

Methods

HITS-CLIP.

Following 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'

AP5fusion2: 5'-GCCTCCCTCGCGCCATCAGAC**AGGGAGGACGATGCGG**-3'

AP5fusion3: 5'-GCCTCCCTCGCGCCATCAGTA**AGGGAGGACGATGCGG**-3'

AP5fusion4: 5'-GCCTCCCTCGCGCCATCAGCT**AGGGAGGACGATGCGG**-3'

AP5fusion5: 5'-GCCTCCCTCGCGCCATCAGGCC**AGGGAGGACGATGCGG**-3'

AP5fusion6: 5'-GCCTCCCTCGCGCCATCAGGA**AGGGAGGACGATGCGG**-3'

Six versions of the 5' fusion primer were designed, each with a unique 'di-tag' (underlined), 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 'Adapter A.'

BP3fusion: 5'-GCCTTGCCAGCCCGCTCAG**CCGCTGGAAGTGA**CTGACAC-3'

One 3' fusion primer was designed consisting of a sequence complimentary to the 3' RNA linker used (bold) and that of the 454 Life Sciences 'Adapter B.'

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 Adapter A as the sequencing primer.

454 Adapter Sequences:

Adapter A: 5'-GCCTCCCTCGCGCCATC-3'

Adapter B: 5'-GCCTTGCCAGCCCGCTCAG-3'

Microarrays.

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

Research Exon Junction Array

Array Design. The research junction array contains probesets 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 cDNA 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 limit some of the noise present in EST libraries by including several rules. ExonWalk requires that every exon and junction be either: 1) Present in cDNA libraries of another organism, 2) Have multiple cDNA GenBank entries supporting it, or 3) 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/cgi-bin/hgTrackUi?g=exonWalk>).

Probe selection regions (PSRs) were created for the exon probesets in a manner analogous to the Affymetrix Exon Arrays (http://www.affymetrix.com/support/technical/technotes/exon_array_design_tech_note.pdf). An exon was divided into multiple PSRs if there was evidence for alternative splice site usage. We targeted 10 perfect match probes for each PSR. All PSRs greater than or equal to 25bp 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 probeset containing 8 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-mers 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 QC methods can be shared by both array types.

Affymetrix MoEx 1.0 ST Exon Arrays for analysis of 3' UTR expression.
Affymetrix exon arrays were probed with neocortex RNA from 4 pairs of P10 WT and Nova2 KO littermate pairs. To identify candidates for Nova2-dependent alternative polyadenylation, core and extended probesets mapping to, or within 10kb of, RefSeq 3' UTRs were first identified. A 3' UTR index (analogous to the Splice Index)⁴⁸ was then calculated by normalizing the 3' UTR probeset level to the core transcript level of the corresponding gene in order to identify differences in specific alternative isoforms, rather than differences due to changes in overall transcript level (see Supplemental Fig. 6 for additional information).

Design Statistics

		Mouse	Research
		Junction Array (MJAY)	
Genome		mm7 (Aug 2005)	
Transcript	Clusters	30,833	
	(genes)		
Total Transcripts		145,993	
Observed Junctions		232,362	
PSRs (exons)		270,632	
Probes per Junction		8	
Probes per PSR (exon)		10	

Bioinformatics.

UCSC Genome tools (<http://genome.ucsc.edu/index.html>) were used extensively. To overlay CLIP tags from littermate animals in Fig. 1a, since 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 chr12 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.

YCA Y analysis

To calculate YCA Y enrichment, the number of observed and expected YCA Ys (length of the sequence divided by 64) was calculated. For each dataset, enrichment was determined by dividing the sum of observed YCA Ys by the sum of expected YCA Ys. To assess YCA Y enrichment, a probability value was determined by calculating a Chi-squared distribution for each dataset.

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.

RNase protection (RPA), RT-PCR, and qRT-PCR assays were performed using neocortex RNA from 3 pairs of P10 WT and Nova2 KO littermates. cDNA 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. Radiolabeled ³²P 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). Radiolabeled probes were gel purified, and

hybridizations and digestions performed as recommended by the manufacturer (Ambion RNAlIII kit). RNase protected fragments and RT-PCR products were separated by electrophoresis on 6% polyacrylamide/7M 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 pGFP (Clontech). The SLC8A1 pA3 (-1250 to +250) was inserted downstream of pA2. To ensure efficient expression in mammalian cells, a CMV promoter was inserted upstream of the GFP ORF. 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, Darnell et al., unpublished), into 293T cells. All transfections were performed in triplicate.

Primer sequences are available on request.

ASPIRE2

Example of the ASPIRE2 algorithm to calculate ΔI

The example gene contains an alternative exon 2, flanked by constitutive exons 1 and 3.

Microarray data provide three types of values to analyze the alternative splicing event:

Median value of probes recognizing constitutive exons: Median (E1, E3).

Median value of probes recognizing exon 2 including isoform: Median (E2, E1-E2, E2-E3).

Median value of probes recognizing exon 2 skipping isoform: Median (E1-E3).

	Sample 1 (three replicates from wildtype brain)	Sample 2 (three replicates from knockout brain)
Median (E1, E3)	300	500
Median (E2, E1-E2, E2-E3)	400	300
Median (E1-E3)	100	300

Normalization:

in = Median (E2, E1-E2, E2-E3) / Median (E1, E3)

ex = Median (E1-E3) / Median (E1, E3)

Ratios:

Rin = in1 / in2

Rex = ex1 / ex2

in	1.33	0.60	Rin = 2.22
ex	0.33	0.60	Rex = 0.55

I : the fraction of exon including isoform

E : the fraction of exon skipping isoform

T: total mRNA expression = 1

Since exon including (I) and excluding (E) isoforms are reciprocal: $T = I + E = 1$

$$I_1 + E_1 = 1 \Rightarrow E_1 = 1 - I_1$$

$$I_2 + E_2 = 1 \Rightarrow I_1/Rin + E_1/Rex = 1 \Rightarrow E_1 = Rex (1 - I_1/Rin)$$

$$1 - I_1 = Rex (1 - I_1/Rin) \Rightarrow I_1 = (Rex - 1) * Rin / (Rex - Rin)$$

$$I2 = I1 / Rin \Rightarrow I2 = (Rex-1) / (Rex - Rin)$$

$$\Delta I = I1 - I2 \Rightarrow \Delta I = (Rex-1) * (Rin-1) / (Rex - Rin)$$

Rather than comparing the values of probesets from individual samples, the fraction of exon including isoform is calculated from ratios Rin and Rex. These values detect splicing change only if reciprocal probe sets change in opposite directions.

For our example above:

$$\Delta I = (0.55-1) * (2.22-1) / (0.55-2.22) = -0.45 * 1.22 / -1.67 = 0.33$$

Prediction: the first sample (wildtype) has 33% more exon included isoform.

ASPIRE2 web server:

ASPIRE2 analysis can be done using the following public server:
<http://www.aspire.edu/>.

Description of the APIRE2 software:

1. Prepare the annotation files:

- a) Prepare **exon_number** file by using the file 'mm7_exonwalk_plus.design.mappin' (provided by Affymetrix), which contains genomic location, number and % transcript hits for the exon or exon-exon junction recognized by each probeset, via the following steps:
 - Annotate the genomic position of JUC probes
 - Annotate the cases where 'none' is written instead of PSR in the F and G columns of JUC rows. Replace these with EX numbers, and insert the corresponding EX rows between the PSR rows to be used for exon numbering in the following step.
 - Genomic exon number for exon probesets (starting with PSR or EX) is numbered sequentially top down when gene strand is + (column C) or bottom up when gene strand is -.
 - If %TR hits is larger than 50%, exon probeset is labeled as C (constitutive), otherwise A. For JUC, C is added only if both contacted exons are labelled with C, and if the JUC doesn't skip an exon (that is E7-E8 is C, but E7-E9 is A).
 - mRNA exon number is also named sequentially, but if the distance between exons is smaller than 40, the number stays the same, but 'var' is added, and if %TR hits is smaller than 50%, the number stays the same. For all alternative exons (A), 'alt' is added to the name.
 - The genomic and mRNA numbers of exon contacted by JUC probesets are numbered.

b) Prepare **results_annotation** file:

- Use mjay.r1.pgf file to get probe_id numbers of probes for each probeset, and their GC contents.
- Use mjay.r1.clf file to add x/y coordinates for each probe.
- Based on common probeSet codes, add from the exon_number file the Transcript Cluster (TC), C/A and genomic exon number information.
- Leave rows 3-25 for background calculation. Sort the remaining rows by TC (to put together all probes for the same gene), and by genomic exon number (to put together the rows for proximal exons).
- Copy TC, Probeset code and genomic exon number from 'exon_number' file.

c) Prepare **ASPIRE_combinations** files 1 and 2, where probeset combinations are determined using genomic exon numbers via the following rules:

- file 1 (high stringency)

rule 1: Only junctions (JUC) that skip one or more exons can be used for Eex.

rule 2: Only junctions (JUC) that share one exon with Eex can be used for Ein. The other exon contacted by Ein JUC is skipped by Eex probeset.

rule 3: If two junctions are used for Ein, they have to share the exon that lacks in Eex.

rule 4: The PSR probeset used for Ein has to be of the exon that is contacted by the Ein junction probesets, but not by the Eex probeset.

rule 5: Ein has to contain two or three probesets (one is not enough).

rule 6: if different probesets are used for Ein, treat them from the very beginning of the ASPIRE script as a single probeset.

rule 7: Ein is named by the exon that is shared by all Ein probesets and is absent in Eex probeset.

- File 2 (low stringency):

This file contains combinations that follow less stringent rules (only one Ein probeset is enough). Important: it will contain only combinations that are NOT in file 1.

rule 1: Only junctions (JUC) that skip one or more exons can be used for Eex.

rule2: If the same Eex was used in file 1, it can only be combined with Ein probesets that were not combined with it in file 1.

rule 3: Ein contains only one probeset (either JUC or PSR).

rule 4: If junction (JUC) is used for Ein, it needs to share one exon with Eex, while the other exon contacted by Ein JUC is skipped by Eex probeset.

rule 5: PSR can only be used for Ein when the Eex skips one or two exons in a row, but no more (i.e., Eex is Ey-Ez, then z-y<3). If PSR is used, it corresponds to the exon that is skipped by Eex.

rule 6: Ein is named by the exon that is contacted by Ein probeset and is absent in Eex probeset.

d) Prepare **output_annotation** files:

- Each time you run the ASPIRE, it will output two separate output files, called fold_change and splicing_change. For the fold change

file, add each transcript into a single row, and include TFC, TFCtest and all the gene-related, but not exon-related annotation (i.e., the columns coming from exon_number and ASPIRE_combinations file). Fold change file includes all transcript clusters that are in the mm7 file, regardless of whether these are in the aspire combination files or not. The splicing change file is prepared by combining data from two aspire combination files, 1 (high) and 2 (low). In the output file, have a separate column that tells whether the data is from high or low file.

- get the following information from ASPIRE_combinations file: TIGR Transcript Cluster, Eex, Ein.
- Based on shared TIGR Transcript Cluster number, add the following information from mm7_exonwalk_plus.design.annotation.map file: Chr, Gene Start, Gene Stop, Strand, Ensembl Gene ID, number of sequenced transcripts
- Based on shared genomic exon number of Eex, add the following information from exon_number file: mRNA exon number, Start, Stop, exon junction sequence.
- Based on shared genomic exon number of Ein, add the following information from exon_number file: mRNA exon number, Start, Stop, % transcript hits
- Based on shared Ensembl Gene ID, download the following information from <http://www.ensembl.org/biomart/martview>: Gene Symbol, Description, Entrez Gene ID,
- Based on shared Gene Symbol, download the following information from <http://david.abcc.ncifcrf.gov/>: Pfam, Gene ontology

2. Data extraction, background subtraction and normalization of data

- a) **USER CHOICE (sample names):** Ask the user to provide sample pair names that will then be used to name the output file (ex: wtCTX_Nova-2/koCTX_Nova-2). For example:
 SAMPLE PAIR1: Sample1 = (ex: wtCTX_Nova-2) / Sample2 = (ex: koCTX_Nova-2)
 SAMPLE PAIR2: Sample1 = (ex: wtCTX_Nova-2) / Sample2 = (ex: wtTHY_Nova-2)
 etc.
- b) **USER CHOICE (replicate number):** Define the number of replicates used for each sample (default= 3). It is allowed to use different number of replicates for the two samples of the same pair.
- c) **USER CHOICE (cel files):** Ask the user to choose the cel file corresponding to each replicate for each sample.
- d) **USER CHOICE (background):** Ask the user to determine the percentile (a choice of 10-90) that will be used for background calculation.

- e) Based on mjay.r1.bgp file (where x/y coordinated of background probes are given), calculate for each set of probes sharing same %GC in each cel file individually the value corresponding o the percentile chosen by the user.
 - f) Place average value of each probe into results_annotation file (where x/y coordinates are defined), and subtract the median value of background probes corresponding to its GC ratio (which is also defined in results_annotation file). If the resulting value is <0, write 0.
 - g) Place data for each pair separately in results file. For example, even if same sample (wtCTX_Nova-2) is chosen for SAMPLE PAIR1 and SAMPLE PAIR2 (see step c above), place the same data twice in results file (this will allow pair-specific filtering in step 3).
 - h) Calculate mean A value of all probes in individual replicates.
 - i) Calculate the mean B of all mean A values.
 - j) Divide values of each probe in one replicate by the mean A value for that replicate and multiply all values by the mean B value (this normalizes for different RNA concentrations that might have been used for each replicate).
 - k) The file is now called 'results', and can be used by the ASPIRE program.
3. Probe filtering: even though the program is able to analyze several pairs at a time (as chosen in step c), the filtering is done separately for each pair!
- a) calculate mean, median and standard deviation (st.dev) for each probe in each Sample (by using values from all chosen replicates)
 - b) **USER CHOICE (probe filtering):** Define the values D1 and D2. The default the values should be **D1 (probe minimal value)=10 (suggest value between 1 and 100), D2 (probe maximal variance)=4 (suggest value between 1 and 50)**
 - c) Remove the probe values from corresponding Sample pair in results file if $(\text{median1} * \text{median2}) / 20 < D1$.
 - d) Remove the probe values from corresponding Sample pair in results file $(\text{mean1} * \text{mean2}) / (\text{stdev1} * \text{stdev2}) < D2$
4. Calculate transcript fold change (TFC) for each transcript cluster - (*this value is analogous to the result obtain by ordinary microarrays.*)
- a) For each probe from probesets labeled as C (constitutive), divide the median probe value of replicates in Sample1 with the median value of replicates in Sample2 to get the probe ratio values (R), for each calculate $\log_2(R)$ and add the values into an additional column of the results file. If the TC contains less than 15 non-filtered probe values from C probesets, then include probes of all the A probesets for this analysis. If no non-filtered probes are left for the gene, then leave cells for TFC and all the other data empty in the output file (do not write 0).

