



Visualizing Different Classes of Structural Variants in Bionano Access Software

DOCUMENT NUMBER:

CG-30548

DOCUMENT REVISION:

B

Effective Date:

07/28/2023

Table of Contents

Legal Notice	4
Patents	4
Trademarks	4
Revision History	5
Overview	6
Distinct Data Visualization Methods	6
Genome Browser	7
<i>De novo</i> Assembly Pipeline	7
Rare Variant Analysis Pipeline	8
Whole Genome Copy Number Variant (CNV) View	9
Variants by Class	9
Aneuploidy	9
Copy Number Variants: Deletions	11
Copy Number Variants: Duplications	14
Structural Variants: Insertions	15
Structural Variants: Translocations	17
Structural Variants: Inversions	19
Absence of Heterozygosity (AOH) and Triploidy	21
Macrosatellite/Microsatellite of Repeat Expansions and Repeat Contractions	23
Complex Variants	24
Haploid/double-Haploid	26
Manually Checking Regions	29

False Positive in the Sotos Region of 5q

Technical Assistance

Legal Notice

For Research Use Only. Not for use in diagnostic procedures.

This material is protected by United States Copyright Law and International Treaties. Unauthorized use of this material is prohibited. No part of the publication may be copied, reproduced, distributed, translated, reverse-engineered or transmitted in any form or by any media, or by any means, whether now known or unknown, without the express prior permission in writing from Bionano Genomics. Copying, under the law, includes translating into another language or format. The technical data contained herein is intended for ultimate destinations permitted by U.S. law. Diversion contrary to U. S. law prohibited. This publication represents the latest information available at the time of release. Due to continuous efforts to improve the product, technical changes may occur that are not reflected in this document. Bionano Genomics reserves the right to make changes to specifications and other information contained in this publication at any time and without prior notice. Please contact Bionano Genomics Customer Support for the latest information.

BIONANO GENOMICS DISCLAIMS ALL WARRANTIES WITH RESPECT TO THIS DOCUMENT, EXPRESSED OR IMPLIED, INCLUDING BUT NOT LIMITED TO THOSE OF MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE. TO THE FULLEST EXTENT ALLOWED BY LAW, IN NO EVENT SHALL BIONANO GENOMICS BE LIABLE, WHETHER IN CONTRACT, TORT, WARRANTY, OR UNDER ANY STATUTE OR ON ANY OTHER BASIS FOR SPECIAL, INCIDENTAL, INDIRECT, PUNITIVE, MULTIPLE OR CONSEQUENTIAL DAMAGES IN CONNECTION WITH OR ARISING FROM THIS DOCUMENT, INCLUDING BUT NOT LIMITED TO THE USE THEREOF, WHETHER OR NOT FORESEEABLE AND WHETHER OR NOT BIONANO GENOMICS IS ADVISED OF THE POSSIBILITY OF SUCH DAMAGES.

Patents

Products of Bionano Genomics® may be covered by one or more U.S. or foreign patents.

Trademarks

The Bionano Genomics logo and names of Bionano Genomics products or services are registered trademarks or trademarks owned by Bionano Genomics in the United States and certain other countries.

Bionano Genomics®, Saphyr®, Saphyr Chip®, Bionano Access®, “VIA™ software” and Bionano EnFocus™ are trademarks of Bionano Genomics, Inc. All other trademarks are the sole property of their respective owners.

No license to use any trademarks of Bionano Genomics is given or implied. Users are not permitted to use these trademarks without the prior written consent of Bionano Genomics. The use of these trademarks or any other materials, except as permitted herein, is expressly prohibited and may be in violation of federal or other applicable laws.

© Copyright 2023 Bionano Genomics, Inc. All rights reserved.

Revision History

REVISION	NOTES
A	Initial release
B	Added new images and content

Overview

Bionano Access® Software enables users to view Saphyr® run results in real time, perform a variety of bioinformatics analyses, and visualize structural variation (SV) that has been detected through optical genome mapping (OGM). This reference document guides users through data visualization by variant class. It covers numerous examples of different SV classes and other variant classes detectable by OGM, including insertions, deletions, duplications, inversions, translocations, ring chromosomes, complex rearrangements, Absence of Heterozygosity (AOH), and triploidy. The purpose of this reference document is to help answer user questions and clarify user results, and to organize data into information meaningful to the user.

NOTE: OGM can detect variation throughout most of the genome but does not include coverage of centromeres, short arms of acrocentric chromosomes, and some exceptionally long paracentric low copy repeat regions.

Distinct Data Visualization Methods

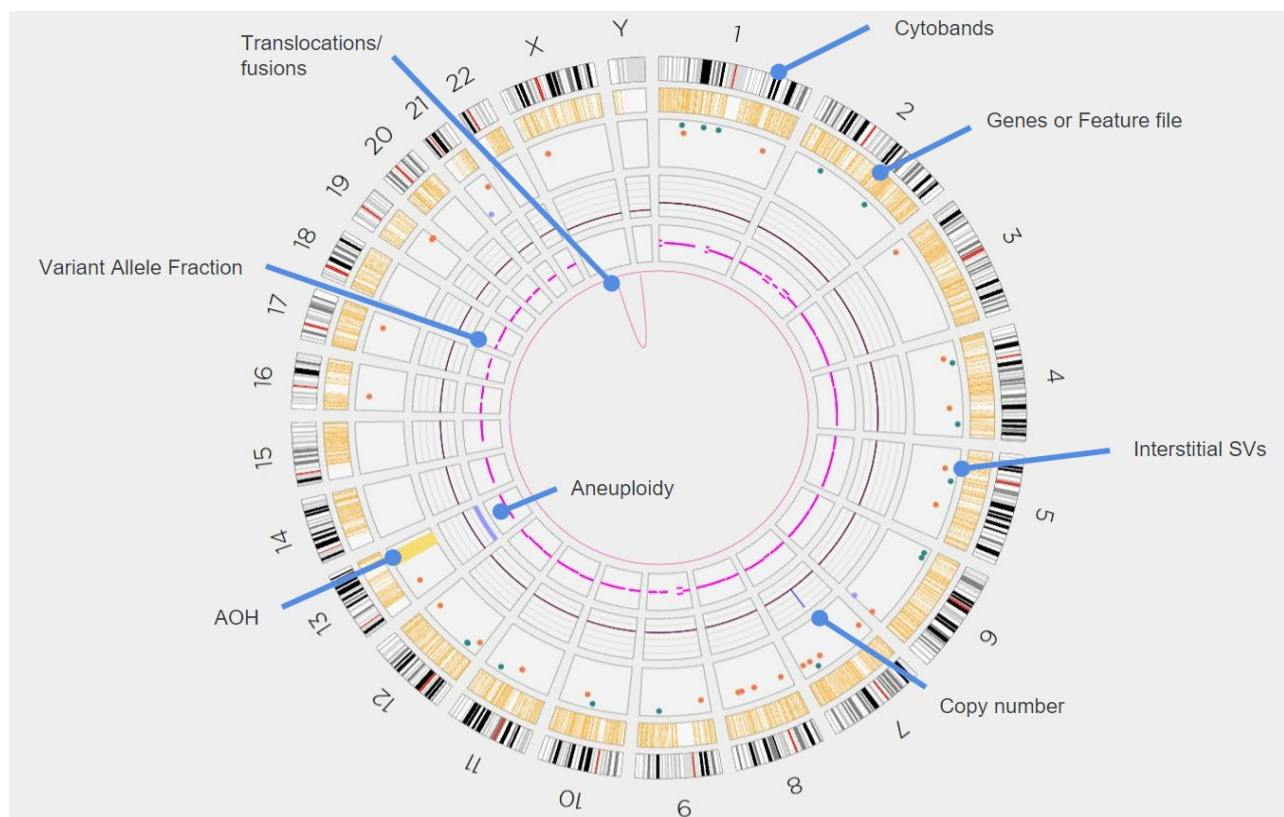


Figure 1. General Circos Plot

In **Figure 1**, the Circos plot provides a whole genome summary of variants detected: copy number (CN) state, regions of Absence of Heterozygosity, and a variant allele fraction profile. This display is filtered dynamically in the user interface of the Access software. The Circos plot also displays chromosome number, cytobands, and,

optionally, gene locations. For more details, please refer to the *Bionano Access Software User Guide* (CG-30142) or to the Bionano Support website.

Genome Browser

There are two algorithm sets for whole genome structural variation detection: the *de novo* Assembly pipeline and the Rare Variant Analysis pipeline (RVA). The *de novo* Assembly pipeline is primarily for germline variants and the RVA is primarily for somatic variants.

De novo Assembly Pipeline

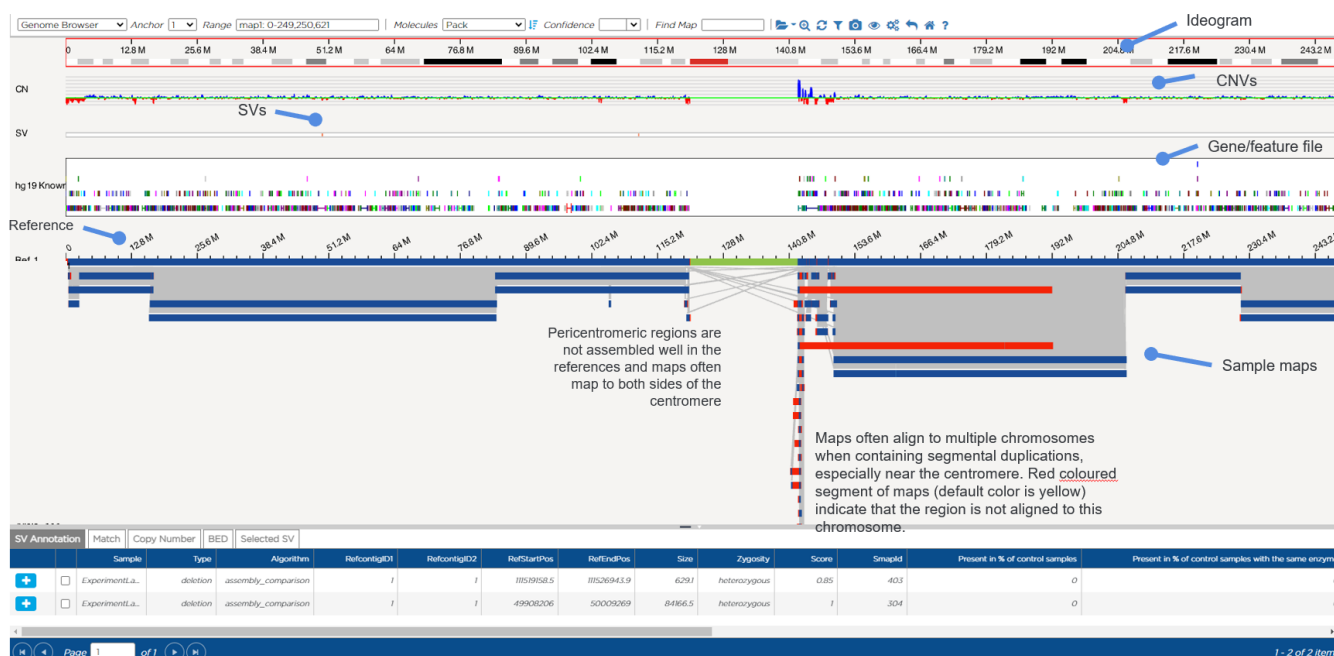


Figure 2. Genome browser – *de novo* Assembly pipeline

In **Figure 2**, the genome browser displays alignment of all assembled maps on single chromosomes. In this example, maps from the *de novo* Assembly pipeline are shown. This pipeline generally assembles contiguous maps, up to chromosome arm length.

NOTE: the default color for non-matching labels is yellow but it can be changed as it has been here, in red, for better resolution and contrast.

The top header displays the types of views, locations on the chromosomes, and customizable options. The top ruler and cytoband track depict what part of the chromosome is being displayed (outlined in red in this example). Cytobands are shown as alternating white, gray, or black bands, with red bands representing centromeres. Other items include:

- CN – track showing copy number baseline, datapoints, and CNV calls.
- SV – track showing structural variants.

- Gene/feature file – when loaded on the Circos plot view, presents BED files containing known hg19 canonical genes.
- Reference chromosome – Ref 1, in this image
- Sample maps that are aligned to the reference.
- Bottom pane - **Variation Annotation** tables

Rare Variant Analysis Pipeline



Figure 3. Genome browser – Rare Variant Analysis pipeline

In **Figure 3**, the genome browser presents the output of the RVA pipeline. The molecules from regions of the genome containing SVs are assembled into maps and each map generally represents a single SV or a cluster of SVs, indicated by the blue pointer. No map is created for parts of the genome matching the reference, hence there are “gaps” with no maps. Here, red colored segments of maps (default color is yellow) indicate that the region is not aligned to an assembled map and the user can infer that there is no SV present.

Whole Genome Copy Number Variant (CNV) View

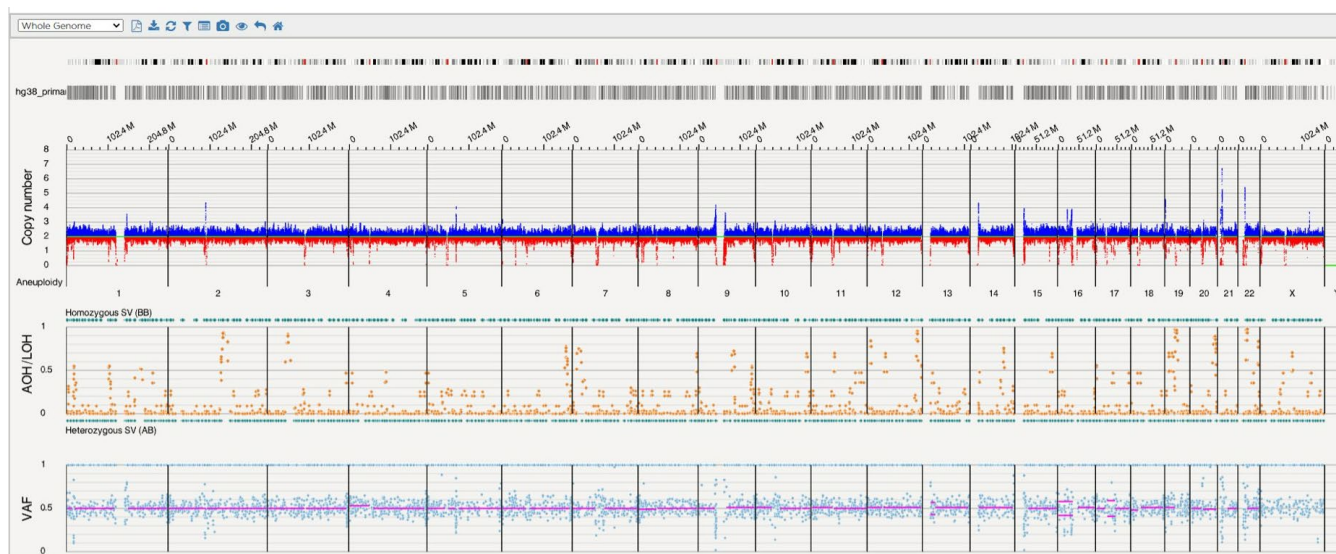


Figure 4. Whole Genome CNV

The whole genome view (top plot) can display the copy number profile calculated by the CNV algorithm based on the number of molecules aligned to the reference genome, as seen in **Figure 4**. This plot highlights CN datapoints centered around a baseline of CN=2, with CNV calls where regions have higher or lower copy coverage compared to most of the genome (generally regions of >500 kbp). The middle plot shows the zygosity of SVs and AOH/LOH data and calls. The final plot on the bottom shows the variant allele fraction (VAF) for each SV. For more detailed information please refer to *the Bionano Access Software User Guide* (CG-30142) or to the Bionano Support website.

Variants by Class

Aneuploidy

Aneuploidy is defined as the condition of having an abnormal number of chromosomes in a diploid cell. For a detailed description of how aneuploidy is called from OGM data, please see the Bionano Genomics document *Bionano Solve Theory of Operations: Structural Variant Calling* (CG- 30110).

