ViralFusionSeq (VFS) vs Virus-Clip

Overview

Published in March 2013, ViralFusionSeq (VFS) [1] was among the first wave of viral integration analysis tools [2-4]. ViralFusionSeq (VFS) is a versatile high-throughput sequencing (HTS) tool for discovering viral integration events and reconstruct fusion transcripts at single-base resolution. VFS was the first tool to combine soft-clipping information, read-pair analysis, and targeted de novo assembly to discover and annotate viral-human fusion events. VFS employs a simple yet effective empirical statistical model is used to evaluate the quality of fusion breakpoints.

Virus-Clip [5] was published much later in Q3 2015, falsely claimed to have demonstrated "*a significant advancement, compared to existing tool*", with a highlight of single-base resolution with focus on speed and memory-efficiency. *In their paper, they evaded a fair comparison with VFS because they failed to execute VFS. The authors have never contacted me for any assistance on installation.*

Technically, Virus-Clip simply executes BWA-Mem, extract soft-clipped sequences and directly ANNOVAR & BLAST the clipped sequences. Its sole reliance on soft-clipped reads with no consideration of sequencing error and erroneous mapping made Virus-Clip a poor choice for next-generation sequencing analysis of viral integration.

In this comparison, I intended to compare Virus-Clip with VFS using the two RNA-Seq data mentioned in their paper. Surprisingly, they did not deposit their sequencing data in public database. Thus, I could only use the dataset CL7R (used in VFS publication, and deposited in NCBI SRA **SRR619716)** to perform a fair comparison. Default setting was used in both tools. The full result reported by **Virus-Clip** and **VFS** are attached as **Table 1** and **Table 2-3**, respectively. Major findings are as follows,

- **Virus-Clip does not report any read-level evidence of viral-integration**. Essentially, multiple reads were merged (a.k.a piled up) and the longest integration sites were reported (See Table 1 below; sequence highlighted in red; indication of merging of more than 1 read based on manual checking of their sequence support in intermediate file. Therefore, sub-clonal differences in viral integration sites would be missed.
- **No quality metrics were used in Virus-Clip's analysis**. Users has to trust the leap of faith when using Virus-Clip.
- **Virus-Clip missed genuine viral integration site**. The real-time PCR validated TRRAP fusion in the CL7R dataset was missed by Virus-Clip due to its sole reliance on soft-clipped sequences.

Conclusion

Users should carefully choose the tool to analyze viral-integration, newer tools are not necessarily better.

References

1. Li, J.-W. et al. "Viralfusionseq: Accurately Discover Viral Integration Events And Reconstruct Fusion Transcripts At Single-Base Resolution". *Bioinformatics* 29.5 (2013): 649-651.
2. Chen, Y. et al. "Virusseq: Software To Identify Viruses And Their Integration Sites Using Next-Generation Sequencing Of Human Cancer Tissue". Bioinformatics 29.2 (2012): 266-267. Web.
3. Wang, Qingguo, Peilin Jia, and Zhongming Zhao. "Virusfinder: Software For Efficient And Accurate Detection Of Viruses And Their Integration Sites In Host Genomes Through Next Generation Sequencing Data". PLoS ONE 8.5 (2013): e64465
4. Wang, Qingguo, Peilin Jia, and Zhongming Zhao. "VERSE: A Novel Approach To Detect Virus Integration In Host Genomes Through Reference Genome Customization". Genome Medicine 7.1 (2015): 2.
   - Note: VERSE is the subsequent improvement made to VirusFinder2
5. Ho, Daniel WH, Karen MF Sze, and Irene OL Ng. "Virus-Clip: A Fast And Memory-Efficient Viral Integration Site Detection Tool At Single-Base Resolution With Annotation Capability". *Oncotarget* 6.25 (2015): 20959-20963.

Tables

Table 1: Result file of Virus-Clip on CL7R dataset

Left_element	Left_chr	Left_pos	Left_seq	Right_element	Right_chr	Right_pos	Right_seq	read_count	Gene_region	Human_gene
Human	chr13	22919137	GTCTGATGTTCAGTTTTCTGAGAAAAGGTGCAAGTTTAAA	Virus	chrVirus	1324	ATCGGCACCGACAACTCTGTTGTCCTCTCTCGGAAATTCACCTCCTTTCCATGGCTGCTAG	1	intergenic	LINC00424(dist=466838),BASP1P1(dist=552032)
Human	chr16	30408354	CAGGTGTGAGCCACCCTGCTGGCCTGGCTTTGCTACTTTCTAACCTTTATAACCTCCAGCAAAAGGA	Virus	chrVirus	1835	CTAATCATCTCATGTTCATGTTCCACTGTTCAAGCCTCCAAGCTGTGCCTTGGGTGGCTTTGGGAGCATGGACATTGACCCGTATAAAGAATTT	11	intronic	ZNF48
Human	chr16	30407192	CCACCGCCCGGAGGAGAGGGAGCCACAGAGAGGCGCCCGCAC	Virus	chrVirus	2348	AGACGACGAGGCAGGTCCCCTAGAAGAAGAACTCCCTCGCCTCGCAGACGAAG	1	exonic	ZNF48
Virus	chrVirus	2443	CGACGAGGCAGGTCCCCTAGAAGAAGAACTCCCTCGCCTCGCAGACGAAGGTCTCAATCGCCGCGTCGCAGAAGATCTCAATCTCGGGAATCT	Human	chr7	105673339	AAATATAATATTTCACCTCTTCCTGAAATGATACAGGAACATTTCAGAACACGCAACTATGTGATCAAAAC	10	UTR3	CDHR3

