ADAR (adenosine deaminase that acts on RNA) editing enzymes target coding and noncoding double-stranded RNA (dsRNA) and are essential for neuronal function. Early studies showed that ADARs preferentially target adenosines with certain 5′ and 3′ neighbours. Here we use current Sanger sequencing protocols to perform a more accurate and quantitative analysis. We quantified editing sites in an ~800-bp dsRNA after reaction with human ADAR1 or ADAR2, or their catalytic domains alone. These large data sets revealed that neighbour preferences are mostly dictated by the catalytic domain, but ADAR2’s dsRNA-binding motifs contribute to 3′ neighbour preferences. For all proteins, the 5′ nearest neighbour was most influential, but adjacent bases also affected editing site choice. We developed algorithms to predict editing sites in dsRNA of any sequence, and provide a web-based application. The predictive power of the algorithm on fully base-paired dsRNA, compared with biological substrates containing mismatches, bulges and loops, elucidates structural contributions to editing specificity.
denosine deaminases that act on RNAs (ADARs) convert adenosines to inosines (A-to-I) in double-stranded regions of viral RNAs, and cellular pre-mRNAs and noncoding RNAs\(^1\). There are thousands of A-to-I editing sites in the human transcriptome\(^2\), in coding and noncoding regions of mRNAs\(^3\). When ADARs target codons they can profoundly affect the proteome. For example, 24 isoforms are possible through varying combinations of editing in 5-HT\(_{1c}\) serotonin receptor pre-mRNA\(^4\). Aberrant editing is linked to depression and suicide\(^5\), cancer\(^6\), and further, ADARs can modulate double-stranded RNA (dsRNA)-mediated gene silencing pathways\(^7\).\(^8\).

Amino (N)-terminal regions of ADARs contain dsRNA-binding motifs (dsRBMs), whereas carboxy (C) termini contain a conserved catalytic domain. A crystal structure of the catalytic domain of human ADAR2 (hADAR2) has been solved\(^9\), as has the nuclear magnetic resonance solution structure of the two dsRBMs of rat ADAR2, in the presence or absence of dsRNA\(^10\).\(^11\).

ADARs target dsRNA of any sequence, but have preferences for certain neighbouring nucleotides. Analyses of Xenopus laevis ADAR1 show a 5’ nearest neighbour preference (Unementhe number of adenosines edited in a dsRNA is affected by dsRNA length and whether base-pairing is interrupted by mismatches, bulges or loops\(^12\). Editing of an AU base pair (bp) creates an IU mismatch, and selectivity is thought to relate to how many mismatches a dsRNA can tolerate before becoming too single stranded to be recognized by an ADAR. In all, 50–60% of adenosines in dsRNAs longer than ~50 bp can be edited before the reaction stops, whereas shorter dsRNAs are edited more selectively, at fewer sites. Internal loops can uncouple helices to turn a long dsRNA into a series of short dsRNAs that are edited more selectively\(^13\). Current paradigms hold that dsRBMs mediate selectivity\(^14\).

Here we use optimized methodology to refine and quantify neighbour preferences of human ADAR1 and ADAR2. Further, by evaluating neighbour preferences of truncated proteins, we determine contributions of the catalytic domain separately from dsRBMs. Using data from in vitro editing of a long perfectly base-paired dsRNA, we develop algorithms for predicting editing sites and provide a web-based programme (http://www.biochem.utah.edu/bass/inosinepredict). Using this algorithm we evaluate the importance of bases beyond nearest neighbours and contributions of RNA structure.

### Results

**Quantification by peak height is relatively accurate.** DNA sequencing data are often reported in Applied Biosystems trace files (‘abi’ chromatograms). Traces from cDNAs of ADAR-edited RNA have been considered to be unquantifiable\(^15\), as earlier dye terminator chemistry resulted in non-uniform peak heights. Advances in chemistry have improved peak-height uniformity\(^16\), but there has been no evaluation of newer outputs to determine adequacy for quantifying editing.

