Supplemental figures for

The rate of *de novo* structural variation is increased in *in vitro* produced offspring and preferentially affects the paternal genome

Young-Lim Lee[1,2,*], Aniek C. Bouwman[2], Chad Harland[3], Mirte Bosse[2], Gabriel Costa Monteiro Moreira[1], Roel F. Veerkamp[2], Erik Mullaart[4], Nadine Cambisano[5], Martien A. M. Groenen[2], Latifa Karim[1,5], Wouter Coppieters[1,5], Michel Georges[1*], Carole Charlier[1,*]

[1]Unit of Animal Genomics, GIGA-R & Faculty of Veterinary Medicine, University of Liège, Liège, Belgium

[2]Wageningen University & Research, Animal Breeding and Genomics, Wageningen, the Netherlands

[3]Livestock Improvement Corporation, Hamilton, New Zealand

[4]CRV B.V., Arnhem, the Netherlands

[5]GIGA Genomics Platform, GIGA Institute, University of Liège, Liège, Belgium

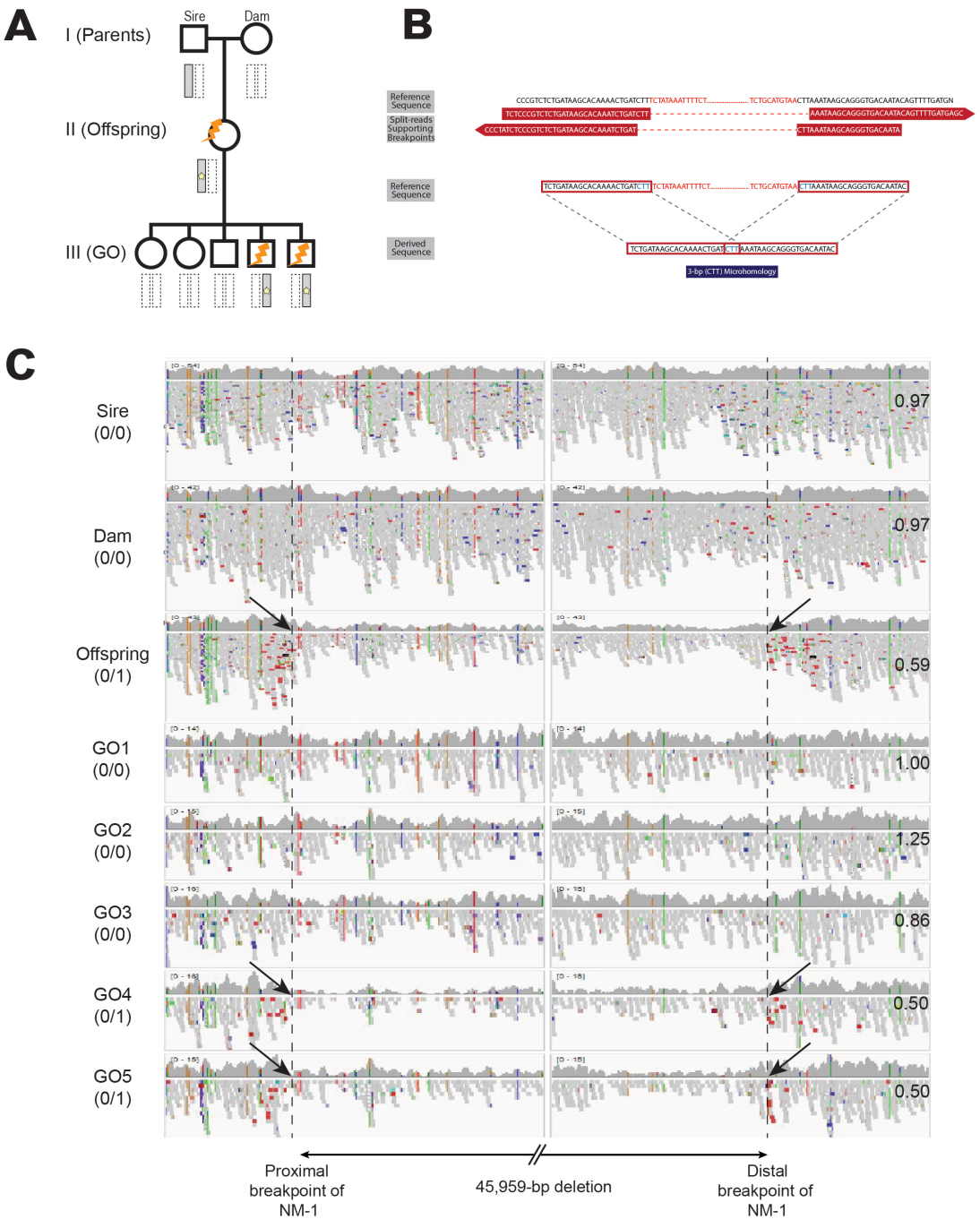
Correspondence:

Michel Georges (michel.georges@uliege.be)

Carole Charlier (carole.charlier@uliege.be)

Young-Lim Lee (younglim.lee@uliege.be)

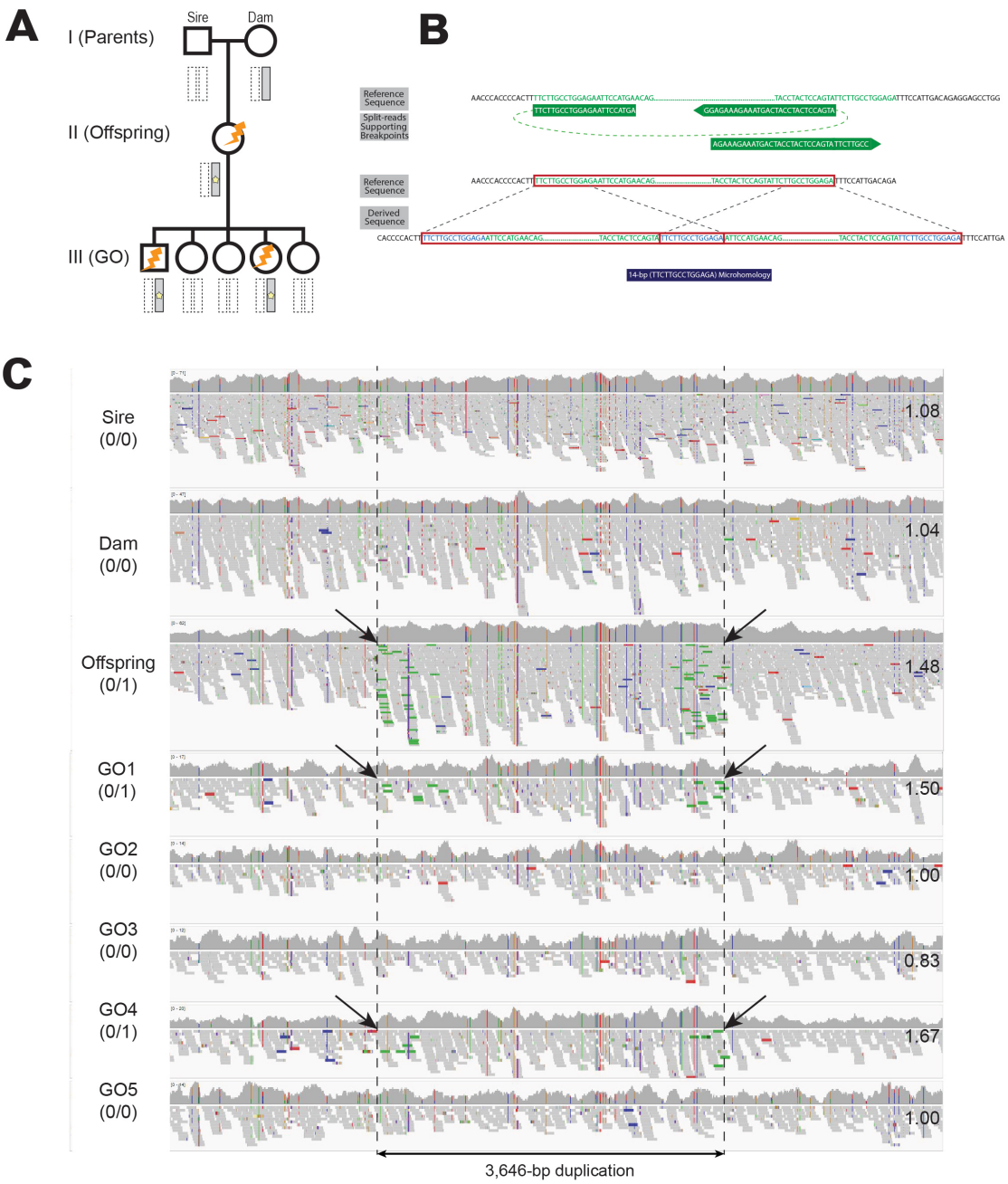
Contents



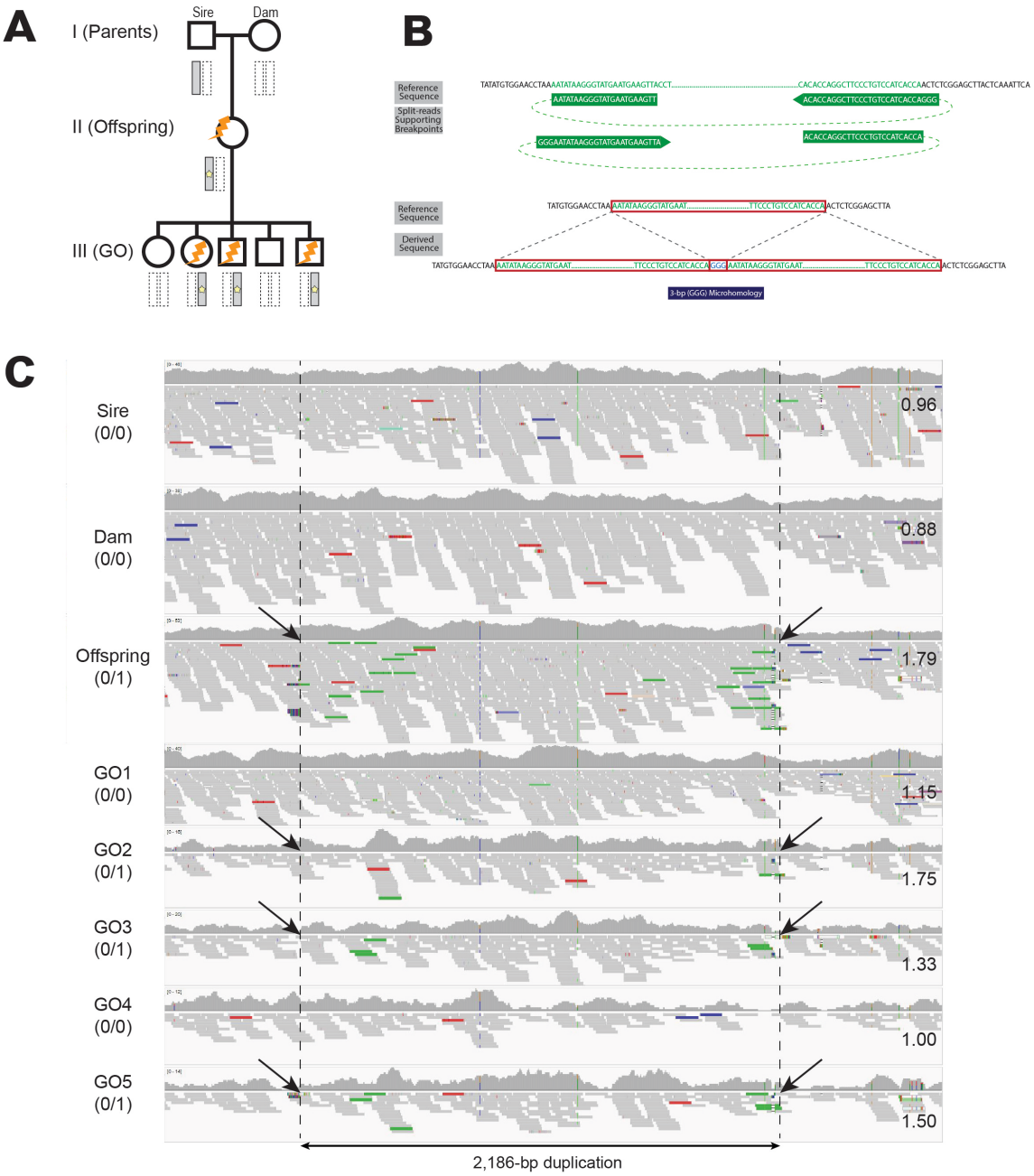
Supplementary Figure S1. Pedigree, breakpoints, and the IGV screen capture of NM-1

A. Depiction of NM-1 in the three-generational (parents-offspring-GO) pedigree. Males are marked with squares and females are marked with circles. The occurrence of the NM-1 was detected in offspring and then transmitted to ≥ 1 GO. Animals carrying NM-1 are marked with the bolt mark. To indicate the paternal origin of NM-1, the bolt mark in the offspring is put on the left side. The haplotype which the dnSV occurred upon is marked with the grey bar, whereas other haplotypes are marked with dotted bar. The dnSV occurred upon paternal haplotype of the offspring, which is indicated by the yellow star symbol on the grey haplotype. Co-segregation of

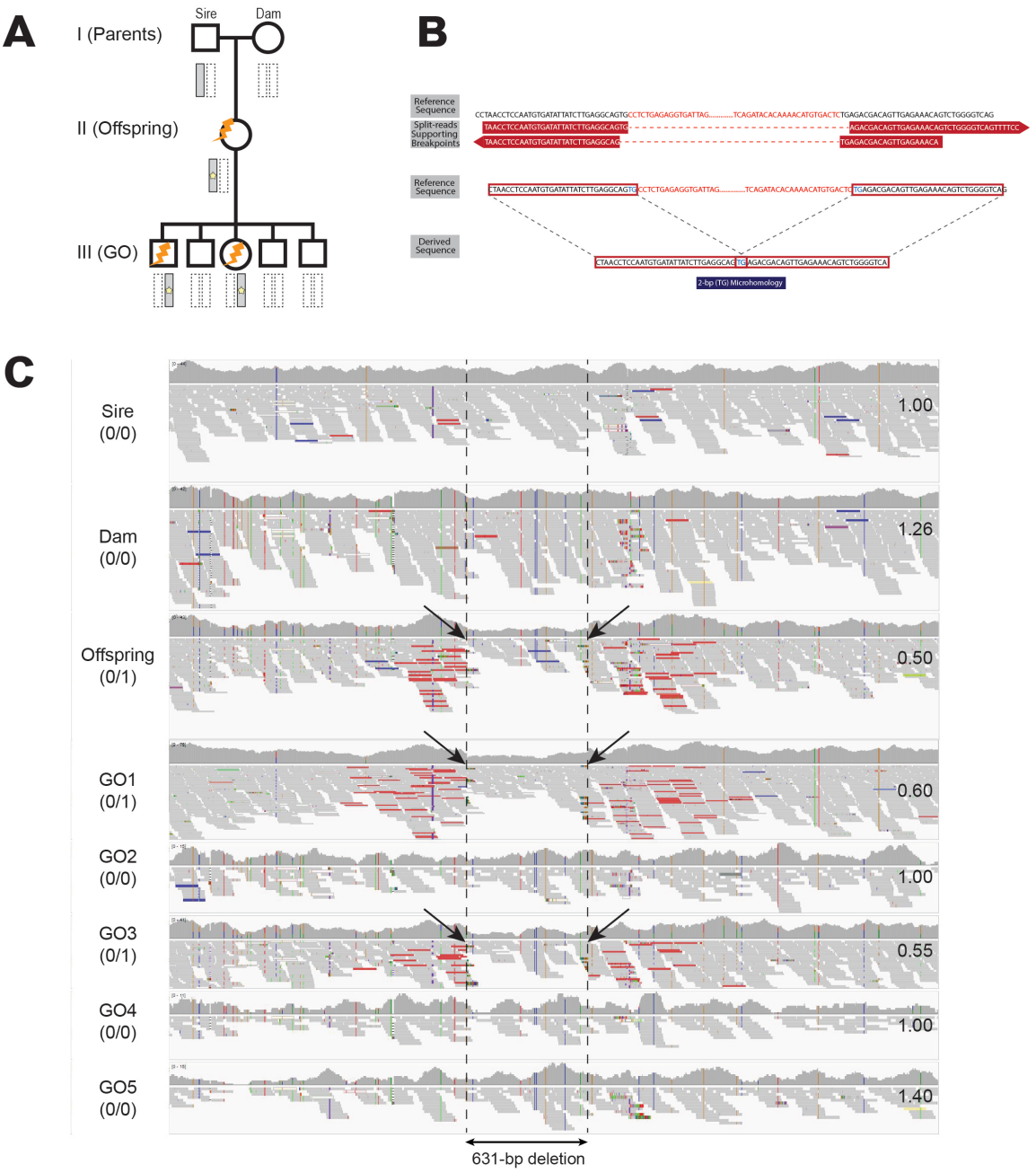
the grey haplotype and the NM-1 indicated perfect linkage between the two. **B.** Characterization of the breakpoints of NM-1. Reference sequence marked with black is unaffected sequence, and the one marked with red is affected sequence. Discordant read-pairs spanning the breakpoints are visualized by red bars with white color sequences (marked with arrows). The lower part shows the breakpoint sequence homology. The sequences shown in blue indicate the homology. For the following figures (Supplementary Figs. 2-19), the same color schemes (red) were used for deletions. In case of duplications, green color was used to mark discordant read-pairs and affected sequence. **C.** Integrative Genome Browser (IGV) (Robinson et al. 2011) screen capture of the NM-1 in the current pedigree. The dotted vertical lines indicate proximal (left one) and distal breakpoints (right one), respectively. The colorful reads indicate soft- or hard- clipped reads thereby indicating a deletion. Left panel shows the animals (sire, dam, offspring, and GO) and their genotypes. The numbers on the right side of the panel are obtained from duphold software (Pedersen and Quinlan 2019). Duphold calculates the fold-coverage change of the SV affected region against the 1-kb flanking sites, which can be useful in interpreting the actual read-depth change occurring in the SV of interest.



Supplementary Figure S2. Pedigree, breakpoints, and the IGV screen capture of NM-2
A. Depiction of NM-2 in the three-generational pedigree. The legend is identical to supplementary figure 1A. **B.** Characterization of the breakpoints of NM-2. The legend is identical to supplementary figure 1B. **C.** IGV (Robinson et al. 2011) screen capture of the NM-2 in the current pedigree. Legend is identical to supplementary figure 1C. The discordant reads (green) aligned at the proximal and distal breakpoints support the presence of NM-2.

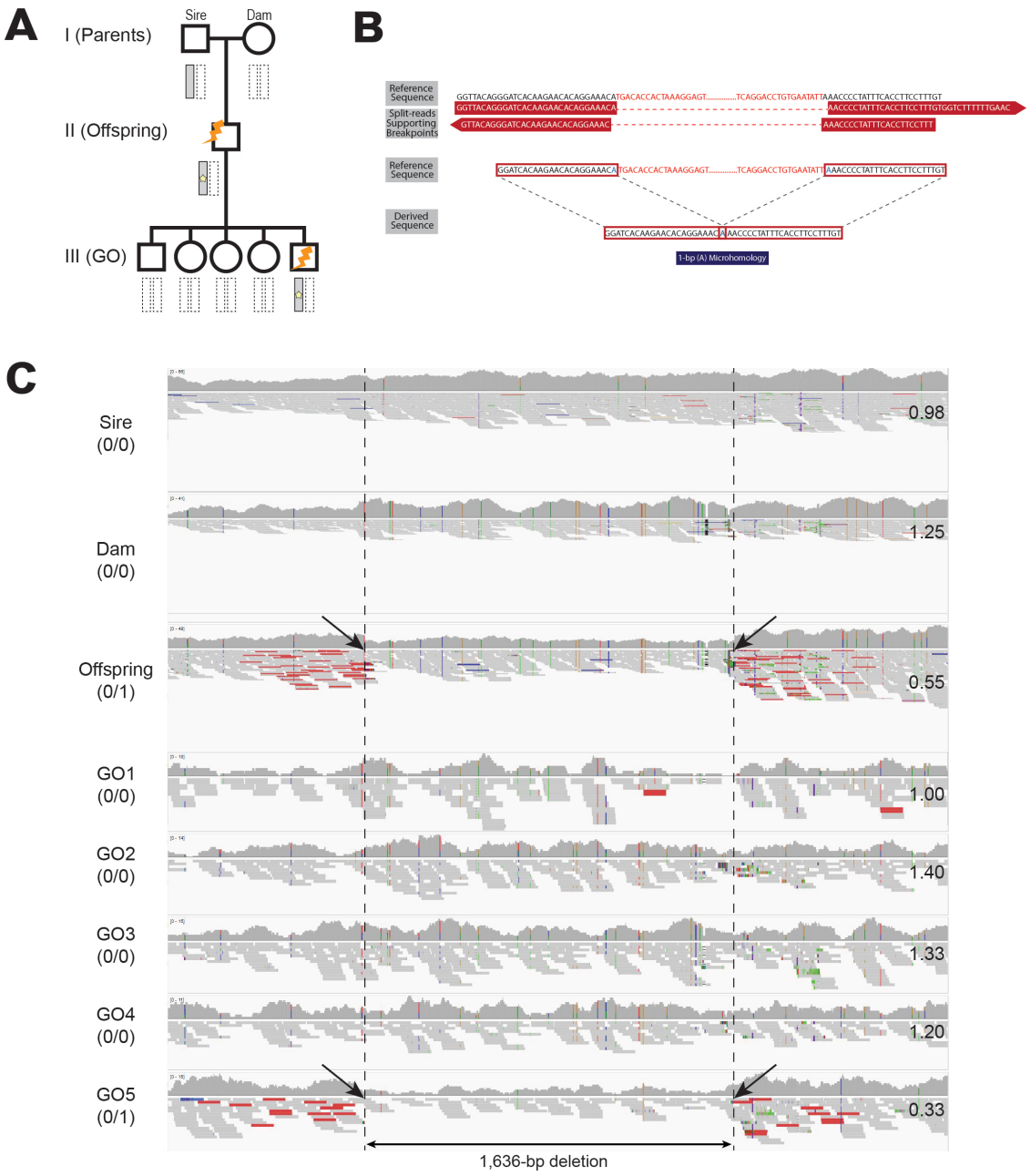


Supplementary Figure S3. Pedigree, breakpoints, and the IGV screen capture of NM-3
A. Depiction of NM-3 in the three-generational pedigree. The legend is identical to supplementary figure 1A. **B.** Characterization of the breakpoints of NM-3. The legend is identical to supplementary figure 1B. **C.** IGV (Robinson et al. 2011) screen capture of the NM-3 in the current pedigree. Legend is identical to supplementary figure 1C. The discordant reads (green) aligned at the proximal and distal breakpoints support the presence of NM-3.



