a

Retroviral integration target motif

ACATTTCTG
GCATTTCT-
AAGCTTTT-
2111**1*

TGGG → TGTT   TACT → CAGCA   CC → ATTA   AAT → TTC   ATTCA → ATG   GAAAATT → TAGTTT   CCTTTGGGT
AACAAATCT
CCTTTGGGT
22222222*

1   194   299   324   507

LTR5_Hs   LTR5   LTR5B   L1MEc   669

kb
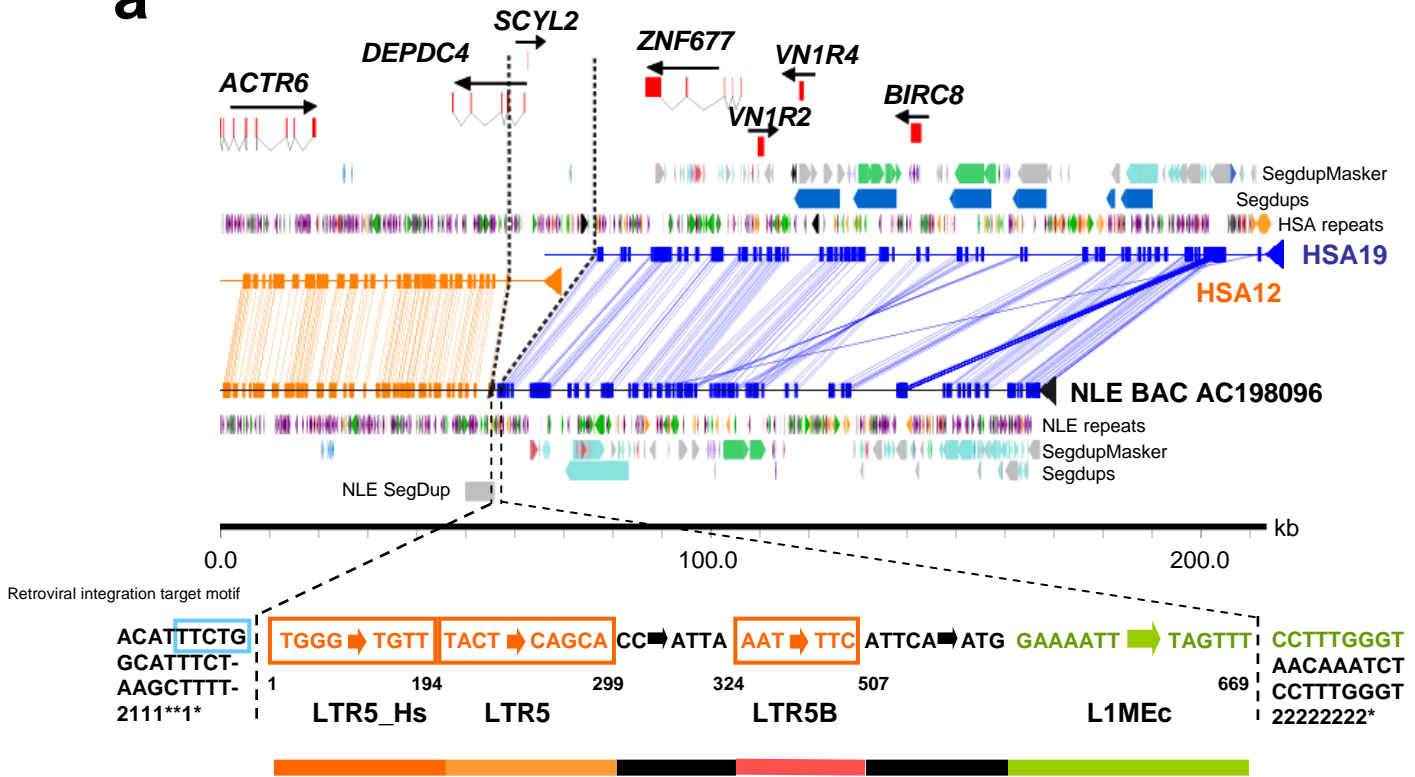
0.0   100.0   200.0
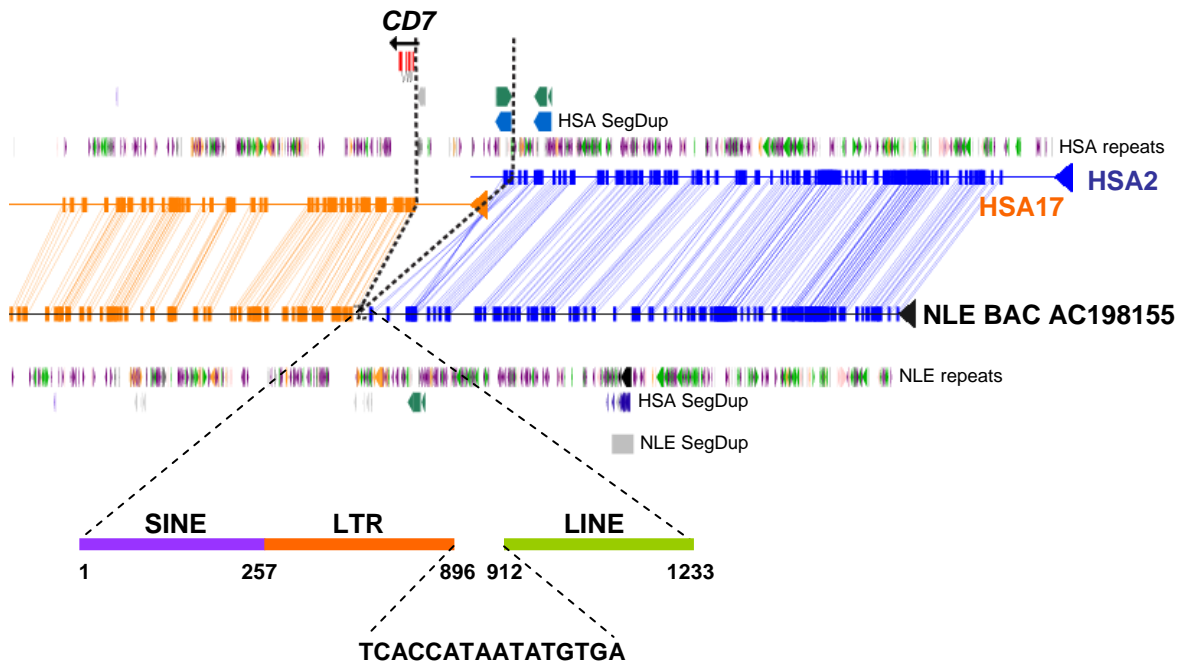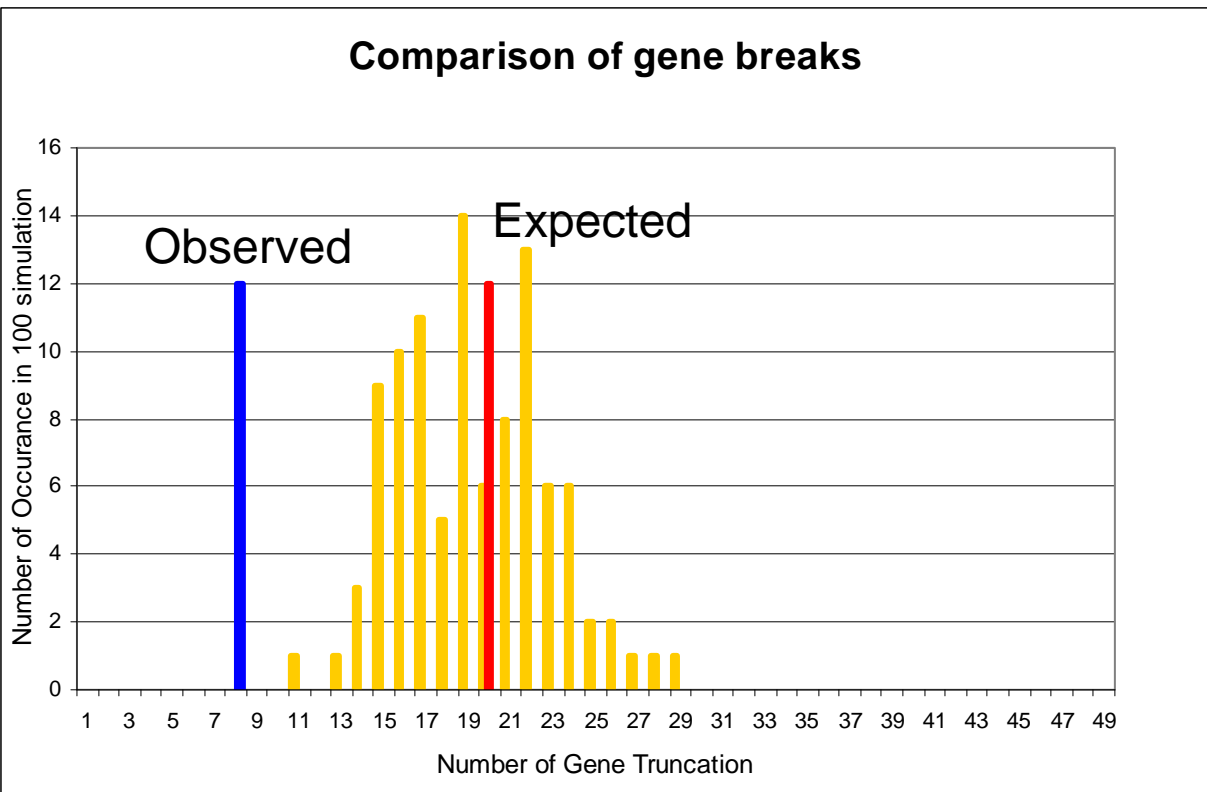
b