To this end, we mixed PCR products representing unedited or edited sequence at known ratios to create a mixture with a defined percentage of edited sequences (see Methods). The mixture was sequenced and chromatograms were quantified by measuring T and C peak heights in strands opposing the edited strand because A/G mixed peaks have more inconsistent heights\(^17\). The percent of the population edited at each site evaluated in the chromatogram was compared with the known ratio of unedited to edited sequences, or ‘true % editing’, in the prepared mixture (Table 1). The least accurate measurements for the 15 sites were those for the 60% edited mixture, which on average was low by 8% (average 52.3 ± 4.5); measuring peak heights rather than volumes gave the least variability (see Supplementary Table S1). The coefficient of variation (ratio of standard deviation to mean) increased at lower % editing (Table 1), and here our methodology did not distinguish between large relative differences that corresponded to small absolute differences (for example, we cannot reliably distinguish the twofold relative difference between 1 and 2% editing). Regardless, the nuclease mapping method previously used to determine ADAR preferences has a standard deviation of 12%, and the more qualitative primer extension method has up to 25% inaccuracy in % editing predicted for each site\(^18\). Thus, the more uniform peak heights associated with current four-dye trace chemistry allowed measurements that were more accurate and precise than previous techniques.

### ADAR nearest neighbour preferences.

Having established that measurements of peak-heights improved accuracy and precision, we used the methodology to analyse neighbour preferences of hADAR1 and hADAR2. We also investigated the contribution of dsRBMs to neighbour preferences, using truncated proteins consisting only of the catalytic domain (hADARD1–D and hADARD2–D).

Titrations were performed to determine the ADAR concentration that gave ~20% overall A-to-I conversion for an internally radiolabelled, 795-bp dsRNA, in 1 h at 30°C. With this % editing, few sites were edited to 100% in the population, ensuring that information was not lost due to saturation. These concentrations were then used in the ADAR preference assay (see Methods), in which non-radiolabelled 795-bp dsRNA was incubated with an ADAR, RNA products purified, and reverse transcribed and amplified with the PCR. PCR products were sequenced, and traces evaluated to determine the percentage of each adenosine edited in the population. These data were used to evaluate neighbour preferences using a binary or quantitative approach.

**Binary approach.** Four-dye sequence traces of cDNA derived from ADAR products have previously been evaluated qualitatively to provide a binary scale of editing within an RNA population. That is,

### Table 1 | True versus measured editing.

| % edited | Measured % edited*
|----------|------------------|
| 0        | 0.04±0.14        |
| 1        | 0.48±0.92        |
| 2        | 0.77±0.99        |
| 5        | 1.80±1.64        |
| 7        | 3.98±2.10        |
| 10       | 6.37±2.89        |
| 15       | 12.59±2.70       |
| 20       | 16.16±3.71       |
| 30       | 25.98±3.49       |
| 40       | 35.70±3.70       |
| 50       | 45.24±4.13       |
| 60       | 52.32±4.51       |
| 70       | 65.41±5.41       |
| 80       | 79.16±2.73       |
| 85       | 86.08±2.47       |
| 90       | 90.42±2.65       |
| 93       | 93.43±2.28       |
| 95       | 95.51±2.24       |
| 98       | 98.30±1.03       |
| 99       | 99.00±0.93       |
| 100      | 99.35±0.55       |

*Standard deviation (±); n=15 editing sites.
Figure 1 | Binary analysis using Two Sample Logo software. (a) Bulk sequencing of the 795-bp dsRNA RT-PCR product allowed measurement of 406 adenosines on the sense and antisense strands combined. The plot arranges each site in order of increasing percentage of editing measured within the population of RT-PCR products. Coloured horizontal lines show mean overall A-to-I conversion of the 795-bp dsRNA incubated with each ADAR: hADAR1 (blue) = 17.8%, hADAR1-D (red) = 22.7%, hADAR2 (green) = 19.1% and hADAR2-D (purple) = 16.4%. For Two Sample Logo analyses (b–f), sequence contexts edited to a greater extent than the mean were scored as enriched, and those edited less than the mean as depleted. Neighbour preferences of the different ADARs were determined from a single incubation, but repeated experiments showed the same relative pattern of editing among the 406 adenosines, even when protein concentrations differed between experiments. (b–f) Logo displays enriched bases above top line and depleted bases below bottom line for neighbouring five bases on both sides of the central edited adenosine. Level of enrichment/depletion is shown by letter heights with reference to scale on the left; y-axes as in (b). Two Sample Logo settings: t-test, show base if P value < 0.005 and no Bonferroni correction. Panels show: (b) Two Sample Logo of Randomized Control; (c) hADAR1; (d) hADAR1-D; (e) hADAR2; and (f) hADAR2-D.

