Supporting Information

Tucker et al. 10.1073/pnas.1108918108

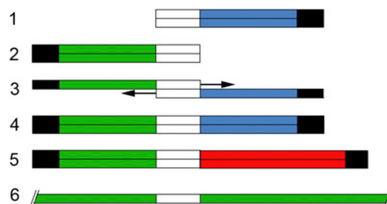

Fig. S1. Proposed mechanism by which a chimeric DNA fragment was introduced into the sequencing library, resulting in detection of the Alu insertion in *male germ cell-associated kinase* (*MAK*) exon 9. Lane 1, a genomic fragment from chromosome 1 (blue) with 12-bp homology to *MAK* exon 9 (white). This fragment has a linker (black) annealed to it in preparation for PCR amplification before exome capture. Lane 2, a genomic fragment from exon 9 of *MAK* (green) with 12-bp perfect homology to a sequence from chromosome 1 (white). This fragment also harbors a library amplification linker (black). Lane 3, during library amplification, the complementary sequences anneal to one another and extend (arrows) to form the blunt-ended chimeric fragment shown as lane 4. Lane 5, the Alu insertion in the proband (red) occurs immediately adjacent to the 12-bp homology with chromosome 1; thus, many library fragments have a portion of exon 9 and a portion of the Alu insertion as shown in lane 5. However, Alu repeats are physically removed from the expanded library before hybridization to the exome capture baits. As a result, the fragment illustrated in lane 5 was absent from the ABI sequencing library (Fig. S2*A*) and noticeably depleted from the Illumina sequencing library (Fig. S2*B*). Lane 6, a portion of the 120-bp exome capture bait from *MAK* exon 9 harboring the 12-bp homology with chromosome 1. Because lane 5 has been removed from the library before exome capture and because the proband has no normal copy of exon 9, the most homologous fragment remaining in the library at the time of exome capture is lane 4. The *MAK* sequences within this lane allowed the reads resulting from it to be mapped to exon 9, but the differences in sequence between the chromosome 1 portion of the chimera (blue) and the true exon 9 sequence were interpreted by the GATK algorithm as two different mutations, ultimately leading to the discovery of the Alu insertion.