- b) Calculate the median, mean, degrees of freedom (n-1) and standard error ($s = \text{sdev} / \sqrt{n-1}$) of the $\log_2(R)$ values. Then calculate the TFC ($\text{TFC} = 2^{\text{median}} = 2$ to the power of median), $\log_2(\text{TFC})$ (=median) and TFC tTest (mean/s).
 - c) Place the transcript fold change (TFC), $\log_2(\text{TFC})$ and TFC tTest into three columns of the TFC output file. Place only TFC into one column of the other two output files (aspire_high and aspire_low).
5. Normalize probe values relative to TFC and combine probe values into probe sets:
 - a) In results file, divide values of probes in Sample1 by the TFC for the corresponding transcript cluster. These are normalized probe values.
(Leave values in Sample2 unchanged. This step will normalize for changes of gene-specific transcription and RNA degradation, so that in continuation, we can specifically analyze changes in alternative splicing.)
 - b) For each probe divide the median probe value of replicates in Sample1 with the median value of replicates in Sample2 to get the probe ratio values (R), for each calculate $\log_2(R)$ and add the values into an additional column of the results file.
6. Probeset filtering: the filtering is done separately for each sample pair!
 - a) **USER CHOICE (Probeset filtering):** Define the value D3 (probeset maximal variance), the default the value should be **D3=0 (suggest value between 0 and 5)**
 - b) Order the rows of the results file corresponding to the Eex and Ein probeset groups from the ASPIRE combination file.
 - c) For each probeset, calculate $\text{PS} = 10 * \text{ABS}(\text{mean}(\log_2(R))) - 5 * \text{the stdev}(\log_2(R))$. Delete both Eex and Ein values if PS for either Eex or Ein is $< D3$ ($\text{PS} < D3$).
 - d) For each probeset, calculate $\text{median}(\log_2(R))$. If $(\text{Eex}(\text{median}(\log_2(R))) * \text{Ein}(\text{median}(\log_2(R))) > 0$, then delete both Eex and Ein values.
7. ΔI calculation
 - a) For each probeset, calculate the degrees of freedom (n-1), standard error ($s = \text{sdev} / \sqrt{n-1}$) and tTest (mean/s) of the $\log_2(R)$ values. Then calculate $dI \text{ tTest} = (\text{ABS}(\text{tTest}(\text{Rex}))) * \text{tTest}(\text{Rin})$.
 - b) Calculate the median(R) ($2^{\text{median}(\log_2(R))} = 2$ to the power of median($\log_2(R)$)). Then calculate ΔI by the following equation:

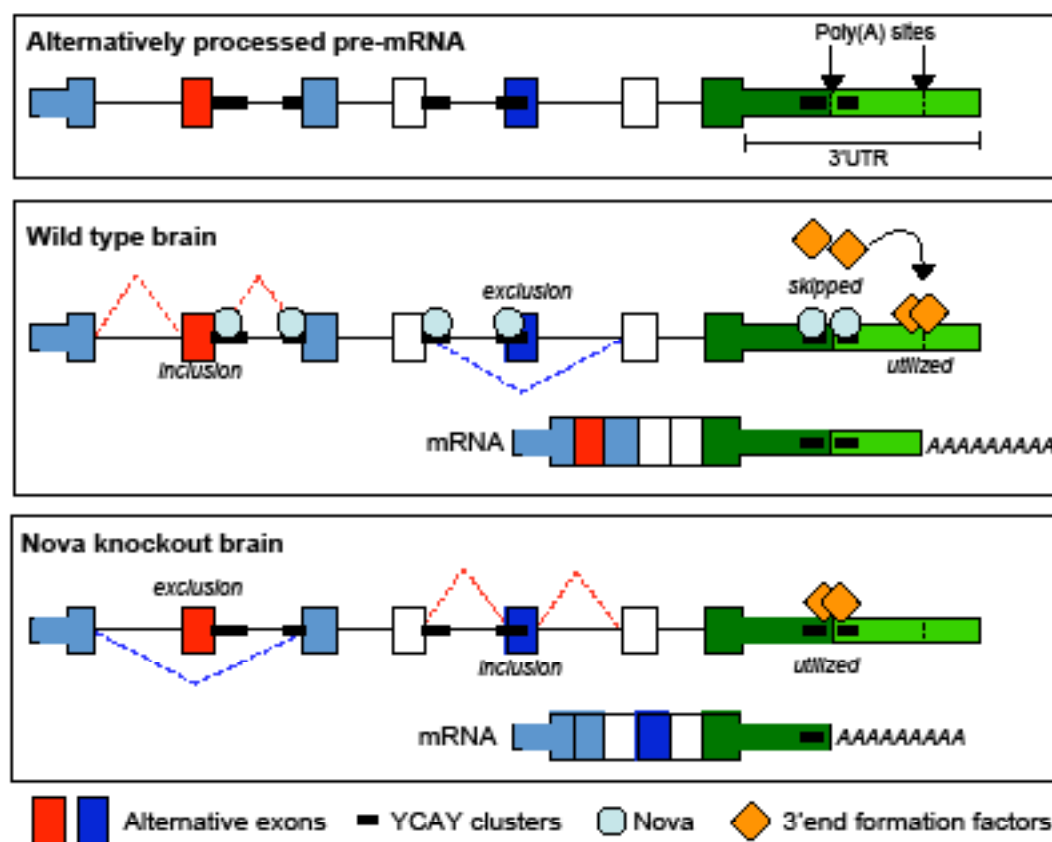
$$I1 = ((\text{median}(\text{Rex}) - 1) * (\text{median}(\text{Rin}))) / (\text{median}(\text{Rex}) - \text{median}(\text{Rin}))$$

$$I2 = I1 / (\text{median}(\text{Rin}))$$

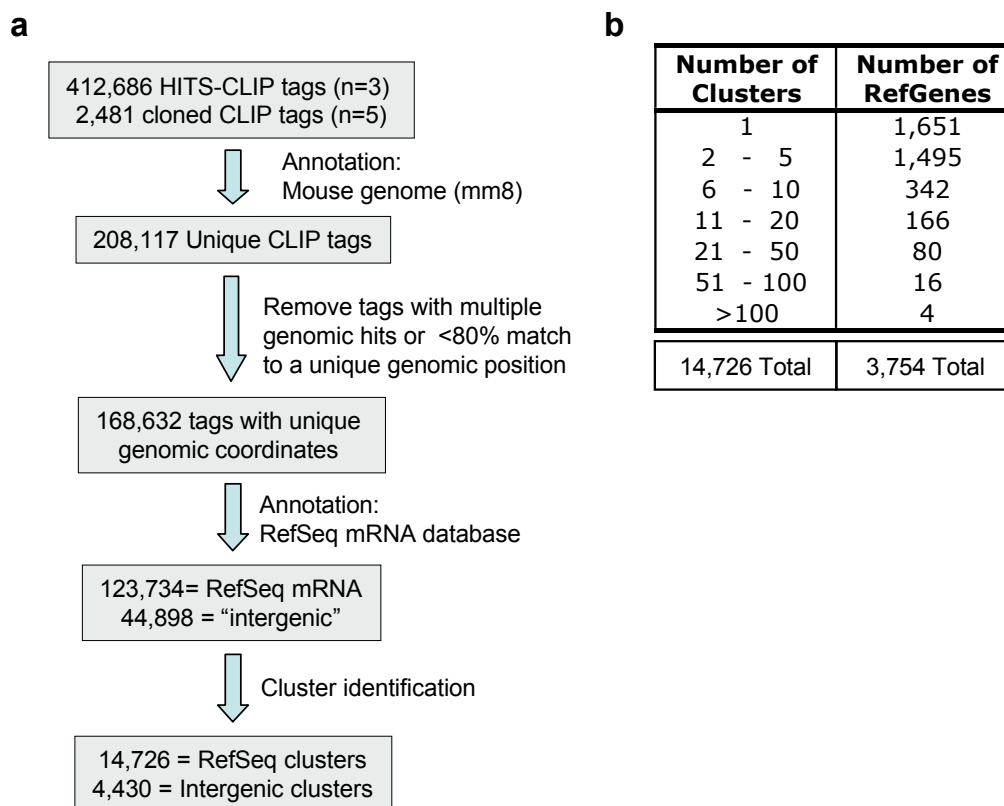
$$dI = I1 - I2$$
 - d) Place I1, I2, dI and dI tTest into the output file

References to Methods

11. Ule, J. et al. CLIP identifies Nova-regulated RNA networks in the brain. *Science* **302**, 1212-1215 (2003).
47. Clark, T. A. et al. Discovery of tissue-specific exons using comprehensive human exon microarrays. *Genome biology* 8 (4), R64 (2007).

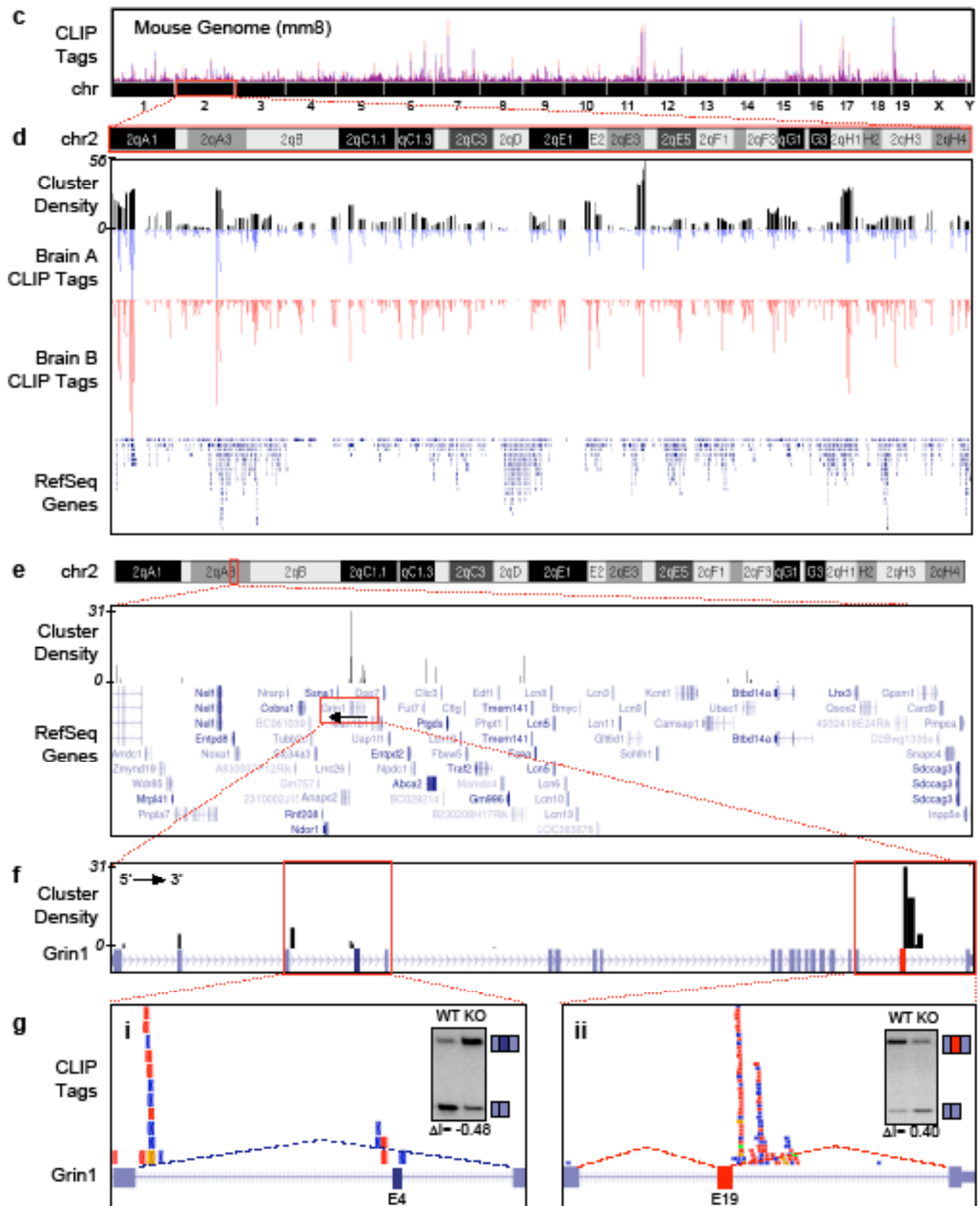


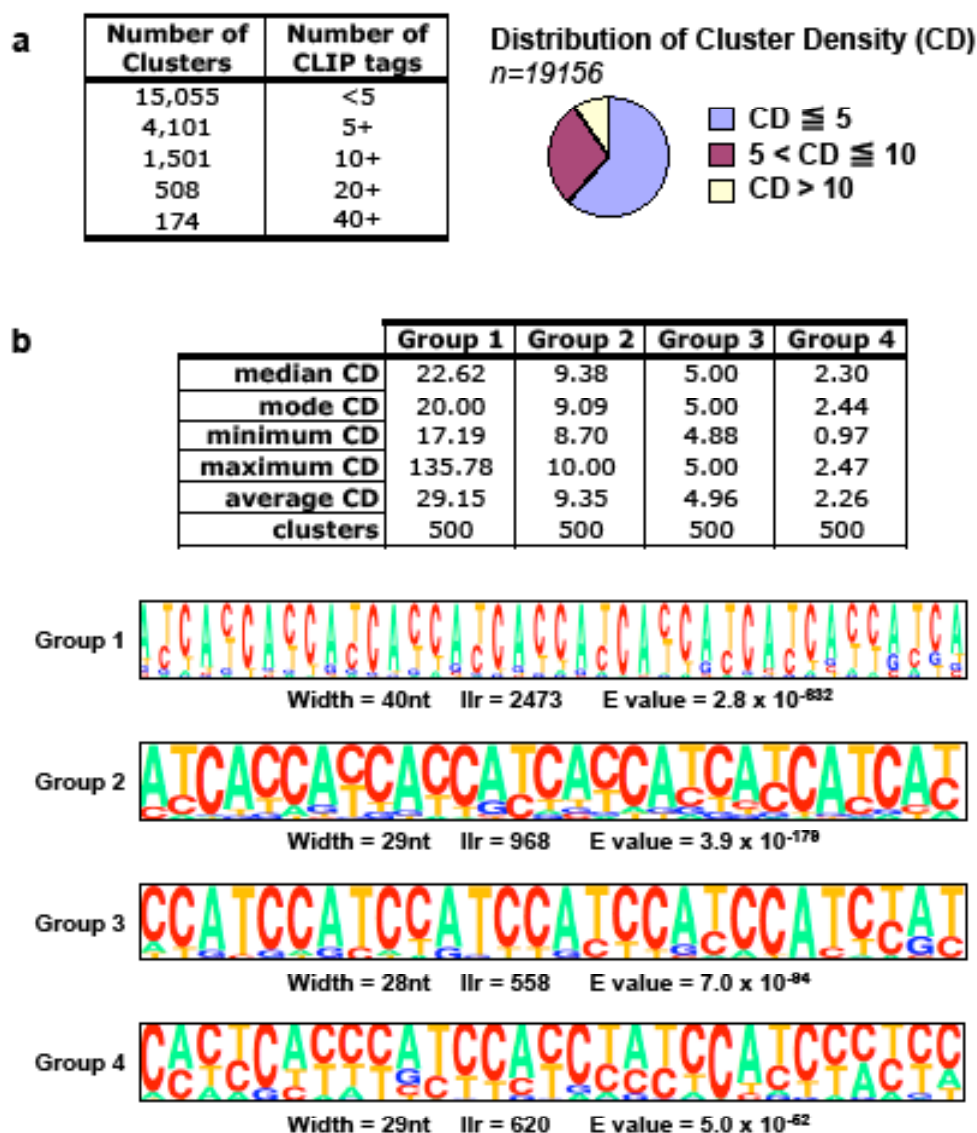
Supplemental Fig. 1. Model for Nova action based on HITS-CLIP data and analysis of Nova2 KO RNA. Nova binds directly to clusters of YCAY elements to mediate either alternative splicing or alternative polyadenylation. This action can be to inhibit core factor binding, as demonstrated for Nova's action to inhibit U1 snRNP¹ or as postulated for its action to inhibit core factors that bind in the neighborhood of poly(A) sites, such as CPSF and CSTF (modeled by orange diamonds). Alternatively, Nova can enhance either exon inclusion or alternative splicing, perhaps through recruitment of auxiliary factors. In either case, the actions are determined by the location of YCAY elements and direct Nova binding.