Table 2: Clipped-Seq result by ViralFusionSeq

HWI-ST977:195:C0N43ACXX:7:2304:20844:31422_Cs_CON_consensus	50 0	50 0	50 0	10 0	0	1 0	1 0	0	45	56	45	CTAATCATCTCATGTTCATGTTCCAACTGTTCAAGCCTCCAAGCTGTGCCTTGGGTG	CCTGGCTTTGCTACTTTCTAACCTTTATAACCTCCAGCAAAAGGA	CCTGGATGCCGGGTCTTCCAAATTACTTCCCACCCAGGTGGCCAGATTCATCAACTCACCCAACACAGAATAGCTTGCCTTAGTGCTGTATGGTGAGGTG	Hepatitis B virus strain CH008, complete genome CH036, CH018, isolate c81 precore/core protein (C) gene, partial cds c80 c53 c50 c49 c44 c14 B577 B571 >gi|1013126622|gb|GQ278135.1| B579 B567 S544 B718 b118-s A02291 promoter region and A01346 pt16N, pt1T, pt1N, 1346.7, 1346.30, 1346.29, 1346.24, 1346.22, 1346.15, 1240.6, 1240.4b, 1240.17a, 951.18a, 951.11a, 274.7, 274.18a, 1346.8, 1346.3, 1346.21, 1346.11, 1240.9b, 1240.7a, 1240.5, 1240.3b, 1240.21a, 1240.1b, 1240.13, 951.8a, 951.5a, 951.2, 951.1a, 951.15, 951.14a, 951.13a, 274.19a, 1418 X (X), (C), core genes, 1304 97 84 30 220 216 109 cxa2075, cxa2074, cxa2029, Occult_HK514, Occult_HK317, Occult_HK526, HK_1323, CX026M(e242), CX026C(e241), BCP-E242-M-1 >gi|605057603|gb|KJ172961.1| BCP-E242-M-2 BCP-E242-E02-M BCP-E242-27-5-8-M BCP-E241-B11-C BCP-E241-B10-C11-2 BCP-E241-B09-C11-2 BCP-E241-B09-C BCP-E241-B08-C11-2 >gi|605057582|gb|KJ172954.1| BCP-E241-B11-C-2 >gi|605057585|gb|KJ172955.1| BCP-E241-B12-C >gi|605057588|gb|KJ172956.1| BCP-E241-C-1 >gi|605057591|gb|KJ172957.1| BCP-E241-C-2 BCP-E241-B07-C11-2 BCP-E241-B07-C BCP-E241-B06-C BCP-E241-B03-C BCP-E241-B02-C A01622 A01339 A01240 B691 (HBx) preC/C (preC/C) B690 B685 B682 B680 B566 30, Cgeno60 Cgeno50 Cgeno38 Cgeno32 Cgeno22 Cgeno19 Cgeno06 clone C28.8 (X) C28.10 C27.2 TL036 TL033 TL030 TL001 migrant 3413 HB419, BH38 M92, S439 J239, J169, J83, I228, 179-96-1, G62K-8, GZ28-1, 2-1, Id, HK1248, HK926, HK352, HK2074, HK1386, HK1402, HK2032, HK947, HK554, HK1159, HK884, DNA, genome, isolate: HBV-VH34-1 2e, C-1858 ea6, ea3, ea2, gene for S protein, cds, isolate:Hawaii 16 GZ-DYH, Q80,	Homo sapiens chromosome 16 clone VMRC53-384H18, complete sequence VMRC53-357D8, Pongo abelii CH276-200C7, Pan troglodytes CH251-74E21, VMRC53-130K8, CH251-138I21, CH276-479H20, zinc finger protein 48 (ZNF48), RefSeqGene on BAC CH17-440E11 from unknown, RP11-297C4,
HWI-ST977:195:C0N43ACXX:7:2111:20146:84766_Cs_CON_consensus	50 0	50 0	50 0	10 0	0	1 0	1 0	0	31	71	31	CTAATCATCTCATGTTCATGTTCCAACTGTTCAAGCCTCCAAGCTGTGCCTTGGTGGCTTTGGAGCATGG	TTTCTAACCTTTATAACCTCCAGCAAAGGA	TGCCCGGGTCTTCCAAATTACTTCCCACCCAGGTGGCCAGATTCATCAACTCACCCCAACACAGAATAGCTTGCCTTAGTGCTGTATGGTGAGGTGGAACAAT	Hepatitis B virus isolate 75-HCC, complete genome	Homo sapiens chromosome 16 clone VMRC53-384H18, complete sequence VMRC53-357D8, Pongo abelii CH276-200C7, Pan troglodytes CH251-74E21, VMRC53-130K8, CH251-138I21, CH276-479H20, zinc finger protein 48 (ZNF48), RefSeqGene on BAC CH17-440E11 from unknown, RP11-297C4,