Figure 2 | Quantitative comparison of editing for different triplets. Bottom plots of a–d show the 16 possible triplet contexts on the x axis with edited A in the centre, ordered according to hADAR1 preferences. 406 adenosines were used to determine the average percentage of the population edited in each triplet context, which is plotted on the y axis and normalized as described (see Methods). The 99% confidence interval (CI) for sample averages is indicated by shading. Top plots show differences in average percentage editing between compared proteins, with values for each triplet shown as black ovals and 99% confidence intervals as vertical lines. Panels show comparisons of triplet preferences for (a) hADAR1 compared with hADAR1-D, (b) hADAR2 compared with hADAR2-D, (c) hADAR1 compared with hADAR2 and (d) hADAR1-D compared with hADAR1-D. See Methods for a description of statistical methodology.

Quantitative approach. Sixteen sequence contexts exist based on 5’ and 3’ nearest neighbours, and we first normalized the data (see Methods), and plotted preferences for the 16 ‘triplets’ using peak heights (Fig. 2). Triplets for all comparisons were arranged left to right on the x axis according to hADAR1 preferences (bottom panels), and differences in % editing plotted separately (top panels). All proteins showed similar trends, and a comparison of triplets along the x axis revealed a clustering of triplets according to identity of the 5’ nearest neighbour. This indicates that the 5’ nearest neighbour has the greatest influence on preferences, confirming conclusions made in our binary analysis (Fig. 1) and in previous reports.17,18

Triplet preferences were almost identical for hADAR1 and hADAR1-D, and very similar between hADAR2 and hADAR2-D, indicating nearest neighbour preferences are largely determined by the catalytic domain. However, hADAR2 showed a greater preference for triplets containing a 3’ G compared with its catalytic domain, hADAR2-D (Fig. 2b), particularly evident in analyses of CAG, AAG and UAG triplets. Thus, although the catalytic domain largely dictates nearest neighbour preferences, for hADAR2, the dsRBMs have a role in discriminating adenosines with a 3’ G.

Triplet comparisons for hADAR1 and hADAR2 (Fig. 2c), and hADAR1-D and hADAR2-D (Fig. 2d), revealed that differences between the catalytic-domain-only proteins do not track with differences between the full-length proteins. This suggests that although
dsRBMs do not contribute substantially to nearest neighbour preferences, the contributions differ for the two ADARs, even on perfectly base-paired dsRNA.

**Best-fit multiplicative models.** Our quantitative analysis provided data for 406 editing sites, an order of magnitude greater than used in previous analyses. Using our larger data set, we set out to create models that more accurately represent nearest neighbour preferences (see Methods). To evaluate the predictive accuracy of various models, Table 2 shows the adjusted coefficient of determination, or \( R^2 \) values associated with hADAR1 and hADAR1-D triplet models (leftmost column of numbers) estimates the % editing of the target adenosine based on the immediate neighbouring 5′ and 3′ bases. This model includes 16 different coefficients to allow the effect of the neighbouring 5′ base to depend on the identity of the neighbouring 3′ base, and conversely, allows the effect of the neighbouring 3′ base to depend on the identity of the neighbouring 5′ base. The remaining models estimate the % editing of the target adenosine based on the identities of 1, 2, 3 or 4 bases on the 5′ and 3′ sides. In contrast to the triplet model, each of the remaining models achieves increased parsimony by invoking the simplifying assumption that the effect of a base at a particular position is not altered by the identities of the bases at other positions.

### Table 2 | Comparison of models for predicting neighbour preferences.

| Protein   | 5′ Old preferences | 3′ Old preferences | 5′ Multiplicative | 3′ Multiplicative |
|-----------|--------------------|--------------------|-------------------|-------------------|
| hADAR1    | U=A>C>G            | None               | U> A=C>G          | G>C>A>U           |
| hADAR2    | ND                 | ND                 | U> A=C>G          | G>C>A>U           |
| hADAR2-D  | U=A>C>G            | U=G>C=A           | G>C>U=A           | C=G>A>U           |
| hADAR1-D  | ND                 | ND                 | ND                | ND                |

*For new nearest neighbour preferences based on two-term model (Table 2, model 3), > indicates a statistically significant difference with \( P \leq 0.05 \), whereas = indicates \( P > 0.05 \); symbols refer to preferences for immediately adjacent bases. Identical relationships were obtained for immediate neighbours using the eight-term model (Table 2, model 6).