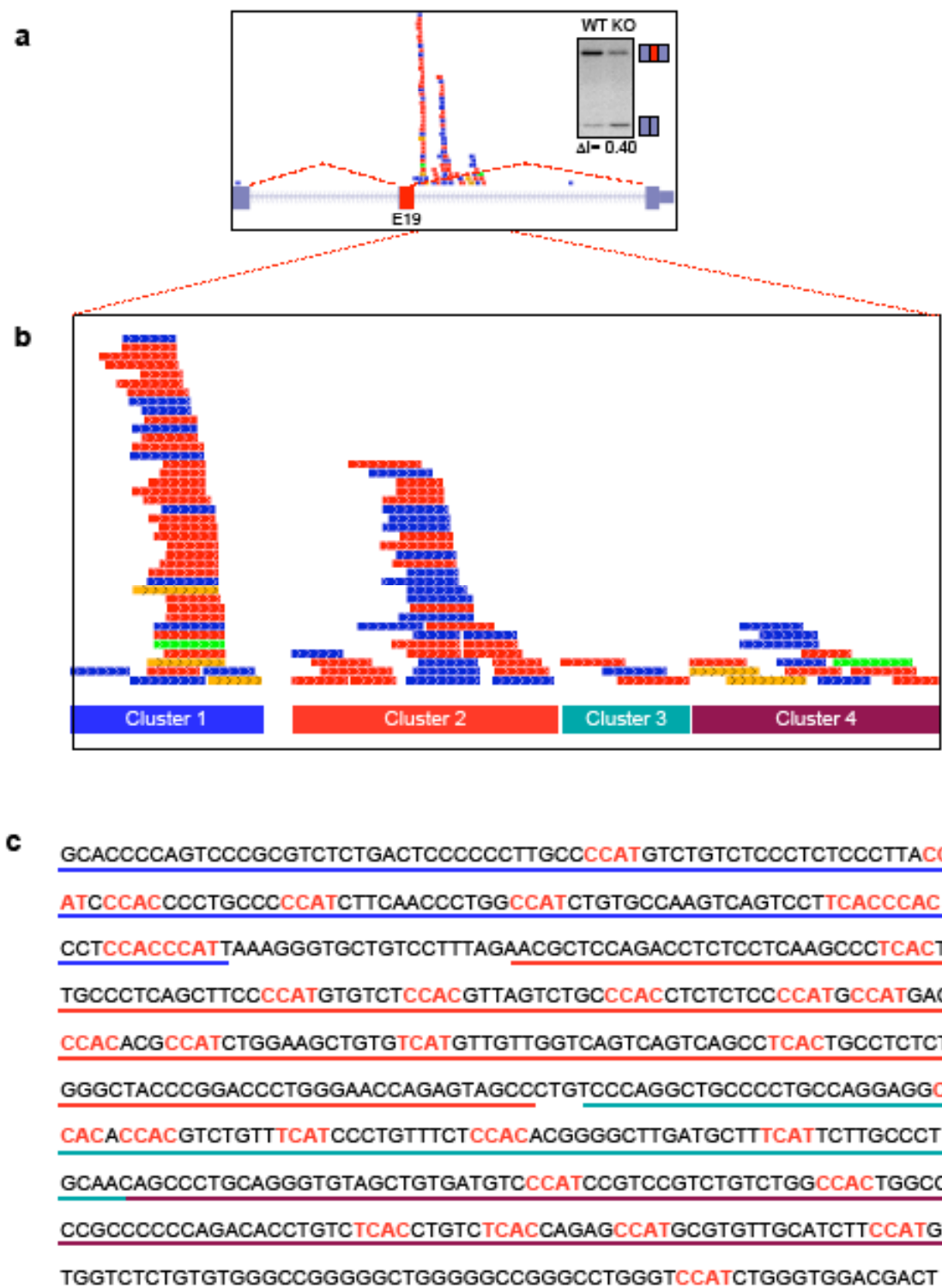
Supplemental Fig. 2. a, flow chart illustrating culling of raw data to unique Nova CLIP tags and clusters. Three HITS-CLIP experiments were done using mouse neocortex. Neocortex expresses predominantly Nova2, with very little Nova12. In two experiments crosslinked RNA-protein complexes above 75kD were taken, which are Nova2 specific bands²; these correspond to Brain A and Brain B tags in Fig. 1. In a third HITS-CLIP experiment done on P9 mouse neocortex, crosslinked RNA-protein complexes above 75kD were taken, and above 55kD were taken, corresponding to a mixture of Nova1 and Nova2 tags. In total, >155,000 tags correspond definitively to Nova2 bound RNAs. **b**, Table illustrating breakdown of number of RefSeq genes (from 3,754 total) harboring a given number of clusters. **c**, Equal numbers of Nova CLIP tags from two littermate animals (red and blue) were overlaid using GenomeGraph (USCS) across the mouse genome (mm8); the Y axis plots the number of tags per 10,000 nt window. **d**, HITS-CLIP map of chromosome 2. Nova cross-linked RNAs are represented as the number and position of tags obtained from the neocortex of two mouse littermates (Brain A/B), and are compared with the position of clusters of tags (overlapping tags, plotted according to position and cluster density (black; #tags/cluster length), and compared with the location of all RefSeq genes. Brain had 46,106 tags present in 10,740 clusters (of minimum size two tags), and Brain B had 100,874 tags in 15,805 clusters. **e-g**, Clusters density and individual Nova cross-linked tags (colored, where each color represents a tag from a different animal) in or near the *Grin1* transcript on chromosome 2; specificity is evident between genes (e; at least 12 of these could be documented to be expressed in brain) and within the gene (f); the position of CLIP tags in *Grin1* correspond to a previously predicted site of Nova-mediated exon inclusion (E19; g(ii)) and a newly predicted site of Nova-mediated exon inhibition (E4; g(i)); both sites were experimentally validated by comparing exon usage in WT versus Nova2 KO neocortex (insets (g)) as described previously³.

Figure 2



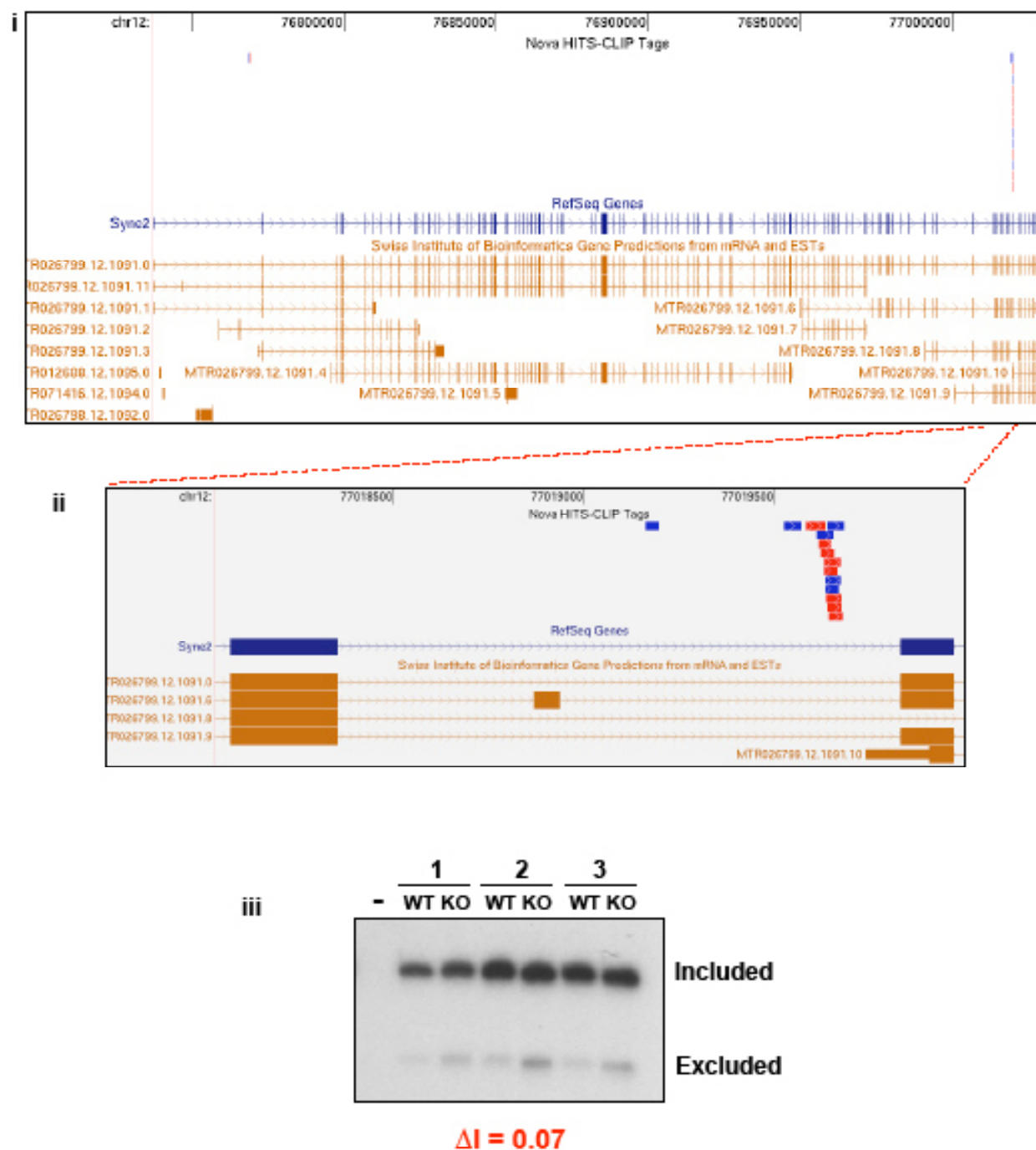


Supplemental Fig. 3. a, The number of clusters that have the indicated cluster density (CD), from total of 19,156 Nova CLIP tag clusters. **b**, Results of MEME analysis four different groups of 500 clusters each that differ in the indicated CD. All clusters, from most to least dense, show enrichment of YCA Ys. The consensus sequences are shown below, with E values showing robust significance, with a strict dependence on CA dinucleotides bounded by pyrimidines, as predicted from the Nova-RNA crystallographic structure. The four most common tetranucleotide motifs were CCAC=17004, CCAT=23422, TCAC=14332, TCAT=19516.



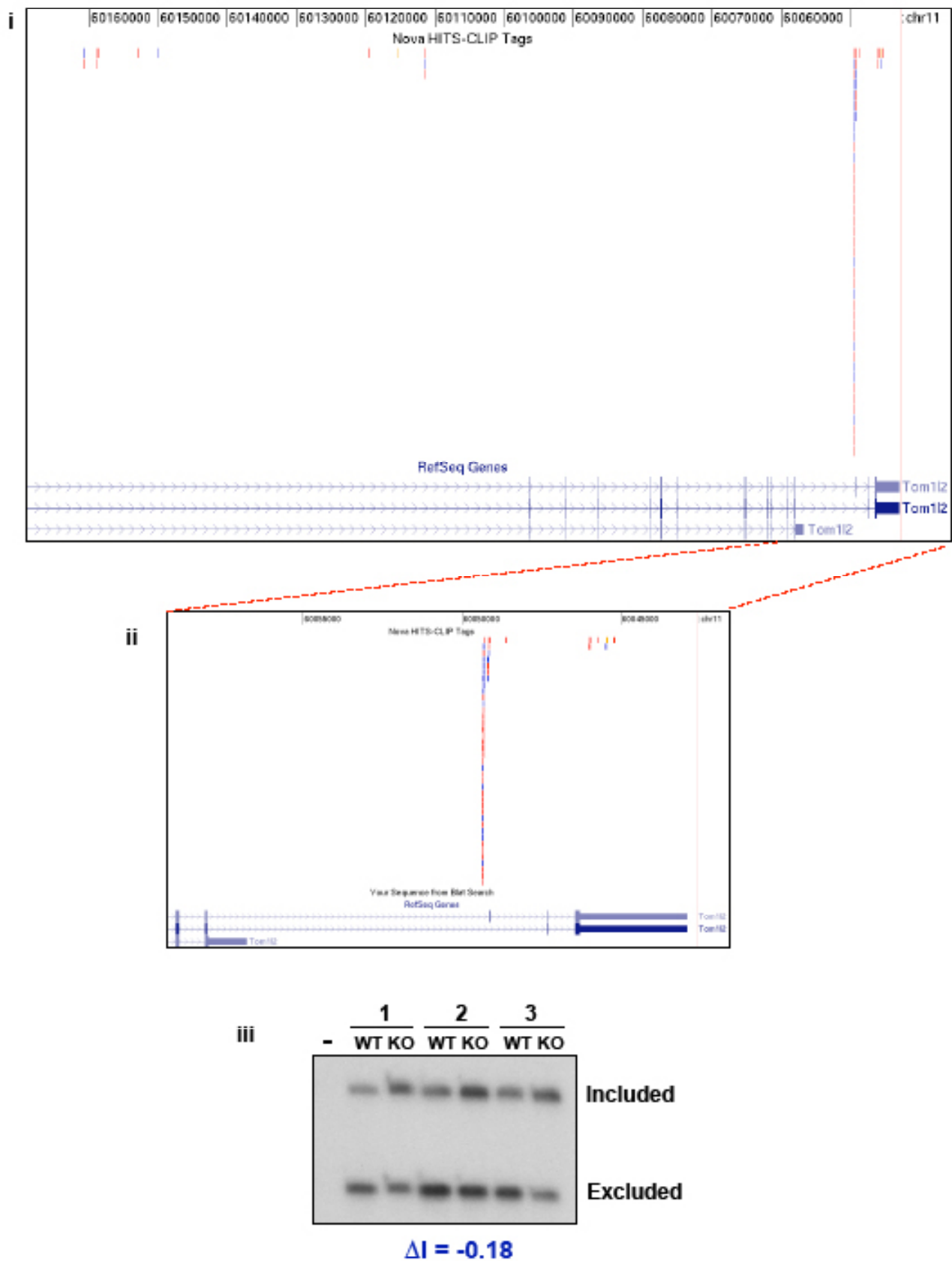
Supplemental Fig. 4. a, *Grin1* pre-mRNA HITS-CLIP clusters are present just downstream of a previously identified Nova-dependent alternate exon (exon 19) whose inclusion correlated with the presence of Nova. These tags clustered in a position consistent with the predicted Nova:RNA interaction map and were rich in YCAY clusters. **b**, Detailed view of clusters downstream of E19, in four regions corresponding to the YCAY-rich sequences underlined in (c). Note that E19, whose inclusion affects trafficking of the NMDA receptors to the synapse, is also referred to as the CI exon or E21 in prior publications⁴ (see comments in Han et al., 2005⁵). Similar nomenclature issues arise with respect to E4, which encodes an alternative extracellular N-terminal domain that regulates receptor pharmacology, and is also referred to as E5 in some papers⁴.

