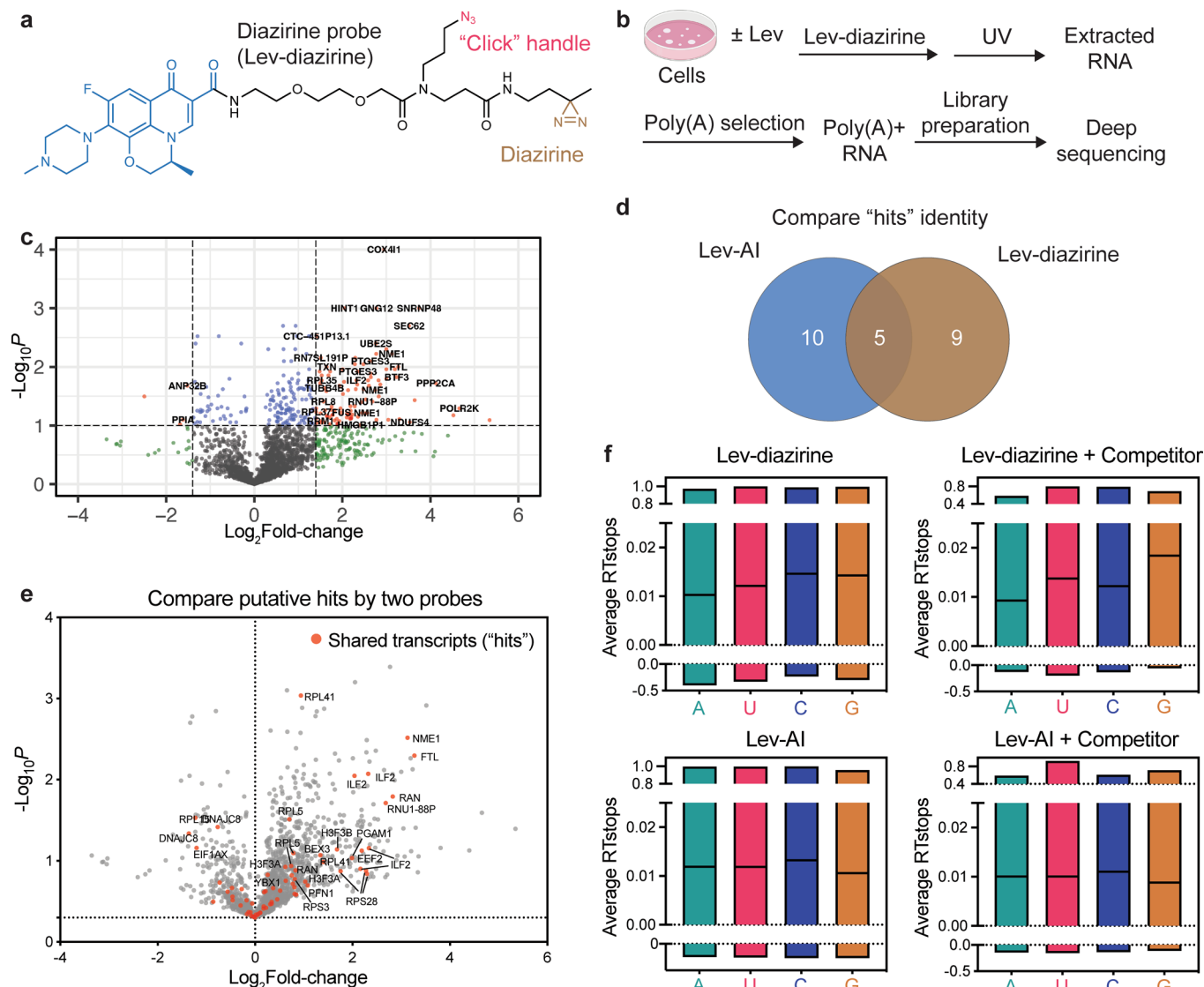
Extended Data Fig. 5 | RBRP complements the existing diazirine-based profiling method. a, Chemical structure of diazirine-conjugated analogue of Levofloxacin (Lev-diazirine). The drug moiety is coloured blue, "click" handle" is coloured red, and diazirine moiety is coloured brown. **b**, Workflow for profiling RNA targets of Lev with Lev-diazirine probe. **c**, Volcano plot showing transcripts that are confidently enriched by Lev-diazirine in HEK293 cells. **d**, Volcano plot showing shared RNA targets that are identified by both Lev-AI and Lev-diazirine. X-axis: log-transformed ratio of transcript abundance in the presence over the absence of unmodified drug. Y-axis: the negative value of log-transformed P-value calculated with DESeq2. **d**, Venn diagram comparing RNA targets that were identified by Lev-AI and Lev-diazirine. **e**. Volcano plot showing shared transcripts ("hits") identified by both RBRP and the existing diazirine-based profiling method. Transcripts that were enriched by Lev-diazirine are coloured grey. Shared transcripts that were identified by both Lev-AI and Lev-diazirine are

