ARTICLES

HITS-CLIP yields genome-wide insights into brain alternative RNA processing

Donny D. Licatalosi[1], Aldo Mele[1], John J. Fak[1], Jernej Ule[3], Melis Kayikci[3], Sung Wook Chi[1], Tyson A. Clark[4], Anthony C. Schweitzer[4], John E. Blume[4], Xuning Wang[2], Jennifer C. Darnell[1] & Robert B. Darnell[1]

Protein–RNA interactions have critical roles in all aspects of gene expression. However, applying biochemical methods to understand such interactions in living tissues has been challenging. Here we develop a genome-wide means of mapping protein–RNA binding sites *in vivo*, by high-throughput sequencing of RNA isolated by crosslinking immunoprecipitation (HITS-CLIP). HITS-CLIP analysis of the neuron-specific splicing factor Nova revealed extremely reproducible RNA-binding maps in multiple mouse brains. These maps provide genome-wide *in vivo* biochemical footprints confirming the previous prediction that the position of Nova binding determines the outcome of alternative splicing; moreover, they are sufficiently powerful to predict Nova action *de novo*. HITS-CLIP revealed a large number of Nova–RNA interactions in 3′ untranslated regions, leading to the discovery that Nova regulates alternative polyadenylation in the brain. HITS-CLIP, therefore, provides a robust, unbiased means to identify functional protein–RNA interactions *in vivo*.

The discovery of RNA molecules with catalytic activity[1] led to the hypothesis that, from the earliest life forms, RNA regulation evolved to have critical roles in living organisms[2–5]. Efforts to develop comprehensive understanding of protein–RNA interactions *in vivo* have combined genetics, bioinformatics, microarray profiling and biochemical approaches. However, the latter have been hampered by methodological problems[6,7]; for example, co-immunoprecipitations can lead to re-association of protein–RNA complexes *in vitro*[8], non-specifically bound RNAs and additional co-precipitating RNA-binding proteins (RNABPs)[9].

We have taken a different approach towards understanding protein–RNA interactions by developing a crosslinking protocol that works in tissues, and can therefore be applied before protein purification. This method, termed CLIP[10,11], uses ultraviolet irradiation to induce covalent crosslinks between protein–RNA complexes *in situ*, allowing rigorous purification of RNABPs along with small fragments of RNA, which can be amplified and sequenced. CLIP has been used to study direct protein–RNA interactions extant in living cells[11–13], including identification of RNA targets[11] for the KH-type RNABP Nova[14,15], and the discovery of hnRNPA1-dependent regulation of a microRNA[12].

Genome-wide efforts to understand Nova function, using exon junction microarrays and bioinformatic analysis of Nova binding sites (YCAY clusters[15], characterized biochemically[16] and crystallographically[17]), suggested that the position of Nova binding to pre-mRNA predicted its action to enhance or inhibit alternate exon inclusion[18]. To identify direct Nova–RNA interactions *in vivo*, we applied high-throughput sequencing methods to CLIP. Here we demonstrate that this approach uncovers new biology in the brain, identifying functional interactions that mediate tissue-specific alternative RNA processing.

Genome-wide protein–RNA interaction maps

We studied Nova–RNA interactions in the mouse neocortex, which expresses the Nova2 protein[19]. We identified 2,481 Nova-bound RNAs (CLIP tags)[10,11] from five experiments using traditional CLIP strategies[10,11] and 412,686 CLIP tags from three experiments using high-throughput pyrosequencing. Tags were filtered to eliminate those with imperfect (<80%) matches to genomic sequences, those with multiple genomic hits or those that were exact duplicates. The resulting set of 168,632 unique tags (≥92% bound to Nova2, and the remainder to either Nova1 or Nova2) included 123,734 tags mapping to messenger-RNA-encoding genes and 44,898 tags mapping to intergenic regions (Supplementary Fig. 2).

As negative controls, we repeated Nova CLIP using *Nova2* knockout brain or an irrelevant antibody, but were unable to amplify PCR products. We also sequenced 43,000 crosslinked RNA tags remaining after Nova immunoprecipitation, corresponding to a sample of all remaining RNABP–RNA interactions in the brain, and compared the frequency of Nova binding sites[15] in these control tags with that in Nova HITS-CLIP tags. Only the latter showed enrichment for YCAY sequences (observed:expected YCAY frequency was 3.56 for Nova tags, compared to 0.99 for the control tags, determined by Chi-squared distribution, $P < 10^{-227}$; see Methods), demonstrating the specificity of CLIP.

HITS-CLIP reproducibility and cluster analysis

We reasoned that one way to distinguish between biologically robust and transient Nova–RNA interactions would be to assess the reproducibility with which HITS-CLIP tags were identified in individual mice. Tags obtained from the neocortex of two P13 littermates showed a remarkable degree of similarity; when equal numbers were aligned across the entire mouse genome in 10-kilobase (kb) windows, a high correlation was evident both graphically (Fig. 1a) and statistically ($R^2 = 0.75$). To assess reproducibility more accurately, we focused on sites containing overlapping tags ('clusters'). A total of 19,156 clusters had at least 2 tags, 508 had 20 or more tags (Supplementary Fig. 3a), and 608 RefSeq transcripts had ≥6 clusters (Supplementary Fig. 2). Inter-animal clusters—sites containing at least one tag from each littermate—were highly reproducible; over 90% (9,697 out of 10,740) of sites containing tags from one animal also had at least one littermate tag. Finally, inspection of individual chromosomes and genes revealed that Nova clusters were highly

[1]Laboratory of Molecular Neuro-Oncology and Howard Hughes Medical Institute, [2]Biocomputing, Information Technology, The Rockefeller University, 1230 York Avenue, New York, New York 10021, USA. [3]MRC Laboratory of Molecular Biology, Cambridge, CB2 0QH, UK. [4]Expression Research, Affymetrix, Inc., Santa Clara, California 95051, USA.

