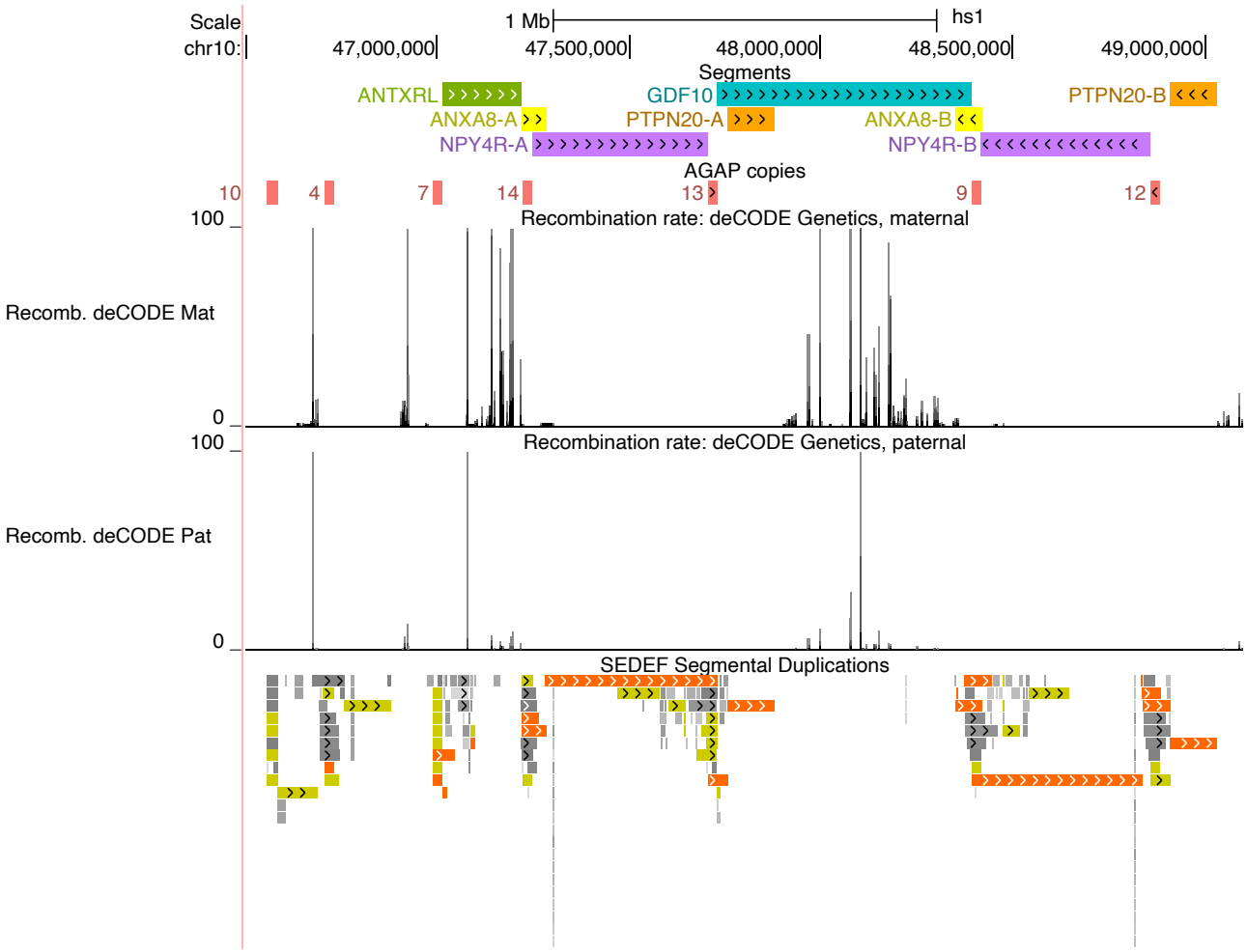
AGAP duplicons associate with structural diversity at Chromosome 10q11.22

Stefania Fornezza[1,4], Vincenza Simona Delvecchio[1,4], William T. Harvey[2], Philip C. Dishuck[2], Evan E. Eichler[2,3], Giuliana Giannuzzi[1]

1. Department of Biosciences, University of Milan, Milan, Italy
2. Department of Genome Sciences, University of Washington School of Medicine, Seattle, WA, USA
3. Howard Hughes Medical Institute, University of Washington, Seattle, WA, USA
4. These authors contributed equally to this work.

Correspondence to:
Giuliana Giannuzzi
University of Milan
Department of Biosciences
Via Giovanni Celoria, 26
20133 Milano, Italy
giuliana.giannuzzi@unimi.it

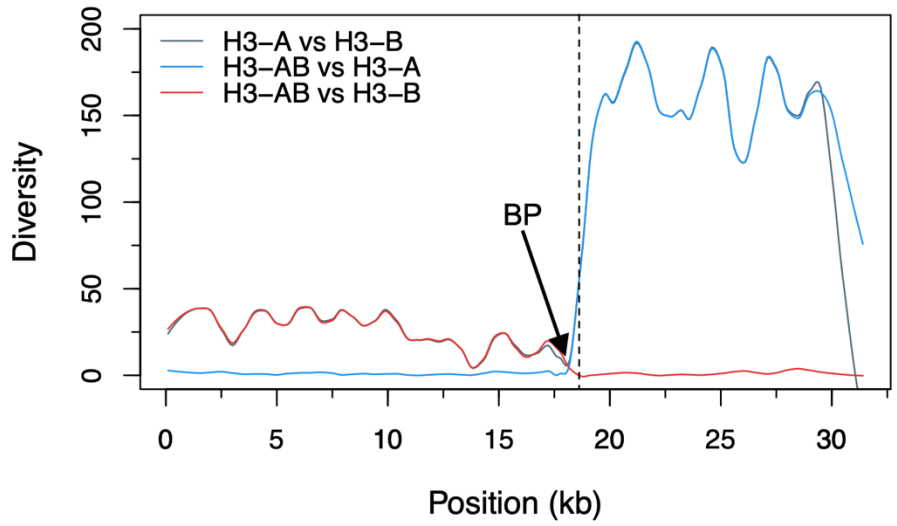
This file contains Supplemental Figures S1-S7.