The magnitude of coefficients in this two-term model, and associated \( P \) values for the significance of the differences between coefficients for different base identities, provide a more quantitative understanding of ADAR neighbour preferences. For example, representing these preferences in a more familiar way, the coefficients of the two-term model (Supplementary Data 1) indicate that hADAR1 has the following preferences: 5′ U> A>C>G and 3′ G>C>A>U, where the difference between 3′ C and A was not statistically significant at \( P \leq 0.05 \), and is thus represented as approximately equal (=), to signify \( P > 0.05 \). Table 3 provides a side-by-side comparison of our refined preferences with those previously published. Although similar, our analyses allow a more quantitative treatment (see Supplementary Data 1), and also reveal a previously undetected 3′ neighbour preference for hADAR1.
The algorithm for this eight-term 1st–4th 5’ and 1st–4th 3’ neighbour fit model is:

\[
\text{% editing} = 20 \times [1\text{st } 5\text{' base coefficient}] \times [2\text{nd } 5\text{' base coefficient}] \times [3\text{rd } 5\text{' base coefficient}] \times [4\text{th } 5\text{' base coefficient}] \times [1\text{st } 3\text{'} base coefficient] \times [2\text{nd } 3\text{'} base coefficient] \times [3\text{rd } 3\text{'} base coefficient] \times [4\text{th } 3\text{'} base coefficient]
\]

(2) with coefficients given in Supplementary Data 1 and visually displayed in Figure 3. To uniquely define coefficient values, all U coefficients with the exception of the first 5’ position were constrained to equal 1. Interestingly, the coefficients for the second 5’ neighbouring base vary substantially from 1 for hADAR1 and hADAR1-D, but not for hADAR2 and hADAR2-D. This suggests that the hADAR1 catalytic domain has structural features that are more interactive with the first and second 5’ nearest neighbours than the hADAR2 catalytic domain.

The P values at the top of each panel in the figure evaluate the null hypothesis that the coefficients of all four bases in the indicated position were identically equal to 1, corresponding to no influence of the bases at that position. The P values reveal a difference between hADAR1 and hADAR2. For hADAR1 and hADAR1-D, the only bases that modelled poorly (P > 0.001) are on the 3’-side of the editing site, after the immediate 3’ neighbour. However, for hADAR2 and hADAR2-D, bases that modelled poorly are on both 5’ and 3’ sides, again excluding the nearest neighbour. This indicates that hADAR1 is not only more sensitive to the second 5’ base identity than hADAR2, but those beyond the second 5’ neighbour.

Evaluating the algorithm on perfectly paired dsRNA. The eight-term algorithms were tested for their ability to predict editing reported for hADAR1 in 36 and 48 bp dsRNAs, and hADAR2 in 61 and 102 bp dsRNAs (Fig. 4; see Supplementary Fig. S1). In the previous report, editing sites were ranked as major (I), minor (i), or below-detection/ unedited (A). Using a best-fit to experimental data, we defined a boundary for scoring edited (I + i), and unedited (A) sites for hADAR1 (9.6%) and hADAR2 (21%) and found that the eight-term regression algorithms successfully ranked most editing sites above most below-detection/ unedited sites (Fig. 4a). The hADAR1 algorithm successfully scored sites for 27 of 37 adenosines (73%) and that for hADAR2, 49 of 76 adenosines (64%), reiterating the accuracy of regression analyses (Table 2, model #6, hADAR1 = 77.1%, hADAR2 = 57.0%).

Because the 795-bp dsRNA is long and perfectly base-paired, effects of termini proximity and selectivity are minimal. Thus, our algorithms reflect neighbour preferences largely free of other contributions. This is emphasized by comparing editing sites predicted by the algorithm with experimentally determined editing sites in substrates in which selectivity has variable roles (Fig. 4b). A previous study compared ADAR1 editing in a short double-stranded...
sequence to editing of the same sequence embedded within a larger dsRNA, either bounded by internal loops or contiguous base pairs. Because of effects of selectivity, only a subset of the predicted sites are edited in the short dsRNA, but almost all predicted sites are edited in the context of a longer molecule. Subtle differences may relate to differences in reaction conditions as duplexes in Figure 4b were edited to completion and mapped using primer extension\(^2\), which only provides semi-quantitative data.

**Roles of dsRBMs and RNA structure in a natural substrate.** We also analysed *in vitro* editing of an RNA mimicking the human 5-HT\(_{2C}\) pre-mRNA, which contains the ‘A–E’ editing sites observed *in vivo* (Fig. 5). The human 5-HT\(_{2C}\) RNA was incubated with each ADAR, and at the highest concentrations tested (see Methods), was edited to a similar overall level by hADAR1 (6.3\%), hADAR1-D (6.4\%), hADAR2 (6.7\%) and hADAR2-D (6.6\%); editing patterns were independent of protein concentration. These concentrations were chosen for comparison, and % editing values are reported in Figure 5. Adenosines are numbered to correspond with positions in the secondary structure, and tabulated sites are shaded to indicate likelihood of editing as predicted by our eight-term model.