coloured red. For **c** and **e**, X-axis: log-transformed ratio of transcript abundance in the presence over the absence of unmodified drug. Y-axis: the negative value of log-transformed P-value calculated with DESeq2 using a two-tailed Wald test. Fold-change=transcript abudance (Lev-diazirine-treated)/transcript abundance (Lev-diazirine + Competitor). **f**, Floating bar graphs comparing the distribution of RT stops of Lev-AI and Lev-diazirine towards four nucleotides. The floating bars represent the range of all data points (minimum to maximum). Lines represent the mean values. Data points are collected from two biologically independent sequencing experiments. All data points for nucleotide in each experiment are compiled for data analysis. Lev-diazirine: $n = 22,958$ (A), 19,154 (U), 19,375 (C), and 18,992 (G). Lev-diazirine + Competitor: $n = 2,420$ (A), 1,778 (U), 2,160 (C), and 1,895 (G). Lev-AI: $n = 425,366$ (A), 389,020 (U), 352,821 (C), and 380,397 (G). Lev-AI + Competitor: $n = 423,930$ (A), 389,653 (U), 354,014 (C), and 377,073 (G).
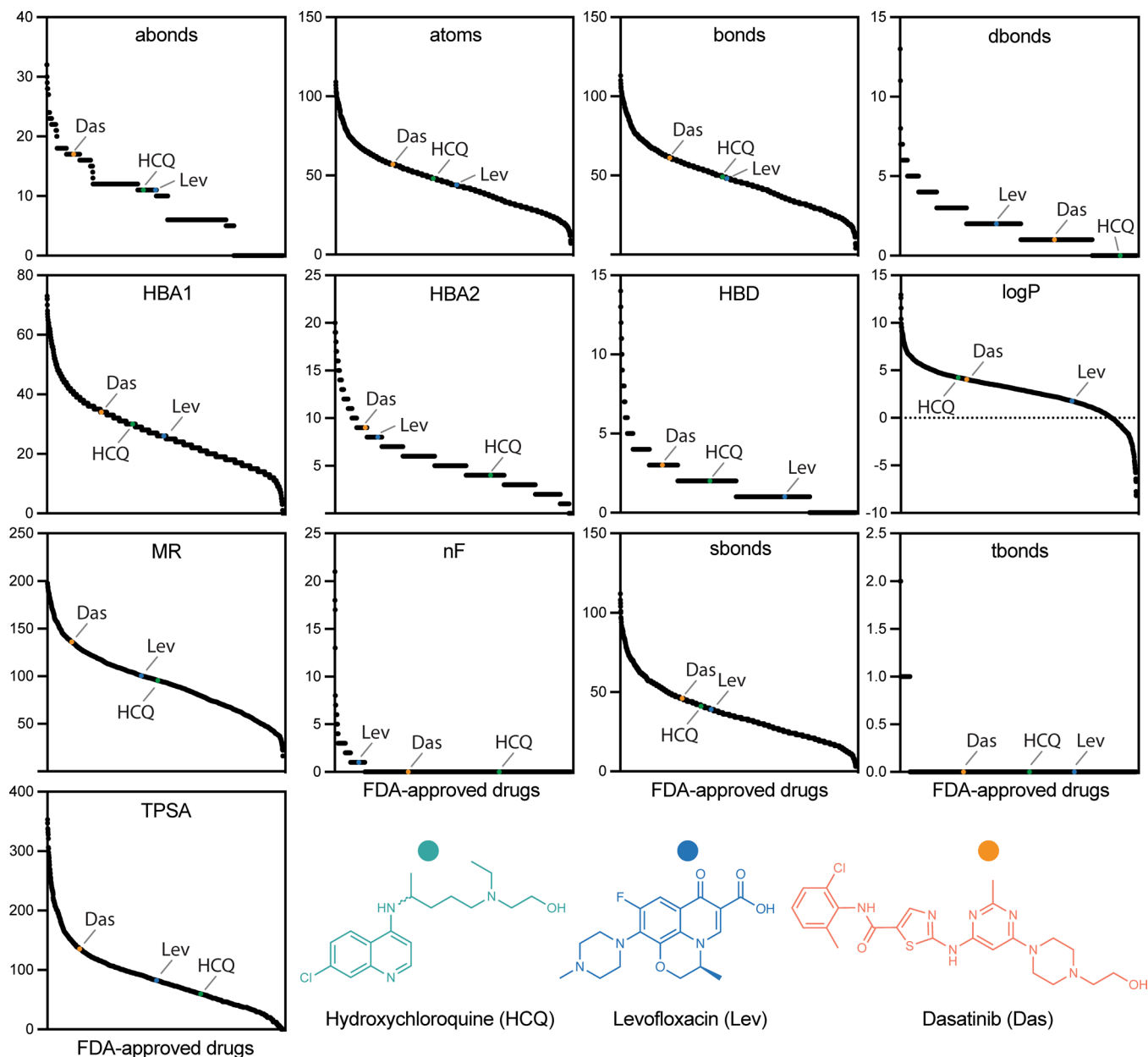
Extended Data Fig. 6 | Plots comparing the structural fingerprints (Open Babel descriptors) of HCQ, Lev, and Das to an acquired list of FDA-approved small-molecule drugs (2076 drugs). Open Babel descriptors are abonds (Number of aromatic bonds), atoms (Number of atoms), bonds (Number of bonds), dbonds (Number of double bonds), HBA1 (Number of Hydrogen Bond Acceptors 1), HBA2 (Number of Hydrogen Bond Acceptors 2), HBD (Number of Hydrogen Bond Donors), logP (Octanol/water partition coefficient), MR (Molar refractivity), MW (Molecular Weight), nF (Number of Fluorine Atoms), sbonds (Number of single bonds), tbonds (Number of triple bonds), and TPSA (Topological polar surface area).
