iScience, Volume *25*

Supplemental information

Detecting tandem repeat variants in coding regions using code-adVNTR

Jonghun Park, Mehrdad Bakhtiari, Bernt Popp, Michael Wiesener, and Vineet Bafna

Supplemental Figures

Figure S1. Comparison between single motif and multiple motif HMM, Related to Figure 1.

An example of a tandem repeat with multiple motifs. When a read is sequenced from the tandem repeat region, single motif HMM would match the variations within motifs with mismatch/insertion/deletion states.

Figure S2. The HMM for VNTRs, Related to Figure 1.

The HMM is composed of three profile HMMs: prefix and suffix profile HMMs to align the flanking regions, and motif profile HMM to align multiple and partial number of motifs. Dotted lines connected to Start and End states indicate the transitions for reads that do not span the entire VNTR region.

Figure S3. IGV screenshots of long read mapping on genomic sample NA12878 showing the small indels detected by code-adVNTR, Related to Figure 2.

Long read validation of three single base heterozygous deletions identified at 2 VNTR loci in ZNF662 and ZNF717 by code-adVNTR using short read.

Figure S4. Comparison of two groups of VNTRs that passed and failed Hardy-Weinberg equilibrium test, Related to Figure 3.

(A) Distribution of VNTR lengths and (B) similarity scores of VNTRs that passed and failed Hardy–Weinberg equilibrium (HWE) test. The similarity score shows the highest number of matched nucleotides between a motif and a flanking sequence normalized by the length of the motif. VNTRs that have similar pattern with flanking regions are challenging because of the ambiguous boundary.

Figure S5. Novel variants detected by code-adVNTR in 4 long coding VNTRs, Related to Figure 3.

Novel variants detected by code-adVNTR within VNTRs in (A) *NEFH*, (B) *FMN2*, (C) *KRT-9*, and (D) *RP1L1* gene.

Supplemental Tables

Table S1. Short coding VNTRs, Related to STAR Methods.

Table S2. Short coding polymorphic VNTRs, Related to STAR Methods.

Table S3. Long coding VNTRs, Related to STAR Methods.

Table S4. Novel small variants found in long coding VNTRs, Related to STAR Methods.

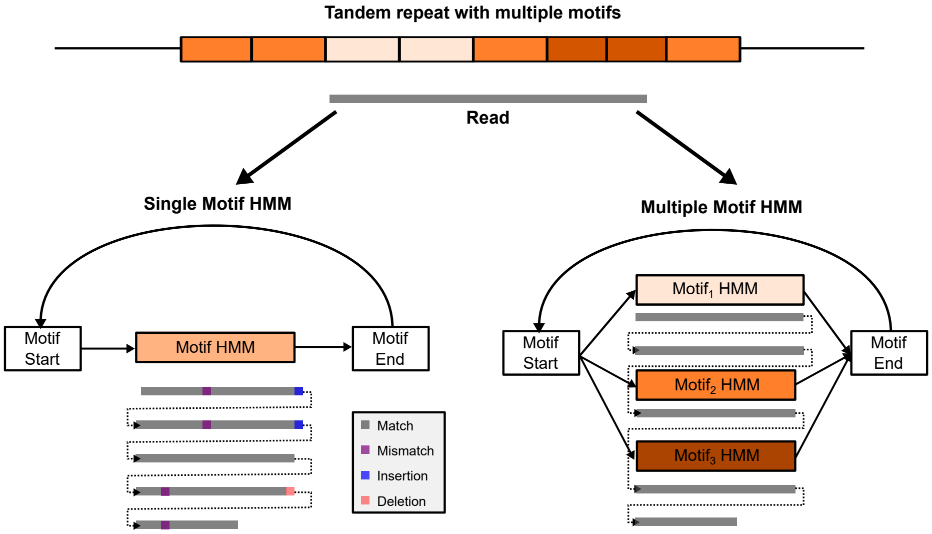

Figure S1: **Comparison between single motif and multiple motif HMM.** An example of a tandem repeat with multiple motifs. When a read is sequenced from the tandem repeat region, single motif HMM would match the variations within motifs with mismatch/insertion/deletion states.