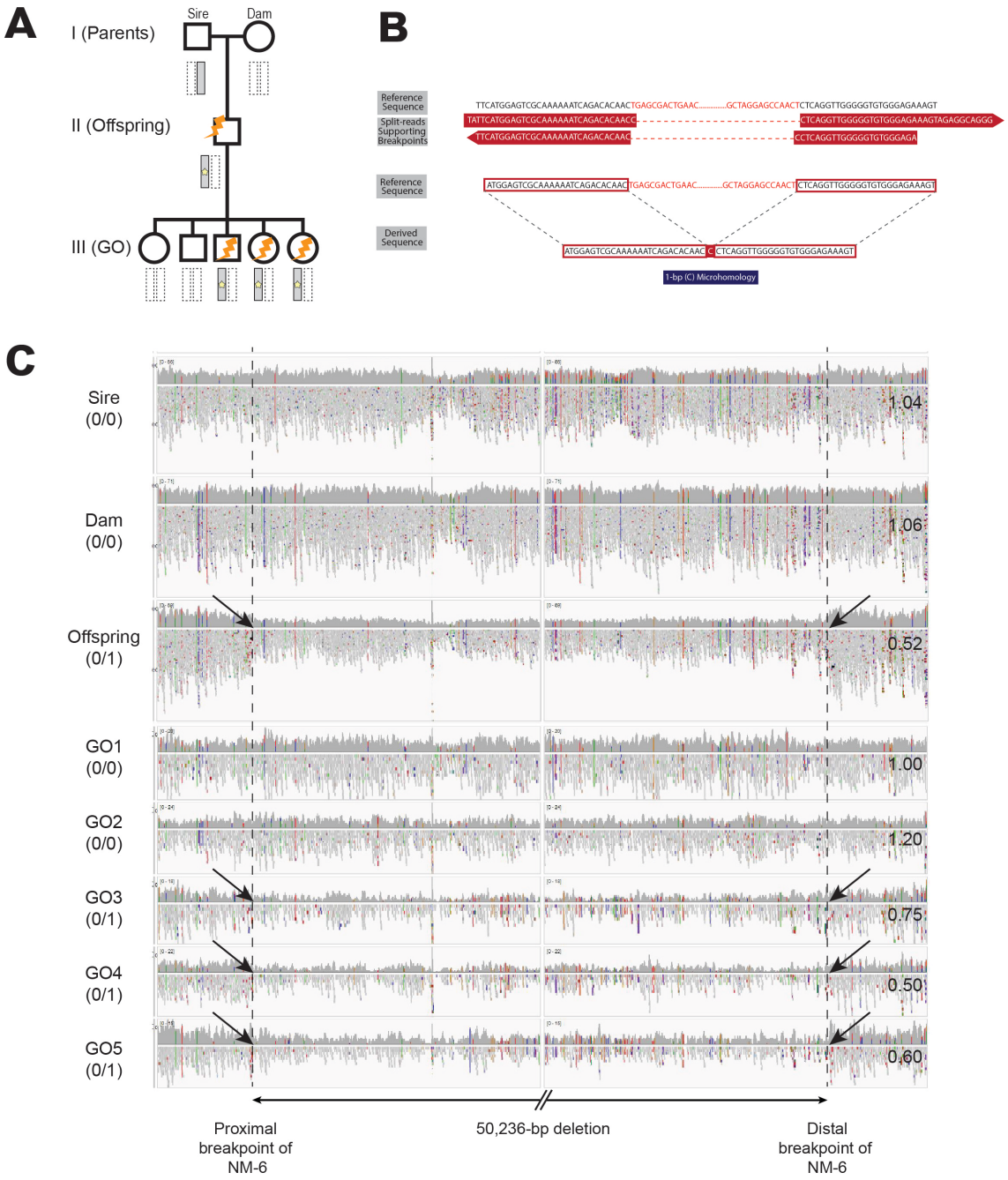
Supplementary Figure S4. Pedigree, breakpoints, and the IGV screen capture of NM-4

A. Depiction of NM-4 in the three-generational pedigree. The legend is identical to supplementary figure 1A. **B.** Characterization of the breakpoints of NM-4. The legend is identical to supplementary figure 1B. **C.** IGV (Robinson et al. 2011) screen capture of the NM-4 in the current pedigree. Legend is identical to supplementary figure 1C. The discordant reads (red) aligned at the proximal and distal breakpoints support the presence of NM-4.



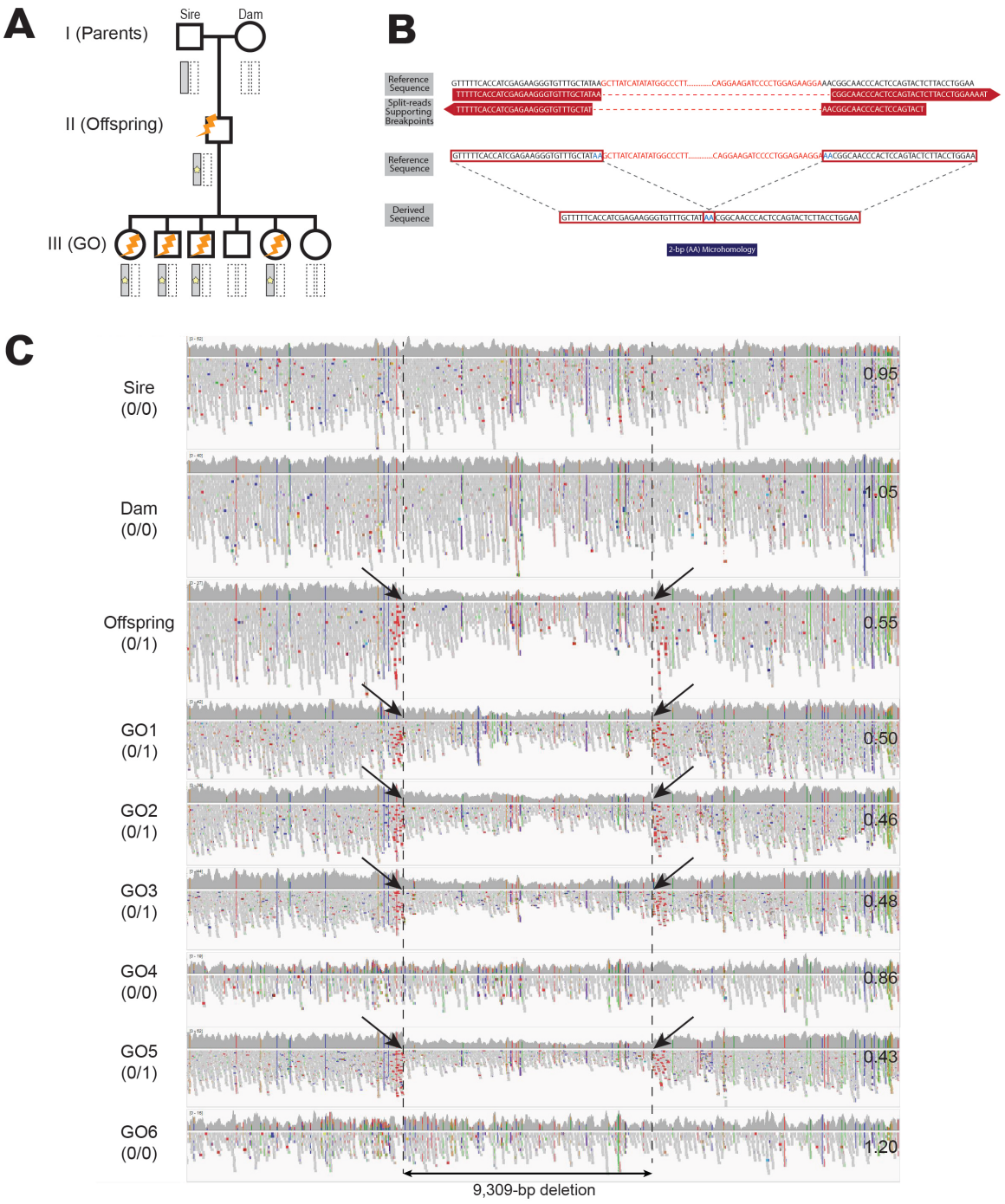
Supplementary Figure S5. Pedigree, breakpoints, and the IGV screen capture of NM-5

A. Depiction of NM-5 in the three-generational pedigree. The legend is identical to supplementary figure 1A. **B.** Characterization of the breakpoints of NM-5. The legend is identical to supplementary figure 1B. **C.** IGV (Robinson et al. 2011) screen capture of the NM-5 in the current pedigree. Legend is identical to supplementary figure 1C. The discordant reads (red) aligned at the proximal and distal breakpoints support the presence of NM-5.



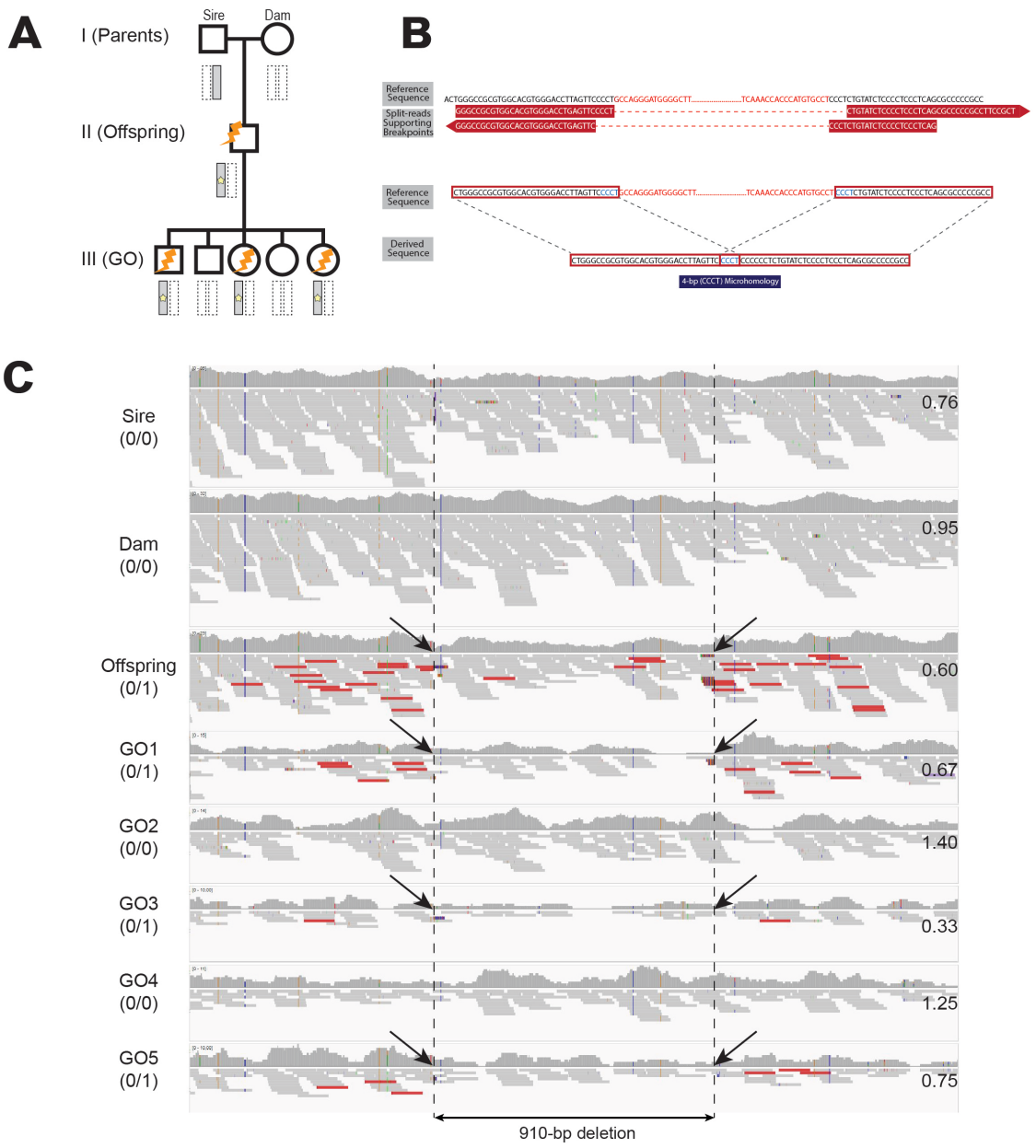
Supplementary Figure S6. Pedigree, breakpoints, and the IGV screen capture of NM-6

A. Depiction of NM-6 in the three-generational pedigree. The legend is identical to supplementary figure 1A. **B.** Characterization of the breakpoints of NM-6. The legend is identical to supplementary figure 1B. **C.** IGV (Robinson et al. 2011) screen capture of the NM-6 in the current pedigree. Legend is identical to supplementary figure 1C. The discordant reads (red) aligned at the proximal and distal breakpoints and the reduced sequencing coverage support the presence of NM-6.



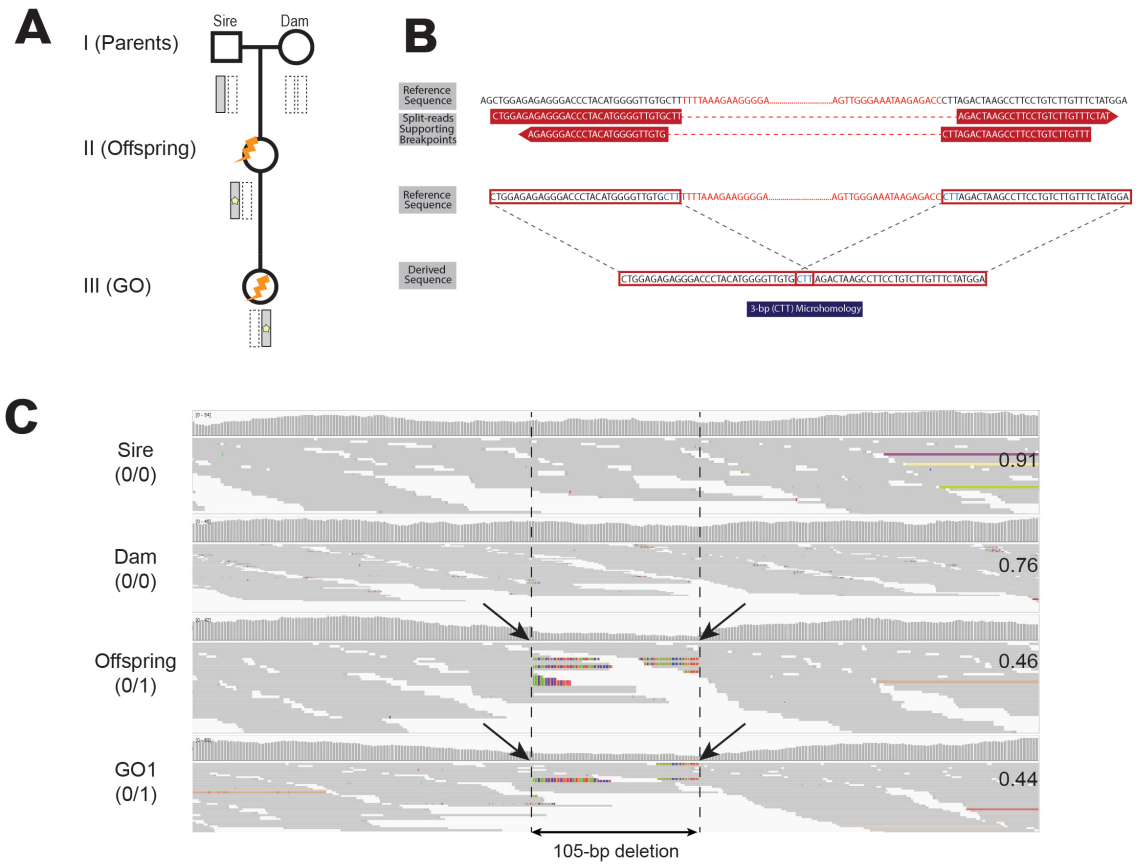
Supplementary Figure S7. Pedigree, breakpoints, and the IGV screen capture of NM-7

A. Depiction of NM-7 in the three-generational pedigree. The legend is identical to supplementary figure 1A. **B.** Characterization of the breakpoints of NM-7. The legend is identical to supplementary figure 1B. **C.** IGV (Robinson et al. 2011) screen capture of the NM-7 in the current pedigree. Legend is identical to supplementary figure 1C. The discordant reads (red) aligned at the proximal and distal breakpoints support the presence of NM-7.



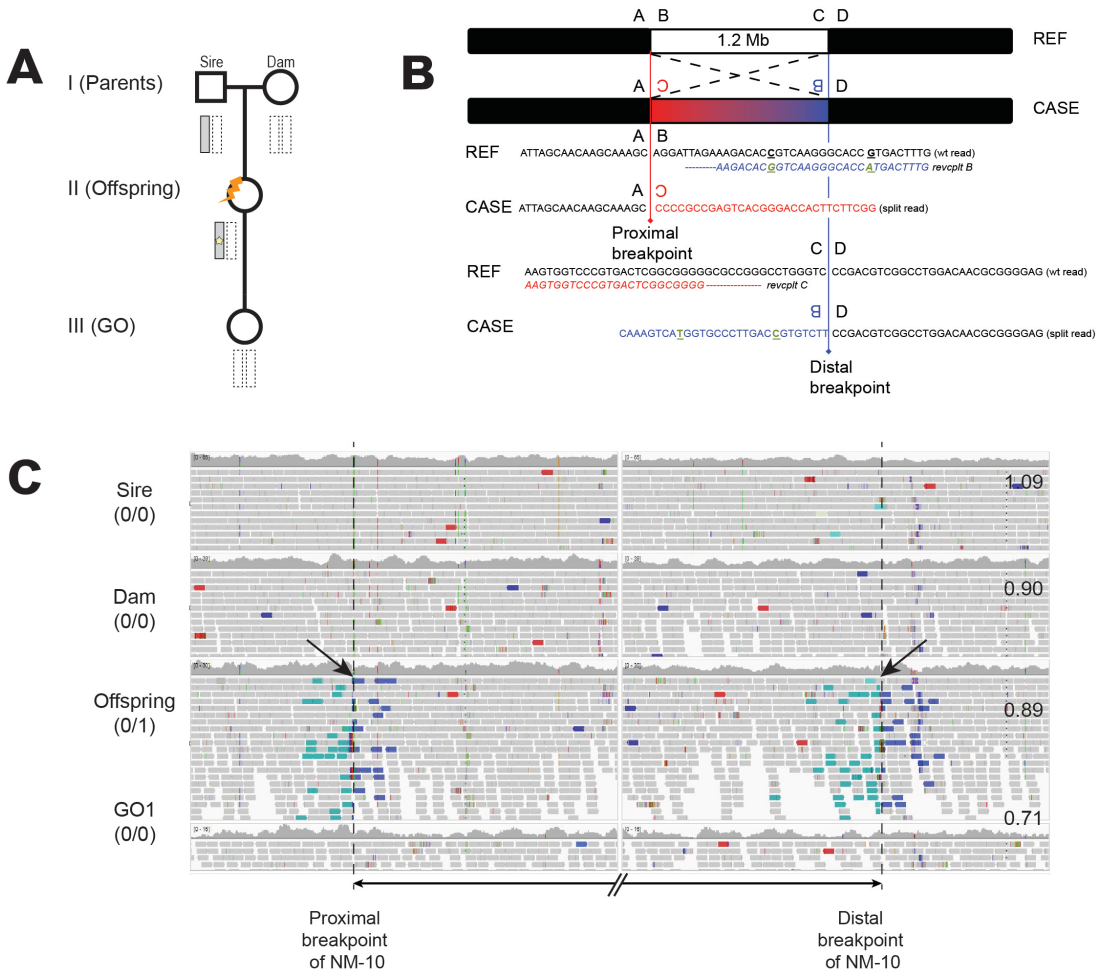
Supplementary Figure S8. Pedigree, breakpoints, and the IGV screen capture of NM-8

A. Depiction of NM-8 in the three-generational pedigree. The legend is identical to supplementary figure 1A. **B.** Characterization of the breakpoints of NM-8. The legend is identical to supplementary figure 1B. **C.** IGV (Robinson et al. 2011) screen capture of the NM-8 in the current pedigree. Legend is identical to supplementary figure 1C. The discordant reads (red) aligned at the proximal and distal breakpoints support the presence of NM-8.



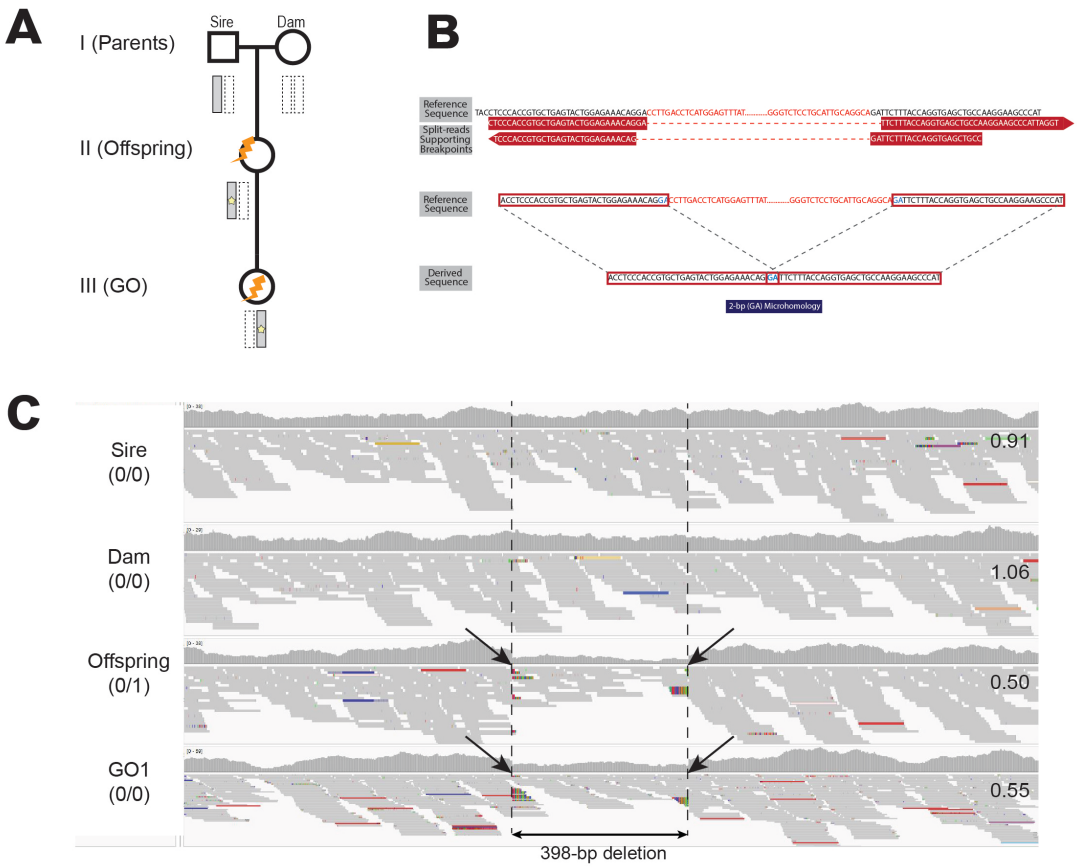
Supplementary Figure S9. Pedigree, breakpoints, and the IGV screen capture of NM-9

A. Depiction of NM-9 in the three-generational pedigree. The legend is identical to supplementary figure 1A. **B.** Characterization of the breakpoints of NM-9. The legend is identical to supplementary figure 1B. **C.** IGV (Robinson et al. 2011) screen capture of the NM-9 in the current pedigree. Legend is identical to supplementary figure 1C. The split reads (marked with multiple colors) aligned at the proximal and distal breakpoints support the presence of NM-9.

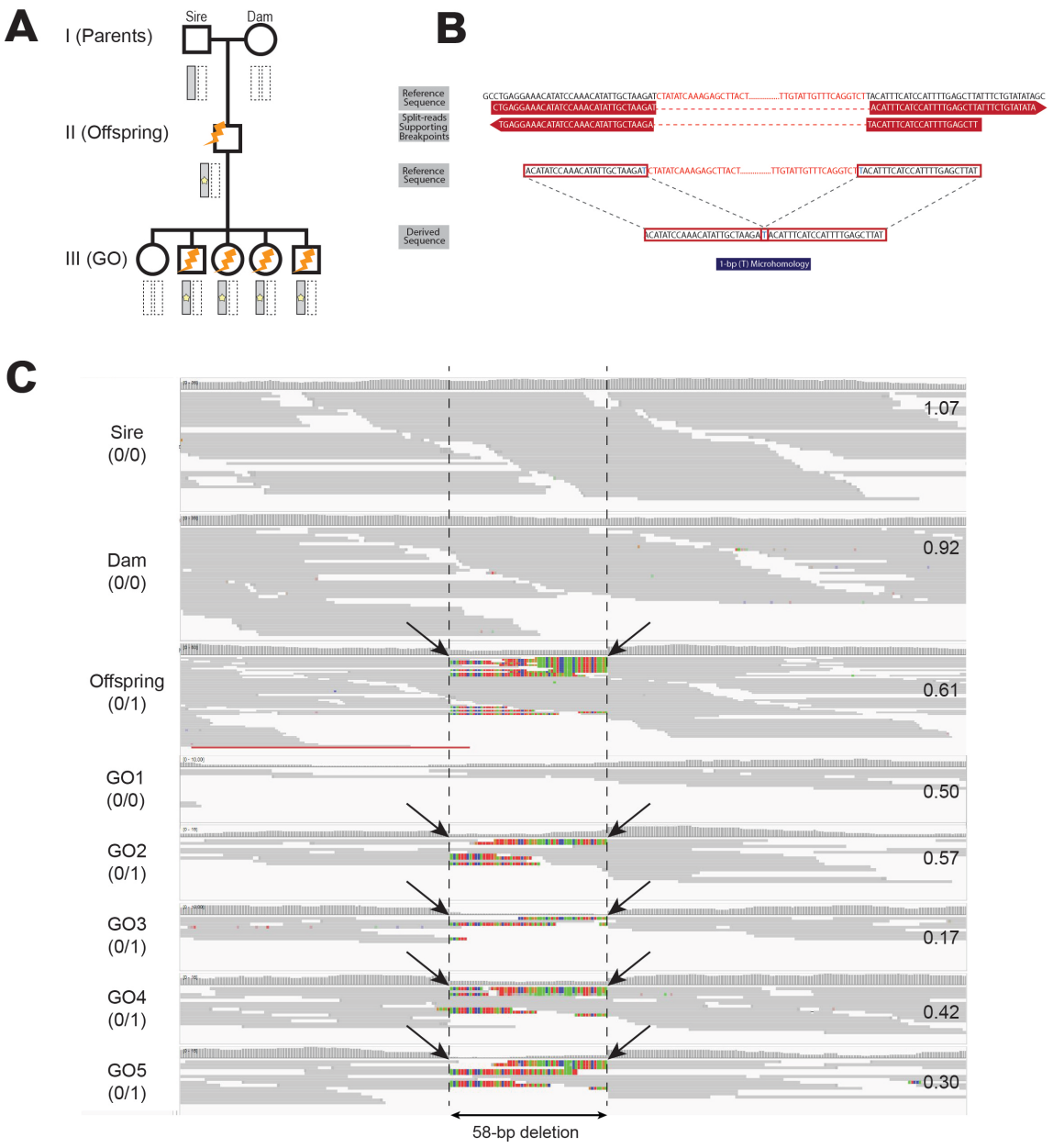


Supplementary Figure S10. Pedigree, breakpoints, and the IGV screen capture of NM-10

A. Depiction of NM-10 in the three-generational pedigree. The legend is identical to supplementary figure 1A. **B.** Characterization of the breakpoints of NM-10. The breakpoint junctions are marked with A, B, C, and D. In the reference genome, the proximal breakpoint is bridged with A-B junction and the distal breakpoint is bridged with C-D junction. In the NM-10 carrier, these junctions are replaced by A-C and B-D, respectively. The sequence below shows wild-type sequence (black) and reverse complementary sequence (blue). The sequence aligned at the C junction in the case is marked with red, and the sequenced aligned at the B junction in the case is marked with blue. The SNPs are marked with bold underlined black (wild-type) green (derived) characters. **C**. IGV (Robinson et al. 2011) screen capture of the NM-10 in the current pedigree. Legend is identical to supplementary figure 1C. The light blue and dark blue reads are discordant read pairs with forward-forward or reverse-reverse orientations, respectively. These discordant read pairs together with the soft-clipped reads (colorful reads) at the breakpoints indicate an inversion.