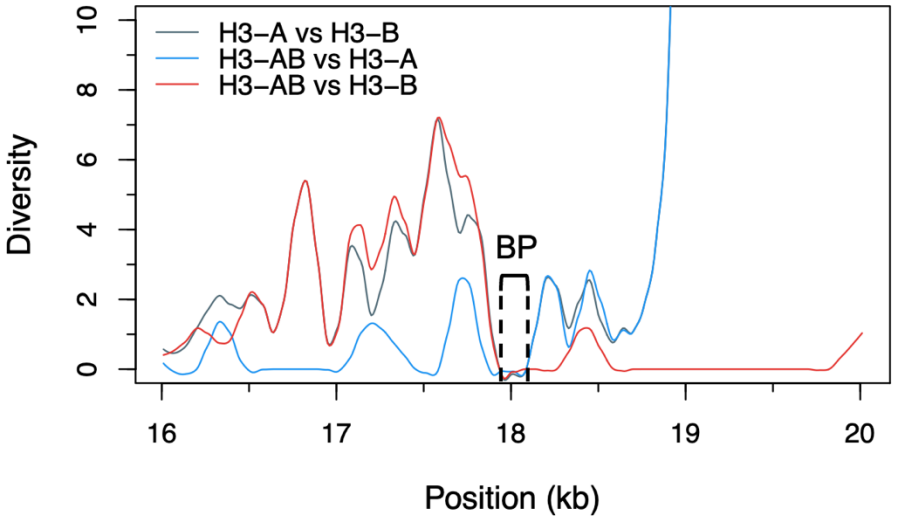
Supplemental Figure S1. Recombination hotspots at chromosome 10q11.22.

View in the UCSC Genome Browser on human T2T-CHM13 v2.0 of the region chr10:46,500,000-49,100,000. Segments track: the locations of the NPY4R-A and NPY4R-B inverted duplications (violet), ANTXRL segment (green), GDF10 segment (blue), ANXA8-A and ANXA8-B inverted duplications (yellow), and PTPN20-A and PTPN20-B inverted duplications (orange). The locations of *AGAP* copies are shown with red boxes, including both protein-coding genes and pseudogenes; maternal and paternal recombination rates from deCODE Genetics are shown; at the bottom, segmental duplication annotation is shown with colors reflecting the level of similarity: light to dark gray for 90-98%; yellow for 98-99%; orange for similarity greater than 99% (Numanagic et al. 2018; Vollger et al. 2022).

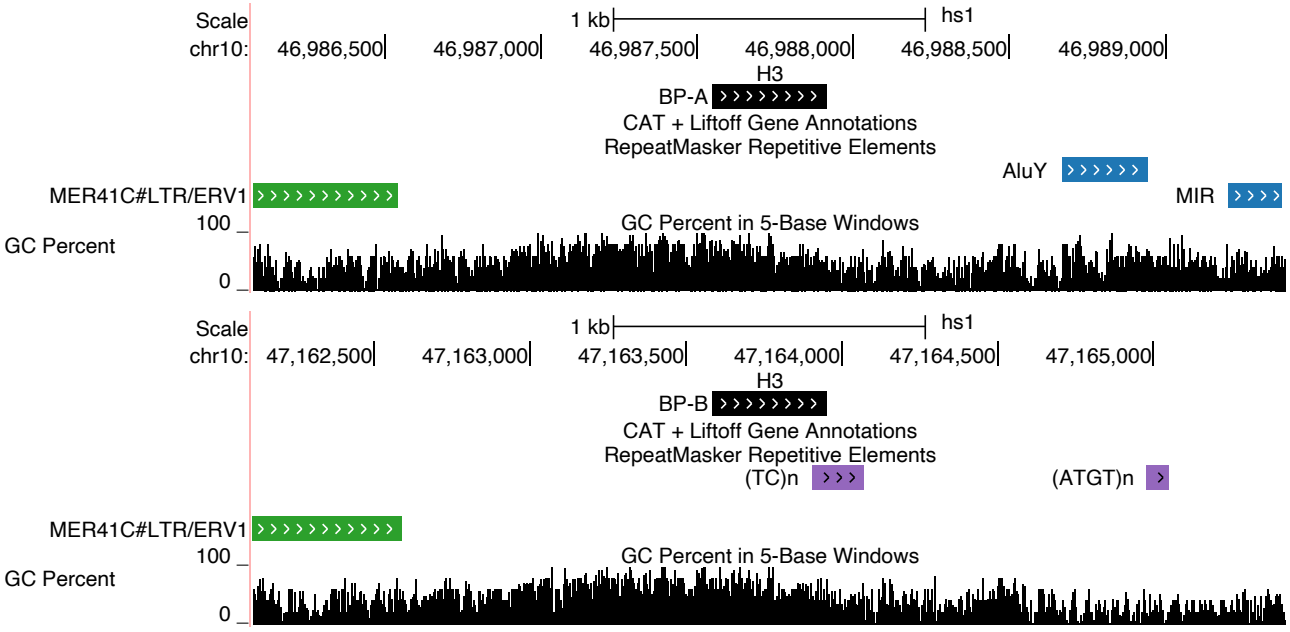
A


H3–A vs H3–B
H3–AB vs H3–A
H3–AB vs H3–B
Diversity
BP
Position (kb)

B


H3–A vs H3–B
H3–AB vs H3–A
H3–AB vs H3–B
Diversity
BP
Position (kb)

C


Scale
chr10:
1 kb
hs1
46,986,500
46,987,000
46,987,500
46,988,000
46,988,500
46,989,000
H3
BP-A
CAT + Liftoff Gene Annotations
RepeatMasker Repetitive Elements
AluY
MER41C#LTR/ERV1
MIR
GC Percent
100
0
GC Percent in 5-Base Windows
Scale
chr10:
1 kb
hs1
47,162,500
47,163,000
47,163,500
47,164,000
47,164,500
47,165,000
H3
BP-B
CAT + Liftoff Gene Annotations
RepeatMasker Repetitive Elements
(TC)n
(ATGT)n
MER41C#LTR/ERV1
GC Percent
100
0
GC Percent in 5-Base Windows

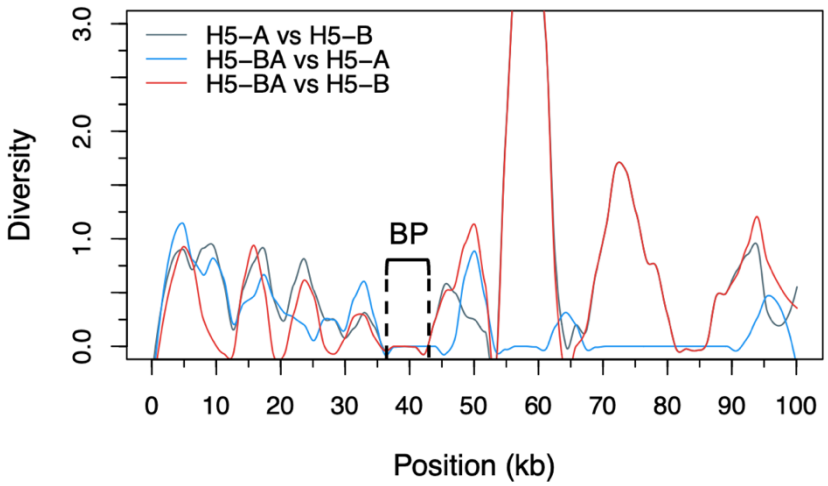
Supplemental Figure S2. H3 breakpoint analysis.