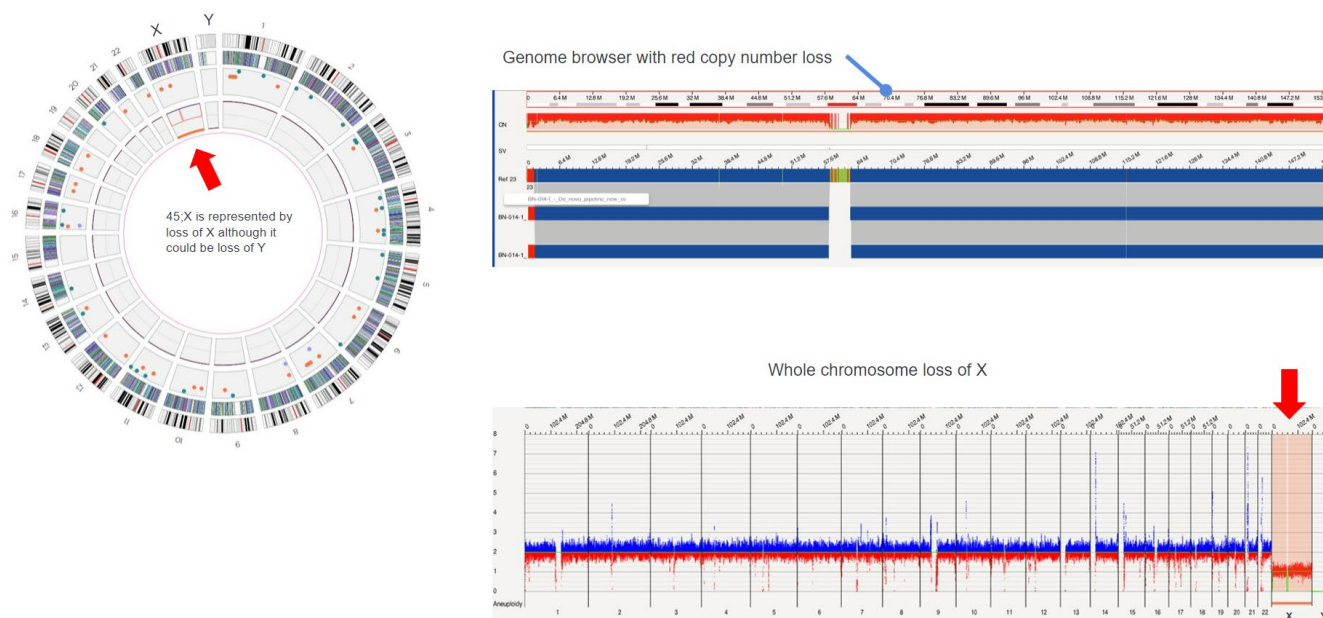


Figure 5. Monosomy – ogm[GRCh38] (X)x1

As shown in **Figure 5**, monosomy, where a single chromosome is represented, can be detected by the CNV algorithm, visualized on the Circos plot as an orange bar in the inner ring (red arrow in this example), and as a copy number call in the first ring (thin red segment(s) inward from the CN baseline). Monosomy can be visualized in the genome browser as red shading and darker red lines projecting down from the top. The whole genome view also indicates loss of chromosome in the same way as the genome browser does. This example shows a 45,X chromosome count which could indicate a loss of X or a loss of Y; the loss of X is arbitrarily indicated as default.

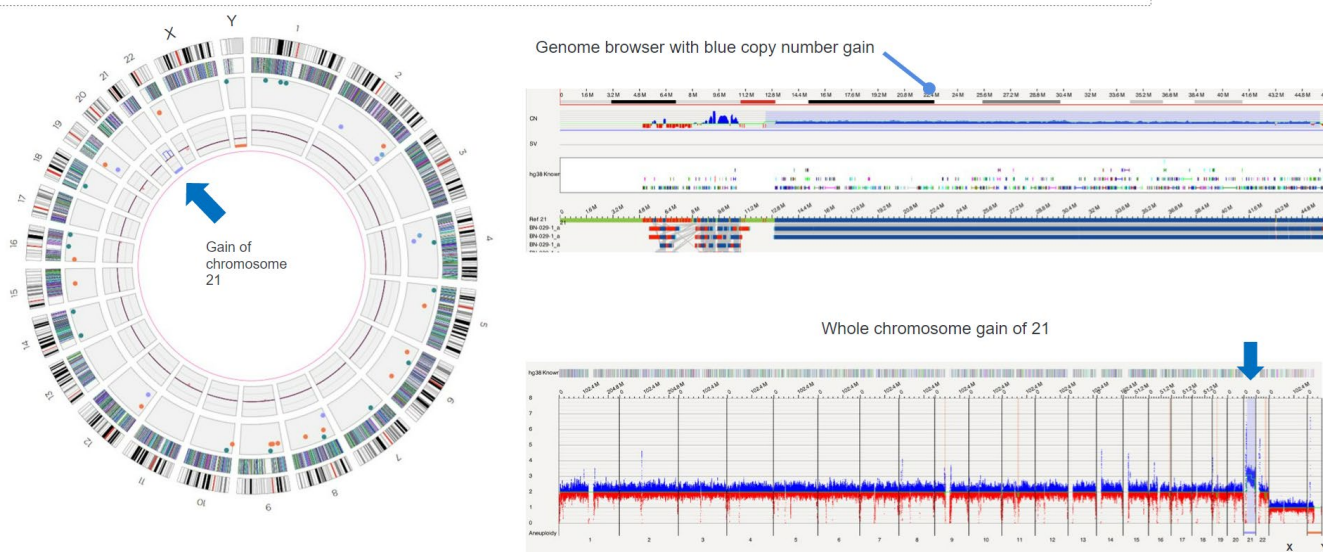


Figure 6. Trisomy – ogm[GRCh38] (21)x3

Trisomy, where three chromosomes are represented, can be detected by the CNV algorithm, visualized on the Circos plot as a lavender bar in the inner ring (blue arrow on Circos plot), and as a CN call in the first ring (thin

blue line), as shown in **Figure 6**. It can be visualized in the genome browser as blue shading and darker blue lines projecting up from the CN status of two. The whole genome view also indicates the gain in the same way as the genome browser does. Since OGM does not detect acrocentric short arm fusions, this profile could represent a true trisomy or an unbalanced Robertsonian translocation.

Copy Number Variants: Deletions

Copy number variation refers to a circumstance in which the number of copies of a specific segment of DNA varies among different individuals' genomes. The individual variants may be short or include thousands of bases. These structural differences may have come about through duplications, deletions or other changes and can affect long stretches of DNA. Such regions may or may not contain a gene(s).



Figure 7. Interstitial deletion – ogm[GRCh38]Xp21.1(31937331_32111402)x0

The deletion in **Figure 7**, depicted in this image by the orange dot, is not called in the CNV pipeline but it is visible in the data. It is not called due to size (165kbp) as the performance of the CNV caller is measured and reported for events 500kbp and above. An interstitial deletion can be seen in the Circos plot as an orange dot, in the genome browser as an SV call (red triangle), and as a horizontal red bar in the SV track. The loss is not called by the CNV algorithm but is supported by the red lines in the CN track.

Large deletions can be seen in the Circos plot in **Figure 9** as a red arrow and connecting lines between sides of the deletion. In the genome browser, large deletions are visualized as a red bar in the CNV track and an intra-chromosomal fusion (if >5 Mbp, or as a deletion if smaller than 5 Mbp) in the SV track.

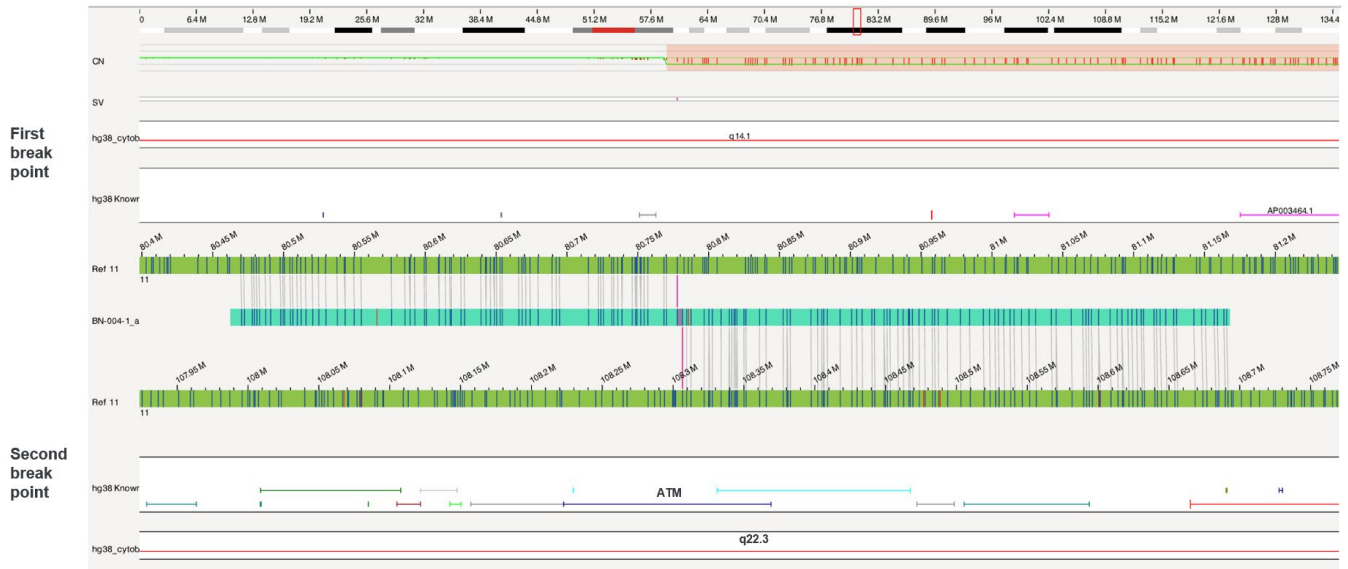


Figure 10. Fusion map from deletion – ogm[GRCh38] t(11;11)(q14.1;q22.3)(80770323_108306632)

By selecting the intra-chromosomal variant from the **Annotation** table, a fusion map depicts both sides of the deletions, visualized vertically in **Figure 10**.

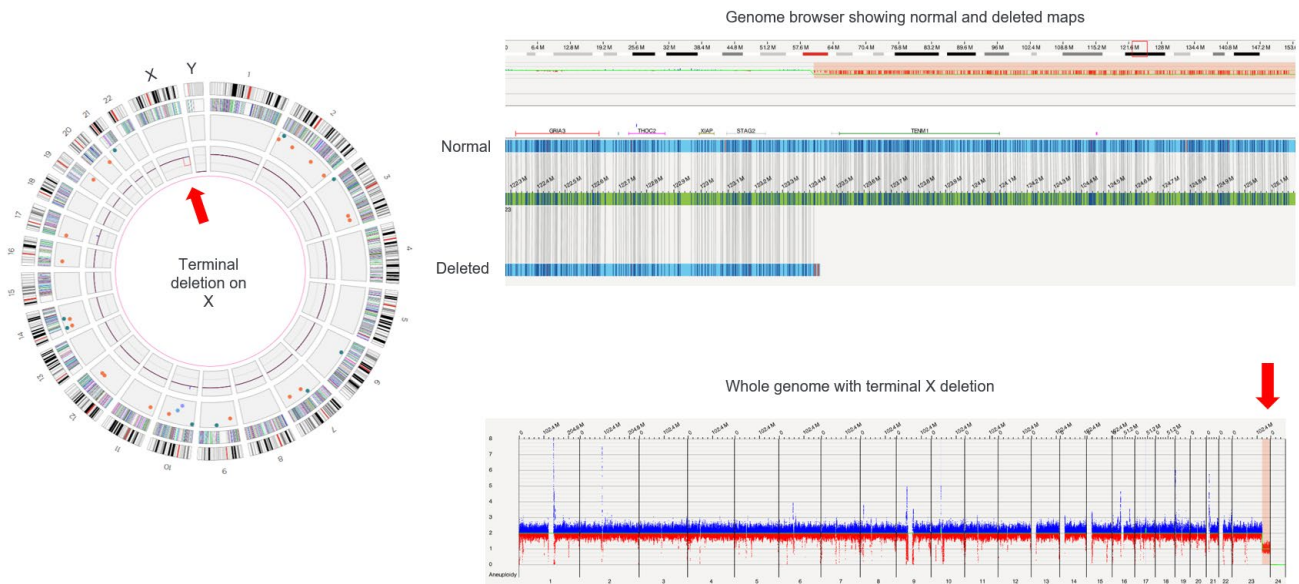


Figure 11. Terminal deletion - ogm[GRCh37] Xq25q28(123414783_155255277)x1

In **Figure 11**, the terminal deletion was detected by the copy number algorithm. However, there is no map that can flank both sides of the deletion, so the SV algorithm was unable to call the deletion. Please refer to *Bionano Solve Theory of Operations: Structural Variant Calling (CG-30110)* for more details.

Copy Number Variants: Duplications

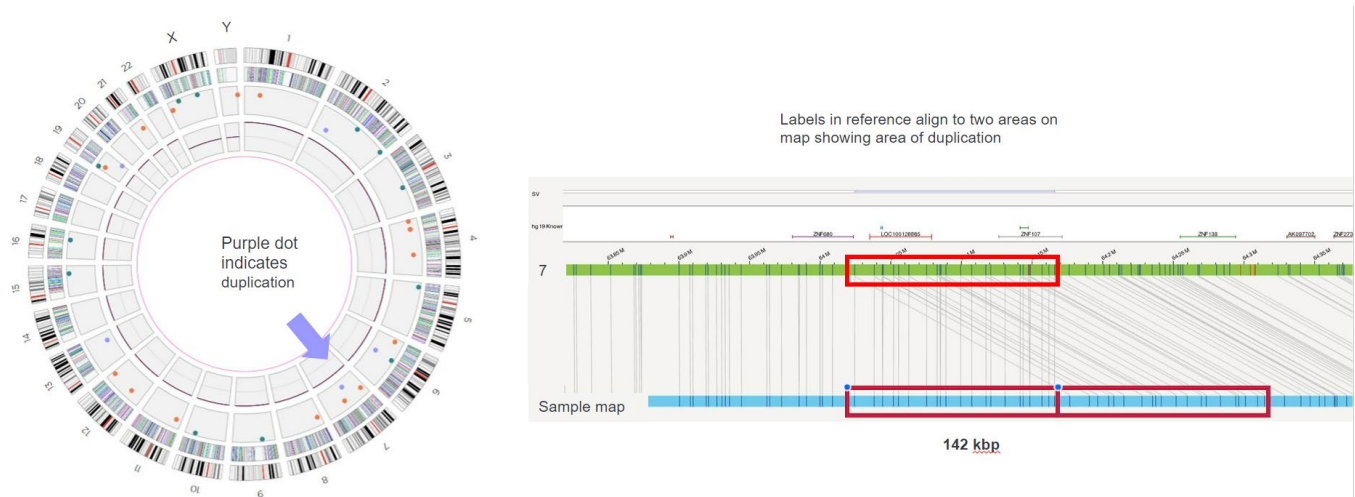


Figure 12. Tandem duplication - ogm[GRCh37] 7q11.21(64024489_64166043)x3

Duplications are indicated by the purple dots in the SV track of the Circos plot in **Figure 12** and by maps with double alignment to the reference in the genome browser. In this example, a tandem direct duplication based on the location adjacent to the first copy and in the same orientation is displayed.



Figure 13. Large duplication - ogm[GRCh38] 2p22.3(32403156_33096390)x3

In **Figure 13**, duplications large enough to be called by SV and CNV algorithms are indicated in the Circos plot as a purple dot in the SV track and blue lines in the copy number track. In the genome browser, large duplications are visualized as a blue bar in the CNV track and a duplication split map in the SV track. Tandem duplications larger than 2 Mbp are called “intrachromosomal fusions” and as blue CNV increase calls (not shown).

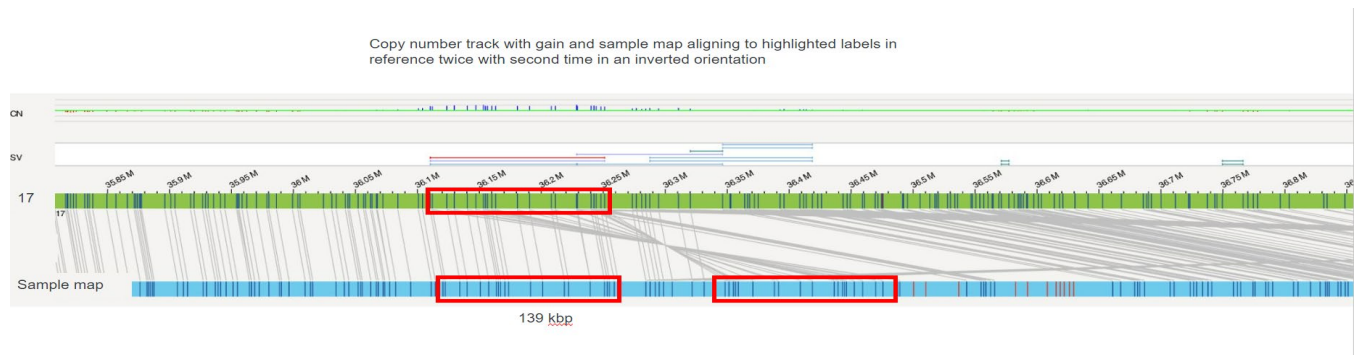


Figure 14. Inverted duplication - ogm[GRCh38] 17q12(36112148_36251175)x3, ogm[GRCh38] inv(17)(q12)(36112148_36251175)

Inverted duplications are shown in **Figure 14** as inverted, duplicated alignments to the reference map.