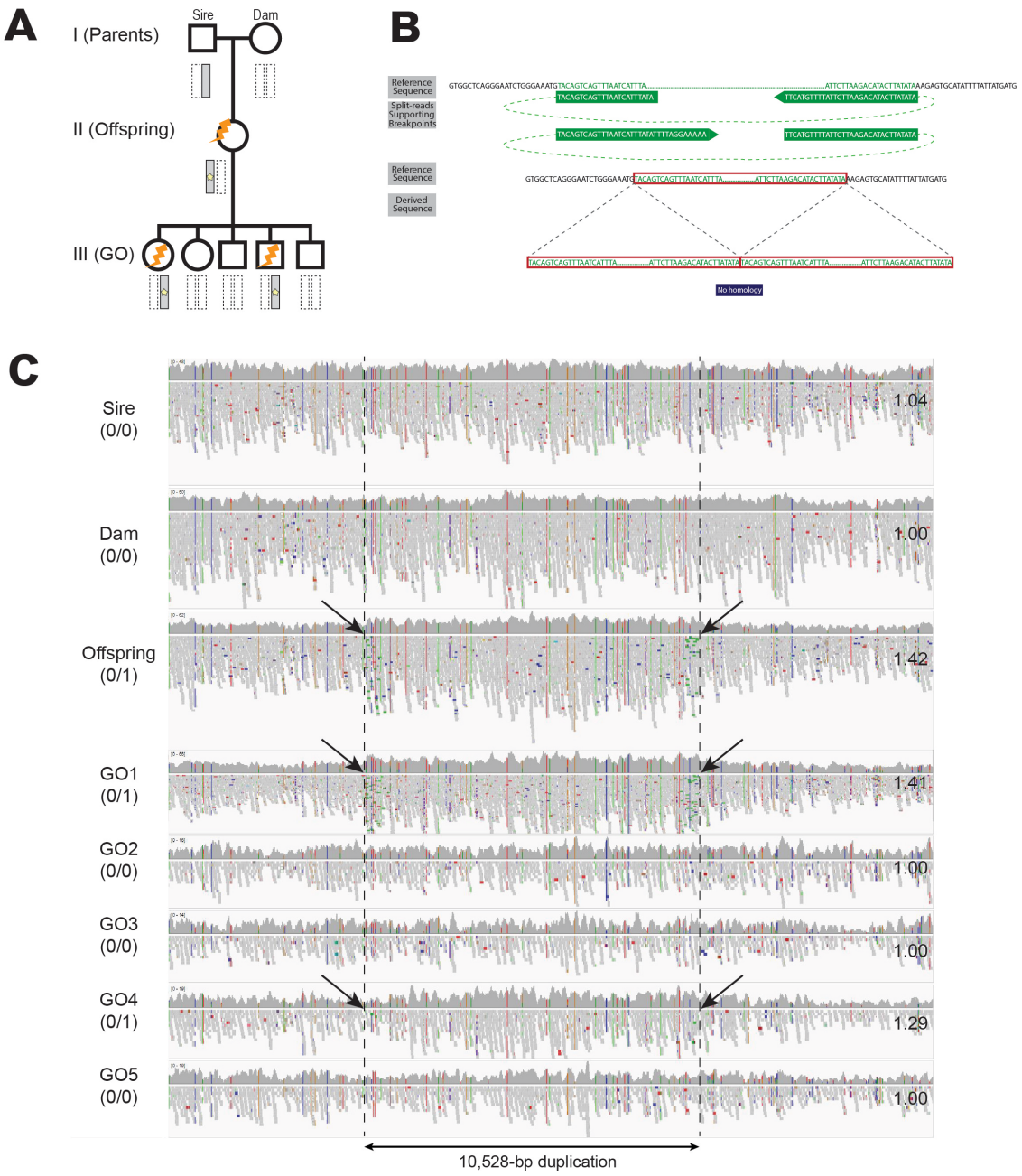


Supplementary Figure S11. Pedigree, breakpoints, and the IGV screen capture of NM-11
A. Depiction of NM-11 in the three-generational pedigree. The legend is identical to supplementary figure 1A. **B.** Characterization of the breakpoints of NM-11. The legend is identical to supplementary figure 1B. **C.** IGV (Robinson et al. 2011) screen capture of the NM-11 in the current pedigree. Legend is identical to supplementary figure 1C. The discordant reads (red) aligned at the proximal and distal breakpoints support the presence of NM-11.

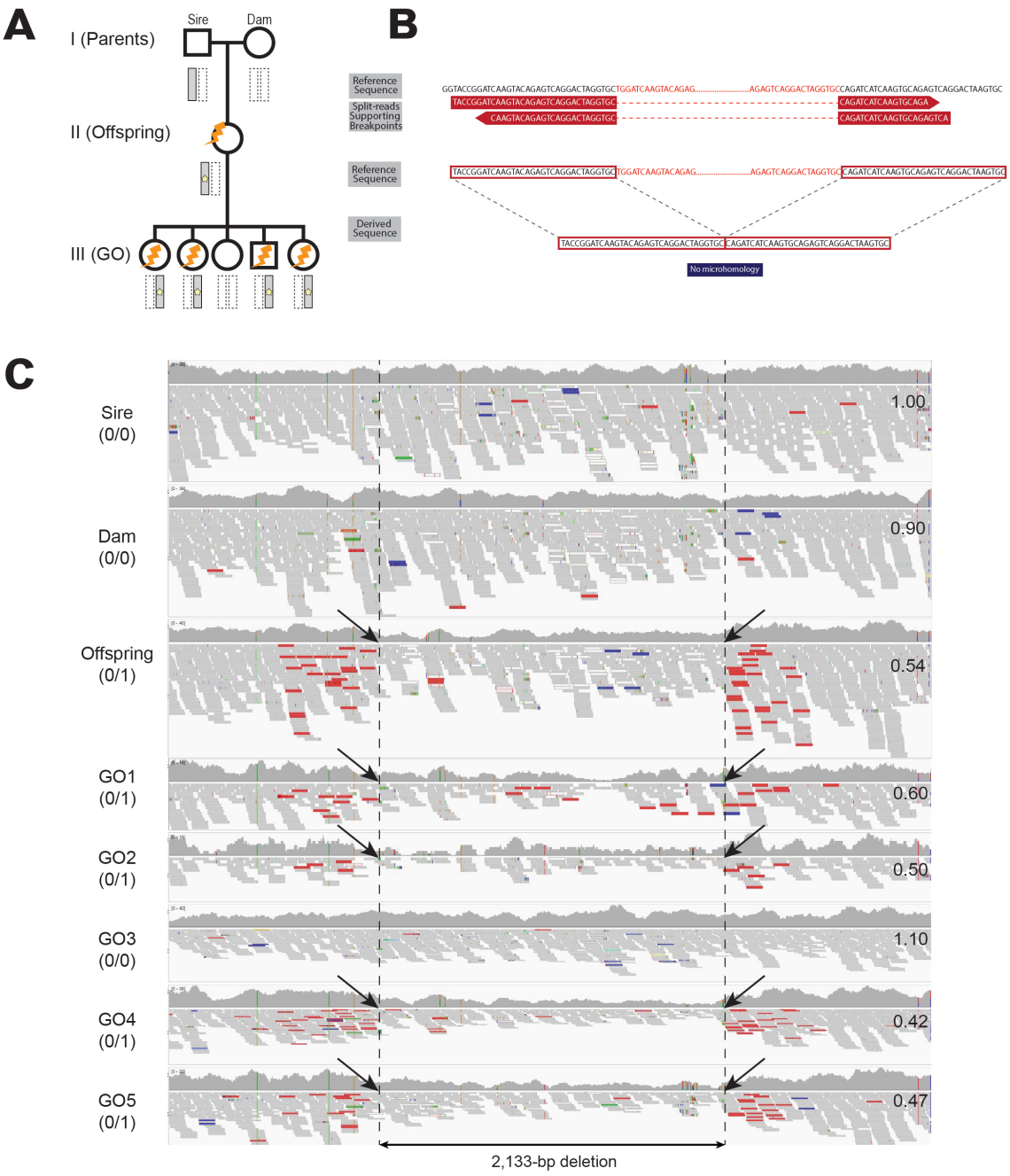


Supplementary Figure S12. Pedigree, breakpoints, and the IGV screen capture of A-1

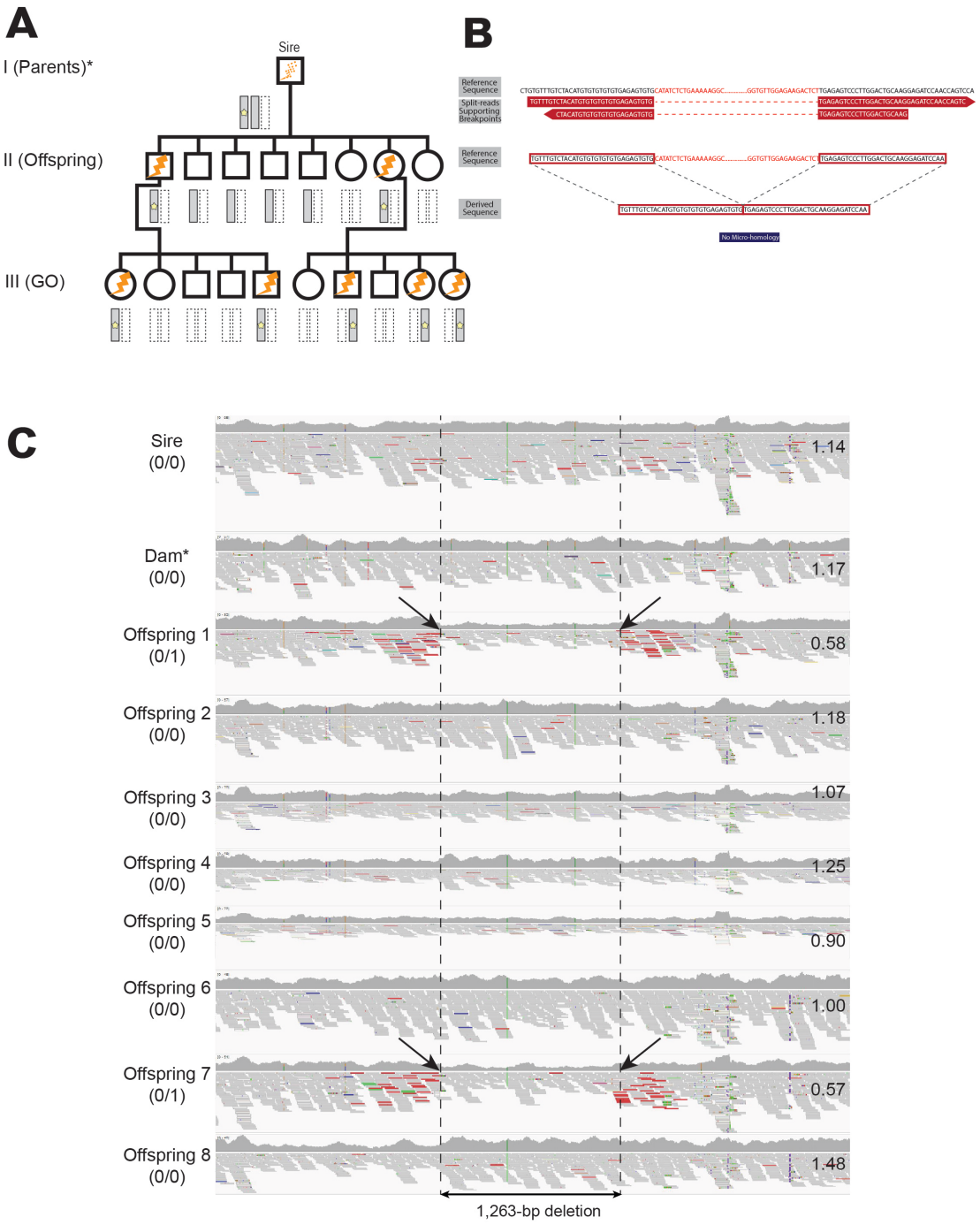
A. Depiction of A-1 in the three-generational pedigree. The legend is identical to supplementary figure 1A. **B.** Characterization of the breakpoints of A-1. The legend is identical to supplementary figure 1B. **C.** IGV (Robinson et al. 2011) screen capture of the A-1 in the current pedigree. Legend is identical to supplementary figure 1C. The discordant reads (red) aligned at the proximal and distal breakpoints support the presence of A-1.



Supplementary Figure S13. Pedigree, breakpoints, and the IGV screen capture of A-2
A. Depiction of A-2 in the three-generational pedigree. The legend is identical to supplementary figure 1A. **B.** Characterization of the breakpoints of A-2. The legend is identical to supplementary figure 1B. **C.** IGV (Robinson et al. 2011) screen capture of the A-2 in the current pedigree. Legend is identical to supplementary figure 1C. The discordant reads (green) aligned at the proximal and distal breakpoints support the presence of A-2.

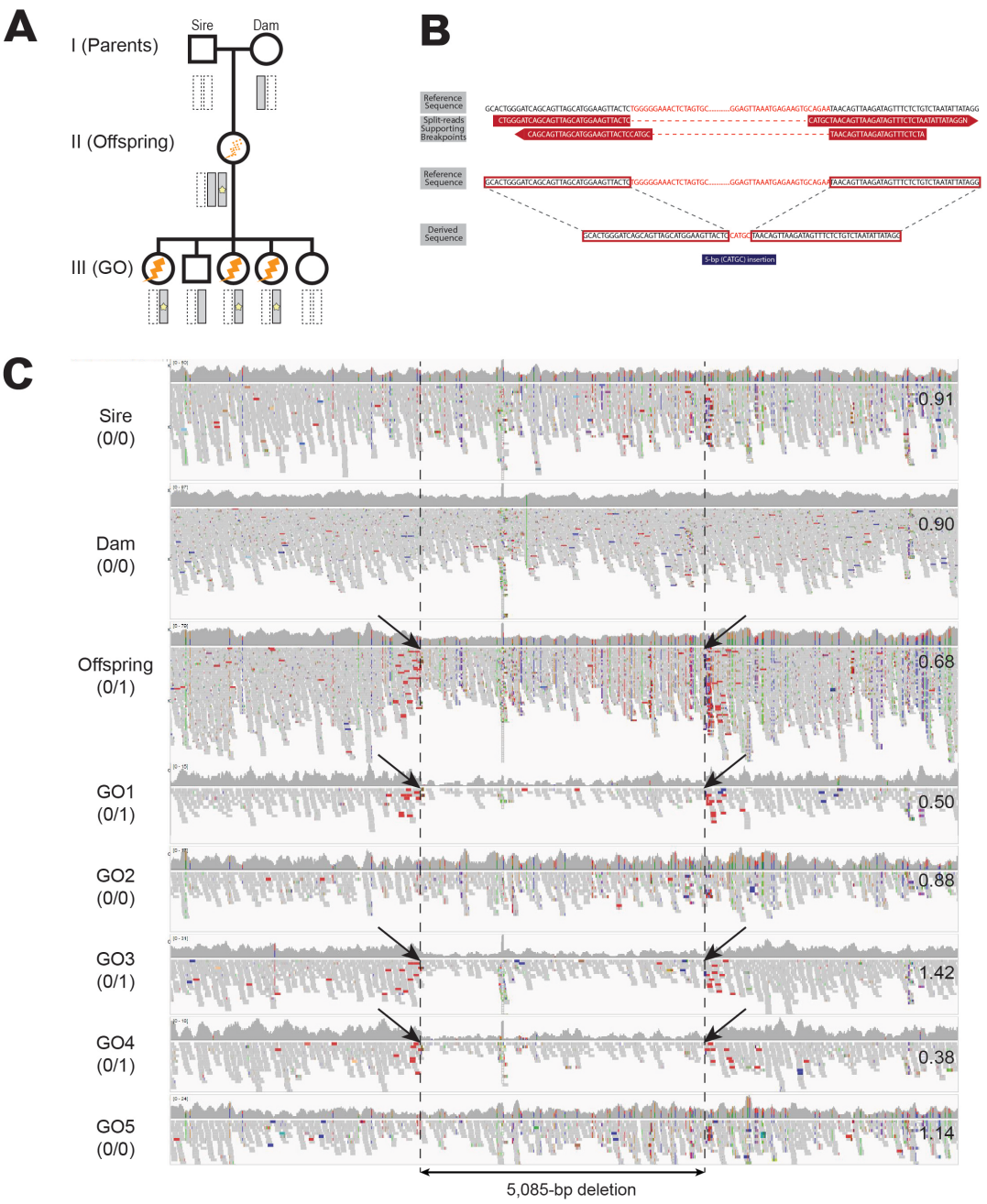


Supplementary Figure S14. Pedigree, breakpoints, and the IGV screen capture of A-3
A. Depiction of A-3 in the three-generational pedigree. The legend is identical to supplementary figure 1A. **B.** Characterization of the breakpoints of A-3. The legend is identical to supplementary figure 1B. **C.** IGV (Robinson et al. 2011) screen capture of the A-3 in the current pedigree. Legend is identical to supplementary figure 1C. The discordant reads (red) aligned at the proximal and distal breakpoints support the presence of A-3.



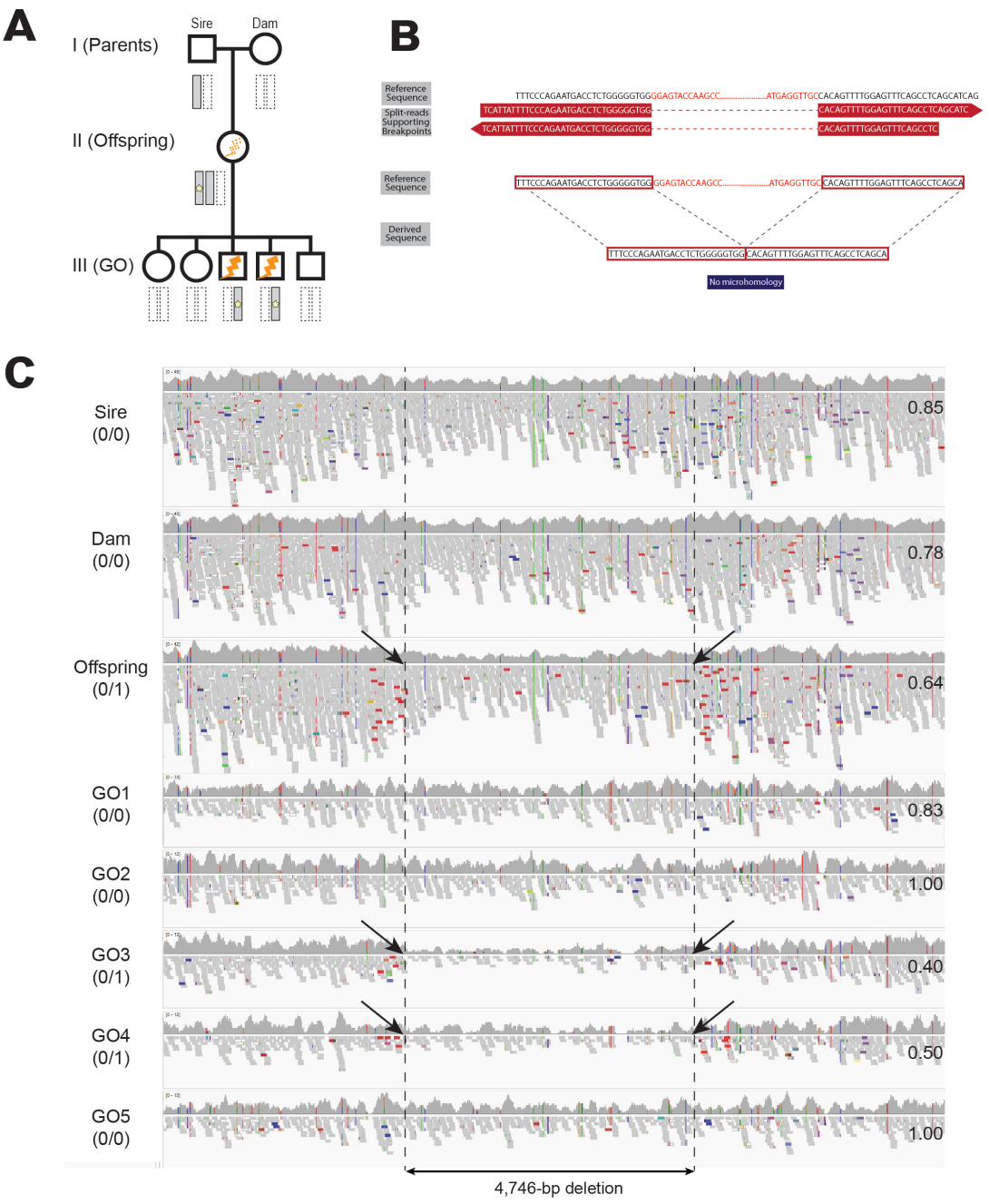
Supplementary Figure S15. Pedigree, breakpoint, and the IGV screen capture of M-1

A. Depiction of M-1 in the three-generational pedigree. The legend is identical to supplementary figure 1A. **B.** Characterization of the breakpoints of M-1. The legend is identical to supplementary figure 1B. **C.** IGV (Robinson et al. 2011) screen capture of the M-1 in the current pedigree. Legend is identical to supplementary figure 1C. For brevity, only sire, dam, and the eight offspring of the sire is visualized (grand-offspring not shown). Dam marked with asterisk is the maternal animal of the 7th offspring. Dam of 2nd offspring was not available, hence not shown. The discordant reads (red) aligned at the proximal and distal breakpoints support the presence of M-1.