Editing at sites previously observed *in vivo* recapitulated well *in vitro*, consistent with studies showing that editing specificity derives from ADAR without a requirement for accessory proteins\(^8\). As observed *in vivo*, sites ‘A’ and ‘B’ were predominantly edited by hADAR1 (ref. 27), and sites ‘C’ and ‘D’ were predominantly edited by hADAR2 (refs 27, 28). The specifcities of the full-length proteins for these sites were mimicked by their deaminase domains, which only provides semi-quantitative data.

In most cases, differences were best understood by considering that structural disruptions in the 5HT\(_{2C}\) RNA substrate uncouple helices to approximate a series of short double-stranded regions\(^8\). Several additional conclusions emerged. First, adenosines at positions 171, 172 and 208 were edited *in vitro* to varying degrees by hADAR1 and hADAR1-D, but not by hADAR2 and hADAR2-D, even though our model predicted greater editing by hADAR2. This indicates that hADAR1 and hADAR2 are affected differently by RNA structure. Further, at these same positions, preferences of the full-length proteins tracked with those of their deaminase domains, implying that the catalytic domain alone can discriminate structural features. Finally, certain positions were edited by the catalytic domain but not by the full-length ADAR (for example, 226, 227), even at sites predicted to be in preferred contexts. Thus, for both ADARs, dsRBMs may sometimes block editing sites. Similarly, adenosines at positions 116 and 118, like site ‘F’, are edited by all
proteins except full-length hADAR1, implying these sites are blocked by dsRBMs of hADAR1, but not those of hADAR2.

Discussion

We show that current protocols for Sanger sequencing allow ADAR editing to be quantified from peak heights of CDNA sequence traces with a decreased error than previous methods (s.d. ≤5%; Table 1). Using this methodology, we refined and quantified neighbour preferences for human ADAR1 and ADAR2. In addition, we applied our methodology to answer questions about ADARs and to generate an algorithm for the de novo prediction of editing sites in dsRNA.

Differences between preferences detailed here and those previously reported (Table 3)17,18 are explained by an increased accuracy and larger sample size, and the different in vitro conditions used. Previous studies used data from dsRNA reacted to completion, thus sacrificing the ability to detect differences between well-edited sites. To overcome this limitation, we reacted 795-bp dsRNA to an intermediate level of editing. Previous studies used dsRNA that was very short compared with the 795-bp dsRNA, incurring effects of duplex termini17,18, and selectivity13. We consider data from the 795-bp dsRNA to reflect neighbour preferences largely free of these effects.

Even with their limitations, previous studies reported neighbour preferences that agree fairly closely with those reported here (Table 3). However, our refinement allowed discrimination between nearest neighbours that were previously thought to be targeted equally well, and also revealed a 3′ nearest neighbour preference for hADAR1. Further, our larger data sets allowed us to construct regression models that allow new insight into ADAR preferences (below).

A prevailing hypothesis is that dsRBMs anchor an ADAR to a dsRNA region, while the catalytic domain provides the specificity that leads to a preference for certain adenosines31. Indeed, chimeric protein constructs of hADAR1 and ADAR2, in which the catalytic domains are exchanged, show specificity that tracks with catalytic domain identity31. By carefully comparing preferences of full-length hADAR1 and hADAR2 with those of their catalytic domains, we confirm that, for most triplet contexts, this hypothesis is true. However, our more quantitative approach allowed us to discern that full-length hADAR2, compared with its catalytic domain, has an increased preference for adenosines with a 3′ G (Figs 2b and 3).

Thus, we find that dsRBMs of hADAR2 contribute to editing specificity. This agrees with nuclear magnetic resonance solution data indicating that serine 258 in the second dsRBM of rat ADAR2 forms a hydrogen bond with the minor groove amino group of the guanosine 3′ to the R/G editing site15. We note, however, that our analyses indicate the catalytic domain, not the dsRBMs, is largely responsible for discriminating adenosines in different sequence contexts.