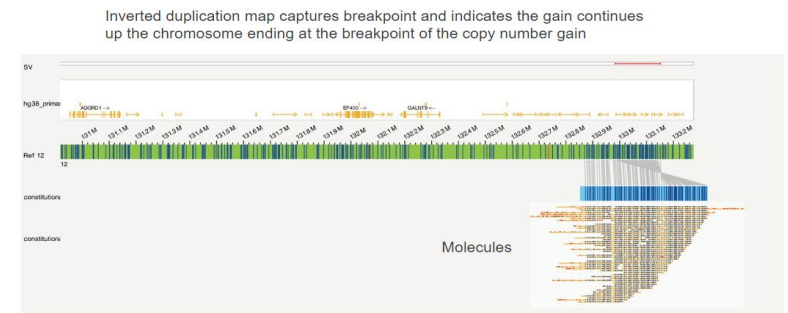
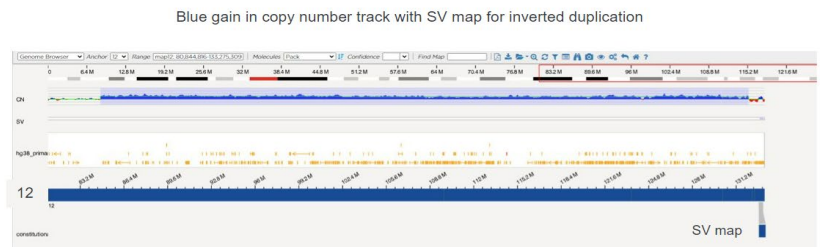
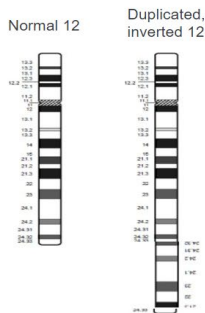
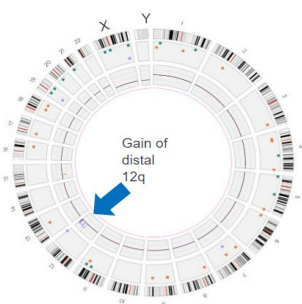
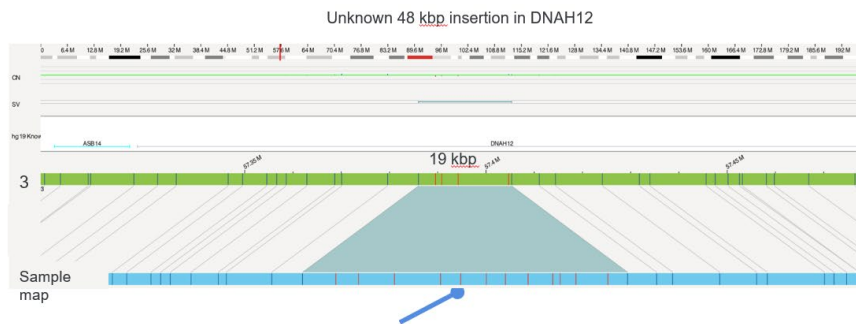
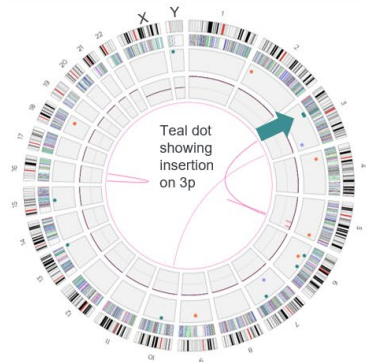


Figure 15. Terminal inverted duplication - ogm[GRCh38] inv(12)(q21.31q24.33)(84677954_133150659)x3

In the Circos plot in **Figure 15**, terminal inverted duplications are shown as lavender dots in the SV track and blue lines in the copy number track. In the genome browser, the variant is represented by a blue bar in the copy number track and by an SV map that contains a duplicated, inverted alignment. The SV map may not extend sufficiently to span the duplicated region, as inverted duplications may be indicative of foldback junctions.

Structural Variants: Insertions

Historically, structural variation is generally defined as a region of DNA approximately 1 kbp and larger in size and can include inversions and balanced translocations or genomic imbalances in the presentation of insertions. However, since the sequencing of human genomes has become routine, the operational spectrum of SVs and CNVs has widened to include smaller events (for example, those >50 bp in length).



Sample map with extra unaligned red labels (default color is yellow) indicating the area is 48 kbp larger than the 19 kbp reference region. The insertion overlaps this region. Note that it is not necessary to have labels in the inserted region to detect an insertion.

Size of insertion is noted in table. Length of region can be determined by differences in start and stop of reference position. However, size of the variant is the difference between the length of the region and the length of the map.

SV Annotation	Match	Copy Number	Selected SV							
	Sample	Type	Algorithm	RefcontigID1	RefcontigID2	RefStartPos	RefEndPos	Size	Score	Smappid
+	ExperimentLa...	insertion	assembly_comparison	3	3	57386006	57405344	48035.9	0.99	1372

Figure 16a. Insertion - ogm[GRCh37] ins(3)(p14.3)(57386006_57405344)

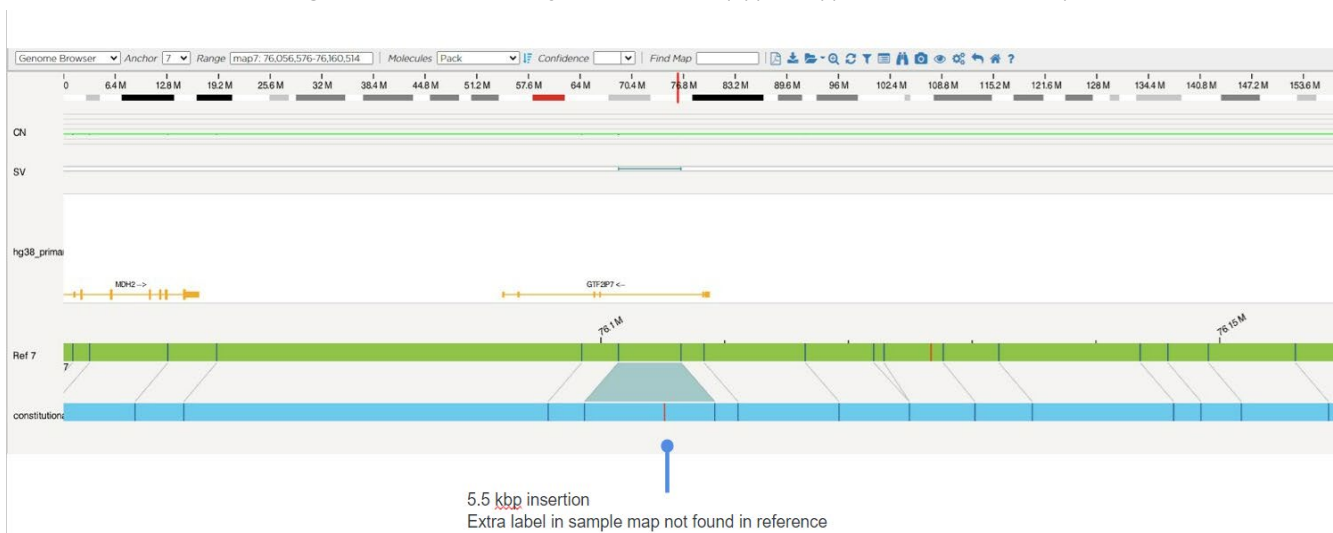


Figure 16b. Small insertion of unknown origin – ogm[GRCh38] ins(7;?)(q11.23;?).

In **Figures 16a and b**, insertions are represented by teal dots in the SV track and by maps that contain boundary labels larger than the reference positions in the genome browser.

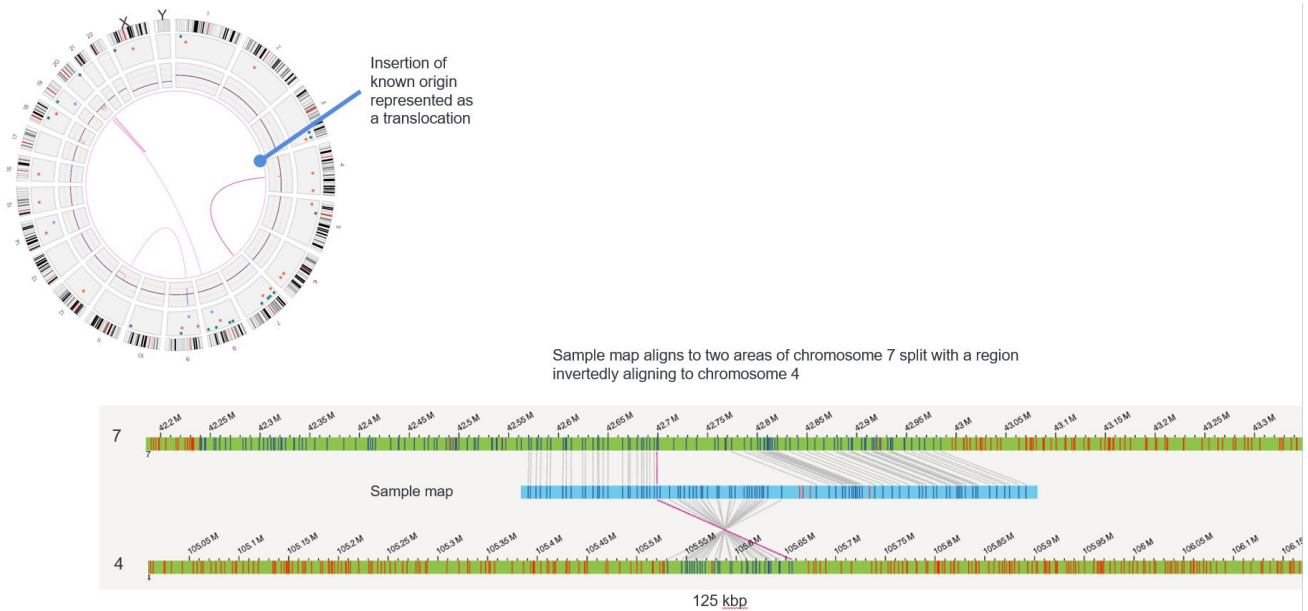


Figure 17. Insertion with known genetic origin - ogm[GRCh38] inv ins(7;4)(p14.1;q24)(42699099; 105656151_105531076)

Insertions of known origin are indicated by magenta translocation lines in the Circos plot in **Figure 17** and by translocation maps in the genome browser.

Structural Variants: Translocations

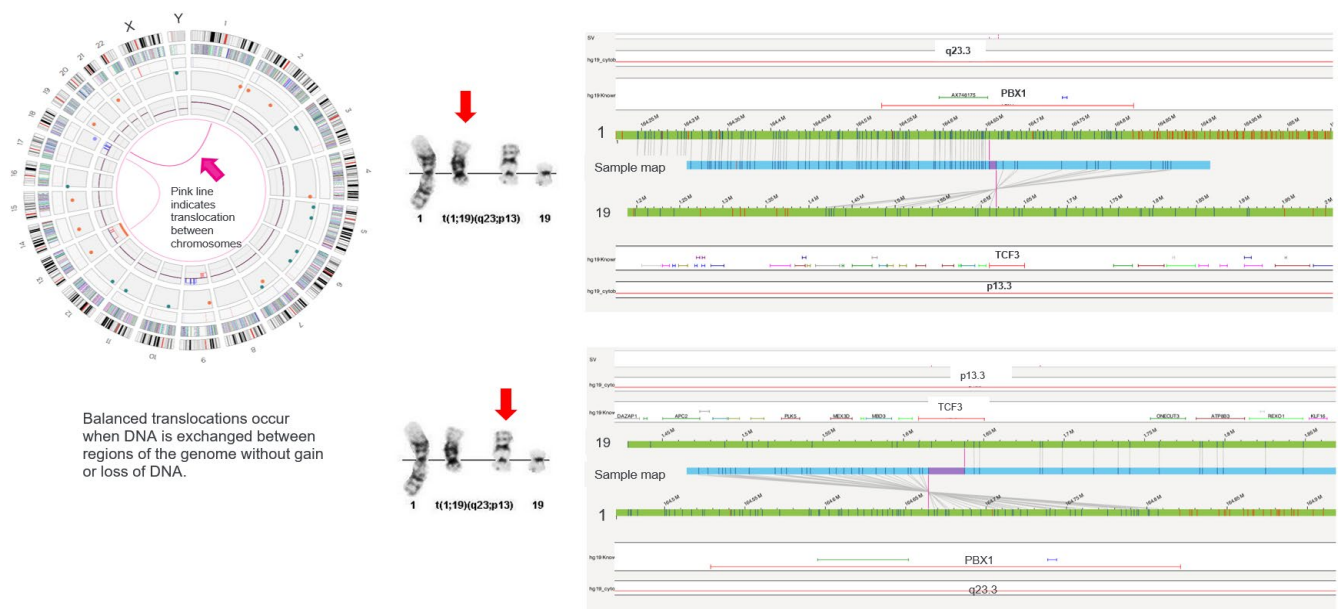


Figure 18. Balanced - ogm[GRCh37] t(1;19)(q23.3;p13.3)(164653451_1617441)

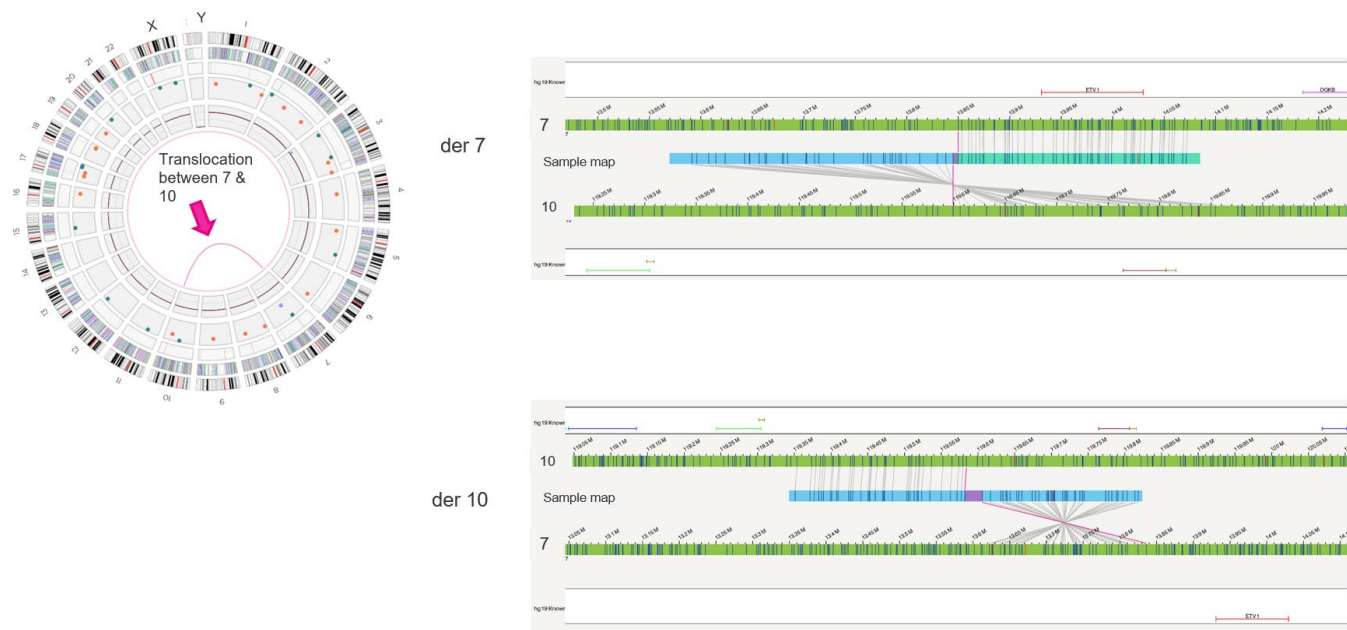


Figure 19. Balanced - ogm[GRCh37] t(7;10)(p21.2;q26.11)(13850028_119599351)

Figures 18 and 19. In both Circos plots above, balanced translocations are indicated by magenta lines connecting the involved chromosomes. Breakpoints and genes can be visualized in the genome browser by viewing the translocation maps.