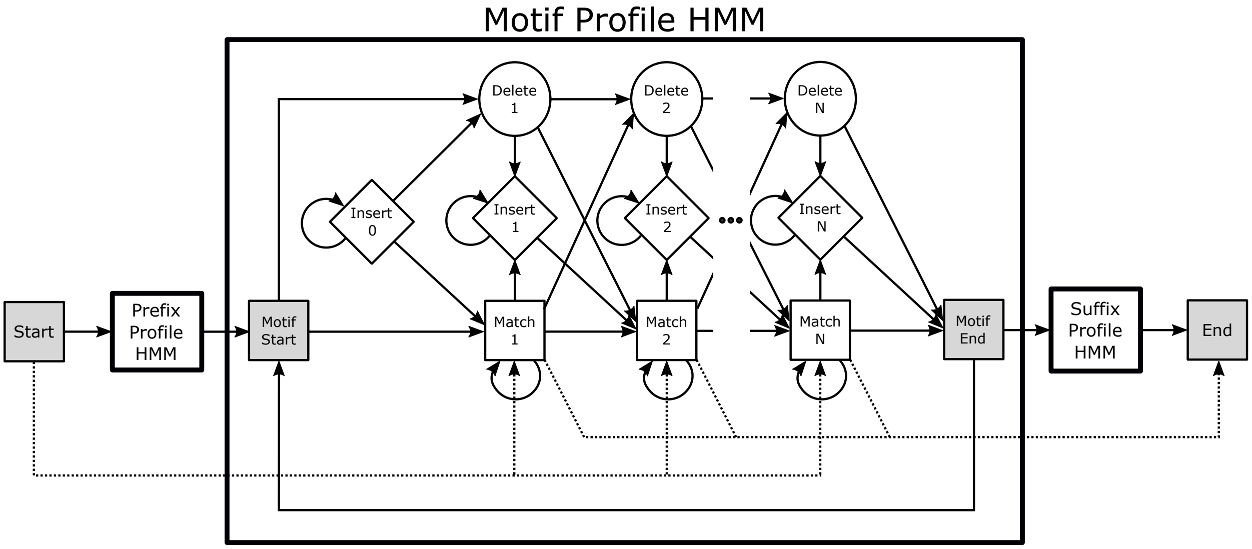

Figure S2: **The HMM for VNTRs.** The HMM is composed of three profile HMMs: prefix and suffix profile HMMs to align the flanking regions, and motif profile HMM to align multiple and partial number of motifs. Dotted lines connected to Start and End states indicate the transitions for reads that do not span the entire VNTR region.

