**Junction characterization**

The total number of reads per sample are listed Table S6. Data were analyzed using CLC Genomic Workbench 8 as previously described10 and junctions were characterized using a Python script (PyCharm Community Edition 2021, JetBrains) as outlined below. The number of reads analyzed from each library are also listed in Table S6.

Junctions were scanned for matches of 10 nucleotides, starting proximal to the break site and searching for an upstream and downstream match corresponding to the smallest possible deletion. These matches establish deletion length to the left and right, respectively. Junctions were then reconstructed as the sequence between the 5’ end of the upstream match and the 3’ end of the downstream match. Junctions were further characterized for insertions (any intervening sequence between the left and right 10 nucleotide matches) and microhomologies (sequence overlap between the left and right 10 nucleotide matches).

Characterized junctions containing base ambiguities (i.e. N, W, S, R, K) were removed from analysis. Furthermore, we excluded junctions with base substitutions in the 3-10 nucleotides proximal to the break site if nucleotides adjacent to the substitution matched the corresponding reference sequence. These substitutions could be attributed to polymerase error during sample amplification and were therefore removed from analysis. Base substitutions in the proximal 1-2 nucleotides were not removed, as it would be difficult to distinguish these from subsequent deletion and insertion generating a true complex indel at those base positions.