Retroviral integration target motif   HERVL-A2/LTR insertion ↓

NLE AC198183   CAGTCACCCCAGCCAACACTGTAACTCGCTTCTTAGACTGTTGTCTTA AAG GTAGTGGAC
HSA Chr4   CAATCACCCCAGCCAACACTGTAACTCCCTTCTTAGACTGTTGACTTAAGTAGTAGAC
HSA Chr22   CCATCAAATCGGCCAAACTTAGGAGA------TCAGATGATT---TGAAAGGTAAGTGAA
               *1 ***111*1*****111*111*111 1111*1***111**1 1*1***2***11 **1

NLE AC198183   TGATTTCATCTTGATAGTCCAGGTCAGTCACCCCAGCCAACACTGTAACTCCCTTCTTAG
HSA Chr4   TGGGCCGGGCGCGGTGGCTCACACCTGTAATCCCAGC--------------------
HSA Chr22   CACCCGCGAAGGGG-ACTCACCACCATGCTCTGTGGACGATCCGGCAACCTCTCTCTGTG
               11 2 1 * 1212211   *2112121111*12 2 2 2 222 2 222 2

NLE AC198183   ACTGTTGACTTAAAGGTAGGAGGAGCCCCAAGTGTCCAGGTGACAATCTTAACTTCCTGT
HSA Chr4   -------ACTT-----TGGGAGG----CCGAG--------------------------
HSA Chr22   --TACTTCCCAAAGAAGCCCGGGAGCT---GGTCTTCAGG-GGCAACCC--GTCACCTTC
               2 2 1*1122   1 111**222 11 1*2 2 2222 2 222 2   2   2 2