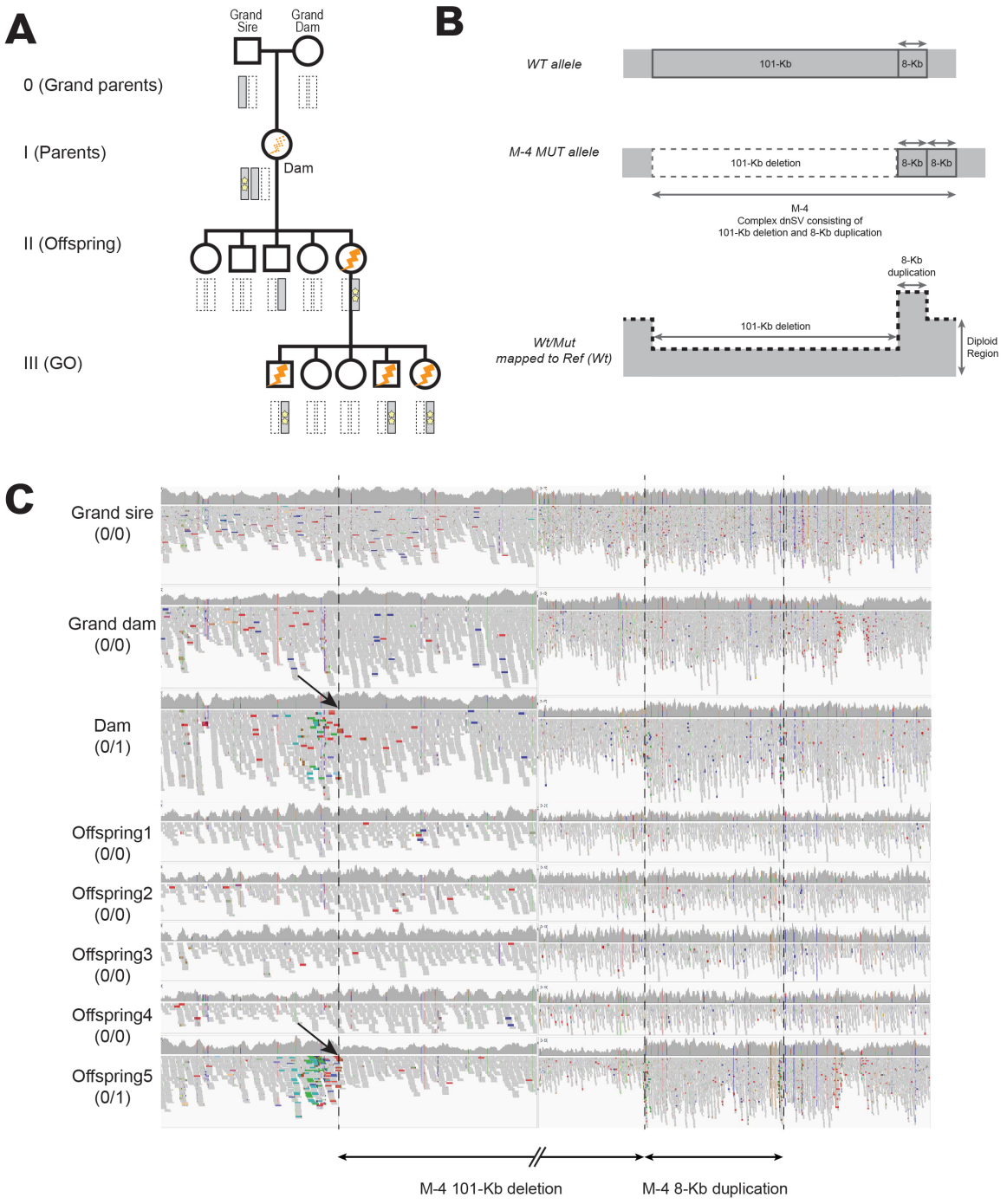
Supplementary Figure S16. Pedigree, breakpoint, and the IGV screen capture of M-2

A. Depiction of M-2 in the three-generational pedigree. The legend is identical to supplementary figure 1A. **B.** Characterization of the breakpoints of M-2. The legend is identical to supplementary figure 1B. **C.** IGV (Robinson et al. 2011) screen capture of the M-2 in the current pedigree. Legend is identical to supplementary figure 1C. The discordant reads (red) aligned at the proximal and distal breakpoints support the presence of M-2.

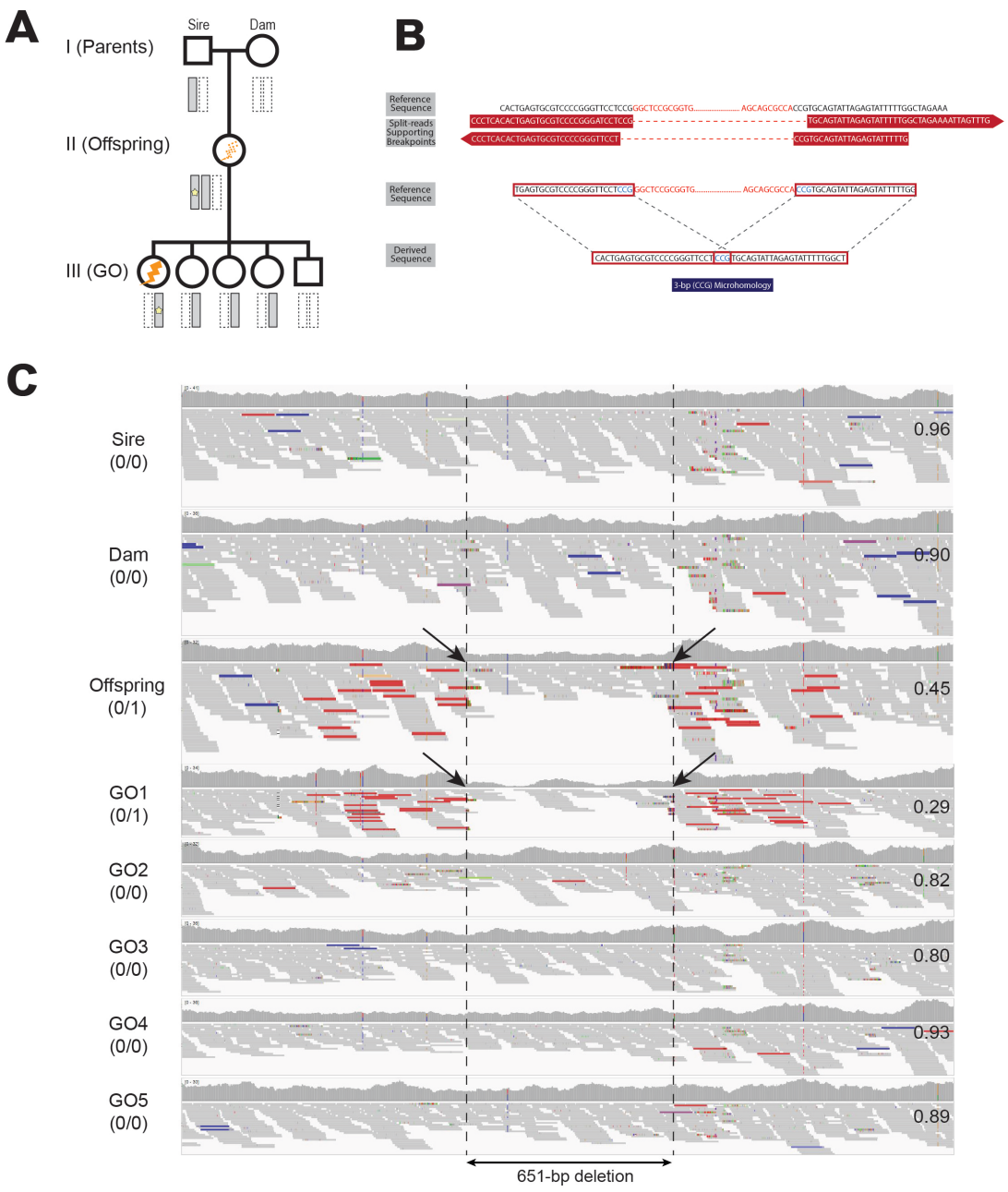


Supplementary Figure S17. Pedigree, breakpoint, and the IGV screen capture of M-3

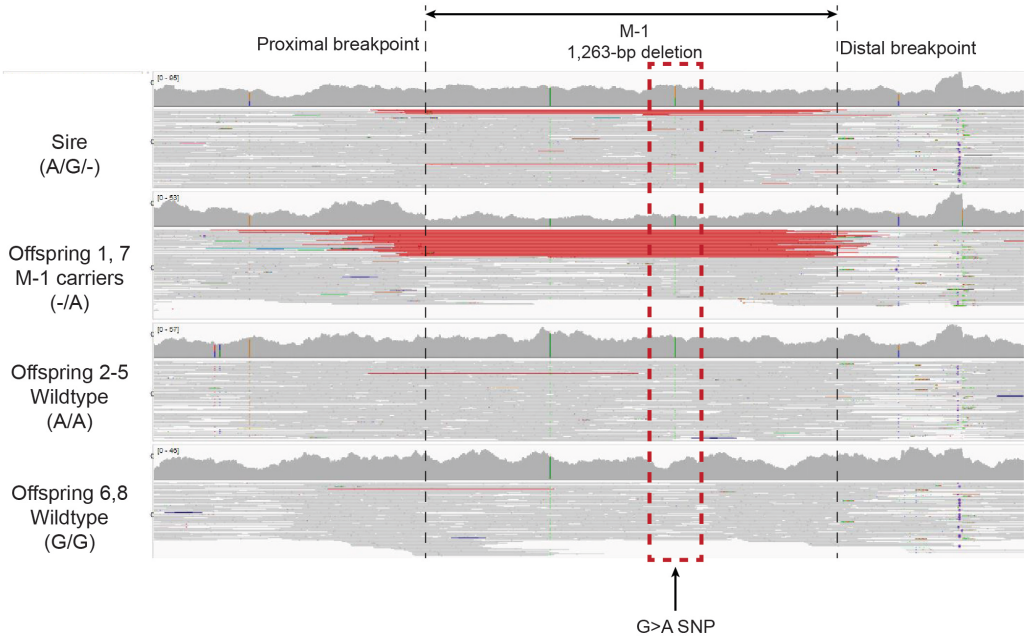
A. Depiction of M-3 in the three-generational pedigree. The legend is identical to supplementary figure 1A. **B.** Characterization of the breakpoints of M-3. The legend is identical to supplementary figure 1B. **C.** IGV (Robinson et al. 2011) screen capture of the M-3 in the current pedigree. Legend is identical to supplementary figure 1C. The discordant reads (red) aligned at the proximal and distal breakpoints support the presence of M-3.



Supplementary Figure S18. Pedigree, breakpoint, and the IGV screen capture of M-4
A. Depiction of M-4 in the three-generational pedigree. The legend is identical to supplementary figure 1A. **B.** Characterization of the structure of the M-4 allele. **C.** IGV (Robinson et al. 2011) screen capture of the M-4 in the current pedigree. Legend is identical to supplementary figure 1C. The discordant reads aligned at the start and the end of the 101-Kb deletion (red) and at the 8-kb duplication (green) support the presence of M-4.

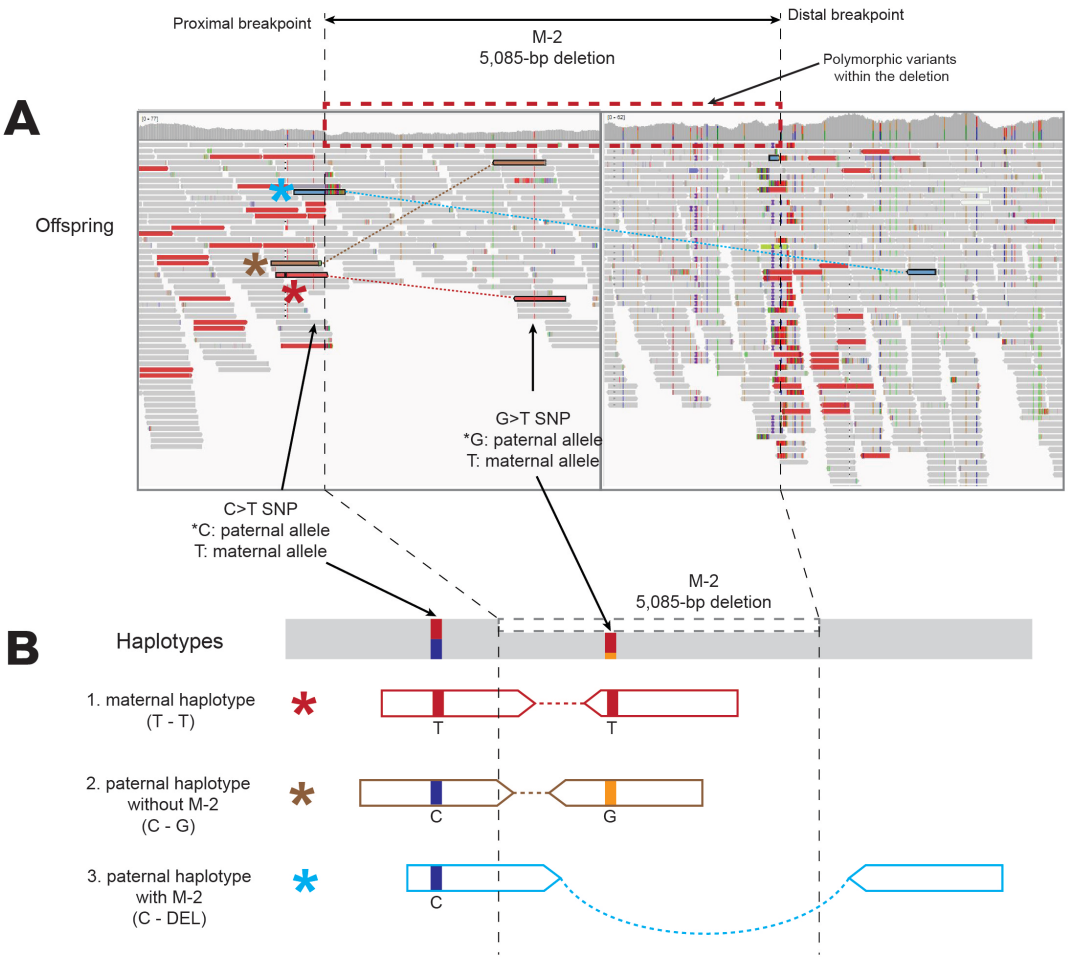


Supplementary Figure S19. Pedigree, breakpoint, and the IGV screen capture of M-5
A. Depiction of M-5 in the three-generational pedigree. The legend is identical to supplementary figure 1A. **B.** Characterization of the breakpoints of M-5. The legend is identical to supplementary figure 1B. **C.** IGV (Robinson et al. 2011) screen capture of the M-5 in the current pedigree. Legend is identical to supplementary figure 1C. The discordant reads (red) aligned at the proximal and distal breakpoints support the presence of M-5.



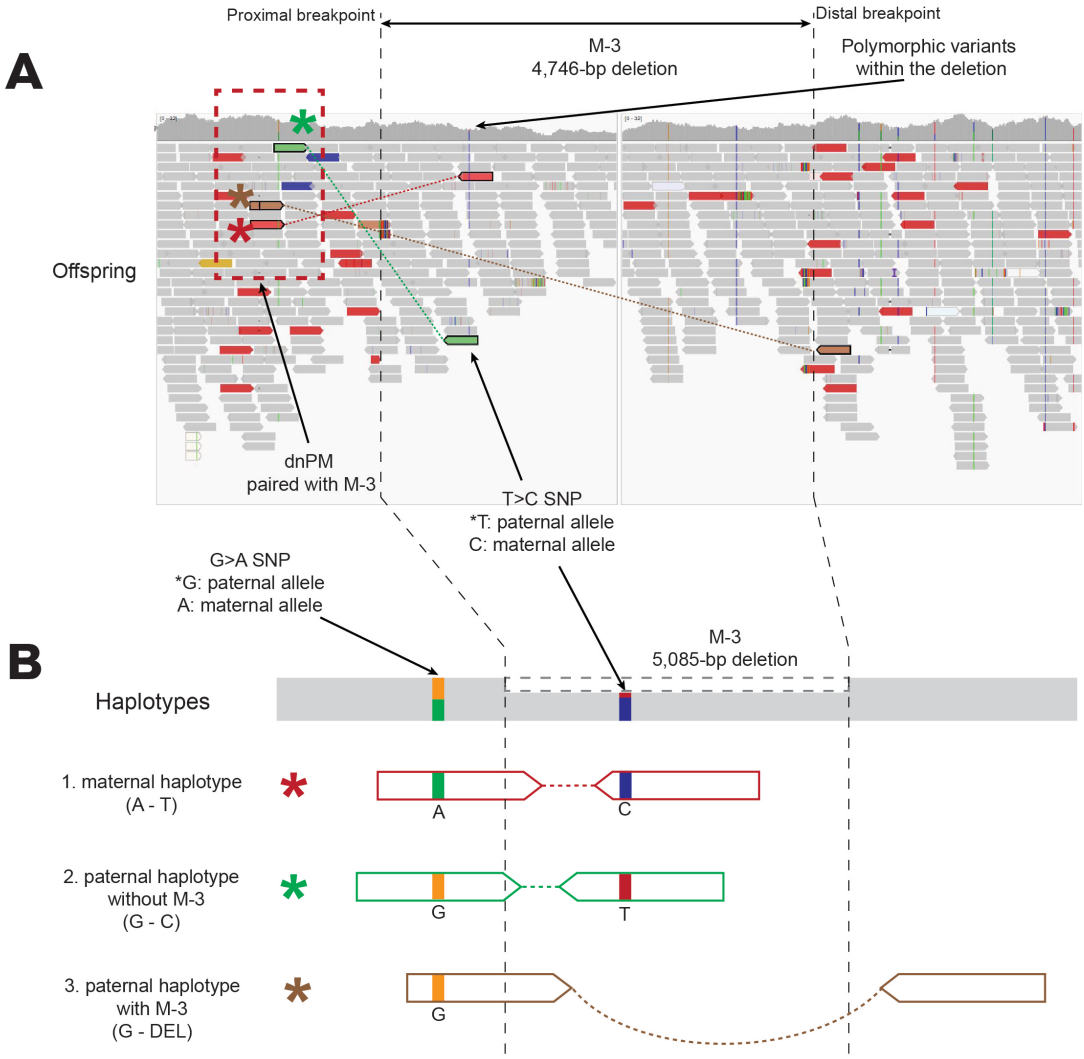
Supplementary Figure S20. Characterization of three haplotypes involving M-1

The presence of a polymorphic G>A SNP (marked with red dotted box) within M-1 in the sire supports the sire is mosaic (carrying A, G, and deletion alleles), and that M-1 is a post-zygotic DNM. The G>A SNP was phased, however was not assigned to either paternal or maternal haplotype of the sire, as there was no grandparents data available. For the eight offspring of the sire, two received the M-1 (Offspring 1 and 7), whereas the remaining six received wild-type allele (Offspring 2-6,8; no deletion). The G>A SNP genotypes of the six offspring were either A/A or G/G, hence confirming that the sire is a mosaic animal, carrying (i) wild-type haplotype (no deletion) with A allele, (ii) wild-type haplotype (no deletion) with G allele, and (iii) a deleted haplotype carrying M-1. By taking into account the long haplotype blocks shared in this pedigree, it was possible to determine that the M-1 occurred on the A allele carrying haplotype (see Supplementary Fig. 15A). The IGV screen capture in the current figure was made with “view as pair” option, unlike the one in Supplementary Fig. 15C, which was made a default mode.



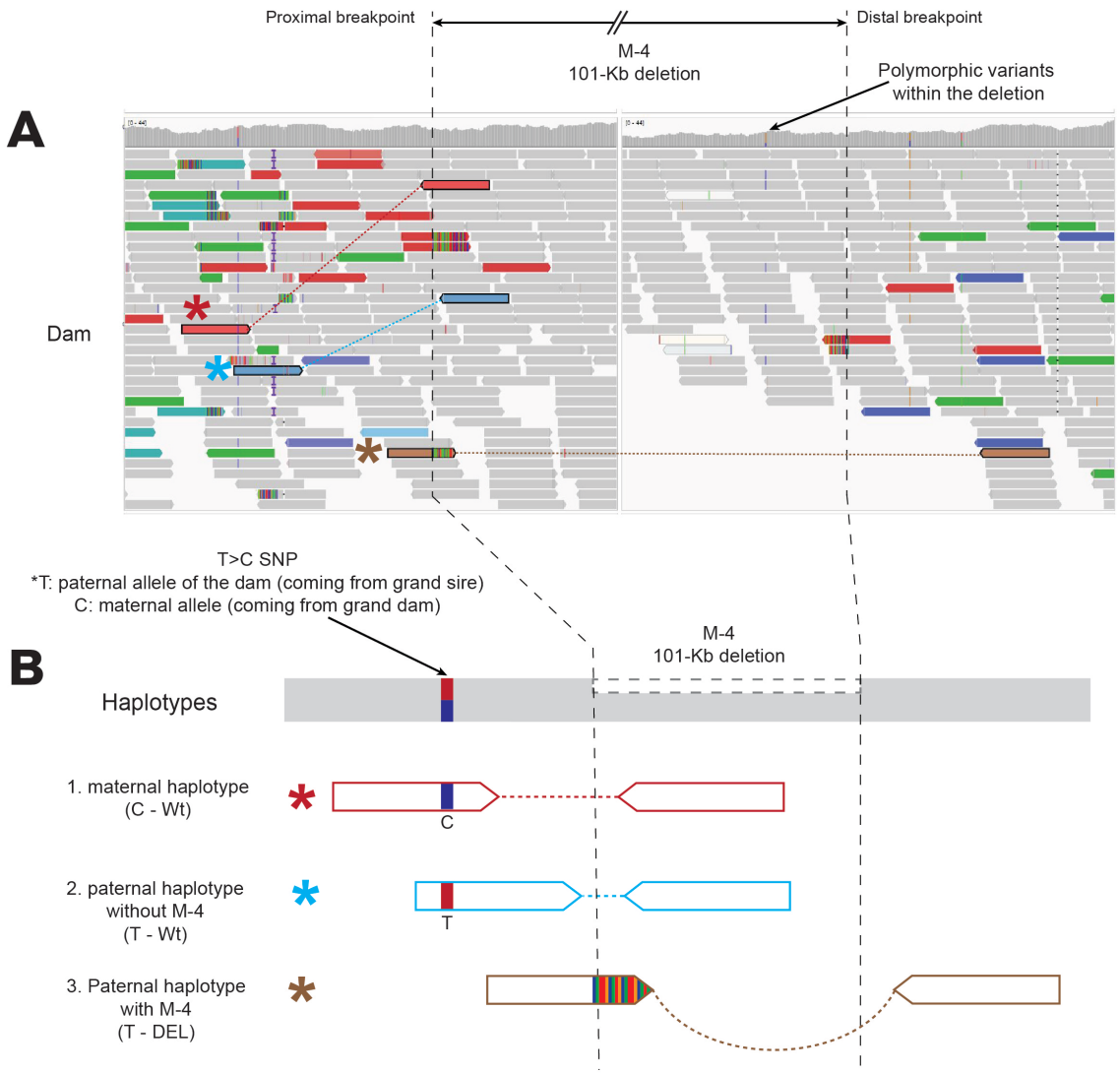
Supplementary Figure S21. Characterization of three haplotypes involving M-2

A. IGV screen capture showing the breakpoints of the M-2. To have sufficient resolution, the mid-part is not shown (indicated by a white vertical bar). Discordant reads supporting the M-2 are filled with red. Red dotted box shows the presence of polymorphic variants within the M-2, thus indicating (i) presence of three haplotypes and (ii) post-zygotic origin of M-2. Reconstruction of reads spanning over the M-2 unravels three haplotypes. These three reads are marked with blue, brown, and red asterisks. The dotted line connects the forward and reverse reads. The blue reads have much larger insert size than the average, supporting M-2. **B.** A schematic drawing of the three haplotypes shown in panel (e). The two variants (C>T SNP and G>T SNP), each located outside and inside of the M-2, respectively, were phased into a maternal haplotype (T-T; red asterisk) and a paternal haplotype (C-G; brown asterisk). Additional to these haplotypes, there is a paternal haplotype with a discordant read-pair, depicted with a curved dotted line (C-DEL; blue asterisk).