Figure 5a



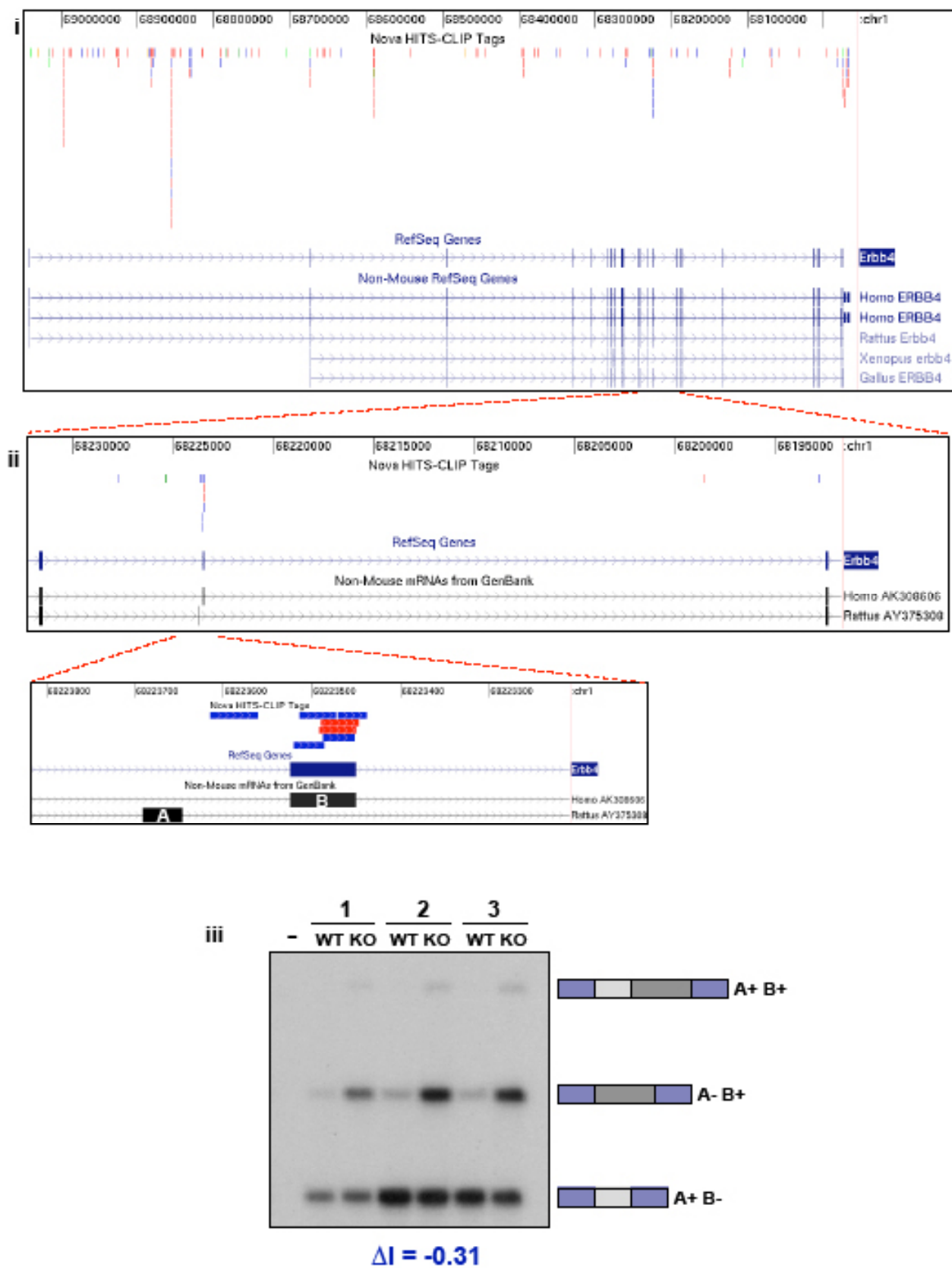
Supplemental Fig. 5. (a-f), Nova CLIP tags mapping to candidate Nova-regulated transcripts are shown in whole gene (i) and region-specific (ii) views which include alternatively spliced isoforms present in different databases. iii, RT-PCR analysis of cortex RNA from 3 pairs of P10 wild type and Nova2 KO littermates, including a RT- control (-). The change in exon inclusion (ΔI) is shown below, with red and blue text representing exons that are preferentially included or excluded in a Nova2-dependent manner respectively.

Figure 5b



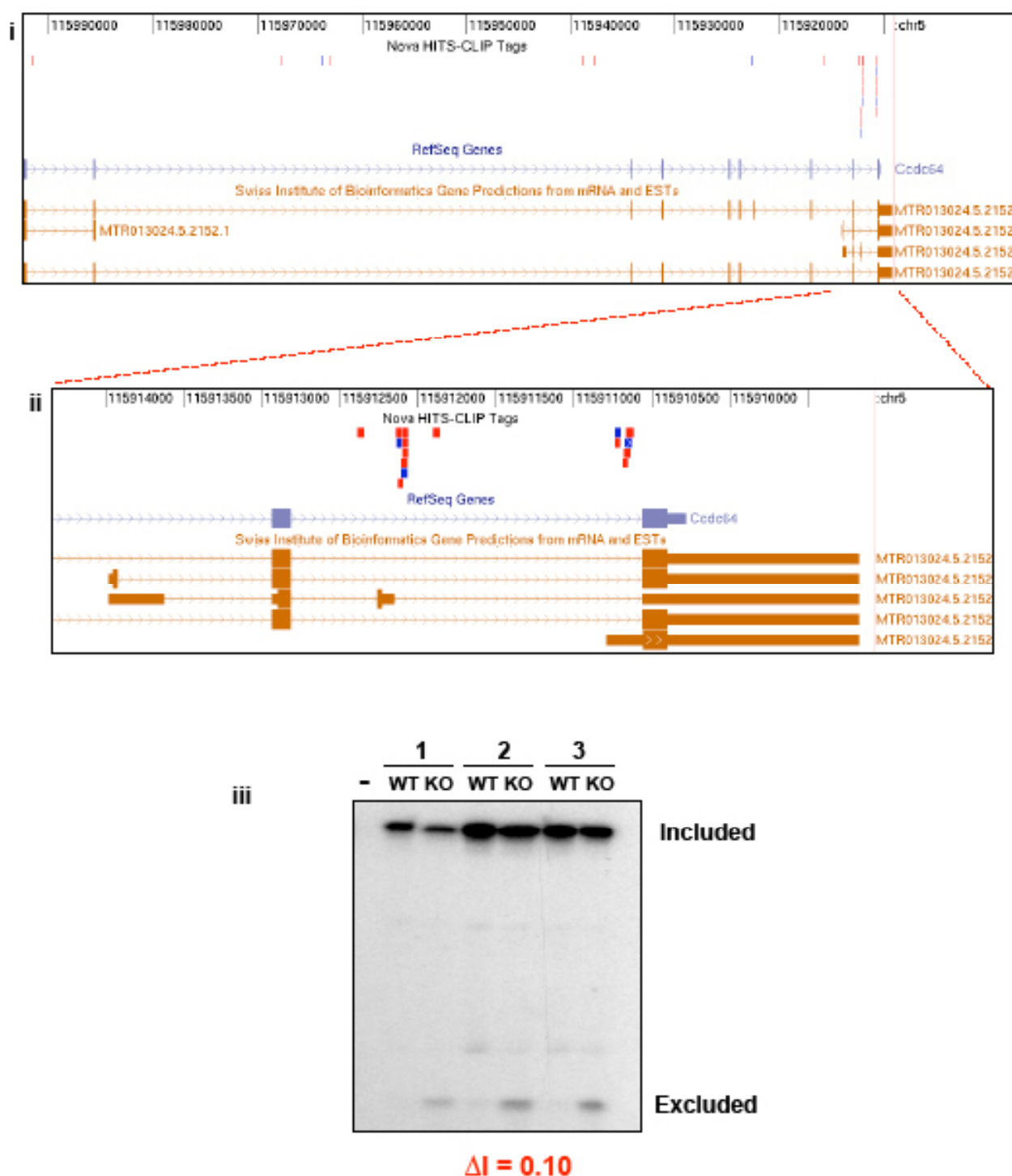
Supplemental Fig. 5. (a-f), Nova CLIP tags mapping to candidate Nova-regulated transcripts are shown in whole gene (i) and region-specific (ii) views which include alternatively spliced isoforms present in different databases. iii, RT-PCR analysis of cortex RNA from 3 pairs of P10 wild type and Nova2 KO littermates, including a RT- control (-). The change in exon inclusion (ΔI) is shown below, with red and blue text representing exons that are preferentially included or excluded in a Nova2-dependent manner respectively.

Figure 5c



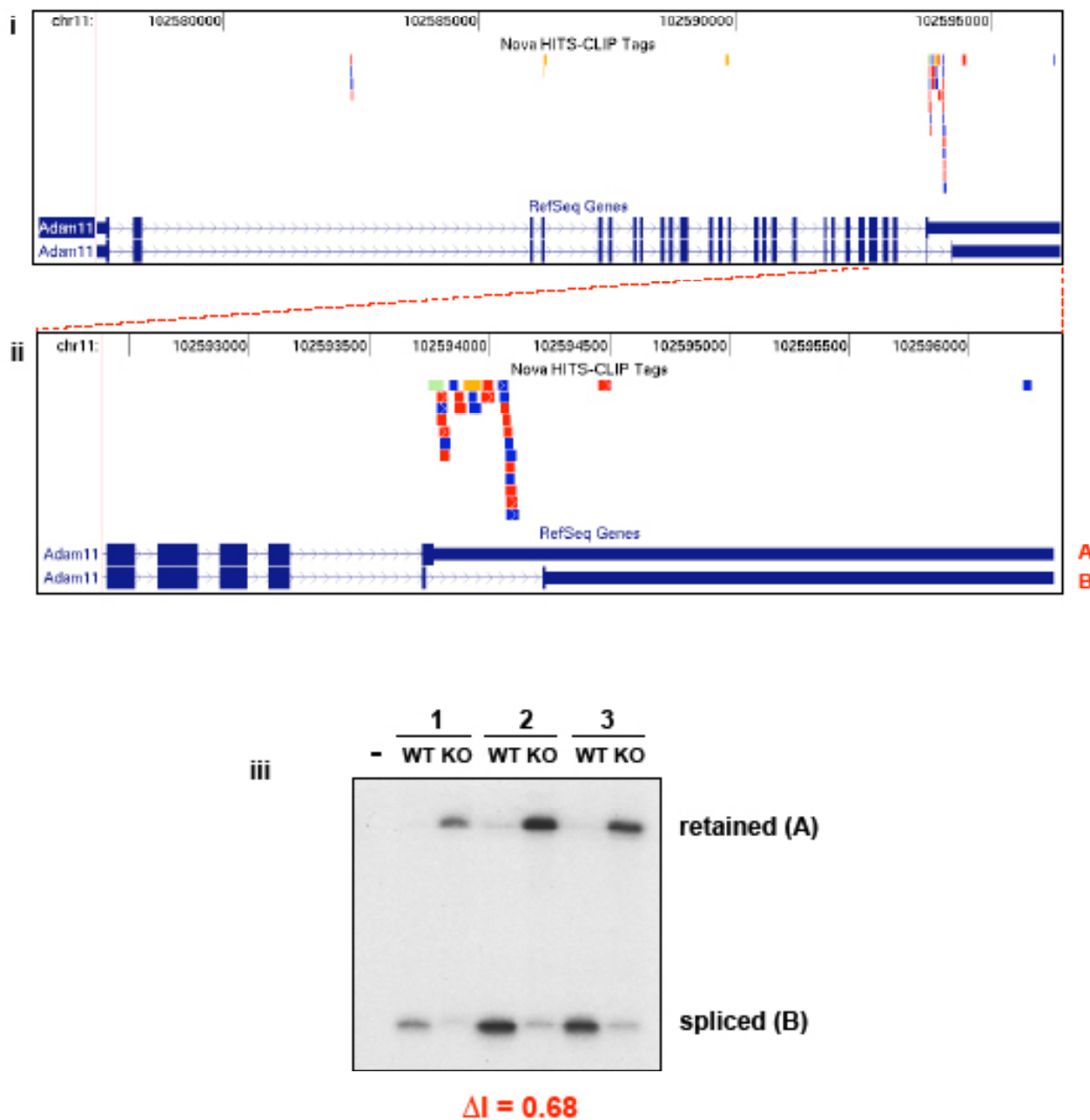
Supplemental Fig. 5. (a-f), Nova CLIP tags mapping to candidate Nova-regulated transcripts are shown in whole gene (i) and region-specific (ii) views which include alternatively spliced isoforms present in different databases. iii, RT-PCR analysis of cortex RNA from 3 pairs of P10 wild type and Nova2 KO littermates, including a RT- control (-). The change in exon inclusion (ΔI) is shown below, with red and blue text representing exons that are preferentially included or excluded in a Nova2-dependent manner respectively.

Figure 5d



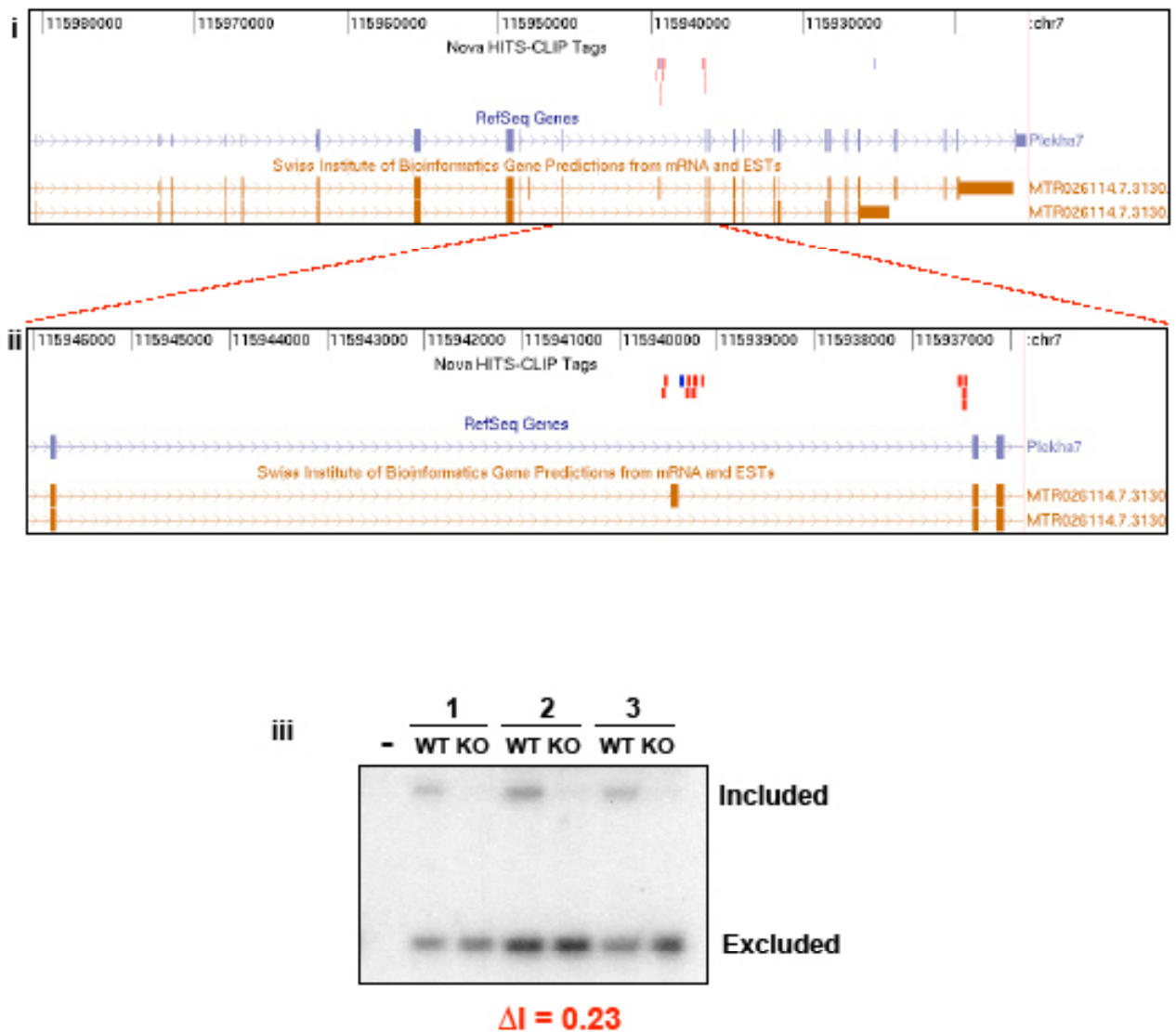
Supplemental Fig. 5. (a-f), Nova CLIP tags mapping to candidate Nova-regulated transcripts are shown in whole gene (i) and region-specific (ii) views which include alternatively spliced isoforms present in different databases. iii, RT-PCR analysis of cortex RNA from 3 pairs of P10 wild type and Nova2 KO littermates, including a RT- control (-). The change in exon inclusion (ΔI) is shown below, with red and blue text representing exons that are preferentially included or excluded in a Nova2-dependent manner respectively.

