Index of Supplementary Information for Harris *et al*.

- Supp. Data 2 → relates to Figure 5: 100vert RBNS data table used for all analyses with sequences, relevant species information, enrichment values, relevant sequence information, and relevant oligo information

SUPPLEMENTARY NOTES

Supp. Note 1

In our nsRBNS experiments, we have observed that human oligos are better bound than mouse oligos on a global scale. This can be observed for all oligos, bound alone, or unbound alone. The precise nature of these differences is difficult to define, however a few factors may contribute. While we can't be certain, one likely explanation might be in a technical challenge of performing iCLIP from mouse brain tissue vs human neuronal cell lines. It is plausible that the identified mouse binding sites were more challenging to derive due to tissue processing and crosslinking efficiencies as well as the abundance of UNK itself. These differences are apparent in some of our comparisons and limited our ability to test the impact of chimerizing mouse bound sequences into human not bound sequences to test if any enhanced UNK binding. Within these analyses, we found that while some regions from mouse bound sequences could enhance human binding, specific positions were not as impactful as we observed when chimerizing human into mouse (**Fig. 4B,C**). Generally, trends of binding patterns were overall weaker for mouse binding sites.

Another potential mechanism driving this is effect are inherent protein differences: Human recombinant protein was used for all *in vitro* experiments; however iCLIP was performed *in vivo* with species-specific proteins[1]. At the amino acid sequence level, human and mouse UNK are 95% identical and 96% similar overall while the RBDs are 99% identical[2,3]. Within the RBDs, only one amino acid is non-similar (Q321 in human, P321 in mouse). Structurally, this amino acid does not lie on the RNA-binding surface and likely does not directly contribute to RNA-protein interactions[4]. However, this single amino acid difference may affect global structural orientation, and therefore indirectly alter RNA-protein interactions *in vivo*. Indeed, previous work on transcription factors has demonstrated preferential binding of human sequences by human protein and vice versa[5].

Cellular context differences: Individual RNA and protein concentrations vary across cell-types and species. Additionally, species-specific alternative splicing can result in sequence differences and isoform expression level changes across cell-types and species and have been largely correlated with genomic evolution[6–9]. These may be particularly important when comparing tissues to cell lines. Cell type mRNA composition and heterogeneity is likely an important consideration when assessing iCLIP from tissue samples. It is plausible that cell-type specific binding events within a tissue dampen the degree to which binding to any one site is detected. Previous work has demonstrated that RNA-protein interactions across tissue culture cells versus isolated tissue can vary drastically, even when similar cell types are considered[10].

Since nsRBNS does not take into consideration any competitory binding, cofactors, salt concentrations, diffusion differences, etc. between cells and purely measures RNA-protein interactions, these factors may be different across species, even when similar cell types are examined.

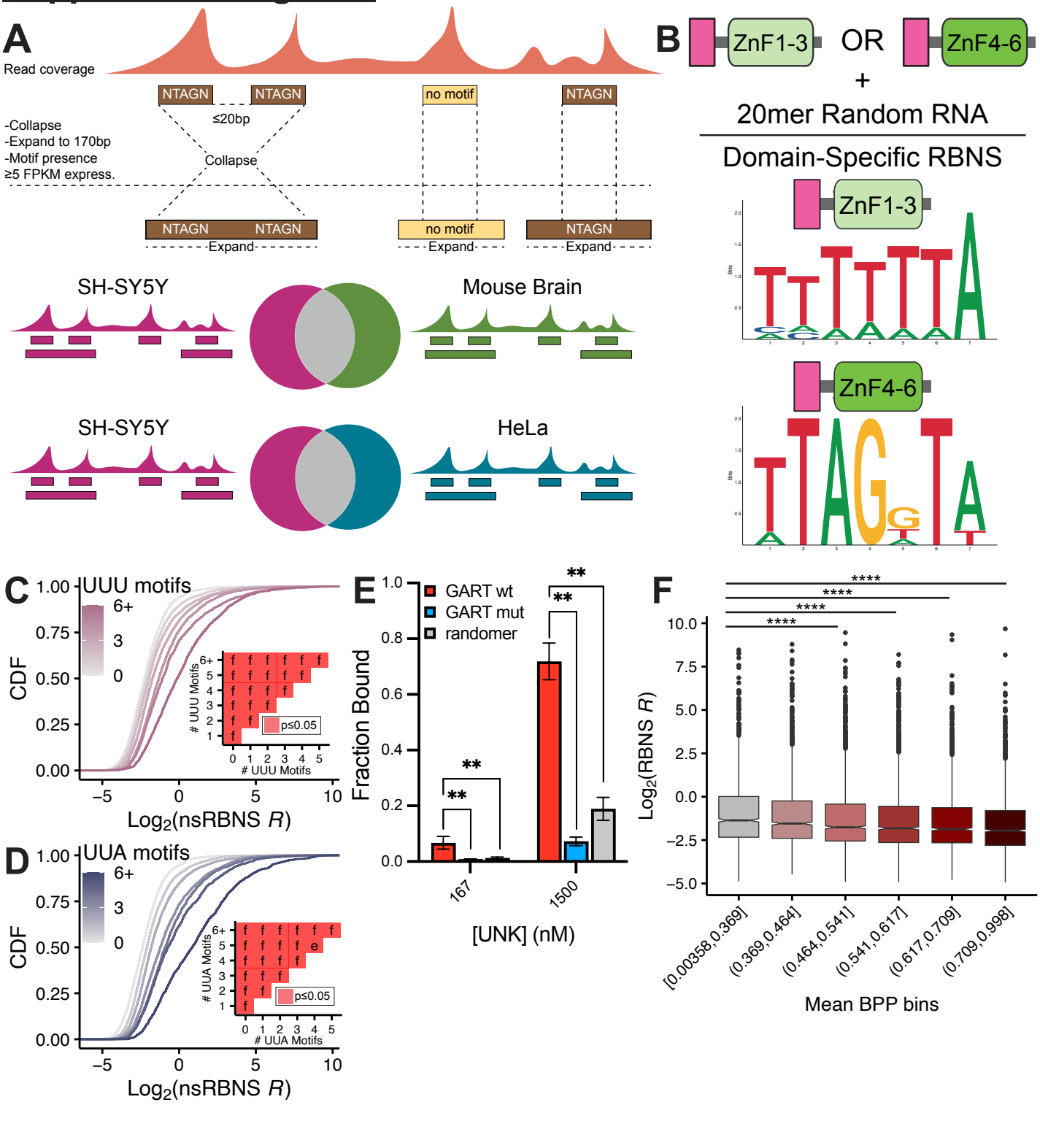
Supp. Note 2

In our 100vert RBNS experiment, we observe a decreased range of enrichments (-3 to 3 on $\log_2$ scale) versus nsRBNS (-6 to 6 on $\log_2$ scale). RBNS is inherently a zero-sum experiment (aka gain must equal loss). While both nsRBNS and 100vert RBNS are impacted by the nature of these experiments, oligo design for 100vert RBNS led to a greater impact (*i.e.*, dampening of signal).