HWI-ST977:195:C0N43ACXX:7:1211:6985:23721_Cs_CON_consensus	50 0	50 0	50 0	10 0	0	1 0	1 0	0	31	71	31	CTAATCATCTCATGTTCATGTTCCAACTGTTCAAGCCTCCAAGCTGTGCCTTGGTGGCTTTGGAGCATGG	TTTCTAACCTTTATAACCTCCAGCAAAGGA	CACAATCACAGCTCACTGCAGCCTCCACCTCCTGGGCTCAAGTGATCCTCCTACCTCAGCCTCCCAGTAGCTGGGACTACAGGTGTGCACCACCATGCCCA	Hepatitis B virus isolate 75-HCC, complete genome	Homo sapiens chromosome 16 clone VMRC53-384H18, complete sequence VMRC53-357D8, Pongo abelii CH276-200C7, Pan troglodytes CH251-74E21, VMRC53-130K8, CH251-138I21, CH276-479H20, zinc finger protein 48 (ZNF48), RefSeqGene on BAC CH17-440E11 from unknown, RP11-297C4,
HWI-ST977:195:C0N43ACXX:7:2211:11842:8694_Cs_CON_consensus	50 0	50 0	50 0	10 0	0	1 0	1 0	0	47	54	47	CTAATCATCTCATGTTCATGTTCCAACTGTTCAAGCCTCCAAGCTGTG	CCCTGCTGGCTACTTTCTAACCTTTATAACCTCCAGCAAAGGA	GGTGGCCAGATTCATCAACTCACCCCAACACAGAATAGCTTGCCTTAGTGCTGTATGGTGAGGTGAACAATGTTCCGGAGACTCTAAGGCCTCCCGATACA	Hepatitis B virus strain CH008, complete genome CH036, CH018, isolate c81 precore/core protein (C) gene, partial cds c80 c53 c50 c49 c44 c14 B577 B571 >gi|1013126622|gb|GQ278135.1| B579 B567 S544 B718 b118-s A02291 promoter region and A01346 pt16N, pt1T, pt1N, 1346.7, 1346.30, 1346.29, 1346.24, 1346.22, 1346.15, 1240.6, 1240.4b, 1240.17a, 951.18a, 951.11a, 274.7, 274.18a, 1346.8, 1346.3, 1346.21, 1346.11, 1240.9b, 1240.7a, 1240.5, 1240.3b, 1240.21a, 1240.1b, 1240.13, 951.8a, 951.5a, 951.2, 951.1a, 951.15, 951.14a, 951.13a, 274.19a, 1418 X (X), (C), core genes, 1304 97 84 30 220 216 109 cxa2075, cxa2074, cxa2029, Occult_HK514, Occult_HK317, Occult_HK526, HK_1323, CX026M(e242), CX026C(e241), BCP-E242-M-1 >gi|605057603|gb|KJ172961.1| BCP-E242-M-2 BCP-E242-E02-M BCP-E242-27-5-8-M BCP-E241-B11-C BCP-E241-B10-C11-2 BCP-E241-B09-C11-2 BCP-E241-B09-C BCP-E241-B08-C11-2 >gi|605057582|gb|KJ172954.1| BCP-E241-B11-C-2 >gi|605057585|gb|KJ172955.1| BCP-E241-B12-C >gi|605057588|gb|KJ172956.1| BCP-E241-C-1 >gi|605057591|gb|KJ172957.1| BCP-E241-C-2 BCP-E241-B07-C11-2 BCP-E241-B07-C BCP-E241-B06-C BCP-E241-B03-C BCP-E241-B02-C A01622 A01339 A01240 561 nonfunctional sequence B691 (HBx) preC/C (preC/C) B690 B685 B682 B680 B566 30, Cgeno60 Cgeno50 Cgeno38 Cgeno32 Cgeno22 Cgeno19 Cgeno06 clone C28.8 (X) C28.10 C27.2 TL036 TL033 TL030 TL001 migrant 3413 HB419, BH38 M92, S439 J239, J169, J83, I228, 179-96-1, G62K-8, GZ28-1, 2-1, Id, HK1248, HK926, HK352, HK2074, HK1386, HK1402, HK2032, HK947, HK554, HK1159, HK884, DNA, genome, isolate: HBV-VH34-1 2e, C-1858 ea6, ea3, ea2, gene for S protein, cds, isolate:Hawaii 16 GZ-DYH, Q80,	Homo sapiens chromosome 16 clone VMRC53-384H18, complete sequence VMRC53-357D8, Pan troglodytes CH251-74E21, VMRC53-130K8, CH251-138I21, zinc finger protein 48 (ZNF48), RefSeqGene on BAC CH17-440E11 from unknown, RP11-297C4,
HWI-ST977:195:C0N43ACXX:7:2111:3592:23479_Cs_CON_consensus	50 0	50 0	50 0	10 0	0	1 0	1 0	0	31	71	31	CTAATCATCTCATGTTCATGTTCCAACTGTTCAAGCCTCCAAGCTGTGCCTTGGTGGCTTTGGAGCATGG	TTTCTAACCTTTATAACCTCCAGCAAAGGA	CACAATCACAGCTCACTGCAGCCTCCACCTCCTGGGCTCAAGTGATCCTCCTACCTCAGCCTCCCAGTAGCTGGGACTACAGGTGTGCACCACCATGCCCA	Hepatitis B virus isolate 75-HCC, complete genome	Homo sapiens chromosome 16 clone VMRC53-384H18, complete sequence VMRC53-357D8, Pongo abelii CH276-200C7, Pan troglodytes CH251-74E21, VMRC53-130K8, CH251-138I21, CH276-479H20, zinc finger protein 48 (ZNF48), RefSeqGene on BAC CH17-440E11 from unknown, RP11-297C4,