We found that a multiplicative model that separately considers neighbour preferences that agree fairly closely with those reported here (Table 3). However, our refinement allowed discrimination between nearest neighbours that were previously thought to be targeted equally well, and also revealed a 3′ nearest neighbour preference for hADAR1. Further, our larger data sets allowed us to construct regression models that allow new insight into ADAR preferences (below). A prevailing hypothesis is that dsRBMs anchor an ADAR to a dsRNA region, while the catalytic domain provides the specificity that leads to a preference for certain adenosines31. Indeed, chimeric protein constructs of hADAR1 and ADAR2, in which the catalytic domains are exchanged, show specificity that tracks with catalytic domain identity31. By carefully comparing preferences of full-length hADAR1 and hADAR2 with those of their catalytic domains, we confirm that, for most triplet contexts, this hypothesis is true. However, our more quantitative approach allowed us to discern that full-length hADAR2, compared with its catalytic domain, has an increased preference for adenosines with a 3′ G (Figs 2b and 3).

Thus, we find that dsRBMs of hADAR2 contribute to editing specificity. This agrees with nuclear magnetic resonance solution data indicating that serine 258 in the second dsRBM of rat ADAR2 forms a hydrogen bond with the minor groove amino group of the guanosine 3′ to the R/G editing site15. We note, however, that our analyses indicate the catalytic domain, not the dsRBMs, is largely responsible for discriminating adenosines in different sequence contexts.

We found that a multiplicative model that separately considers the identity of 5′ and 3′ nearest neighbours gives as good a fit to editing data as triplet identities. This suggests that the ADAR active site interrogates these positions independently. Further, multiplicative models that considered base identities beyond nearest neighbours showed increased fit (Table 2), indicating that editing site choice is influenced by more than nearest neighbours. Finally, the regression modelling indicated that, for all proteins studied, 5′ bases have more influence on editing than 3′ bases.

Our analysis revealed that hADAR1 is more influenced by bases 5′ of an editing site than hADAR2 (Fig. 3, P values). At the surface of the hADAR2 catalytic pocket are amino acids that are disorderd in the crystal structure14, and show poor conservation with hADAR1. The hADAR1a sequence (GALFDKSCSDRAMESTESRHYPVFENPKQGK) is also slightly longer than the analogous hADAR2a sequence (ARIFSPHEPEEADVPRHNPDRKARGQ). In the hADAR2-D crystal structure, this region is predicted to be close to the site being edited, and thus, is a good candidate for mediating the increased sensitivity of hADAR1 to 5′ neighbours.

We developed a web-based application based on our eight-term model (http://www.biochem.utah.edu/bass/inosinepredict; Supplementary Software). The algorithm was developed by fitting to experimentally determined editing sites in a long perfectly base-paired dsRNA, and approximates ADAR preferences in the absence of the effects of RNA structure. ADARs target dsRNA formed from sense–antisense transcripts15, or that introduced into an organism to mediate RNA interference14, and we envision our algorithm facilitating researchers in the identification of such sites. That said, although our algorithm represents an advance, the R values (Table 2) emphasize that its predictive power is still limited. Predictions should be treated cautiously, especially for hADAR2, or for approximating editing under conditions different from those used here. However, we envision the limitations of our model are key to its improvement. For example, application of our algorithm to ADAR substrates in which RNA structure mediates editing site choice will facilitate studies to define how structure affects editing, setting the stage for future algorithms that take such features into account.

Methods

Protein purification. Expression constructs included a N-terminal 10-histidine tag followed by a TEV protease site, then the ADAR CDNA, ligated into the YEpTOP2PGAL1 vector16. hADAR2 and hADAR2-D vectors were constructed as described35, using a hADAR2a CDNA template36, with the hADAR2-D construct encoding residues 299–701 of hADAR2a34. hADAR1 and hADAR1-D vectors were similarly constructed from the nuclear hADAR1a isoform, which initiates at Met296 of the hADAR1id isoform31. The hADAR1-D construct encodes residues 528–931 of hADAR1a. Proteins were expressed in Saccharomyces cerevisiae and purified as described4, with modifications specified in Supplementary Methods. hADAR2, hADAR2-D and hADAR1-D were purified to >98% as estimated by SYPRO Red staining of SDS–polyacrylamide gels with BSA standards4, and stored in storage buffer A (20 mM Tris–HCl, pH 8.0, 100 mM NaCl, 1 mM 2-mercaptoethanol, 15% glycerol), hADAR1 was stored in storage buffer B (50 mM Tris–HCl, pH 8.0, 200 mM KCl, 3 mM EDTA, 0.01% NP-40, 10% glycerol and 1 mM DTT) and purified to 80%, twice the purity previously achieved for hADAR1 (ref. 18).