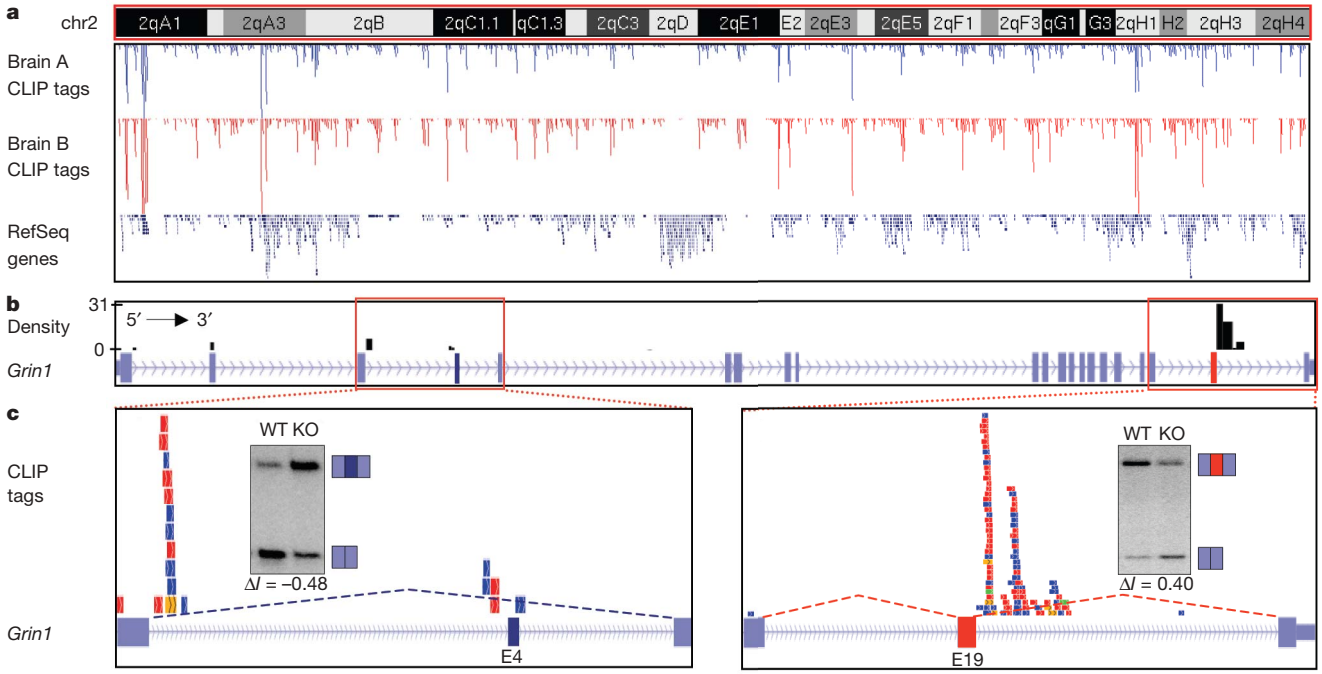

Figure 1 | HITS-CLIP genome-wide map of Nova–RNA binding sites. **a**, Chromosome 2 (chr2) RefSeq genes and CLIP tags from the neocortex of two mouse littermates (brain A, 46,106 tags in 10,740 clusters; brain B, 100,874 tags in 15,805 clusters). **b**, Cluster density (number of tags per cluster length; black) in the *Grin1* transcript. **c**, *Grin1* E4 (left) and E19 (right) tags, one colour per biological replicate, predict Nova-dependent exon skipping and inclusion, respectively (experimental validation[20] is shown). KO, knockout; WT, wild type.

reproducible in both position and extent of crosslinking, and were specific to a subset of brain-expressed RefSeq genes (Fig. 1 and Supplementary Fig. 2).

To determine how faithfully Nova CLIP-tag clusters reflect previously defined consensus Nova binding sites[15], we analysed them for consensus motifs by MEME analysis, and found they were significant enriched in YCAY motifs (Supplementary Fig. 3). This was evident across all 19,156 Nova CLIP-tag clusters (3.9-fold; $P < 10^{-227}$), and in tags associated with functional Nova interactions (see below). Taken together, these observations indicate that HITS–CLIP reproducibly identifies discrete, YCAY-rich, Nova binding sites in mouse brain RNAs, and suggests that these binding sites may point to positions of functional Nova–RNA interactions.

HITS-CLIP analysis of splicing targets

HITS-CLIP offered a chance to compare predicted sites of Nova–RNA regulation derived from bioinformatic and microarray analysis[11,18,20] with interaction sites observed by *in vivo* crosslinking. A total of 39 previously validated[20] Nova2-regulated transcripts harboured Nova CLIP tags (ranging from 1 to 96 tags) within 3 kb of the alternative exon local region (bounded by the constitutive splice donor and acceptor exons) and 34 of these harboured CLIP-tag clusters. The position and YCAY content (4.1-fold enrichment; $P < 10^{-156}$) of these clusters was consistent with the predicted Nova bioinformatic map[18]. For example, YCAY-rich HITS-CLIP clusters were present downstream of the known Nova2 target *Grin1* exon 19 (E19; Fig. 1b, c (right))[20], in a position previously predicted by the Nova bioinformatic map[18] (Supplementary Fig. 4).

We also observed HITS-CLIP tags in *Grin1* upstream of an alternative exon (exon 4; E4) that was not a previously known Nova target. The position of these tags predicted Nova-dependent inhibition of E4 inclusion, which was confirmed experimentally (Fig. 1b, c (left)), suggesting that HITS-CLIP might provide a general means to identify new sites of protein–RNA regulation. Six additional transcripts with Nova HITS-CLIP clusters near regulated splice sites were tested; each was aberrantly spliced in *Nova2* knockout compared to wild-type brain in a manner conforming to the Nova bioinformatic map (Supplementary Fig. 5).

To assess further how the position of Nova binding is related to the outcome of such splicing events, we analysed Nova HITS-CLIP tags in Nova-regulated exons newly identified using an updated version of exon-junction microarrays[20] harbouring probe sets for exon junctions in ~145,000 transcripts. Arrays were interrogated with RNA from wild-type or *Nova2* null neocortex, and results analysed with ASPIRE2, a revision of the ASPIRE algorithm[20] that searches for reciprocal changes in exon-included and exon-excluded probe sets. We identified 32 out of 45 previously validated[20] Nova2-dependent exons, as well as 46 new candidates with $|\Delta I|$ (the change in fraction of alternative exon inclusion[20]) values ranging from 0.19 to 0.60 and with characteristics seen previously[20] (Supplementary Fig. 6 and Supplementary Tables 1 and 2). To simplify subsequent analysis, we focused on 35 cassette exons, and confirmed that alternative splicing was Nova2-dependent in 7 out of 7 cassette exons assayed (Supplementary Fig. 4).

We generated a map in which we placed all 1,085 Nova CLIP tags identified from a total of 71 Nova2-regulated cassette exons (43 validated targets, and 28 newly predicted targets with $\Delta I > 0.2$ and ΔI-*t*-test > 25; see Methods) onto a single composite pre-mRNA (Fig. 2a and Supplementary Fig. 7). These tags spanned 11.5 kb, but were very heavily concentrated around splice sites, in positions that corresponded extremely well with the bioinformatically predicted Nova map[18], and with previous biochemical analysis of Nova-dependent splicing[21–23] (Fig. 2a). Furthermore, clusters in these regions showed a 3.4-fold enrichment in YCAY elements ($P < 10^{-174}$), with 72 of 123 clusters containing at least 3 YCAY elements within 30 nucleotides (nt), consistent with previous biochemical data[21–23].