Table 3: Paired-end result by VialFusionSeq

Name	c1	c2	Read ID 1	d	str	CIGAR1	f1	f2	f3	f4	f5	f6	f7	f8	f9	f10	Sequence 1	p1	p2	p3	Read ID 2	q	str2	CIGAR2	r	s1	s2	Gene annotation	t1	t2	u	v1	v2	v3	Repeat annotation	w	Sequence 2
Cs_CON_consensus	1872	1973	HWI-ST977:195:C0N43ACXX:7:1314:20027:59146/2	0	-	101M	3	0	0	0	0	0	0	0	0	1	GGAAAAAGTCAGAAGGCAAAAAAGAGAGTAACTCCACAGAAGCTCCAAATTCTTTATACGGGTCAATGTCCATGCTCCCAAAGCCACCCAAGGCACAGCTTG	28901242	28901343	28901343	HWI-ST977:195:C0N43ACXX:7:1314:20027:59146/1	37	+	101M	2	28900684	28900693	RN5S89;rRNA;KNOWN;RNA, 5S ribosomal 89 [Source:HGNC Symbol;Acc:42887];p23.2	5499	0	F	34589	28900496	28900495	MER91A;346;+;DNA;hAT-Tip100	647	CTTCGGTCAACATGAAGGAGCAGCCACAGCCATTTCAAGGTCACTTGTGGAATTGCGTCCACGTGGACTCCGGTGACTCCTGTGCTCTTCCTTTTACTCCG
HBV_3215	1872	1973	HWI-ST977:195:C0N43ACXX:7:1214:2860:38341/1	0	-	101M	3	0	0	0	0	0	0	0	0	1	GGAAAAAGTCAGAAGGCAAAAAAGAGAGTAACTCCACAGAAGCTCCAAATTCTTTATCGGGTCAATGTCCATGCTCCAAAGCCACCCAAGGCACAGCTTG	28901242	28901343	28901343	HWI-ST977:195:C0N43ACXX:7:1214:2860:38341/2	37	+	101M	2	28900684	28900693	RN5S89;rRNA;KNOWN;RNA, 5S ribosomal 89 [Source:HGNC Symbol;Acc:42887];p23.2	5499	0	F	34589	28900496	28900495	MER91A;346;+;DNA;hAT-Tip100	647	CTTCGGTCAACATGAAGGAGCAGCCACAGCCATTTCAAGGTCACTTGTGGAATTGCGTCCACGTGGACTCCGGTGACTCCTGTGCTCTTCCTTTTACTCCG
Cs_CON_consensus	1872	1973	HWI-ST977:195:C0N43ACXX:7:1308:14770:8730/2	0	-	101M	3	0	0	0	0	0	0	0	0	1	GGAAAAAGTCAGAAGGCAAAAAGAGAGTAACTCCACAGAAGCTCCAAATTCTTTATACGGGTCAATGTCCATGCTCCAAAGCCACCCAAGGCACAGCTTG	28901242	28901343	28901343	HWI-ST977:195:C0N43ACXX:7:1308:14770:8730/1	37	+	101M	2	28900684	28900693	RN5S89;rRNA;KNOWN;RNA, 5S ribosomal 89 [Source:HGNC Symbol;Acc:42887];p23.2	5499	0	F	34589	28900496	28900495	MER91A;346;+;DNA;hAT-Tip100	647	CTTCGGTCAACATGAAGGAGCAGCCACAGCCATTTCAAGGTCACTTGTGGAATTGCGTCCACGTGGACTCCGGTGACTCCTGTGCTCTTCCTTTTACTCCG
HBV_3215	1872	1973	HWI-ST977:195:C0N43ACXX:7:1203:2476:70775/1	0	-	101M	3	0	0	0	0	0	0	0	0	1	GGAAAAAGTCAGAAGGCAAAAAGAGAGTAACTCCACAGAAGCTCCAAATTCTTTATACGGGTCAATGTCCATGCTCCAAAGCCACCCAAGGCACAGCTTG	28901242	28901343	28901343	HWI-ST977:195:C0N43ACXX:7:1203:2476:70775/2	37	+	101M	2	28900684	28900693	RN5S89;rRNA;KNOWN;RNA, 5S ribosomal 89 [Source:HGNC Symbol;Acc:42887];p23.2	5499	1	F	34589	28900496	28900495	MER91A;346;+;DNA;hAT-Tip100	647	CTTCGGTCAACATGAAGGAGCAGCCACAGCCATTTCAAGGTCACTTGTGGAATTGCGTCCACGTGTACTCCGGTGACTCCTGTGCTCTTCCTTTTACTCG
HBV_3215	1872	1973	HWI-ST977:195:C0N43ACXX:7:1211:5204:96420/2	0	-	101M	3	0	0	0	0	0	0	0	0	1	GGAAAAAGTCAGAAGGCAAAAAGAGAGTAACTCCACAGAAGCTCCAAATTCTTTATCGGGTCAATGTCCATGCTCCAAAGCCACCCAAGGCACAGCTTG	28901242	28901343	28901343	HWI-ST977:195:C0N43ACXX:7:1211:5204:96420/1	37	+	101M	2	28900684	28900693	RN5S89;rRNA;KNOWN;RNA, 5S ribosomal 89 [Source:HGNC Symbol;Acc:42887];p23.2	5499	0	F	34589	28900496	28900495	MER91A;346;+;DNA;hAT-Tip100	647	CTTCGGTCAACATGAAGGAGCAGCCACAGCCATTTCAAGGTCACTTGTGGAATTGCGTCCACGTGGACTCCGGTGACTCCTGTGCTCTTCCTTTTACTCCG
HBV_3215	1872	1973	HWI-ST977:195:C0N43ACXX:7:1111:12847:42793/2	0	-	101M	3	0	0	0	0	0	0	0	0	1	GGAAAAAGTCAGAAGGCAAAAAGAGAGTAACTCCACAGAAGCTCCAAATTCTTTATCGGGTCAATGTCCATGCTCCAAAGCCACCCAAGGCACAGCTTG	28901242	28901343	28901343	HWI-ST977:195:C0N43ACXX:7:1111:12847:42793/1	37	+	101M	2	28900684	28900693	RN5S89;rRNA;KNOWN;RNA, 5S ribosomal 89 [Source:HGNC Symbol;Acc:42887];p23.2	5499	0	F	34589	28900496	28900495	MER91A;346;+;DNA;hAT-Tip100	647	CTTCGGTCAACATGAAGGAGCAGCCACAGCCATTTCAAGGTCACTTGTGGAATTGCGTCCACGTGGACTCCGGTGACTCCTGTGCTCTTCCTTTTACTCCG