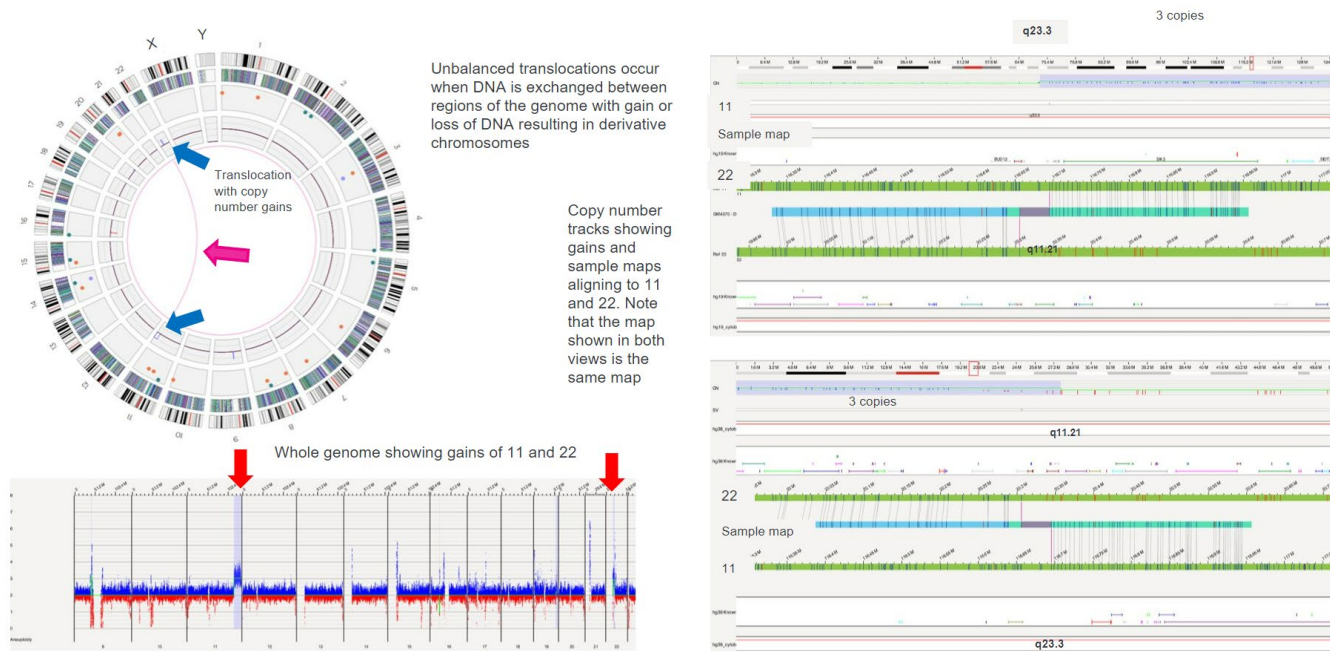
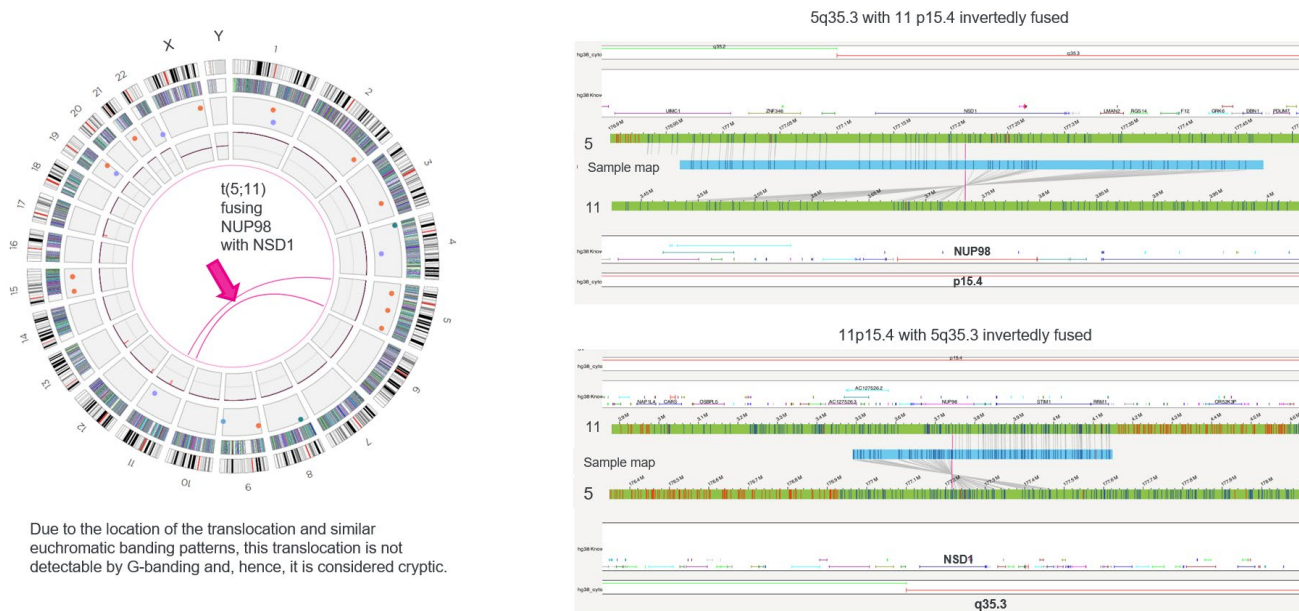


Figure 20. Unbalanced translocation - ogm[GRCh37] t(11;22)(q23.3;q11.21)(116695396_20305577),ogm[GRCh37] 11q23.3q25(116682645_134939459)x3,ogm[GRCh37] 22q11.1q11.21(16695297_20356873)x3

In **Figure 20**, unbalanced translocations are indicated by magenta lines connecting the involved chromosomes and blue and/or red lines in the copy number track, associated with the magenta lines. In the genome browser, the translocation maps, with associated breakpoints and genes, can be visualized with the copy number gains and losses found in the copy number track. Copy number changes can also be visualized with the whole genome view.



Due to the location of the translocation and similar euchromatic banding patterns, this translocation is not detectable by G-banding and, hence, it is considered cryptic.

Figure 21. Cryptic translocation - o[m][GRCh38] – t(5;11)(q35.3;p15.4)(177212867_3734712)

Cryptic translocations, as seen in **Figure 21**, can be viewed as magenta lines in the Circos plot and translocation maps in the genome browser view. Translocations are considered cryptic because of their specific location in the chromosome, not detectable by other molecular genetic strategies.

Structural Variants: Inversions

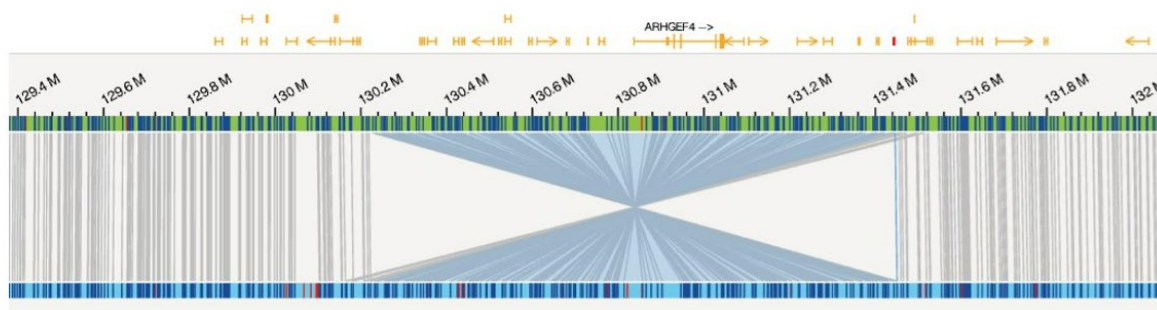


Figure 22. Simple inversion

Inversions are visualized in the genome browser at an inverted orientation of the sample map labels to the reference map labels. **Figure 22** displays inversions < 5 Mbp.

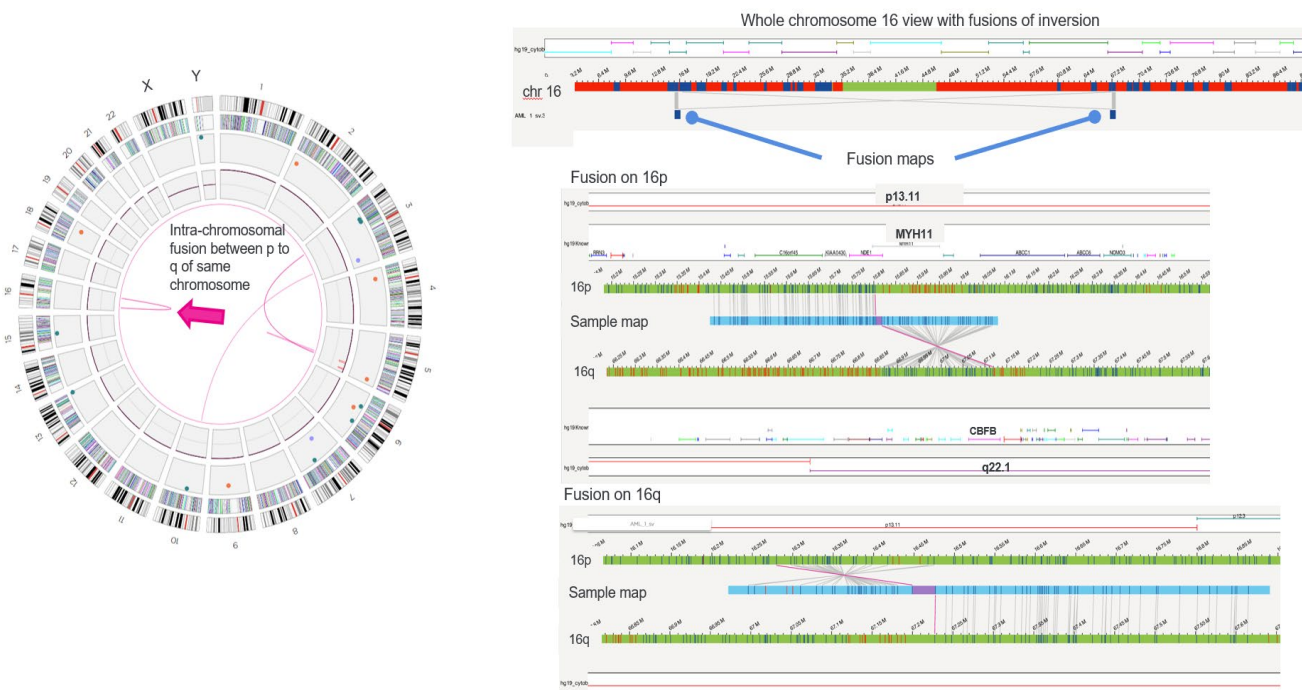


Figure 23. Pericentric inversion/fusion - ogm[GRCh37] inv(16)(p13.11q22.1)(15803116_67118000)

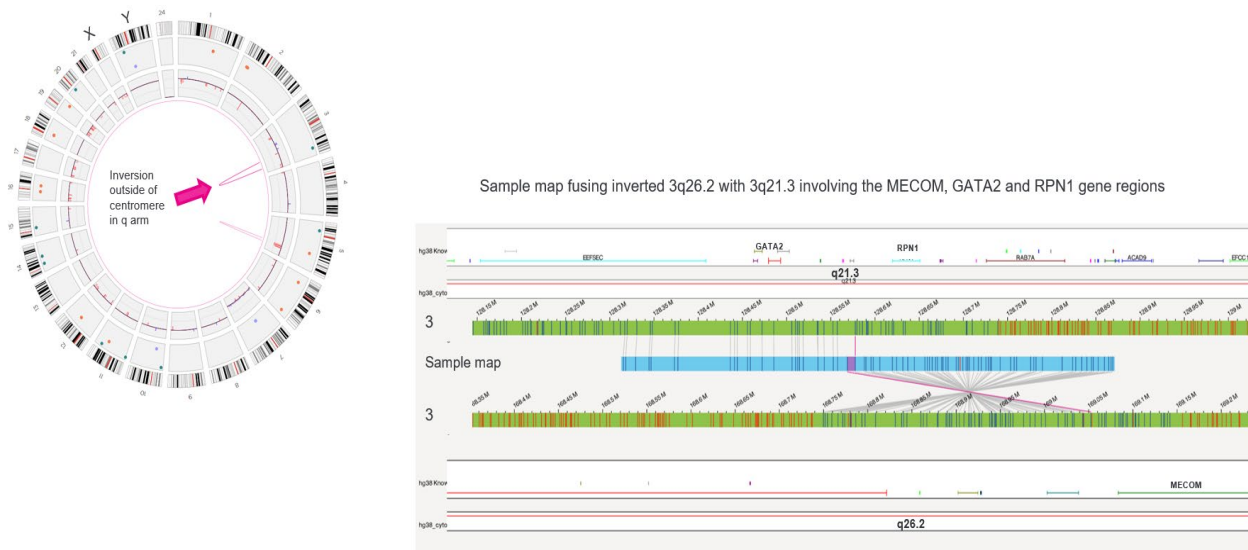


Figure 24. Paracentric inversion - ogm[GRCh38] inv(3q21.3q26.2)(128577641_169052490)

Large pericentric (**Figure 23**) and paracentric (**Figure 24**) inversions are indicated by magenta lines in the Circos plot. The sides of the inversions are captured by intra-chromosomal fusion maps that can be visualized in the genome browser.

- OGM cannot differentiate cis or trans fusion pairs
- **inv(16)(p13q22) CBFβ/MYH11**
- **t(16;16)(p13;q22) CBFβ/MYH11**
- **del(16)(q22) CBFβ/MYH11 (not shown)**

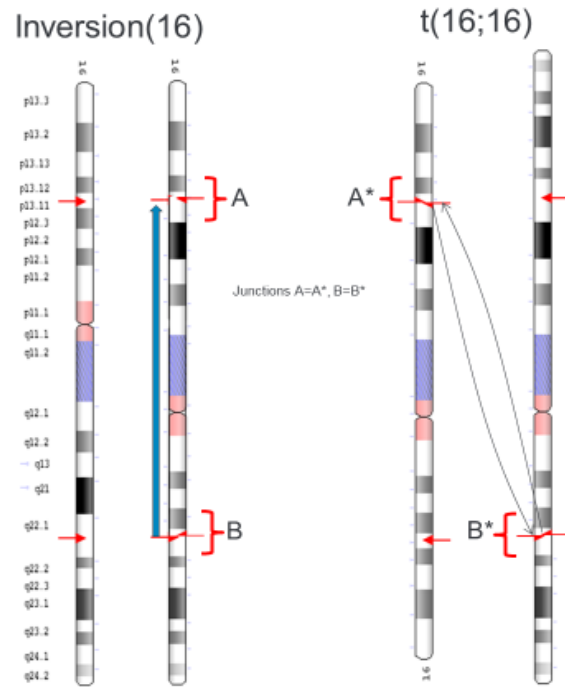


Figure 25. Inv(16) vs. t(16;16)

Optical genome mapping cannot differentiate between a translocation involving two homologs or an inversion happening at the same breakpoints (see **Figure 25**).

Absence of Heterozygosity (AOH) and Triploidy

Absence of heterozygosity (AOH) is a well-recognized genomic change by which heterozygous alleles are absent, rendering the sequence effectively homozygous. This homozygosity can result in human diseases including congenital and somatic disorders due to absence of wild-type alleles or biparental expression of an imprinted region. Triploidy is a chromosome abnormality that occurs when there is an extra set of chromosomes present in each cell. The signs and symptoms associated with triploidy vary but may include a variety of birth defects and unusually small size.

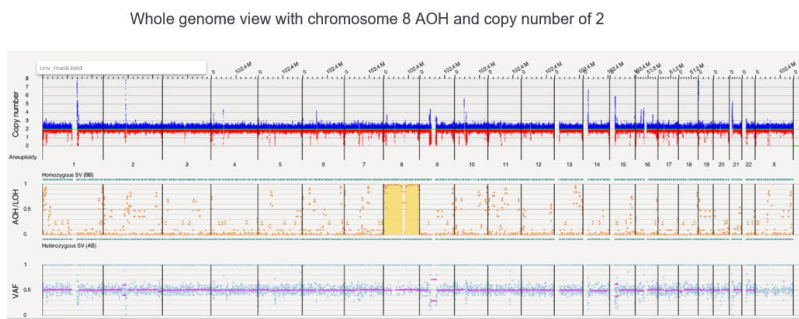
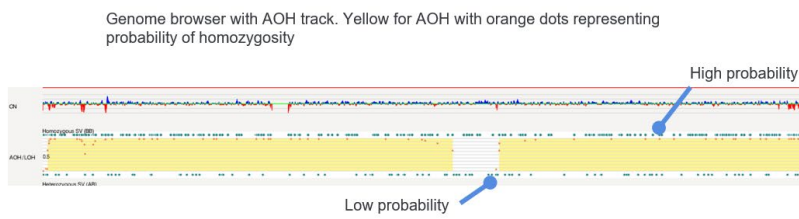
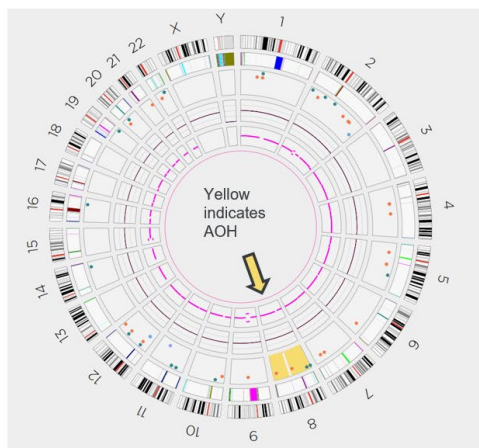


Figure 26. AOH: Chromosome 8

In **Figure 26**, a yellow region in the AOH track shows two identical homologs in chromosome 8, demonstrating absence of heterozygosity.

