Figure S1.

(A) SHAPE profiles for the RB1 5′ UTR for WT, the structural controls C4A and C166U, and the retinoblastoma-associated mutations G17C and G18U. WT has a very similar profile to the structural controls, while the disease-associated mutations differ from WT on the 5′ end of the sequence. Values are the mean reactivity across five repeats. Asterisks indicate positions where the reactivity could not be determined. **(B)** The representative structures from Figure 2, plotted with additional SHAPE reactivity information for each nucleotide indicated by the color. Each structure is compatible with that sequence's reactivity profile.

Figure S2.

We performed transcription factor motif analysis in conjunction with an analysis of publicly available ChIP-seq data to characterize the intergenic region upstream of the RB1 gene and identify potential transcriptional regulators and transcription factor binding sites (TFBSs). **(A)** Genomic region upstream of the RB1 gene (in hg19 coordinates). Small red bars indicate the point-source peak of each ChIP-seq experiment. Grayscale indicates the ChIP-seq score with higher values corresponding to greater enrichment and certainty. We found 61 ChIP-seq point-source peaks, corresponding to 32 unique proteins, within 200 nucleotides of the annotated transcription start site. The gray, dark red, and navy lines indicate the transcription start site, translation start site, and the end of the first exon respectively. **(B)** A majority of the point-source peaks (74%; red lines, bottom) are upstream of the transcription start site, indicating that this gene is

conventionally regulated. **(C)** To locate TFBSs likely to be disrupted by these mutations, we selected ChIP-seq point-source peaks within 20 nucleotides of G17C/G18U. Out of the 61 point-source peaks, 11 (18%) are within 20 nucleotides of G17C/G18U. These 11 point-source peaks, representing putative TFBSs affected by these mutations, correspond to 9 unique transcription factors. **(D)** We compared the sequence motifs for these transcription factors to the sequence around the mutations and found that only EGR1 and MAZ (2 out of 32 transcription factors, 6%) are likely to bind around these mutations.

Figure S3.

RB1 protein alignment based on CDS sequences from each organism found through NCBI BLAST. Much like the sequence of the RB1 5′ UTR, the RB1 coding sequence is highly conserved across the organisms shown in the alignment.

Figure S4.

(A) SHAPE reactivity profiles for cow (left) and manatee (right). Values represent mean reactivity across five repeats. **(B)** Representative structures from Figure 5A (left: cow, right: manatee), with nucleotide colors representing SHAPE reactivity. The representative structures are compatible with the overall reactivities found by SHAPE structural probing.

Figure S5.

(A) Mean Firefly luciferase activity, **(B)** Firefly luciferase RNA abundance, and **(C)** Renilla luciferase activity for each RB1 5′ UTR construct. Error bars represent standard error of the mean. Firefly luciferase activity is correlated with Firefly luciferase RNA levels, while the Renilla luciferase activity as a transfection control demonstrates that transfection efficiency is not a confound in this experiment. All values are relative to an empty vector control.