HK2100	2238	2380	HWI-ST977:195:C0N43ACXX:7:1104:13127:68273/2	0	+	101M	0	1	0	0	0	0	0	0	0	1	CACTCCTCCGCTTACAGACCACCAAATGCCCCTATCTTACAACACTTCCTCACACTTCCGGAAACTACTGTTGTTAGACGACGAGGCAGGTCCCTAGAAGAAGAACTC	7885577672778	9555277657866	9555277657866	HWI-ST977:195:C0N43ACXX:7:1104:13127:68273/1	37	-	101M	7	9847550566	9615586566	TRRAP;protein_coding;KNOWN;transformation/transcription domain-associated protein [Source:HGNC Symbol;Acc:12347];q22.1	0	1	T	0	98557292175	85527921175	AluSp;2441;+;SINE;Alu	137	TGGGGCCACCTCCCCGATTTTGCAAAGCTCCTTTTCTTCATATGCCATACAAATAGCAATTGCATCAATGAGAACCAAAGGATCCATTCCTTTCGGAGCCATT
Cs_CON_consensus	2290	2380	HWI-ST977:195:C0N43ACXX:7:2106:9531:42948/1	0	+	98M	0	1	0	0	0	0	0	0	0	1	GCTTACAGACCACCAAATGCCCCTATCTTATCAACACTTCCGGAAACTACTGTTGTTAGACGACGAGGCAGGTCCCTAGAAGAAGAACTCCCTCGCC	7885577332125553	9555277322557654	9555277322557654	HWI-ST977:195:C0N43ACXX:7:2106:9531:42948/2	37	-	101M	7	9847550566	9615586566	TRRAP;protein_coding;KNOWN;transformation/transcription domain-associated protein [Source:HGNC Symbol;Acc:12347];q22.1	0	0	F	9447	985532131361	985532131361	MER5B;307;-;DNA;hAT-Charlie	0	CTCGAGTCCATGCTTTTCAAAAGATCCGCACGTGTTCTGATGCAGCCAGGCACTGAGAGACATCTGGAAAACCAGTTTCATGCCCTCAATACTGCCAGGCT
HBV_3215	2237	2381	HWI-ST977:195:C0N43ACXX:7:2110:3171:65250/1	0	+	101M	0	1	0	0	0	0	0	0	0	1	GTGGATTCGCACTCCTCCCGCTTACAGACCACCAAATGCCCCTATCTTATCAACACTTCCGGAAACTACTGTTGTTAGACGACGAGGCAGGTCCCTAGAA	78855773212553	95552773225584	95552773225584	HWI-ST977:195:C0N43ACXX:7:2110:3171:65250/2	37	-	101M	7	9847550566	9615586566	TRRAP;protein_coding;KNOWN;transformation/transcription domain-associated protein [Source:HGNC Symbol;Acc:12347];q22.1	0	0	F	9447	985532131361	985532131361	MER5B;307;-;DNA;hAT-Charlie	0	CTCGAGTCCATGCTTTTCAAAAGATCCGCACGTGTTCTGATGCAGCCAGGCACTGAGAGACATCTGGAAAACCAGTTTCATGCCCTCAATACTGCCAGGCT
Cs_CON_consensus	1897	1958	HWI-ST977:195:C0N43ACXX:7:1111:14729:16111/1	23	+	101M	1	0	0	0	0	0	0	0	0	1	CACTGTTCAAGCCTCCAAGCTGTGCCTTGGGTGGCTTTGGAGCATGGACATTGACCCGTATAAAGAATTTGGAGCTTCTGTGGAGTTACTCTCTTTTTGC	7885322155553	9555847157556	9555847157556	HWI-ST977:195:C0N43ACXX:7:1111:14729:16111/2	37	-	101M	7	9847550566	9815586566	TRRAP;protein_coding;KNOWN;transformation/transcription domain-associated protein [Source:HGNC Symbol;Acc:12347];q22.1	0	0	F	919	985532131361	985532131361	MER5B;307;-;DNA;hAT-Charlie	0	GACTGAACCTGAATCTCCAAGGAAAGGACTCGAGTCCATGCTTTTCAAAAGATCCGCACGTGTTCTGATGCAGCCAGGCACTGAGAGACATCTGGAAAACC
Cs_CON_consensus	1895	1958	HWI-ST977:195:C0N43ACXX:7:21	23	+	101M	1	0	0	0	0	0	0	0	0	1	CACTGTTCAAGCCTCCAAGCTGTGCCTTGGGTGGCTTTGGAGCATGGACATTGACCCGTATAAAGAATTTGGAGCTTCT	7885322125	9555847157	9555847157	HWI-ST977:195:C0N43ACXX:7:21	37	-	101M	7	9847550566	9815586560	TRRAP;protein_coding;KNOWN;transformation/transcription domain-associated protein [Source:HGNC Symbol;Acc:12347];q22.1	0	0	F	919	98553131322	98553131322	MER5B;307;-;DNA;hAT-Charlie	0	GACTGAACCTGAATCTCCAAGGAAAGGACTCGAGTCCATGCTTTTCAAAAGATCCGCACGTGTTCTGATGCAGCCAGG

