FMRF targets distinct mRNA sequence elements to regulate protein expression

Manuel Ascano Jr1, Neelanjana Mukherjee2, Pradeep Bandaru1, Jason B. Miller1, Jeffrey D. Nusbaum1, David L. Corcoran2, Christine Langlois3, Mathias Munschauer1, Scott Dewell4, Markus Hafner1, Zev Williams1,3, Uwe Ohler2* & Thomas Tuschl1

Fragile X syndrome (FXS) is a multi-organ disease that leads to mental retardation, macro-orchidism in males and premature ovarian insufficiency in female carriers. FXS is also a prominent monogenic disease associated with autism spectrum disorders (ASDs). FXS is typically caused by the loss of fragile X mental retardation 1 (FMR1) expression, which codes for the RNA-binding protein FMRP. Here we report the discovery of distinct RNA-recognition elements that correspond to the two independent RNA-binding domains of FMRP, in addition to the binding sites within the messenger RNA targets for wild-type and I304N mutant FMRP isoforms and the FMRP paralogues FXR1P and FXR2P (also known as FXR1 and FXR2). RNA-recognition-element frequency, ratio and distribution determine target mRNA association with FMRP. Among highly enriched targets, we identify many genes involved in ASD and show that FMRP affects their protein levels in human cell culture, mouse ovaries and human brain. Notably, we discovered that these targets are also dysregulated in Fmr1−/− mouse ovaries showing signs of premature follicular overdevelopment. These results indicate that FMRP targets share signalling pathways across different cellular contexts. As the importance of signalling pathways in both FXS and ASD is becoming increasingly apparent, our results provide a ranked list of genes as basis for the pursuit of new therapeutic targets for these neurological disorders.

Most clinical cases of FXS are a result of a hyper-expansion and methylation of CGG repeats within the promoter of FMR1, leading to a loss of its expression1-3. The FMR1 RNA-binding protein family has three members, FMRP, FXR1P and FXR2P, which possess two centrally located KH domains and a carboxy-terminal arginine-glycine-rich region implicated in mRNA binding4,7. FMR1 codes for multiple protein isoforms, but is predominantly expressed as a 69 kDa protein (isoform 7)4,9. Isoform 1 and six other alternative splice variants include exon 12, with isoform 1 coding for the full-length protein (71 kDa). Exon 12 insertion lengthens the second KH (KH2) RNA-binding domain, possibly influencing FMRP RNA-binding specificity or affinity. The I304N mutation, first described in a FXS patient, is also located in the KH2 domain and is reported to attenuate association with RNA and polysomes14-16. FMR1-family proteins are implicated in various RNA processes including RNA subcellular localization by facilitating nucleo-cytoplasmic shuttling13 and association with motor proteins14-16, and are also suggested to mediate translational regulation12,17. Given the critical role of FMR1 in human cognition and premature ovarian insufficiency18,19, there have been intensive efforts towards the identification of the RNA targets of FMRP, with the view that their discovery would shed light on the array of related disorders and provide options for molecular therapy15-26. No precise RNA-recognition element (RRE) has been defined and very few bona fide mRNA targets are confirmed22.

RNA target sites of the FMR1 protein family

To identify the binding sites of FMR1-family proteins (Fig. 1a and Supplementary Fig. 1), we first compared photocrosslinking methods28-30 using stable Flag-haemagglutinin (HA)-tagged FMRP isoform 7 in HEK293 cells (Fig. 1b), as these cells and human brain share 90% of expressed genes according to a comparison of existing RNA-seq data sets31-33 (Supplementary Fig. 2). The difference between FMR1 levels in the experimental system and the brain is 1.3-fold, as calculated using RNA-seq data and the quantified expression of FMR1 in our stable cells. We found that 4-thiouridine (4SU) photoactivatable ribonucleoside-enhanced crosslinking and immunoprecipitation (PAR-CLIP) provided the highest yield of crosslinked RNAs, and this approach was used for all FMR1-family proteins (Fig. 1c). Complementary DNA libraries were generated after PAR-CLIP and Illumina-sequenced (Supplementary Table 1). Genome-aligned reads were grouped by PARalyzer34 to identify segments of RNA that represented peaks of T-to-C conversion, termed binding sites. PARalyzer separated closely spaced binding sites connected by overlapping reads and yielded a median RNA segment length of 33 nucleotides (Supplementary Fig. 3). FMRP isoforms 1 and 7 bound to approximately 80,000–100,000 sites, of which >85% mapped to ~6,000 mRNAs (Supplementary Tables 1, 2 and http://fmrp.rockefeller.edu). FXR1P and FXR2P protein binding sites are comprised within FMRP binding sites, with an overlap of 95% (Supplementary Table 3).