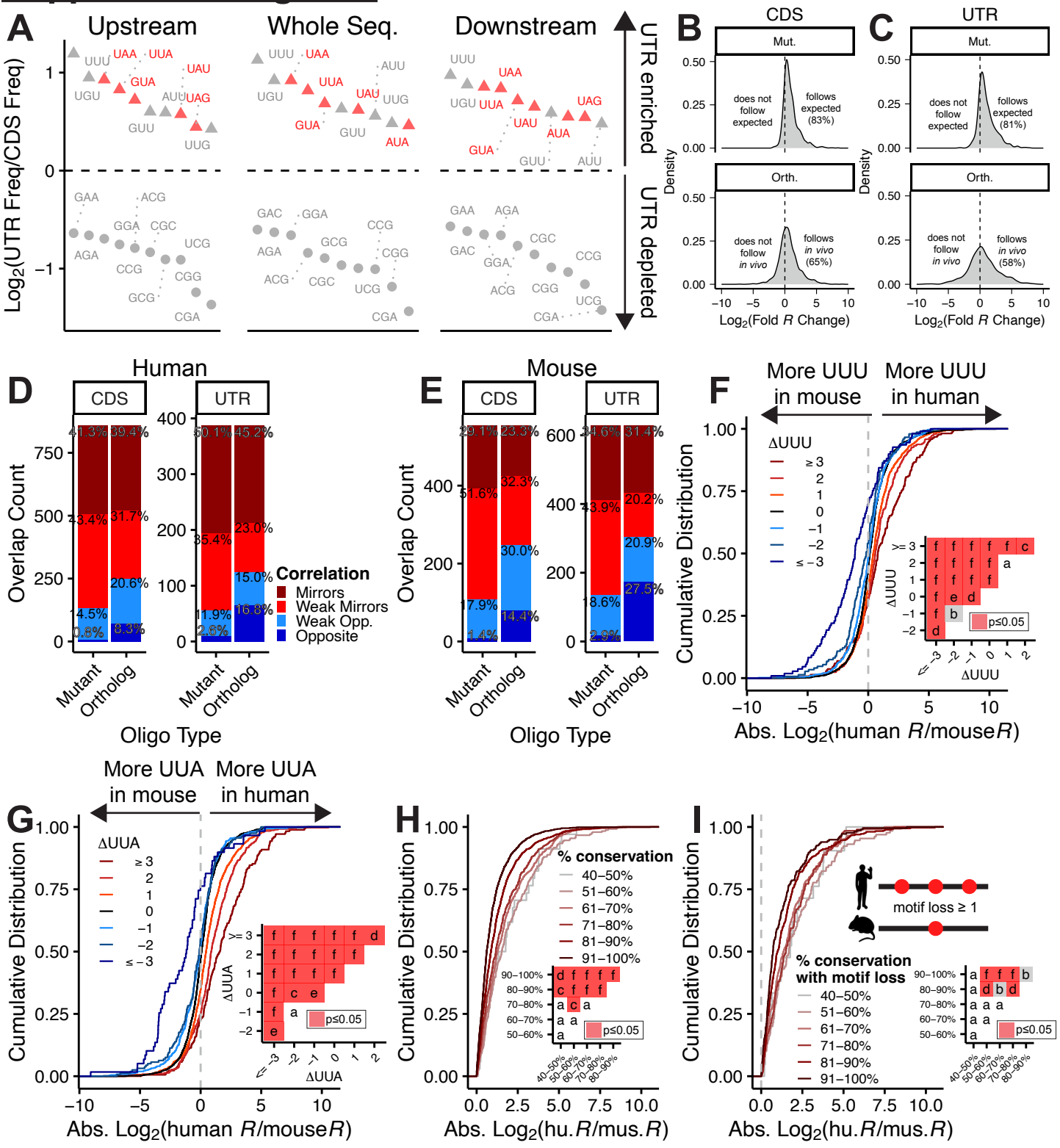
For comparison, in nsRBNS only ~5,000 (20%) out of ~25,000 oligos were predicted to bind, whereas ~7,000 (28%) had unknown binding capabilities, while the remaining ~13,000 (52%) were expected to be non-binders. While non-binders were primarily included as controls and for validation purposes, this allowed for less competition within the oligo pool, and therefore a larger separation of enrichment values.

In contrast, only ~10% of expected non-binders were included for 100vert RBNS. Further, 100vert RBNS was designed from the top bound oligos in nsRBNS, further narrowing expected binding differences across oligos. This prevalence of high-capacity binders led to an earlier saturation of UNK protein with high competition potential between the oligos. The increased competition dampened the overall signal of the experiment, resulting in a decreased range of enrichments measured.