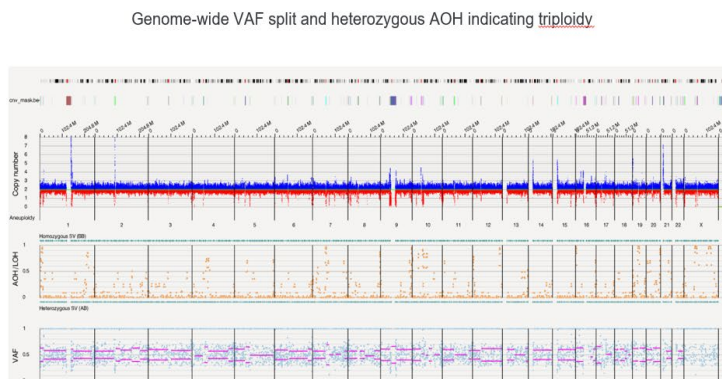
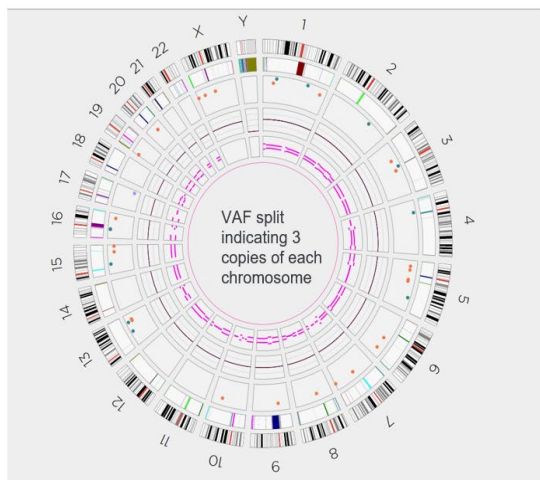


Figure 27. Polyploid - Triploidy

Triploidy— the magenta lines indicate that the variant allele frequency (VAF) is split in the Circos plot in **Figure 27** where there are three copies of each chromosome in this 69XXX sample. This represents the observation of three alleles across the genome in a triploidy case (i.e., 69XXX, 69XXY, 69XYY). In triploidy, for heterozygous loci, there are always two of one allele and one of another (three of the same allele were homozygous), resulting in 33% or 66% variant allele fraction. This is represented by the pair of magenta lines around the inner ring. **NOTE:** a diploid sample will have only one magenta line.

Macrosatellite/Microsatellite of Repeat Expansions and Repeat Contractions

Bionano EnFocus pipelines are designed to characterize target loci. Currently, there are two EnFocus tools in the Bionano Access software suite: EnFocus Fragile X and EnFocus FSHD. In both pipelines, molecules aligning to the regions of interest are extracted and assembled. The resulting consensus maps can be used to characterize the repeats, which include repeat expansions (polymorphic nucleotide sequences scattered throughout the human genome) and repeat contractions (specific for the normal allele). Please refer to the *Fragile X* (CG-30457) and *FSHD* (CG-30321) *Theory of Operation* analysis documents.

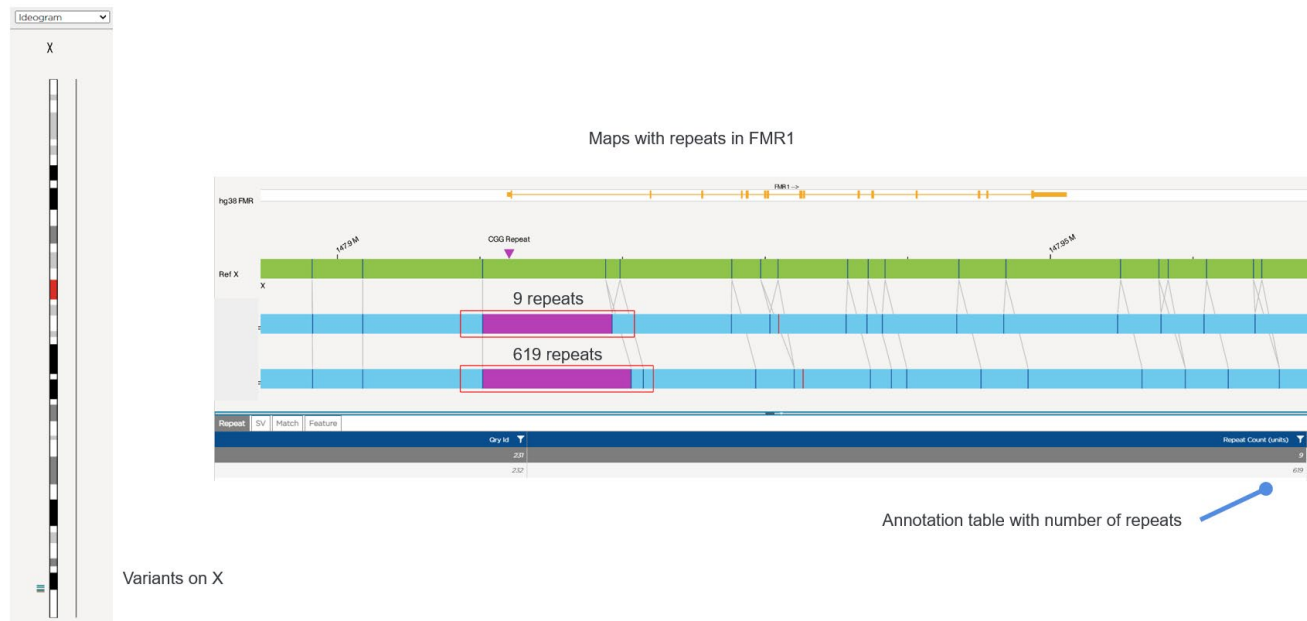


Figure 28. Fragile X – Repeat Expansion

Fragile X expansion information and maps can be visualized with the EnFocus Fragile X analysis tool, as seen in **Figure 28**. The gene structure of FMR1 is shown at the top of the figure. The label intervals on the consensus maps that overlap the trinucleotide locus are colored in magenta. The number of repeats is computed by comparing the interval lengths on the maps and the interval lengths on the reference, subtracting a constant region (around the repeat expansion region), and converting to repeats (triplet repeat is three nucleotides per repeat unit). In this example, the repeat counts are 9 and 619. Please see *Bionano Solve Theory of Operations: Structural Variant Calling* (CG-30110) for more detail.



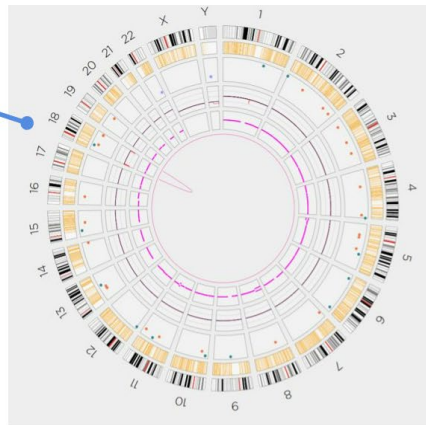
Figure 29. FSHD – Repeat contraction.

Repeat contractions related to FSHD1 can be assessed with the EnFocus FSHD analysis tool. **Figure 29** depicts the two loci with D4Z4 macrosatellite repeats on 4q and 10q. The map regions that overlap the D4Z4 repeats on chromosomes 4 and 10 are colored in magenta. The lengths of the repeats are determined by comparing the interval lengths of the maps and the reference. The right most labels beyond the D4Z4 region are used to determine the A and B haplotypes.

Complex Variants

Complex variants are genomic rearrangements comprising multiple structural variants, typically involving three or more breakpoint junctions. Examples are isochromes and chromothripsis as these contribute to human genomic variation and can cause Mendelian disease.

Deletions at both ends of chromosome 18



Map showing fusion of deletion breakpoints and proximally moving alignment

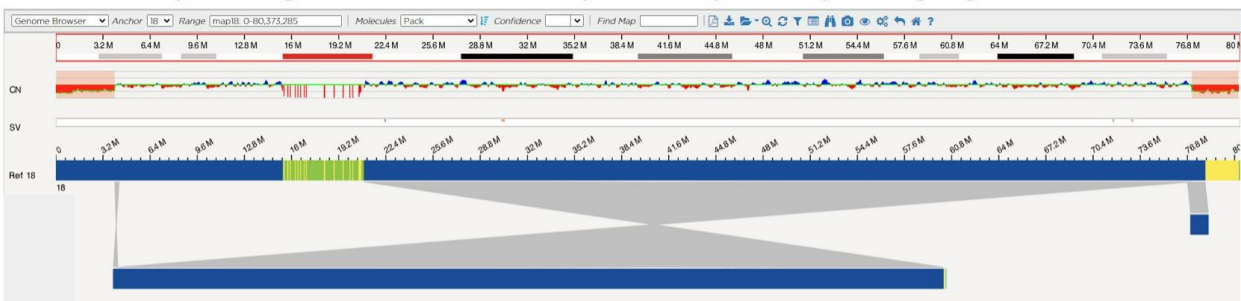
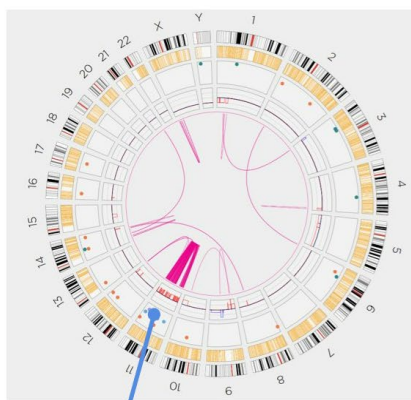


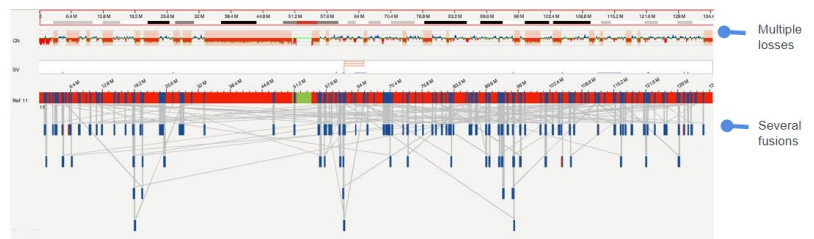
Figure 30. Ring chromosome

Ring chromosomes can be detected by OGM and are often indicated as shown in **Figure 30**. Here, a sub-telomeric fusion is displayed in the Circos plot by an intrachromosomal connection of the two ends of chromosome 18, and by CNV losses coordinated with the intrachromosomal fusion. In the genome browser, the copy number losses and the inverted fusion map are visualized.

Some variants are complex and may be hard to reconstruct the chromosome structure



Chromothripsis



With several breakpoints, putative gene fusions can be identified in the annotation table and visualized in the genome browser



Figure 31. Complex rearrangements

In the Circos plot seen in **Figure 31**, chromothripsis (thripsis, in translation from Greek, means destruction into small parts) can be inferred by observation of many magenta translocation lines within a single chromosome and/or between a few chromosomes. In the genome browser, intrachromosomal fusion maps and/or interchromosomal translocation maps, and the putative gene fusions in the **Annotation** tables, can provide additional information about the structure of chromothripsis.

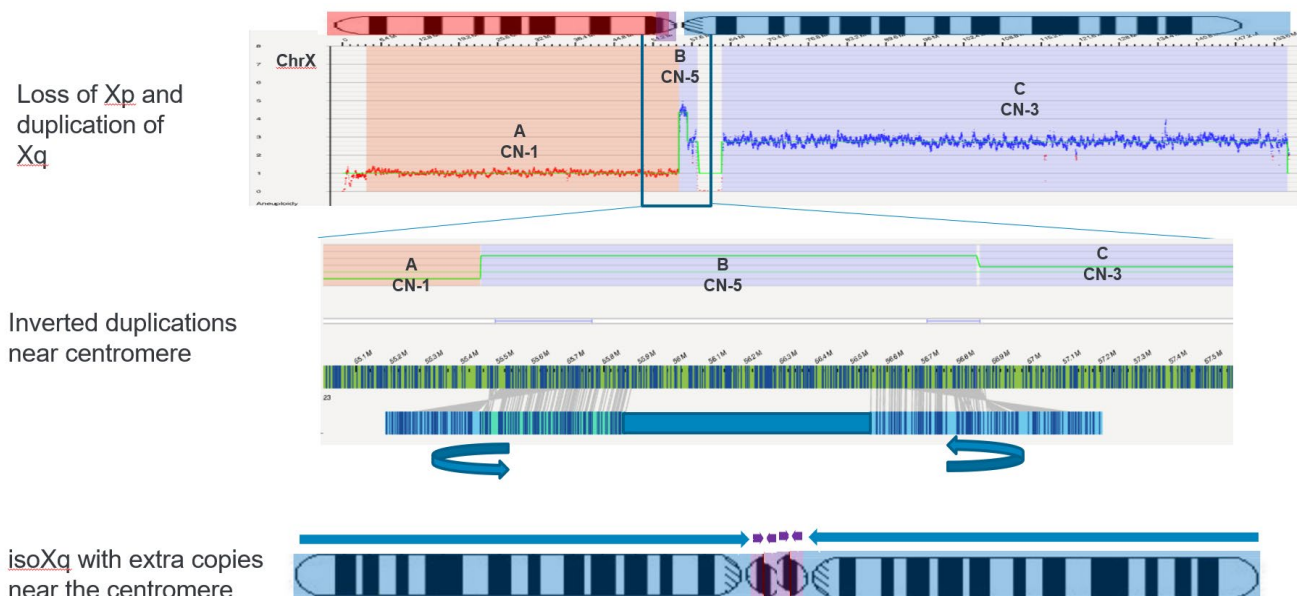


Figure 32. Complex X

Isochromosomes can be detected by gains and losses of whole chromosome arms within the same chromosome, as depicted in **Figure 32**.

Haploid/double-Haploid

For samples with abnormal baseline ploidy, the copy number algorithm may normalize the baseline ploidy incorrectly. This can be detected by whole chromosome gains of even numbers of copies. If this is suspected, then running the *de novo* pipeline to detect LOH could reveal haploidy or double haploidy. **Figure 33** suggests a doubling of a haploid event evidenced by 4-copy gains of chromosomes 14, 18, and 21. In **Figure 34**, the *de novo* assembly shows LOH of all chromosomes except for the 4-copy gains which further suggests the haploid/double-haploid genome.