We also noted some HITS-CLIP tags in unanticipated regions. For example, we observed frequent binding of Nova in intronic sequences upstream of Nova-regulated exons. However, binding to these sites was only robust in a limited number of transcripts (Fig. 2a and Supplementary Fig. 7). To generate a map representative of

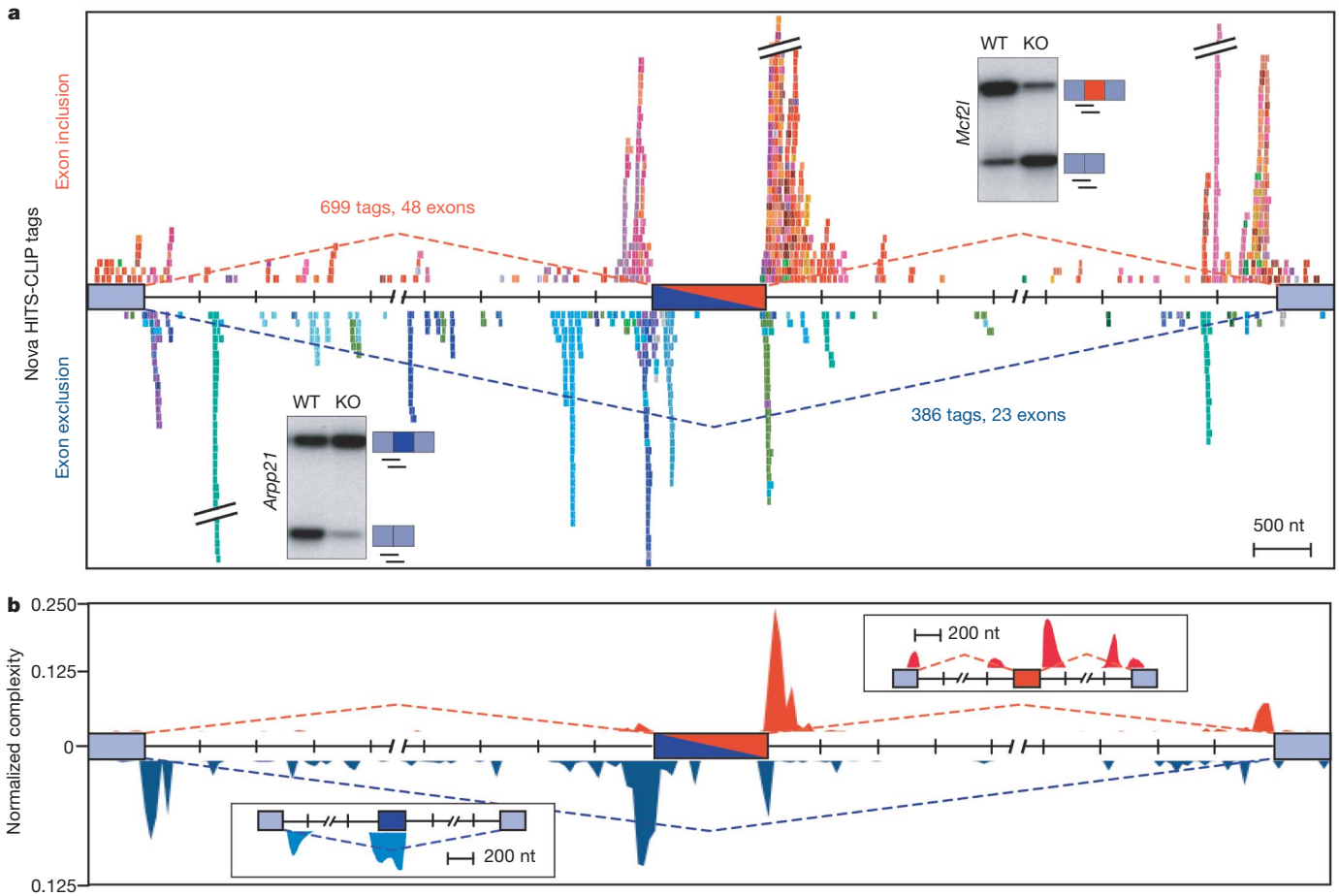

Figure 2 | Nova–RNA interaction maps associated with Nova-dependent splicing regulation. **a**, CLIP tags around all known Nova-regulated cassette exons, with one colour per transcript. Tags were mapped onto a composite transcript containing an alternative (dark blue/red box) and flanking constitutive (light blue box) exons. Tags are from transcripts showing Nova-dependent exon inclusion (top panel) or exclusion (bottom panel); representative examples of experimentally validated target RNAs (*Arpp21* and *Mcf2l*) are shown (insets). **b**, Normalized complexity map (see Methods) of Nova–RNA interactions recapitulate predicted maps[18] (insets) for Nova-dependent exon inclusion (red) or exclusion (blue).

consensus Nova action, we normalized our data, first to the number and distribution of CLIP tags between transcripts, and then to the number of different transcripts with tags at a given position (complexity). This allowed us to focus on potential regulatory binding sites common to several transcripts. This 'normalized complexity' map (Fig. 2b) demonstrated that Nova CLIP tags corresponded very precisely to the bioinformatically predicted sites of Nova action (Fig. 2b, insets). We conclude that HITS-CLIP confirms the hypothesis that Nova binding occurs directly on YCAY-rich elements near splice sites *in vivo*, and that the position of such Nova binding determines the outcome of Nova-dependent splicing regulation.

Nova regulates alternative polyadenylation

We next explored whether other HITS-CLIP clusters might reveal new Nova functions. Analysis of the genomic position of Nova clusters revealed that 23% of Nova HITS-CLIP tags mapped to intergenic regions (Fig. 3a). To examine the possibility that these tags may correspond to previously undescribed isoforms of RefSeq genes with alternative terminal exons, we examined the distance between intergenic clusters and neighbouring RefSeq genes. There was an exponential increase in the cumulative number of tags within 10 kb downstream of known stop codons, compared to a linear increase within 10 kb upstream of start codons (717 versus 101 clusters were present within 10 kb of the stop or start codon, respectively; Fig. 3b). This suggests that, in addition to binding known 3′ untranslated regions (UTRs; Fig. 3a), Nova binds to unannotated 3′ UTR extensions of known genes. Within 3′ UTRs, tags were enriched near poly(A) sites, and to a lesser degree near stop codons (Fig. 3c). A large number of clusters were positioned within a few hundred nucleotides of poly(A) sites (Fig. 3d)—a region that contains core and potential auxiliary elements controlling transcript termination and poly(A) site use[24,25].

These observations suggested that Nova might function in a second pre-mRNA processing event in the mouse brain—regulated poly(A) site use (alternative polyadenylation)—a process about which little is known. We analysed alternative polyadenylation by hybridizing Affymetrix exon arrays with *Nova2* wild-type versus knockout brain RNA, and screened for changes in alternate 3′ UTR relative to total mRNA abundance (Supplementary Fig. 8). We identified 297 transcripts with such differences ($\geq$1.5-fold; $P < 0.05$); 43 contained 100 3′ UTR CLIP-tag clusters, and these were preferentially present near poly(A) sites (Fig. 3d).