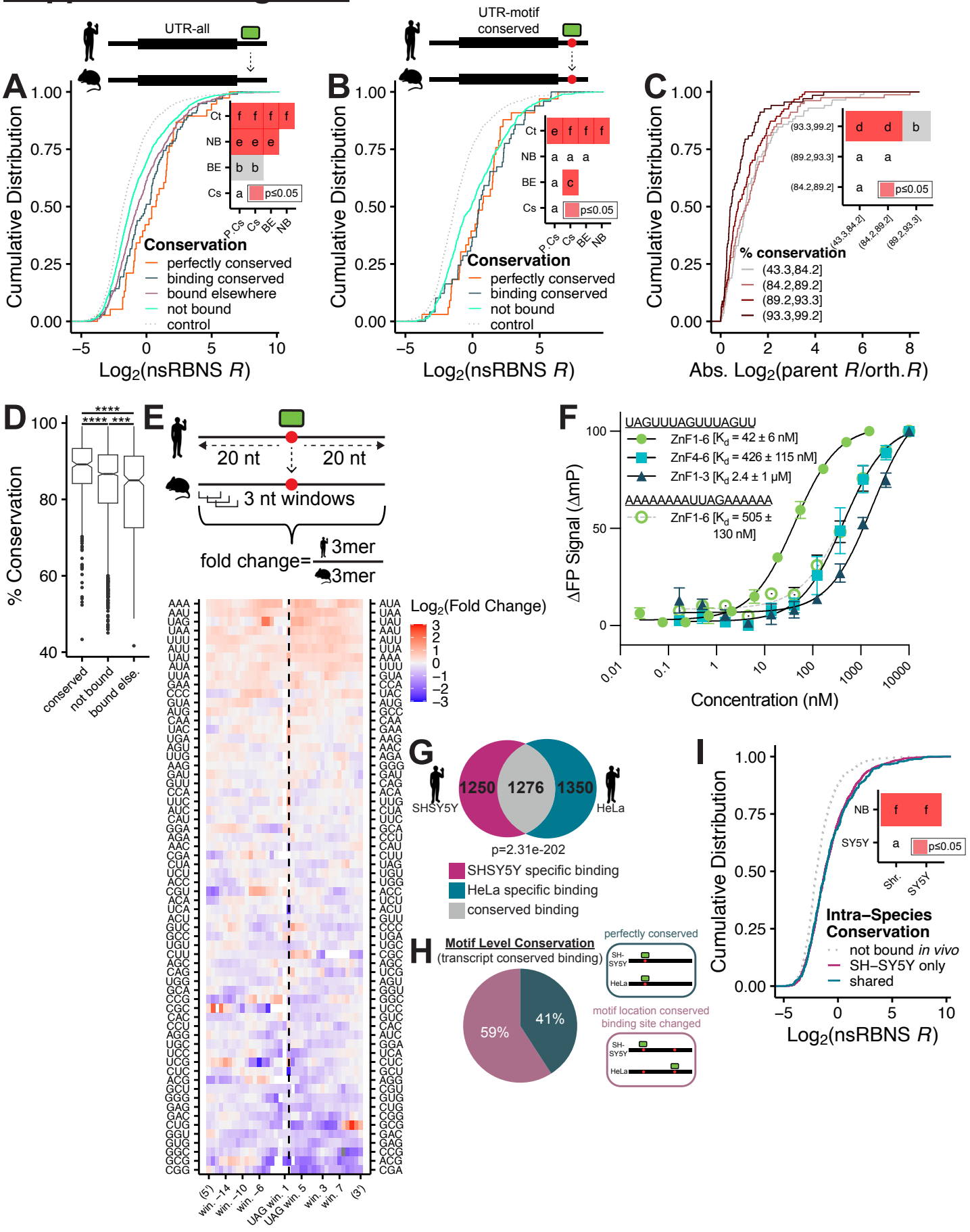
Supplemental Figure 1

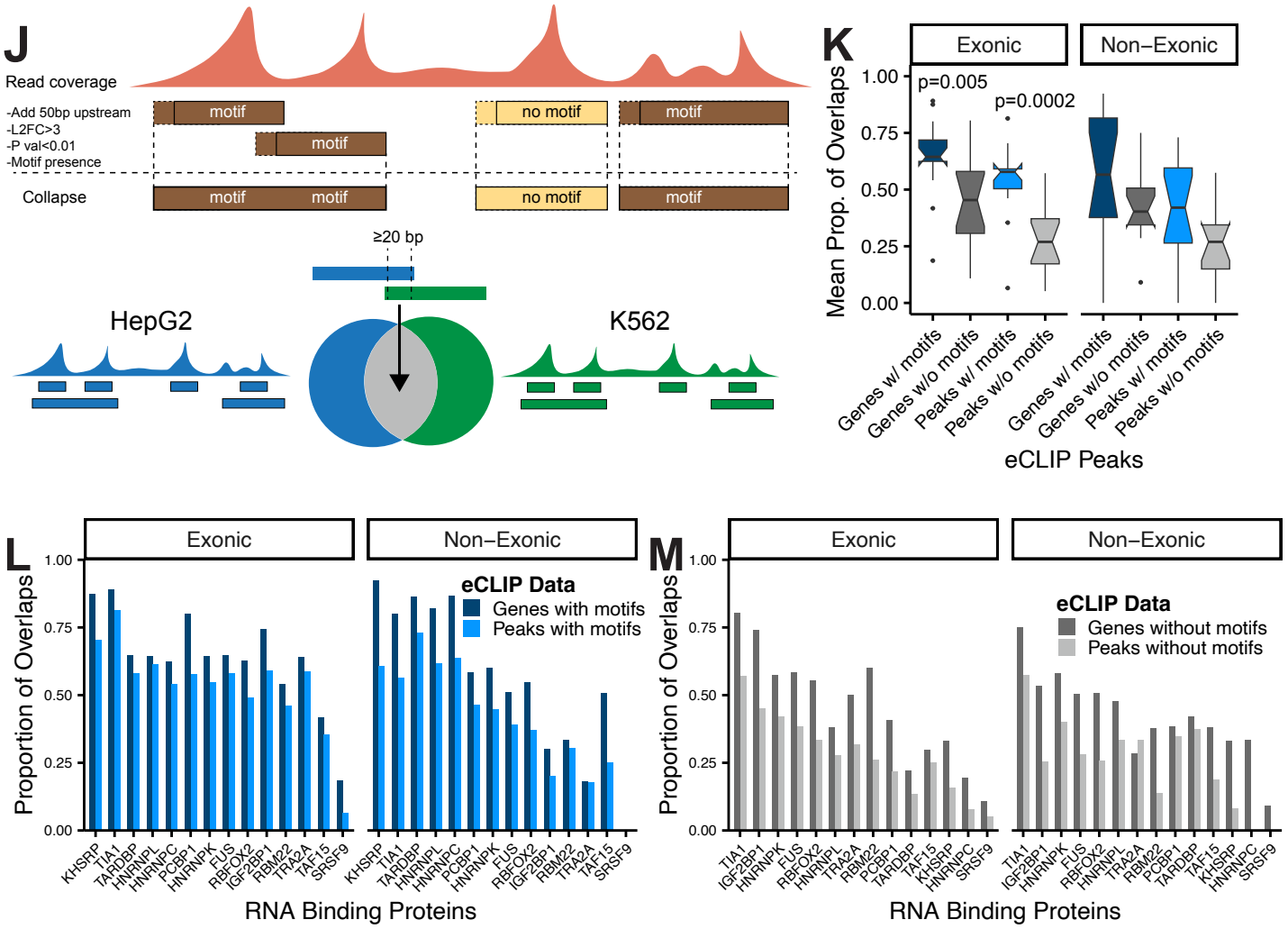


Supplemental Figure 2

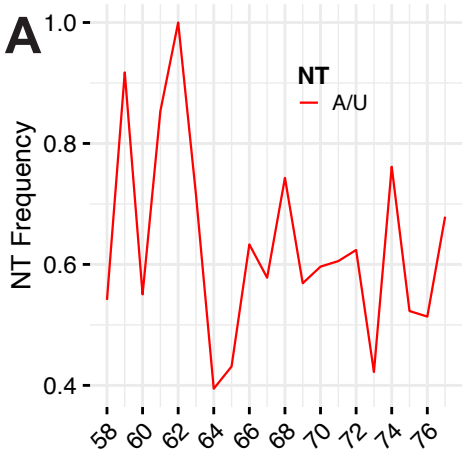


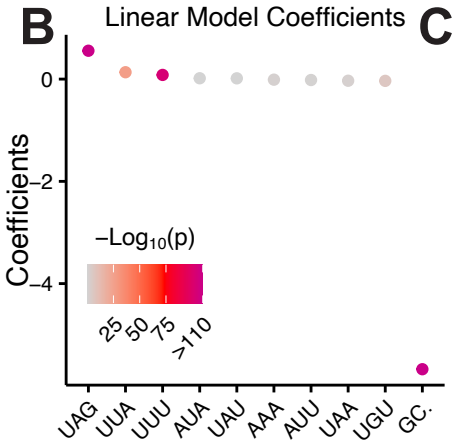
Supplemental Figure 3




J
Read coverage
-Add 50bp upstream
-L2FC>3
-P val<0.01
-Motif presence
motif
motif
no motif
motif
Collapse
motif
motif
no motif
motif
≥20 bp
HepG2
K562
K
Exonic
Non−Exonic
p=0.005
p=0.0002
Mean Prop. of Overlaps
1.00
0.75
0.50
0.25
0.00
Genes w/ motifs
Genes w/o motifs
Peaks w/ motifs
Peaks w/o motifs
eCLIP Peaks
L
Exonic
Non−Exonic
eCLIP Data
Genes with motifs
Peaks with motifs
Proportion of Overlaps
KHSRP
TIA1
TARDBP
HNRNPL
HNRNPC
PCBP1
HNRNPK
FUS
RBFOX2
IGF2BP1
RBM22
TRA2A
TAF15
SRSF9
RNA Binding Proteins
M
Exonic
Non−Exonic
eCLIP Data
Genes without motifs
Peaks without motifs
Proportion of Overlaps
TIA1
IGF2BP1
HNRNPK
FUS
RBFOX2
HNRNPL
TRA2A
RBM22
PCBP1
TARDBP
TAF15
KHSRP
HNRNPC
SRSF9
RNA Binding Proteins

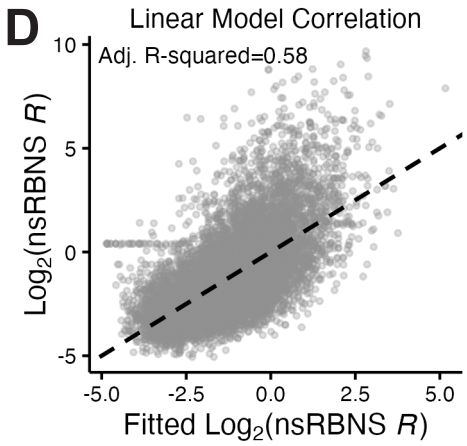
Supplemental Figure 4

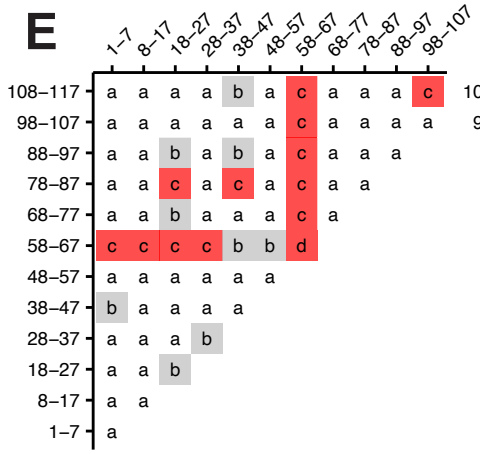


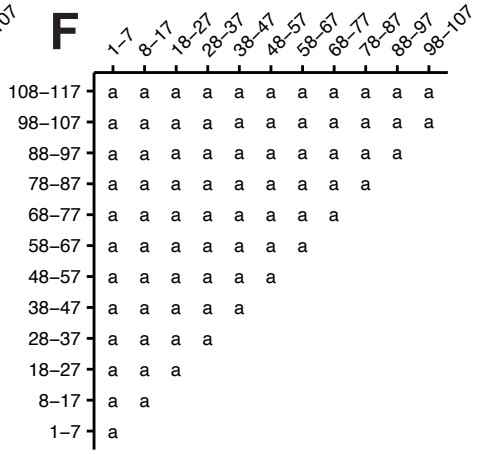


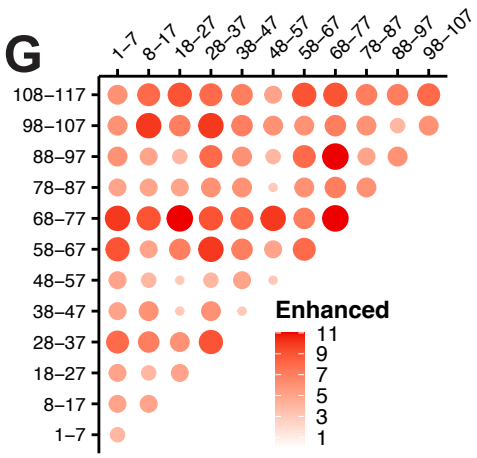
C

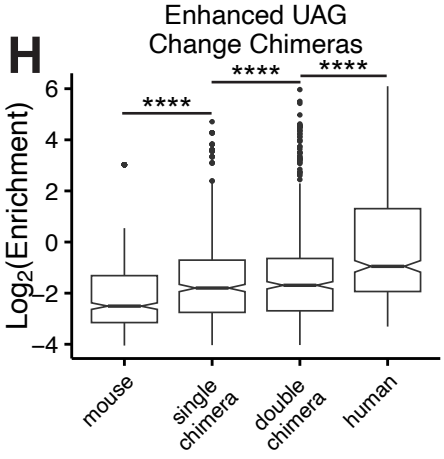
*k*mer	Coefficients	-Log_{10}(p)
UAG	0.555	>110
UUA	0.1345	27.39
UUU	0.0818	103.38
AUA	0.0165	0.74
UAU	0.0147	0.68
AAA	-0.0113	2.00
AUU	-0.0178	1.21
UAA	-0.0322	2.15
UGU	-0.0347	6.70
GC-content	-5.6743	>110