Karyotype:
27<1n>,XX,+14,+18,
+21[20]

RVA

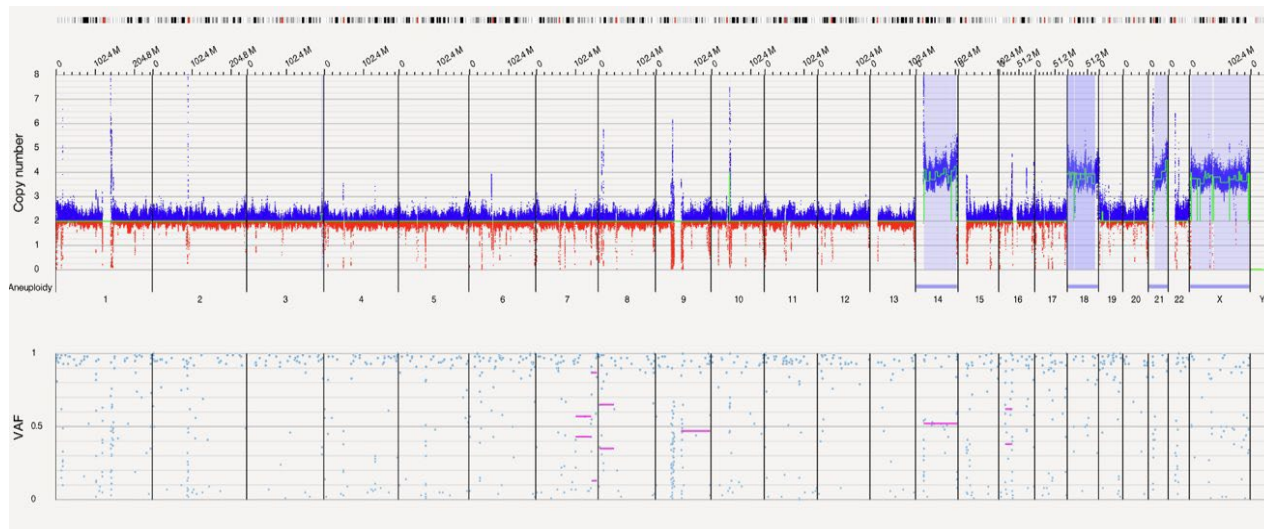


Figure 33. Doubling of a haploid event

De novo

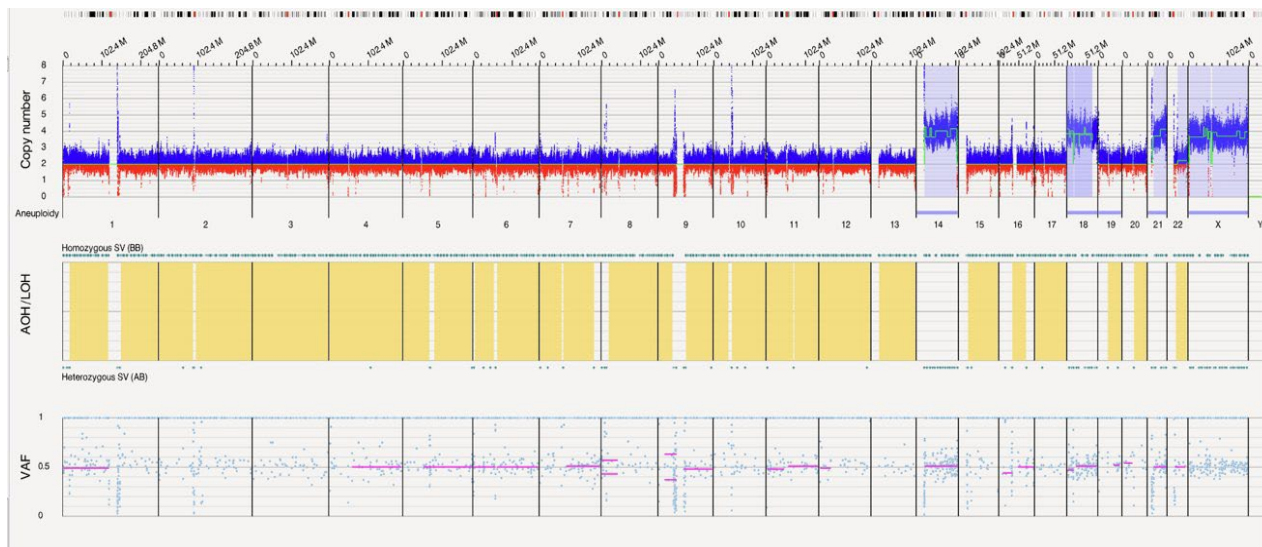


Figure 34. Haploid/double-haploid genome

HYPERDIPLOID

Visualizing the whole genome, with the relative losses and gains, can assist in determining off-center data. In **Figure 35**, copy number changes for chromosomes 5q, 7p, 9, 13, 15-17, 20, and Y are suggestive of two copies and, in relation, chromosomes 2-4, 6, 8, 14, 18, 19, and X could have three copies while chromosomes 1, 5p, 7q, 10-12, and 22 could have four copies. Chromosome 21 might have five copies. Using the increase in data points and LOH found in the *de novo* assembly can assist in determining copy number. In **Figure 36**, in addition to their relative height on the copy number scale, most chromosomes thought to be at two copies have LOH and few VAF

data points whereas most three copy chromosomes do not contain LOH, and four copy gains have many heterozygous data points.

Karyotype: 74~76<3n>,XXY,+X,+Y,+1,-
 5,i(7)(q10),-
 9,+10,+11,+12,add(12)(q24)x2,-13,-
 15,-16,-17,-
 20,+21,+21,+21,+22,+22,+3mar[cp17]

RVA

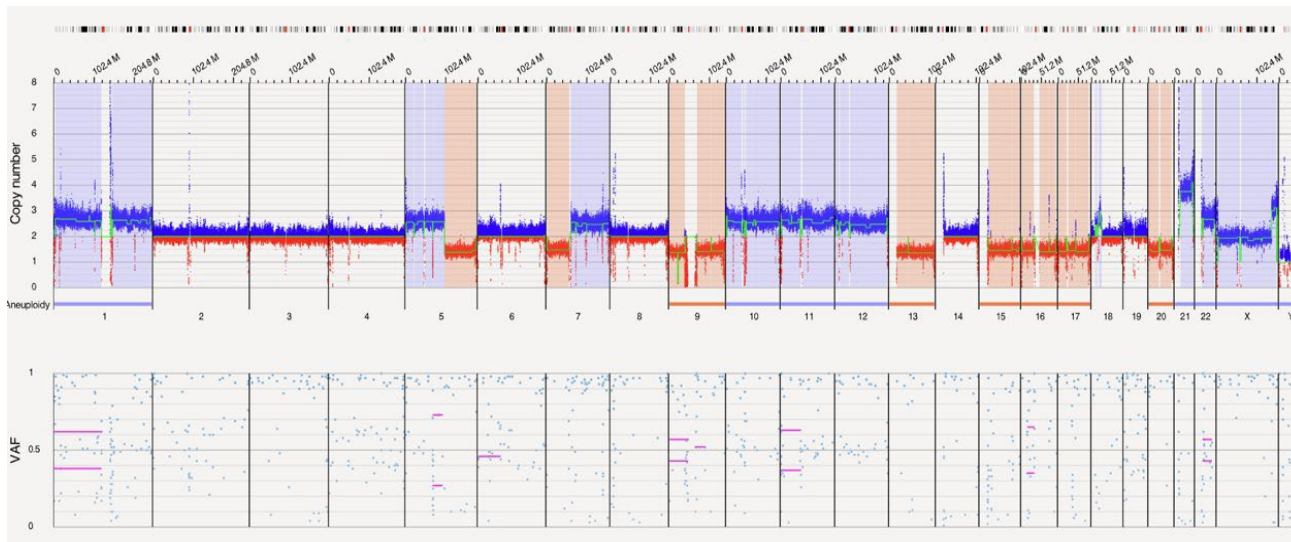


Figure 35. RVA of a hyperdiploid event

De novo

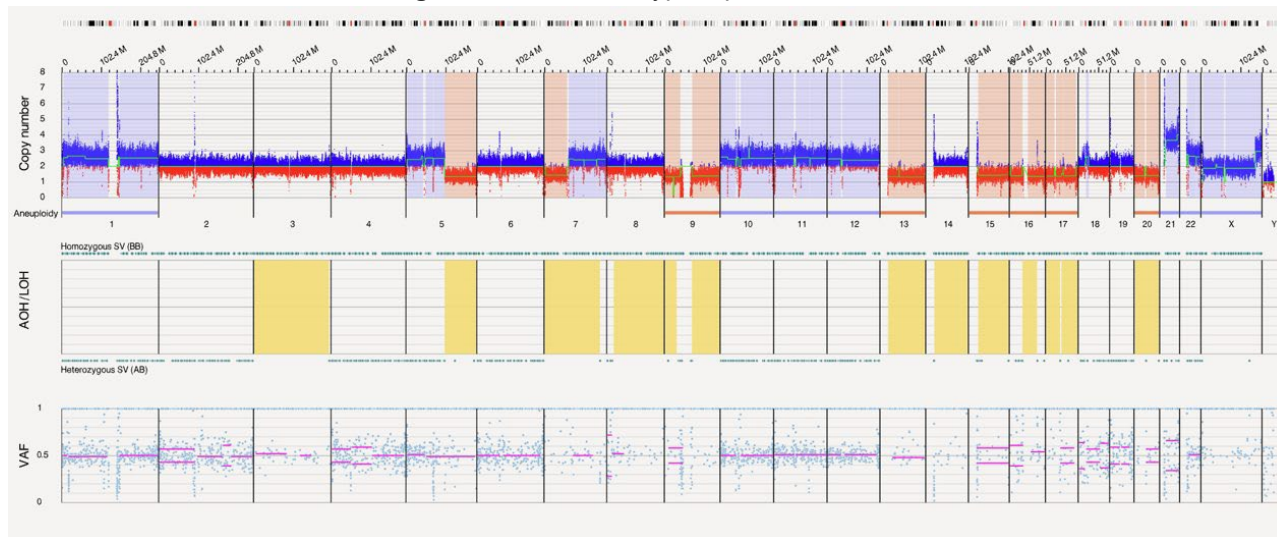


Figure 36. De novo analysis of a hyperdiploid event

HYPODIPLOID

Using the whole genome view, the total number of chromosomes can be assessed. In **Figure 37**, the number of losses, with a few gains, will average approximately forty-one chromosomes. Due to the mosaic nature of losses

and gains, there could be more than one clone in this cancer sample.

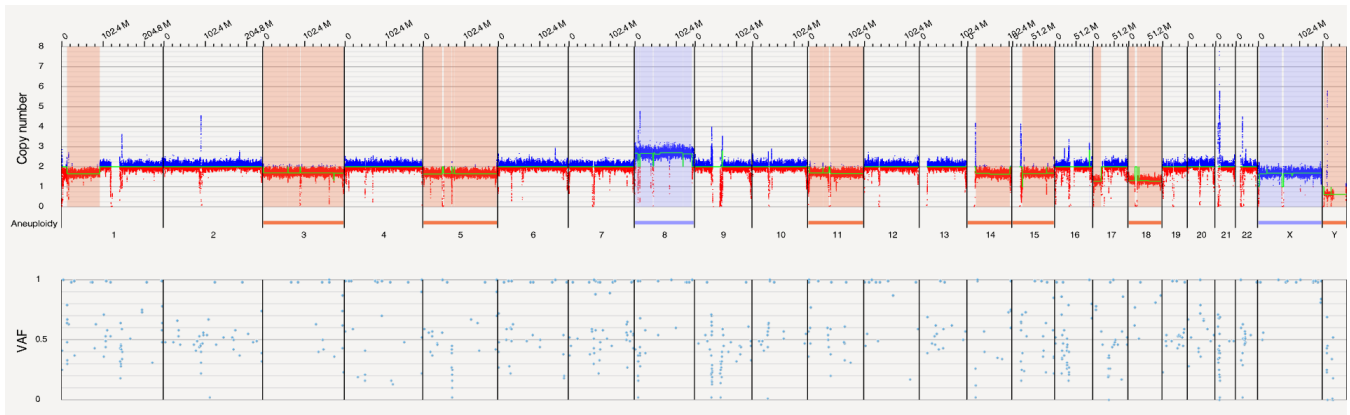


Figure 37. Hypodiploid example

TETRAPLOIDY-SOLID TUMOR

As discussed earlier in the section “Absence of Heterozygosity (AOH) and Triploidy,” ploidy differences from diploid genomes can be inferred. For tetraploid genomes, multiple 4-copy gains in the whole genome view suggests a doubling event. A *de novo* assembly can be used to determine if the two copies contain LOH. In **Figure 38**, chromosomes 1p, 3, most of 4, 5, 6, 8-11, 13, 15-18, and 22 suggest two copies with LOH and, in relation, chromosomes 2, 7, 12, 14, 20, and 21 are four copies and chromosomes 1q, 19, and X are three copies.

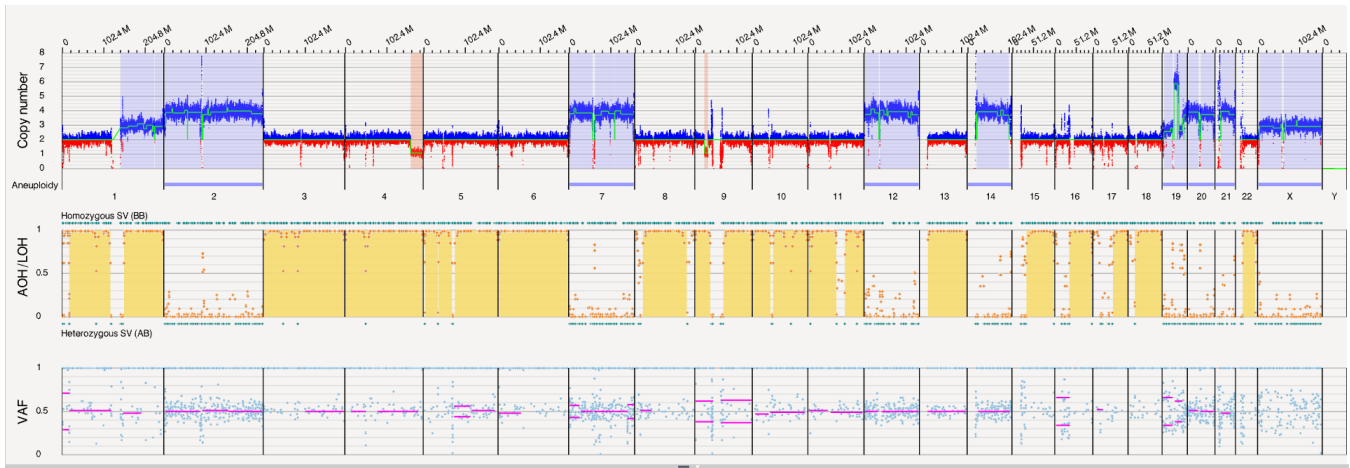


Figure 38. Tetraploidy example

Manually Checking Regions

Certain genomic regions should be reviewed for variants that have been affected by low confidence scores or the controls database (see examples of Hematological Regions in **Table 1** and Constitutional Regions in **Table 2**). Manually assessing the maps, alignment, molecules, and the CNV track can assist in determining the validity of variants in difficult regions.

Table 1. Manually Checked Regions – Hematological Regions

Hematological Regions
7p12.2 – IKZF1

14q32 - IGH
Xp22.33 – PAR1, <i>CRLF2</i>

Table 2. Manually Checked Regions – Constitutional Regions

Constitutional Regions (hg38)
chr15:22,787,850-23,051,049 – deletion <i>NIPA1/2</i>
chr16:21935203-22455963 – deletion 16p12.1
chr16:29595531-30188534 – duplication 16p11.2
chrX/Y:335,785-1,419,488 – deletion PAR1
chrX:154396222-154653579 duplication Xq28

7p12.2 DELETION - *IKZF1*

According to the *Bionano Solve Theory of Operation: Variant Annotation Pipeline* (CG-30190), five molecules of support are recommended for putative SVs. **Figure 39** shows an *IKZF1* deletion that was filtered out due to four molecules of support. Manual inspection of the gene region revealed the deletion.

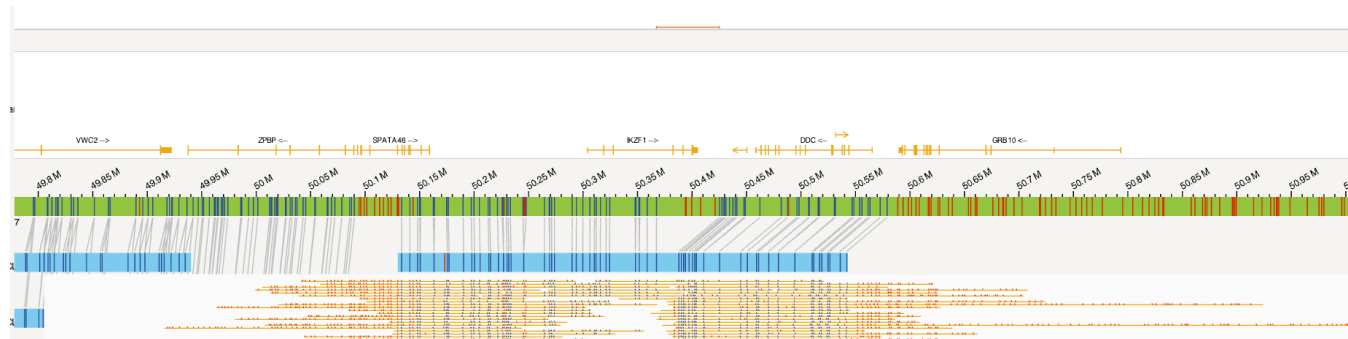


Figure 39. *IKZF1* deletion

16p12.1 DELETION

In **Figure 40**, combining the SV and CNV variant information, a deletion in a known microdeletion region is detected. This deletion was found in ~20% of the controls database.