A) Diversity plots of H3 breakpoint region compared with the putative SDs mediating the deletion. Here, as the breakpoint occurs towards the end of the SD blocks, the sequence was extended downstream. A horizontal dashed line marks the end of the segmental duplication alignment.
The plot shows the pairwise diversity in 500 bp sliding windows with 100 bp increment between the H3 breakpoint regions (H3-AB) and the original "A" (blue line) and "B" (red line) SDs. The gray line refers to the comparison between the parental SDs (H3-A versus H3-B), as a reference. The arrow indicates the location of the breakpoint as the point of switch in relative diversity between the blue and red lines.

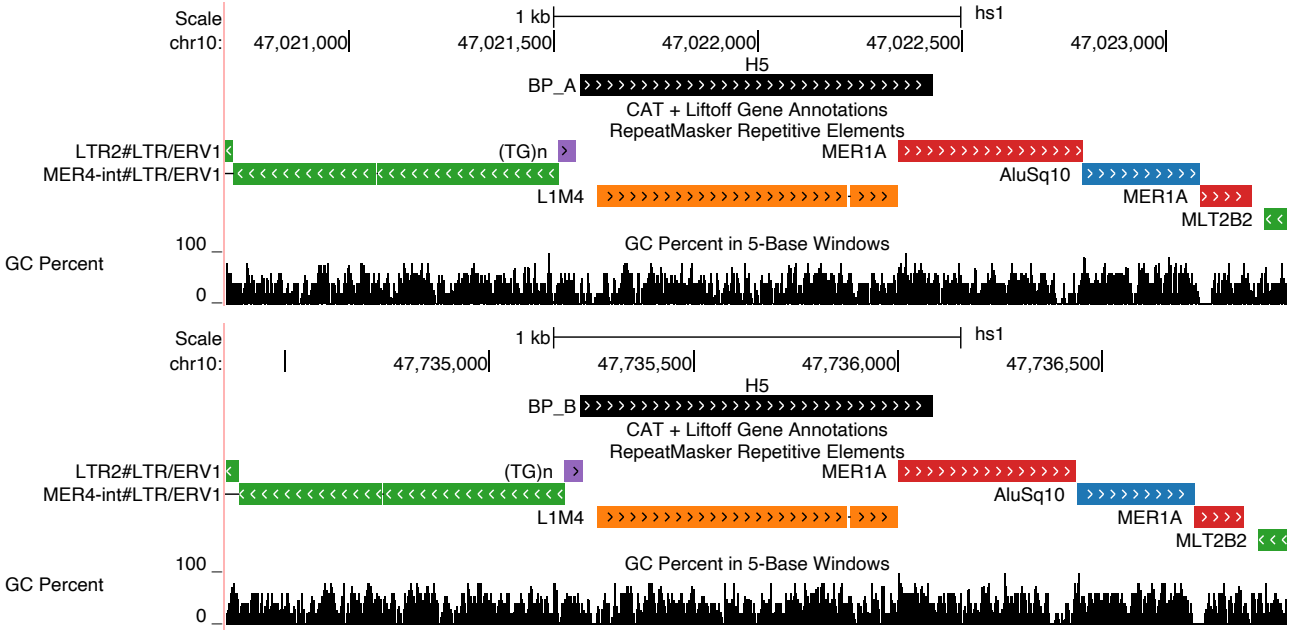
B) Diversity plot of a region zoomed around the H3 breakpoint. The plot was obtained using the PopGenome R package, with width of 100 and jump of 10. H3-AB: sequence of H3 putative breakpoint region; H3-A and H3-B: sequences of segmental duplications that triggered the H3 rearrangement. The gray line refers to the comparison of H3-A versus H3-B, as a reference; the blue line refers to the comparison of H3-AB versus H3-A; the red line refers to the comparison of H3-AB versus H3-B. The restricted breakpoint interval of 369 bp is marked by dotted lines.

C) View in UCSC Genome Browser on human T2T-CHM13 v2.0 of gene annotation, repetitive elements, and GC content at the H3 breakpoints in the proximal and distal SD copies triggering the rearrangement.