We tested poly(A) site use in two candidates, *Cugbp2* and *Slc8a1*. Both have microarray-predicted Nova-dependent changes in 3′ UTR usage (1.5- and 2-fold, respectively), and both contained CLIP tags near poly(A) sites (Fig. 4a and Supplementary Fig. 9). RNase protection analysis (RPA) demonstrated that use of these poly(A) sites was increased in *Nova2* knockout brain (Fig. 4a, e and Supplementary Fig. 9); ΔC (the change in percentage of transcripts cleaved at the relevant poly(A) site, analogous to ΔI)[18] for these transcripts was 0.22–0.25 (for example, 41% to 66% use of poly(A) site 2 (pA2) in *Cugbp2* transcripts in wild-type versus *Nova2* knockout brain; Fig. 4a), comparable in magnitude to Nova-dependent changes in alternative exon usage. Furthermore, the increase in proximal poly(A) use in *Cugbp2*

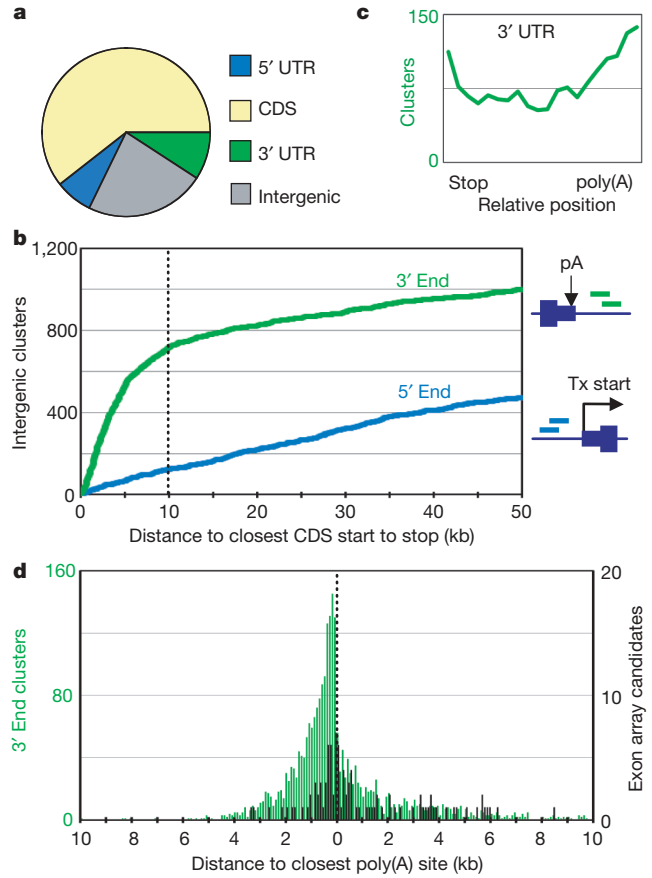

Figure 3 | Nova CLIP tags cluster near polyadenylation sites. **a**, Distribution of Nova CLIP-tag clusters (RefSeq mm8; CDS refers to introns/exons between UTRs). **b**, Intergenic clusters were plotted relative to the closest annotated start (5′ end) or stop (3′ end) codon. pA, poly(A) site; Tx, transcription. **c**, Distribution of 2,465 clusters in 3′ UTRs (defined as distance from stop codon to closest transcript termination site). **d**, CLIP tag clusters relative to the closest poly(A) site (transcript end in UCSC known genes), plotted as clusters per 50 nt in 2,465 3′ UTRs (green bars) or in 43 Nova-dependent 3′ UTRs identified with exon arrays (black bars).

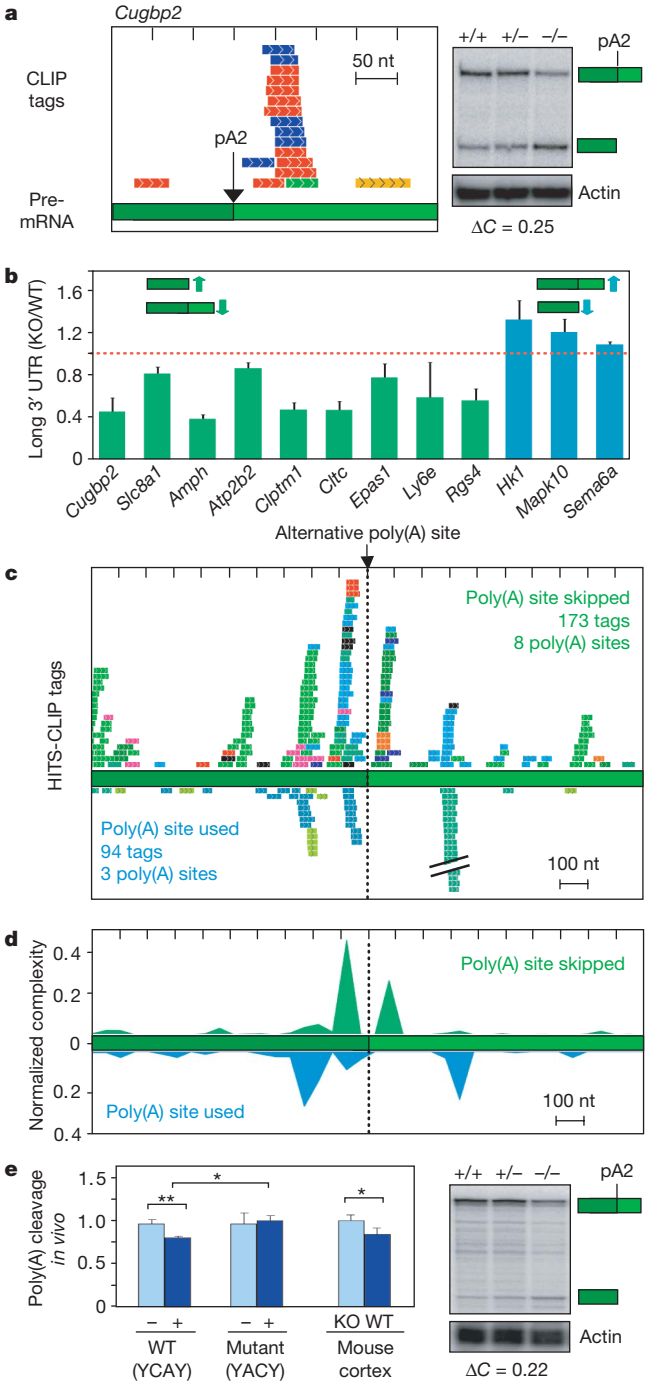

Figure 4 | Nova regulates alternative polyadenylation. **a**, Nova CLIP tags near *Cugbp2* pA2; each colour represents a biological replicate. Right: RPA measuring cleavage at pA2 in wild-type (WT, +/+), *Nova2* heterozygous (+/−) and knockout (KO, −/−) P10 neocortex, with actin as a normalization control and ΔC (see text) shown. **b**, qRT–PCR validation of Nova-dependent poly(A) regulation, using RNA from three WT or *Nova2* KO littermates, presented as long 3′ UTR isoform relative to total transcript abundance, normalized to WT (1.0; dotted line); error bars represent standard deviation. **c**, **d**, Composite (**c**) and normalized complexity (**d**) maps for Nova-regulated alternative polyadenylation sites (as in Fig. 2). **e**, Nova binds YCAY elements to directly regulate alternative polyadenylation. Left: qRT–PCR analysis of cells transfected with WT or mutant (three YACY mutations) poly(A) reporter constructs co-transfected with a control (−) or Nova2 expressing (+) plasmid, or WT versus *Nova2* KO neocortex ($*P < 0.02$, $**P < 0.01$); error bars represent standard deviation. Right panel: RPA for cleavage of *Slc8a1* pA2, otherwise as described in **a**.

and *Slc8a1* transcripts in *Nova2* knockout brain was associated with reciprocal decreases in processing at distal poly(A) sites, suggesting that changes in the relative levels of alternatively polyadenylated *Cugbp2* and *Scl8a1* mRNAs are not due to differences in isoform stability, but result directly from aberrant poly(A) site use in the *Nova2* knockout.