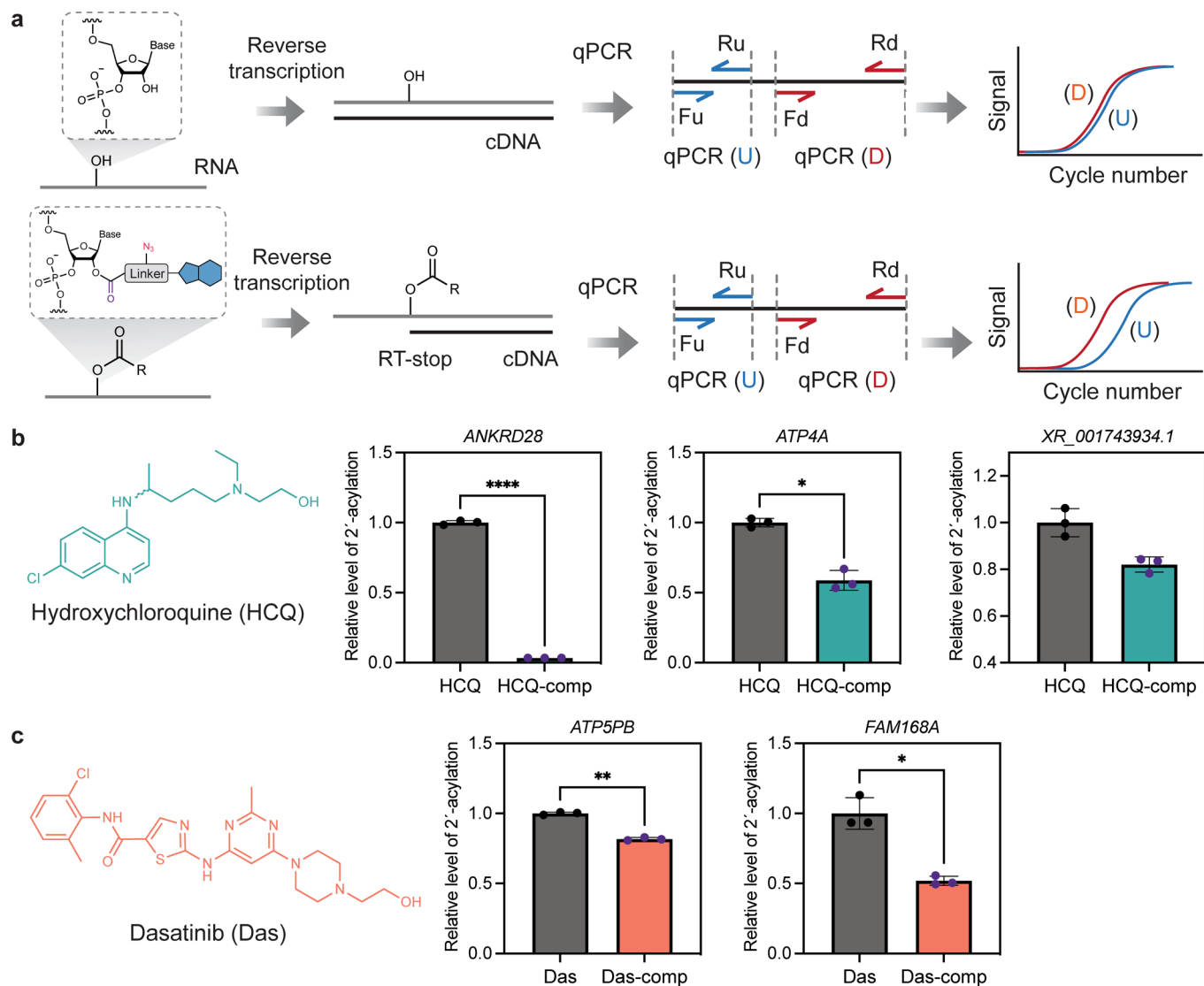
Extended Data Fig. 7 | qPCR independently validates the transcriptome interactions of Hydroxychloroquine (HCQ) and Dasatinib (Das) at several transcriptome binding sites in HEK293 cells. a, Workflow showing the strategies of validating competable 2´-OH acylation sites with qPCR (Top panel). **b-c**, qPCR validated and quantified the relative level of 2´-OH acylation at the drug-binding loci of HCQ (**b**) and Das (**c**). Data represent mean ± s.e.m., n = 3 biologically independent experiments. Statistical significance was calculated with two-tailed Student's t-tests: *$P < 0.05$, **$P < 0.01$, ****$P < 0.0001$. P values are <0.0001 (*ANKRD28*), 0.0198 (*ATP4A*), 0.0011 (*ATP5PB*), and 0.0254 (*FAM168A*), respectively.