A



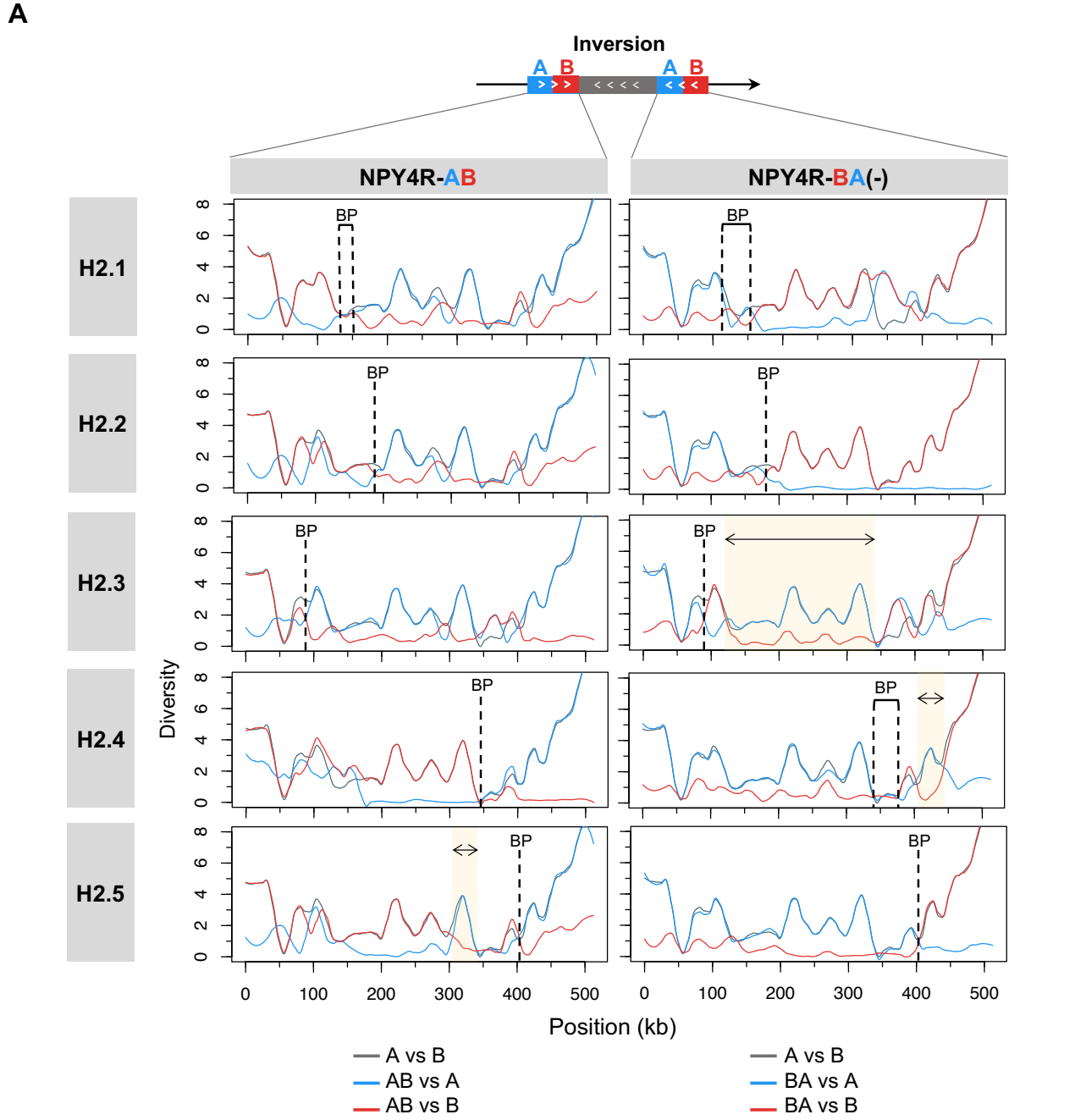
B

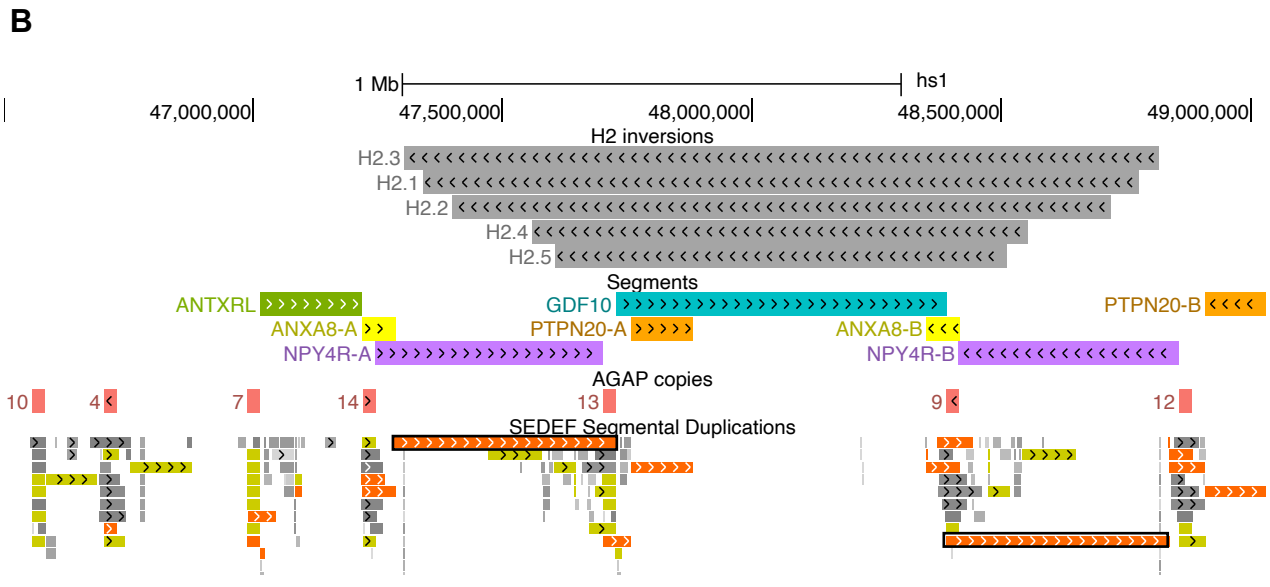


Supplemental Figure S3. H5 breakpoint analysis.

A) Diversity plot of a region zoomed around the H5 breakpoint. The plot was obtained using the PopGenome R package, with width of 100 and jump of 10. H5-AB: sequence of H5 putative breakpoint region; H5-A and H5-B: sequences of segmental duplications that triggered the H5 rearrangement. The gray line refers to the comparison of H5-A versus H5-B, as a reference; the blue line refers to the comparison of H5-AB versus H5-A; the red line refers to the comparison of H5-AB versus H5-B. The restricted breakpoint interval of 867 bp is marked by dotted lines.

B) View in UCSC Genome Browser on human T2T-CHM13 v2.0 of gene annotation, repetitive elements, and GC content at the H5 breakpoints in the proximal and distal SD copies triggering the rearrangement.


A
Inversion
A B
A B
NPY4R-AB
NPY4R-BA(-)
H2.1
H2.2
H2.3
H2.4
H2.5
BP
Diversity
0 2 4 6 8
0 100 200 300 400 500
Position (kb)
A vs B
AB vs A
AB vs B
A vs B
BA vs A
BA vs B

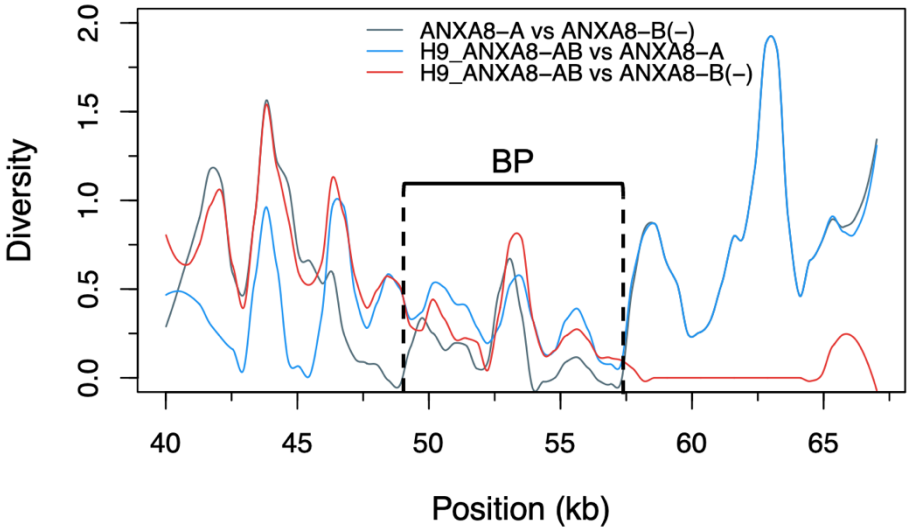

B
1 Mb
hs1
47,000,000
47,500,000
48,000,000
48,500,000
49,000,000
H2 inversions
H2.3
H2.1
H2.2
H2.4
H2.5
Segments
ANTXRL
ANXA8-A
NPY4R-A
GDF10
PTPN20-A
PTPN20-B
ANXA8-B
NPY4R-B
AGAP copies
10
4
7
14
13
9
12
SEDEF Segmental Duplications