We used quantitative polymerase chain reaction with reverse transcription (qRT–PCR) to measure the relative abundance of alternative poly(A) isoforms of 29 additional candidate Nova targets (from Fig. 3d). Twelve transcripts had significant changes in levels of alternatively polyadenylated transcripts ($P < 0.04$ in 11 out of 12; Fig. 4b). These transcripts did not change in overall transcript abundance (data not shown); nine were consistent with a Nova-dependent action to block, and three to enhance, use of the adjacent poly(A) site. Seventeen transcripts had either no change in poly(A) site usage in *Nova2* knockout brain (most of these were low abundance (at least in alternate 3′ isoforms)), and/or had confounding changes in overall steady-state transcript levels, and two transcripts had 3′ UTR changes as well as alternate splicing of terminal exons. We mapped Nova CLIP tags from the 12 Nova-regulated 3′ UTRs onto a composite transcript containing an alternative polyadenylation site (Fig. 4c), and onto a normalized complexity Nova–RNA 3′ UTR interaction map (Fig. 4d). This revealed reproducible Nova binding to discrete YCAY-rich (3.5-fold, $P < 10^{-227}$) regions flanking Nova-regulated alternative poly(A) sites. Taken together, the quantitative analysis of transcript levels and the HITS-CLIP map demonstrate that Nova

binds to YCAY-rich elements flanking poly(A) sites and is necessary for their proper regulation in mouse brain.

To test whether Nova binding to these 3′ UTRs is sufficient to suppress poly(A) site use, we generated a green fluorescent protein (GFP) reporter construct containing parts of the *Slc8a1* 3′ UTR harbouring alternative poly(A) sites (Supplementary Fig. 9b), as well as mutant constructs in which YCAY elements were mutated to YACY (a sequence to which Nova does not bind[15]). Co-transfection of Nova2-expressing constructs with these reporters into 293T cells (which do not express Nova[22]) demonstrated a Nova- and YCAY-dependent reduction in alternative poly(A) site usage of the same magnitude and direction as seen in wild-type versus *Nova2* knockout neocortex (Fig. 4e). Taken together, these results identify direct Nova–RNA interactions in the 3′ UTR that regulate brain-specific alternative RNA processing.

Discussion

Genome-wide screens have been used to establish correlations between the action of RNABPs and biological diversity[6,7,26–30], but are unable to identify direct sites of RNA regulation. HITS-CLIP provides a general solution to this problem by generating a transcriptome-wide biochemical 'footprint' of protein–RNA interactions in living tissues. This in turn allows a direct comparison of predicted (for example, microarray or bioinformatically derived) and observed (HITS-CLIP) sites of action, and thereby provides a new platform for deriving functional RNABP maps and for assessing models of protein–RNA regulation.

HITS-CLIP extends our transcriptome-wide understanding of Nova–RNA interactions, which was previously limited to bioinformatic analysis of YCAY clusters within 200 nt of alternate or bounding constitutive exons[18]. Analysis of HITS-CLIP tags mapping to 71 Nova-regulated exons (Fig. 2) yielded a more refined map of Nova action. Over 91% of the normalized Nova binding associated with exon inclusion (Fig. 2b) occurred within 500 nt of either the alternative 5′ or the constitutive 3′ splice sites, whereas 74% of the normalized Nova binding associated with exon exclusion (Fig. 2b) occurred within 500 nt of the constitutive 5′ splice donor or surrounding the alternate exon. This strengthens the conclusion that the position of Nova–RNA interaction determines the outcome of splicing, an observation that may extend to splicing factors more generally[31]. Importantly, the strength of these correlations suggests that the HITS-CLIP map is sufficiently robust to predict protein–RNA regulation, as shown for seven new Nova splicing targets (Fig. 1d and Supplementary Fig. 5).

Whereas most of the Nova-regulated sites conform to a general set of rules based on direct Nova binding, there are also clear exceptions. For example, Nova binds robustly but in atypical positions in several regulated transcripts (for example, *Brsk2* and *Rap1gap*; Fig. 2a, b and Supplementary Fig. 7). Such examples may point to new mechanisms of Nova action, which may include interactions with other RNABPs. For example, Ptbp2 interacts with Nova and other RNABPs such as KSRP (also known as Khsrp) to modulate the outcome of alternative splicing[32,33]. In addition, RNA structure may regulate or be affected by interaction with other factors, as suggested by the ability of the splicing factor Mbnl1 to stabilize RNA hairpins[34], from analysis of splicing defects in *Mapt* that underlie frontotemporal dementia with Parkinsonism[35], and from structural studies of competition between hnRNP F and Ptbp1 binding to the *c-src* transcript (F. Allain, personal communication).

The unbiased nature of HITS-CLIP led to the unexpected identification of Nova binding near poly(A) sites and the recognition of its role in regulating alternative polyadenylation in the brain. The presence of such tissue-specific factors was postulated after the recognition of differential polyadenylation of immunoglobulin M heavy chain transcripts in B cells[36] and of *Calca* (calcitonin/calcitonin-related polypeptide) pre-mRNA in neurons[37]. Alternative poly(A) sites are present in ~50% of human genes[38], and their regulation is believed to have an important role in tissue and developmental mRNA regulation[39,40] as well as in human disease[41]. In particular, brain mRNAs seem to be preferentially processed at promoter-distal poly(A) sites to generate long 3′ UTRs[42,43]. Interestingly, in 9 of 12 instances examined (Fig. 4b), Nova promoted the production of mRNAs with long 3′ UTRs. Thus, one important action of Nova may be to generate long 3′ UTRs in neurons, which may be subject to regulation by microRNAs or other RNABPs.

Numerous links have been made between pre-mRNA splicing and 3′-end processing[44,45], such as the observation that the splicing factor sex-lethal can regulate polyadenylation by competing with Cstf64 (also known as CstF2) for RNA binding[46]. Although we found two transcripts that had both Nova-dependent changes in splicing and polyadenylation, Nova can mediate splicing-independent alternative polyadenylation. For example, Nova suppresses the *Slc8a1* pA2 site in an intron-less transcript (Fig. 4), and *Cugbp2* and *Slc8a1* alternative polyadenylation was not coupled to alternative splicing in brain (unpublished data).

