

Supplementary File 1

Class 1 breakpoints

AC198097 TAGCGGGCAGAGCCAAGCAGGGCTAAACTCCAACCCAGGGCCTGACGACAGGCACATCCAGCTAATGTTTGTATTTTAGTAG
HSA Chr7 TAGCGGGCAGAGCCAAGCAGGGCTAAACTCCAACCCAGGGCCTGAGTGGACAAATTTCTTTGCTCGTTGCTGTGTATTA-
HSA Chr2 AACAGAGCGATTCT---CCTGCCTCAGCCTCCTGAGTAGCTGGGACGACAGGCACATGCCCGGTTAATGTTTATTTTAGTAG
1*11*1**1111*1111*11*1**1*1*11*111111*111 *22*222*2**2**122* 2 2222***1*2*2*2**2*

AC198100 CAGAGGAGTGAAGAATTATAGCTCCTAACAGGGGAACAATAGGTAGACTTCTAAGGTTTATGACTCTAGTCCCTG
HSA Chr9 CAGAGGAGTGAAGAATTATAGCTCCTAACAGGGGAACAATGA-----AACATTTGTATTACAA-----G
HSA Chr9 TATAGTTGAGTACCAACTCAGCTGC-AATAAAGGAATGCCAGGTAGACTTCTAAGGTTTATGACTCTAGTCTCTG
1*1**11*1*1*11* 1 1****1*1*1*11121*1111122222222222**22***2*2*22*222 22*

AC198101 CAAATAGGCCCTACTTTTCAAAAATACATTTATCAATGTTGTATACACTTCTACTAAAAATACAAATATTAGCCGGGTGTGCTG
HSA Chr8 CAAATAGGCCCTACTTTTCAAAAATACATTTATCAATGTTGTATACAC-----AATGTAATACTTTGCTGG-GTAGTG
HSA Chr8 TGAAG----TCAGGAGTTCGAGACCAGCCTGACCAACATGGTGAACATCTCTACTAAAAATACAAATATTAGCCGGGCATGCTG
11*1111111* 1111***1*1*11*111*1*1***11*1*11122222222222**22***2*222*2** 1*22**

AC198102 CAGTGAGTAATCTTGGCACCTTTGTCAAAAATCAGTTGGCTGTGGAGTAGTGACAAGATGATGTGAGGCTTGGTGATGCTTGT
HSA Chr16 CAGTGAGTAGTCTTGGCACCTTTGTCAAAAATCAGTTGGCTATAGATACATGG-----ATTAATTTTGCATTCCCTGT
HSA Chr5 -----AGAAGT--TGGGATAGTAGTTA-----CTCTTTGTGGAGTAGTGACAACATGATGTGAGGCTTGGTGATGCTTGT
11111**121*11***1*111*1*1*11111111*111*2*2**2222*2222 222222*2*222*2*222*2*2**

AC198146 TGATAGCAAATAAGTCTCACAAAGACCTGATGGGTTTATCAGGTTAAAACTAAAGCACAAACAGACACATTAGCCTAGGCCTA
HSA Chr3 TGATAGCAAATAAGTCTCACAAAGACCTTATGGGTTTATCAGG-----GATTTCCTACTTTGCATCCTCCTCATTTTCTCTTG
HSA Chr8 TTTTAAAACTTTTTTGTAAAACTAAAGCACAAATTTTTATTTTAGTTTTT---ACCAGACACATTAGCCTAGGCCTA
111*111*11*1**1*111111**1* 1*11111*1*111*222222*2*2222*222222**22*2*2*222*222*2*2

AC198151 AATAATCAAAAACCATGAATGACAGAAGAGACCAGACCAGTTTCTCAAATATGCTCCACAAGTAATGTGATTTCTGC
HSA Chr10 AATAATCAAAAACCATGAATGACAGAAGAGACCAGACCAG-----AAAATAT-----ATGAACAGAAACCATCTCC
HSA Chr10 CAGGA--GGCCACTGT-AGTGTGGATATCTCTGGATCAGTTTCTCAAATCATTTCCACAGGTAATGTCTATTCTCC
1*11*111111**11*1*1*111**11111111**1**222*22***21** 2222*22122*2222 2**2** *

AC198152 ACAACTTCGGGAGCCATCACTTATTTTTGTTGTGTTCCAGAAGGTTGCCCAGCAGCCCGCTGCGGGCTGGGGGTCCGGGA
HSA Chr1 ACAACTTCGGGAGCCATCACTTATTTTTGTTGTGTTCCAGAAGGTTCTCATAGACTGCAGTGTGAATAGTGTCTCCATTA
HSA Chr19 GTGACCTTGAGCCTCGTCTCCGCCAGGCCGGCTCGCGGAAGGGGCCCTAGCAGCCCGCTGCGGGCTGGGGG-TCCGGGA
111*111*1*111*1*1*1*11111111*11*11*11*1****1*2*2**2222*22*22*2*2222*2*22 **222*

Topo IIv site (84% similarity)