Nearly all mRNA-binding sites were located in exons (>90%) (Fig. 2a) and distributed between coding sequence (CDS) and 3’ untranslated regions (UTRs) (>95%, total), with slightly more CDS sites, similar to distributions seen for other cytoplasmic RNA-binding proteins38. The computational sequence analysis method cERMIT35 revealed two key RREs, ACUK and WGGA (in which K = G or U and W = A or U) (Fig. 2b and Supplementary Fig. 4). Together, ACUK and WGGA RREs were found in ≥50% of mRNA-binding sites in isoforms 1 and 7, occurring exclusively or together within the same binding site (Fig. 2c). Remaining binding sites typically contained close derivatives of either RRE.
isoforms 1 and 7 vary by the presence of exon 12 (black) within KH2. The I304N mutation (red asterisk) is located within the KH2 domain. The arginine–glycine-rich region (RG; orange bars) is also implicated in RNA binding. The lengths of proteins in amino acids are indicated. We established stable inducible cell lines expressing Flag-HA–tagged wild-type and I304N mutants of FMRP (isoforms 1 and 7), and its paralogues FXR1P and FXR2P (ref. 47). Flag-HA epitope-tagged wild-type and I304N mutants of FMRP (isoforms 1 and 7), and its paralogues FXR1P and FXR2P (ref. 47). RNA–FMRP covalently bound to Flag-HA–FMRP (arrow) was determined after crosslinked HEK293 cells expressing Flag-HA-tagged FMRP isoform 7 were separated by SDS–polyacrylamide gel electrophoresis (PAGE). The migrations of protein mass standards are indicated. Enrichment of radiolabelled RNA relating with the number of RREs within a PARalyzer-defined binding site. An RNA segment containing nine WGGAAs bound almost two orders of magnitude tighter than those containing one WGGA, whereas binding of ACUK-containing RNAs varied only fivefold. EMSAs using RNAs representing target sites within PAR-CLIP-identified targets showed enrichment by RIP-chip, of 646 transcripts were twofold enriched but not identified as PAR-CLIP targets. We used binding-site information obtained by PAR-CLIP to infer the salient features for stable association in RIP-chip (Fig. 4, Supplementary Fig. 8 and Supplementary Table 6). Increasing frequency of WGGAs and ACUK-containing elements led to greater RIP-chip enrichment, in agreement with in vitro affinity measurements. On average, top targets contained more RRE binding sites (18 per transcript) compared to the least-enriched targets (13 per transcript).

Figure 2 | Analysis of FMR1-family protein mRNA-binding sites. a, Distribution of binding sites within mRNA targets of the FMR1 protein family. b, Two key RREs were inferred from FMRP isoform 1 and 7 binding sites; top, ACUK; bottom, WGGA. c, Distribution of FMRP binding sites, colour-coded on the basis of cERMIT-inferred RREs, across representative targets. Open boxes and thin lines indicate CDS and UTRs, respectively. Numbers indicate nucleotide number. RPKM, reads per kilobase of mature transcript per million mapped reads.
Regulatory impact of FMRP on its mRNA targets

To assess the effect of FMRP binding sites on mRNA stability, small interfering (siRNA) knockdown of FMR1 or the FMR1 family was performed and mRNA-expression profiles were analysed by microarray. We found no evidence for FMRP affecting target mRNA abundance (data not shown).