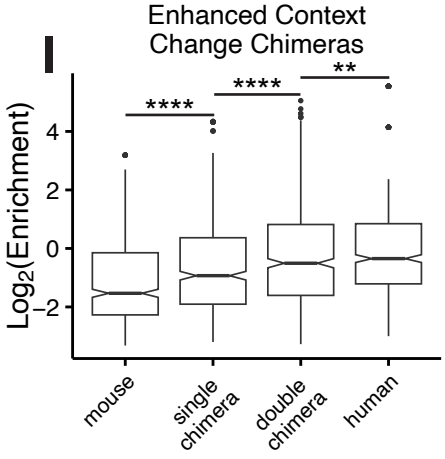




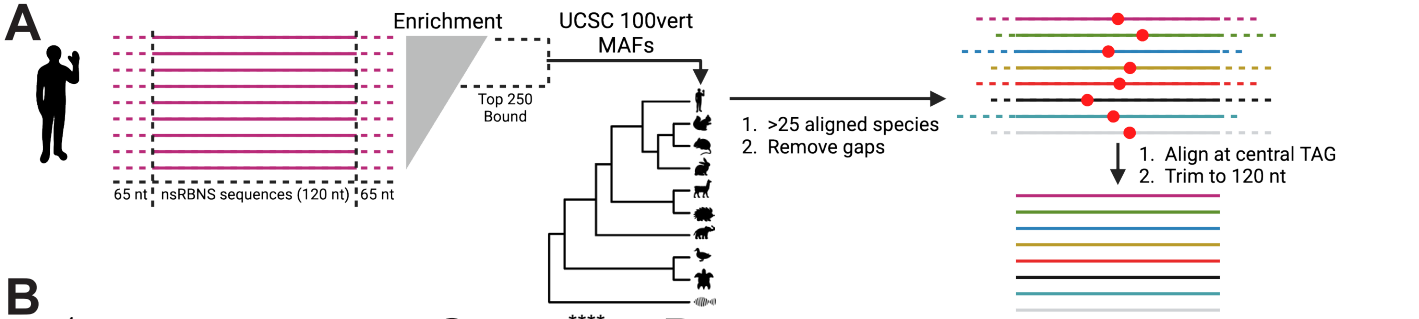


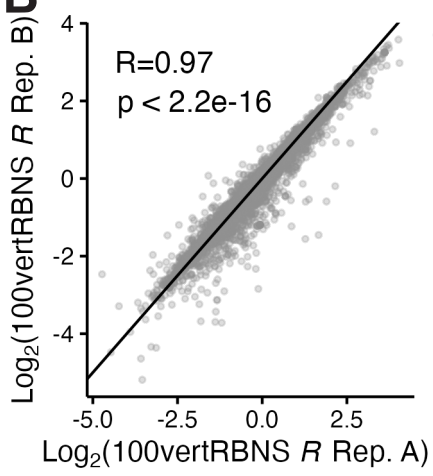


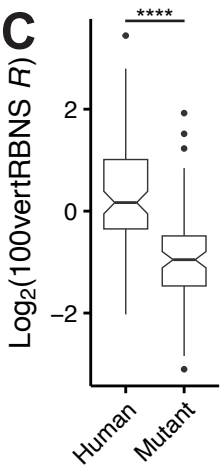


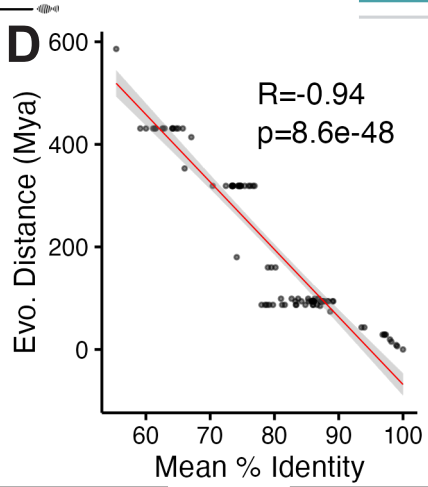


Supplemental Figure 5









E

	Mean $\Delta Log_2(R)$	Mean % ID
Evo. Dist.	-0.86	-0.94
Mean $\Delta Log_2(R)$		0.92