Figure 5e

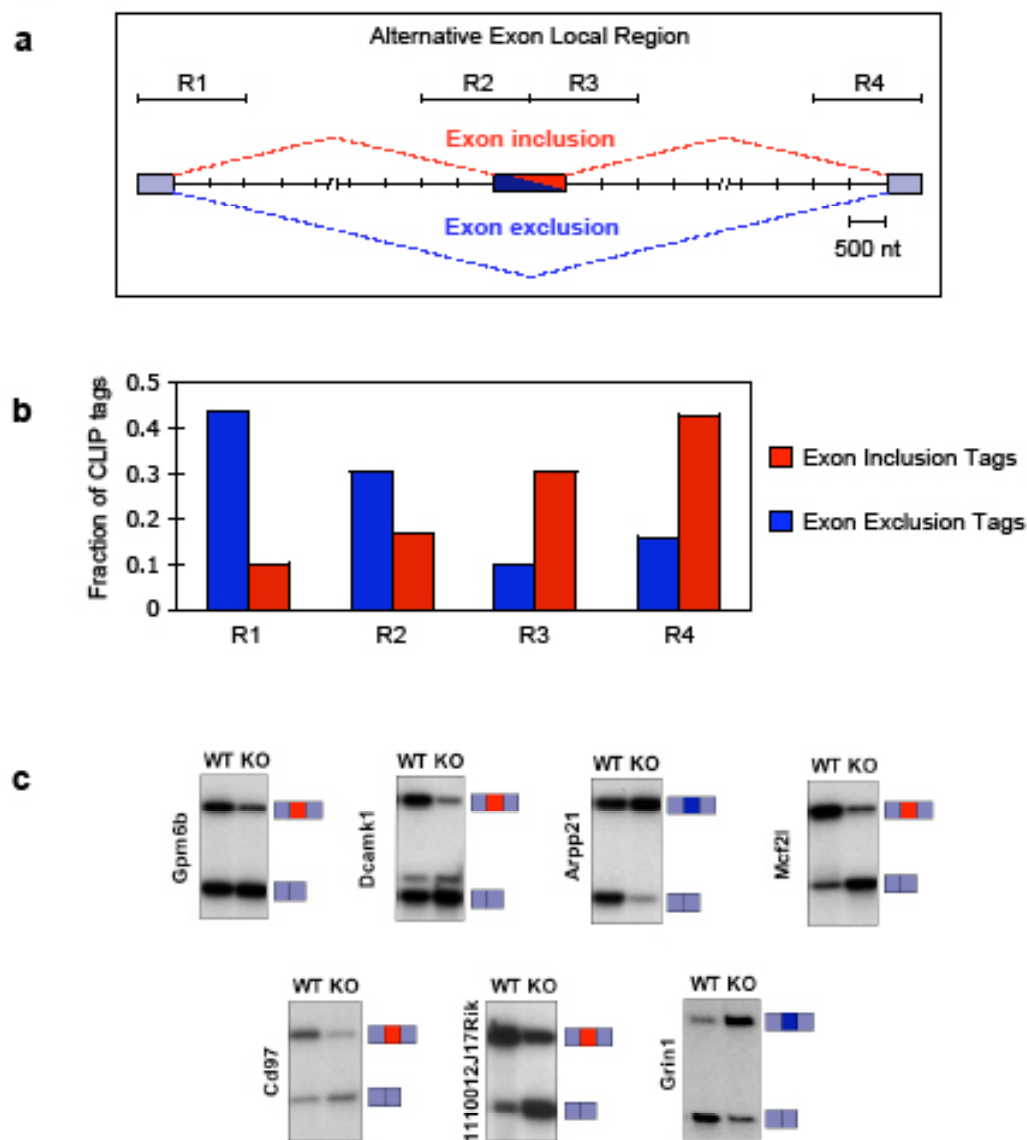


Supplemental Fig. 5. (a-f), Nova CLIP tags mapping to candidate Nova-regulated transcripts are shown in whole gene (i) and region-specific (ii) views which include alternatively spliced isoforms present in different databases. iii, RT-PCR analysis of cortex RNA from 3 pairs of P10 wild type and Nova2 KO littermates, including a RT- control (-). The change in exon inclusion (ΔI) is shown below, with red and blue text representing exons that are preferentially included or excluded in a Nova2-dependent manner respectively.

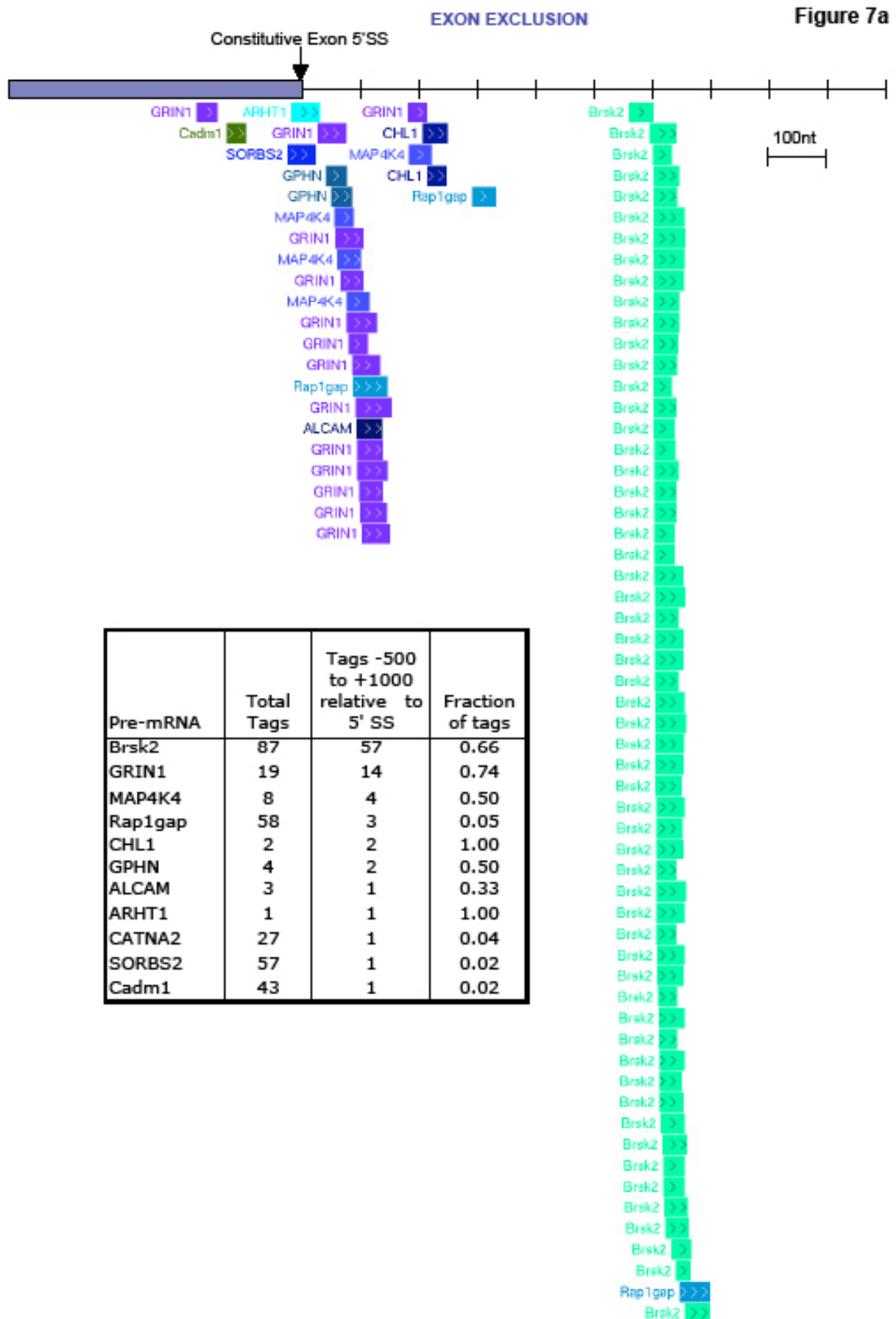
Figure 5f



Supplemental Fig. 5. (a-f), Nova CLIP tags mapping to candidate Nova-regulated transcripts are shown in whole gene (i) and region-specific (ii) views which include alternatively spliced isoforms present in different databases. iii, RT-PCR analysis of cortex RNA from 3 pairs of P10 wild type and Nova2 KO littermates, including a RT- control (-). The change in exon inclusion (ΔI) is shown below, with red and blue text representing exons that are preferentially included or excluded in a Nova2-dependent manner respectively.



Supplemental Fig. 6. **a**, Schematic of regions (R1 – R4) used to analyze CLIP tag abundance in new Nova-regulated exons in this Figure and in Supplemental Table 1. **b**, Chart illustrating the fraction of CLIP tags present in R1-4 broken down by tags present in Nova-dependent exon included or excluded transcripts. **c**, RT-PCR autoradiograms from analysis of 7 new candidate Nova-dependent alternate exons; for each, primers bounding the alternate exon were used to generate a larger (exon included) and smaller (exon excluded) isoform.



Supplemental Fig. 7. a-h, closeup views from Figure 2a of all CLIP tags present in regions R1-R4 from 71 Nova-dependent exon excluded (a-d) and included (e-h) alternatively spliced transcripts. In each, a Table is given summarizing CLIP tag data for each transcript.

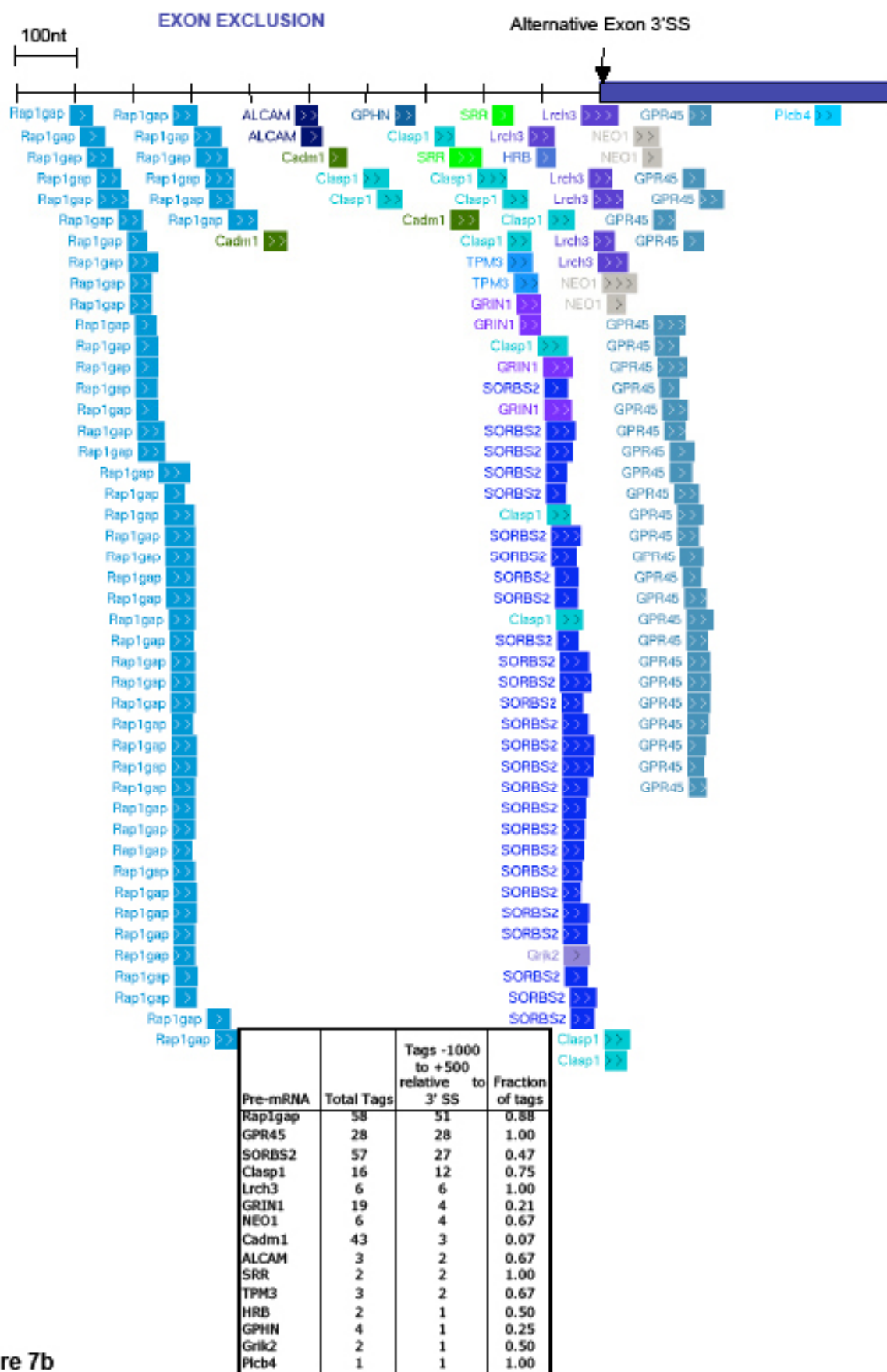


Figure 7b

Supplemental Fig. 7. a-h, closeup views from Figure 2a of all CLIP tags present in regions R1-R4 from 71 Nova-dependent exon excluded (a-d) and included (e-h) alternatively spliced transcripts. In each, a Table is given summarizing CLIP tag data for each transcript.