A panel of FMRP targets was selected on the basis of enrichment in RIP-chip, low-to-intermediate expression in RNA-seq, similar abundance in the human brain (using published microarray^28 and RNA-seq data sets^32,33) and with documented neurological and human disease relevance, and then analysed by quantitative western blot to determine protein levels as a function of FMRP expression (Fig. 4e). FXR2P, HUWE1, KDM5C and MTOR protein levels, among others, showed up to 50% reduction in protein levels upon expression of FMRP in HEK293. We analysed lysates prepared from human post-mortem brains. Four FXS brains (Supplementary Fig. 9) were available with age/sex/anatomic-matched controls from prefrontal cortical, hippocampal and cerebellar regions. Although only four out of eight antibodies yielded quantifiable bands in brain lysates, we observed a general trend of elevated target protein expression levels in FXS brains. This is the inverse of FMRP-overexpression effects in HEK293, and consistent with FMRP affecting the protein levels of its mRNA targets.

The mRNA targets identified here are from a human transcriptome in which most genes are comparably expressed in the human brain (Supplementary Fig. 2). We discovered ASD-related and numerous other genes implicated in neuronal disorders associated with FXS and validated representatives by EMSA, RIP-chip and immunoblot. We found genes involved in Angelman, Prader–Willi, Rett, and Cornelia de Lange syndromes. Interestingly, the ASD and Angelman syndrome-associated gene UBE3A ubiquitinates ARC and SACS^40; ARC is a well-known target and here we identify SACS as a targeted transcript. These findings potentially provide the molecular link to tie together elements of clinically overlapping disorders, principally setting a molecular target framework for characterizing the connections between FXS and its associated phenotypes.

Although fragile-X-related diseases are primarily considered to be CNS disorders, at least two other target organs are affected: the testes and ovaries. We reasoned that changes in FMR1 expression lead to dysregulation of largely overlapping targets of shared across all affected organs. Thus, dysregulated genes and pathways in the brain might also contribute to phenotypes in the testes and ovaries. We therefore examined the ovaries of Fmr1<sup>−/−</sup> mice^41 as CNS and macroorchidism phenotypes had been reported, yet ovary development had largely been under-investigated. Ovaries from Fmr1<sup>−/−</sup> mice were markedly larger by 3 weeks post-birth compared to wild-type controls (Fig. 5a, b). At 12 and 18 weeks post-birth, knockout ovaries were 22% and 72% larger by mass compared to age-matched controls, respectively. Notably, we found increased protein levels of Tsc2, Sash1 and Mtor (Fig. 5c). As a role for the Mtor pathway in the regulation of ovarian development has been previously identified, it is tempting to
and from 18-week-old a greater than expected number of follicles compared to wild type (\(\text{WT}\)). As for human samples, Pabpc1 was used for normalization (\(\text{WT}\)), morphological (\(\text{Fmr1}\)), and ratio control as it was a gene with PAR-CLIP binding sites but showed no target protein level difference from control (%).

**Figure 4 | RRE-dependent enrichment criteria for FMRP association with mRNAs.** a-d. Cumulative distribution fraction plots of FMRP targets on the basis of indicated criteria. Transcripts were grouped and colour-coded on the basis of indicated bins. Non-targets are mRNAs transcripts with zero PAR-CLIP binding sites, although detectable in the array; total denotes the sum of non-targets and PAR-CLIP-identified targets detectable by RIP-chip. RIP-chip binding sites, within the human \(\text{Pik3ca}\) transcript. However, we find that it is a less-enriched RIP-chip target compared to \(\text{MTOR}\) and \(\text{TSC2}\), whose protein levels appear regulated in an FMRP-dependent manner. Interestingly, recent evidence demonstrated that \(\text{Tsc}\) mutant mice\(^{44-46}\), which have increased \(\text{Mtor}\) activity, had impaired mGluR-induced long-term synaptic depression and protein synthesis compared to \(\text{Fmr1}\) mice; crossing \(\text{Tsc}\) with \(\text{Fmr1}\) mice corrected the phenotypes\(^{44}\). Given our results it is likely that the decreased protein synthesis observed in \(\text{Tsc}\) mice is due to the \(\text{Mtor}\) pathway inducing increased FMRP regulatory activity, a consequence which can be alleviated through the removal of FMRP itself in \(\text{Tsc}\) mice. However, it should be noted that FMRP can also associate with transcripts of ERK pathway components. Therapeutic targeting of the \(\text{MTOR}\) pathway has become an important goal, but must be further guided by additional functional analysis, particularly of FMRP targets upstream and downstream of \(\text{MTOR}\) and interconnected signalling pathways (Supplementary Fig. 10). Combined, our validation work in \(\text{Fmr1}\) knockout mouse ovaries and in human brain demonstrate that the effect of FMRP binding to specific target genes identified in cell culture is extensible to physiologically relevant contexts.
METHODS SUMMARY