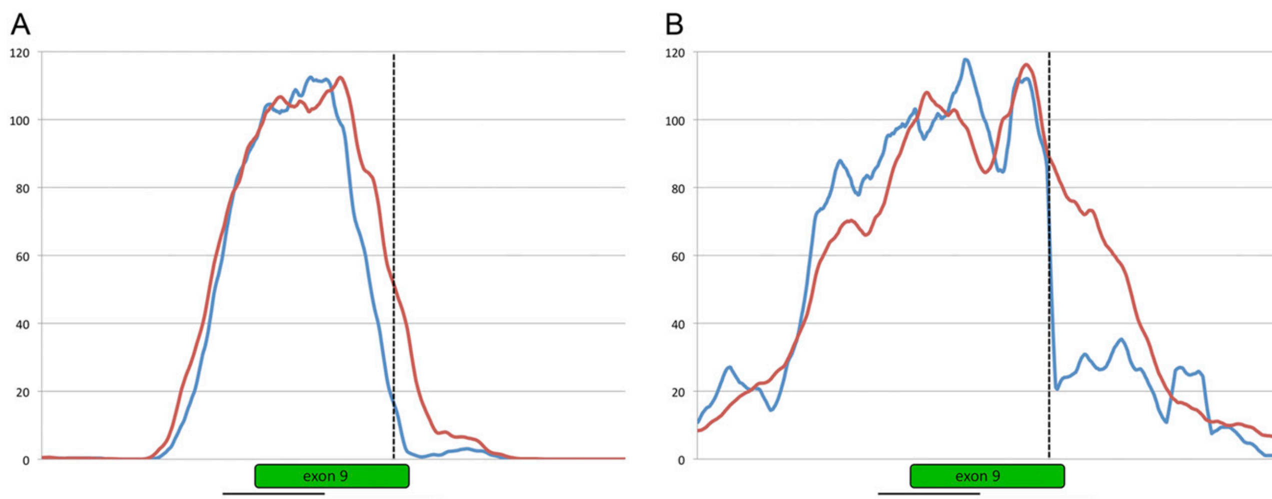

Fig. S2. Physical removal of Alu sequences before exome sequencing reduces the sequence coverage near the Alu insertion in exon 9. (*A*) This graph shows the relative frequency of each exon 9 nucleotide in the ABI exome sequence data of the proband with the Alu insertion (blue line) compared with similar data averaged from seven other individuals who do not harbor the Alu insertion (red line). For each patient, the median frequency for a nucleotide in exon 9 was set to 100, and all other frequencies were scaled accordingly. The extent of exon 9 is shown as a green box at the bottom, whereas the sequence compositions of the two exome capture baits are shown as black lines below the exon symbol. The vertical dashed line corresponds to the site of the Alu insertion. Note that the exclusion of Alu sequences before exome capture resulted in complete removal of normal sequences from the 3′ end of exon 9, thereby allowing the chimeric fragment shown in Fig. S1, row 4 to be captured. (*B*) This graph is identical to *A* except that the data are derived from the paired-end Illumina sequencing experiment. The larger fragment size used in the Illumina protocol results in more nonexonic nucleotides being captured and sequenced. In addition, the physical Alu blocking was less complete than in the ABI experiment, allowing some Alu–*MAK* junction sequences to be captured and sequenced. However, the Burrows-Wheeler Aligner mapping algorithm removed these Alu sequences from the alignment (note the abrupt reduction in coverage at the Alu insertion point), and as a result, the Alu insertion was not detected with the Illumina experiment.

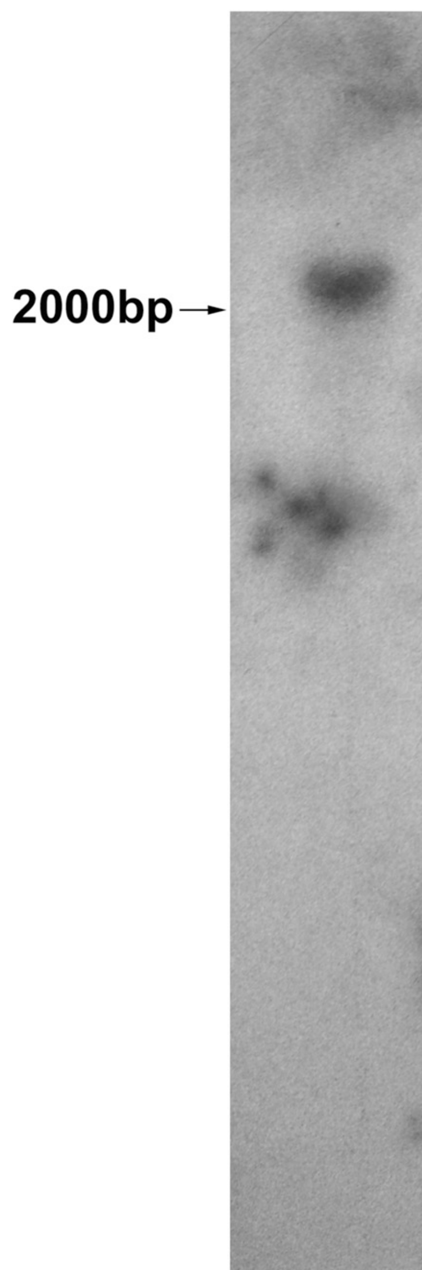

Fig. S3. Northern blot human retinal RNA. Northern blotting of RNA extracted from the retina of a normal human eye donor revealed the largest and most abundant MAK transcript to be slightly larger than 2,000 nt in length.

```
Human:     CCCGTTGGGGCAGAACTTGCTTTCAAAAGGAGCAATGCAGAAGAAAGCATAATTAAACCAATCGAAAAACTATCATGC
Mouse:     TCCATGGGGGCGGACCTTTCCTTCAAGAGGAGTAACGCAGAAGACAGCATAATTAAACCGATTGAAAACCTGTCATGC
Zebrafish: GCGCCCATTGGAGGAGCACCAATGTCCAGAATCAGCCCAG--------------------------------------
Frog:      CCCATTGGAGGAGGCATTGGTTACAATAGAACTAATGGAG--------------------------------------

Human:     AATGAAACTTTTCCTGAAAAATTAGAGGACCCACAAGGAAATCTTGGAAGTTATGCTACTTACAATCAGTCAGGATAT
Mouse:     ACTGGGAAGTCTGCCGAGCAGCTAGAGGACCCACAAGGAAACCTTGGGAGCTACACCACTTACAACCAGACGGGATAC
Zebrafish: -------------------------------------GAAACTTTGTAACTACAACATACAATCTCTCTGGAGGATAC
Frog:      -------------------------------------GCATTTTAGGAGGTTTTGTAGGTCCTTCCTACAACTCAACT
```