Name	c1	c2	Read ID	v1	str	v2	v3	v4	v5	v6	v7	v8	v9	v10	v11	v12	v13	Sequence	p1	p2	p3	Read ID 2	q1	str2	q2	q3	q4	q5	Gene annotation	r1	r2	r3	s1	s2	s3	Sequence 2	t1	Repeat	t2	Sequence 3
sen sus			09:2266:81049/1															GTGGAGTTACTCTCTTTTTTGC	1 8 6	2 8	09:2266:81049/2											5 5 6	8 6			1 9 5	3 5 1			CACTGAGAGACATCTGGAAAACC
Cs_CON_con sen sus	1 8 5 1	1 9 5 2	HWI-ST977:195:C0N43ACXX:7:2315:3942:18450/1	3 7	+	1 0 1 M	1	0	0	0	0	0	0	0	0	1	ATGTTCCACTGTTCAAGCCTCCAAGCTGTGCCTTGGGTGGCTTTTGGAGCATGGACATTGACCCGTATAAAGAATTTGGAGCTTCTGTGGAGTTACTCTCTT	7	9 8 8 3 3 2 1 8 4	9 8 5 3 2 2 9 5	HWI-ST977:195:C0N43ACXX:7:2315:3942:18450/2	3 7	-	1 0 1 M	7	9 8 4 7 5 0 5 6	9 8 6 1 0 8 6 6	TRRAP;protein_coding;KNOWN;transformation/transcription domain-associated protein [Source:HGNC Symbol;Acc:12347];q22.1	0	0	F	9 1 6	9 8 5 2 1 9 5 6	9 8 5 3 3 5 1	MER5B;307;-;DNA;hA T-Charlie	0	ATGGACTGAACCTGAATCTCCAAGGAAAGGACTCGAGTCCATGCTTTTCAAAAGATCCGGGCACTGAGAGACATCTGGAAA
HK 210 0	2 2 9 1	2 3 9 2	HWI-ST977:195:C0N43ACXX:7:2309:10280:91267/1	0	+	1 0 1 M	0	1	0	0	0	0	0	0	0	1	CTTACAGACCACCAAATGCCCCTATCTTATCAACACTTCCGGAAACTACTGTTGTTAGACGACGAGGCAGGTCCCCTAGAAGAAGAACTCCCTGCCCTCGC	7	9 8 5 3 2 1 9	9 8 5 3 2 9 5	HWI-ST977:195:C0N43ACXX:7:2309:10280:91267/2	3 7	-	1 0 1 M	7	9 8 4 7 5 0 5 6	9 8 6 1 0 8 6 6	TRRAP;protein_coding;KNOWN;transformation/transcription domain-associated protein [Source:HGNC Symbol;Acc:12347];q22.1	0	0	F	9 1 1	9 8 5 3 2 1 9 5 6	9 8 5 3 3 5 1	MER5B;307;-;DNA;hA T-Charlie	0	CCCAAATGGACTGAACCTGAATCTCCAAGGAAAGGACTCGAGTCCATGCTTTTCAAAAGATCCGCACGTGTTCTGATGCAGCCAGGCACTGAGAGACATCT
Cs_CON_con sen sus	2 2 9 1	2 3 9 2	HWI-ST977:195:C0N43ACXX:7:1316:21068:4592/1	0	+	1 0 1 M	0	1	0	0	0	0	0	0	0	1	CTTACAGACCACCAAATGCCCCTATCTTATCAACACTTCCGGAAACTACTGTTGTTAGACGACGAGGCAGGTCCCCTAGAAGAAGAACTCCCTGCCCTCGC	7	9 8 5 3 2 1 8 0	9 8 5 3 2 9 0	HWI-ST977:195:C0N43ACXX:7:1316:21068:4592/2	3 7	-	1 0 1 M	7	9 8 4 7 5 0 5 6	9 8 6 1 0 8 6 6	TRRAP;protein_coding;KNOWN;transformation/transcription domain-associated protein [Source:HGNC Symbol;Acc:12347];q22.1	0	0	F	9 1 1	9 8 5 3 2 1 9 5 6	9 8 5 3 3 5 1	MER5B;307;-;DNA;hA T-Charlie	0	CCCAAATGGACTGAACCTGAATCTCCAAGGAAAGGACTCGAGTCCATGCTTTTCAAAAGATCCGCACGTGTTCTGATGCAGCCAGGCACTGAGAGACATCT
Cs_CON_con sen sus	1 8 5 6	1 9 5 7	HWI-ST977:195:C0N43ACXX:7:2304:18664:3429/1	2 3	+	1 0 1 M	1	0	0	0	0	0	0	0	0	1	CCACTGTTCAAGCCTCCAAGCTGTGCCTTGGGTGGCTTTGGAGCATGGACATTGACCCGTATAAAGAATTTGGAGCTTCTGTGGAGTTACTCTCTTTTTTG	7	9 8 5 3 2 1 9	9 8 5 3 2 9 5	HWI-ST977:195:C0N43ACXX:7:2304:18664:3429/2	3 7	-	1 0 1 M	7	9 8 4 7 5 0 5 6	9 8 6 1 0 8 6 6	TRRAP;protein_coding;KNOWN;transformation/transcription domain-associated protein [Source:HGNC Symbol;Acc:12347];q22.1	0	0	F	9 1 0	9 8 5 2 1 9 5 6	9 8 5 3 3 5 1	MER5B;307;-;DNA;hA T-Charlie	0	ACCCAAATGGACTGAACCTGAATCTCCAAGGAAAGGACTCGAGTCCATGCTTTTCAAAAGATCCGCACGTGTTCTGATGCAGCCAGGCACTGAGAGACATC
Cs_CON_con sen sus	1 8 8 7	1 9 8 7	HWI-ST977:195:C0N43ACXX:7:1216:18676:74965/2	2 5	+	1 0 1 M	5	0	0	0	0	0	0	0	0	1	GTGGCTTTGGAGCATGGACATTGACCCGTATAAAGAATTTGGAGCTTCTGTGGAGTTACTCTCTTTTTTGCCTTCTGACTTTTTCCCTTCTATTCGTGACA	7	9 8 5 3 2 2 2 3	9 8 5 3 2 2 2 4	HWI-ST977:195:C0N43ACXX:7:1216:18676:74965/1	3 7	-	1 0 1 M	7	9 8 4 7 5 0 5 6	9 8 6 1 0 8 6 6	TRRAP;protein_coding;KNOWN;transformation/transcription domain-associated protein [Source:HGNC Symbol;Acc:12347];q22.1	0	0	F	8 7 7	9 8 5 3 2 1 9 5 6	9 8 5 3 3 5 1	MER5B;307;-;DNA;hA T-Charlie	0	GGGAAGCTTTCAGACAGTGGATTCCCAGGTCTCACCCAAATGGACTGAACCTGAATCTCCAAGGAAAGGACTCGAGTCCATGCTTTTCAAAAGATCCGCAC