Methods are described in greater detail in Supplementary Information. FVB129P2 fragile X mice were a gift from S. Zukin. Gateway plasmids (Invitrogen) generated in this study will be deposited in http://www.addgene.org. FlpIn T Rex HEK293 (Invitrogen) inducible stable cell lines were generated per manufacturer’s instructions. The titres, source and use of antibodies used in this study are listed in Supplementary Information. PAR-CLIP was performed essentially as described, except that the second RNase T1 digestion was omitted following the immunoprecipitation. Recombinant wild-type and mutant FMRP isoform 1 proteins were purified using a baculovirus expression system (Invitrogen). EMSAs and western blots were performed as described in [146, 247, 261, 2011]. Parameters of computational analyses are described in Supplementary Information and in the relevant sections at http://fmrp.rockefeller.edu/. Relevant data sets, including raw data, are available at http://fmrp.rockefeller.edu/ and gene expression omnibus (GSE39686).

Full Methods and any associated references are available in the online version of the paper.

Received 29 June; accepted 2 November 2012.

Published online 12 December 2012.

1. Verkerk, A. J. et al. Identification of a gene (FMR-1) containing a CCG repeat coincident with a breakpoint cluster region in fragile X syndrome. Cell 65, 905–914 (1991).
2. Kremer, E. J. et al. Mapping of DNA instability at the fragile X to a trinucleotide repeat sequence (CCG)n. Science 252, 1711–1714 (1991).
3. Vincent, A. et al. Human and murine FMR-1: alternative splicing and translational initiation downstream of the CGG-repeat. Nature Genet. 4, 244–251 (1993).
4. Verkerk, A. J. et al. Alternative splicing in the fragile X gene FMR1. Hum. Mol. Genet. 2, 399–404 (1993).
5. De Boule, K. et al. A point mutation in the FMR-1 gene associated with fragile X mental retardation. Nature Genet. 3, 31–35 (1993).
6. Siomi, H. Mi., Choi, M., Siomi, M.C., Nussbaum, R.L. & Dreyfuss, G. The protein product of the fragile X gene, FMR1, has characteristics of an RNA-binding protein. Cell 74, 291–298 (1993).
7. Ashley, C.T., J. Wilkinson, K.D., Reines, D. & Warren, S.T. FMR1 protein: conserved RNP family domain and selective RNA binding. Science 262, 563–566 (1993).
8. Brown, V. et al. Purified recombinant Fmrp exhibits selective RNA binding as an intrinsic property of the fragile X mental retardation protein. J. Biol. Chem. 278, 15521–15527 (1998).
9. Darnell, J. C. et al. Fragile X mental retardation protein targets G quartet mRNAs important for neuronal function. Cell 107, 489–499 (2001).
10. Ashley, C.T. et al. Human and murine FMR-1: alternative splicing and translational initiation downstream of the CGG-repeat. Nature Genet. 4, 244–251 (1993).
11. Verkerk, A. J. et al. Alternative splicing in the fragile X gene FMR1. Hum. Mol. Genet. 2, 399–404 (1993).
12. De Boule, K. et al. A point mutation in the FMR-1 gene associated with fragile X mental retardation. Nature Genet. 3, 31–35 (1993).
13. Siomi, H. Mi., Choi, M., Siomi, M.C., Nussbaum, R.L. & Dreyfuss, G. Essential role for KH domains in RNA binding: impaired RNA binding by a mutation in the KH domain of FMR1 that causes fragile X syndrome. Cell 77, 33–39 (1994).
14. Feng, Y. et al. FMRP associates with polyribosomes as an mRNA and the I304N mutation of severe fragile X syndrome abolishes this association. Mol. Cell 1, 109–118 (1997).
15. Friel, R.A., Benseler, R.E., Hua, J., Bogerd, H.P. & Cullen, B.R A nuclear role for the Fragile X mental retardation protein. EMBO J. 15, 5408–5414 (1996).
16. Ling, S.C., Fahrner, P.S., Greenough, W.T. & Geld, V.I. Transport of Drosophila fragile X mental retardation protein-containing ribonucleoprotein granules by kinesin-1 and cytoplasmic dynein. Proc. Natl Acad. Sci. USA 101, 17428–17433 (2004).
17. Davidovic, L. et al. The fragile X mental retardation protein is a molecular adapter between the neurospecific KIF3C kinesin and dendritic RNA granules. Nature Genet. 16, 3047–3058 (2007).
18. Dichter, L.B. et al. Antisense, A. L. N. Sanger, R.H. & Bassell, G.J A direct role for FMRP in activity-dependent dendritic mRNA transport links filopodial-spool spine morphogenesis to fragile X syndrome. Dev. Cell 14, 926–939 (2008).
19. Corbin, F. et al. The fragile X mental retardation protein is associated with poly(A)+ mRNA in actively translating polyribosomes. Hum. Mol. Genet. 6, 1465–1472 (1997).
20. Oostra, B.A. & Willemsen, R. Fragile X syndrome: from gene to clinic. Cell 146, 247–261 (2011). 21. Krueger, D. & Bear, M.F. Toward fulfilling the promise of molecular medicine in fragile X syndrome. Annu. Rev. Med. 62, 411–429 (2011).
METHODS

