



Additional file 1. IGV screenshot of the dimorphic HERV-W locus 18q21.1_W2. The top panel shows the position on the chromosome and hg38 coordinates. The Coverage track shows the number of reads that covers each base at the locus. The numbers within brackets [0-96] shows the range of reads covering the visualized region. The Reads track shows a sample of the reads mapped to that locus. The colored reads are discordant reads whose mates have unexpected insert size or inconsistent orientation. The RepeatMasker track shows the position of transposable elements and repeats identified in that chromosomal interval. The presence of discordant reads with mates mapping to internal HERV-W coding sequence is used to infer the presence of provirus at the locus. The increase in read coverage pertaining to the solo LTR region is also indicative of presence of provirus, as the reads from 2 LTRs of a provirus allele map to the solo LTR. The candidate is experimentally verified to be homozygous for provirus at that locus.