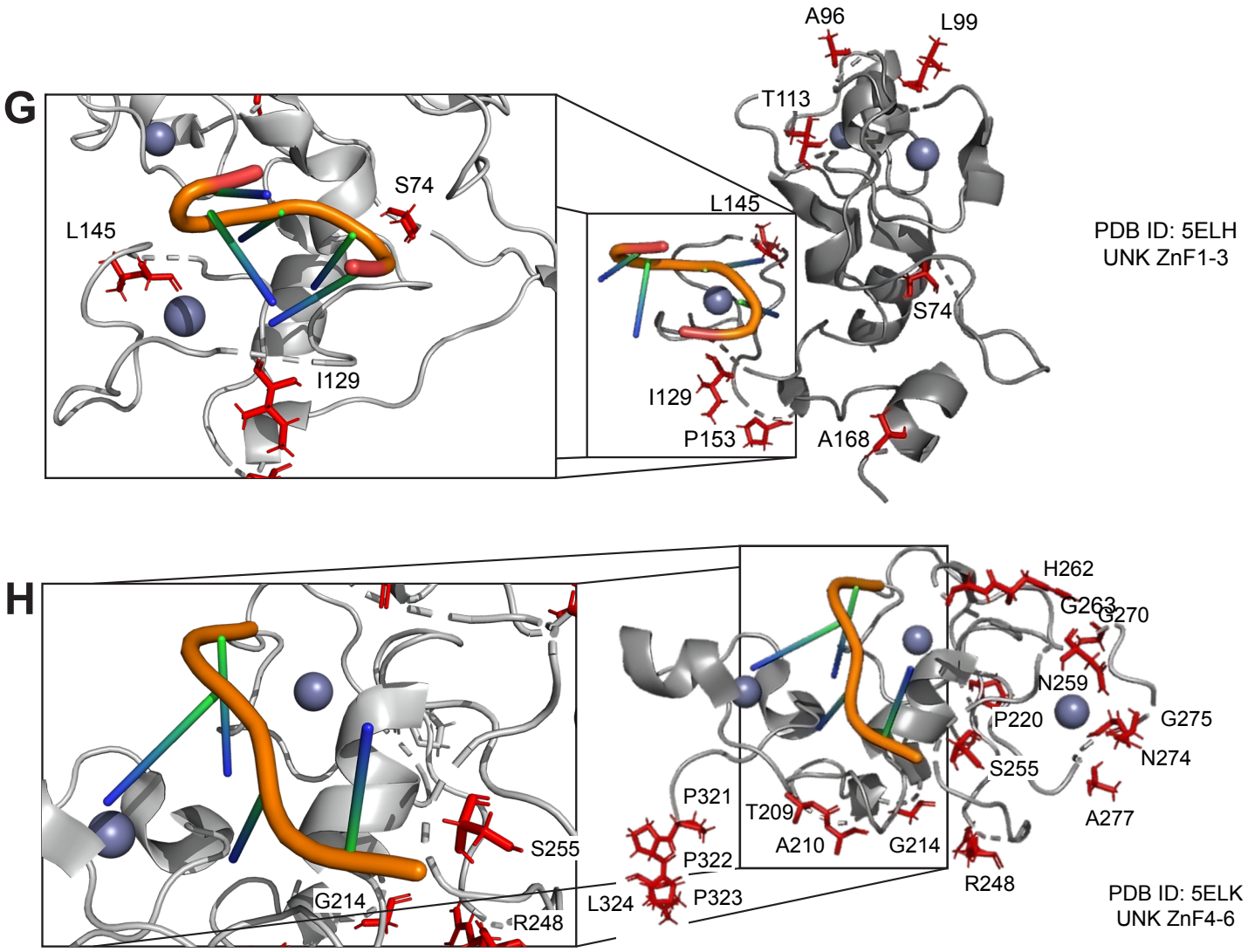
F

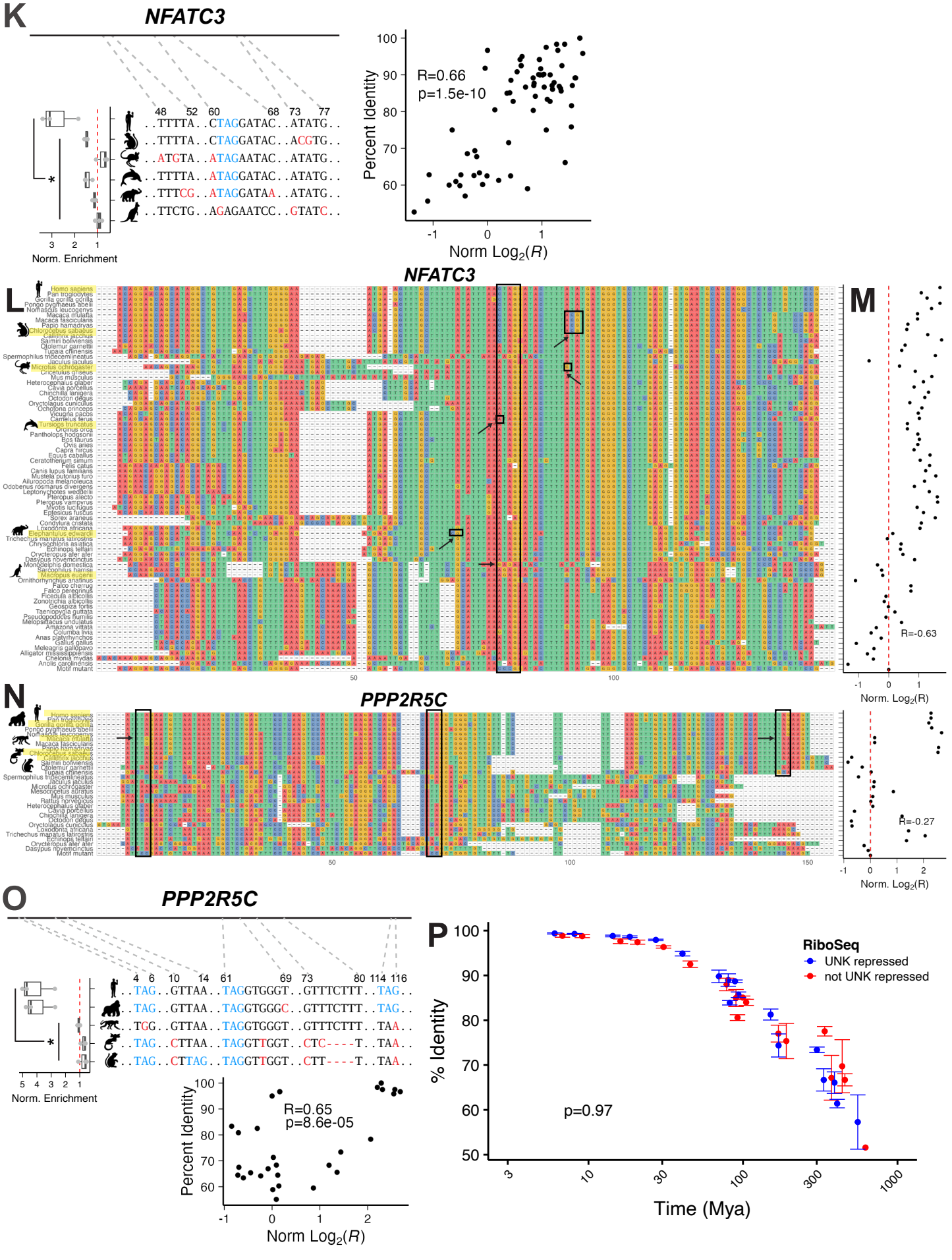
```
                H. sapiens:   31 PQHYTYLKEFRTEQCPLFVQHKCTQHRPYTCFHWHFVNQRRRRSIRRRDGTFNYSPDVYCTKYDEATGLCPEGDECPFLHRTTGDTERR 119
                M. mulatta:   31 PQHYTYLKEFRTEQCPLFVQHKCTQHRPYTCFHWHFVNQRRRRSIRRRDGTFNYSPDVYCTKYDEATGLCPEGDECPFLHRTTGDTERR 119
                C. jacchus:   31 PQHYTYLKEFRTEQCPLFVQHKCTQHRPYTCFHWHFVNQRRRRSIRRRDGTFNYSPDVYCTKYDEATGLCPEGDECPFLHRTTGDTERR 119
M. musculus:                  31 PQHYTYLKEFRTEQCPLFVQHKCTQHRPYTCFHWHFVNQRRRRSIRRRDGTFNYSPDVYCTKYDEATGLCPEGDECPFLHRTTGDTERR 119
S. scrofa:                    31 PQHYTYLKEFRTEQCPLFVQHKCTQHRPYTCFHWHFVNQRRRRSIRRRDGTFNYSPDVYCTKYDEATGLCPEGDECPFLHRTTGDTERR 119
E. caballus:                  31 PQHYTYLKEFRTEQCPLFVQHKCTQHRPYTCFHWHFVNQRRRRSIRRRDGTFNYSPDVYCTKYDEATGLCPEGDECPFLHRTTGDTERR 119
P. alecto:                    31 PQHYTYLKEFRTEQCPLFVQHKCTQHRPYTCFHWHFVNQRRRRSIRRRDGTFNYSPDVYCTKYDEATGLCPEGDECPFLHRTTGDTERR 119
F. albicollis:                14 GRSRRYLKEFRTEQCPLFVQHKCTQHRPYTCFHWHFVNQRRRRSIRRRDGTFNYSPDIYCTKYDETTGICPEGDECPFLHRTTGDTERR 102
A. mississippiensis:          29 PQHYTYLKEFRTEQCPLFVQHKCTQHRPYTCFHWHFVNQRRRRSIRRRDGTFNYSPDIYCTKYDETTGICPEGDECPFLHRTTGDTERR 117
X. tropicalis:                32 PQHYTYLKEFRTEQCPLFVQHKCTQHRPYTCFHWHFVNQRRRRSVRRRDGTFNYSPDIYCTKYDETTGICPDSDECPFLHRTTGDTERR 120
T. nigroviridis:              32 PQHYTYLKEFRTEQCPLFVQHKCTQHRPFSCFHWHFLNQRRRRPIRRRDGTFNYSPDIYCTKYDEGTGTCSDGDECPFLHRTAGDTERR 120
O. latipes:                   45 PQHYTYLKEFRTEQCPLFVQHKCTQHRPFSCFHWHFLNQRRRRPIRRRDGTFNYSPDVYCTKYDEGTGACPEGDDCPFLHRTAGDTERR 133
D. rerio:                     47 PQHYTYLKEFRTEQCPLFVQHKCTQHRPFTCFHWHFLNQRRRRPIRRRDGTFNYSPDVYCVKYDEGTGTCPDGDECPFLHRTAGDTERR 135
RNA Binding AAs:                                *                                                          *   *

H: 120 YHLRYYKTGICIHETDSKGNCTKNGLHCAFAHGPHDLRSPVYDIRELQAMEALQNGQTTVEGSIEGQSAGAASHAMIEKILSEEPRWQETAYVLGNYKTEPCKKPPRL 227
M: 120 YHLRYYKTGICIHETDSKGNCTKNGLHCAFAHGPHDLRSPVYDIRELQAMEALQNGQTTVEGSIEGQSAGAASHAMIEKILSEEPRWQETAYVLGNYKTEPCKKPPRL 227
C: 120 YHLRYYKTGICIHETDSKGNCTKNGLHCAFAHGPHDLRSPVYDIRELQAMEALQNGQTTVEGSIEGQSAGAASHAMIEKILSEEPRWQETAYVLGNYKTEPCKKPPRL 227
M: 120 YHLRYYKTGICIHETDSKGNCTKNGLHCAFAHGPHDLRSPVYDIRELQAMEALQNGQTTVEGSIEGQSAGAASHAMIEKILSEEPRWQETAYVLGNYKTEPCKKPPRL 227
S: 120 YHLRYYKTGICIHETDSKGNCTKNGLHCAFAHGPHDLRSPVYDIRELQAMEALQNGQTTVEGSIEGQSAGAASHAMIEKILSEEPRWQETAYVLGNYKTEPCKKPPRL 227
E: 120 YHLRYYKTGICIHETDSKGNCTKNGLHCAFAHGPHDLRSPVYDIRELQAMEALQNGQTTVEGSIEGQSAGAASHAMIEKILSEEPRWQETAYVLGNYKTEPCKKPPRL 227
P: 120 YHLRYYKTGICIHETDSKGNCTKNGLHCAFAHGPHDLRSPVYDIRELQAMEALQNGQTTVEGSIEGQSAGAASHAMIEKILSEEPRWQETAYVLGNYKTEPCKKPPRL 227
F: 103 YHLRYYKTGICIHETDSKGNCTKNGVHCAFAHGPHDLRSPVYDIRELQAMEALQNGQTTSEGGIEGQSAVAASHAMIEKILSEEPRWQETTYVLGNYKTEQCKKPPRL 210
A: 118 YHLRYYKTGICIHETDSKGNCTKNGVHCAFAHGPHDLRSPVYDIRELQAMEALQNGQTTSEGGIEGQSAVAASHAMIEKILSEEPRWQDTTYVLGNYKTEQCKKPPRL 225
X: 121 YHLRYYKTGICIHETDSKGHCTKNGQHCAFAHGPHDLRSPVYDIRELQAMEALQNGQPTLDGGMEGQIAMAASHAMIEKILGEEPRWQDTTYVLGHYKTEQCKKPPRL 228
T: 121 YHLRYYKTGSCIHETDAKGHCSKNGSHCAFAHGSHDLRSPVYDIREVQVMES-QGGVGATEG--DGQSGQAASTALIEKIVSEEPRWQDHNYVLSHYKTELCKKPPRL 225
O: 134 YHLRYYKTGSCIHETDAKGHCSKNGSHCAFAHGSHDLRTPVYDIREVQVMESQGGAGSGEGSGGDLQSGQAASTALIEKILSEEPRWQDNGYVLSHYKTELCKKPPRL 241
D: 136 YHLRYYKTGSCIHETDGKGHCSKNGPHCAFAHGSHDLRSPVYDIREVQVLEAQATTGLTEGSSGEGQSGVVASTALIEKILSEDPRWQDNSFVLSHYKTELCKKPPRL 243
R:       *   *  *** *                *  * **                                                    *

H: 228 CRQGYACPYYHNSKDRRRSPRKHKYR----------------------------SSPCPNVKHGDEWGDPGKCENGDACQYCHTRTEQQFHPEIYKSTKCNDMQQSGSC 308
M: 228 CRQGYACPYYHNSKDRRRSPRKHKYR----------------------------SSPCPNVKHGDEWGDPGKCENGDACQYCHTRTEQQFHPEIYKSTKCNDMQQSGSC 308
C: 228 CRQGYACPYHHNSKDRRRSPRKHKYR----------------------------SSPCPNVKHGDEWGDPGKCENGDACQYCHTRTEQQFHPEIYKSTKCNDMQQSGSC 308
M: 228 CRQGYACPYYHNSKDRRRSPRKHKYR----------------------------SSPCPNVKHGDEWGDPGKCENGDACQYCHTRTEQQFHPEIYKSTKCNDMQQAGSC 308
S: 228 CRQGYACPYYHNSKDRRRSPRKHKYR----------------------------SSPCPNVKHGDEWGDPGKCENGDACQYCHTRTEQQFHPEIYKSTKCNDMQQSGSC 308
E: 228 CRQGYACPYYHNSKDRRRSPRKHKYR----------------------------SSPCPNVKHGDEWGDPGKCENGDSCQYCHTRTEQQFHPEIYKSTKCNDMQQSGSC 308
P: 228 CRQGYACPYYHNSKDRRRSPRKHKYR----------------------------SSPCPNVKHGDEWGDPGKCESGDACQYCHTRTEQQFHPEIYKSTKCNDMQQSGSC 308
F: 211 CRQGYACPYYHNSKDRRRSPRKHKYR----------------------------SSPCPSVKHGDEWGDPSKCENGDSCQYCHTRTEQQFHPEIYKSTKCNDMQQSGSC 291
A: 226 CRQGYACPYYHNSKDRRRSPRKHKYR----------------------------SSPCPSVKHGDEWGDPTKCENGDACQYCHTRTEQQFHPEIYKSTKCNDMQQSGSC 306
X: 228 CRQGYACPYYHNSKDRRRSPRQNKYR----------------------------SSPCPNVKHGDEWGDPSKCESGDSCQYCHTRTEQQFHPEIYKSTKCNDMQQSGSC 309
T: 226 CRQGYACPYYHNSKDRRRSPHKHKYRCFGITSVVFILLQMHFNSSFAVRCFHRALPCPAVKQSEEWGDPSKCEGAEVCQYCHTRTEQQFHPEIYKSTKCNDMQQCGSC 333
O: 242 CRQGYACPYFHNSKDRRRSPHKHKYR----------------------------ALPCPAVKQSEEWGDPSKCEGAEACQYCHTRTEQQFHPEIYKSTKCNDMQQSGSC 322
D: 244 CRQGYACPYYHNSKDRRRSPHKHKYR----------------------------ALPCPSVKHSDEWGDPSKCEGGEGCQYCHTRTEQQFHPEIYKSTKCNDIQQSGNC 324
R:                                                                                                  *   **   *  ***  *

H: 309 PRGPFCAFAHVEQPPLSDDLQPSSAVS 335
M: 309 PRGPFCAFAHVEQPPLSDDLQPSSAVS 335
C: 309 PRGPFCAFAHVEQPPLSDDLQPSSAVS 335
M: 309 PRGPFCAFAHIEPPPLSDDVQPSSAVS 335
S: 309 PRGPFCAFAHVEQPPLGDDLQPSSAVS 335
E: 309 PRGPFCAFAHVEQPPLSDDLQPSSSVS 335
P: 309 PRGPFCAFAHVEQPPLSDDPQPSSAVS 335
F: 292 PRGPFCAFAHVEQPALSEDLQQSSAVS 318
A: 307 PRGPFCAFAHVEQPPLNEDLQPSSAVS 333
X: 310 PRGPFCAFAHVEQSLVCDDLQSPSVVS 336
T: 334 PRGPFCAFAHIESKLLWYRSNPFIVAR 360
O: 323 PRGPFCAFAHADKPF------------ 337
D: 325 PRGPFCAFAHLEKVS------------ 339
R:       *  *  **
```