Supplemental Figure S4. Breakpoint definition in five haploid genomes showing the H2 haplotype.

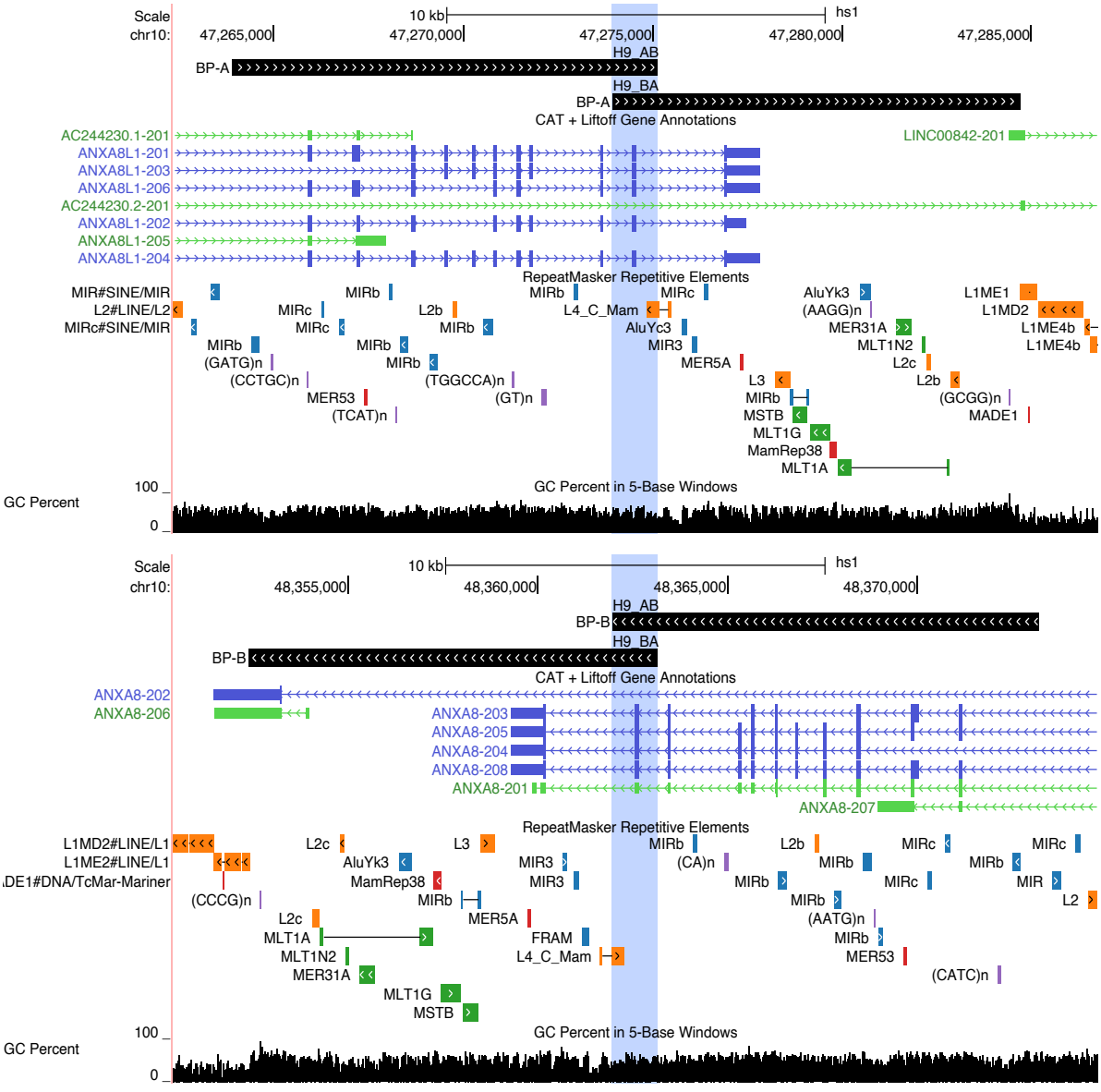
A) Diversity plots of the H2 NPY4R-AB duplicon (*left*) or H2 NPY4R-BA duplicon, reverse strand (*right*) compared with the H1 NPY4R-A (blue line) or NPY4R-B (red line) duplicons. The grey line refers to the comparison between H1 NPY4R-A and NPY4R-B, as a reference. Dotted lines designate the breakpoint region that is consistent between AB and BA plots. Segments where the similarity is unexpectedly higher to the other SD copy are highlighted in yellow.

B) View in the UCSC Genome Browser of inversion segments in H2 haplotypes. The region shown is chr10:46,500,000-49,100,000 (T2T-CHM13 v2.0). Black squares in the SEDEF Segmental Duplications track indicate the SD pair generating the inversions.