The Nova HITS-CLIP map offers insight into the mechanism of poly(A) site selection in the brain. Changes in the accessibility of core (for example, CPSF and CstF) or auxiliary factors to interact with *cis* elements surrounding the poly(A) site underlie the regulation of alternative polyadenylation[24,25,47,48]. We find no evidence that Nova regulates transcripts encoding such factors (including subunits of CPSF, CstF, CFI and CFII). Instead, our data point to Nova as a *trans*-acting factor that binds YCAY elements flanking regulated poly(A) sites, and that the position of Nova binding may determine whether it acts to promote or inhibit poly(A) site use (Fig. 4). For example, Nova CLIP tags overlap the canonical CPSF and/or CstF binding sites within 30 nt of the *Cugbp2* and *Slc8a1* poly(A) sites, which are suppressed by Nova. In contrast, in transcripts in which Nova enhances poly(A) site use, it binds to more distal elements, where it may antagonize the action of auxiliary factors. Therefore the position of Nova 3′ UTR binding may determine the outcome of poly(A) site selection in a manner analogous to its action on splicing regulation (Supplementary Fig. 1).

In summary, HITS-CLIP offers a powerful new platform for studying RNA regulation *in vivo*. This genome-wide biochemical approach complements bioinformatic, microarray and genetic studies. HITS-CLIP is able to identify biologically relevant interactions, providing a focus on direct protein–RNA contacts as critical points for understanding RNABP function. The unbiased nature of the platform holds the potential for new discovery, including the elucidation of preferred binding sequences and the identification of regulated RNA substrates. Identifying Nova as the first vertebrate factor to regulate alternative polyadenylation in mouse brain demonstrates that a single factor can regulate different aspects of tissue-specific RNA metabolism. Finally, the reproducible nature of HITS-CLIP suggests that it provides a robust platform to explore RNABP-dependent mechanisms of gene expression in complex and dynamic scenarios.

METHODS SUMMARY

HITS-CLIP. CLIP was performed on mouse *Nova2* wild-type and knockout (CD1) brains as described[11]. After PCR amplification, high-throughput sequencing was performed (454 Life Sciences).

Microarrays. For analysis of Nova-dependent alternate splice and alternate 3′ UTR variants, a custom exon junction array (Affymetrix) or MoEx 1.0 ST Affymetrix exon arrays, respectively, were used.

Bioinformatics. CLIP tags and clusters were analysed with BED or WIG formatted custom tracks using the UCSC Genome Browser and Genome Graph tools (http://genome.ucsc.edu). Composite maps were generated by determining the distance between tags and closest splice sites within the alternative exon local region and converted to coordinates in a BED format custom track, with tags from each gene assigned different colours. MEME sequence analysis was done using tools available at http://meme.sdsc.edu. ASPIRE2 was based on ASPIRE (ref. 20).

Biochemical and transfection assays. Biochemical assays were done using biological triplicate sibling mice, unless otherwise noted. RPAIII kits from Ambion were used, and RT–PCR experiments were done as described[21–23], with modifications described in Methods; qPCR was done with a MyIQ BioRad thermal cycler and data analysed as described in Methods. Wild-type or mutant GFP

alternative polyadenylation reporters were transfected into 293T cells in the presence or absence of pNova2 (described in Methods).

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 5 May; accepted 3 October 2008.
Published online 2 November 2008.