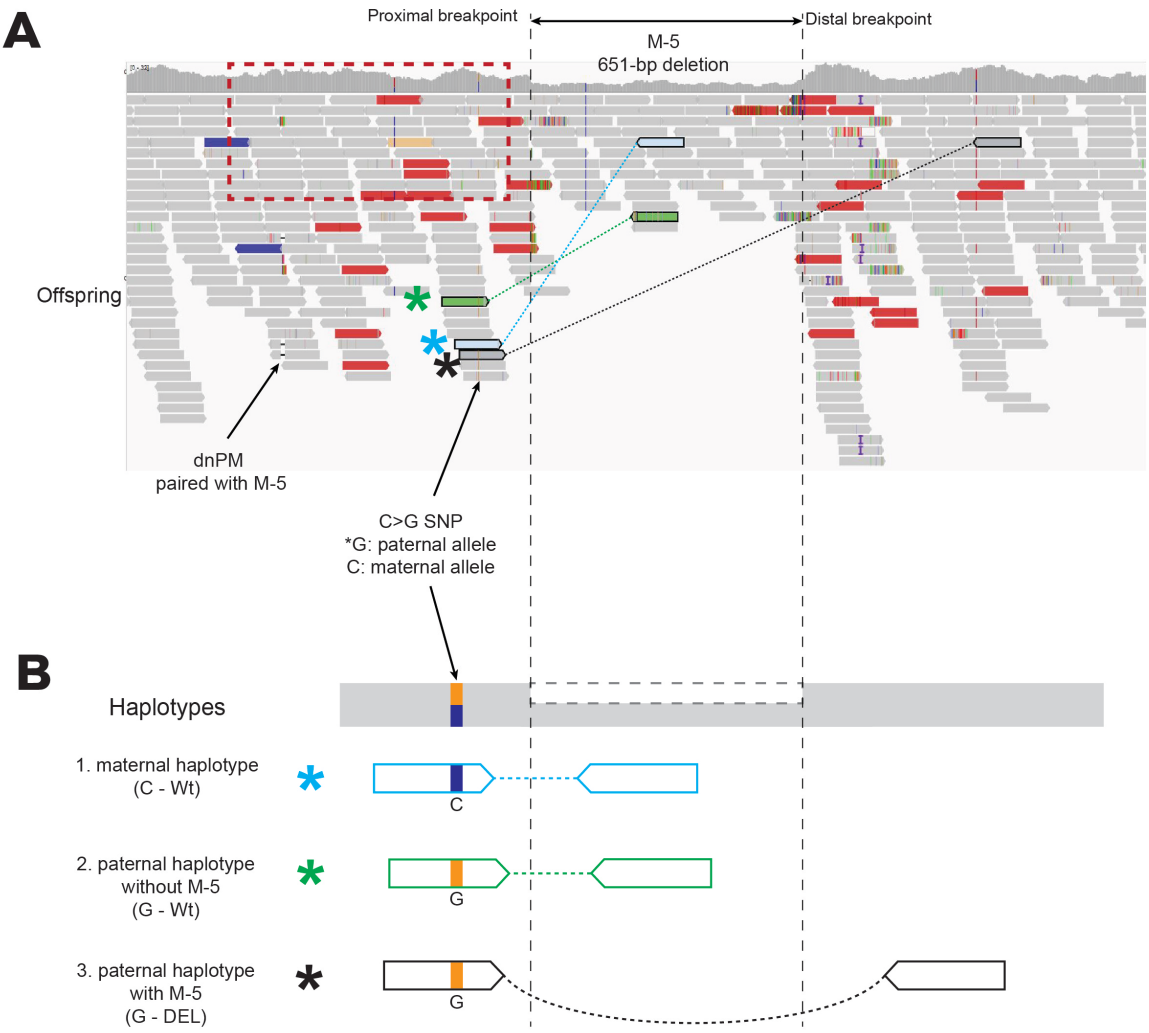
Supplementary Figure S22. Characterization of three haplotypes involving M-3

A. IGV screen capture showing the breakpoints of the M-3. To have sufficient resolution, the mid-part is not shown (indicated by a white vertical bar). Discordant reads supporting the M-3 are displayed as red. A polymorphic variant within the M-3 (only one is shown in the figure), is marked with an arrow. This indicates (i) presence of three haplotypes and (ii) post-zygotic origin of M-3. Reconstruction of reads spanning over the M-3 unravels three haplotypes involving M-3. These three reads are marked with green, brown, and red asterisks. The dotted line connects the forward and reverse reads. The brown reads have much larger insert size than the average, supporting M-3. The red dotted box shows the presence of dnPM (see Supplementary Fig. 29). **B.** A schematic drawing of the three haplotypes shown in panel (b). The two variants (G>A SNP and T>C SNP), each located outside and inside of the M-3, respectively, were phased into a maternal haplotype (A-C; red asterisk) and a paternal haplotype (G-T; green asterisk). Additional to these haplotypes, there is a paternal haplotype with a discordant read-pair, depicted with a curved dotted line (G-DEL; brown asterisk).



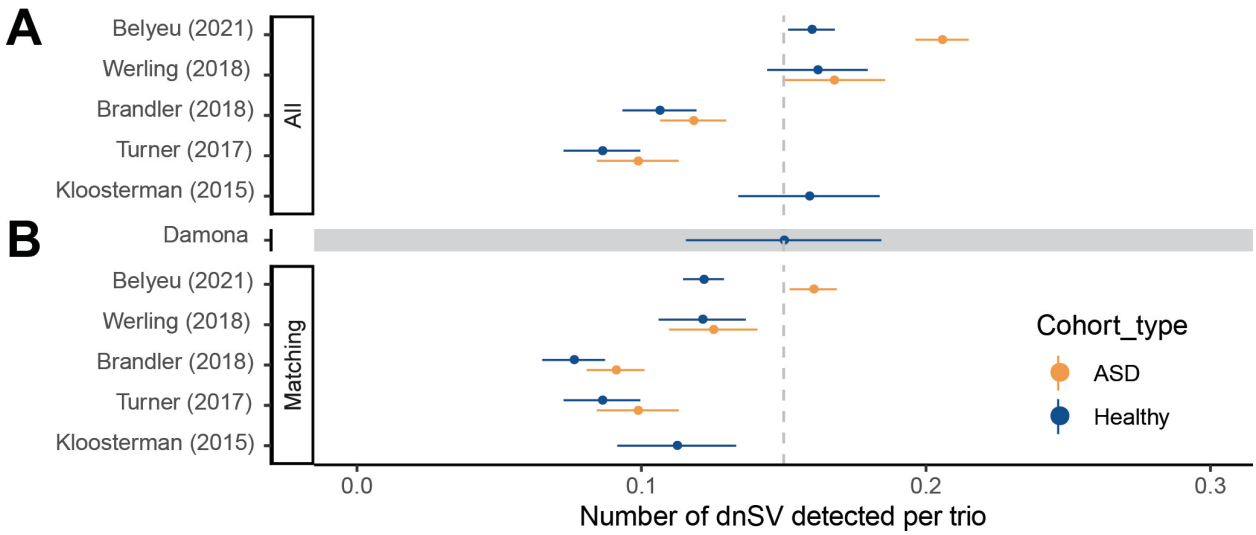
Supplementary Figure S23. Characterization of three haplotypes involving M-4

A. IGV screen capture showing the breakpoints of the 101-kb deletion of the M-4. To have sufficient resolution, the mid-part is not shown (indicated by a white vertical bar). Discordant reads supporting the M-4 are displayed as red. A polymorphic variant within the M-4 (only one is shown in the figure), is marked with an arrow. This indicates (i) presence of three haplotypes and (ii) post-zygotic origin of M-4. Reconstruction of reads spanning over the M-4 unravels three haplotypes involving M-4. These three reads are marked with red, blue, and brown asterisks. The dotted line connects the forward and reverse reads. The brown reads have much larger insert size than the average, supporting M-4. **B.** A schematic drawing of the three haplotypes shown in panel (e). The variant (T>C SNP), each located outside of the M-4 were phased into a maternal haplotype (C-Wt; red asterisk) and a paternal haplotype (T-Wt; blue asterisk). Additional to these haplotypes, there is a paternal haplotype with a discordant read-pair, depicted with a curved dotted line (T-DEL; brown asterisk).



Supplementary Figure S24. Characterization of three haplotypes involving M-5

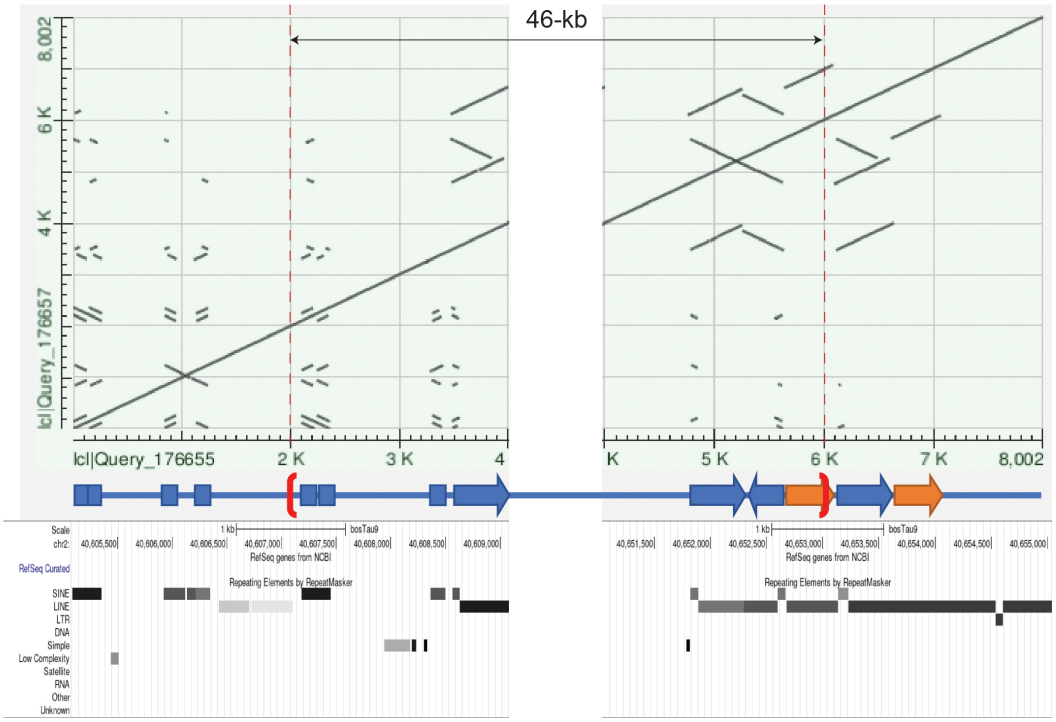
A. Reconstruction of reads spanning over the M-5 unravels three haplotypes involving M-5. These three reads are marked with green, blue, and black asterisks. The dotted line connects the forward and reverse reads. The black reads have much larger insert size than the average, supporting M-5. The red dotted box shows the presence of a dnPM (see Supplementary Fig. 30). **B.** A schematic drawing of the three haplotypes shown in panel (f). A variant (C>G SNP) located outside and of the M-5 was phased into a maternal haplotype (C-Wt; green asterisk) and a paternal haplotype (G-Wt; blue asterisk). Additional to these haplotypes, there is a paternal haplotype with a discordant read-pair, depicted with a curved dotted line (G-DEL; black asterisk).



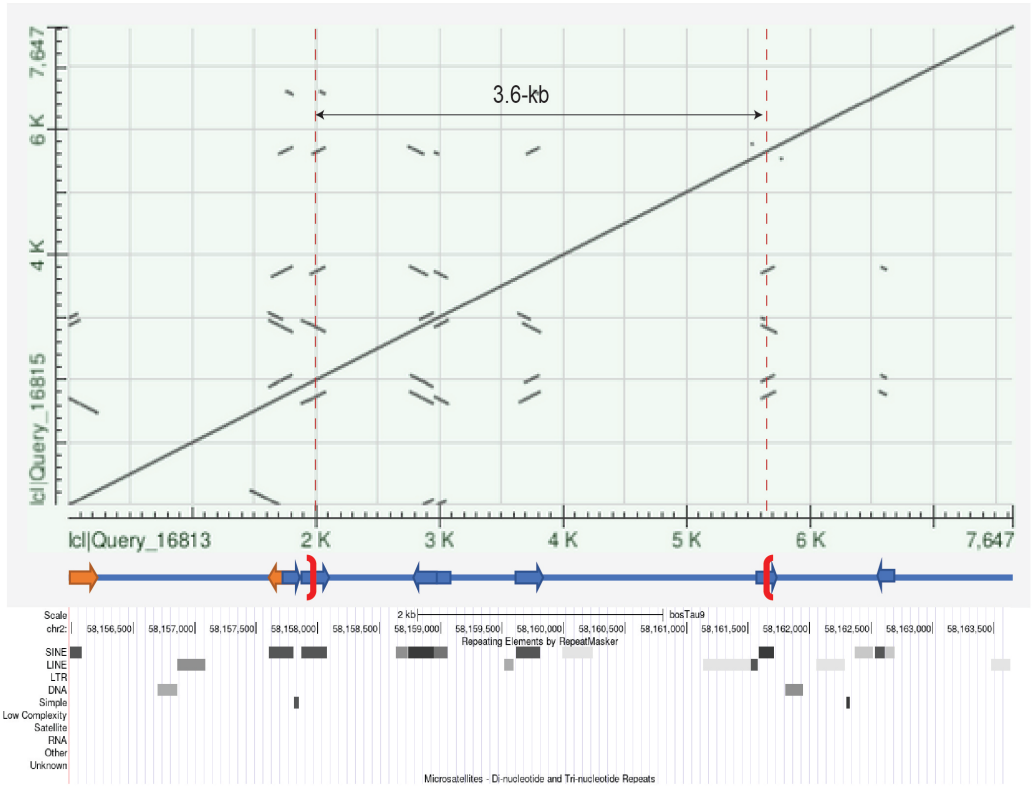
Supplementary Figure S25: Comparison of the number of dnSV detected per trio between the Damona pedigree and human studies.

The number of dnSV per trio in the Damona pedigree was compared with those from human studies, which used families with healthy offspring (blue: healthy) and families with offspring affected with autism spectrum disorder (ASD: orange). The horizontal bars correspond to the 95% confidence intervals of the estimates. **A.** ("All"), the numbers take diverse types of structural variants into account (e.g., mobile elements insertions). **B.** ("Matching") the numbers were recomputed using only deletions, duplications and inversion, the types of SV that were used in this study on the Damone pedigree.

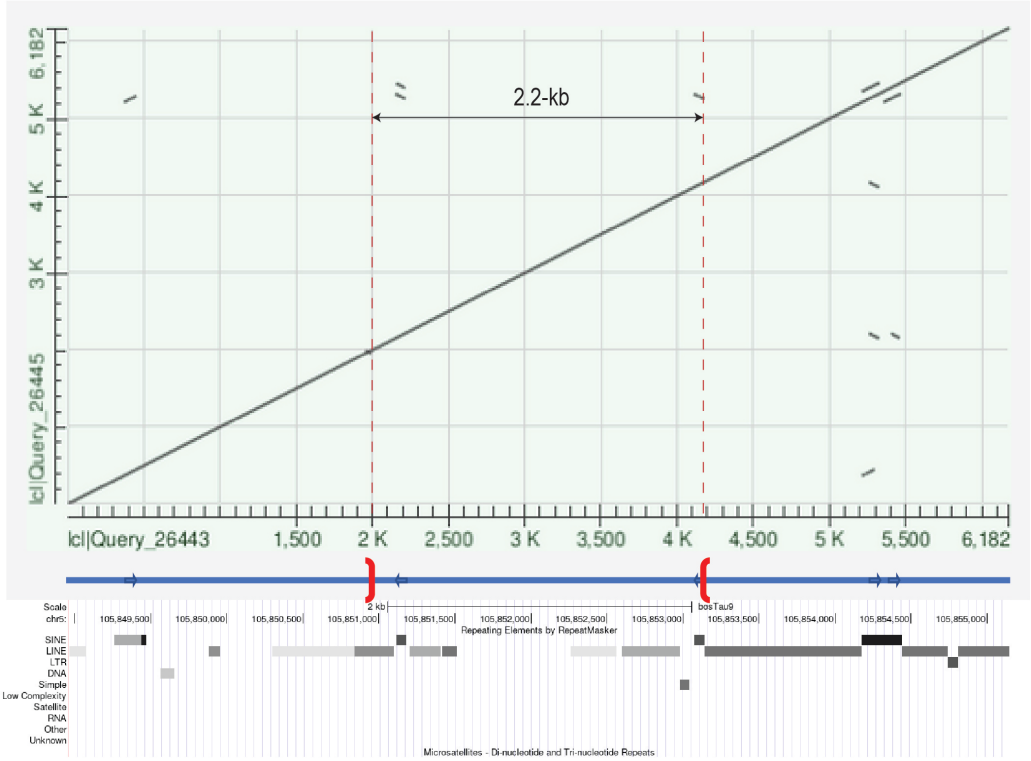
A (NM-1)


46-kb
lcl|Query_176655
2 K
3 K
4
5 K
6 K
7 K
8,002
RefSeq genes from NCBI
Repeating Elements by RepeatMasker
SINE
LINE
LTR
DNA
Simple
Low Complexity
Satellite
RNA
Other
Unknown

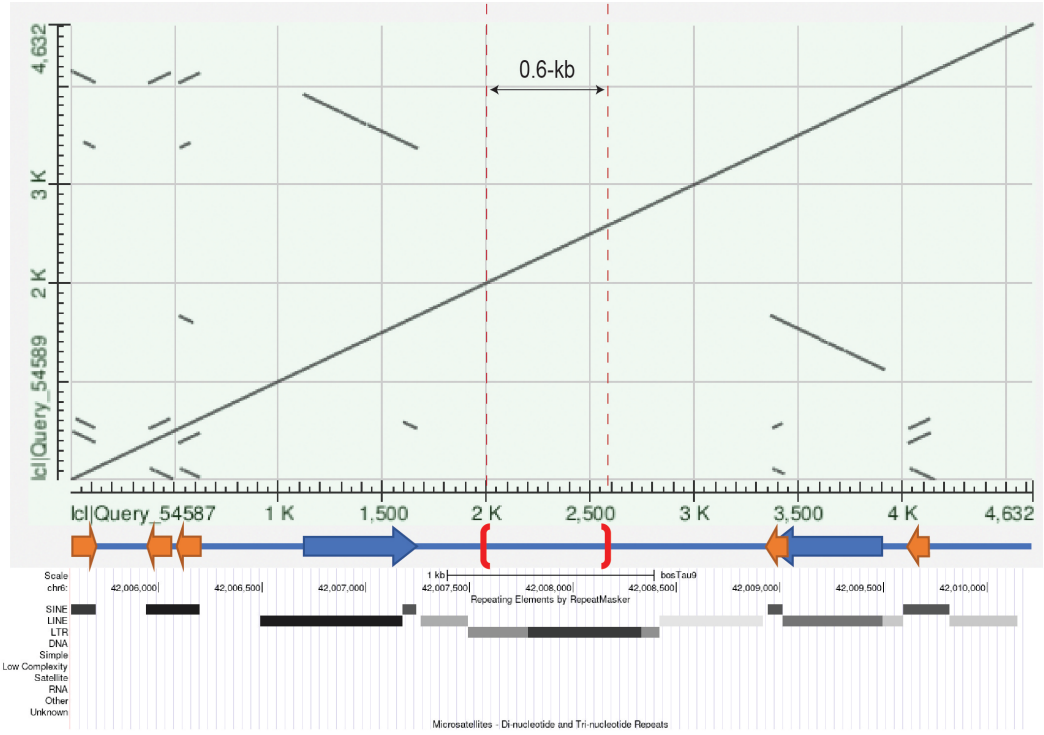
B (NM-2)


3.6-kb
lcl|Query_16813
2 K
3 K
4 K
5 K
6 K
7,647
Repeating Elements by RepeatMasker
SINE
LINE
LTR
DNA
Simple
Low Complexity
Satellite
RNA
Other
Unknown
Microsatellites - Di-nucleotide and Tri-nucleotide Repeats

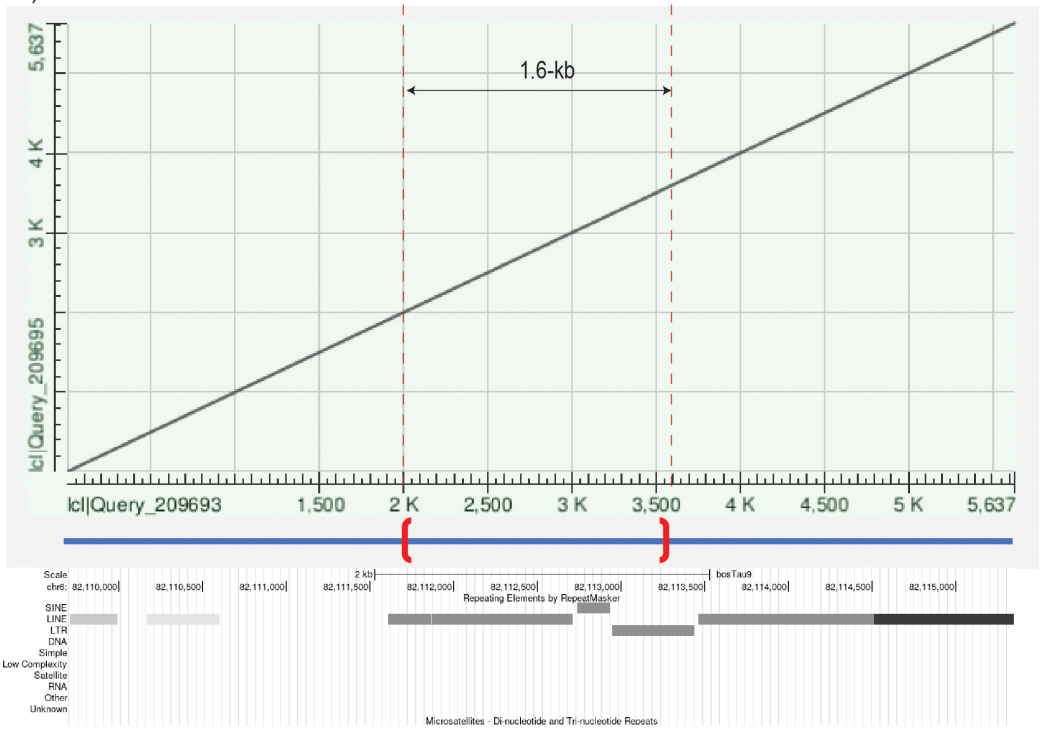
C (NM-3)


2.2-kb
lcl|Query_26445
lcl|Query_26443
1,500
2 K
2,500
3 K
3,500
4 K
4,500
5 K
5,500
6,182
Repeating Elements by RepeatMasker
SINE
LINE
LTR
DNA
Simple
Low Complexity
Satellite
RNA
Other
Unknown
Microsatellites - Di-nucleotide and Tri-nucleotide Repeats

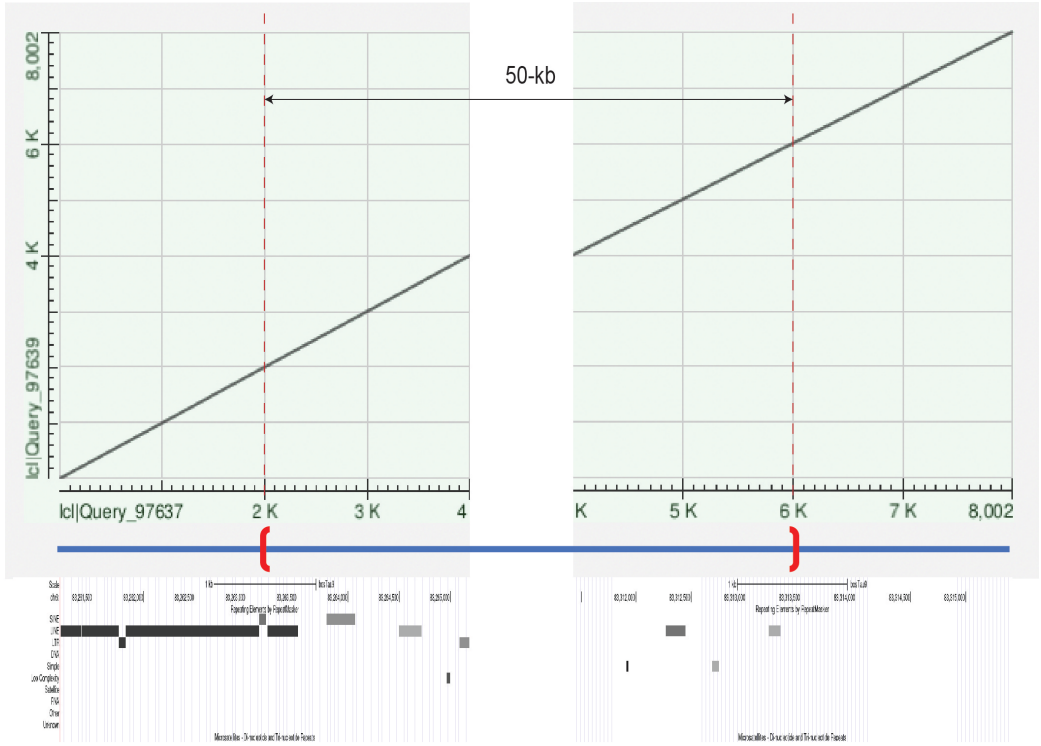
D (NM-4)