G
A96
L99
T113
S74
L145
L145
S74
I129
I129
P153
A168
PDB ID: 5ELH
UNK ZnF1-3
H
H262
G263
G270
N259
P220
G275
N274
S255
A277
P321
T209
A210
G214
P322
P323
R248
L324
S255
G214
R248
PDB ID: 5ELK
UNK ZnF4-6
I
ATP1B1
J
R=-0.51
Norm. Log2(R)


K
NFATC3
48 52 60 68 73 77
..TTTTA..CTAGGATAC..ATATG..
..TTTTA..CTAGGATAC..ACGTG..
..ATGTA..ATAGAATAC..ATATG..
..TTTTA..ATAGGATAC..ATATG..
..TTTCG..ATAGGATAA..ATATG..
..TTCTG..AGAGAATCC..GTATC..
Norm. Enrichment
Percent Identity
R=0.66
p=1.5e-10
Norm Log2(R)
L
NFATC3
M
R=-0.63
Norm. Log2(R)
N
PPP2R5C
R=-0.27
Norm. Log2(R)
O
PPP2R5C
4 6 10 14 61 69 73 80 114 116
..TAG..GTTAA..TAGGTGGGT..GTTTCTTT..TAG..
..TAG..GTTAA..TAGGTGGGC..GTTTCTTT..TAG..
..TGG..GTTAA..TAGGTGGGT..GTTTCTTT..TAA..
..TAG..CTTAA..TAGGTTGGT..CTC----T..TAA..
..TAG..CTTAG..TAGGTTGGT..CTT----T..TAA..
Norm. Enrichment
Percent Identity
R=0.65
p=8.6e-05
Norm Log2(R)
P
% Identity
p=0.97
Time (Mya)
RiboSeq
UNK repressed
not UNK repressed

SUPPLEMENTARY LEGENDS

Supp. Fig. 1. Design and validation of natural sequence RNA Bind-n-Seq. A) Diagram of iCLIP peak overlap analysis method between human neuronal cells (SH-SY5Y) and mouse brain tissue or human HeLa cells. B) Design and motif logos of RBNS with ZnF1-3 and ZnF4-6. C-D) Cumulative distribution function of $\log_2$ nsRBNS enrichment of all oligos separated by C) UUU and D) UUA motif content. Insets show significance values for all comparisons via KS test and corrected for multiple comparisons via the BH procedure. Red denotes significant ($p \leq 0.05$). Values are as follows: e ($p \leq 0.001$), f ($p \leq 0.0001$). E) *in vitro* qPCR fraction bound for wild type and mutant *GART* RNA oligos incubated with UNK at 167 and 1500 nM. Significance was determined via one-sided Wilcox test ($p \leq 0.01$; n=5 technical replicates). Error bars represent SEM. F) Box and whisker plot of $\log_2$ nsRBNS enrichment of all oligos separated by quantile-binned mean base pair probability (BPP) of the central region (54-64). Significance determined via KS test and corrected for multiple corrections via the BH procedure. All comparisons to lowest bin are significant ($p \leq 0.0001$). Centre line denotes median (50th percentile) with bounds of box representing 25th to 75th percentiles and the whiskers denoting 5th to 95th percentiles. Outliers are denoted as individual points.

Supp. Fig. 2. Analysis of species-specific binding patterns. A) Scatter plot of the $\log_2$ *k*mer frequency fold change (UTR/CDS) of the top and bottom ten 3mers of all (left) motif-upstream (center) whole sequence and (right) motif-downstream sequences colored by UNK bound *k*mer as identified via RBNS[11]. B-C) Density plot of *in vivo* binding versus *in vitro* binding patterns for "motif mutant" and "orthologous" oligos versus *in vivo* bound oligos for B) CDS and C) UTR oligos. D-E) Bar plot of *in vivo* binding versus *in vitro* binding patterns for "motif mutant" and "orthologous" oligos versus D) human and E) mouse *in vivo* bound oligos. "Mirrors" correlation defined as ≥ 2-fold change of *in vivo* bound over *in vivo* unbound, "weak mirrors" defined as < 2 and ≥ 1-fold change, "weak opposite" defined as < 1 and > 0.5 fold change, and "opposite" defined as ≤ 0.5

fold change. F-G) Cumulative distribution function of $\log_2$ fold nsRBNS enrichment change of *in vivo* bound over *in vivo* not bound oligos separated by F) ΔUUU and G) ΔUUA content. Insets show significance values for all comparisons via KS test and corrected for multiple comparisons via the BH procedure. Red denotes significant (p≤0.05). Values are as follows: a (ns), b (p≤0.1), c (p≤0.05), d, (p≤0.01), e (p≤0.001), f (p≤0.0001). H-I) Cumulative distribution function of $\log_2$ fold nsRBNS enrichment change of *in vivo* bound over *in vivo* not bound oligos separated by percent conservation for H) all and I) *k*mer loss cross-species comparisons. Insets show significance values for all comparisons via KS test and corrected for multiple comparisons via the BH procedure. Red denotes significant (p≤0.05). Values are as follows: a (ns), b (p≤0.1), c (p≤0.05), d, (p≤0.01), f (p≤0.0001).