RNA preparation. Radiolabelled and non-radiolabelled 795- bp dsRNA encoding chloramphenicol acetyl transferase (CAT) was prepared as described3. The dsRNA has 22 nt 5′ overhangs at each terminus. Human 5-HT2C pre-mRNA template was cloned de novo with a T7 RNA polymerase promoter into the pUC18 vector (Fermentas all primers in Supplementary Table S2). Transcription was as for 795-bp dsRNA4. RNA (sequence in Supplementary Methods) was gel purified, boilded (2 min) and refolded for hybridization of 795-bp dsRNA4; editing was identical without gel purification or refolding.

Four-dye-trace bulk sequencing quantification. CDNA populations from reverse transcription PCR (RT–PCR) of editing products were bulk sequenced in one reaction rather than sequencing individually cloned molecules. Thus, editing sites appear as mixed peaks in traces. Four-dye-traces sequences in abi file format were processed using BioEdit (http://www.mbio.ncsu.edu/BioEdit/bioedit.html; File > Batch Export of Raw Sequence Trace Data). Text file outputs were opened and evaluated in Microsoft Excel (Microsoft) Editing sites were quantified by measuring maximal height of T peaks (unedited) and C peaks (edited) and calculating proportion of the population edited at each site (100% × [C height/ (T height + C height)]). For peaks without a clear maximal height, shoulder shape and distances between distinct peaks were used as guides to manually select a shoulder value as the maximal peak height.

For method validation, standard techniques were used to clone a transcription template that differed from the antisense CAT template6 in that certain adenosines were changed to guanosines (edited). Primer pair 31/32, flanking the CAT coding region, was used to PCR amplify edited and unedited CAT antisense templates. PCR products were gel purified and concentrations determined by ultraviolet spectrophotometry, using precise extinction coefficients, calculated as described3. PCR products were mixed in known ratios to mimic prescribed levels of editing at certain adenosines, then sequenced (Primer 55; GENEWIZ).

ADAR assays. For ADAR activity assays, radiolabelled 795-bp dsRNA was reacted in 22.5 mM HEPES-Tris–HCl pH 7.5 (20°C), 40 mM KCl, 10 mM MgCl2, 0.5 mM DTT, 0.1 mM 2-mercaptoethanol, 0.01% NP-40 and 1 μg/μl Promega RNasin Plus (Promega), for 1 h at 30°C. Varying concentrations (nM–μM) of hADAR2 and hADAR2-D were incubated with 1 nM 795-bp dsRNA, and hADAR1 and hADAR1-D with 0.1 nM 795-bp dsRNA, to determine conditions that provided ~20% overall A-to-I conversion, as determined by thin layer chromatography4. For the ADAR preference assay, non-radiolabelled 795-bp dsRNA was reacted as in the ADAR activity assay. ADAR concentrations were chosen to give ~20%
Proteinase K (NEB) and SDS and purified 41. Primer 90 was used for reverse
concentration, while the RNA concentration was kept at 0.1 nM; hADAR1 0.5,
2 . Hundley, H. A. & Bass, B. L. ADAR editing in double-stranded UTRs and other
RNA editing by adenosine deaminases that act on RNA 1 . Bass, B. L. RNA editing by
adenosine deaminases that act on RNA 12875 – 12884 (2000).
Bass, B. L. RNA editing and hypermethylation by adenosine deamination. Trends Biochem. Sci. 22, 157–162 (1997).
Lehmann, K. A. & Bass, B. L. The importance of internal loops within RNA
structures of ADAR1 and ADAR2 double-stranded RNA binding motifs. Structure 14, 345–355 (2006).
Polson, A. G. & Bass, B. L. Preferential selection of adenosines for modification
by double-stranded RNA adenosine deaminase. EMBO J. 13, 5701–5711 (1994).
Lehmann, K. A. & Bass, B. L. Double-stranded RNA adenosine deaminases
ADAR1 and ADAR2 have overlapping specificities. Biochemistry 39, 12875–12884 (2000).
Steif, R. et al. The structure of the ADAR2 dsRBM–RNA complex reveals a sequence-specific readout of the minor groove. Cell 143, 225–237 (2010).
Steif, R., Xu, M., Skrisovska, L., Emsen, R. B. & Allain, F. H. Structure and
specific RNA binding of ADAR2 double-stranded RNA binding motifs. Structure 14, 345–355 (2006).
Purpeisov, V., Hurwitz, J. & Liakouska, L. M. Two Sample Logo: a graphical
sequence logo for secondary structure and domain sequence alignments. Bioinformatics 22, 1536–1537 (2006).
Hall, J. A. G. & Bass, B. L. Identification of adenosine deaminase editing
sites by denaturing HPLC. Nucleic Acids Res. 34, 1269–1278 (2006).
Felsenfeld, G. & Greider, C. W. Evidence that RNA editing is required for the
maintenance of the human X chromosome. Nature 315, 102–105 (1985).
Wong, S. K., Satoh, S. & Lazzarini, D. W. A subcellular localization assay for
adenosine deaminase knockdown in mammalian cells. Nucleic Acids Res. 32, 1150–1156 (2004).
Felsenfeld, G. & Greider, C. W. The human adenosine deaminase gene
is a germline suppressor of X chromosome inactivation. Nature 315, 102–105 (1985).
Felsenfeld, G. & Greider, C. W. Evidence that RNA editing is required for the
maintenance of the human X chromosome. Nature 315, 102–105 (1985).
Wong, S. K., Satoh, S. & Lazzarini, D. W. A subcellular localization assay for
adenosine deaminase knockdown in mammalian cells. Nucleic Acids Res. 32, 1150–1156 (2004).
Felsenfeld, G. & Greider, C. W. The human adenosine deaminase gene
is a germline suppressor of X chromosome inactivation. Nature 315, 102–105 (1985).
36. Macbeth, M. R. & Bass, B. L. Large-scale overexpression and purification of ADARs from Saccharomyces cerevisiae for biophysical and biochemical studies. *Methods Enzymol.* **424**, 319–331 (2007).