0.6-kb
lcl|Query_54589
lcl|Query_54587
1 K
1,500
2 K
2,500
3 K
3,500
4 K
4,632
Repeating Elements by RepeatMasker
SINE
LINE
LTR
DNA
Simple
Low Complexity
Satellite
RNA
Other
Unknown
Microsatellites - Di-nucleotide and Tri-nucleotide Repeats

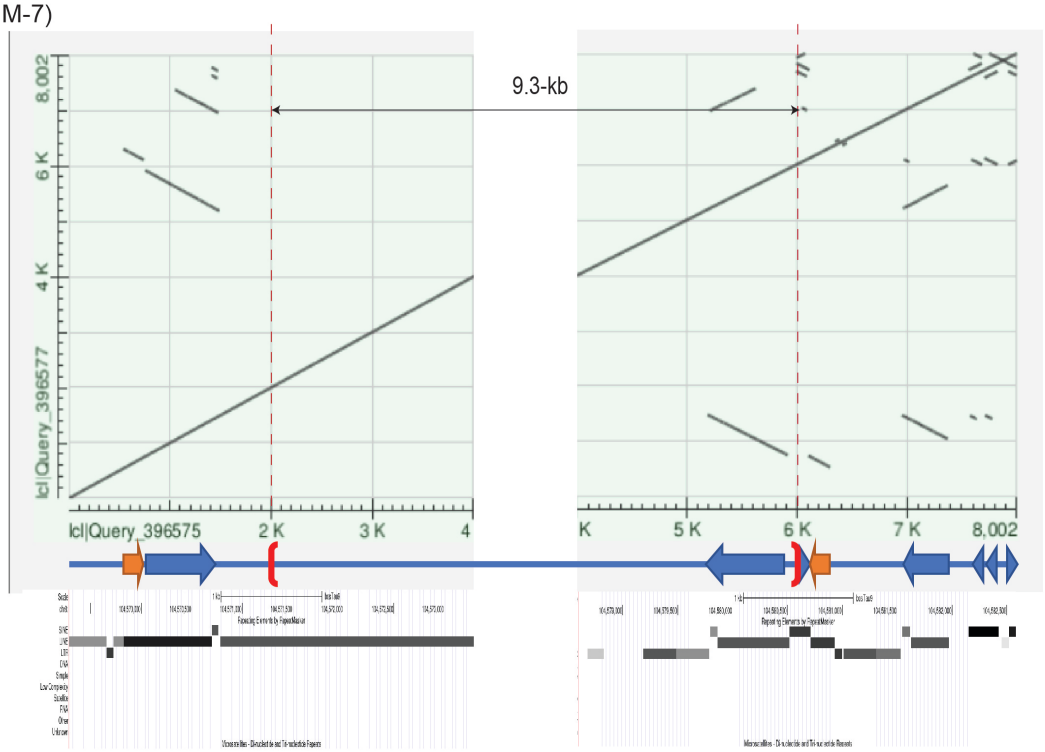
E (NM-5)


1.6-kb
lcl|Query_209695
5,637
4 K
3 K
lcl|Query_209693
1,500
2 K
2,500
3 K
3,500
4 K
4,500
5 K
5,637
Scale
chr6:
2 kb
bosTau9
Repeating Elements by RepeatMasker
SINE
LINE
LTR
DNA
Simple
Low Complexity
Satellite
RNA
Other
Unknown
Microsatellites - Di-nucleotide and Tri-nucleotide Repeats

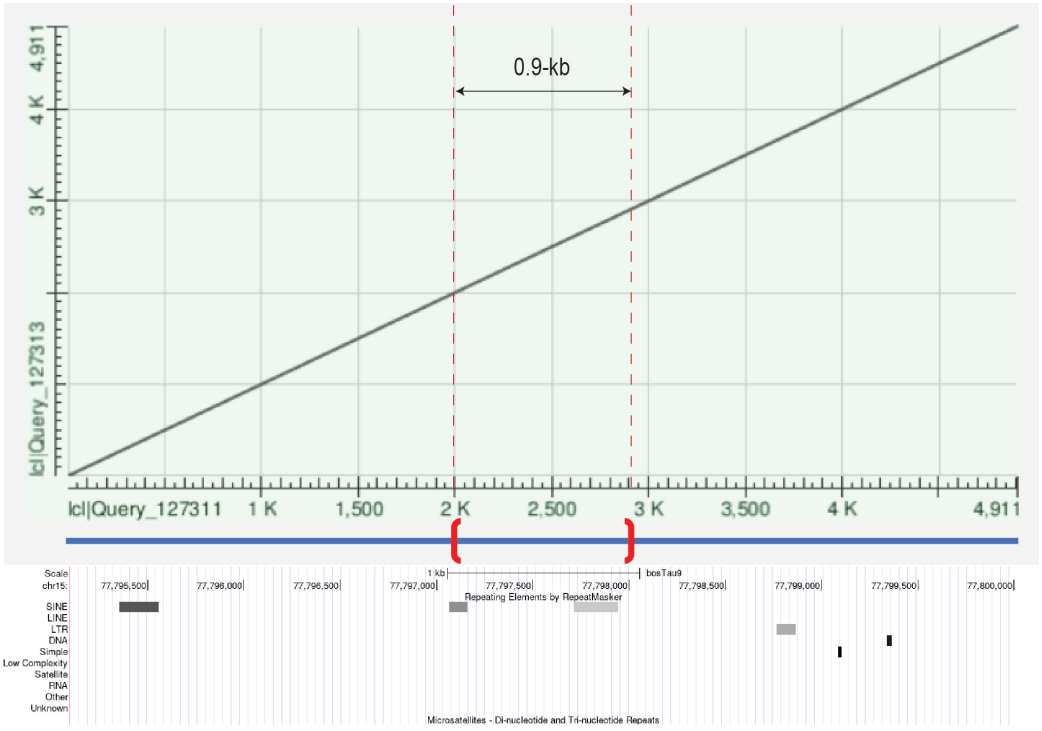
F (NM-6)


50-kb
lcl|Query_97639
8,002
6 K
4 K
lcl|Query_97637
2 K
3 K
4
K
5 K
6 K
7 K
8,002

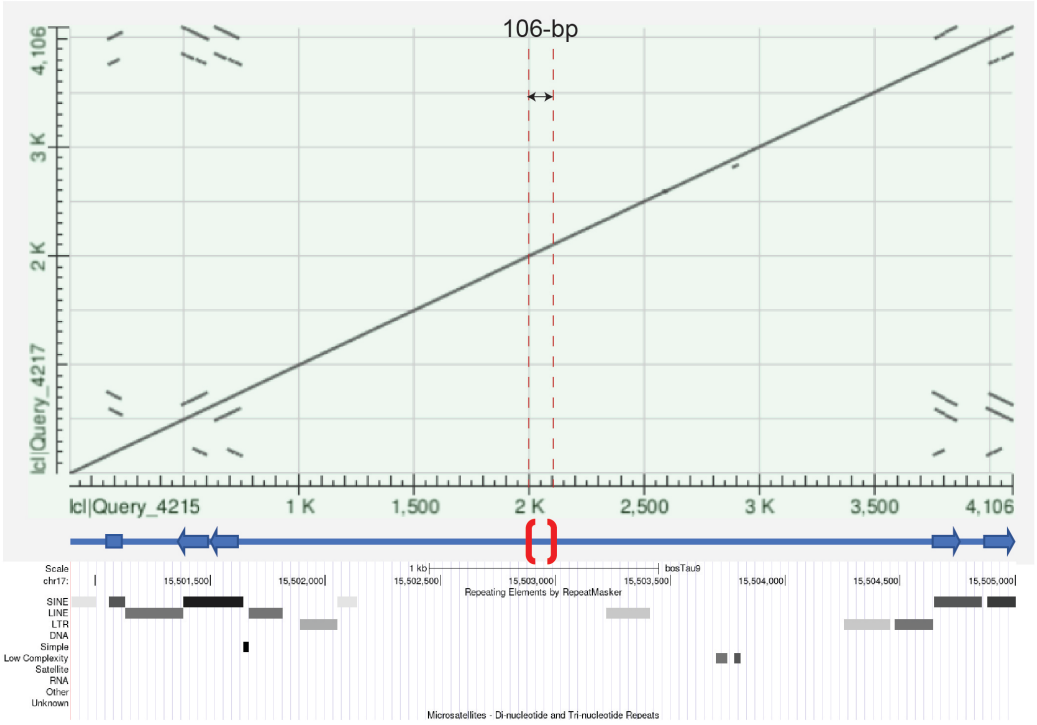
G (NM-7)


9.3-kb
lcl|Query_396577
8,002
6 K
4 K
lcl|Query_396575
2 K
3 K
4
K
5 K
6 K
7 K
8,002

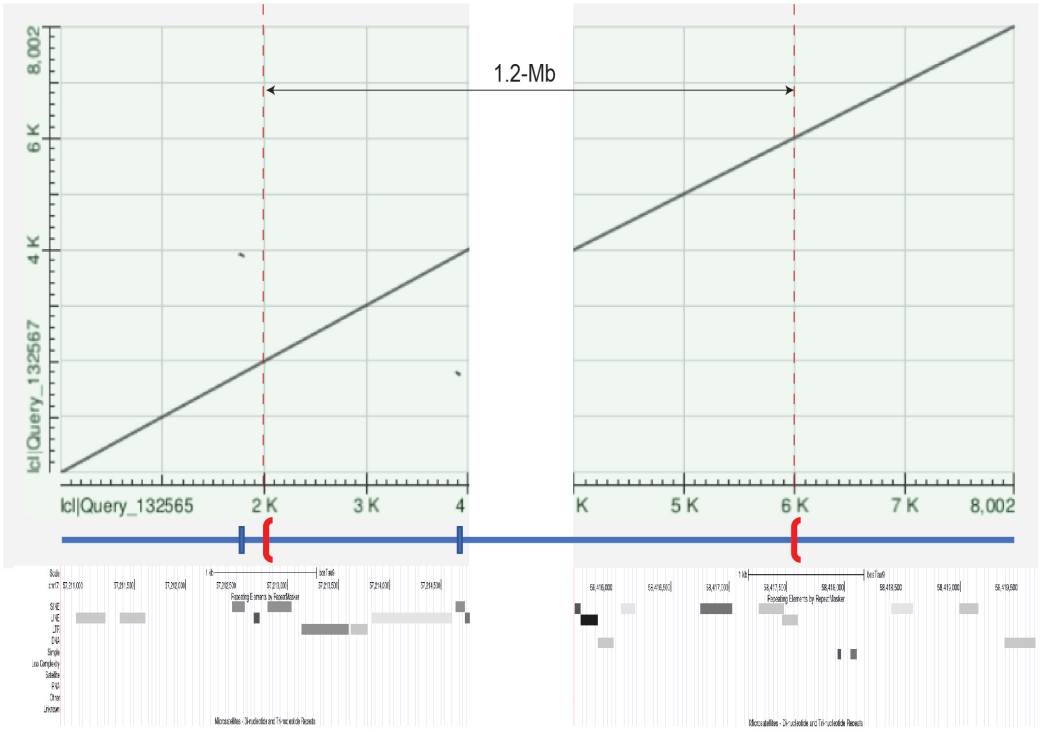
H (NM-8)


0.9-kb
lcl|Query_127313
4,911
4 K
3 K
lcl|Query_127311
1 K
1,500
2 K
2,500
3 K
3,500
4 K
4,911
Scale
chr15:
77,795,500
77,796,000
77,796,500
77,797,000
77,797,500
77,798,000
77,798,500
77,799,000
77,799,500
77,800,000
1 kb
bosTau9
Repeating Elements by RepeatMasker
SINE
LINE
LTR
DNA
Simple
Low Complexity
Satellite
RNA
Other
Unknown
Microsatellites - Di-nucleotide and Tri-nucleotide Repeats

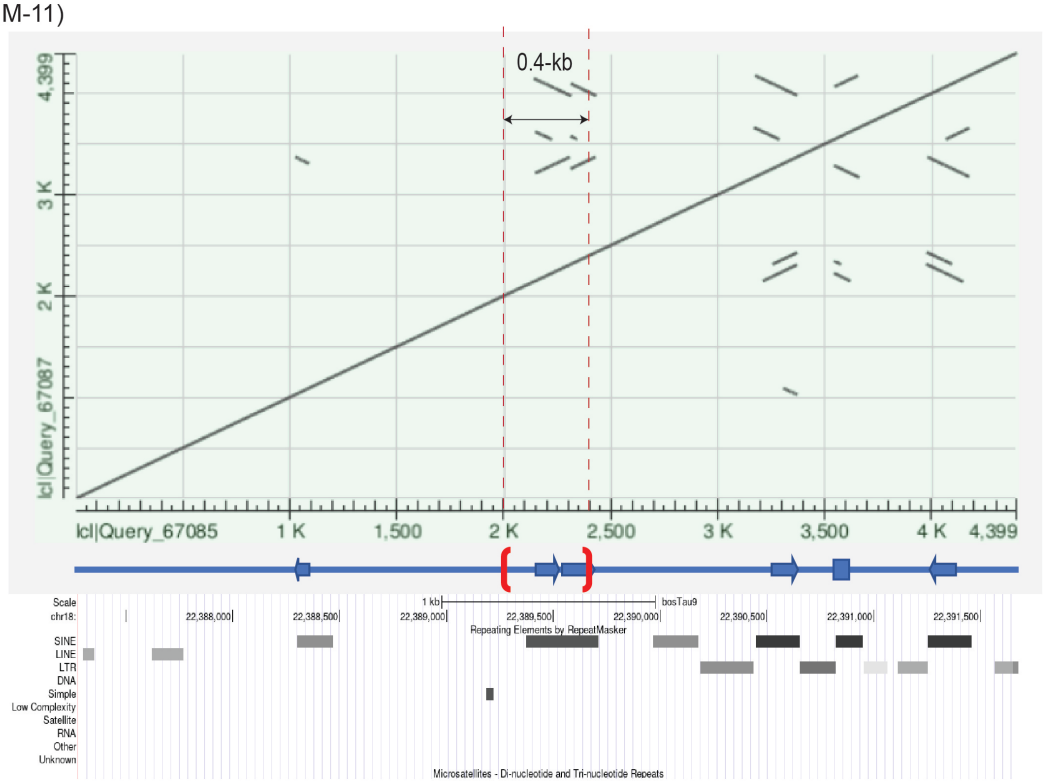
I (NM-9)


106-bp
lcl|Query_4217
4,106
3 K
2 K
lcl|Query_4215
1 K
1,500
2 K
2,500
3 K
3,500
4,106
Scale
chr17:
15,501,500
15,502,000
15,502,500
15,503,000
15,503,500
15,504,000
15,504,500
15,505,000
Repeating Elements by RepeatMasker
SINE
LINE
LTR
DNA
Simple
Low Complexity
Satellite
RNA
Other
Unknown
Microsatellites - Di-nucleotide and Tri-nucleotide Repeats

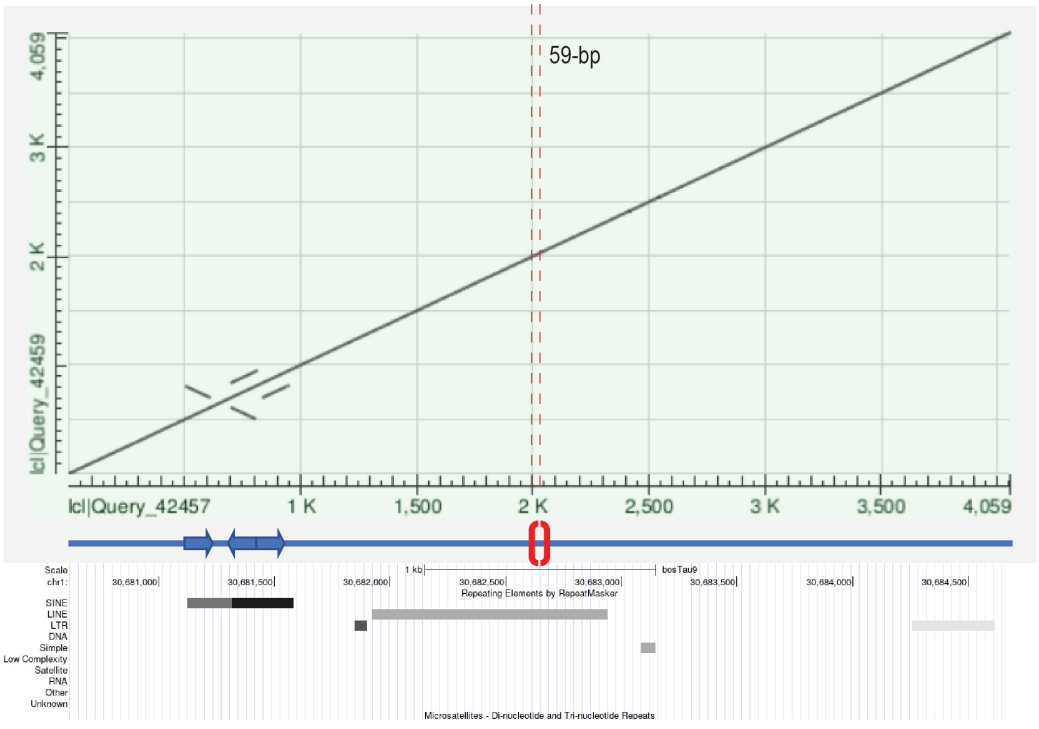
J (NM-10)


1.2-Mb
lcl|Query_132567
8,002
6 K
4 K
lcl|Query_132565
2 K
3 K
4
K
5 K
6 K
7 K
8,002

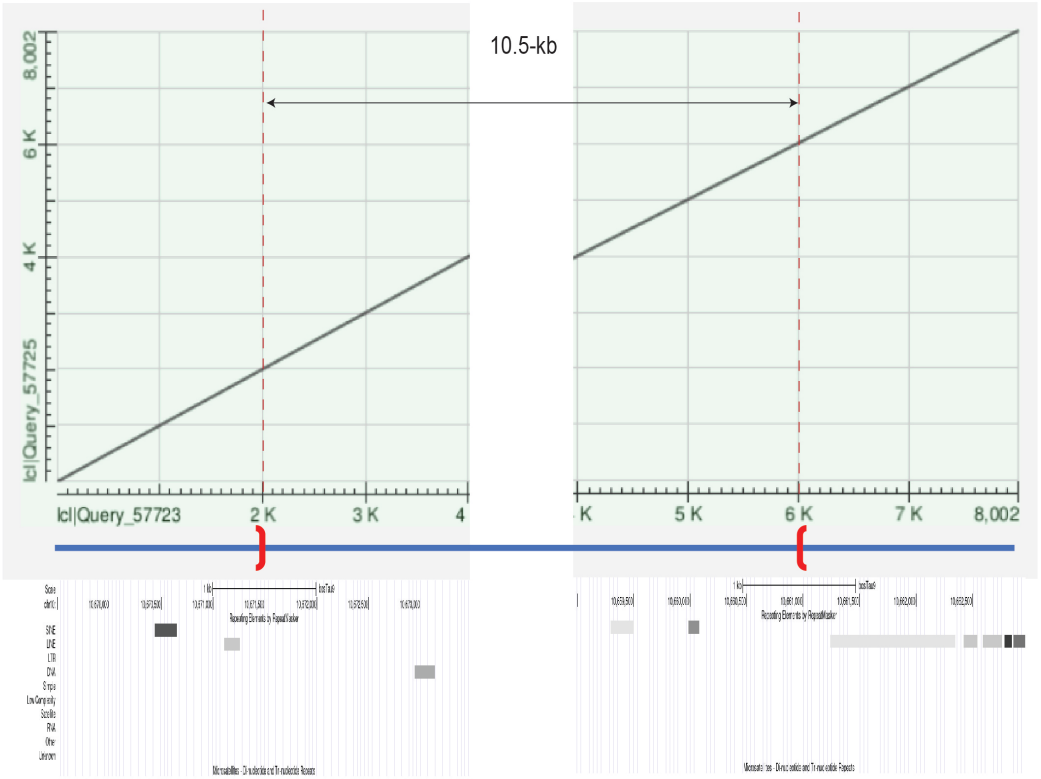
K (NM-11)


0.4-kb
lcl|Query_67087
4,399
3 K
2 K
1,500
2,500
3,500
4 K
Scale
chr18:
22,388,000
22,388,500
22,389,000
22,389,500
22,390,000
22,390,500
22,391,000
22,391,500
1 kb
bosTau9
Repeating Elements by RepeatMasker
SINE
LINE
LTR
DNA
Simple
Low Complexity
Satellite
RNA
Other
Unknown
Microsatellites - Di-nucleotide and Tri-nucleotide Repeats

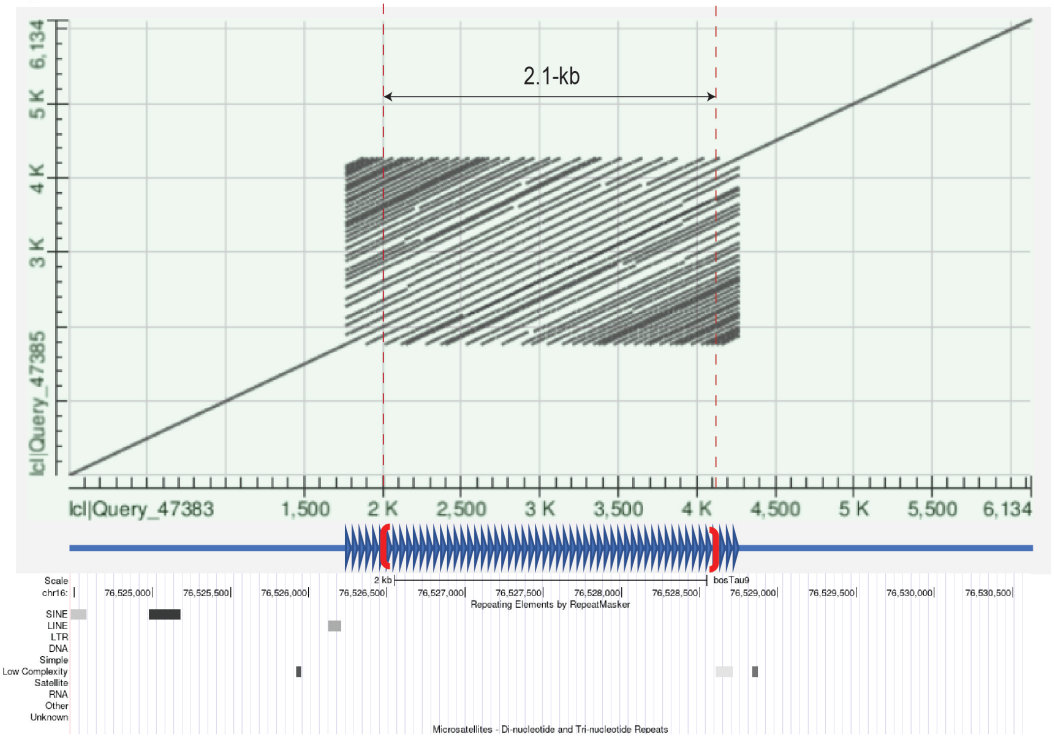
L (A-1)


59-bp
lcl|Query_42459
lcl|Query_42457
4,059
3 K
2 K
1 K
1,500
2,500
3,500
Scale
chr1:
30,681,000
30,681,500
30,682,000
30,682,500
30,683,000
30,683,500
30,684,000
30,684,500
1 kb
bosTau9
Repeating Elements by RepeatMasker
SINE
LINE
LTR
DNA
Simple
Low Complexity
Satellite
RNA
Other
Unknown
Microsatellites - Di-nucleotide and Tri-nucleotide Repeats