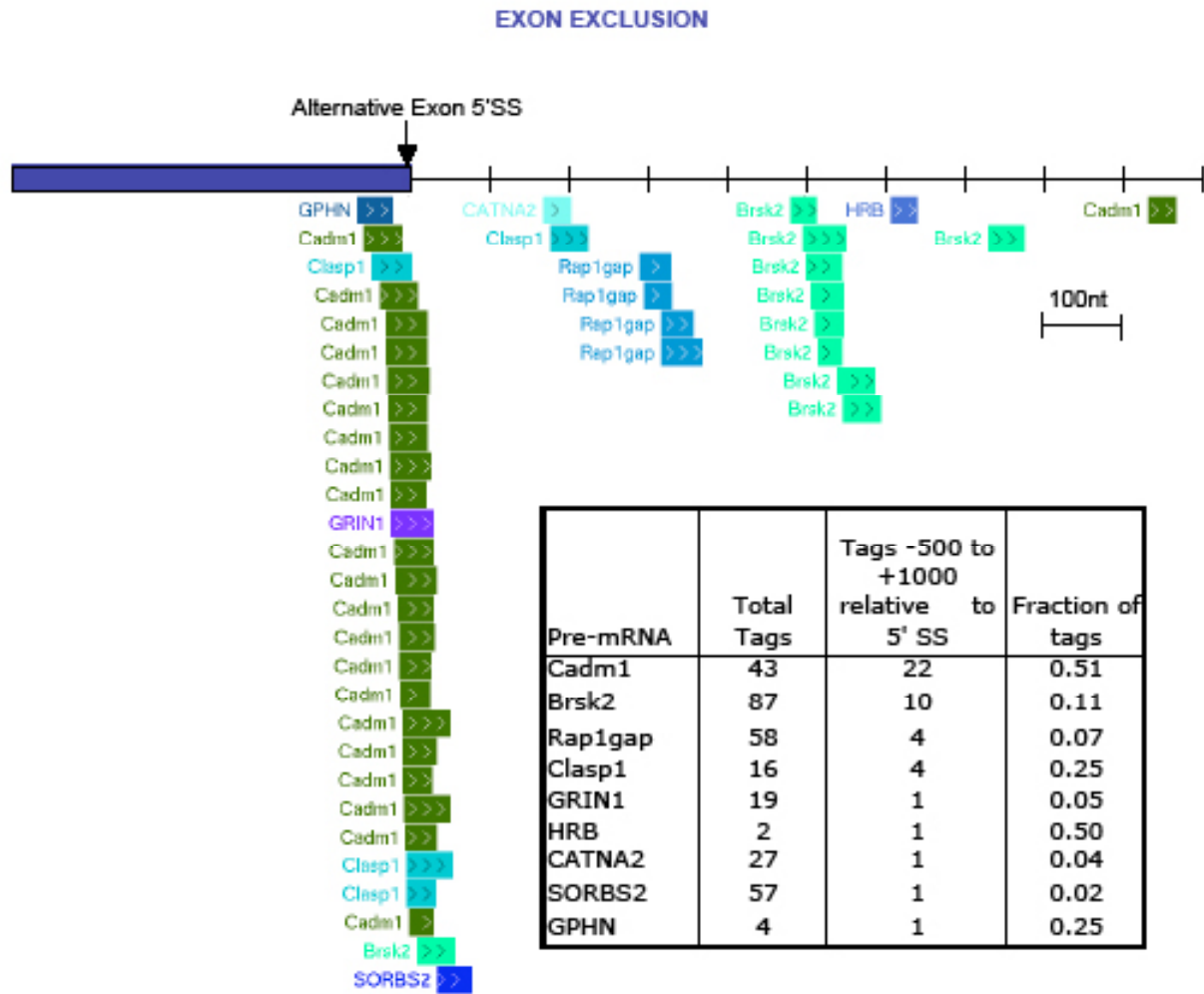


Figure 7c

Supplemental Fig. 7. a-h, closeup views from Figure 2a of all CLIP tags present in regions R1-R4 from 71 Nova-dependent exon excluded (a-d) and included (e-h) alternatively spliced transcripts. In each, a Table is given summarizing CLIP tag data for each transcript.

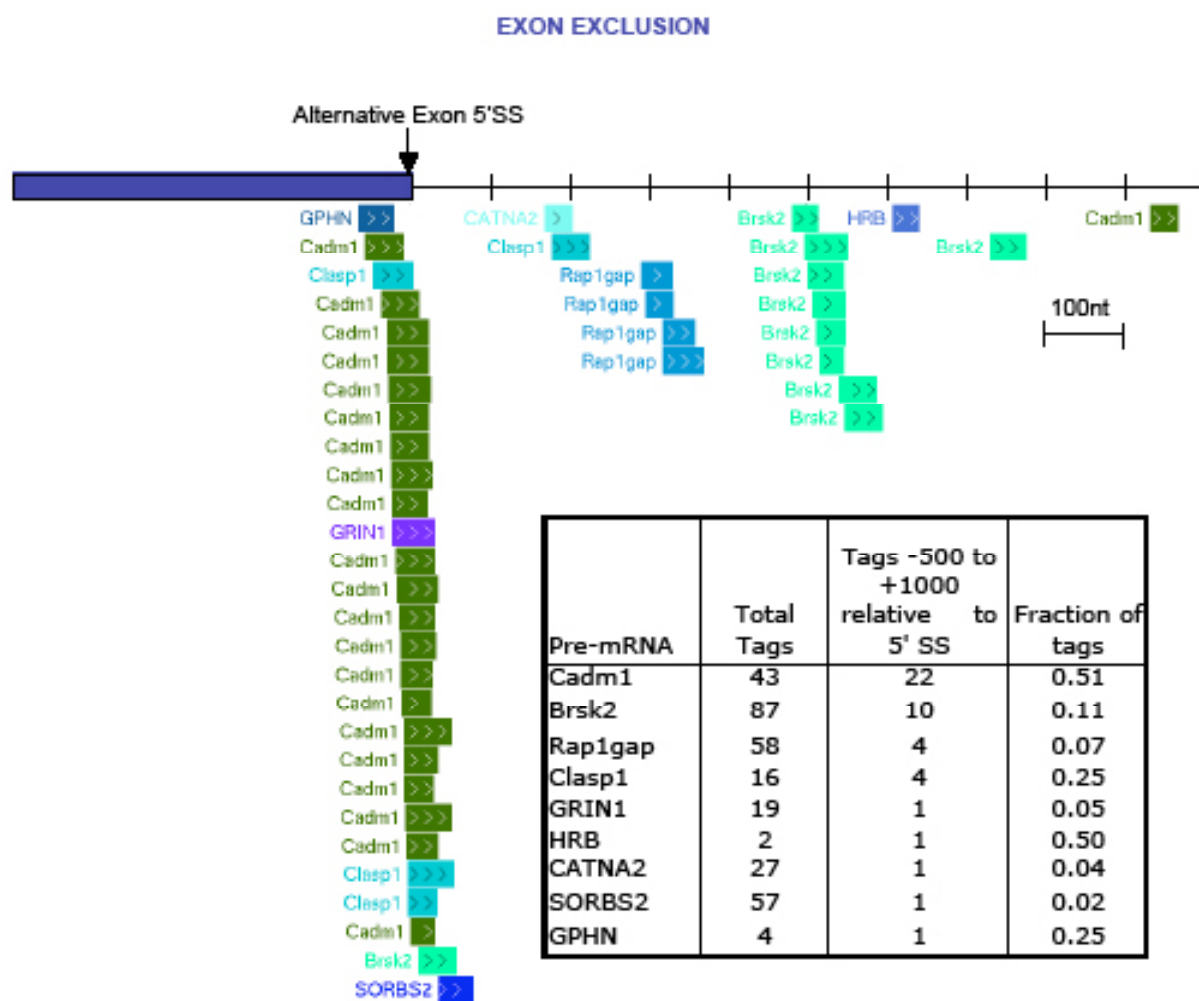


Figure 7c

Supplemental Fig. 7. a-h, closeup views from Figure 2a of all CLIP tags present in regions R1-R4 from 71 Nova-dependent exon excluded (a-d) and included (e-h) alternatively spliced transcripts. In each, a Table is given summarizing CLIP tag data for each transcript.

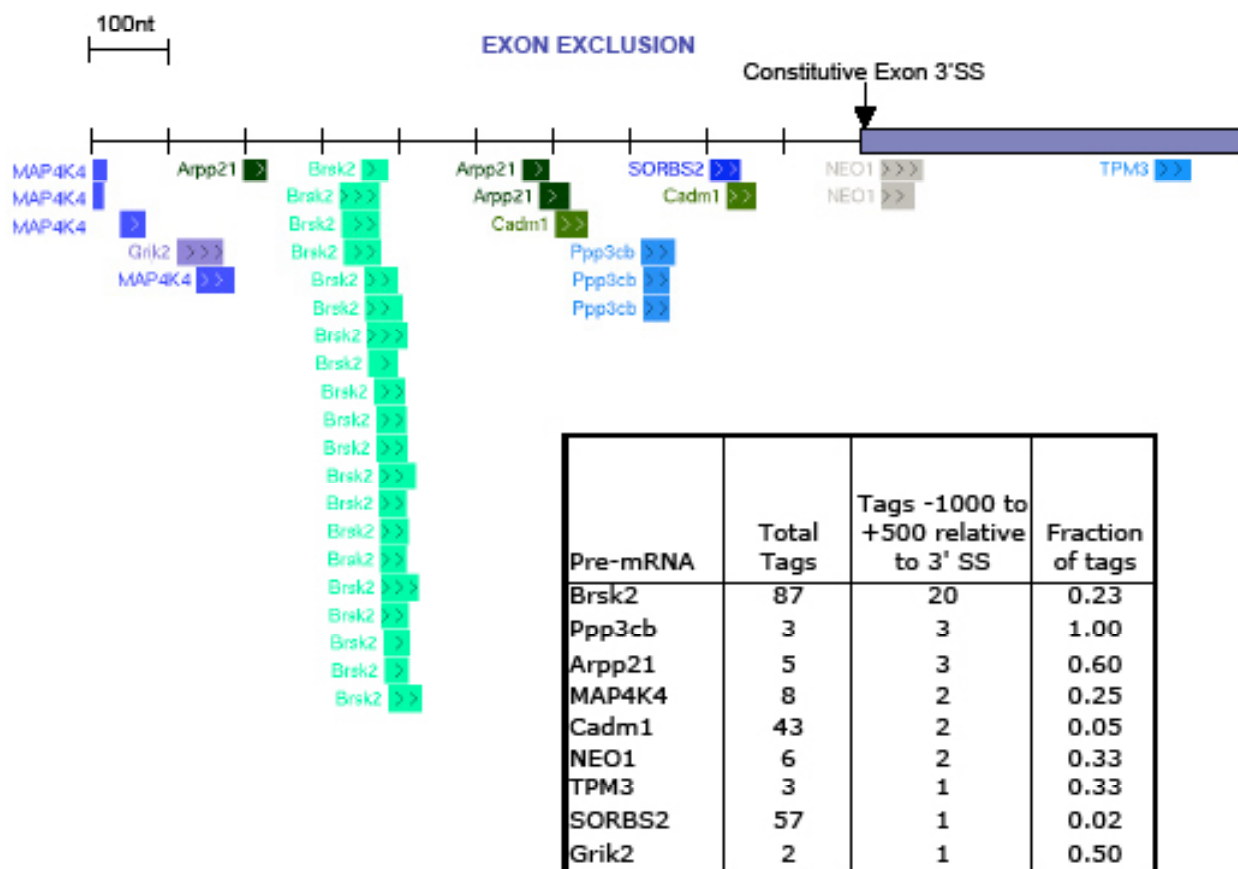


Figure 7d

Supplemental Fig. 7. a-h, closeup views from Figure 2a of all CLIP tags present in regions R1-R4 from 71 Nova-dependent exon excluded (a-d) and included (e-h) alternatively spliced transcripts. In each, a Table is given summarizing CLIP tag data for each transcript.

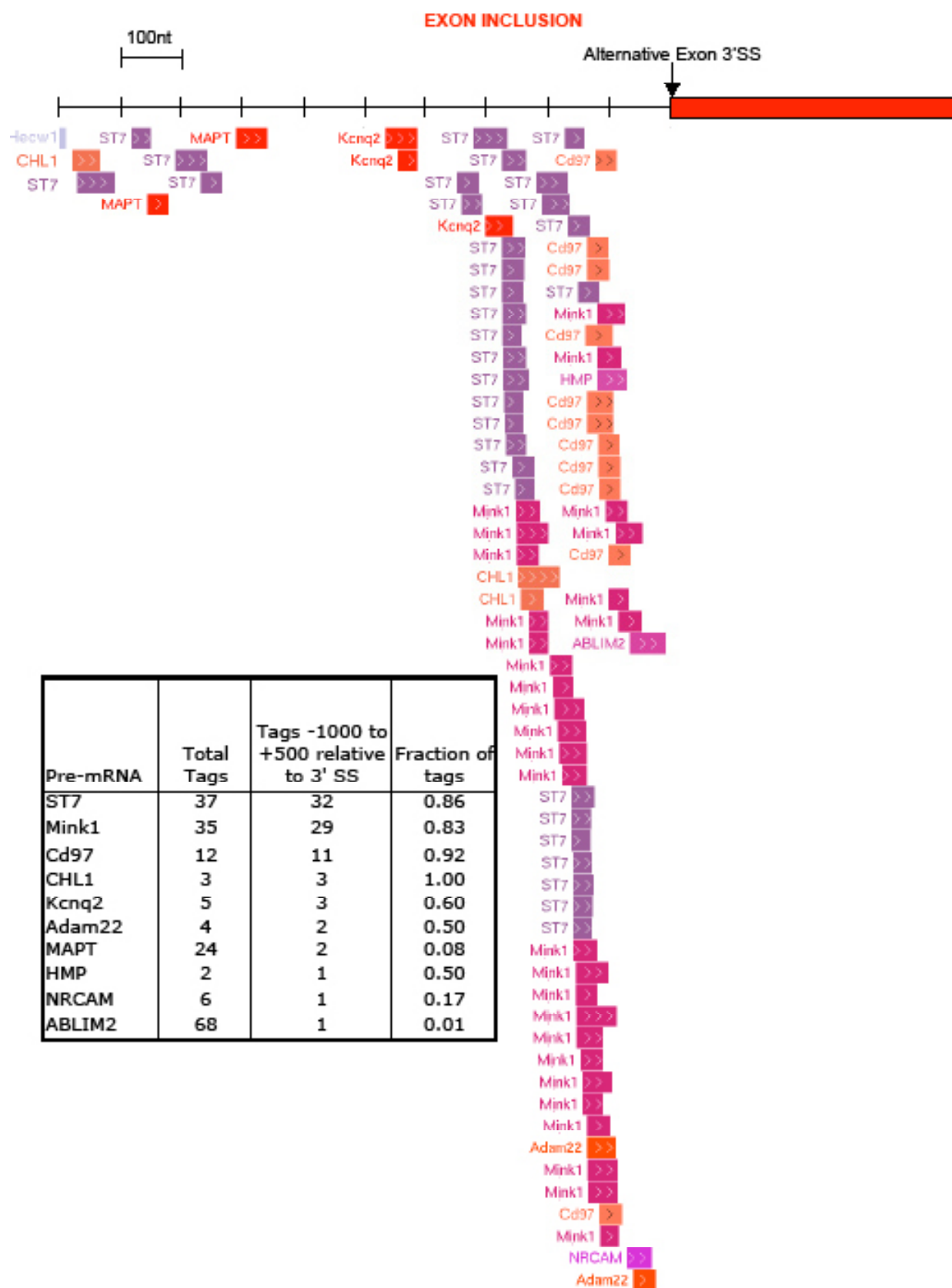


Figure 7f

Supplemental Fig. 7. a-h, closeup views from Figure 2a of all CLIP tags present in regions R1-R4 from 71 Nova-dependent exon excluded (a-d) and included (e-h) alternatively spliced transcripts. In each, a Table is given summarizing CLIP tag data for each transcript.