37. George, C. X. & Samuel, C. E. Human RNA-specific adenosine deaminase ADAR1 transcripts possess alternative exon 1 structures that initiate from different promoters, one constitutively active and the other interferon inducible. *Proc. Natl Acad. Sci. USA* **96**, 4621–4626 (1999).

38. Bass, B. L. & Weintraub, H. A developmentally regulated activity that unwinds RNA duplexes. *Cell* **48**, 607–613 (1987).

39. Tatuurov, A. V., You, Y. & Owczarzy, R. Predicting ultraviolet spectrum of single stranded and double stranded deoxyribonucleic acids. *Biophys. Chem.* **133**, 66–70 (2008).

40. Macbeth, M. R., Lingam, A. T. & Bass, B. L. Evidence for auto-inhibition by the N terminus of hADAR2 and activation by dsRNA binding. *RNA* **10**, 1563–1571 (2004).

41. Hough, R. F. & Bass, B. L. Purification of the Xenopus laevis double-stranded RNA adenosine deaminase. *J. Biol. Chem.* **269**, 9933–9939 (1994).

42. Wherry, R. J. A new formula for predicting the shrinkage of the coefficient of multiple correlation. *Ann. Math. Stat.* **2**, 440–457 (1931).

43. Mathews, D. H., Sabina, J., Zuker, M. & Turner, D. H. Expanded sequence dependence of thermodynamic parameters improves prediction of RNA secondary structure. *J. Mol. Biol.* **288**, 911–940 (1999).

44. Zuker, M. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res.* **31**, 3406–3415 (2003).

**Acknowledgments**

We are grateful to Dr David Nix of the University of Utah Health Science Center Bioinformatics Core, and Rachel Bookman for assistance in creating the InosinePredict Web application. We thank Dr Mark R. Macbeth for providing the hADAR1-D expression vector. This work was supported by funds from the National Institute of General Medical Sciences to B.L.B. (R01GM044073), and a postdoctoral fellowship to J.M.E. from the American Foundation for Suicide Prevention. T.G. is supported by the University of Utah Study Design and Biostatistics Center, with funding in part from the Public Health Services (UL1-RR025764, C06-RR11234) from the National Center for Research Resources.

**Author contributions**

J.M.E. performed all biochemical experiments, integrated data from biochemical and statistical analyses, and wrote a draft of the paper. T.G. performed all statistical analyses, designed the models, and wrote and edited certain sections of the paper. B.L.B. oversaw all analyses and edited and prepared the final manuscript.

**Additional information**

**Supplementary Information** accompanies this paper at http://www.nature.com/naturecommunications

**Competing financial interests:** The authors declare no competing financial interests.

**Reprints and permission** information is available online at http://npg.nature.com/reprintsandpermissions/

**How to cite this article:** Eggington, J. M. et al. Predicting sites of ADAR editing in double-stranded RNA. *Nat. Commun.* 2:319 doi: 10.1038/ncomms1324 (2011).

**License:** This work is licensed under a Creative Commons Attribution-NonCommercial-Share Alike 3.0 Unported License. To view a copy of this license, visit http://creativecommons.org/licenses/by-nc-sa/3.0/