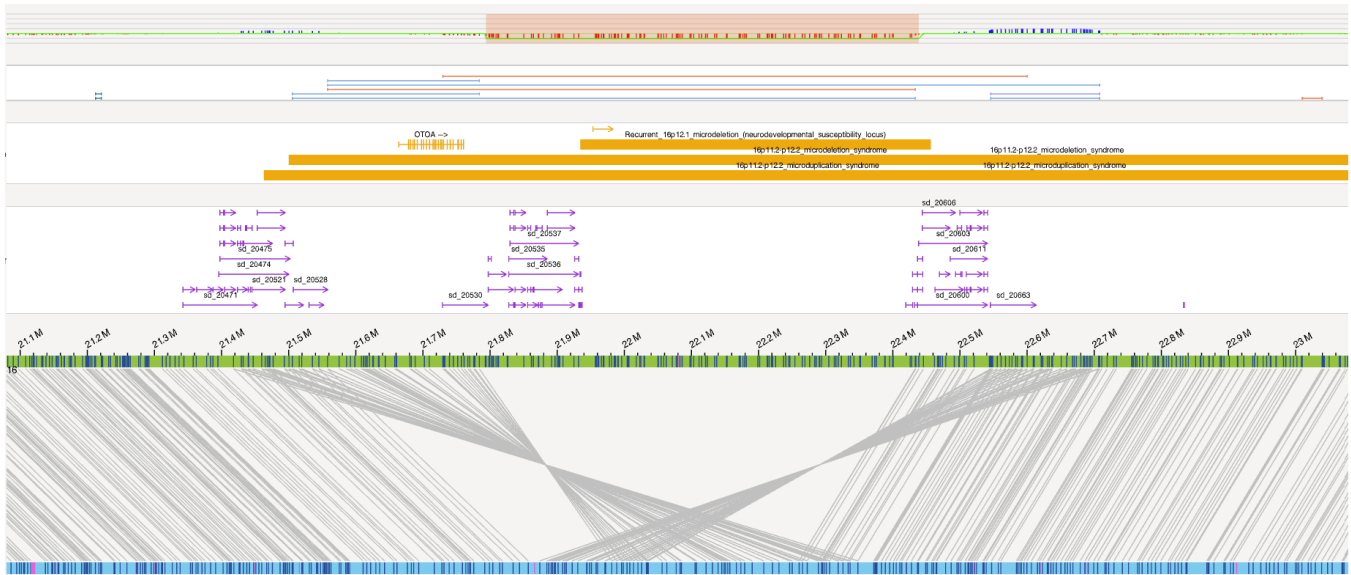


Figure 40. 16p12.1 deletion

16p11.2 MICRODUPLICATION SYNDROME

Segmental duplications can affect map alignment and the calling of SVs that have passed recommended filtering. As shown in **Figure 41**, an SV was not called even though a duplication map was created. Using the copy number variant in the CNV track assists in map interpretation to reveal a true duplication.



Figure 41. Segmental duplications

22q TERMINAL DELETION - SHANK3

In **Figure 42**, no SV or CNV call was made to indicate a deletion. Closer inspection of map molecules shows a group of blunted molecules indicating deleted DNA (green arrow).



Figure 42. 22q terminal deletion

Xq28 MICRODUPLICATION – FEMALE (CARRIER)

Similar duplications to the duplication shown in **Figure 43** are present in the controls database. If the reference allele is “ABC,” then the real inverted duplication looks like “AB~~B~~C”. As shown below, there should be distinct A-B and B-C junctions in the molecules.

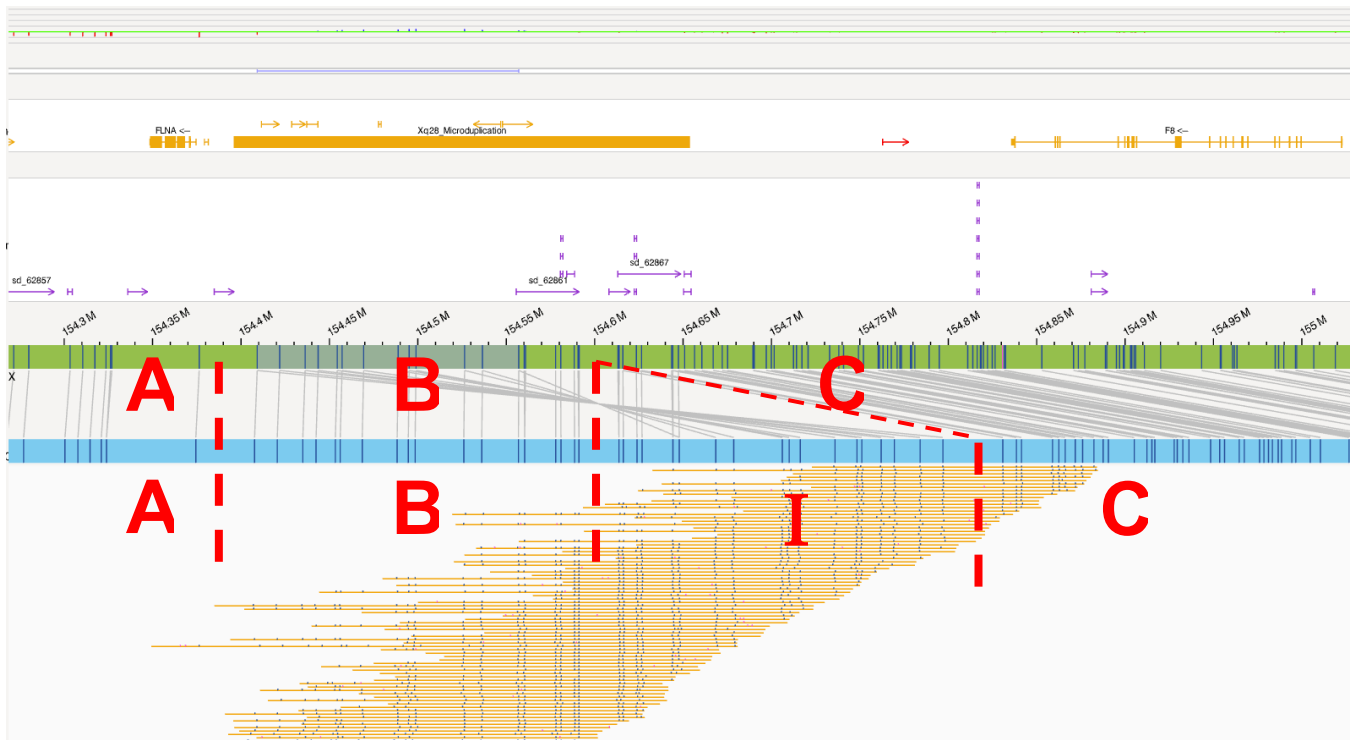


Figure 43. Microduplication in a female carrier

XQ28 MICRODUPLICATION– MALE (AFFECTED)

Figure 44 shows a similar Xq28 duplication in a male. Assessing the validity of the variant is conducted in the same manner as for females. Due to only 1 X in males, the gain in the coverage track is more pronounced though with no CNV called.

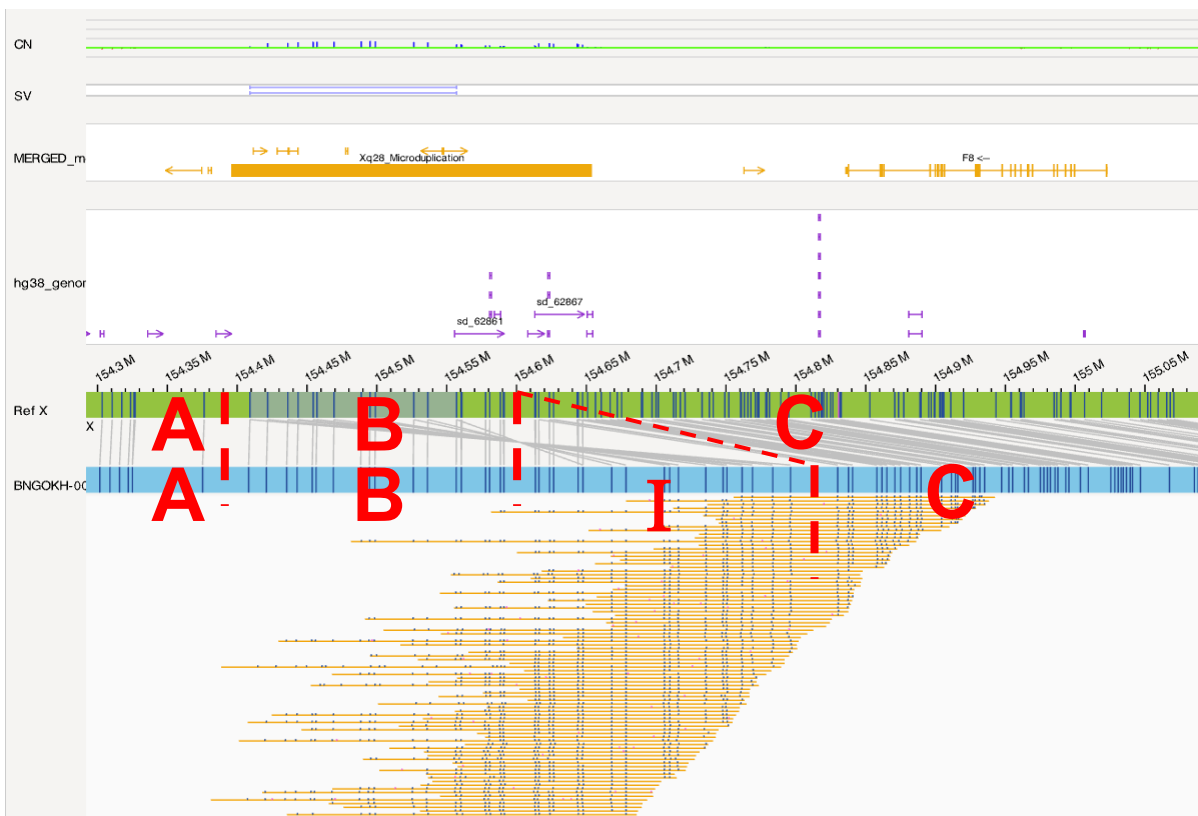
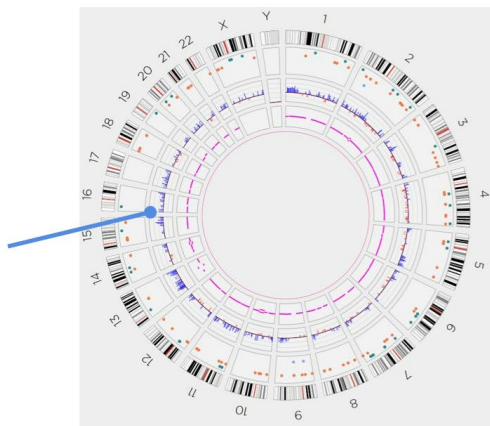


Figure 44. Microduplication in a male carrier

Interpreting Conflicting Variant Information

Multiple small gains and losses in Circos plot and whole genome views indicate poor quality data. In addition, out-of-range quality control variation metrics warn of systematic bias.



Assembly CNV Statistic	Value
Percent above expected (2 Mbp window) <20% acceptable	43%
Percent above expected (6 Mbp window) <20% acceptable	80%

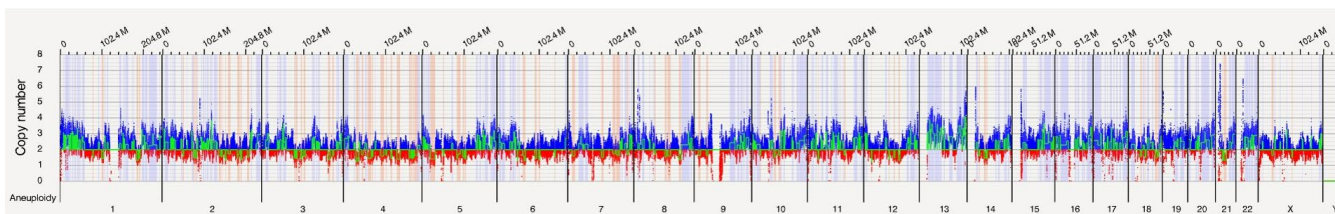


Figure 45. Noisy copy number data

In **Figure 45**, multiple small copy number changes are seen in the Circos plot and in the whole genome view, coupled with poor CNV statistics indicating noisy copy number data. Caution should be used when interpreting copy number changes. In **Figure 45**, it is impossible to distinguish true copy number change from background noise; therefore, the CNV data is unusable. CNV noise is usually due to ineffective removal of DNA-bound protein (DLE-1 or otherwise) from the DNA molecules, causing bias as to which molecules can travel into chip nanochannels. Using fresh, properly stored Proteinase K usually solves the problem. **NOTE:** the quality of the SV is not affected by such noise in coverage depth.

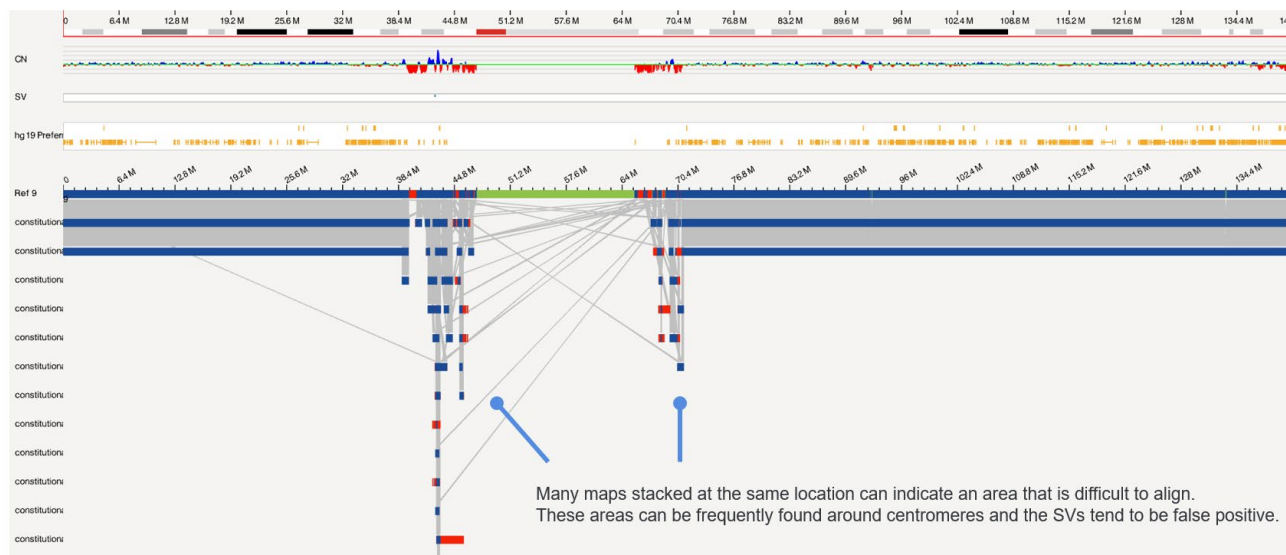


Figure 46. Stacks of maps

As seen in **Figure 46**, in the genome browser view, multiple stacked maps indicate the presence of large low copy repeats that have ambiguous and/or non-unique mapping, seen in repetitive areas of the genome and especially around centromeres. Caution should be used interpreting SVs in these areas.