Cs_CON_con sen sus	1 8 7 7 1	1 9 7 2	HWI-ST977:195:C0N43ACXX:7:2114:15201:93191/1	0	+	1 0 1 M	3	0	0	0	0	0	0	0	0	1	CCAAGCTGTGCCTTGGGTGGCTTTGGAGCATGGACATTGACCCGTATAAAGAATTTGGAGCTTCTGTGGAGTTACTCTTTTTTGCCTTCTGACTTTTTC	7	9 8 5 3 2 2 2 5	9 8 5 3 2 2 6	HWI-ST977:195:C0N43ACXX:7:2114:15201:93191/2	3 7	-	1 0 1 M	7	9 8 4 7 5 0 5 6	9 8 6 1 0 8 6 6	TRRAP;protein_coding;KNOWN;transformation/transcription domain-associated protein [Source:HGNC Symbol;Acc:12347];q22.1	0	0	F	8 7 5	9 8 5 2 1 9 5 6	9 8 5 3 3 5 1	MER5B;307;-;DNA;hA T-Charlie	0	CTGGGAAGCTTTCAGACAGTGGATTCCCAGGTCTCACCCAAATGGACTGAACCTGAATCTCCAAGGAAAGGACTGAGTCCATGCTTTTCAAAAGATCCGC
HK 210 0	2 2 6 1	2 3 6 1	HWI-ST977:195:C0N43ACXX:7:1216:15134:44837/1	0	-	1 0 1 M	0	1	0	0	0	0	0	0	0	1	CCTGCCTCGTCGTCTAACAACAGTAGTTTCCGGAAGTGTTGATAAGATAGGGGCATTTGGTGGTCTGTAAGCGGGAGGAGTGCGAATCCACACTCCAAAAG	1 6	3 0 4 0 6 7 7	3 3 0 0 7 5 4	HWI-ST977:195:C0N43ACXX:7:1216:15134:44837/2	3 7	+	1 0 1 M	1 6	3 3 8 0 9 5 4	3 3 4 7 1 2	1-Sep;protein_coding;KNOWN;septin 1 [Source:HGNC Symbol;Acc:2879];p11.2, ZNF48;protein_coding;KNOWN;zinc finger protein 48 [Source:HGNC Acc:13114];p11.2	0	0	T	3 0 3 0 6 9 0 5	3 3 0 4 6 9 2	GC_rich;24;+;Low_complexity;Low_complexity	0	GGGAGGCGGCGCCCGTACCTGGGAACCCGACGCGCGCCCGGGCCCGCGCTGGCCGGCAGAGGGCAGCGGCAGGGCGCCGCGGGCGCACACATGGCTGCGTGACC
Cs_CON_con sen sus	1 9 0 0 7 8	1 2 0 0 7 8	HWI-ST977:195:C0N43ACXX:7:1307:19154:64924/1	1 8	-	1 0 1 M	3	1	0	0	0	0	0	0	0	1	AGCAGAAGCGGTGTCGAGGAGATCACGAATAGAAGGGAAAAGTCAGAAGGCAAAAAGAGAGTAACTCCACAGAAGCTCCAAATTCTTTATACGGGTCAA	1 6	3 0 4 0 8 1 1 1 7	3 3 0 4 8 1 1 2 7	HWI-ST977:195:C0N43ACXX:7:1307:19154:64924/2	3 7	+	1 0 1 M	1 6	3 0 3 8 9 4 2 7	3 3 4 1 2 9 2	ZNF48;protein_coding;KNOWN;zinc finger protein 48 [Source:HGNC Symbol;Acc:13114];p11.2	0	0	F	4 3 3 2	3 0 4 0 8 0 9	3 0 0 0 3 0 2	AluJb;2153;-;SINE;Alu	0	CTCCTACCTCAGCCTCCCAGTAGCTGGGACTACAGGTGTGCACCACCATGCCCAGCTAATTTTTGTATTTTTTTTTGTAGAGTTAGGGTCTCACCATGTTG
Cs_CON_con sen sus	1 8 9 8 9	1 8 9 9	HWI-ST977:195:C0N43ACXX:7:1306:6304:78455/1	3 7	-	1 0 1 M	3	0	0	0	0	0	0	0	0	1	AGATCACGAATAGAAGGGAAAAAGTCAGAAGGCAAAAAGAGAGTAACTCCACAGAAGCTCCAAATTCTTTATACGGGTCAATGTCCATGCTCCAAAGCCA	1 6	3 0 4 0 8 1 1 2	3 3 0 4 8 1 1 6 2	HWI-ST977:195:C0N43ACXX:7:1306:6304:78455/2	3 7	+	9 9 M	1 6	3 0 3 8 9 1 4 2 7	3 3 4 1 2 9 9	ZNF48;protein_coding;KNOWN;zinc finger protein 48 [Source:HGNC Symbol;Acc:13114];p11.2	0	0	F	3 8 9	3 0 4 0 8 0 9	3 0 0 0 3 0 2	AluJb;2153;-;SINE;Alu	0	CCATGCCCAGCTAATTTTTGTATTTTTTTTTGTAGAGTTAGGGTCTCACCATGTTGCCCAGGCTGGTTTCTAACTCCTGGGCTTAAGTGATCCACCCAC
Cs_CON_con sen sus	1 9 0 4 4 8	1 2 0 4 9	HWI-ST977:195:C0N43ACXX:7:1301:1832:12750/1	1 8	-	1 0 1 M	3	1	0	0	0	0	0	0	0	1	GGTGAACAATGTTCCGGAGACTCTAAGGCCTCCCGATACAGAGCAGAAGCGGTGTCGAGGGAGATCACGAATAGAAGGGAAAAAGTCAGAAGGCAAAAAGA	1 6	3 0 4 0 8 2 3 0 5	3 3 0 0 4 8 2 3 9 6	HWI-ST977:195:C0N43ACXX:7:1301:1832:12750/2	3 7	+	1 0 1 M	1 6	3 0 3 8 9 1 4 2 7	3 3 4 1 2 9 9	ZNF48;protein_coding;KNOWN;zinc finger protein 48 [Source:HGNC Symbol;Acc:13114];p11.2	0	0	F	3 4 4	3 0 0 0 8 3 2	3 3 0 4 9 2	MIRb;339;+;SINE;MIR	0	TCTCACCATGTTGCCCAGGCTGGTTTCTAACTCCTGGGCTTAAGTGATCCACCCACCTCGGCCTCCCAAAGTGCTAGGATTACAGGTGTGAGCCACCCTGC
Cs_CON_con sen sus	1 9 0 4 5	1 2 0 4 6	HWI-ST977:195:C0N43ACXX:7:1216:16878:72783/1	1 8	-	1 0 1 M	3	1	0	0	0	0	0	0	0	1	GAACAATGTTCCGGAGACTCTAAGGCCTCCCGATACAGAGCAGAAGCGGTGTCGAGGAGATCACGAATAGAAGGGAAAAAGTCAGAAGGCAAAAAGAGAG	1 6	3 0 4 0 8 2 3 0 8	3 3 0 4 8 2 3 9 2	HWI-ST977:195:C0N43ACXX:7:1216:16878:72783/2	3 7	+	1 0 1 M	1 6	3 0 3 8 9 1 4 2 7	3 3 4 1 2 9 9	ZNF48;protein_coding;KNOWN;zinc finger protein 48 [Source:HGNC Symbol;Acc:13114];p11.2	0	0	F	3 4 1	3 0 0 0 8 3 8	3 3 0 4 9 2	MIRb;339;+;SINE;MIR	0	CACCATGTTGCCCAGGCTGGTTTCTAACTCCTGGGCCTTAAGTGATCCACCCACCTCGGCCTCCCAAAGTGCTAGGATTACAGGTGTGAGCCACCCTGCTGG
Cs_CON_	1 9 0	1 2 0	HWI-ST977:195:C0N43A	1 8	-	1 0	3	1	0	0	0	0	0	0	0	1	GAACAATGTTCCGGAGACTCTAAGGCCTCCCGATACAGAGCAGAAGCGGTGTCGAGG	1 6	3 0 0 4	3 3 0 0 4	HWI-ST977:195:C0N43A	3 7	+	1 0 1 M	1 6	3 0 0 3	3 0 0 4	ZNF48;protein_coding;KNOWN;zinc finger protein 48	0	3	F	3 4 1	3 0 0 4 4	3 0 0 4 4	MIRb;339;+;SINE;MIR	0	CACCATGTTGCCCAGGCTGGTTTCTAACTCCTGGGCTTAAGTGATCCACCCACCTCGGC