Fig. S4. The cDNA sequence for codons 520–571 of the human full-length retinal transcript of *MAK*. The cDNA sequence for the homologous regions of mouse, zebrafish, and frog are shown below the human sequence. Exon 12 (shown in red) is present in mouse but not zebrafish or frog. An updated sequence corresponding to the retinal transcript including human exon 12 (codons 532–557) has been submitted to National Center for Biotechnology Information's RefSeq database.

Table S1. Prioritization of exome variants

	Solid	Illumina
All variants	26,712	23,043
Quality >50	17,156	20,346
Coding and splice variants	14,915	17,148
dbSNP	2,734	3,500
Human exome databases	1,013	1,163
Local database	586	556
Family structure	413	420

Table S2. List of gene-specific primer pairs

Primer sequence	Primer name
GGAGTGTGCTTATGGGCAAG	Human MAK exon 1 forward
TCGATTATATCCGGCAGAGG	Human MAK exon 8 reverse
ACCGAAATGTTGAATTGGGA	Human MAK exon 7 forward
GGGGGTTTATTTCCTTTCCA	Human MAK exon 11 reverse
CTCGACAGGGGAAAACAAGA	Human MAK exon 10 forward
ACTGAGGGAATGGGCTGTG	Human MAK exon 13 reverse
CCCTTCGTCAAATCATCTGG	Human MAK exon 8 forward
TCTTGTTTTCCCCTGTCGAG	Human MAK exon 10 reverse
TGTACTGATGGGCAAGAGCAAT	Mouse MAK exon 1 forward
ACCGTCTCCGGGCTCCCTTAT	Mouse MAK exon 9 reverse
TCCCAATGCCAGTAGCGAGGC	Mouse MAK exon 7 forward
CTACCGGTGGCCTCCATACTTA	Mouse MAK exon 14 reverse
GATGGAACCTATGGCAGCCTTC	Zebrafish MAK exon 1 forward
CCAACTCAGGACCCATACATAAC	Zebrafish MAK exon 6 reverse
GAATCATGCTAATGTGGTGA	Zebrafish MAK exon 4 forward
AAGGTTTAGCTTTGCCAGTT	Zebrafish MAK exon 14 reverse
CACATTTAGCACTGTGACTAAG	Zebrafish MAK exon 11 forward
GTTCTGCTCTCCTTTTCCAT	Zebrafish MAK exon 15 reverse
AACCGATATACAACCATGAAAC	*Xenopus* laevis MAK exon 2 forward
AAGATCTGGTCAACCTCACTTG	*Xenopus* laevis MAK exon 6 reverse
TAATGTATCAAATTTTACAAGG	*Xenopus* laevis MAK exon 4 forward
CGATGGAAATGGGATCTGAC	*Xenopus* laevis MAK exon 10 reverse
ATATGTAACAAAAATGCACTC	*Xenopus* laevis MAK exon 8 forward
ATGGCCTCCATATTTGGCTA	*Xenopus* laevis MAK exon 14 reverse
TGAGGGGATGTGTTTCCTTC	Human CRX forward
GCCAAGGCGTTGACAGAATA	Human CRX reverse
CTCTACGACGTGGACGGTAA	Human recoverin forward
CAAACTGGATCAGTCGCAGA	Human recoverin reverse
CCGGCAGAAGATTGTAGAGC	Human Pax6 forward
GCCCGTTCAACATCCTTAGT	Human Pax6 reverse
GAGAAGGCATTCAACGAAGC	Human Chx10 forward
TCCTTGGCTGACTTGAGGAT	Human Chx10 reverse