A


ANXA8–A vs ANXA8–B(–)
H9_ANXA8–AB vs ANXA8–A
H9_ANXA8–AB vs ANXA8–B(–)
BP
Diversity
0.0
0.5
1.0
1.5
2.0
40
45
50
55
60
65
Position (kb)

B

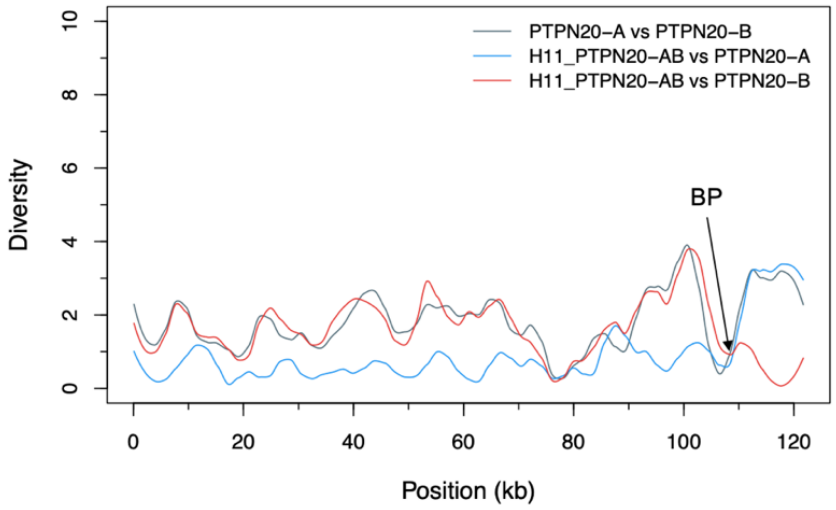

Scale
chr10:
10 kb
hs1
47,265,000
47,270,000
47,275,000
47,280,000
47,285,000
H9_AB
BP-A
H9_BA
BP-A
CAT + Liftoff Gene Annotations
AC244230.1-201
ANXA8L1-201
ANXA8L1-203
ANXA8L1-206
AC244230.2-201
ANXA8L1-202
ANXA8L1-205
ANXA8L1-204
LINC00842-201
RepeatMasker Repetitive Elements
MIR#SINE/MIR
L2#LINE/L2
MIRc#SINE/MIR
MIRb
MIRc
L2b
MIRb
MIRb
MIRb
(GATG)n
(CCTGC)n
MER53
(TCAT)n
MIRb
(TGGCCA)n
(GT)n
MIRb
MIRc
L4_C_Mam
AluYc3
MIR3
MER5A
L3
MIRb
MSTB
MLT1G
MamRep38
MLT1A
AluYk3
(AAGG)n
MER31A
MLT1N2
L1ME1
L1MD2
L1ME4b
L1ME4b
L2c
L2b
(GCGG)n
MADE1
100
GC Percent
0
GC Percent in 5-Base Windows
Scale
chr10:
10 kb
hs1
48,355,000
48,360,000
48,365,000
48,370,000
H9_AB
BP-B
H9_BA
BP-B
CAT + Liftoff Gene Annotations
ANXA8-202
ANXA8-206
ANXA8-203
ANXA8-205
ANXA8-204
ANXA8-208
ANXA8-201
ANXA8-207
RepeatMasker Repetitive Elements
L1MD2#LINE/L1
L1ME2#LINE/L1
DE1#DNA/TcMar-Mariner
(CCCG)n
L2c
MLT1A
MLT1N2
MER31A
MLT1G
MSTB
L2c
AluYk3
MamRep38
L3
MIRb
MER5A
MIR3
MIR3
FRAM
L4_C_Mam
MIRb
(CA)n
MIRb
MIRb
MIRc
MIRb
(AATG)n
MIRb
MER53
(CATC)n
MIRc
MIRc
MIRb
MIR
L2
100
GC Percent
0
GC Percent in 5-Base Windows

Supplemental Figure S5. H9 breakpoint analysis.

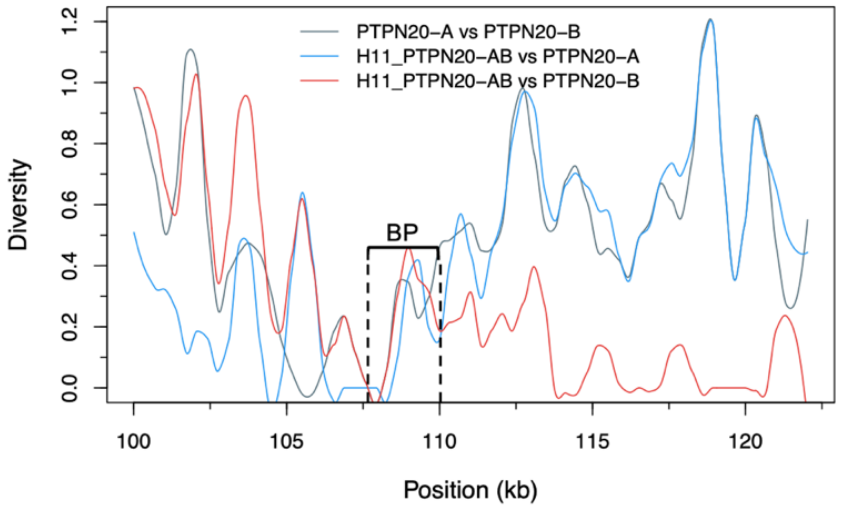
A) Diversity plot of a region zoomed around the H9 breakpoint. The plot was obtained using the PopGenome R package, with width of 100 and jump of 10. H9_ANXA8-AB: sequence of H9 breakpoint region; ANXA8-A and ANXA8-B(-): sequences of segmental duplications that triggered the H9 rearrangement. The gray line refers to the comparison of ANXA8-A versus ANXA8-B(-), as a reference; the blue line refers to the comparison of H9_ANXA8-AB versus ANXA8-A; the red line refers to the comparison of H9_ANXA8-AB versus ANXA8-B(-). The breakpoint interval of 11.5 kb is marked by dotted lines.

B) View in UCSC Genome Browser on human T2T-CHM13 v2.0 of gene annotation, repetitive elements, and GC content at the H9 breakpoints in the proximal and distal SD copies triggering the rearrangement. The final identified breakpoint region, given by the intersection between AB and BA breakpoints, is highlighted in blue.