NLE AC198183   TTAATGGAATCGCTGTTGTGTCTCCTGGTGGCAGCATTCCTCCCTTTGAAACTAAGACCT
HSA Chr4   --
HSA Chr22   CTAG-GGAACTGCATTAGCATGAGCTG-----------CTTCTTTTTAAGA----AACGT
               22   2222   22 2 2 2   222   2 22 222 2 2   22 2

NLE AC198183   TTAAGCCAGCAGAACGTAATGTCATGGGAACGGGAAGCAAAAATTTTGCTAGTGACTCTC
HSA Chr4   ----GCGGGCAGATCACGAGGTCAGGAGATCGAGA--------CCATCCCAG-------C
HSA Chr22   TTAAG---GCCGGGCACGGTGGCTCACGCCTATAATCCCAGCATTTTG--GGAGGCCGAG
               2222*1   **1*1 *   12*1*1 1 *1 11 1* 2 2 2222*21 1* 2 2   1

NLE AC198183   TAAGAGTGATGATGAGTGGTGCCACCTCCACCCCTTGATTCCTGGACCTGTGAATCCTGG
HSA Chr4   TAAAAACGGTGA---------------------------------------AACCCCGT
HSA Chr22   GCAGGCGGATCACCAG--GTCAGGAGTTCAAAACCAGCCTGACCAACATGTGAAACCCCA
               11*21 *2*1* 22 22   2 22   2 2 2   22 2222** ** 1

Retroviral integration target motif

NLE AC198183   CTATGAGAGAAATAGT AAGA ACACAAAAAATAAGCTGGGCATGGTGGCGTGCACCTGTAAT
HSA Chr4   CTCTACTAAAAATA-------CA-AAAAATTAGCCGGGCGTAGTGGCGGGCGCCTGTAGT
HSA Chr22   TCTC-------TACTAAAAACACAAAAAATAAGCTGGGCATGGTGGCGTGCACCTGTAAT
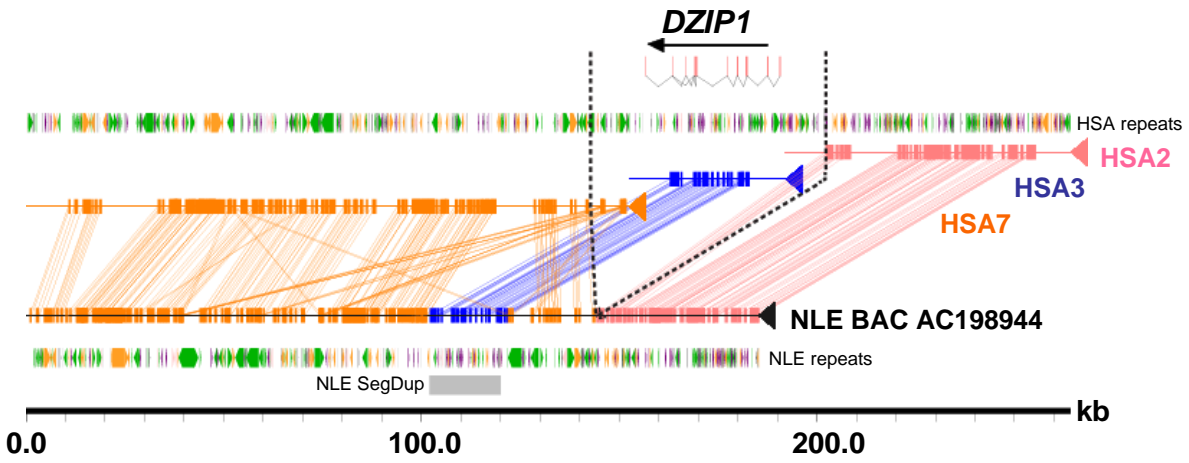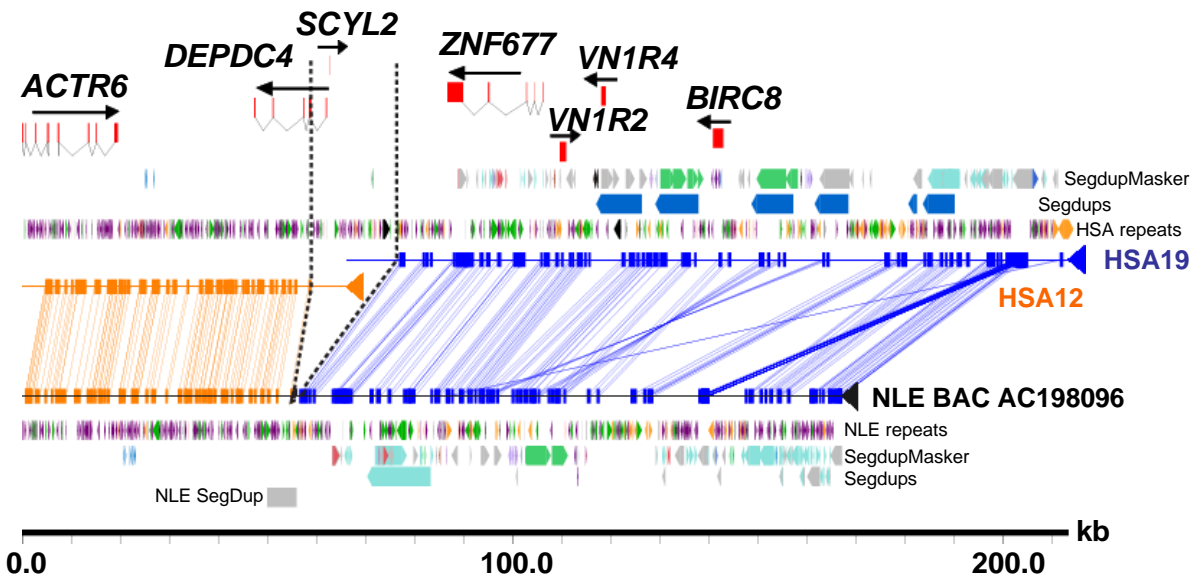   11 1 1 111 ** 222 22**2*****2***2****2*2*****2**2******2*