nature research

Corresponding author(s):   Eric T. Kool

Last updated by author(s):   Jun 21, 2023

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☒	☐	A description of all covariates tested
☒	☐	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. F, t, r) with confidence intervals, effect sizes, degrees of freedom and P value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☐	☒	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	Data were collected with softwares from the instrument manufacturers. StepOne software V2.3 was used on StepOnePlus real-time PCR system. VNMRJ (Version 4.2) was used on the NMR. Gen5 software (Version 3.08) was used on a BioTek Synergy HT plate reader. DNA fragment analysis was performed on a ABI 3130xl Genetic Analyzer.
Data analysis	Modified icSHAPE pipeline (https://github.com/qczhang/icSHAPE) and RBRP scripts (https://github.com/linglanfang/RBRP) was used for bioinformatic analysis.GraphPad Prism 9 (Version 9.5.0) was used for common plots. MestReNova (Version 14.2.3-29241) was used for NMR data processing. ImageStudioLite (Version 5.2.5) was for gel image processing. DNA fragment analysis data were analyzed with Peak Scanner Software 2 (Version 2.0). IGV tracks were processed with wiggletools (Version 1.2).

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:
- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

All sequencing data are available through the Gene Expression Omnibus (GEO) under accession GSE229331. Data supporting the findings of this study are available in the article, Supplementary Information, and Supplementary files S1-S4. Source data and bedgraphs of RBRP scores are also freely available at figshare at https://doi.org/10.6084/m9.figshare.20326824.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences   ☐ Behavioural & social sciences   ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	All biochemical and cellular assays were independently conducted at least three times (n≥3) unless otherwise stated, including in vitro translation, RT-qPCR, cell-free YBX1 reporter assay, and drug titration experiments. Western-blot analysis was conducted twice with similar results. Melting curve determination with circular dichroism spectroscopy was highly reproducible, therefore the sample size was n=2. RBRP was conducted with two biologically independent replicates (n=2) following a previous publication (Nature volume 519, pages486–490 (2015)). The RPKM and RT stops between two biological replicates show high concordance and the bioinformatics analyses were performed at RNA regions with high concordance between replicates.
Data exclusions	No data were excluded from the analyses.
Replication	All attempts at replication were successful. In vitro translation, RT-qPCR, cell-free YBX1 reporter assay, and drug titration experiments were independently repeated at least three times (n≥3) and showed similar results. Western-blot analysis in Fig.5a was independently conducted twice and showed similar results.
Randomization	All samples and cells were randomly allocated into experimental groups.
Blinding	All samples and cells were randomly allocated into experimental groups. Thus, no blinding is required.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☐	☒ Antibodies
☐	☒ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☒	☐ Human research participants
☒	☐ Clinical data
☒	☐ Dual use research of concern

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Antibodies

Antibodies used	Anti-YBX1 antibody (D299) (Cell Signaling Technology; #4202S) (dilution: 1:2000) Anti-GAPDH antibody (D4C6R) (Cell Signaling Technology; #97166) (dilution: 1:2000)
Validation	Anti-YBX1 antibody: https://www.cellsignal.com/products/primary-antibodies/yb1-d299-antibody/4202?site-search-type=Products&N=4294956287&Ntt=ybx1&fromPage=plp Anti-GAPDH antibody:https://www.cellsignal.com/products/primary-antibodies/gapdh-d4c6r-mouse-mab/97166?site-search-type=Products&N=4294956287&Ntt=gapdh&fromPage=plp

Eukaryotic cell lines

Policy information about cell lines

Cell line source(s)	HEK293 cells were obtained from ATCC.
Authentication	The morphology check by microscope was performed to authenticate the cell line.
Mycoplasma contamination	Cell lines were not tested for mycoplasma contamination.

Commonly misidentified lines
(See ICLAC register)

None.