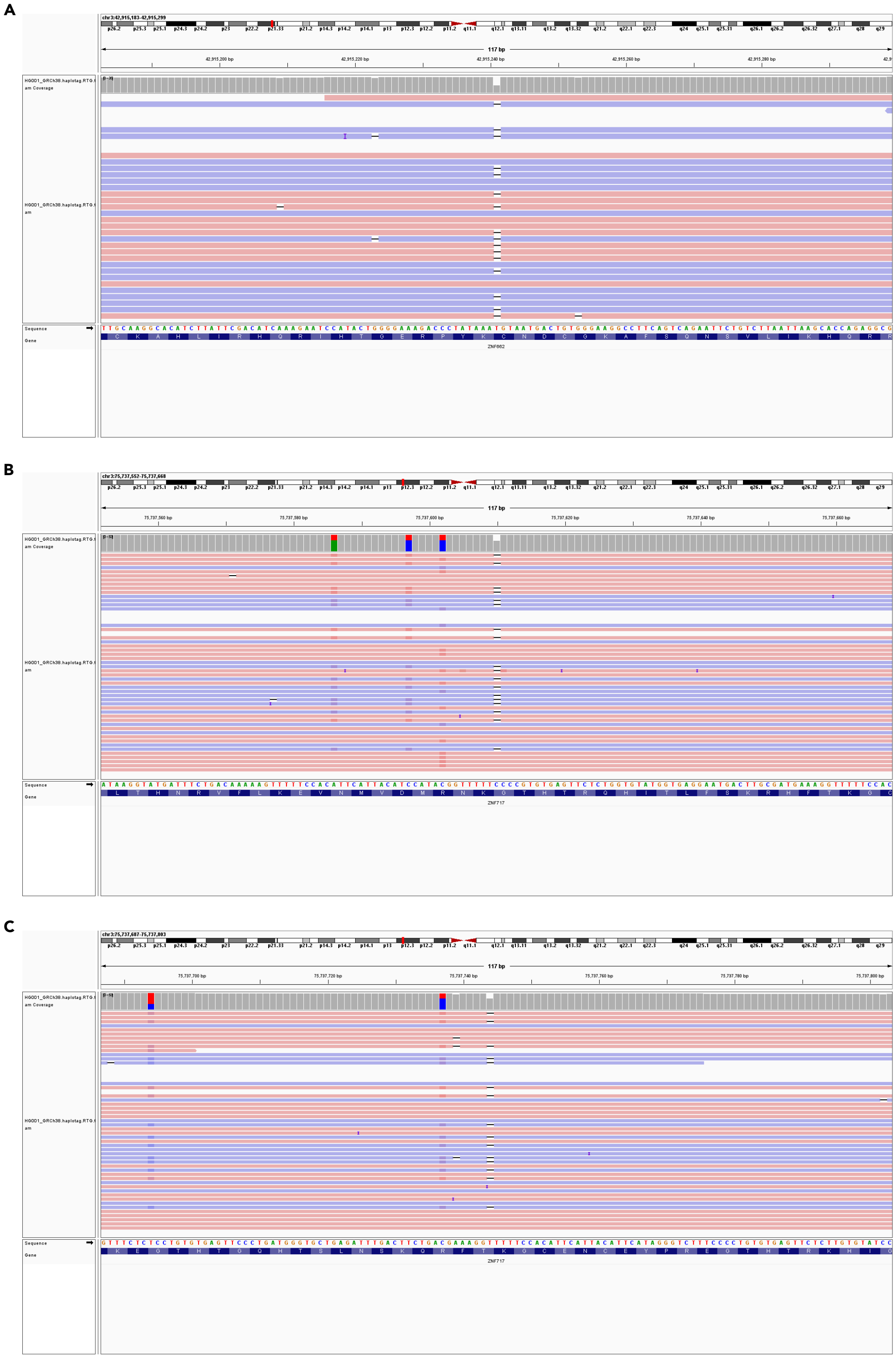
A
chr3:42,915,183-42,915,299
117 bp
ZNF662
B
chr3:75,737,552-75,737,668
117 bp
ZNF717
C
chr3:75,737,687-75,737,803
117 bp
ZNF717
Sequence
Gene

Figure S3: **IGV screenshots of long read mapping on genomic sample NA12878 showing the small indels detected by code-adVNTR.** Long read validation of three single base heterozygous deletions identified at 2 VNTR loci in ZNF662 and ZNF717 by code-adVNTR using short read.

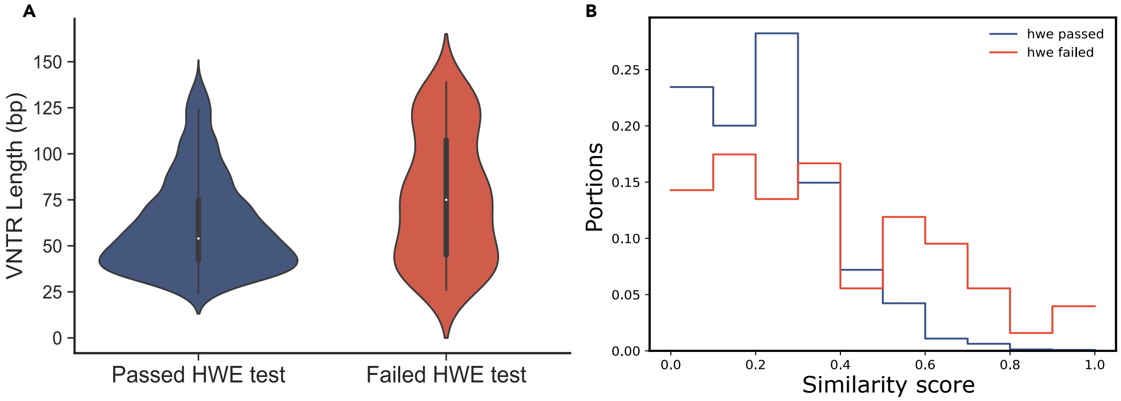

Figure S4: **Comparison of two groups of VNTRs that passed and failed Hardy-Weinberg equilibrium test.** (A) Distribution of VNTR lengths and (B) similarity scores of VNTRs that passed and failed Hardy–Weinberg equilibrium (HWE) test. The similarity score shows the highest number of matched nucleotides between a motif and a flanking sequence normalized by the length of the motif. VNTRs that have similar pattern with flanking regions are challenging because of the ambiguous boundary.