1. Zaug, A. J. & Cech, T. R. The intervening sequence RNA of *Tetrahymena* is an enzyme. *Science* **231,** 470–475 (1986).
2. de Duve, C. Co-chairman's remarks: the RNA world: before and after. *Gene* **135,** 29–31 (1993).
3. Maizels, N. & Weiner, A. M. The 'last ribo-organism' was no breakthrough. *Nature* **330,** 616 (1987).
4. Gilbert, W. The RNA world. *Nature* **319,** 618 (1986).
5. Sharp, P. A. On the origin of RNA splicing and introns. *Cell* **42,** 397–400 (1985).
6. David, C. J. & Manley, J. L. The search for alternative splicing regulators: new approaches offer a path to a splicing code. *Genes Dev.* **22,** 279–285 (2008).
7. Moore, M. J. & Silver, P. A. Global analysis of mRNA splicing. *RNA* **14,** 197–203 (2008).
8. Mili, S. & Steitz, J. A. Evidence for reassociation of RNA-binding proteins after cell lysis: implications for the interpretation of immunoprecipitation analyses. *RNA* **10,** 1692–1694 (2004).
9. Darnell, J. C., Mostovetsky, O. & Darnell, R. B. FMRP RNA targets: identification and validation. *Genes Brain Behav.* **4,** 341–349 (2005).
10. Ule, J., Jensen, K., Mele, A. & Darnell, R. B. CLIP: a method for identifying protein–RNA interaction sites in living cells. *Methods* **37,** 376–386 (2005).
11. Ule, J. *et al.* CLIP identifies Nova-regulated RNA networks in the brain. *Science* **302,** 1212–1215 (2003).
12. Guil, S. & Caceres, J. F. The multifunctional RNA-binding protein hnRNP A1 is required for processing of miR-18a. *Nature Struct. Mol. Biol.* **14,** 591–596 (2007).
13. van der Brug, M. P. *et al.* RNA binding activity of the recessive parkinsonism protein DJ-1 supports involvement in multiple cellular pathways. *Proc. Natl Acad. Sci. USA* **105,** 10244–10249 (2008).
14. Buckanovich, R. J., Posner, J. B. & Darnell, R. B. Nova, the paraneoplastic Ri antigen, is homologous to an RNA-binding protein and is specifically expressed in the developing motor system. *Neuron* **11,** 657–672 (1993).
15. Darnell, R. B. Developing global insight into RNA regulation. *Cold Spring Harb. Symp. Quant. Biol.* **71,** 321–327 (2006).
16. Buckanovich, R. J., Yang, Y. Y. & Darnell, R. B. The onconeural antigen Nova-1 is a neuron-specific RNA-binding protein, the activity of which is inhibited by paraneoplastic antibodies. *J. Neurosci.* **16,** 1114–1122 (1996).
17. Lewis, H. A. *et al.* Sequence-specific RNA binding by a Nova KH domain: implications for paraneoplastic disease and the fragile X syndrome. *Cell* **100,** 323–332 (2000).
18. Ule, J. *et al.* An RNA map predicting Nova-dependent splicing regulation. *Nature* **444,** 580–586 (2006).
19. Yang, Y. Y. L., Yin, G. L. & Darnell, R. B. The neuronal RNA binding protein Nova-2 is implicated as the autoantigen targeted in POMA patients with dementia. *Proc. Natl Acad. Sci. USA* **95,** 13254–13259 (1998).
20. Ule, J. *et al.* Nova regulates brain-specific splicing to shape the synapse. *Nature Genet.* **37,** 844–852 (2005).
21. Jensen, K. B. *et al.* Nova-1 regulates neuron-specific alternative splicing and is essential for neuronal viability. *Neuron* **25,** 359–371 (2000).
22. Dredge, B. K. & Darnell, R. B. Nova regulates GABA(A) receptor gamma2 alternative splicing via a distal downstream UCAU-rich intronic splicing enhancer. *Mol. Cell. Biol.* **23,** 4687–4700 (2003).
23. Dredge, B. K., Stefani, G., Engelhard, C. C. & Darnell, R. B. Nova autoregulation reveals dual functions in neuronal splicing. *EMBO J.* **24,** 1608–1620 (2005).
24. Hu, J., Lutz, C. S., Wilusz, J. & Tian, B. Bioinformatic identification of candidate *cis*-regulatory elements involved in human mRNA polyadenylation. *RNA* **11,** 1485–1493 (2005).
25. Zhao, J., Hyman, L. & Moore, C. Formation of mRNA 3′ ends in eukaryotes: mechanism, regulation, and interrelationships with other steps in mRNA synthesis. *Microbiol. Mol. Biol. Rev.* **63,** 405–445 (1999).
26. Ben-Dov, C., Hartmann, B., Lundgren, J. & Valcarcel, J. Genome-wide analysis of alternative pre-mRNA splicing. *J. Biol. Chem.* **283,** 1229–1233 (2008).
27. Blencowe, B. J. Alternative splicing: new insights from global analyses. *Cell* **126,** 37–47 (2006).
28. Wang, G. S. & Cooper, T. A. Splicing in disease: disruption of the splicing code and the decoding machinery. *Nature Rev. Genet.* **8,** 749–761 (2007).
29. Johnson, J. M. *et al.* Genome-wide survey of human alternative pre-mRNA splicing with exon junction microarrays. *Science* **302,** 2141–2144 (2003).
30. Keene, J. D. RNA regulons: coordination of post-transcriptional events. *Nature Rev. Genet.* **8,** 533–543 (2007).
31. Zhang, C. *et al.* Defining the regulatory network of the tissue-specific splicing factors Fox-1 and Fox-2. *Genes Dev.* **22,** 2550–2563 (2008).
32. Markovtsov, V. *et al.* Cooperative assembly of an hnRNP complex induced by a tissue-specific homolog of polypyrimidine tract binding protein. *Mol. Cell. Biol.* **20,** 7463–7479 (2000).
33. Polydorides, A. D., Okano, H. J., Yang, Y. Y., Stefani, G. & Darnell, R. B. A brain-enriched polypyrimidine tract-binding protein antagonizes the ability of Nova to regulate neuron-specific alternative splicing. *Proc. Natl Acad. Sci. USA* **97,** 6350–6355 (2000).
34. Yuan, Y. *et al.* Muscleblind-like 1 interacts with RNA hairpins in splicing target and pathogenic RNAs. *Nucleic Acids Res.* **35,** 5474–5486 (2007).
35. Grover, A. *et al.* 5′ splice site mutations in *tau* associated with the inherited dementia FTDP-17 affect a stem–loop structure that regulates alternative splicing of exon 10. *J. Biol. Chem.* **274,** 15134–15143 (1999).
36. Early, P. *et al.* Two mRNAs can be produced from a single immunoglobulin mu gene by alternative RNA processing pathways. *Cell* **20,** 313–319 (1980).
37. Rosenfeld, M. G. *et al.* Production of a novel neuropeptide encoded by the calcitonin gene via tissue-specific RNA processing. *Nature* **304,** 129–135 (1983).
38. Iseli, C. *et al.* Long-range heterogeneity at the 3′ ends of human mRNAs. *Genome Res.* **12,** 1068–1074 (2002).
39. Edwalds-Gilbert, G., Veraldi, K. L. & Milcarek, C. Alternative poly(A) site selection in complex transcription units: means to an end? *Nucleic Acids Res.* **25,** 2547–2561 (1997).
40. Sandberg, R., Neilson, J. R., Sarma, A., Sharp, P. A. & Burge, C. B. Proliferating cells express mRNAs with shortened 3′ untranslated regions and fewer microRNA target sites. *Science* **320,** 1643–1647 (2008).
41. Danckwardt, S., Hentze, M. W. & Kulozik, A. E. 3′ end mRNA processing: molecular mechanisms and implications for health and disease. *EMBO J.* **27,** 482–498 (2008).
42. Burge, C. B. *et al.* Alternative isoform regulation in human tissue transcriptomes. *Nature* doi:10.1038/nature07509 (this issue).
43. Zhang, H., Lee, J. Y. & Tian, B. Biased alternative polyadenylation in human tissues. *Genome Biol.* **6,** R100 (2005).
44. Maniatis, T. & Reed, R. An extensive network of coupling among gene expression machines. *Nature* **416,** 499–506 (2002).
45. Proudfoot, N. J., Furger, A. & Dye, M. J. Integrating mRNA processing with transcription. *Cell* **108,** 501–512 (2002).
46. Gawande, B., Robida, M. D., Rahn, A. & Singh, R. *Drosophila* Sex-lethal protein mediates polyadenylation switching in the female germline. *EMBO J.* **25,** 1263–1272 (2006).
47. Takagaki, Y., Seipelt, R. L., Peterson, M. L. & Manley, J. L. The polyadenylation factor CstF-64 regulates alternative processing of IgM heavy chain pre-mRNA during B cell differentiation. *Cell* **87,** 941–952 (1996).
48. Veraldi, K. L. *et al.* hnRNP F influences binding of a 64-kilodalton subunit of cleavage stimulation factor to mRNA precursors in mouse B cells. *Mol. Cell. Biol.* **21,** 1228–1238 (2001).
49. Clark, T. A. *et al.* Discovery of tissue-specific exons using comprehensive human exon microarrays. *Genome Biol.* **8,** R64 (2007).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements The authors are grateful to members of the Darnell and Ule laboratories and J. Richter for critical discussions and review of the manuscript, B. Friedman for suggesting the use of exon arrays, F. Allain for communicating unpublished results, and M. Suarez-Farinas for help with statistics. This work was supported by NIH R01 NS34389 (R.B.D.) and the Howard Hughes Medical Institute. R.B.D. is an HHMI Investigator.

Author Contributions D.D.L. and R.B.D. wrote the paper. D.D.L., A.M. and J.J.F. performed the biochemical and CLIP experiments. J.U. and M.K. developed ASPIRE2 and analysed exon junction array data. D.D.L., S.W.C., X.W. and R.B.D. did bioinformatic analysis. D.D.L, J.C.D. and R.B.D. analysed the data. T.A.C., A.C.S. and J.E.B. developed the exon junction microarray.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare competing financial interests: details accompany the full-text HTML version of the paper at www.nature.com/nature. Correspondence and requests for materials should be addressed to R.B.D. (darnelr@rockefeller.edu).

METHODS

HITS-CLIP. After an initial RT–PCR using DNA primers complimentary to RNA linkers (previously described)[11], an additional PCR reaction was performed using the following fusion primers. AP5fusion1, 5′-GCCTCCCTCGCGCCATCAGCG**AGGGAGGACGATGCGG**-3′. Five additional versions of this 5′ fusion primer were designed, each with a unique 'di-tag' (at the position underlined above; fusion primers 2–6 with AC, TA, CT, GC and GA, respectively), providing the ability to sequence multiple experiments simultaneously. In bold is the sequence complimentary to the 5′ RNA linker used. The remaining sequence at the 5′ end is that of the 454 Life Sciences 'Adaptor A.'