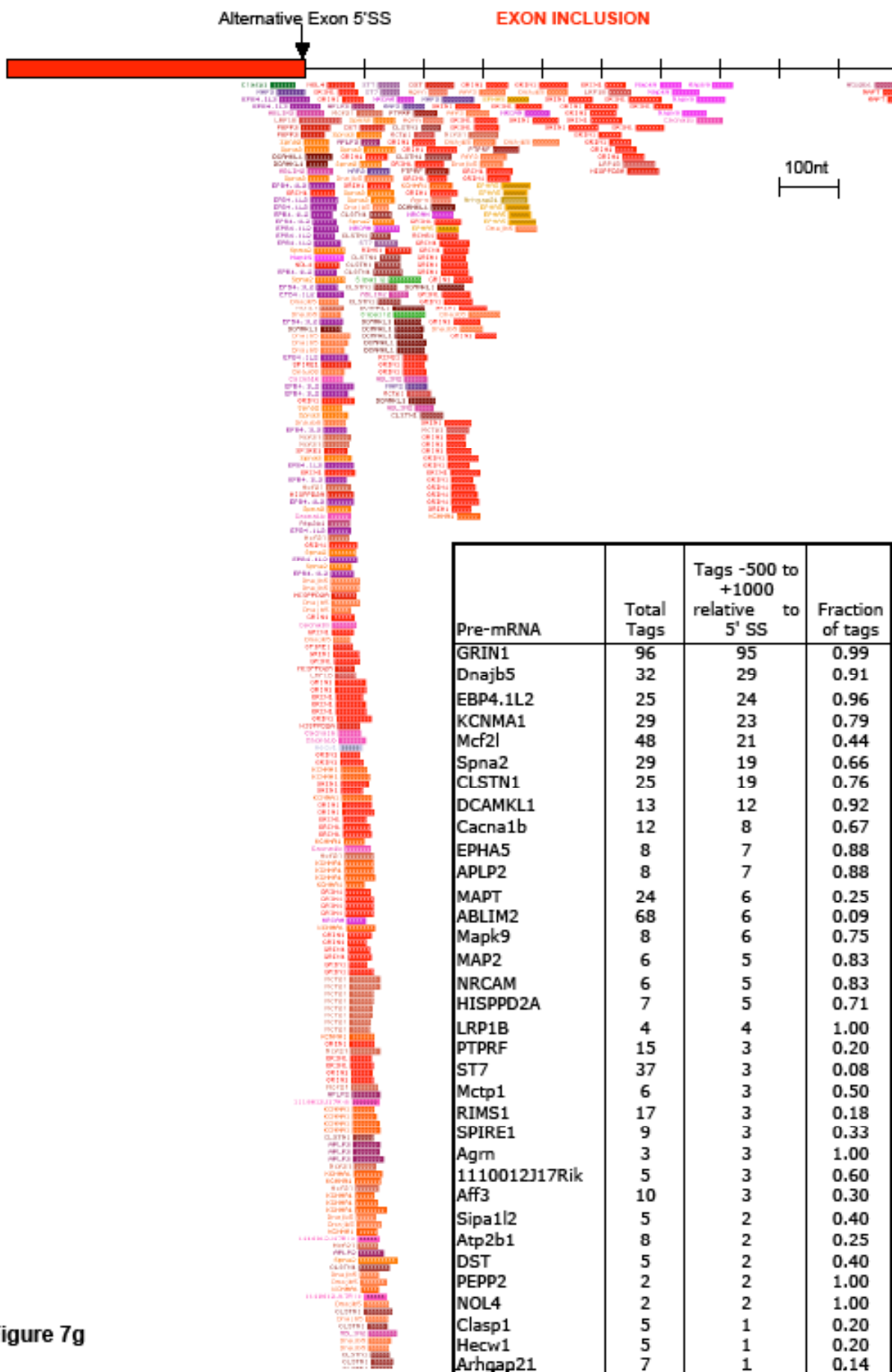


Figure 7g

Supplemental Fig. 7. a-h, closeup views from Figure 2a of all CLIP tags present in regions R1-R4 from 71 Nova-dependent exon excluded (a-d) and included (e-h) alternatively spliced transcripts. In each, a Table is given summarizing CLIP tag data for each transcript.

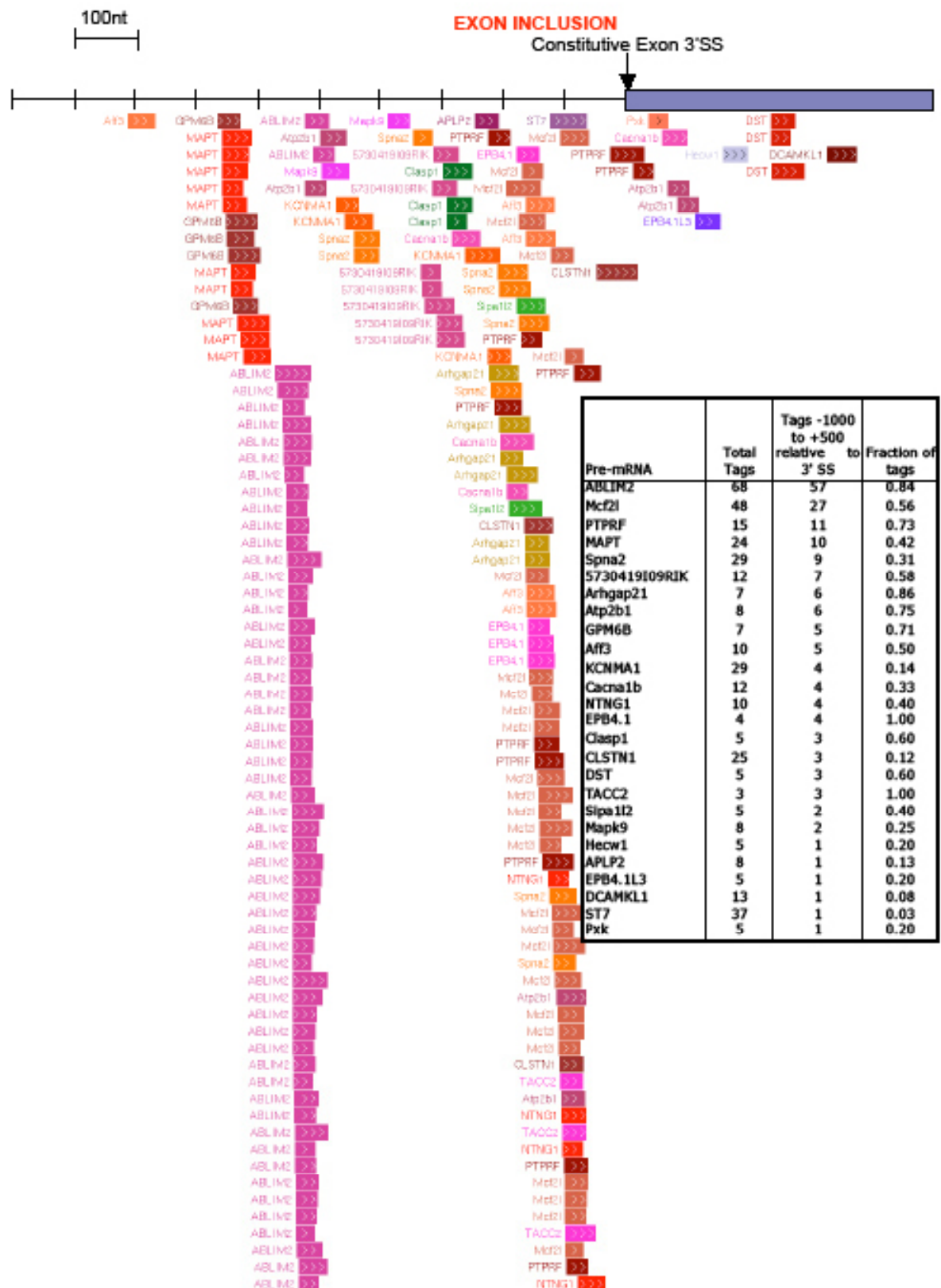
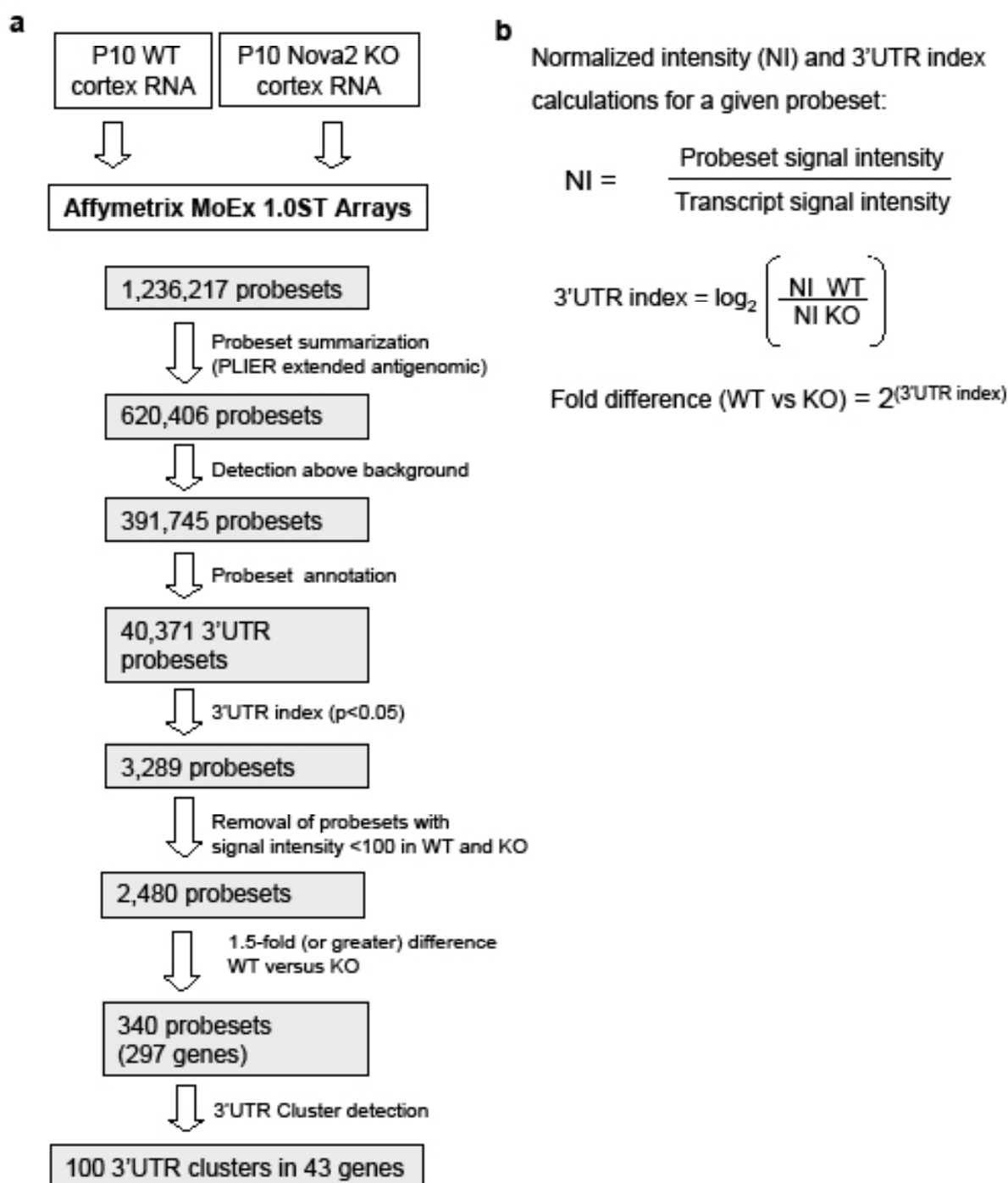


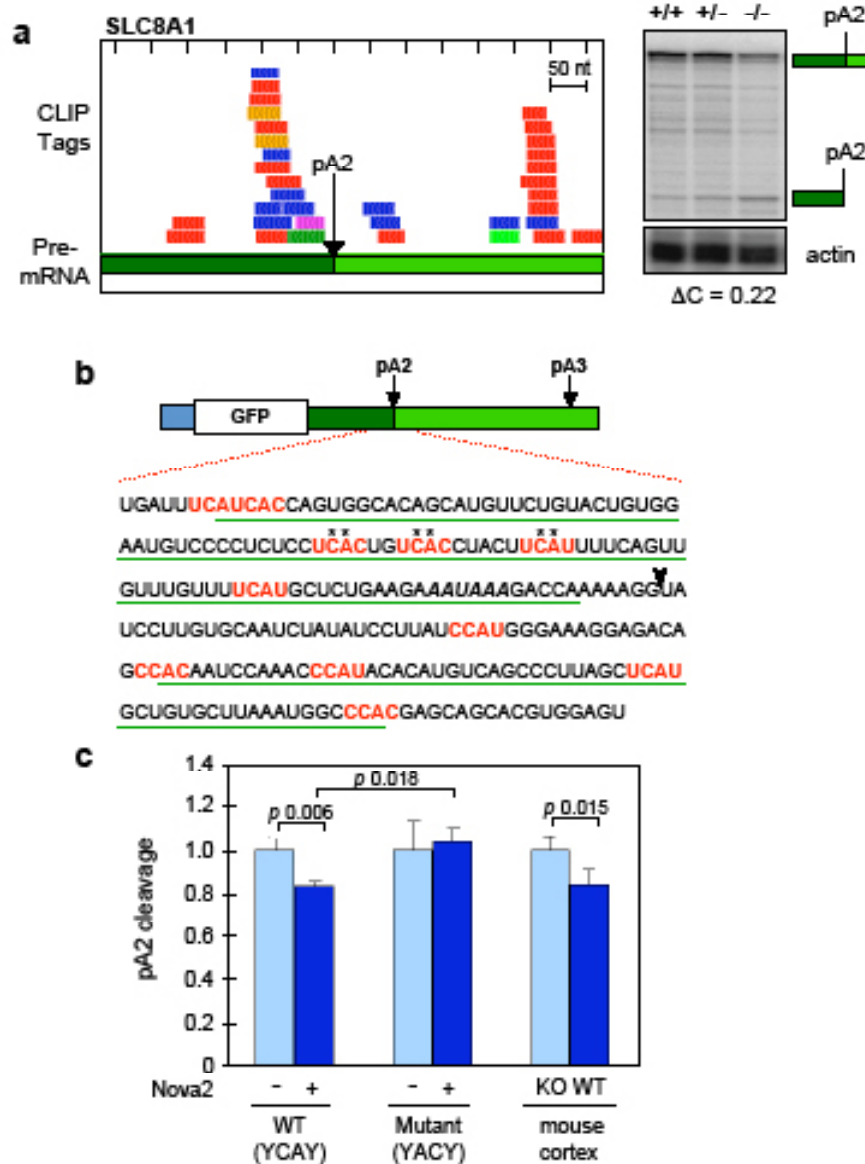
Figure 7h

Supplemental Fig. 7. a-h, closeup views from Figure 2a of all CLIP tags present in regions R1-R4 from 71 Nova-dependent exon excluded (a-d) and included (e-h) alternatively spliced transcripts. In each, a Table is given summarizing CLIP tag data for each transcript.

Figure 8



Supplemental Fig. 8. a, Flow chart outlining the use of Affymetrix Exon Arrays used to screen for alternative poly(A) site use. **b**, Overview of method used to generate the 3' UTR index.