CD7

HSA SegDup

HSA repeats

HSA2

HSA17

NLE BAC AC198155

NLE repeats

HSA SegDup

NLE SegDup

SINE	LTR	LINE
1	257	896 912	1233

TCACCATAATATGTGA
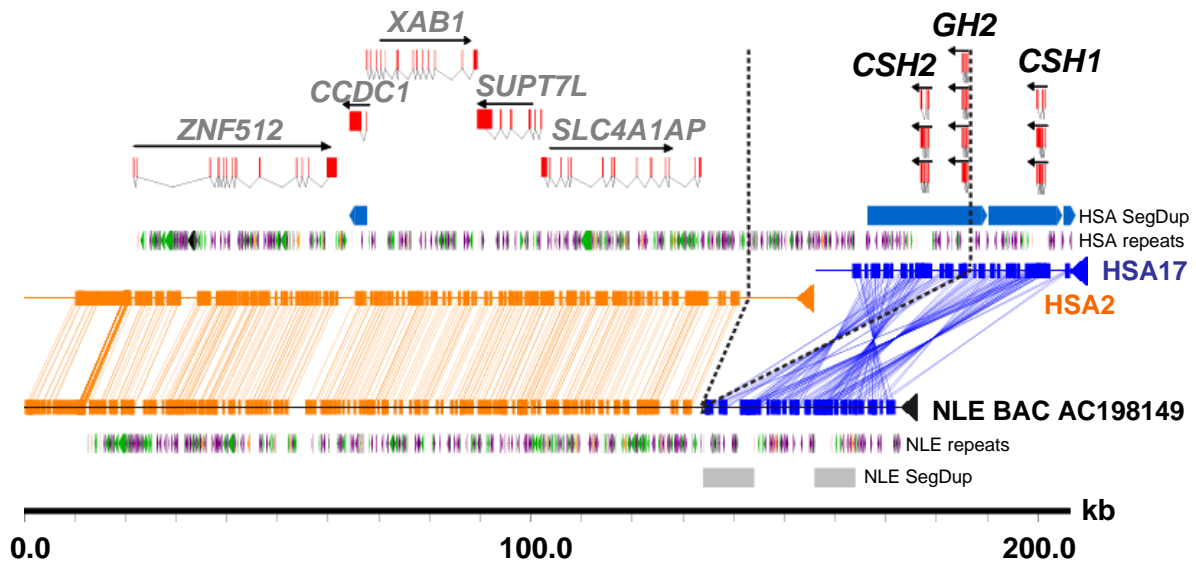
Supplementary Fig. 2
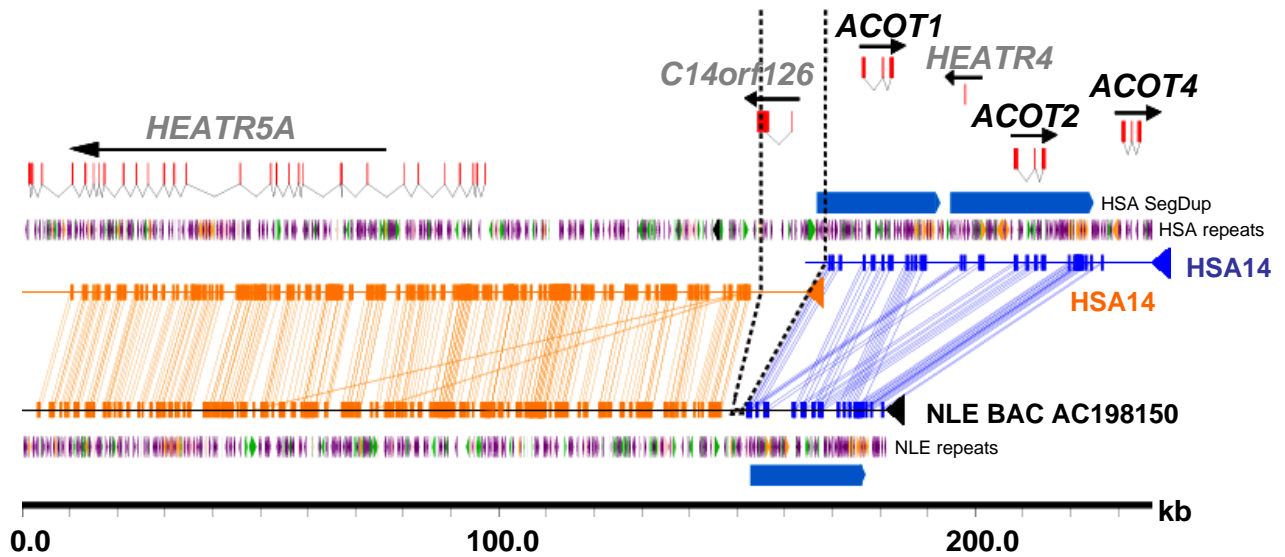
Comparison of gene breaks