Continuation data table (page 5 of 5). Column headers are not printed on this page.

Name	Pos1	Pos2	Read ID (left)	Q	Str	CIG	c1	c2	c3	c4	c5	c6	c7	c8	c9	c10	Sequence (left)	N	Coord1a	Coord1b	Read ID (right)	Sc	Str2	CIG2	N2	Coord2a	Coord2b	Annotation	x	y	F	z	Coord3a	Coord3b	Repeat	w	Sequence (right)
consensus	45	46	CXX:7:1213:16551:80800/1			1M											AGATCACGAATAGAAGGGAAAAAGTCAGAAGGCAAAAAAGAGAG		08208	08308	CXX:7:1213:16551:80800/2			1M		89427	11129	[Source:HGNC Symbol;Acc:13114];p11.2					08342	08492			CTCCCAAAGAGCTAGGATTACAGGCGGGAGCCACCCTGCTGG
HBV_321S	1992	2046	HWI-ST977:195:C0N43A CXX:7:2208:13062:58237/2	0	-	54M	2	1	0	0	0	0	0	0	0	1	GGACAATGTTCCGGAGACTCTAAGGCCTCCCGATACAGAGCAGAAGCGGTGTCG	16	3040824	3040835	HWI-ST977:195:C0N43A CXX:7:2208:13062:58237/1	37	+	101M	16	303819427	304081429	ZNF48;protein_coding;KNOWN;zinc finger protein 48 [Source:HGNC Symbol;Acc:13114];p11.2	0	0	F	325	3040883092	3040884092	MIRb;339;+;SINE;MIR	0	CTGGTTTCTAACTCCTGGGCTTAAGTGATCCACCCACCTCGGCCTCCCAAAGTGCTAGGATTACAGGTGTGAGCCACCCTGCTGGCCTGGCTTTGCTACTT
Cs_CON_consensus	1982	2083	HWI-ST977:195:C0N43A CXX:7:1316:16655:37260/1	37	-	101M	1	1	0	0	0	0	0	0	0	1	ACACAGAATAGCTTGCCTTAGTGCTGTATGGTGAGGTGAACAATGTTCCGGAGACTCTAAGGCCTCCCGATACAGAGCAGAAGCGGTGTCGAGGAGATCAC	16	3040820	3040835	HWI-ST977:195:C0N43A CXX:7:1316:16655:37260/2	37	+	101M	16	303819427	304081429	ZNF48;protein_coding;KNOWN;zinc finger protein 48 [Source:HGNC Symbol;Acc:13114];p11.2	0	0	F	299	3040883092	3040884092	MIRb;339;+;SINE;MIR	0	GATCCACCCACCTCGGCCTCCCAAAGTGCTAGGATTACAGGTGTGAGCCGCCTGCTTTGCTACTTTCTAACCTTTATAACCTCCAGCAAA
Cs_CON_consensus	1967	2068	HWI-ST977:195:C0N43A CXX:7:1113:16483:35469/1	37	-	101M	3	1	0	0	0	0	0	0	0	1	CCTTAGTGCTGTATGGTGAGGTGAACAATGTTCCGGAGACTCTAAGGCCTCCCGATACAGAGCAGAAGCGGTGTCGAGGAGATCACGAATAGAAGGGAAAA	16	3040824	3040836	HWI-ST977:195:C0N43A CXX:7:1113:16483:35469/2	37	+	101M	16	303819427	304081429	ZNF48;protein_coding;KNOWN;zinc finger protein 48 [Source:HGNC Symbol;Acc:13114];p11.2	0	4	F	288	3040883092	3040884092	MIRb;339;+;SINE;MIR	0	CTCGGCCTCCCAAAGTGCTAGGATTACAGGTGTGAGCCACCCTGCTGGCTGGCTTTGCTACTTTCTAACCTTTATAACCTCCAGCAAAAGGACTAATCAT
HK2100	2441	3442	HWI-ST977:195:C0N43A CXX:7:1114:7857:22466/1	0	-	101M	5	0	1	0	0	1	0	1	0	0	GTTTGTGAGTGATTGGAGGTTGGGGACTGCAAATTTTGGCCAGGACACGTGGGTGCTCCCCTAGAAAATTGAGAGAAGTCCACCACGAGTCTAGACTAGC	16	5516330	5516307	HWI-ST977:195:C0N43A CXX:7:1114:7857:22466/2	37	+	101M	16	516789647	516806043	NA;pseudogene;KNOWN;NA;q12.1	4929197	0	F	4219898	5516330	5516305	L2c;217;+;LINE;L2	58	CCTGCTTTCTTTCCTATGTGTTTGATTTGAACTTCACTTCCAGGGAAAAAGAGTCTGATTGGCCTAGGTTGTGTCATGTGACTTCCCAGGGTCATGTGAT
Cs_CON_consensus	2775	3376	HWI-ST977:195:C0N43A CXX:7:1213:17368:83971/1	0	-	101M	3	0	1	0	0	1	0	1	0	0	CCAGCGATAACCAGGACAAGTTGGAGGACAAGAGGTTTGTGAGTGATTGGAGGTTGGGGACTGCAAATTTGGCCAGGACACGTGGGTGCTCCCCTAGAA	16	5516330	5516348	HWI-ST977:195:C0N43A CXX:7:1213:17368:83971/2	37	+	101M	16	516789647	516806043	NA;pseudogene;KNOWN;NA;q12.1	4929197	1	F	4219898	5516330	5516305	L2c;217;+;LINE;L2	0	GTGACTTCCCAGGGTCATGTGATGGGCATCCCTTCTAGGATCACACAGTGAGGCAGGTCAGACCCCAGCAACATATGTCCACTGCCCCTGCTTAACACTCA
Cs_CON_consensus	2775	3376	HWI-ST977:195:C0N43A CXX:7:2306:11724:40968/1	0	-	101M	3	0	1	0	0	1	0	1	0	0	CCAGCGATAACCAGGACAAGTTGGAGGACAAGAGGTTTGTGAGTGATTGGAGGTTGGGGACTGCAAATTTGGCCAGGACACGTGGGTGCTCCCCTAGAA	16	5516330	5516348	HWI-ST977:195:C0N43A CXX:7:2306:11724:40968/2	37	+	101M	16	516789647	516806043	NA;pseudogene;KNOWN;NA;q12.1	4929197	0	F	4219898	5516330	5516305	L2c;217;+;LINE;L2	0	GTGACTTCCCAGGGTCATGTGATGGGCATCCCTTCTAGGATCACACAGTGAGGCAGGTCAGACCCCAGCAACATATGTCCACTGCCCCTGCTTAACACTCA
HBV_321S	2775	3376	HWI-ST977:195:C0N43A CXX:7:2204:2228:31732/2	0	-	101M	3	0	1	0	0	1	0	1	0	0	CCAGCGATAACCAGGACAAGTTGGAGGACAAGAGGTTTGTGAGTGATTGGAGGTTGGGGACTGCAAATTTGGCCAGGACACGTGGGTGCTCCCCTAGAA	16	5516330	5516348	HWI-ST977:195:C0N43A CXX:7:2204:2228:31732/1	37	+	101M	16	516789647	516806043	NA;pseudogene;KNOWN;NA;q12.1	4929197	0	F	4219898	5516330	5516305	L2c;217;+;LINE;L2	0	GTGACTTCCCAGGGTCATGTGATGGGCATCCCTTCTAGGATCACACAGTGAGGCAGGTCAGACCCCAGCAACATATGTCCACTGCCCCTGCTTAACACTCA