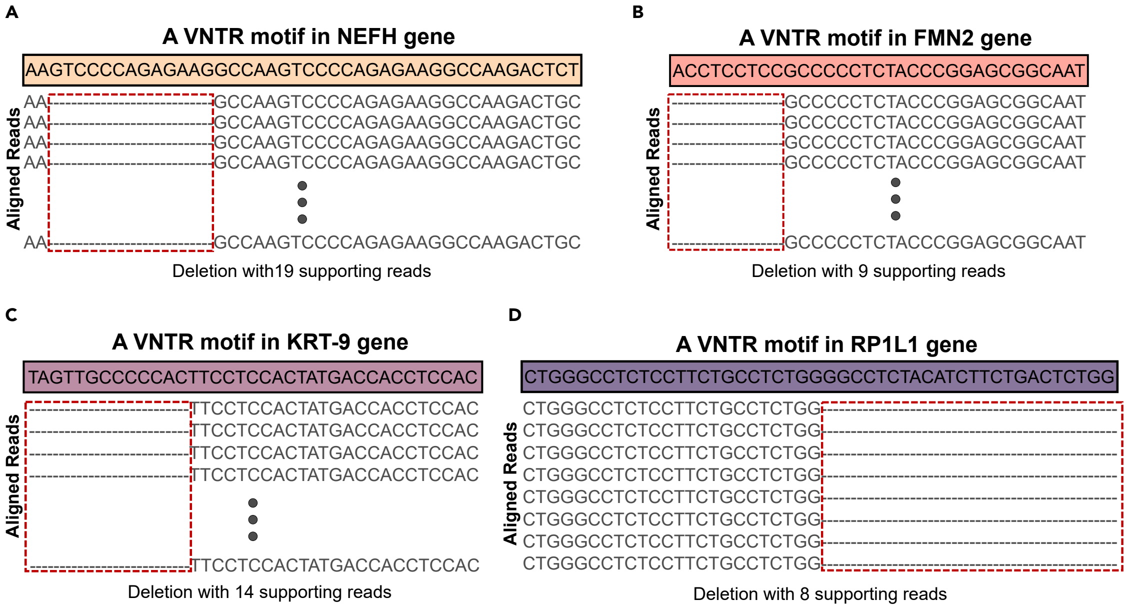
A
A VNTR motif in NEFH gene
AAGTCCCCAGAGAAGGCCAAGTCCCCAGAGAAGGCCAAGACTCT
Aligned Reads
AA
GCCAAGTCCCCAGAGAAGGCCAAGACTGC
Deletion with19 supporting reads
B
A VNTR motif in FMN2 gene
ACCTCCTCCGCCCCCTCTACCCGGAGCGGCAAT
Aligned Reads
GCCCCCTCTACCCGGAGCGGCAAT
Deletion with 9 supporting reads
C
A VNTR motif in KRT-9 gene
TAGTTGCCCCCACTTCCTCCACTATGACCACCTCCAC
Aligned Reads
TTCCTCCACTATGACCACCTCCAC
Deletion with 14 supporting reads
D
A VNTR motif in RP1L1 gene
CTGGGCCTCTCCTTCTGCCTCTGGGGCCTCTACATCTTCTGACTCTGG
Aligned Reads
CTGGGCCTCTCCTTCTGCCTCTGG
Deletion with 8 supporting reads

Figure S5: **Novel variants detected by code-adVNTR in 4 long coding VNTRs.** Novel variants detected by code-adVNTR within VNTRs in (A) *NEFH*, (B) *FMN2*, (C) *KRT-9*, and (D) *RP1L1* gene.