Empirical p value = calculated by number of times the observed gene disruptions is seen by simulations divided by the number of simulations = 0/100, *P<0.001*
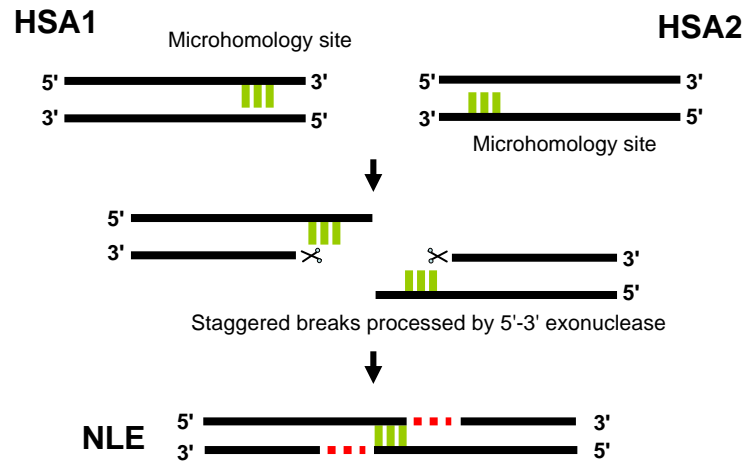
Supplementary Figure 3

a

DZIP1

HSA repeats

HSA2
HSA3
HSA7

NLE BAC AC198944

NLE repeats

NLE SegDup

kb

0.0   100.0   200.0

b

ACTR6   DEPDC4   SCYL2   ZNF677   VN1R4   BIRC8

VN1R2

SegdupMasker
Segdups
HSA repeats

HSA19
HSA12

NLE BAC AC198096

NLE repeats
SegdupMasker
Segdups

NLE SegDup

kb

0.0   100.0   200.0

Supplementary Figure 4

a

ZNF512 CCDC1 XAB1 SUPT7L SLC4A1AP CSH2 GH2 CSH1

HSA SegDup
HSA repeats
HSA17
HSA2
NLE BAC AC198149
NLE repeats
NLE SegDup

0.0   100.0   200.0 kb

b

HEATR5A C14orf126 ACOT1 HEATR4 ACOT2 ACOT4

HSA SegDup
HSA repeats
HSA14
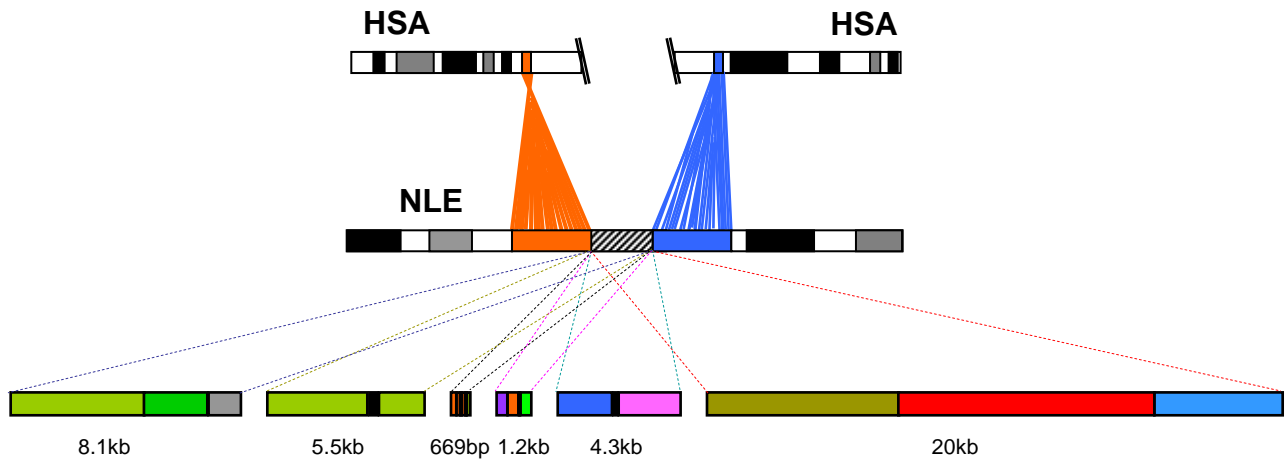HSA14
NLE BAC AC198150
NLE repeats

0.0   100.0   200.0 kb

Supplementary Fig. 5

Mechanism for microhomology-mediated end-joining

Supplementary Figure 6

Class II mosaic insertions