Table S1. Oligonucleotides used in this study.

Mutagenesis oligonucleotides:	
DONSON_pR211C_S	CAAACTCTCCTCTGAGCTCTGTTGTACCTTCCAGCAG
DONSON_pR211C_AS	CTGCTGGAAGGTACAACAGAGCTCAGAGGAGAGTTTG
DONSON_pY270C_S	GTGAGCTTTACTTCTCTATGTAATTTGCTGAAGACAAAAC
DONSON_pY270C_AS	GTTTTGTCTTCAGCAAATTACATAGAGAAGTAAAGCTCAC
DONSON_pP545L_S	GAGCAACTTAGTCAAATACTGTTACTTGGGAAATCATC
DONSON_pP545L_AS	GATGATTTCCCAAGTAACAGTATTTGACTAAGTTGCTC
DONSON_pP224S_S	CTTATCTATTGGCTCCACTCTGCTTTGTCTTGGCTAC
DONSON_pP224S_AS	GTAGCCAAGACAAAGCAGAGTGGAGCCAATAGATAAG
DONSON_pF165S_S	GTCTATCAAAACGCGACTCCTTTCCACCTCTTCTCAACCC
DONSON_pF165S_AS	AAGGGTTGAGAAGAGGTGCTAAGGAGTCGCGTTTTGATAG
RT-PCR oligonucleotides:	
Rat_INS2_Ex2_F	CCTGCTCATCCTCTGGGAGC
Rat_INS2_Ex3_R	AGGTCTGAAGGTCACGGGCC
Rat-DONSON_Ex5_R	CAGTTGTGCCACTTCATTTC
Gateway cloning oligonucleotides:	
DONSON_attB1_F1	GGGGACAAGTTTCTACAAAAAAGCAGGCTTACCCCCAACAACAAATTCC
DONSON_attB2_R1	GGGGACCACTTTGTACAAGAAAGCTGGGTACTGCCCATGAAACTGATCC
Sequencing oligonucleotides:	
DONSON_seq_F1	GTACTGAGTTACCCGTAGATTGGTCTATC
DONSON_seq_F2	GACATACTTTCTATCAAGCTGCG
DONSON_seq_R1	ACAGATTCAGGTCTGTGATCC
DONSON_seq_R2	TGCCCAGGTAAAGGGTTGAG
DONSON_seq_F3	TCATGTTAGCTGATAAAATCTTTAGG
DONSON_seq_R3	TGCAAAGGAGAGCATTTCAC
DONSON_seq_F4	GGCTTACCATTACATCTGGCTC
DONSON_seq_R4	CCTTCTGCCATTAACGTATCC
DONSON_seq_F5	GCGAGTCCTCATAACCAATTTC
DONSON_seq_R5	AAAAGTTTGCGGACTGCTG
DONSON_seq_F6	TGTCTCATGCATCTCAGTTGG
DONSON_seq_R6	ATTTCCCATTTGCAAGGAAG
EGFP_seq_F1	GGAGTACAACTACAACAGCC
EGFP_seq_R1	TCCGTCGGAGCCGCACTAC
pEGFP_seq_F1	CAGAGCTGGTTTAGTGAACCGTCAG
pEGFP_seq_R1	AGGCACAGTCGAGGCTGATC
M13F	GTAAAACGACGGCCAG
M13R	CAGGAAACAGCTATGAC

Table S2. Summary statistics from the Long Ranger pipeline.

Summary Statistics	P1	P2	P3	Average
Input DNA				
Mean molecule length (μ)	34,015	31,391	30,067	31,824
DNA in molecules > 20kb	71.70%	72.30%	71.20%	71.73%
Corrected estimate of DNA loaded	1.21 ng	1.33 ng	1.03 ng	1.19 ng
GEMs				
Number detected	1,670,049	1,652,325	1,734,425	1,685,600
N50 linked-reads per molecule	20	17	20	19
Phasing				
SNPs phased	98.30%	98.70%	98.90%	98.63%
Longest phase block	9,385,247	18,466,034	4,636,973	10,829,418
N50 phase block length	1,568,908	3,329,364	873,873	1,924,048
Sequencing				
Number reads	774,189,712	763,832,020	808,535,762	782,185,831
Mean depth	33.7X	33.2X	34.6X	33.8X
Mapped reads	97.20%	97.10%	96.10%	96.80%
Zero coverage	0.12%	0.12%	0.95%	0.40%
Total variants (SNPs and Indels)	5,606,271	6,481,408	5,858,991	5,982,223

A

B

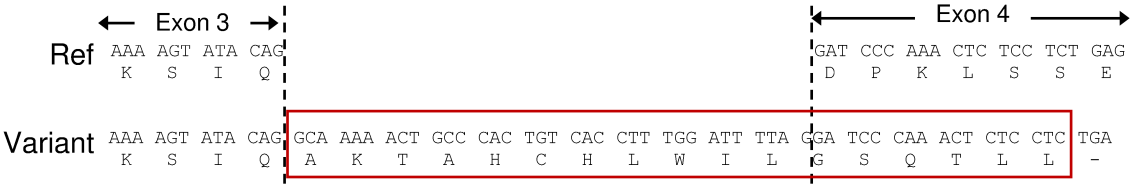

Figure S1. *In silico* splice site predictions for *DONSON* c.607-36G>A. (A) Screenshot from the Alamut Visual v2.11 software showing the predicted splicing effect of the variant. The top box represents the reference sequence with a G at position c.607-36 while the bottom box represents the variant sequence with an A at this position. The sequence encoding *DONSON* exon 4 is highlighted in blue. Probabilities for the use of the splice sites are indicated for four different algorithms, where the green bars below the sequence represent predicted acceptor splice sites and the dark blue bars above represent predicted donor splice sites. The figure shows that *DONSON* c.607-36G>A likely introduces a new splice acceptor splice site 34 bp upstream of the canonical splice acceptor site for exon 4. (B) Nucleotide and translated protein sequences of the mRNA transcripts expected for the reference sequence ('Ref') and the *DONSON* c.607-36G>A variant ('Variant'). Sequences encoded by exons 3 and 4 are indicated and exon/intron boundaries are shown with a dashed line. The intronic sequence likely included in the variant mRNA would alter the reading frame and introduce a stop codon after 17 amino acids (red box), likely causing nonsense mediated decay of this transcript.