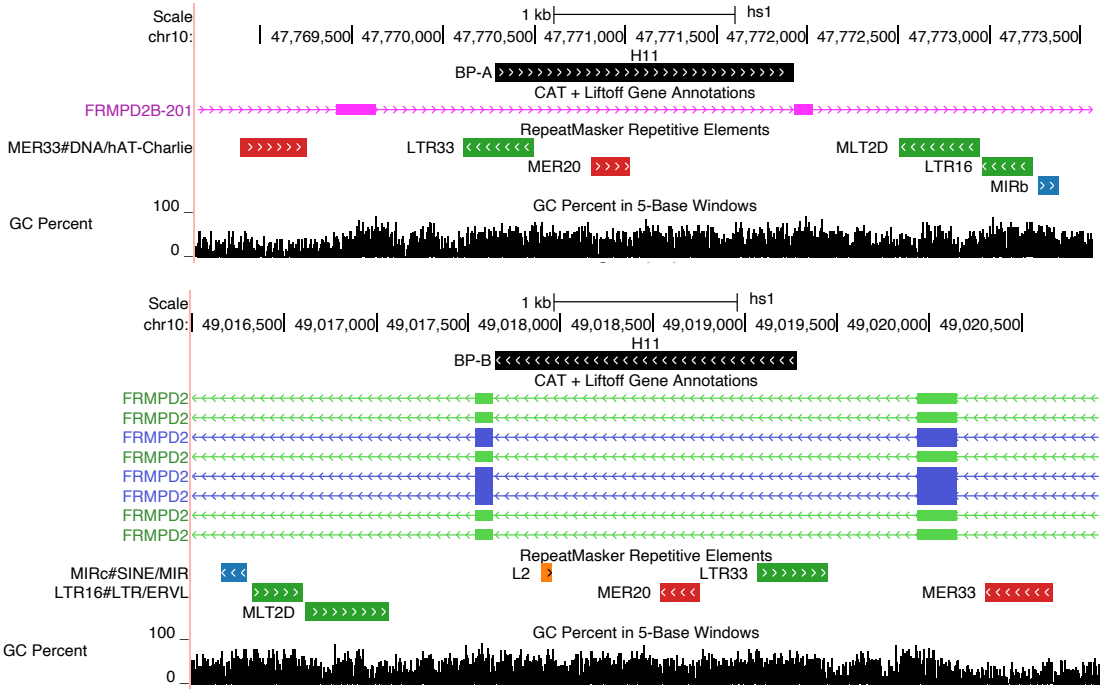
A


PTPN20–A vs PTPN20–B
H11_PTPN20–AB vs PTPN20–A
H11_PTPN20–AB vs PTPN20–B
Diversity
BP
Position (kb)

B


PTPN20–A vs PTPN20–B
H11_PTPN20–AB vs PTPN20–A
H11_PTPN20–AB vs PTPN20–B
Diversity
BP
Position (kb)

C


Scale
chr10:
1 kb
hs1
47,769,500 47,770,000 47,770,500 47,771,000 47,771,500 47,772,000 47,772,500 47,773,000 47,773,500
H11
BP-A
CAT + Liftoff Gene Annotations
FRMPD2B-201
RepeatMasker Repetitive Elements
MER33#DNA/hAT-Charlie
LTR33
MER20
MLT2D
LTR16
MIRb
GC Percent
GC Percent in 5-Base Windows
100
0
Scale
chr10:
1 kb
hs1
49,016,500 49,017,000 49,017,500 49,018,000 49,018,500 49,019,000 49,019,500 49,020,000 49,020,500
H11
BP-B
CAT + Liftoff Gene Annotations
FRMPD2
RepeatMasker Repetitive Elements
MIRc#SINE/MIR
LTR16#LTR/ERVL
L2
LTR33
MER20
MER33
MLT2D
GC Percent
GC Percent in 5-Base Windows
100
0

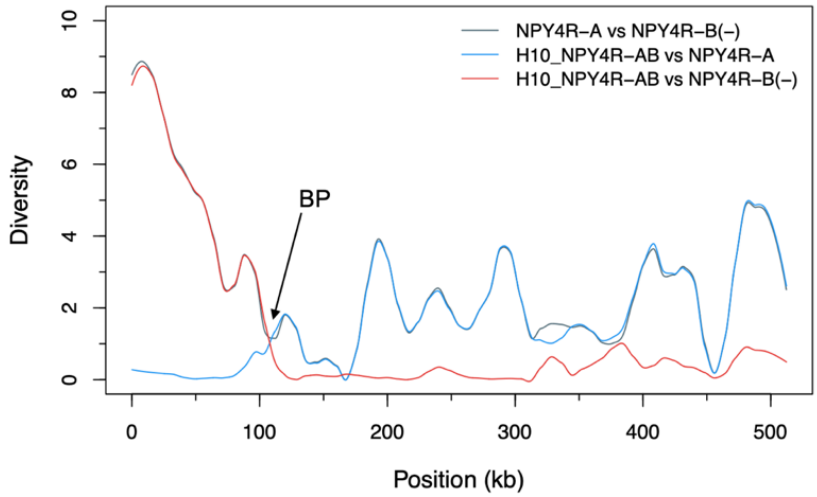
Supplemental Figure S6. H11 breakpoint analysis.

A) Diversity plot of H11 putative breakpoint region. The plot was obtained using the PopGenome R package, with width of 500 and jump of 100. H11_PTPN20-AB: sequence of H11 putative breakpoint region; PTPN20-A and PTPN20-B: sequences of segmental duplications that triggered the H11 rearrangement. The gray line refers to the comparison of PTPN20-A versus PTPN20-B, as a reference; the blue line refers to the comparison of H11_PTPN20-AB versus PTPN20-A; the red line refers to the comparison of H11_PTPN20-AB versus PTPN20-B. The breakpoint interval is marked by dotted lines.

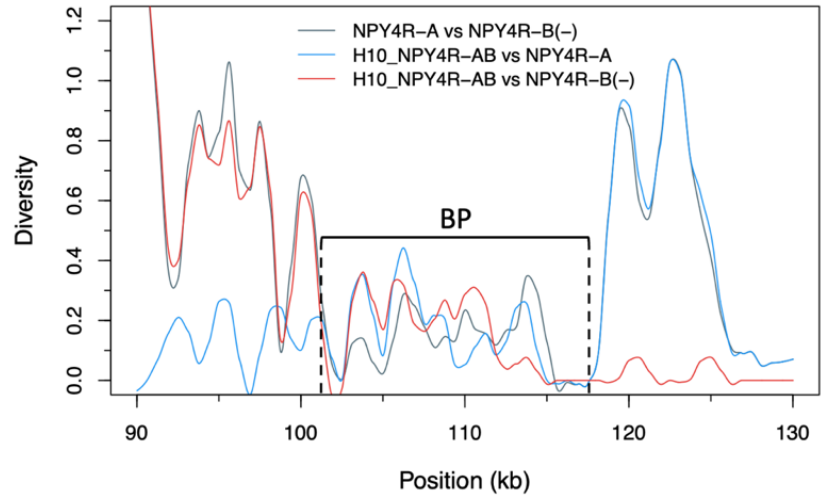
B) Diversity plot of a region zoomed around the H11 breakpoint. The plot was obtained using the PopGenome R package, with width of 100 and jump of 10. H11_PTPN20-AB: sequence of H11 putative breakpoint region; PTPN20-A and PTPN20-B: sequences of segmental duplications that triggered the H11 rearrangement. The gray line refers to the comparison of PTPN20-A versus PTPN20-B, as a reference; the blue line refers to the comparison of H11_PTPN20-AB versus PTPN20-A; the red line refers to the comparison of H11_PTPN20-AB versus PTPN20-B. The breakpoint interval of 1.6 kb is marked by dotted lines.