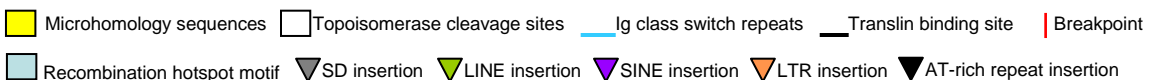
Immunoglobulin class switch repeats

AC198875 AACACAATATTTGTTTAAAGGCTTATACTACTGTTTCAAGTAATCTCCTGCCTCAGCCTCTCGAAGTGCCGAG
HSA Chr12 AACACAATATTTGTTTAAAGGCTTGCCAAAAACA--CAAACAAAGCCTTAAACCCCAAACCTTTAAAGTATAAAA
HSA Chr19 TCTCACTTTGTT-GCCAGGGCTCTTGAACCTCTGGTTTCAAGTGATTCTCCGTCCTCAGCCCTCGAAGTGCTGAG
1111**11*1**1*111*11**1*2*2 *221222222222**221*22**2222*2**22* 2*22***22 2*2

Translin binding site

Topo IIv site (88% similarity)

AC198944 ATGGTTACCAATATTTTCCCTACAGTGTCTAAATAGACTTGTAGCTTTCTTGTTCAACCTGTTTATCTGTTAAG
HSA Chr9 ATGGTTACCAATATTTTCCCTACAGTGTCTAA-TAIGTGTTAATTGC---ATTTAGAGTATTTTTCATCTCAG
HSA Chr6 AAAATAAAATAAAATAAAATTAAGTACGCGAATAGACTTGTAGCATTCTTGCTCAACCTGTTTATCTGTTAAG
*111*11*11*1**11*1111*11***11111*2*2 2222*2**2212222221*2*222*2***2*222*2**



Class 2 breakpoints (breakpoint intervals between 9 bp – 669 bp)

```

AC198144 CCTTTTCTCTGTGGCCCTTTTCTTCAATCCCATCTACTATTAGCTTCCCTCATCGTTCAAAGGTGGAGAATTGAATGATGAAA
HSA Chr5 CTTTTTCTCTGTGGCCCTTTTCTTCAATCCCATCTACTCATGGTCTTCCCTCTC---AAGGTCTCACAACCTCCCTACC--AG
HSA Chr16 CTTTGATTAGCGGCCAAATTAAGGATTAGGTCCTATGCTAAGATGGTCAGGGC-CTCAAAGGTGGAGAATTGAATGATGAAA
* **111**1*1****111**1111*1*1111****11 212* 211* 1* 222**2*2222*2**22*222*22222*2
    
```

Topo IIv site (84% similarity)

```

AC198149 TCGCCTGAGGTTAGGAGTTCGAGACCAGCTCGAACTCACTAGGATGTTGGGAGAAAGCAGCGAGAGAGAAGGGGCCAGGGCA
HSA Chr2 TCGCCTGAGGTCAGGAGTTCGAGACCAGCTGGCCAATATGGTGAACCCTGCTCTACAAAAATACAAAAATAGCCA
HSA Chr17 TGGCCCCAGGCCTAAACATGAGGCCATAAATCATGCAGAGAAAACAGGTGAGGAGCAAGCAGCGAGAGAGAAGGGGCCAGG-TA
*1**11**1 111*11*11**1*1*11 1 1 1 21 2222222*222*22*2*2*2*2*22222222* 1*
    
```

Translin binding site

Topo IIv site (84% similarity)

```

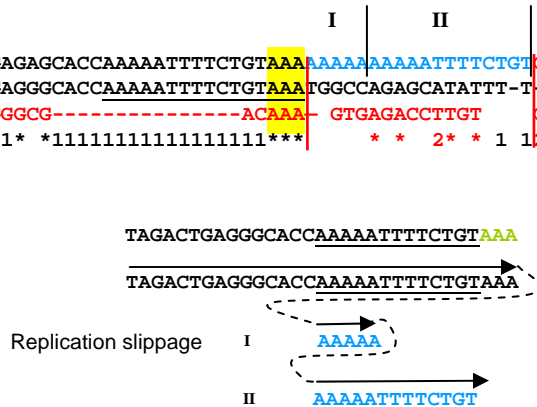
AC198526 ATCAGACTACTTAGGCTCAAATTCCACAACACCACTTAAAGGCTGTGTGACCTTGTACAGGTTTTTTAACCT
HSA Chr3 ATCAGACTACTTGGGCTCAAATTAC----TTCTATGAAAGTGATATCTCATTATCTCCATTTTATAGATG
HSA Chr3 GCCCACCACCACCTGAATTCGAGGCCCACTTAAAGGCTGTGTGACCTTGTCAAGTTTCTTAACT
11*11**1**11 11*111****212 2 1*22*2****22*2*2*22*22222 2***12**2222
    
```

Translin binding site

```

AC198147 TAGACTGAGAGCACCAAAAATTTCTGTAAA AAAAA AAAAAATTTCTGTCTAAAAA AAAAAACTCAGGACATCTGGAC
HSA Chr1 TAGACTGAGGGCACCAAAAATTTCTGTAAA TGGCCAGAGCATATTT-T-----AGGCTTGCAGGCCATATGGTT
HSA Chr1 TAGCCTGGGCG-----ACAAA- GTGAGACCTTGT CTAAAAA AAAAAACTCAGGACATCTGGAC
***1***1* *11111111111111111*** ** * 2* * 1 1 2222222222*2222222222****2***2***22
    
```