Supp. Fig 3. Analysis of species-specific syntenic motif level binding patterns. A-B) Cumulative distribution function of $\log_2$ nsRBNS enrichment of control (light grey; dotted), not bound (teal), bound elsewhere (purple), binding conserved (blue), and perfectly conserved (orange) for A) all UTR and B) motif conserved UTR oligos. Insets show significance values for all comparisons via KS test and corrected for multiple comparisons via the BH procedure. Red denotes significant (p≤0.05), and gray denotes nearing significant (p≤0.1). Values are as follows: a (ns), b (p≤0.1), c (p≤0.05), e (p≤0.001), f (p≤0.0001). C) Cumulative distribution function of $\log_2$ fold nsRBNS enrichment change (parent/ortholog) of “binding conserved” oligos pairs separated by percent conservation. Inset shows significance values for all comparisons via KS test and corrected for multiple comparisons via the BH procedure. Red denotes significant (p≤0.05). Values are as follows: a (ns), b (p≤0.1), d (p≤0.01). D) Boxplot of percent conservation of binding conserved, not bound, and bound elsewhere oligo pairs where the parent was bound in human, but the aligned orthologous region was unbound. Significance was determined via KS tests and corrected for multiple comparisons via the BH procedure. Statistical marks are as follows: *** — p ≤ 0.001, **** — p ≤ 0.0001. Centre line denotes median (50th percentile) with bounds of box

representing 25th to 75th percentiles and the whiskers denoting 5th to 95th percentiles. Outliers are denoted as individual points. E) Heat map of 3mer human over mouse enrichment upstream and downstream of central UAG in orthologs bound better in humans. F) Delta fluorescence polarization binding curves (n=3 technical replicates) for UNK ZnF1-6 (green circle), ZnF1-3 (blue triangle), and ZnF4-6 (teal square) incubated with a tri-UAG-containing RNA oligo graphed with delta fluorescence polarization binding curves for UNK ZnF1-6 (hollow green circle) incubated with a mono-UAG-containing RNA oligo. Each curve was normalized to its minimum and maximum fluorescence polarization signal to produce delta fluorescence polarization values. Data are presented as mean values +/- SD. G) Transcript level conservation of iCLIP UNK hits between human neuronal cells (SH-SY5Y) and human epithelial cells (HeLa). Significance determined via hypergeometric test. H) Motif level conservation of iCLIP UNK hits between human neuronal cells (SH-SY5Y) and human epithelial cells (HeLa). I) Cumulative distribution function of $\log_2$ nsRBNS enrichment of human not bound *in vivo* (light grey; dotted), SH-SY5Y-specific oligos (purple), and SH-SY5Y and HeLa shared oligos (blue) Inset shows significance values for all comparisons via KS test and corrected for multiple comparisons via the BH procedure. Red denotes significant ($p \leq 0.05$). Values are as follows: a (ns), b ($p \leq 0.1$), d, ($p \leq 0.01$). J) Diagram of eCLIP peak overlap analysis between HepG2 and K562 data. K) Mean proportion of overlaps of bound eCLIP genes and peaks between available HepG2 and K562 data[12] for 14 RBPs with and without known motifs from RBNS[11]. Data split into exonic and non-exonic peaks (see **Methods**). Significance determined via KS tests. Centre line denotes median (50th percentile) with bounds of box representing 25th to 75th percentiles and the whiskers denoting 5th to 95th percentiles. Outliers are denoted as individual points. L) Overlap of bound eCLIP genes and peaks between available HepG2 and K562 data[12] for 14 RBPs with known motifs from RBNS[11]. Data split into exonic and non-exonic peaks (see **Methods**). M) Overlap of bound eCLIP genes and peaks between available HepG2 and K562 data[12] for 14 RBPs without known motifs from RBNS[11]. Data split into exonic and non-exonic peaks (see **Methods**).

Supp. Fig. 4. Analysis of regional impacts on binding. A) A/U nucleotide frequency of the central region (pos58-77) of enriched "Context Change" chimeras. B-D) Linear modeling of all natural (non-mutated, non-chimeric) sequences. B) Plot of linear model coefficients for top UNK motifs as defined by RBNS[11], colored by $-\log_{10}$ p. C) Table of linear model coefficients and $-\log_{10}$ p for top UNK motifs. D) Correlation of fitted $\log_2$ nsRBNS enrichment via linear model versus observed $\log_2$ nsRBNS enrichment. E-F) Heat map of significance values (p) for all single and double E) "UAG Change" and F) "Context Change" chimeras with significant binding changes ($p \leq 0.05$, red) and nearing significant ($p \leq 0.1$, grey). Values are as follows: a (ns), b ($p \leq 0.1$), c ($p \leq 0.05$), d ($p \leq 0.01$), e ($p \leq 0.001$), f ($p \leq 0.0001$). Significance was determined via paired, one-sided Wilcox test and corrected for multiple comparisons via the BH procedure. G) Heat map of number of "Context Change" double chimeras enhanced upon chimerization. H-I) Box and whisker plot of $\log_2$ enrichment of mouse, single chimera, double chimera, and human oligos where both the single and double chimeras showed improved binding over mouse for H) "UAG Change" and I) "Context Change" chimeras. Significance was determined via paired, one-sample, Wilcox test. Statistical marks are as follows: ** — $p \leq 0.01$, **** — $p \leq 0.0001$. Centre line denotes median (50th percentile) with bounds of box representing 25th to 75th percentiles and the whiskers denoting 5th to 95th percentiles. Outliers are denoted as individual points.

Supp. Fig. 5. Evolutionary Conservation of Binding. A) Design of 100 vertebrate DNA pool. B) Correlation plot of two experimental 100 vertebrate nsRBNS replicates. Pearson's correlation coefficient included. C) Box and whisker plot of $\log_2$ 100vertRBNS enrichment for human and total motif mutants. Significance was determined via paired, one-sided Wilcox test. Statistical marks are as follows: **** — $p \leq 0.0001$. Centre line denotes median (50th percentile) with bounds of box representing 25th to 75th percentiles and the whiskers denoting 5th to 95th percentiles. Outliers are denoted as individual points. D) Evolutionary distance in millions of years (Y-axis)

versus mean percent RNA sequence identity (X-axis) for each aligned oligo. Pearson's correlation coefficient included. E) Full correlation of evolutionary distance in millions of years, mean percent RNA sequence identity, and mean delta $\log_2$ 100vertRBNS enrichment. F) Multiple sequence alignment of the RNA-binding domains of UNK across select vertebrates. Individual ZnF domains are highlighted via black bar. Blue denotes similar amino acids while red denotes non-similar sequence divergence as predicted by BLAST. Asterisks denote direct RNA-contacting residues as predicted based on Murn *et al.,* 2016[4] G) Structure of UNK ZnF1-3 from Murn *et al.,* 2016[4] with less stringently conserved residues highlighted in red. Note: only I129 is predicted to have direct RNA contacts based on the crystal structure. H) Structure of UNK ZnF4-6 from Murn *et al.,* 2016[4] with less stringently conserved residues highlighted in red. Note: no evolving residues are predicted to have direct RNA contacts based on the crystal structure. I) Full multiple sequence alignment for *ATP1B1*. Related to figure 5E. J) Normalized $\log_2$ 100vertRBNS enrichment of *ATP1B1*. Pearson's correlation coefficient included. K) (left) Multiple sequence alignment for *NFATC3 for Homo sapiens, Chlorocebus sabaeus, Jaculus jaculus, Tursiops truncatus, Elephantulus edwardii*, *and Macropus eugenii* with normalized 100vertRBNS enrichment by species (n=3 technical replicates). Centre line denotes median (50th percentile) with bounds of box representing 25th to 75th percentiles and the whiskers denoting 5th to 95th percentiles. All data included as individual points. Significance determined via one-sided Wilcox tests. (right) Percent RNA sequence identity (Y-axis) versus normalized delta $\log_2$ nsRBNS enrichment (X-axis). Pearson's correlation coefficient included. L) Full multiple sequence alignment for *NFATC3*. M) Normalized $\log_2$ 100vertRBNS enrichment of *NFATC3*. Pearson's correlation coefficient included. N) (left) Multiple sequence alignment for *PPP2R5C*. (right) Normalized $\log_2$ 100vertRBNS enrichment of *PPP2R5C*. Pearson's correlation coefficient included. O) (top) Multiple sequence alignment for *PPP2R5C* for *Homo sapiens, Gorilla gorilla gorilla, Macaca mulatta, Callithrix jacchus,* and *Saimiri boliviensis* with normalized 100vertRBNS enrichment by species (n=3 technical replicates). Centre line denotes median (50th percentile) with bounds of

box representing 25th to 75th percentiles and the whiskers denoting 5th to 95th percentiles. All data included as individual points. Significance determined via one-sided Wilcox tests. (bottom) Percent RNA sequence identity (Y-axis) versus normalized delta $\log_2$ 100vertRBNS enrichment (X-axis). Pearson's correlation coefficient included. P) Plot of percent identity by evolutionary distance, X-axis plotted on $\log_{10}$ scale. Error bars show SEM. Data was separated by regulation as determined via RiboSeq where blue is higher than average $\log_2$ fold change (>-0.9) and red is less than average $\log_2$ fold change (<-0.9). Significance was determined via KS test. Data are presented as mean values +/- SD.