**Animals.** Wild-type and knockout FVB129P2 female mice were a gift from S. Zukin and were maintained under standard conditions. Female mice, 21–30 days old, were killed by cervical dislocation and the ovaries were aseptically removed. Dissection of the ovary from the fat pad and the bursa was done using a stereomicroscope at ×2 magnification. Excised ovaries were gently washed using PBS, weighed on an analytical balance and used either for paraffin embedding and immunohistological studies or for protein extraction. All procedures were approved by the Rockefeller University Institutional Animal Care and Use Committee (IACUC).

**Histology.** Mouse ovaries were fixed in 4% paraformaldehyde overnight at 4°C, embedded in paraffin, cut into 4-μm sections, and applied to Super frost/ Plus slides (Fisher Scientific). Haematoxylin staining was performed by the Molecular Cyto genetic core facility of Memorial Sloan-Kettering Cancer Center.

**Plasmids.** Plasmids pENTR4-FMRP isoform 7, -FXRP1 and -FXRP2 were generated by restriction enzyme digestion and ligation of the respective PCR products into pENTR4 (Invitrogen). pENTR4-FMRP isoform 1 was obtained by restriction enzyme digestion and ligation of the S25s base pair PCR product, corresponding to exon 12 and a portion of exons 11 and 13, into the pENTR4-FMRP isoform 7 plasmid. pENTR4-I304N FMRP mutant isoforms were generated by QuakChange (Agilent) according to the manufacturer’s instructions. The pENTR4-FMRP isoforms were subsequently recombined into the pPT/TO/Flag/HAGEDEST designation vectors using Gateway LR recombine in according to manufacturer’s protocol (Invitrogen). All plasmids generated in this study will be deposited at http://www Addgene.org.

**Cell lines and culture.** Generation and maintenance of stable FMR1-family proteins expressed in FlpIn TrEx HEK293 cells were lines as performed in ref. 47. Expression of FMR1-family proteins was induced by supplementing medium with 1 μg/ml doxycycline for 16 hr, prior to any analysis.

**Oligodeoxynucleotides and DNA subcloning.** The following oligodeoxynucleotides were used for PCR and subcloning (restriction sites are underlined): FMR1 isoform 7 (accession version NM_002024.4) was used at a titre of 1:3,000, in conjunction with the appropriate species antibodies raised against rabbit or mouse immunoglobulin (Dako, P0448 and MAB2160). Horseradish-peroxidase-conjugated polyclonal goat secondary antibodies against test proteins and PABPC1 were done in parallel from the lanes of control samples after loading normalization (with PABPC1).