Supplemental Fig. 9. Nova binds YCA Y elements to directly regulate alternative polyadenylation. **a**, Nova regulation of SLC8A1, otherwise as described in (4a). Nova HITS-CLIP tags clustered upstream and downstream of SLC8A1 poly(A) site 2 (pA2). Different colored bars represent tags from individual CLIP experiments. Right panel: RNase protection assay measuring cleavage at pA2 in WT (+/+), Nova2 heterozygous (+/-) and Nova2 KO (-/-) mouse P10 neocortex total RNA; an actin probe was used as a normalization control and DC given (as defined in the text). **b**, Cartoon of minigene construct used to assay Nova-dependent regulation of polyadenylation; a GFP reporter is present upstream of ~3kb of 3' UTR from the SLC8A1 transcript. The 240 bp sequence around the pA2 site (arrowhead) regulated in mouse brain is shown; underlined sequence represents footprint of clusters in this region; YCA Y elements are red; the CPSF binding site is in bold; asterisks represent sites where YCA Y elements were mutated to YACY. **c**, qRT-PCR results from cells transfected with either WT or mutant (3 YACY mutations; similar data was seen mutating a fourth binding site; data not shown) poly(A) reporter construct in the presence or absence of a cDNA expressing Nova2, or WT or Nova2 KO mouse cortex.

Supplemental References

- 1.** Ule, J. et al. An RNA map predicting Nova-dependent splicing regulation. *Nature* **444**, 580-586 (2006).

- 2.** 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* **95**, 13254-13259 (1998).

- 3.** Ule, J. et al. Nova regulates brain-specific splicing to shape the synapse. *Nat Genet* **37**, 844-852 (2005).

- 4.** Lee, J. A. et al. Depolarization and CaM kinase IV modulate NMDA receptor splicing through two essential RNA elements. *PLoS Biol* **5**, e40 (2007).

- 5.** Han, K., Yeo, G., An, P., Burge, C. B. & Grabowski, P. J. A combinatorial code for splicing silencing: UAGG and GGGG motifs. *PLoS Biol* **3**, e158 (2005).

Supplemental Table 1

RefSeq ID	Gene Symbol	Exon	Chr	Exon Start	Exon End	ΔI	ΔI Test	RT-PCR	Local Tags	Cassette Exon	Total Local Cassette Tags	R1	R2	R3	R4	Nova-dependent
NM_009655	Alcam	E11	chr16	52188138	52188176	-0.22	-123.67		Y	Y	3	1	2	0	0	Exclusion
NM_028755	Arpp21	E11	chr9	111992312	111992371	-0.32	-169.92	-0.38	Y	Y	5	0	0	0	3	Exclusion
NM_001009930	Brsk2	E19	chr7	141810232	141810279	-0.29	-41.34		Y	Y	87	57	0	10	20	Exclusion
NM_001081276	Clasp1	E20	chr1	120363681	120363743	-0.32	-31.84		Y	Y	16	0	12	4	0	Exclusion
NM_008169	Grin1	E3	chr2	25133352	25133414	-0.33	-151.16	-0.35	Y	Y	19	14	4	1	0	Exclusion
NM_010472	Hrb	E6	chr1	82759032	82759151	-0.22	-182.06		Y	Y	2	0	1	1	0	Exclusion
NM_001081255	Lrch3	E16	chr16	32915553	32915660	-0.31	-27.82		Y	Y	6	0	6	0	0	Exclusion
NM_001081219	Myo1a	E20	chr10	127120247	127120310	-0.21	-36.3		N	N						Exclusion
NM_008914	Ppp3cb	E10	chr14	19300026	19300052	-0.33	-308.19		Y	Y	3	0	0	0	3	Exclusion
NM_021536	Arht1	E18	chr11	80073709	80073831	-0.19	-50.39		Y	Y	1	1	0	0	0	Exclusion
NM_021544	Scn5a	E5	chr9	119391797	119391888	-0.2	-25.33		N	N						Exclusion
NM_020618	Smarce1	E3	chr11	99041128	99041232	-0.27	-61.17		Y	Y	3	0	0	0	0	Exclusion
NM_178252	Snx26	E1	chr7	30242945	30242972	-0.41	-47.8		Y	N						Exclusion
NM_172752	Sorbs2	E11	chr8	47284086	47284142	-0.26	-75.44		Y	Y	57	1	27	1	1	Exclusion
NM_029081	5730419109Rik	E22	chr6	142977546	142977596	0.21	42.34		Y	Y	12	2	0	0	7	Inclusion
NM_177678	Abliim2	E10	chr5	36156651	36156683	0.33	313.27		Y	Y	68	1	1	6	57	Inclusion
NM_001007221	Adam22	E25	chr5	8101079	8101186	0.19	25.35		Y	Y	4	0	2	0	0	Inclusion
NM_010678	Aff3	E2	chr1	38483473	38483547	0.35	54.85		Y	Y	10	0	0	3	5	Inclusion
NM_194341	Ap1gbp1	E17	chr11	83855851	83855886	0.21	139.69	0.36	N	N						Inclusion
NM_019978	Dcamk11	E15	chr3	55609779	55609852	0.21	330.21		Y	Y	13	0	0	12	1	Inclusion
NM_133833	Dst	E88	chr1	34246407	34246478	0.23	110.34		Y	Y	5	0	0	2	3	Inclusion
NM_199012	Fchs2	E11	chr7	101120389	101120460	0.22	31.84		N	N						Inclusion
NM_023122	Gpm6b	E1	chrX	161697670	161697789	0.26	127.02		Y	Y	7	0	0	0	5	Inclusion
NM_001081348	Hecw1	E1	chr13	14226560	14226617	0.29	159.08		Y	Y	5	0	0	1	1	Inclusion
NM_173760	Hisppd1	E24	chr1	99551509	99551685	0.22	40.24		N	N						Inclusion
NM_178795	Hisppd2a	E26	chr2	121018293	121018355	0.26	31.32		Y	Y	7	0	0	5	0	Inclusion
NM_029673	Hmp	E3	chr6	71782682	71782777	0.21	50.6		Y	Y	2	0	1	0	0	Inclusion
NM_010838	Mapt	E8	chr11	104134247	104134339	0.2	51.11		Y	Y	24	2	2	6	10	Inclusion
NM_178076	Mcf2l	E1	chr8	12949436	12949653	0.2	25.62		Y	N						Inclusion
NM_178076	Mcf2l	E26	chr8	13013892	13013984	0.52	65.21	0.6	Y	Y	48	0	0	21	27	Inclusion
NM_030174	Mctp1	E6	chr13	77195318	77195356	0.39	52.17		Y	Y	6	1	0	3	0	Inclusion
NM_001045959	Mink1	E14	chr11	70422826	70422912	0.22	118.71		Y	Y	35	6	29	0	0	Inclusion
NM_008632	Map2	E8	chr1	66349940	66350074	0.23	105.43		Y	Y	6	1	0	5	0	Inclusion
NM_175089	Nek1	E16	chr8	63942170	63942301	0.33	30.31		N	N						Inclusion
NM_199024	Nol4	E8	chr18	22912754	22912945	0.41	71.33		Y	Y	2	0	0	2	0	Inclusion
NM_176930	Nrcam	E6	chr12	45415665	45415721	0.6	439.12		Y	Y	6	0	1	5	0	Inclusion
NM_144920	Pepp2	E10	chr6	140542753	140542941	0.28	96.56		Y	Y	2	0	0	2	0	Inclusion
NM_145458	Pxx	E17	chr14	6951534	6951566	0.45	51.57		Y	Y	5	1	0	0	1	Inclusion
NM_053270	Rims1	E17	chr1	22383750	22383932	0.26	28.71		Y	Y	17	12	0	3	0	Inclusion
NM_009159	Sfrs5	E5	chr12	81867473	81867510	0.21	76.73		N	N						Inclusion
NM_001081337	Sipa1l2	E20	chr8	128309283	128309336	0.39	53.24		Y	Y	5	1	0	2	2	Inclusion
NM_001085371	Speg	E3	chr1	75268135	75268312	0.27	26.62		Y	N						Inclusion
NM_194355	Spire1	E7	chr18	67638210	67638248	0.24	47.55		Y	Y	9	0	0	3	0	Inclusion
NM_001081242	Tln2	E30	chr9	67026207	67026252	0.29	44.98		Y	N						Inclusion
NM_011925	Cd97	E4	chr8	86622008	86622142	0.37	4.16	0.47	Y	Y	12	1	11	0	0	Inclusion
NM_172963	1110012J17Rik	E11	chr17	66245417	66245486	0.33	3.58	0.49	Y	Y	5	0	0	3	0	Inclusion

Supplemental Table 1. Table enumerating the top newly identified 46 Nova-dependent alternate exon candidates using the Research Exon Junction Array and ASPIRE2, along with delineation of ΔI, ΔI-tTest (see Supplemental Methods), the presence of CLIP tags within the alternative exon local regions (see Supplemental Fig. 6).

Supplemental Table 2

GO Miner analysis: GO Category Cellular Component

GO Cellular Component	TOTAL GENES	CHANGED EXONS	CHANGED GENES	ENRICHMENT	LOG10(p)	FALSE DISCOVERY RATE	GENE NAMES
GO:0042995 cell projection	271	19	16	8.47	-10.67	<0.001	ALCAM, ANK3, CADM1, CHL1*, CLSTN1, CTNNA2, GRIN1*, HRB, KCNMA1, MAPT, MCF2L*, MTAP2, MYO1A, NRCAM, TLN2, TPM3
GO:0043005 neuron projection	108	12	10	13.28	-8.58	<0.001	ALCAM, ANK3, CADM1, CHL1*, GRIN1*, KCNMA1, MAPT, MTAP2, NRCAM, TPM3
GO:0005856 cytoskeleton	520	18	16	4.41	-6.57	<0.001	ABLIM2, ARHGAP21, CLASP1*, CTNNA2, DST, EPB4.1L2, EPB4.1L3, GPHN, GRIN1*, MAPT, MTAP2, MYO1A, SPIRE1, SPNA2, TLN2, TPM3
GO:0030424 axon	60	6	6	14.35	-5.49	<0.001	ALCAM, ANK3, CADM1, KCNMA1, MAPT, NRCAM
GO:0030054 cell junction	251	11	10	5.72	-5.13	<0.001	ARHGAP21, CADM1, CTNNA2, DST, GPHN, GRIN1*, RIMS1, SCN5A, SPNA2, TLN2
GO:0045202 synapse	182	9	8	6.31	-4.48	<0.001	ANK3, CADM1, CLSTN1, GPHN, GRIN1*, KCNMA1, RIMS1, TLN2
GO:0005886 plasma membrane	903	18	17	2.7	-4.07	0.003	ALCAM, ARHGAP21, ATP2B1, CADM1, CTNNA2, DST, GPHN, GRIN1*, KCNMA1, MYO1A, PTPRF, PXX, RIMS1, SCN5A, SNX26, SPNA2, TLN2
GO:0044459 plasma membrane part	659	15	14	3.05	-3.92	0.004	ALCAM, ARHGAP21, ATP2B1, CADM1, CTNNA2, DST, GPHN, GRIN1*, KCNMA1, MYO1A, RIMS1, SCN5A, SPNA2, TLN2
GO:0043228 non-membrane-bounded organelle	933	19	17	2.61	-3.89	0.003	ABLIM2, ARHGAP21, CLASP1*, CTNNA2, DST, EPB4.1L2, EPB4.1L3, GPHN, GRIN1*, MAPT, MTAP2, MYO1A, SMARCE1, SPIRE1, SPNA2, TLN2, TPM3
GO:0043232 intracellular non-membrane-bounded organelle	933	19	17	2.61	-3.89	0.003	ABLIM2, ARHGAP21, CLASP1*, CTNNA2, DST, EPB4.1L2, EPB4.1L3, GPHN, GRIN1*, MAPT, MTAP2, MYO1A, SMARCE1, SPIRE1, SPNA2, TLN2, TPM3
GO:0015629 actin cytoskeleton	119	6	6	7.23	-3.78	0.003	ABLIM2, CTNNA2, DST, EPB4.1L2, MYO1A, TPM3
GO:0030425 dendrite	45	6	4	12.75	-3.59	0.003	CADM1, CHL1*, GRIN1*, MTAP2
GO:0016323 basolateral plasma membrane	55	4	4	10.43	-3.26	0.006	CADM1, DST, MYO1A, TLN2
GO:0009925 basal plasma membrane	8	2	2	35.87	-2.89	0.011	DST, MYO1A
GO:0045178 basal part of cell	9	2	2	31.88	-2.78	0.011	DST, MYO1A
GO:0045211 postsynaptic membrane	76	5	4	7.55	-2.73	0.011	CLSTN1, GPHN, GRIN1*, KCNMA1
GO:0044463 cell projection part	41	4	3	10.5	-2.55	0.016	GRIN1*, KCNMA1, MTAP2
GO:0044456 synapse part	86	5	4	6.67	-2.53	0.015	CLSTN1, GPHN, GRIN1*, KCNMA1
GO:0005912 adherens junction	42	3	3	10.25	-2.52	0.016	CTNNA2, SPNA2, TLN2
GO:0045177 apical part of cell	42	4	3	10.25	-2.52	0.016	CHL1*, KCNMA1, MYO1A
GO:0030016 myofibril	44	3	3	9.78	-2.46	0.014	ABLIM2, SPNA2, TPM3
GO:0043292 contractile fiber	44	3	3	9.78	-2.46	0.014	ABLIM2, SPNA2, TPM3
GO:0030426 growth cone	15	3	2	19.13	-2.33	0.018	GRIN1*, TPM3
GO:0030427 site of polarized growth	15	3	2	19.13	-2.33	0.018	GRIN1*, TPM3
GO:0005913 cell-cell adherens junction	16	2	2	17.93	-2.27	0.022	CTNNA2, SPNA2
GO:0019898 extrinsic to membrane	58	3	3	7.42	-2.12	0.029	EPB4.1L2, EPB4.1L3, GPHN
GO:0014704 intercalated disc	21	2	2	13.66	-2.04	0.035	SCN5A, SPNA2
GO:0031252 leading edge	63	4	3	6.83	-2.02	0.036	CTNNA2, MCF2L*, TLN2
GO:0016020 membrane	3430	36	33	1.38	-2.01	0.034	ADAM22, ALCAM, AP1GBBP1, APLP2, ARHGAP21, ARHT1, ATP2B1, CADM1, CHL1*, CLSTN1, CTNNA2, DST, EPB4.1L2, EPB4.1L3, EPHA5, GPHN, GPM6B, GRIN1*, IMMT, KCNMA1, LRP1B, MAPT, MCF2L*, MCTP1, MYO1A, NRCAM, PTPRF, PXX, RIMS1, SCN5A, SNX26, SPNA2, TLN2
GO:0016324 apical plasma membrane	24	2	2	11.96	-1.92	0.040	KCNMA1, MYO1A
GO:0005911 intercellular junction	71	3	3	6.06	-1.88	0.042	CADM1, CTNNA2, SPNA2

* denotes gene with multiple Nova-regulated exons

Supplemental Table 2. GOMiner analysis of genes containing Nova2-regulated alternative exons identified by ASPIRE2. All Nova-regulated genes identified by ASPIRE2 ($|DI| > 0.19$, $|DI-tTest| > 25$) were analyzed by GO Miner (discover.nci.nih.gov/gominer/). Comparisons were made with a list of all genes expressed at P10 in cortex as determined experimentally by exon array analysis (data not shown). All results with a false discovery rate less than 0.05 are shown. Gene names of previously validated Nova regulated transcripts (using ASPIRE) are in bold. There is essentially complete overlap in GO categories of Nova regulated transcripts previously identified and those newly identified by ASPIRE2; Nova regulates alternative exons in transcripts encoding synaptic proteins, consistent with previous observations¹.

Supplemental References

1. Ule, J. et al. Nova regulates brain-specific splicing to shape the synapse. *Nat Genet* **37**, 844-852 (2005).