One 3′ fusion primer (BP3fusion) was designed consisting of a sequence complimentary to the 3′ RNA linker used (bold) and that of the 454 Life Sciences 'Adaptor B.' BP3fusion: 5′-GCCTTGCCAGCCCGCTCAG**CCGCTGGAAGTGACTGACAC**-3′.

PCR amplification was performed using Accuprime Pfx (Invitrogen) and ranged between 10 and 15 cycles. The product was then run on a 2% agarose gel and purified using QiaEx II beads (Qiagen). A total of 100 ng of DNA was submitted for sequencing per run. The sequencing of CLIP tags was performed using 454 Life Sciences Adaptor A as the sequencing primer. The 454 adaptor sequences are as follows: Adaptor A, 5′-GCCTCCCTCGCGCCATC-3′; Adaptor B, 5′-GCCTTGCCAGCCCGCTCAG-3′.

Microarrays. RNA samples for all microarrays (Exon junction and MoEx 1.0 ST) were prepared using Whole Transcript Sense Target Labelling Assay and reagents (Affymetrix).

The research exon junction array contains probe sets for all exons and exon–exon junctions observed within transcripts in the input data. As input, the design uses transcript annotations from RefSeq (NCBI35), Ensembl (version 38) and ExonWalk. All transcripts were mapped to the August 2005 version of the mouse genome (NCBI 35, mm7). The array was designed primarily to interrogate well-annotated exons and splicing events from known genes. ExonWalk is a program that merges complementary DNA evidence together to predict full-length isoforms, including alternative transcripts. It was designed to incorporate the richness of transcript variation present in cDNA sequences but to limit some of the noise present in expressed-sequence-tag libraries by including several rules. ExonWalk requires that every exon and junction be: present in cDNA libraries of another organism, have multiple cDNA GenBank entries supporting it, or be evolving like a coding exon as determined by Exoniphy. More information on ExonWalk is available through the UCSC Genome Browser (http://genome.ucsc.edu/index.html).

Probe selection regions (PSRs) were created for the exon probe sets in a manner analogous to the Affymetrix exon arrays (http://www.affymetrix.com/support/technical/technotes/exon_array_design_technote.pdf). An exon was divided into multiple PSRs if there was evidence for alternative splice site usage. We targeted ten perfect match probes for each PSR. All PSRs greater than or equal to 25 bp were represented by at least one probe. All exon–exon junctions observed in the input transcripts (both alternative and constitutive junctions) are interrogated with a probe set containing eight perfect match probes that are tiled in one-base increments from the −4 to the +4 position (relative to the joining event).

Similar to the Affymetrix exon arrays (below), the probes on the array are all 25-base polymers and are designed for sense-strand target. The research junction array contains the same control and background probes as the Affymetrix exon arrays, such that data processing and array quality control methods can be shared by both array types.

Design statistics were as follows: Genome, mm7 (Aug 2005); transcript clusters (genes), 30,833; total transcripts, 145,993; observed junctions, 232,362; PSRs (exons), 270,632; and probes per junction, 8; probes per PSR (exon), 10.

Affymetrix MoEx 1.0 ST exon arrays for analysis of 3′ UTR expression were probed with neocortex RNA from four pairs of P10 wild-type and *Nova2* knockout littermate pairs. To identify candidates for Nova2-dependent alternative polyadenylation, core and extended probe sets mapping to, or within 10 kb of, RefSeq 3′ UTRs were first identified. A 3′ UTR index (analogous to the Splice Index)[49] was then calculated by normalizing the 3′ UTR probe set level to the core transcript level of the corresponding gene to identify differences in specific alternative isoforms, rather than differences due to changes in overall transcript level (see Supplementary Fig. 6 for additional information).

Bioinformatics. UCSC Genome tools (http://genome.ucsc.edu/index.html) were used extensively. To overlay CLIP tags from littermate animals in Supplementary Fig. 2c, because there were different numbers of CLIP tags in each experiment, we chose a computer-generated random subset of tags from the larger set to create sets with equal numbers of tags. A single outlier of intergenic tags (of unknown significance) on chromosome 12 was removed to generate more readable data, leaving a comparison of ~49,000 tags in each group. Graphs were generated with the UCSC Genome Browser or Genome Graph tools, after importing each tag set as a custom track.

YCAY analysis. To calculate YCAY enrichment, the number of observed and expected YCAYs (length of the sequence divided by 64) was calculated. For each data set, enrichment was determined by dividing the sum of observed YCAYs by the sum of expected YCAYs. To assess YCAY enrichment, a probability value was determined by calculating a Chi-squared distribution for each data set.

Normalized complexity map. To adjust for differences in CLIP tag numbers between different transcripts, the total tag number for each was normalized to 1.0. For each 50-nt window, the fraction of these tags present in individual transcripts was determined, and the sum of the fractions for all transcripts in each window was multiplied by the complexity (the number of different transcripts with a CLIP tag in the 50-nt window). As a result, regions with a high CLIP tag number and low complexity are minimized compared to regions with a smaller number of CLIP tags shared by multiple genes.

Biochemical and transfection assays. RPA, RT–PCR and qRT–PCR assays were performed using neocortex RNA from three pairs of P10 wild-type and *Nova2* knockout littermates. Complementary DNA of total RNA was generated using random hexamers and Superscript III (Invitrogen). For RT–PCR of each candidate splicing target, different cycle numbers were tested to ensure linear amplification of PCR products. Radiolabelled ^{32}P α-dCTP was added to the PCR reactions for the last two cycles. RPA probes were *in vitro* transcribed from linearized plasmids containing cloned sequences corresponding to the poly(A) sites in Cugbp2 (−275 to +68 nt) and Slc8a1 (−218 to +136 nt), and actin probes were synthesized from pTRI-Beta-Actin-mouse (Ambion). Radiolabelled probes were gel purified, and hybridizations and digestions performed as recommended by the manufacturer (Ambion RNAIII kit). RNase protected fragments and RT–PCR products were separated by electrophoresis on 6% polyacrylamide/7 M urea gels. qRT–PCR analysis of alternative polyadenylation candidates was performed using primers pairs located upstream and downstream of the poly(A) site being assayed. qRT–PCR analyses were performed in experimental triplicate using iTaq SYBR Green reaction mixes and iQ5 Real-Time PCR machines (BioRad).

The alternative poly(A) reporter plasmid was constructed by insertion of a PCR-amplified fragment corresponding to the Slc8a1 pA2 (−250 to +250 relative to the cleavage site) into plasmid vector pGFP (Clontech). The Slc8a1 pA3 (−1250 to +250) was inserted downstream of pA2. To ensure efficient expression in mammalian cells, a cytomegalovirus promoter was inserted upstream of the GFP open reading frame. YCAY elements were converted to YACY by PCR site-specific PCR mutagenesis. Wild-type or mutant poly(A) reporter plasmids were transfected (Lipofectamine-2000), with or without a Nova2-expressing plasmid (pNova2, R.B.D. *et al.*, unpublished), into 293T cells. All transfections were performed in triplicate. Primer sequences are available on request. For ASPIRE2, see Supplementary Information.