**Immunoblotting and quantitative chemiluminescence analyses.** Prior to this procedure, each primary antibody was tested singularly. Fold differences were calculated by dividing the mean intensity of the protein of interest in experimental samples over the mean intensity of that protein of interest in Control samples after loading normalization (with PABPC1). For PAR-CLIP cdna library generation. Concurrent with the addition of doxycycline, 100 μM 4SU or 6SG was added to the cells and incubated for 16 hr, prior to 365 nm ultraviolet irradiation. Irradiation 15-μm plates were used for each PAR-CLIP experiment. Isolation of crosslinked RNAs, and sequencing of subsequently generated cdNA libraries were done essentially as described in ref. 28, except that wild-type and I304N FMRP isoform 1 and 7 PAR-CLIPs were performed with omission of the Rhase T1 digest following the immunoprecipitation in order to increase the probability for recovering G-rich binding sites.

**Recombinant baculoviral expression and purification of wild-type and I304N PAR-CLIP FMRP.** Wild-type and I304N mutant versions of pENTR4-Flag-HA-FMRP-His6 were constructed and then recombined into the pDEST8 baculoviral expression vector by Gateway LR recombine in. Baculoviral production and amplification was performed as described in the Bac-to-Bac Manual (Invitrogen). S9 cells were used for the recombinant virus production, amplification and expression of recombinant FMRP proteins. Recombinant viruses were amplified to ~2 × 10^8 plaque-forming units per ml ml. In brief, 1 × 10^8 S9 cells were infected at a multiplicity of infection (MOI) of 5 and placed back into fresh Supplemented Grace’s Insect Medium (supplemented with 10% FBS, 0.1% Pluronic-F68, 100 μg/ml penicillin, 100 μg/ml streptomycin) into a spinner flask at a final concentration of 1 × 10^6 cells per ml medium for 3–4 days prior to collection. Cells were lysed in loading buffer (50 mM Tris–HCl, pH 8.0, 1 M KCl, 5 mM MgCl2, 5 mM imidazole, 10% glycerol, 0.1% Triton X-100, 1 μM β-mercaptoethanol, 1× EDTA-free Protease Inhibitor Cocktail (Roche)), centrifuged (19,000g, 10 min, 4°C) and resulting supernatants filtered through a 5-μm membrane ( Pall) prior to loading onto an X 16 column (GE Life Science) packed with 10 ml of cobalt immobilized metal affinity chromatography resin (Clontech) pre-equilibrated in loading buffer and attached to an AKTA Explorer FPLC. Six column volume washes (loading buffer with 13 mM imidazole) were passed through the column and then step eluted (loading buffer with 300 mM imidazole), collecting 1-ml-sized fractions. The peak of fractions containing FMRP were determined by SDS–PAGE and Coomassie-staining of the gel. The peak fractions were pooled and dialysed overnight into buffer containing 20 mM Tris–HCl, pH 7.65, at 25°C, 300 mM KCl, 5 mM MgCl2, 50% glycerol, 0.1% Triton X-100, 1 mM DTT. Aliquots of FMRP were stored at – 80°C. One litre of baculoviral-infected S9 cells typically yielded 1 mg of pure FMRP.

**[γ-32P]ATP 5 s radiolabeling of oligoribonucleotides.** Synthetic oligoribonucleotides (10 pmol, 18–60-nucleotide size ranges and listed in Supplementary Tables 4 and 5) were combined with 5 pmol (γ-32P)ATP (6,000 Ci mmol−1) in a 10-μl reaction buffer containing 70 mM Tris–HCl, pH 7.6, at 25°C, 10 mM MgCl2, 30% glycerol, 10% DMSO, 0.1 mg/ml BSA, 200 mM KCl, 5 mM MgCl2, 10 mM dithiothreitol (DTT), 100 mg/ml BSA (Ambion) and incubated at 37°C. The mixture was placed onto ice. 5′ T4 polynucleotide kinase (NEB) was then added and incubated for 15 min at 37°C. Thereafter, non-radiolabelled ATP was added to a final concentration of 1 mM and incubated at 37°C for an additional 5 min at 37°C. Following addition of 40 μl H2O, samples were incubated at 95°C for 30 s, then loaded onto Microspin G25 columns (GE Life Science) and centrifuged at 730g for 1 min, according to manufacturer’s instructions. The volume of the eluate was increased to 100 μl by the addition of water for a final concentration of 100 mM and stored at –20°C.