Topo II site on reverse strand (88% similarity)



 Microhomology sequences	 Topoisomerase cleavage sites	 Ig class switch repeats	 Translin binding site	 Breakpoint	
 Recombination hotspot motif	 SD insertion	 LINE insertion	 SINE insertion	 LTR insertion	 AT-rich repeat insertion

Class 2 breakpoints (breakpoint intervals between 9 bp – 669 bp)

154 bp
AT-rich repeat insertion

AC198099 TTATTAATGAGATTTTTATCTAAGGAGGGTTTCTACCTCTCGATGGAGATAGTATATATATATATATATATCTTGAGCAAAGACA
 HSA Chr19 TTATTAATGAGATTTTTATCTAAGGAGGGTTTCTGCCTCTTGATGGAGATACTTAACATGACTAAACATGGATTGGGCAAAGCG
 HSA Chr12 T-GTATGTGAAACTGTTACATAAAATAAGAGTTAGAA-----AAGAGAGAATATATATATATATATATATATCTTGAGCAAAGACA
 *11*111**1*1*1***11***11*1*11**11211111 1*11****122*22*2**222*2*2**222*2*****22*2
 Breakpoint

84 bp
AT-rich repeat insertion

AC198945 TAATTGAATGCCCTTTATTTCTTCTCCTGCCTGATTGCCAGCCTTCTTGCCTTACTGGACAGCACTTCAGCAGTGCAGT
 HSA Chr10 TAATTGAATACCTTTATTTCTTCTCCTGCCTAATTACCAG--TTTTGGCCATTGATGACTGCTGTGGTTTTGT
 HSA Chr4 CTAT---ACCTTATTTAATATAATTTGGATTACTAACACAGCCTTCTTGCCTTACTGGACAGCACTTCAGCAGTGCAGT
 11*111*1 111***1*11 *11*111*11* *1* 1***22**2***2*2*2*2*2222***2*2*22*2222**
 Breakpoint

669 bp
LTR insertion

Retroviral integration target motif

AC198096 AGAGCATCCTGAAAAAATTAATATACATTTCTGCTTTGGGTTTTAAGTGGCTCTCTTCAATTCCTTTTTTGGTT
 HSA Chr12 AGAGCATCCTGAAAAAATTAATATGCATTTCTAACAAATCTAGATTATACAAAAATATGCAATTTTTTCAGTTTTG
 HSA Chr19 GGTGTAGTCTCCCAAGCTTTATTCGAAGCTTTT-CCTTTGGGTTTTAAGTGGCTCTCTTCAATTCCTTTTTTGGCT
 1*1*1*11**1111**11**1*1*112111**1*1*2222222*222*22222222222* **2*22**222*2212
 Breakpoint

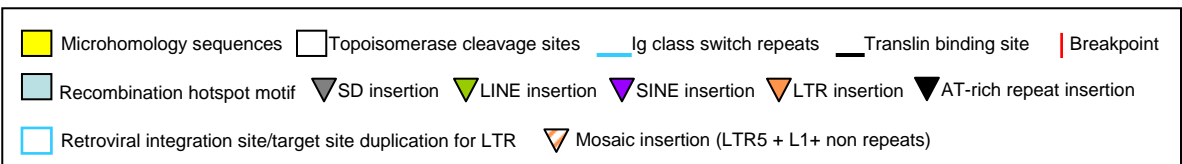
313 bp
LTR insertion

Retroviral integration site/TSD

AC198183 TCACCCAGCCAACACTGTAACCTCGCTTCTTAGACTGTTGTCTTAAGAGAACACAAAAATAAGCTGGGCATGGTGGCGTGCAC
 HSA Chr4 TCACCCAGCCAACACTGTAACCTCCTTCTTAGACTGTTGACTTAAAG-----CA-AAAAATTAGCCGGGCGTAGTGGCGGGCGC
 HSA Chr22 TCAAAATCGGCCAAACTTAGGAGA-----TCAGATGATT---TGAAAGAAAAACAAAAATAAGCTGGGCATGGTGGCGTGCAC
 111*1**111*111*111 1111*1***111*1* 1*1****22 22**2*****2***2***2*2*****2*2*
 Breakpoint

210 bp
AluY insertion

AC198150 TAGCAATCTAATTGAATGGAAAAGGAACCTTCAGTAACAGAGAACCTTCTAAAACCTAGAGA
 HSA Chr14 CAGCAATCTAATTCAATGGAAAAGGAACCTCTTTCATGAACCACTTCTTTAAAAAGCGAAG
 HSA Chr14 TAACTAAAAAAAAGAATGTAGAGGGAGATAGTAACAGAGAACCTTCTAAAACCTAGAGA
 1*1111111**11 *111111*11*111111*222*22*222*22*2*2*****222*22
 Breakpoint



Class 2 breakpoints (breakpoint intervals >1 kb – 20 kb)