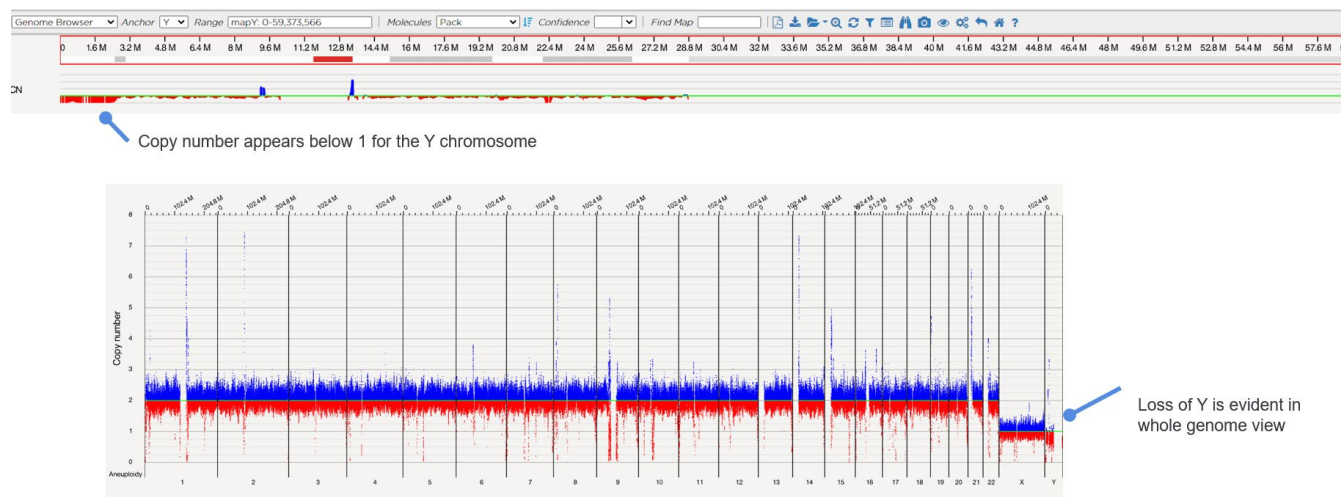
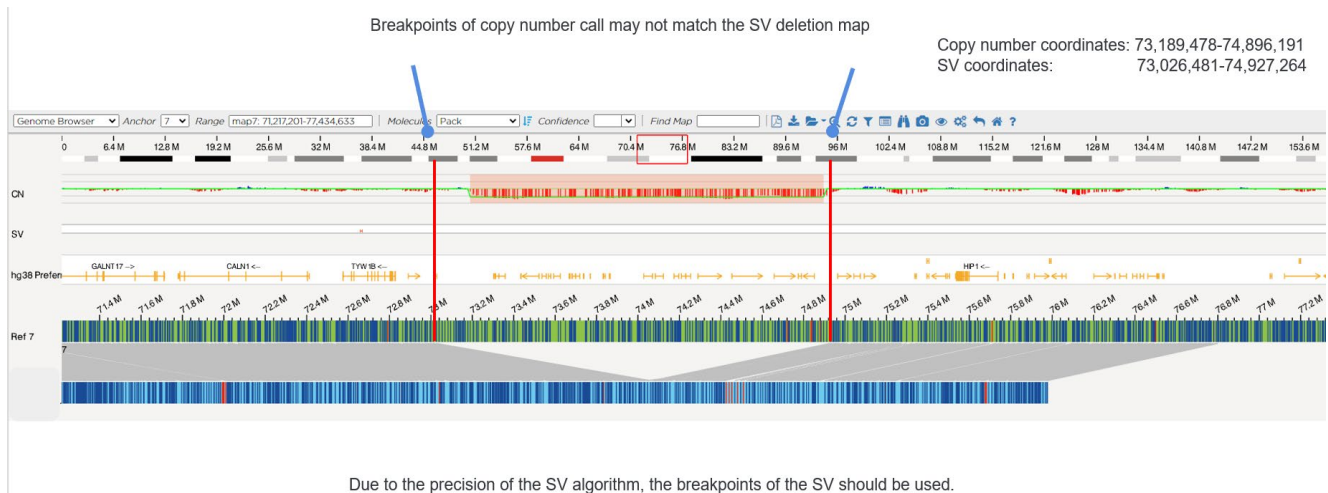


Figure 47. Masked variants – Y chromosome

The variability of coverage profile in chrY in control samples is substantially higher relative to other chromosomes. In fact, a sizable portion of the Y chromosome is masked due to high variance. Aneuploidy can still be called in such variable regions, but users will have to ignore the mask and use the confidence score as a guide to visually look for actual events. In **Figure 47**, the coverage of chrY is notably lower than the other chromosomes, and the confidence scores (scores of 1) pass the recommended threshold. The chrY loss is a true positive.



Due to the precision of the SV algorithm, the breakpoints of the SV should be used.
Figure 48. Differences between copy number tracks and SV map breakpoints

As indicated by the red bar in the copy number track and the deleted sample map in **Figure 48**, copy number changes >500 kbp should be picked up by both algorithms, except for terminal deletions and duplications. However, the two algorithms have called different breakpoints in this example. In general, the SV algorithm breakpoints are more precise and should take precedence over the CNV coordinates.

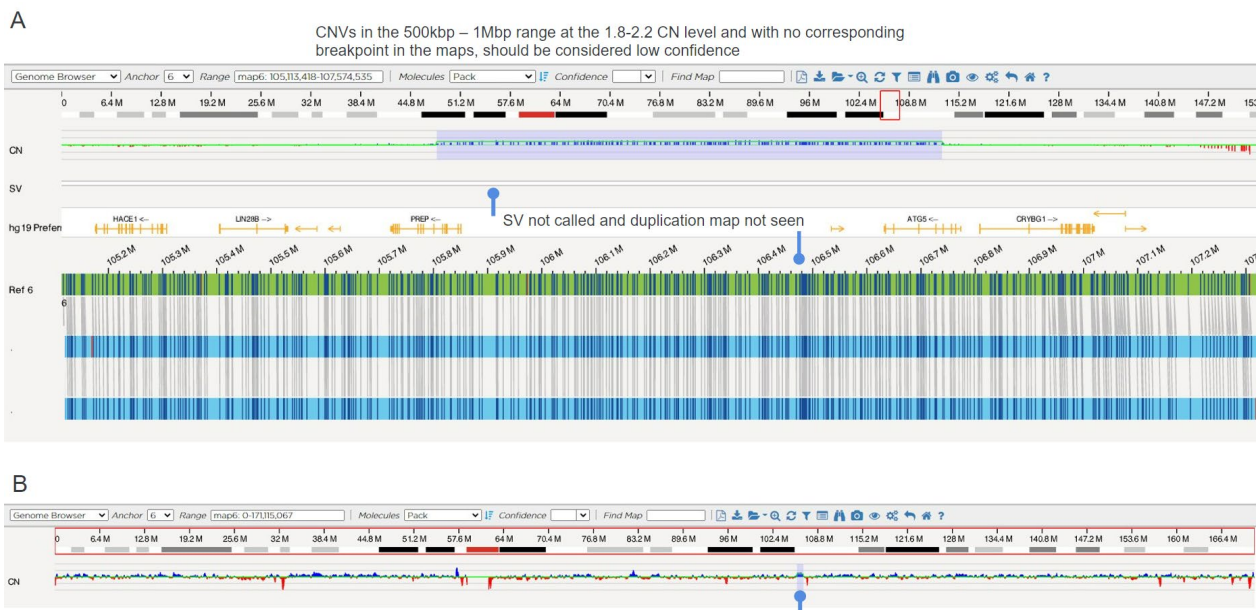
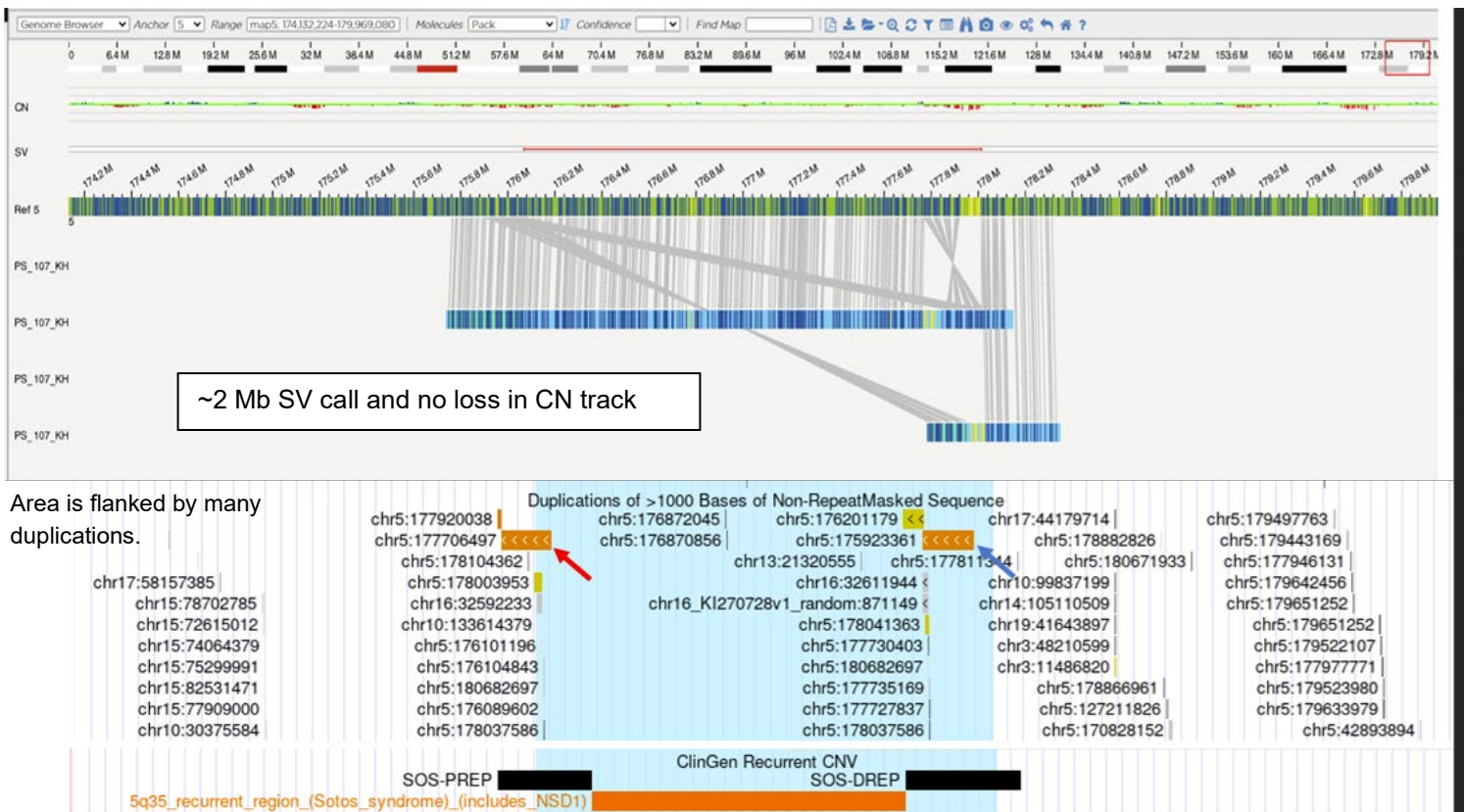


Figure 49 A and B. False positive copy number call

Figure 49 A. Here, a 933.1 kbp copy number gain (with a copy number of 2.470) has been detected in a constitutional/non-cancer sample. However, this example is likely to be a false positive. First, this is a constitutional sample, and a non-whole number copy number gain (i.e., 2.47 rather than 3) is unexpected. Second, interstitial copy number variants >500 kbp should be detectable by both the CN and SV algorithms. It is unlikely that the SV algorithm would miss a gain occurring at a cellular fraction of 47%. Finally, as seen in **Figure 49 B**, the coverage gain of the call is not notably higher than the rest of the chromosome.

False Positive in the Sotos Region of 5q

In **Figure 50**, a ~2 Mb deletion is detected by a map and shown in the SV track while there is no deletion in the CNV track. As shown by the genome browser, this area is flanked by duplications which can cause misalignment. Combing the information provided by the CNV and SV algorithm, false positive deletions in the Sotos region can be identified.



Misaligned translocation maps may be detected by shorter length, many maps in the area, other maps aligning to only one chromosome, no reciprocal map, and similar label pattern to one chromosome may misalign



Figure 50. False positive translocations

In the genome browser view, translocation maps should be reviewed to determine validity. As seen in **Figure 50**, misaligned maps do occur. The map that is called the translocation (or intra-chromosomal fusion) is a short map, and the alignment to 174 Mbp of Chr1 consists of only a few labels. Furthermore, that label pattern is mostly identical to the longer map just above, which called an inversion based on a much more robust alignment. Therefore, the translocation is likely a false positive due to a misalignment to 174 Mbp and the inversion is likely valid.

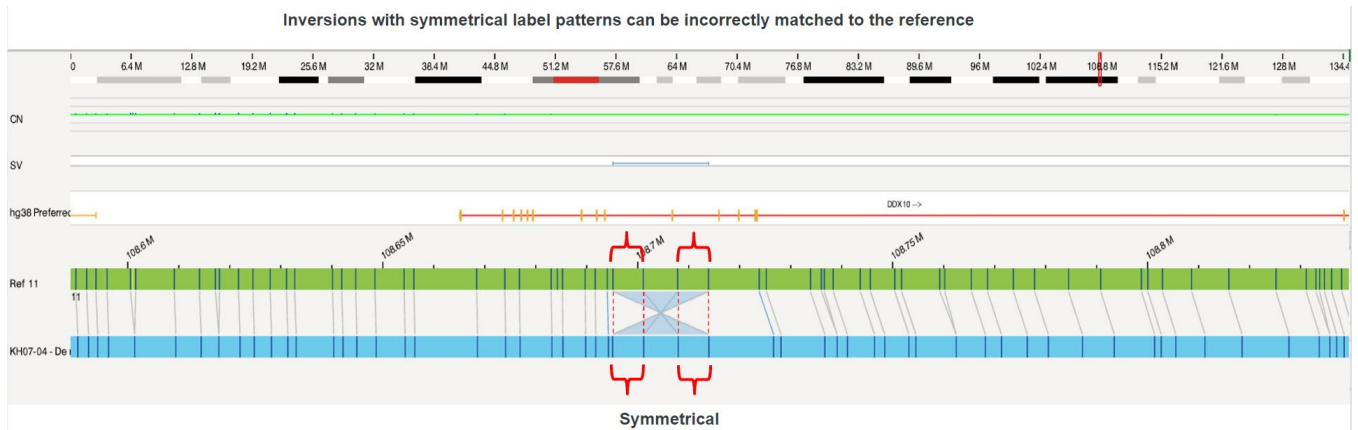


Figure 51. False positive inversions

In the genome browser view, the pattern for inversion maps should be reviewed for validity, as incorrect matching is seen in **Figure 51**. Care needs to be taken when there is a perfectly symmetrical label pattern in inversions. Here, the inverted alignment could have been aligned vertically with the reference, resulting in a simple insertion event without the inversion.

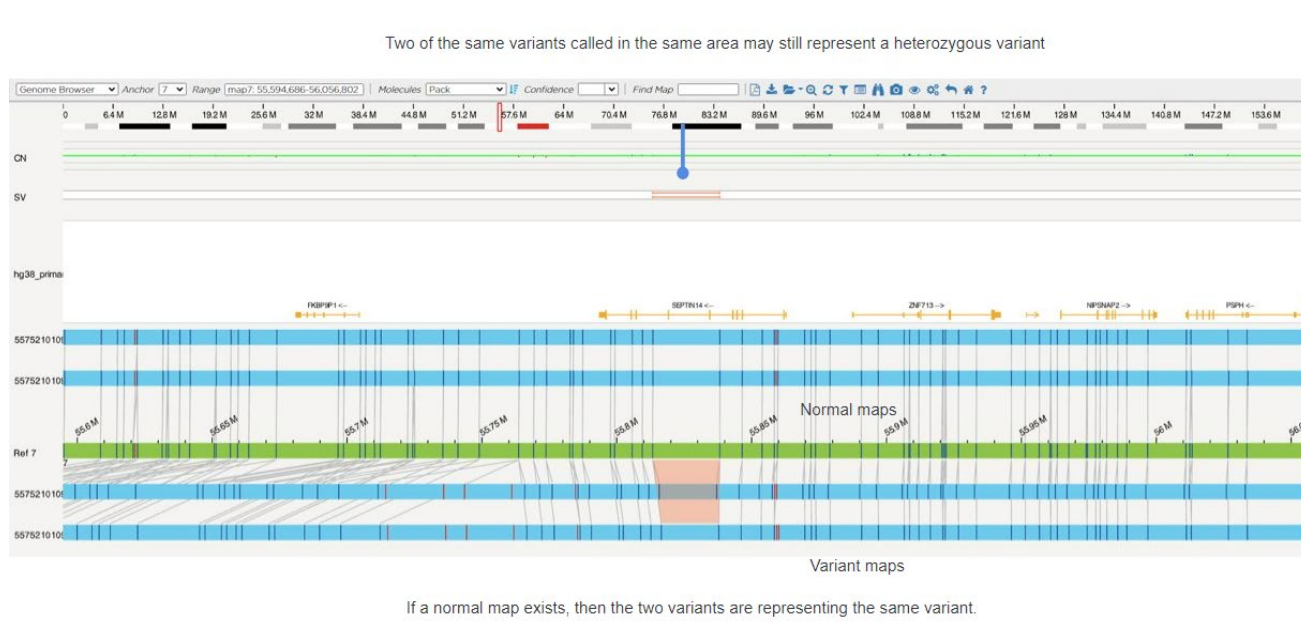


Figure 52. The homozygous variant is heterozygous.

To rule out redundant maps indicating homozygosity, view all maps for the region in the genome browser view. **Figure 52** illustrates this point, where two deletions are called by the two variant maps in the bottom. However, the deletion is heterozygous because there is a reference allele captured by two normal maps. Due to over-splitting (in both the *de novo* Assembly and the Rare Variant pipelines), redundant maps representing the same allele may be produced. Therefore, two of the same variants called in the same area may still represent a heterozygous variant. Refer to the zygosity or variant allele fraction value to infer zygosity.

Technical Assistance

For technical assistance, contact Bionano Genomics Technical Support.

You can retrieve documentation on Bionano products, SDS's, certificates of analysis, frequently asked questions, and other related documents from the Support website or by request through e-mail and telephone.

TYPE	CONTACT
Email	support@bionanogenomics.com
Phone	Hours of Operation: Monday through Friday, 9:00 a.m. to 5:00 p.m., PST US: +1 (858) 888-7663
Website	www.bionanogenomics.com/support
Address	Bionano Genomics, Inc. 9540 Towne Centre Drive, Suite 100 San Diego, CA 92121