**EMSAs.** 1 nM radiolabelled RNA was combined with FMRP varying in concentration from 0–1 μM in 20 μl EMSA buffer (20 mM Tris–HCl, pH 7.6, 300 mM KCl, 5 mM MgCl2, 35% glycerol, 1 mM DTT, 0.1 μg/ml RNAse (Promega), 100 ng/ml transfer RNA, 0.1 mg/ml acetylated BSA (Ambion)) and incubated at 30°C for 1 h in 1.5 ml passivated (50 μg/ml acetylated BSA, 27°C, 1 h) and siliconized microcentrifuge tubes. After 1 h, 5 μl of EMSA loading buffer (50% glycerol, bromophenol blue in EMSA buffer) was added. Alternatively,
for supershift assays, 1 μg antibody was added to each reaction and incubated for an additional 30 min prior to addition of EMSA loading buffer. Reactions were separated by native PAGE using 6% or 10% (bottom)/6% (top) step-gradient gels (24 × 15 cm). Polyacrylamide gels were polymerized using 49:1 acrylamide:bisacrylamide in Tris-glycine buffer (25 mM Tris base, 0.2 M Glycine), and pre-run for 30 min at 4 °C at 300 V in the same buffer. Reactions were loaded and the species were separated for 2 h at 300 V at 4 °C. The 32P radioactive signal was detected using phosphorimager screens (1 h to overnight exposure) and the signal quantified using ImageGauge software. Curves and binding constants were calculated using Kaleidagraph v4.03 (Synergy) software.

**Pre-processing reads.** Details describing specific pre-processing parameters are available at http://fmrp.rockefeller.edu. Raw Illumina reads were stripped of 3′ adapters using the FASTX-Toolkit (http://hannonlab.cshl.edu/fastx_toolkit/). Reads shorter than 13 nucleotides or containing an ambiguous nucleotide were discarded. Processed reads were aligned to the reference genome (GRCh37/hg19) by the Bowtie algorithm (0.12.7), allowing for two alignment errors. For each read, only the best mismatch-stratum was reported for up to ten different locations; T-C mismatches with the genomic sequence were subtracted from the mismatch count for each of the mapped locations. After the conversion subtraction, reads that mapped to only one genomic location were retained for downstream analysis.

**PARalyzer and annotation.** For each library, PARalyzer was used to identify binding sites as described previously in ref. 34. See FMR.ini file at fmrp.rockefeller.edu for details about exact parameters. PARalyzer-derived clusters were annotated using an in-house software annotator, using annotation data from the following sources (all tables available upon request): (1) protein-coding transcript: lincRNA, misc_RNA, Mt_rRNA, Mt_trRNA, RNA, smRNA and snRNA annotation were downloaded from biomart/Ensembl version GRCh37 (http://jun2011.archive.ensembl.org/biomart/martview); (2) repetitive element annotation was downloaded from the RepeatMasker track of UCSC genome table browser (http://genome.ucsc.edu/); Repeat elements of the class ribosomal RNA, small nuclear RNA and tRNA were combined with Ensembl annotation for those categories described above; (3) piwi-interacting RNA annotation was downloaded from functional RNAdb version 3.4 (http://www.ncrna.org/FRNadb/files/hg18_grn.zip) and converted to hg19 using the liftOver utility; (4) microRNA annotation was downloaded from miRBase Sequence version 17 (http://www.mirbase.org/). Binding sites overlapping multiple annotation categories were assigned a single annotation according to the FMR.yaml file available at fmrp.rockefeller.edu (higher number given more priority).

**cERMIT motif and RRE hierarchical analyses.** We used cERMIT41 to identify motifs enriched in clusters exhibiting higher T-to-C conversion evidence (for detailed parameters see cERMIT template file at fmrp.rockefeller.edu). Unlike many motif-discovery tools, cERMIT used the full data set to discover motifs with high evidence scores. We ranked binding sites mapping to 3′ UTR, 5′ UTR, coding or intron annotation categories by the log2(T-to-C conversion events). Motifs found by cERMIT were filtered for low-complexity motifs (motifs for which any single nucleotide was present in 80% or more locations). To define core RREs, the position weight matrices of the remaining motifs were clustered and aligned using a modified version of the STAMP48 tool that only considers the forward strand (available upon request).