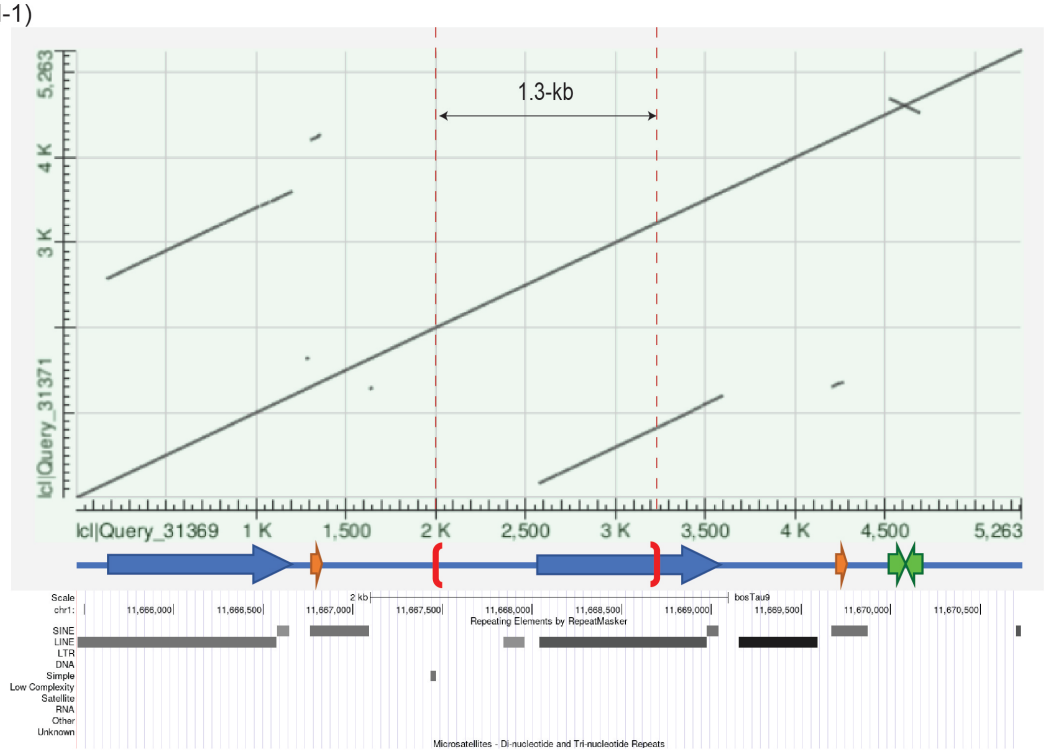
M (A-2)


10.5-kb
lcl|Query_57725
8,002
6 K
4 K
lcl|Query_57723
2 K
3 K
4
K
5 K
6 K
7 K
8,002
Repeating Elements by RepeatMasker
Microsatellites - Di-nucleotide and Tri-nucleotide Repeats

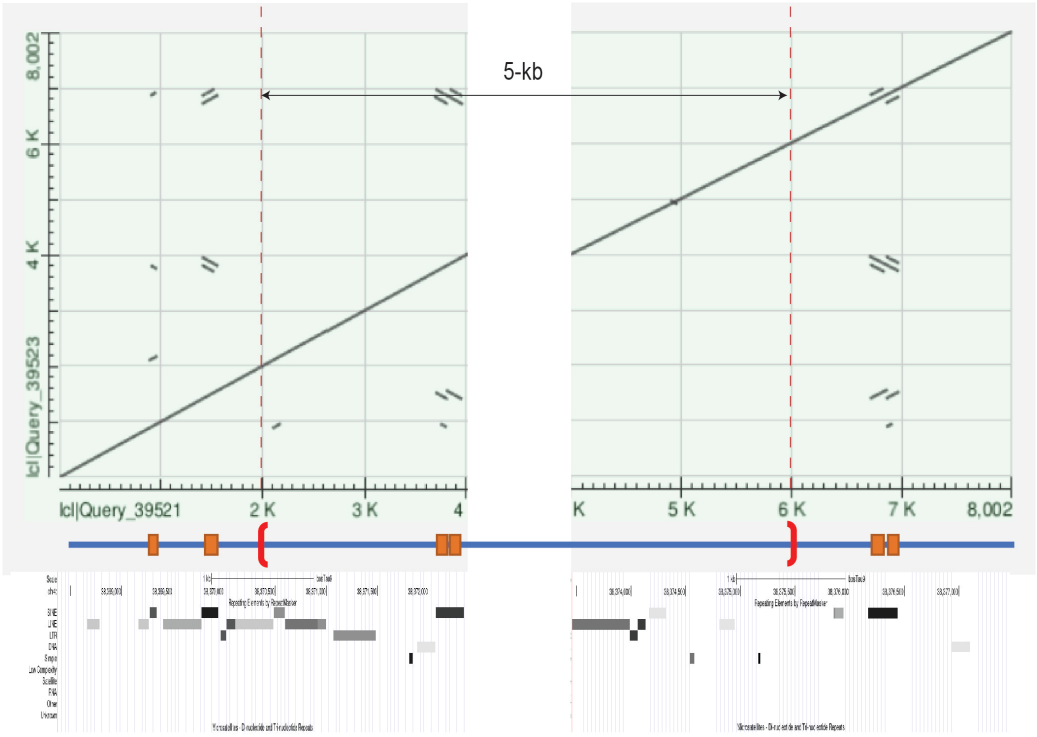
N (A-3)


2.1-kb
lcl|Query_47385
6,134
5 K
4 K
3 K
lcl|Query_47383
1,500
2 K
2,500
3 K
3,500
4 K
4,500
5 K
5,500
6,134
Scale
chr16:
76,525,000
76,525,500
76,526,000
76,526,500
76,527,000
76,527,500
76,528,000
76,528,500
76,529,000
76,529,500
76,530,000
76,530,500
Repeating Elements by RepeatMasker
SINE
LINE
LTR
DNA
Simple
Low Complexity
Satellite
RNA
Other
Unknown
Microsatellites - Di-nucleotide and Tri-nucleotide Repeats

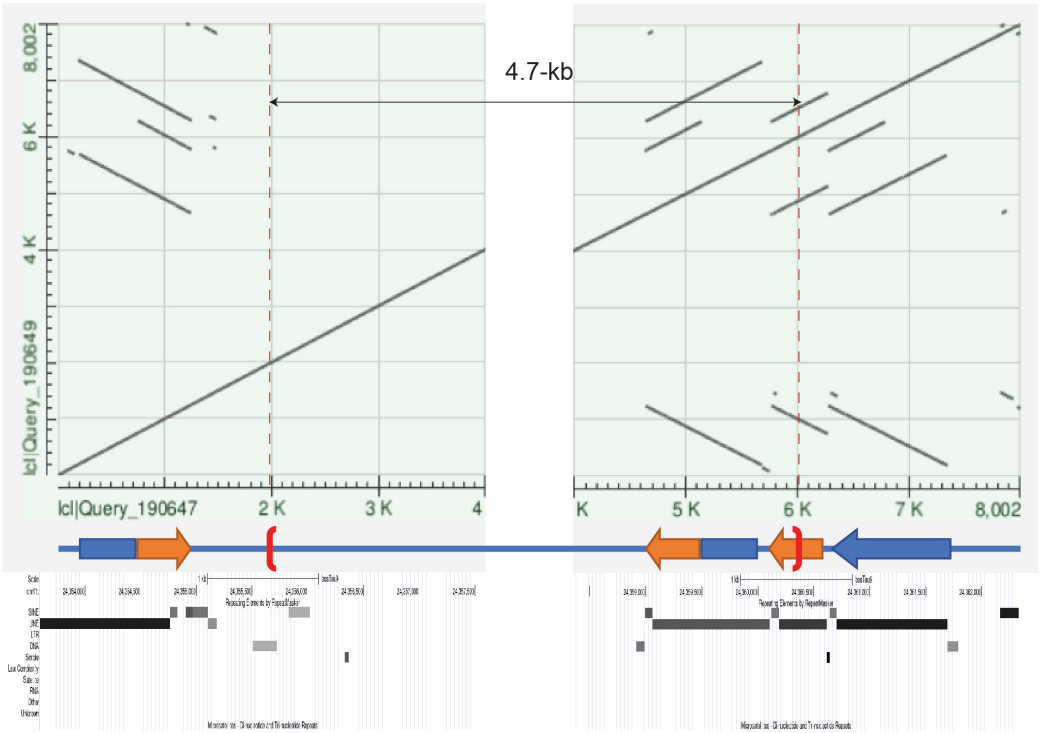
O (M-1)


1.3-kb
lcl|Query_31371
5,263
4 K
3 K
lcl|Query_31369
1 K
1,500
2 K
2,500
3 K
3,500
4 K
4,500
5,263
Scale
chr1:
11,666,000
11,666,500
11,667,000
11,667,500
11,668,000
11,668,500
11,669,000
11,669,500
11,670,000
11,670,500
2 kb
bosTau9
Repeating Elements by RepeatMasker
SINE
LINE
LTR
DNA
Simple
Low Complexity
Satellite
RNA
Other
Unknown
Microsatellites - Di-nucleotide and Tri-nucleotide Repeats

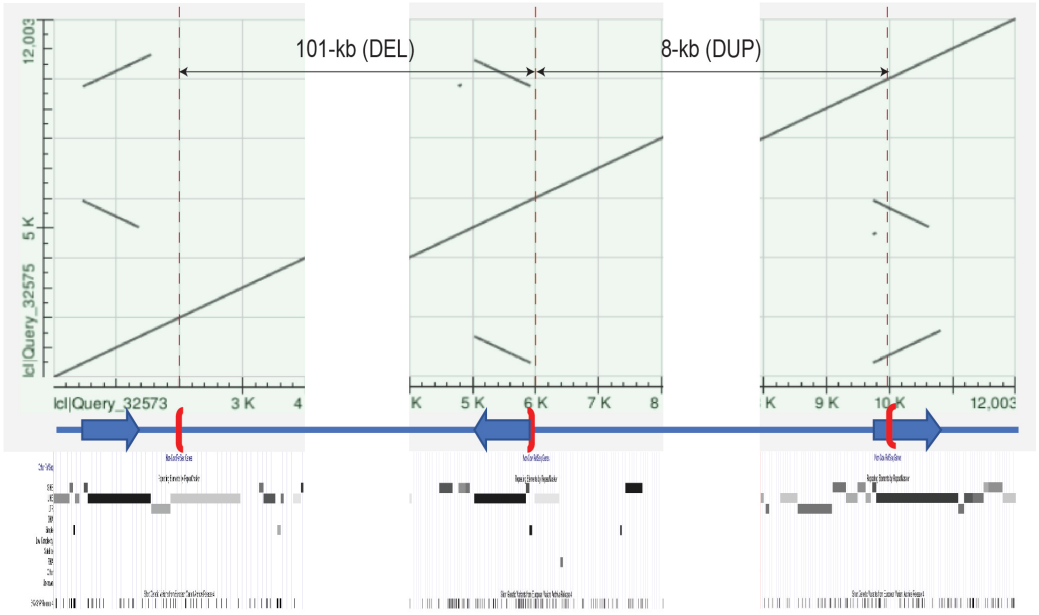
P (M-2)


5-kb
lcl|Query_39523
8,002
6 K
4 K
lcl|Query_39521
2 K
3 K
4
K
5 K
6 K
7 K
8,002

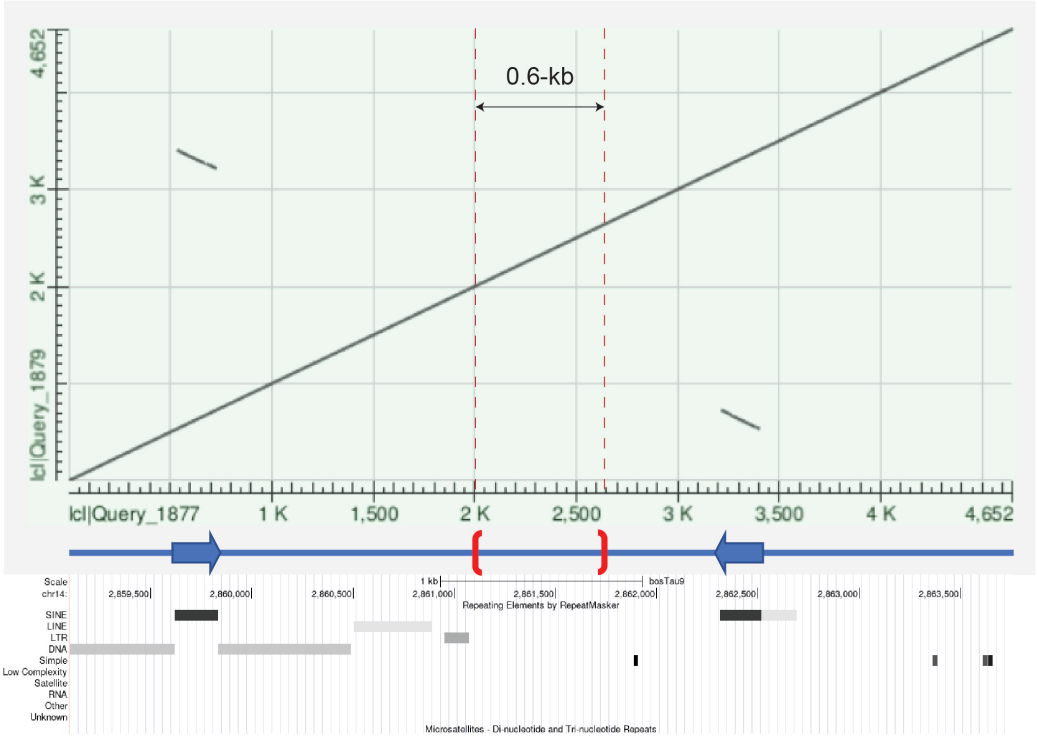
Q (M-3)


4.7-kb
lcl|Query_190649
8,002
6 K
4 K
lcl|Query_190647
2 K
3 K
4
5 K
6 K
7 K
8,002

R (M-4)


101-kb (DEL)
8-kb (DUP)
lcl|Query_32575
12,003
5 K
lcl|Query_32573
3 K
4
5 K
6 K
7 K
8
9 K
10 K
12,003

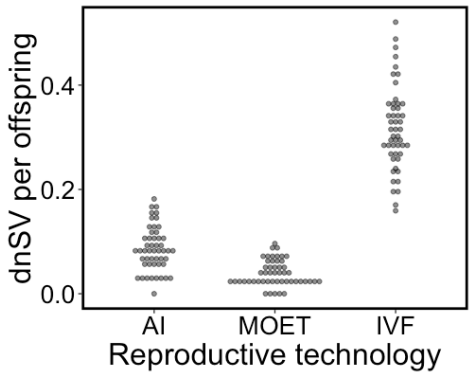
S (M-5)



Supplementary Figure S26. Characterization of repeat sequences spanning the breakends of the 19 dnSV

We evaluated potential dnSV formation mechanisms, with an aim to identify ones formed via non-allelic homologous recombination (NAHR), by investigating repeat sequences surrounding the breakends. Each panel consists of a dot plot, a schematic drawing marking breakends of dnSV and repeats (arrows) and known bovine repeats in the corresponding region. Dot plots were obtained from BLAST (blastn; blast.ncbi.nlm.nih.gov) by aligning two identical sequences that span over breakends (2-kb before and after the proximal and distal breakends, respectively). The red inward facing brackets indicate breakends of deletions, the red outward facing brackets indicate breakends of duplications, and the right-side facing brackets indicate breakends of inversions. Red dotted vertical lines indicate breakends.

A

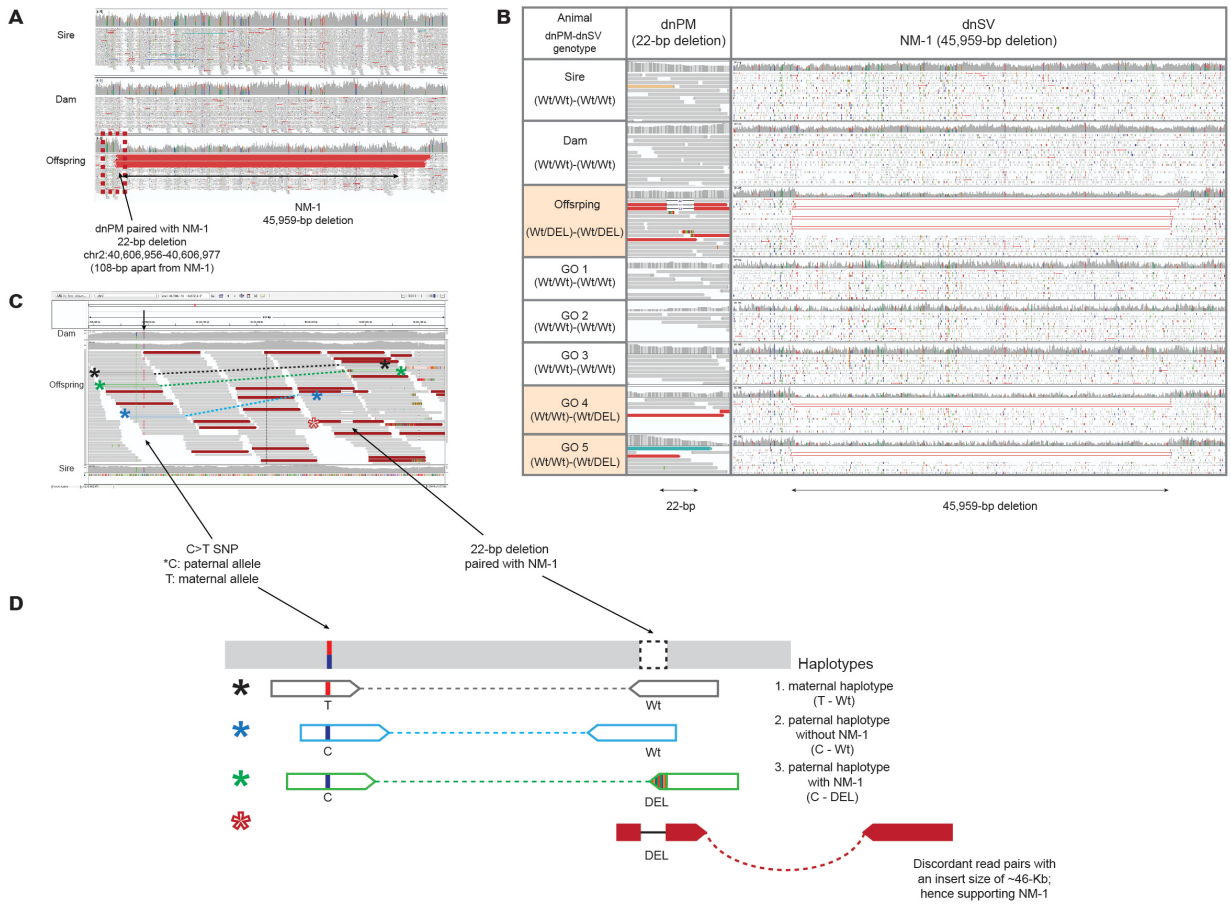


B

```
Call:
lm(formula = Nunber.of.dnSV.detect ~ Sample.Type.Prob + Repro.Prob +
    Age.Sire.in.years + Age.Dam.in.years + Sequ.Depth.Prob +
    Sequ.Depth.Sire + Sequ.Depth.Dam, data = Damona)

Residuals:
     Min       1Q   Median       3Q      Max
-0.40973 -0.23842 -0.06011  0.01672  1.68408

Coefficients:
                           Estimate Std. Error t value Pr(>|t|)
(Intercept)               -0.362429   0.360963  -1.004   0.3174
Sample.Type.ProbSperm     -0.025460   0.100668  -0.253   0.8008
Repro.ProbFlushed embryo  -0.045857   0.088849  -0.516   0.6067
Repro.ProbIVF              0.218360   0.088979   2.454   0.0156 *
Age.Sire.in.years         -0.002786   0.020118  -0.138   0.8901
Age.Dam.in.years          -0.022603   0.033280  -0.679   0.4983
Sequ.Depth.Prob           -0.009394   0.007267  -1.293   0.1987
Sequ.Depth.Sire            0.011707   0.005276   2.219   0.0284 *
Sequ.Depth.Dam             0.018425   0.010319   1.786   0.0767 .
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.381 on 118 degrees of freedom
Multiple R-squared:  0.1503,    Adjusted R-squared:  0.09266
F-statistic: 2.608 on 8 and 118 DF,  p-value: 0.01156
```

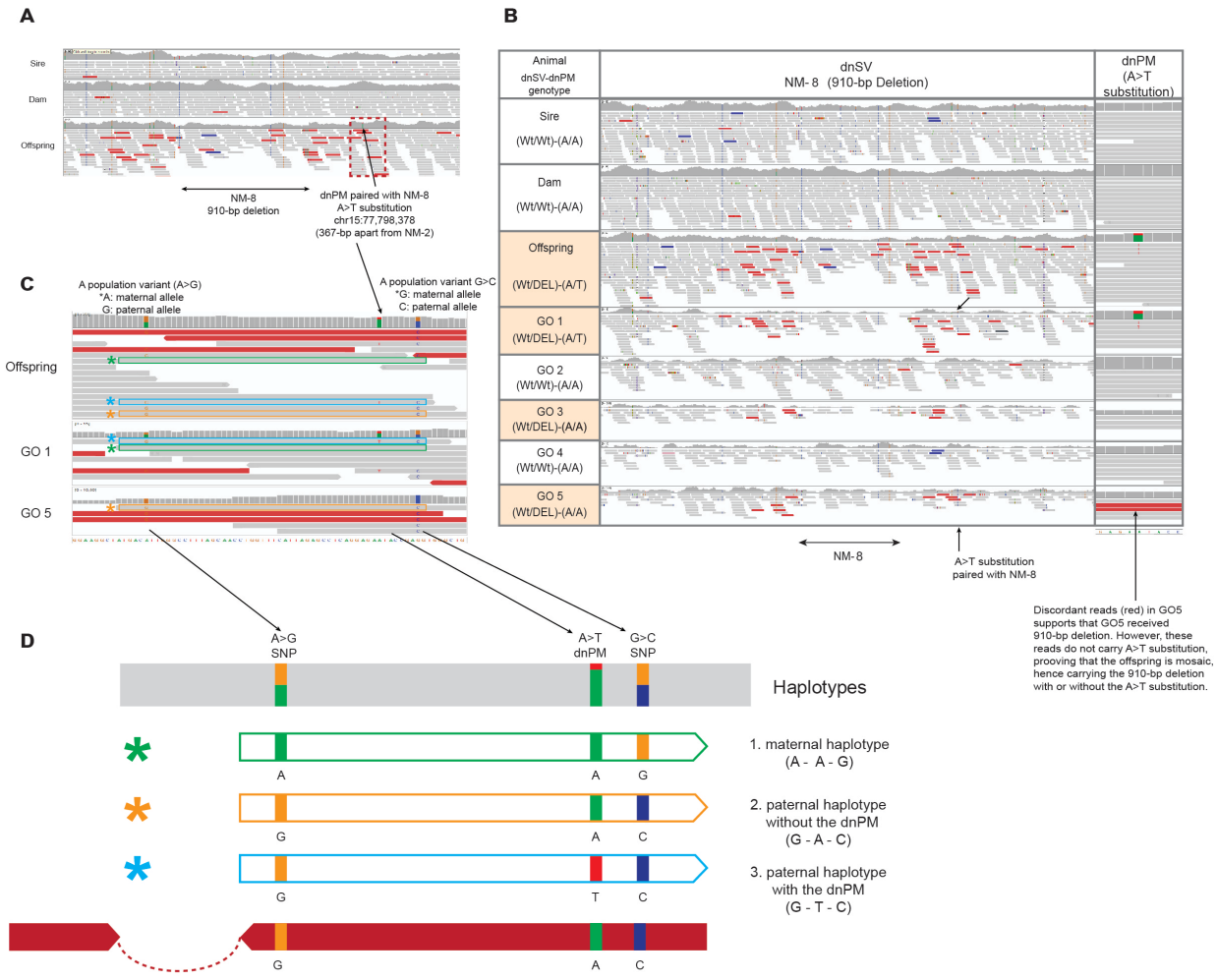
Supplementary Figure S27. Effect of ART and sire’s sequence depth on dnSV rate

A. Confidence intervals for the number of dnSV detected on average per offspring obtained by AI, MOET or IVF were estimated by bootstrapping. We sampled 127 “offspring” with replacement from the complete list (Supplemental Table 2) of Damona probands one hundred times and computed for each bootstrapped sample the number of dnSV detected per animal by ART category. The distributions of the obtained numbers are shown, showing the marked effect of IVF. **B.** Testing the effect of sample type of the proband (sperm or blood), reproductive technique used to produce the proband (AI, MOET, or IVF), age of the sire at reproduction (in years), age of the dam at reproduction (in years), sequence depth of offspring, sire, and dam on the number of dnSV detected using a linear model implemented with the lm() function in R. The output of the analysis is shown.