SUPPLEMENTARY TABLES

Supp. Table 1. Primer sequences for cloning of recombinant UNK domains

Primer name	Sequence
pGEX-GST-SBP-UNK ZnF1-3 forward primer	GCGTGAACCGGGATCCCCGCAGCACTACACGT
pGEX-GST-SBP-UNK ZnF1-3 reverse primer	ACGATGCGGCCGCTCAATTCTGCAAGGCCTCC
pGEX-GST-SBP-UNK ZnF4-6 forward primer	GCGTGAACCGGGATCCCCTCGGTGGCAAGAGA
pGEX-GST-SBP-UNK ZnF4-6 reverse primer	ACGATGCGGCCGCTCAGGACACAGCTGAGGAA

Supp. Table 2. Primer sequences for RBNS reverse transcription and PCR

Primer name	Sequence
nsRBNS RT primer	TCCTTGGCACCCGAGAATTCCANNNNNNNNTGATGCTCAATCCTCTGTTG
RBNS index primer	CAAGCAGAAGACGGCATACGAGAT[N_6 barcode]GTGACTGGAGTTCCTTGGCACCCGAGAATTCCA
RBNS reverse primer	AATGATACGGCGACCACCGAGATCTACACGTTCAGAGTTCTACAGTCCGA

Supp. Table 3. Accession numbers for ENCODE RNA-seq data

Encode ID	Cell line	Rep
ENCFF267RKD,ENCFF455VYN	K562	1
ENCFF606ZTR,ENCFF444KCV	K562	2
ENCFF713MNU,ENCFF478DZZ	HepG2	1
ENCFF936SLY,ENCFF446UEC	HepG2	2

Supp. Table 4. Accession numbers for ENCODE eCLIP data

Encode ID	RBP Target	Cell Line
ENCFF073PCD	HNRNPL	HepG2
ENCFF077OSY	HNRNPL	K562
ENCFF082QGS	RBFOX2	HepG2
ENCFF103PRM	HNRNPL	HepG2
ENCFF105GZJ	TRA2A	HepG2
ENCFF121XCN	PCBP1	K562
ENCFF127WMZ	KHSRP	K562
ENCFF145YYK	IGF2BP1	HepG2
ENCFF150ZOO	RBM22	K562
ENCFF185IDD	IGF2BP1	K562
ENCFF230QOU	HNRNPC	HepG2
ENCFF241AOZ	HNRNPK	K562

ENCFF253ZSN	RBM22	HepG2
ENCFF288MWL	HNRNPK	HepG2
ENCFF327JJE	SRSF9	HepG2
ENCFF348TPU	TIA1	K562
ENCFF374XQF	KHSRP	K562
ENCFF378HWC	TIA1	HepG2
ENCFF383ZAQ	TAF15	HepG2
ENCFF390PJW	RBFOX2	HepG2
ENCFF401YRZ	HNRNPL	K562
ENCFF402AIE	TRA2A	K562
ENCFF409DPS	TAF15	K562
ENCFF421FJD	TARDBP	HepG2
ENCFF432ASF	SRSF9	HepG2
ENCFF443KJS	FUS	K562
ENCFF502OYV	PCBP1	HepG2
ENCFF526OQL	HNRNPC	K562
ENCFF534YQS	HNRNPC	HepG2
ENCFF545NBF	RBM22	HepG2
ENCFF565ILV	FUS	K562
ENCFF606RXB	TARDBP	K562
ENCFF611AHG	KHSRP	HepG2
ENCFF613UUR	SRSF9	K562
ENCFF618ZPP	TRA2A	K562
ENCFF626XAA	TARDBP	HepG2
ENCFF664RLU	HNRNPC	K562
ENCFF669TNM	TARDBP	K562
ENCFF674TKN	TIA1	K562
ENCFF685MZA	FUS	HepG2
ENCFF705SDK	IGF2BP1	HepG2
ENCFF754XAQ	HNRNPK	HepG2
ENCFF766DUS	PCBP1	K562
ENCFF779OIO	RBFOX2	K562
ENCFF824IDO	HNRNPK	K562
ENCFF853FGC	IGF2BP1	K562
ENCFF856EHA	TIA1	HepG2
ENCFF860QZG	SRSF9	K562
ENCFF899HGF	RBM22	K562
ENCFF899ZEH	TAF15	HepG2
ENCFF910WLP	RBFOX2	K562
ENCFF914VUW	PCBP1	HepG2
ENCFF951IBI	FUS	HepG2
ENCFF966KQG	TRA2A	HepG2
ENCFF988MWD	KHSRP	HepG2
ENCFF996BXS	TAF15	K562

Supp. Table 5. Sequences for synthetic UAG fluorescence polarization RNA oligos

Oligo name	Sequence

Mono-UAG FP RNA	AAAAAAAAUUAGAAAAAA /6-FAM/
Tri-UAG FP RNA	UAGUUUAGUUUAGUU /6-FAM/

Supp. Table 6. Sequences for synthetic *GTPB4* chimeras flourescence polarization RNA oligos

Oligo name	Sequence
Human *GTPB4* FP RNA	UACAUCGCAUUAGACAUUUU /6-FAM/
Mouse *gtpb4* FP RNA	UUCAUCGGAUUCGACAUUUU /6-FAM/
Chi58-67 *gtpb4* FP RNA	UUCAUCGCAUUAGACAUUUU /6-FAM/

Supp. Table 7. Final RNA sequences for qPCR binding assay

Oligo name	Sequence
WT GART RNA	GAGUUCUACAGUCCGACGAUCGUUAUCUCCAACAAAGCCGC AGUAGCUGGGUUAGAUAAAGCGGAAAGAGCUGGUAUUCCCA CUAGAGUAAUUAAUCAUAAACUGUAUAAAAAUCGUGUAGAAU UUGACAGUGCAAUUGACCAACAGAGGAUUGAGCAUCA
Mutant GART RNA	GAGUUCUACAGUCCGACGAUCGUUAUCUCCAACAAAGCCGC AGUAGCUGGGUUAGAUAAAGCGGAAAGAGCUGGUAUUCCCA CCCGAGUAAUUAAUCAUAAACUGUAUAAAAAUCGUGUAGAAU UUGACAGUGCAAUUGACCAACAGAGGAUUGAGCAUCA

Supp. Table 8. qPCR primers used for qPCR binding assay

Primer name	Sequence
qPCR_FWD	GAGTTCTACAGTCCGACGATC
qPCR_REV	TGATGCTCAATCCTCTGTTG

SUPPLEMENTARY REFERENCES

1. Murn, J. *et al.* Control of a neuronal morphology program by an RNA-binding zinc finger protein, Unkempt. *Genes Dev* **29**, 501–512 (2015).
2. UniProt. UniProt: the Universal Protein knowledgebase in 2023. *Nucleic Acids Res* **51**, D523–D531 (2023).
3. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *J Mol Biol* **215**, 403–410 (1990).
4. Murn, J., Teplova, M., Zarnack, K., Shi, Y. & Patel, D. J. Recognition of distinct RNA motifs by the clustered CCCH zinc fingers of neuronal protein Unkempt. *Nat Struct Mol Biol* **23**, 16–23 (2016).
5. Wilson, M. D. *et al.* Species-specific transcription in mice carrying human chromosome 21. *Science (1979)* **322**, 434–438 (2008).
6. Merkin, J., Russell, C., Chen, P. & Burge, C. B. Evolutionary dynamics of gene and isoform regulation in Mammalian tissues. *Science (1979)* **338**, 1593–1599 (2012).
7. Barbosa-Morais, N. L. *et al.* The evolutionary landscape of alternative splicing in vertebrate species. *Science (1979)* **338**, 1587–1593 (2012).
8. Barr, K. A., Rhodes, K. L. & Gilad, Y. The relationship between regulatory changes in cis and trans and the evolution of gene expression in humans and chimpanzees. *Genome Biol* **24**, 1–21 (2023).
9. Sudmant, P. H., Alexis, M. S. & Burge, C. B. Meta-analysis of RNA-seq expression data across species, tissues and studies. *Genome Biol* **16**, 1–11 (2015).
10. Perez-Perri, J. I. *et al.* The RNA-binding protein landscapes differ between mammalian organs and cultured cells. *Nat Commun* **14**, 2074 (2023).
11. Dominguez, D. *et al.* Sequence, structure, and context preferences of human RNA binding proteins. *Mol Cell* **70**, 854–867 (2018).

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