C) View in UCSC Genome Browser on human T2T-CHM13 v2.0 of gene annotation, repetitive elements, and GC content at the H11 breakpoints in the proximal and distal SD copies triggering the rearrangement.

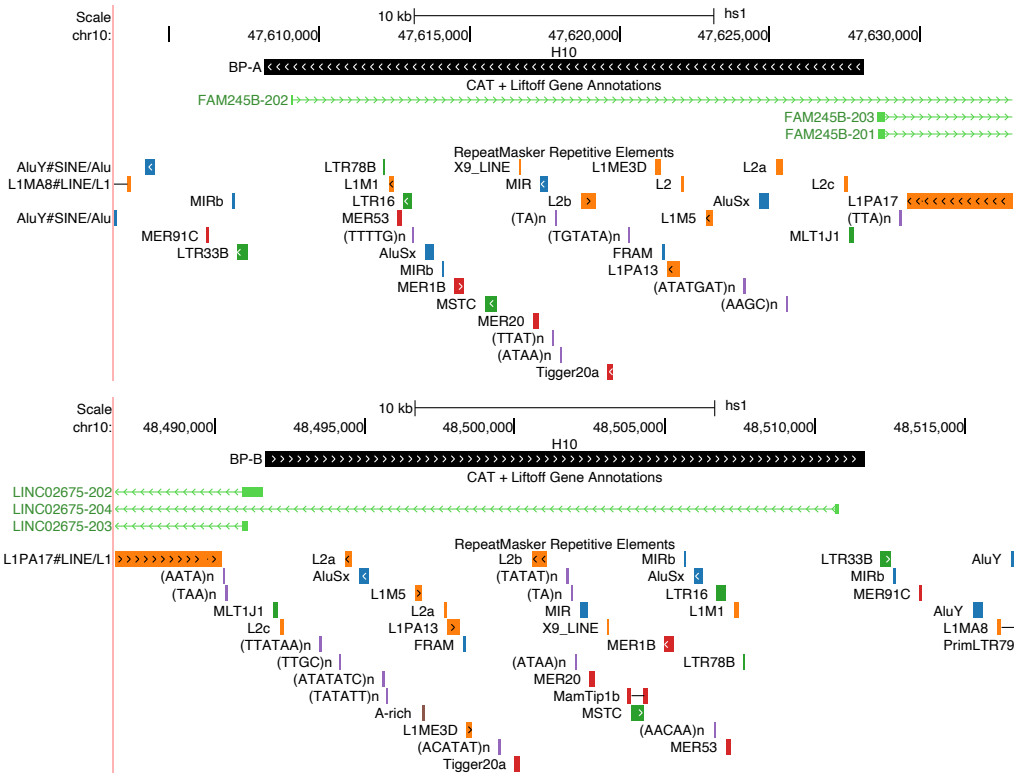
A


NPY4R–A vs NPY4R–B(–)
H10_NPY4R–AB vs NPY4R–A
H10_NPY4R–AB vs NPY4R–B(–)
BP
Diversity
Position (kb)

B


NPY4R–A vs NPY4R–B(–)
H10_NPY4R–AB vs NPY4R–A
H10_NPY4R–AB vs NPY4R–B(–)
BP
Diversity
Position (kb)

C


Scale
chr10:
10 kb
hs1
47,610,000
47,615,000
47,620,000
47,625,000
47,630,000
H10
BP-A
CAT + Liftoff Gene Annotations
FAM245B-202
FAM245B-203
FAM245B-201
RepeatMasker Repetitive Elements
AluY#SINE/Alu
L1MA8#LINE/L1
MIRb
AluY#SINE/Alu
MER91C
LTR33B
LTR78B
L1M1
LTR16
MER53
(TTTTG)n
AluSx
MIRb
MER1B
MSTC
MER20
(TTAT)n
(ATAA)n
Tigger20a
X9_LINE
MIR
L2b
(TA)n
(TGTATA)n
FRAM
L1PA13
(ATATGAT)n
(AAGC)n
L1ME3D
L2
L1M5
L2a
AluSx
L2c
L1PA17
(TTA)n
MLT1J1
Scale
chr10:
10 kb
hs1
48,490,000
48,495,000
48,500,000
48,505,000
48,510,000
48,515,000
H10
BP-B
CAT + Liftoff Gene Annotations
LINC02675-202
LINC02675-204
LINC02675-203
RepeatMasker Repetitive Elements
L1PA17#LINE/L1
(AATA)n
(TAA)n
MLT1J1
L2c
(TTATAA)n
(TTGC)n
(ATATATC)n
(TATATT)n
A-rich
L1ME3D
(ACATAT)n
Tigger20a
L2a
AluSx
L1M5
L2a
L1PA13
FRAM
L2b
(TATAT)n
(TA)n
MIR
X9_LINE
(ATAA)n
MER20
MamTip1b
MSTC
MIRb
AluSx
LTR16
L1M1
MER1B
LTR78B
(AACAA)n
MER53
LTR33B
MIRb
MER91C
AluY
L1MA8
PrimLTR79

Supplemental Figure S7. H10 breakpoint analysis.

A) Diversity plot of H10 putative breakpoint region. The plot was obtained using the PopGenome R package, with width of 500 and jump of 100. H10_NPY4R-AB: sequence of H10 putative breakpoint region; NPY4R-A and NPY4R-B: sequences of segmental duplications that triggered the H10 rearrangement. The gray line refers to the comparison of NPY4R-A versus NPY4R-B, as a reference; the blue line refers to the comparison of H10_NPY4R-AB versus NPY4R-A; the red line refers to the comparison of H10_ NPY4R-AB versus NPY4R-B. The breakpoint interval is marked by dotted lines.

B) Diversity plot of a region zoomed around the H10 breakpoint. This was obtained using the PopGenome R package, with width of 100 and jump of 10. H10_NPY4R-AB: sequence of H10 putative breakpoint region; NPY4R-A and NPY4R-B: sequences of segmental duplications that triggered the H10 rearrangement. The gray line refers to the comparison of NPY4R-A versus NPY4R-B, as a reference; the blue line refers to the comparison of H10_NPY4R-AB versus NPY4R-A; the red line refers to the comparison of H10_ NPY4R-AB versus NPY4R-B. The breakpoint interval of 20 kb is marked by dotted lines.

C) View in UCSC Genome Browser on human T2T-CHM13 v2.0 of gene annotation, repetitive elements, and GC content at the H10 breakpoints in the proximal and distal SD copies triggering the rearrangement.