**Wild-type versus I304N comparative analysis.** Because isoform 1 and 7 I304N libraries contained substantially more reads, and consequently binding sites, than their respective wild-type libraries, we randomly sub-sampled reads from the I304N libraries such that each subset would have at least as many and no greater than ~10% more binding sites as the wild-type library. To compare RRE read depth between wild-type and I304N libraries, we first identified the genomic coordinates and underlying read depth for each instance of either ACUK or WGGA within a binding site from the wild-type library. Next, we determined the read-depth ratio comparing the wild-type RRE read depth with the read depth of the wild-type RRE coordinates in each I304N subset. We calculated the average of the I304N versus wild-type read depth ratio for each instance of an RRE in wild-type binding sites labelled as ACUK, mixed and WGGA.

**RIP-chip and analyses.** RIP-chip using Human Genome U133 Plus 2.0 (Affymetrix) was performed essentially as described in ref. 49. Fifty 15-cm plates of doxycycline-induced (1 μg ml−1, 16–20 h) stable cells expressing Flag-HA–FMRP isoform 1 were used. Cleared supernatants were split into two equal fractions for technical replicates. 1/100 of each replicate volume was saved for total RNA isolation. The remaining lysate was incubated with magnetic beads pre-conjugated with anti-Flag M2 antibodies (10 μl conjugated beads per 1 ml of lysate). After 1-h incubation at 4 °C, beads were washed three times with IP wash buffer then re-suspended in 1 bead volume of ×1 proteinase K buffer containing 0.6 mg m−1 proteinase K, and incubated at 65 °C for 20 min. RNA was then isolated by phenol–chloroform extraction then ethanol precipitation, and finally dissolved in 6 μl H2O. RNA concentration was determined using a NanoDrop UV–Vis spectrophotometer. RNA from the pre-immunoprecipitated aliquots were Trizol extracted, then similarly prepped using the RNeasy Mini Kit (Invitrogen) was used according to manufacturer’s instructions and recovered RNA was ethanol-purified and finally dissolved in 6 μl H2O. RNA concentration was determined using a NanoDrop UV–Vis spectrophotometer. RNA from the pre-immunoprecipitated aliquots were Trizol extracted, then similarly prepped using the RNeasy Mini Kit. Robust multi-array average (RMA) background correction and robust quantile normalization was applied to RIP-chip microarray data using the AffyR package42. After filtering for detection, the median of all probes mapping to a gene was used to calculate gene-level expression values. A log2-fold enrichment (LFE) was calculated by subtracting the log2 (RIP gene-level expression) from the log2 (lysate gene-level expression). These LFE values were binned on the basis of properties determined by PAR-CLIP experiments and plotted in MATLAB using the empirical cumulative distribution function. For analyses involving ACUK and WGGA RREs, genes containing low-complexity motifs were excluded to determine the sole contributions of ACUK and WGGA RREs. PARalyzer-defined binding sites overlapping repetitive elements were not used for summarizing gene-level site information. The significance of enrichment for each bin of data in comparison to the total set of enrichments was determined by the Kolmogorov–Smirnov test. Bonferroni correction was applied to account for multiple comparisons.

47. Hoell, J. et al. in Handbook of RNA Biochemistry Vol. 1 (eds Hartmann, R.K., Binderief, A., Schön, A. & Westhof, E.) (Wiley-VCH, 2013).
48. Mahony, S., Auron, P. E. & Benos, P. V. DNA familial binding profiles made easy: comparisons of various motif alignment and clustering strategies. PLoS Comp. Biol. 3, e61 (2007).
49. Landthaler, M. et al. Molecular characterization of human Argonaute-containing ribonucleoprotein complexes and their bound target mRNAs. RNA 14, 2580–2596 (2008).
50. Gautier, L., Cope, L., Bolstad, B. M. & Irizarry, R. A. affy-analysis of Affymetrix GeneChip data at the probe level. Bioinformatics 20, 307–315 (2004).