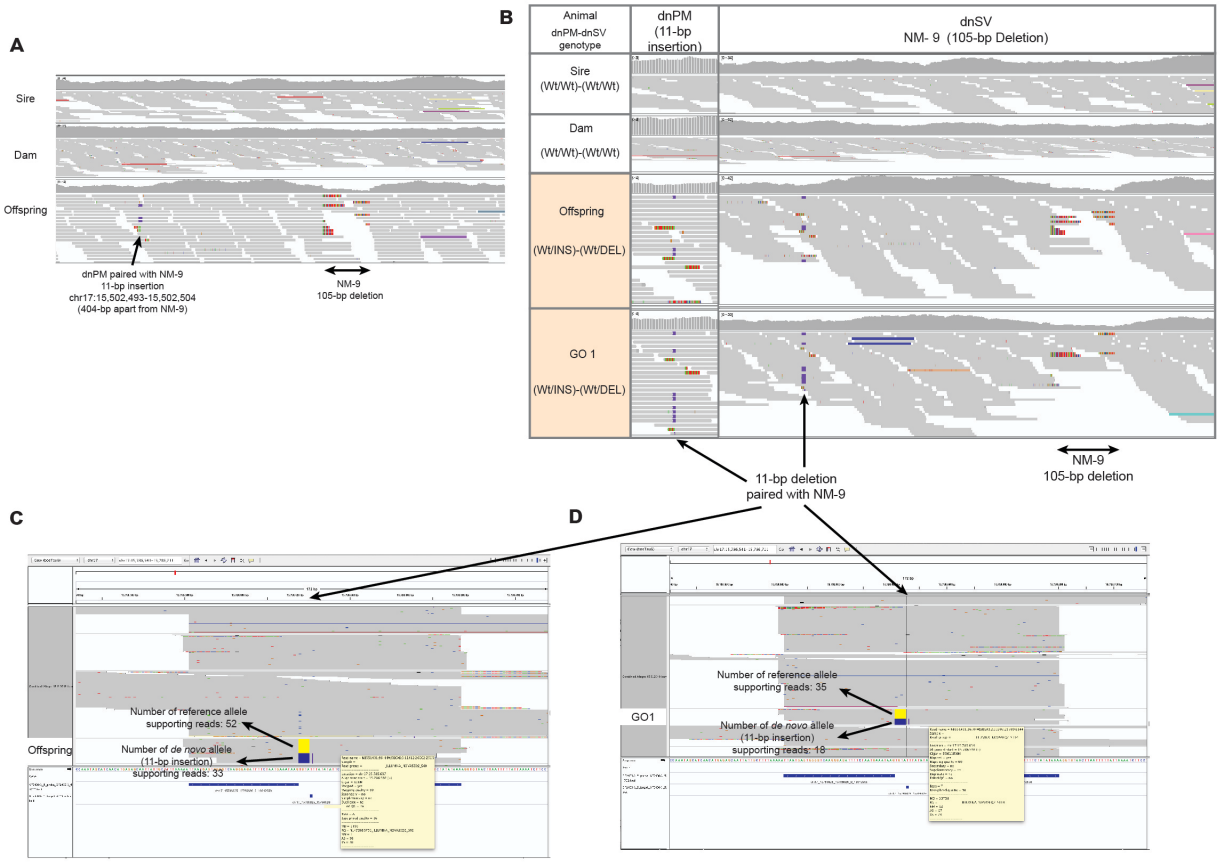
Supplementary Figure S28. Characterization of the cluster mutation involving NM-1

A. An overview of the cluster DNM consisting of NM-1 and 22-bp deletion. The red horizontal stripes shown in the offspring mark the discordant reads that span over the breakpoints of NM-1. The 22-bp deletion is located 108-bp from the proximal breakpoint of NM-1. The region encompassing the 22-bp deletion (marked with the red dotted box) is zoomed in the panel C. **B**. Genotypes of the cluster DNM in the current pedigree. Wt stands for wild-type and DEL stands for deletion. The offspring carries both NM-1 and the paired 22-bp deletion, however GO 4 and GO 5 of which NM-1 was transmitted did not receive the 22-bp deletion. Animals carrying NM-1 are marked with translucent orange. Pictures on the right side shows the IGV screen capture at the 22-bp deletion and the NM-1. **C.** IGV screen capture at the 22-bp deletion. Reconstruction of reads spanning over the 22-bp deletion and C>T SNP at the vicinity unravels three haplotypes. These three read pairs are marked with black, green, and blue asterisks. A read marked with red asterisk is a forward read that carries the 22-bp deletion and its read pair is mapped at the distal breakpoint of the NM-1, thus the read pair together supports both the 22-bp deletion and NM-1. **D.** A schematic drawing of the three haplotypes shown in panel (c). The two variants present in this region (C>T SNP and 22-bp deletion) can be phased into a maternal haplotype carrying T-Wt (black asterisk) and a paternal haplotype carrying C-Wt (blue asterisk). Additional to these haplotypes, there is a paternal haplotype with the 22-bp deletion (C-DEL; green asterisk), indicating post-zygotic origin of the 22-bp deletion. The discordant reads marked with red show that there is haplotype supporting both NM-1 and the 22-bp deletion concurrently.



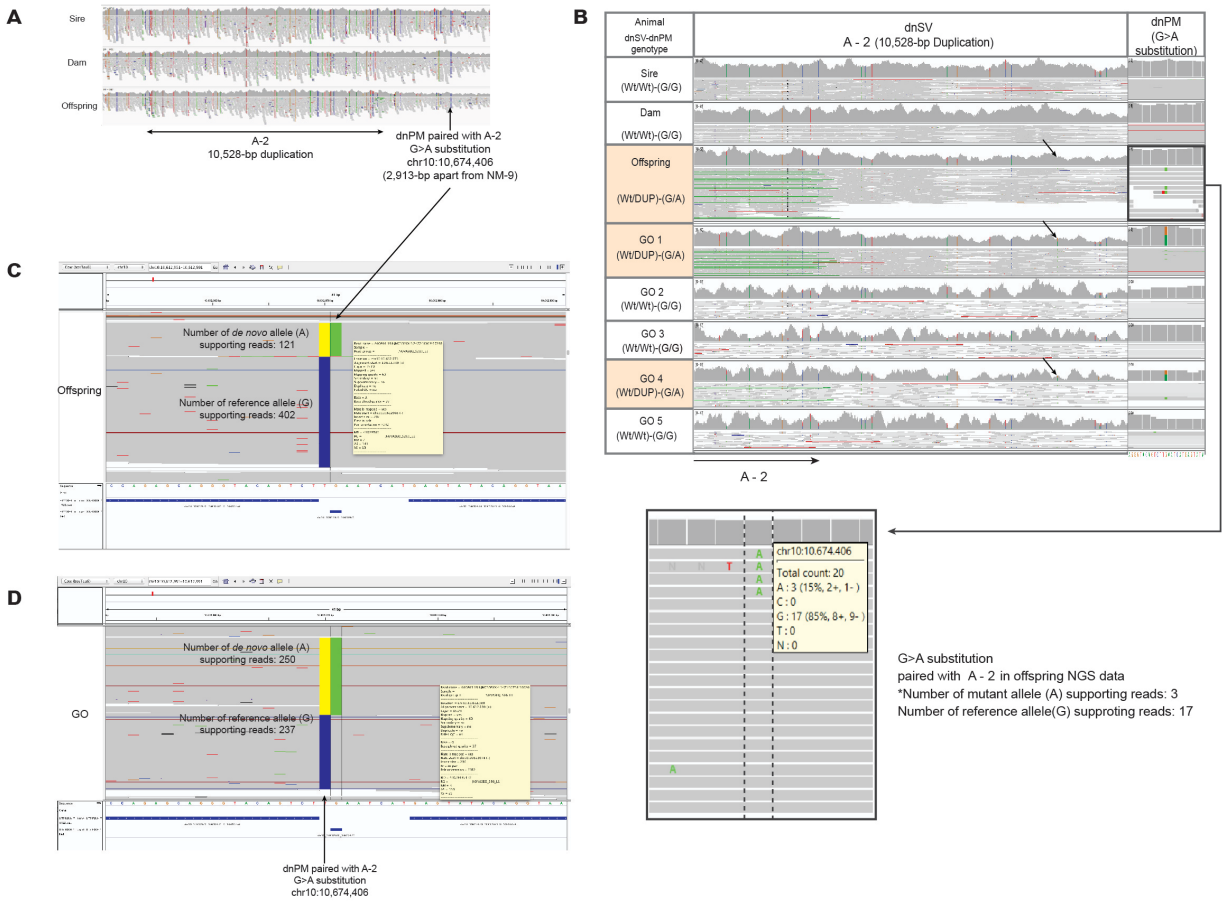
Supplementary Figure S29. Characterization of the cluster mutation involving NM-8

A. An overview of the cluster DNM consisting of NM-8 and A>T substitution. The A>T substitution is located 367-bp from the distal breakpoint of NM-8. The region encompassing the A>T substitution is (marked with the red dotted box) is zoomed in the panel (c). **B.** Genotypes of the cluster DNM in the current pedigree. Wt stands for wild-type and DEL stands for deletion. The offspring carries both NM-8 and the T allele (mutated allele). Of the five GOs, three GOs (GOs 1/3/5), received DEL allele, however only GO1 also received the T allele. Animals carrying NM-8 are marked with translucent orange. Pictures on the right side show the IGV screen capture at the NM-8 and the A>T substitution. **C.** IGV screen capture at the A>T substitution. Reconstruction of reads spanning over the A>T substitution and two other SNPs (A>G and G>C) at the vicinity unravels three haplotypes. These three reads are marked with green, blue, and orange asterisks. **D.** A schematic drawing of the three haplotypes shown in panel (c). The three variants present in this region can be phased into a maternal haplotype carrying A-A-G alleles (green asterisk) and a paternal haplotype carrying G-A-C alleles (orange asterisk). Additional to these haplotypes, there is a paternal haplotype with the dnPM (T allele; G-T-C; blue asterisk), indicating post-zygotic origin of the A>T substitution. Below the drawings of three haplotypes is shown a discordant read pair (filled with red), which can be observed in GO 5 (See panel (b)). GO 5 carries NM-8, but not the dnPM.



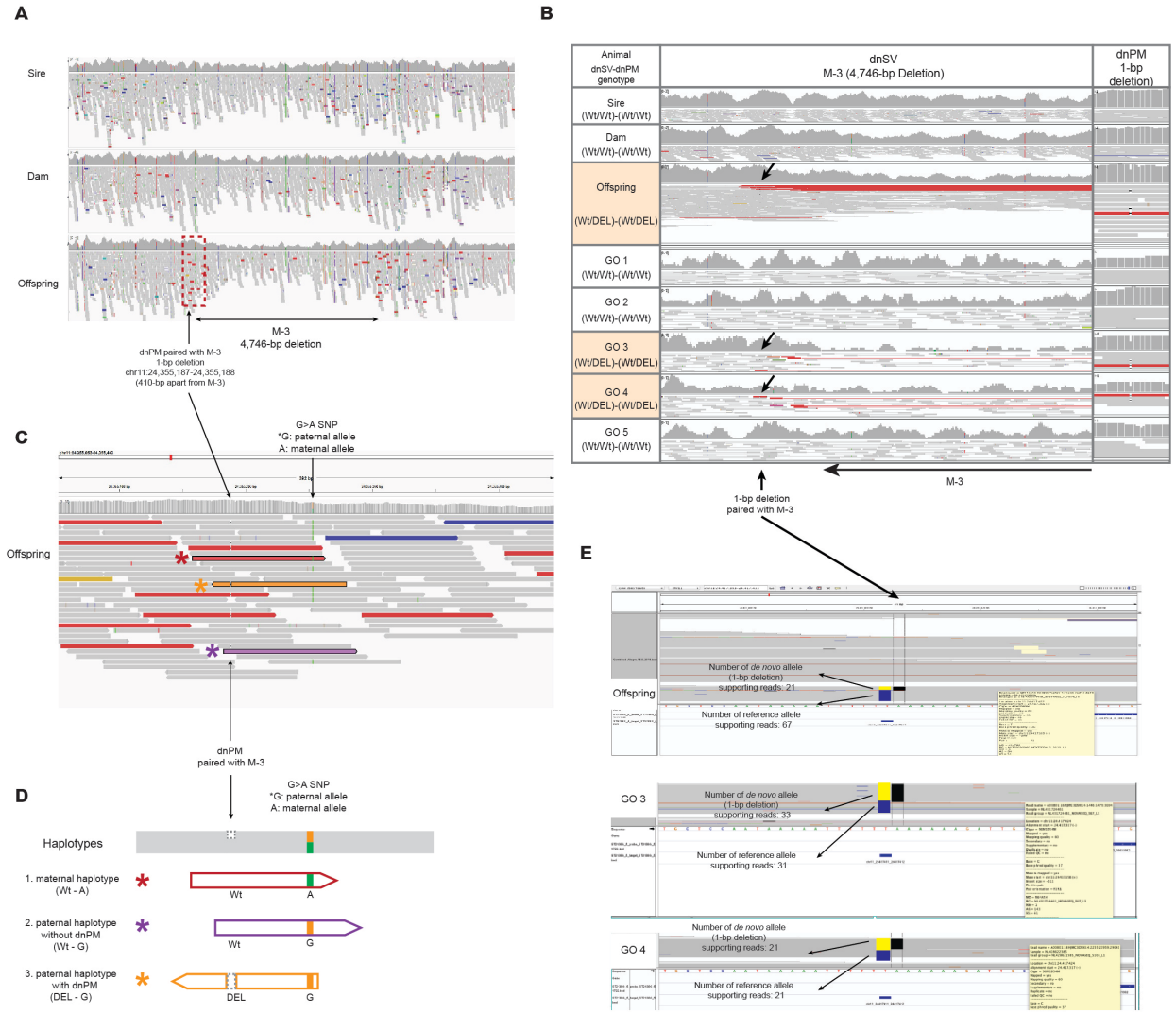
Supplementary Figure S30. Characterization of the cluster mutation involving NM-9

A. An overview of the cluster DNM consisting of NM-9 and a 11-bp insertion. The 11-bp insertion is located 404-bp from the proximal breakpoint of NM-9. **B.** Genotypes of the cluster DNM in the current pedigree. Wt stands for wild-type, DEL stands for deletion, INS stands for insertion. The offspring carries both NM-9 and the 11-bp insertion allele (mutated allele). The GO received both NM-9 and the 11-bp INS. Animals carrying NM-10 are marked with translucent orange. Pictures on the right side show the IGV screen capture at the 11-bp insertion and NM-9. **C.** Amplicon sequencing validation data obtained for the 11-bp insertion in the offspring. In this experiment, the number of reads supporting this reference (WT) and the mutated alleles (INS) were 52 and 33, respectively, thus underlining allelic dosage deviating from ~0.5. **D.** Amplicon sequencing validation data obtained for the 11-bp insertion in the GO. In this experiment, the number of reads supporting this reference (WT) and the mutated alleles (INS) were 35 and 18, respectively, hence the allelic dosage deviating from ~0.5. Thus, taking the allelic dosage shown in the offspring and GO together, the allelic imbalance (deviating from ~0.5) may arise from technical artefacts, instead of an indication of post-zygotic mutation in the offspring.



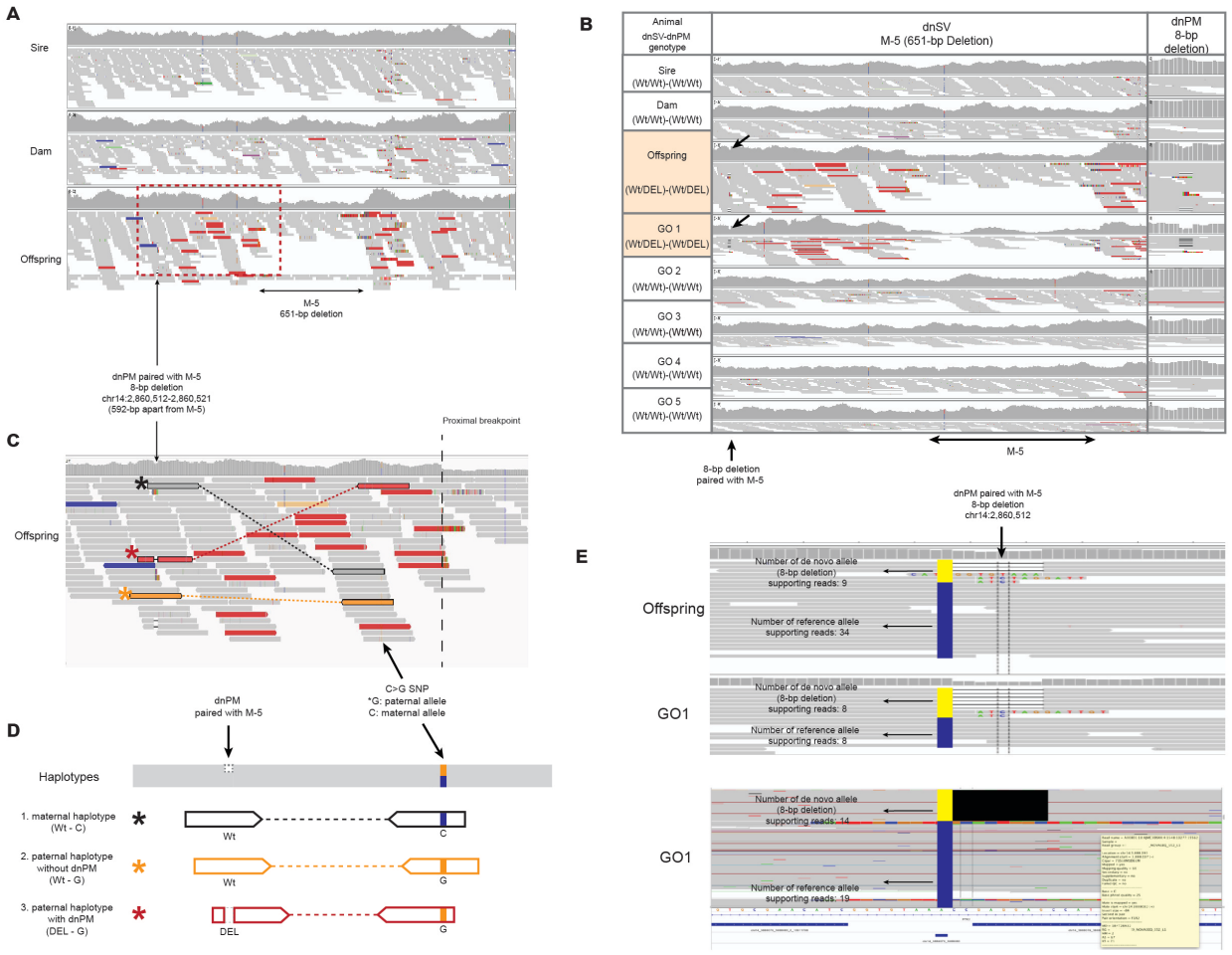
Supplementary Figure S31. Characterization of the cluster mutation involving A-2

A. An overview of the cluster DNM consisting of A-2 and G>A substitution. The G>A substitution is located 2,913-bp from the distal breakpoint of A-2. **B.** Genotypes of the cluster DNM in the current pedigree. Wt stands for wild-type and DUP stands for duplication. The offspring carries both A-2 and the A allele (mutated allele). Of the five GOs, two GOs (GOs 4 and 5), received both DUP and A alleles. Animals carrying A-2 are marked with translucent orange. Pictures on the right side show the IGV screen capture at the A-2 distal breakpoint and the G>A substitution. The G>A substitution occurred in the offspring was zoomed in (marked with the black outline). The number of reads supporting the reference (G) and the mutated alleles (A) in the NGS data were 17 and 3, respectively. **C.** Amplicon sequencing validation data obtained for the G>A substitution in the offspring. In this experiment, the number of reads supporting this reference (G) and the mutated alleles (A) were 402 and 121, respectively, thus revealing an allelic dosage deviating from ~0.5. **D.** Amplicon sequencing validation data obtained for the G>A substitution in one of the GOs. In this experiment, the number of reads supporting the reference (G) and the mutated alleles (A) were 237 and 250, respectively, hence the allelic dosage is ~0.5.



Supplementary Figure S32. Characterization of the cluster mutation involving M-3

A. An overview of the cluster DNM consisting of M-3 and a 1-bp deletion. The 1-bp deletion is located 410-bp from the proximal breakpoint of M-3. The region encompassing the 1-bp deletion is (marked with the red dotted box) is zoomed in the panel (c). **B.** Genotypes of the cluster DNM in the current pedigree. Wt stands for wild-type and DEL stands for deletion. The offspring carries both M-3 and the 1-bp deletion allele (mutated allele). Of the five GOs, two (GOs 3 and 4) received both M-3 and the 1-bp deletion. Animals carrying M-3 are marked with translucent orange. Pictures on the right side show the IGV screen capture at the M-3 and the 1-bp deletion. **C.** IGV screen capture at the 1-bp deletion. Reconstruction of reads spanning over the 1-bp deletion and a SNP (G>A) at the vicinity unravels three haplotypes. These three reads are marked with red, orange, and purple asterisks. **D.** A schematic drawing of the three haplotypes shown in panel (c). The two variants (Wt>DEL and G>A SNP), located outside of M-3, were phased into a maternal haplotype (Wt-A; red asterisk) and a paternal haplotype (Wt-G; purple asterisk). Additional to these haplotypes, there is a paternal haplotype with the dnPM (DEL-G; orange asterisk). **E**. Amplicon sequencing validation data obtained for the 1-bp deletion in the offspring and GOs 3 and 4. The allelic dosage in offspring was inferior to the one shown in GO.



Supplementary Figure S33. Characterization of the cluster mutation involving M-5

A. An overview of the cluster DNM consisting of M-5 and an 8-bp deletion. The 8-bp deletion is located 592-bp from the proximal breakpoint of M-5. The region encompassing the 8-bp deletion is (marked with the red dotted box) is zoomed in the panel (c). **B.** Genotype of the cluster DNM in the current pedigree. Wt stands for wild-type and DEL stands for deletion. The offspring carries both M-5 and the 8-bp deletion allele (mutated allele). Of the five GOs, one (GO 1) received both M-5 and the 8-bp deletion. Animals carrying M-5 are marked with translucent orange. Pictures on the right side show the IGV screen capture at the M-5 and the 8-bp deletion. The 8-bp deletion is highlighted with the arrows. **C.** IGV screen capture at the 8-bp deletion. Reconstruction of reads spanning over the 8-bp deletion and a SNP (C>G) at the vicinity unravels three haplotypes. These three reads are marked with black, red, and orange asterisks. **D.** A schematic drawing of the three haplotypes shown in panel (c). The two variants (Wt>DEL and C>G SNP), located outside of M-5, were phased into a maternal haplotype (Wt-C; red asterisk) and a paternal haplotype (Wt-G; orange asterisk). Additional to these haplotypes, there is a paternal haplotype with the dnPM (DEL-G; red asterisk). **E**. The NGS data shows that the number of reads supporting the reference (Wt) and the mutated alleles (8-bp deletion) for the offspring were 34 and 9. The lower panel shows the NGS data of the GO 1, where the number of reads supporting the reference (Wt) and the mutated alleles (8-bp DEL) for the GO were 8 and 8, respectively. Below, amplicon sequencing validation data obtained for the 8-bp deletion in GO 1 is shown. In this experiment,

the number of reads supporting the reference and the mutated alleles (8-bp DEL) were 19 and 14, respectively.

References

Belyeu JR, Brand H, Wang H, Zhao X, Pedersen BS, Feusier J, Gupta M, Nicholas TJ, Baird L, Devlin B, et al. 2021. De novo structural mutation rates and gamete-of-origin biases revealed through genome sequencing of 2,396 families. *Am J Hum Genet* **108**: 1–11.

Brandler WM, Antaki D, Gujral M, Kleiber ML, Whitney J, Maile MS, Hong O, Chapman TR, Tan S, Tandon P, et al. 2018. Paternally inherited cis-regulatory structural variants are associated with autism. *Science* **360**: 327–331.

Kloosterman WP, Francioli LC, Hormozdiari F, Marschall T, Hehir-Kwa JY, Abdellaoui A, Lameijer EW, Moed MH, Koval V, Renkens I, et al. 2015. Characteristics of de novo structural changes in the human genome. *Genome Res* **25**: 792–801.

Pedersen BS, Quinlan AR. 2019. Duphold : scalable , depth-based annotation and curation of high-confidence structural variant calls. *Gigascience* **8**: 1–5.

Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, Getz G, Mesirov JP. 2011. Integrative genomics viewer. *Nat Biotechnol* **29**: 24–26.

Turner TN, Coe BP, Dickel DE, Hoekzema K, Nelson BJ, Zody MC, Kronenberg ZN, Hormozdiari F, Raja A, Pennacchio LA, et al. 2017. Genomic Patterns of De Novo Mutation in Simplex Resource Genomic Patterns of De Novo Mutation in Simplex Autism. *Cell* 171(3):710-715.e12.

Werling DM, Brand H, An J, Stone MR, Zhu L, Glessner JT, Collins RL, Dong S, Layer RM, Markenscoff-papadimitriou E, et al. 2018. An analytical framework for whole-genome sequence association studies and its implications for autism spectrum disorder. *Nat Genet* **50**: 727–736.