Analysis Quick Start: Annotated Rare Variant Calling

Document Number: 30375

Document Revision: A

Table of Contents

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 in 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®, Irys®, IrysView®, IrysChip®, IrysPrep®, IrysSolve®, Saphyr®, Saphyr Chip®, Bionano Access®, 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.

Revision History

Revision	Release Date	Notes
A	03/12/2021	Initial release

Rare Variation Detection - Introduction

Bionano Access supports different workflows for the detection of structural variants in human genome (Figure 1). As per Bionano technology, a rare variant is defined as a variant that is present in low abundance in the sample and are not present in the reference molecule. To identify such variants effectively, a dedicated bioinformatics pipeline is used that locally aligns molecules to a reference, assembles those with putative differences into consensus maps, and identifies the structural variations with less computational burden. For more information on each of these workflows, Theory of Operation :Structural Variant Calling - 30110, Theory of Operation - Variant Annotation Pipeline - 30190, Theory of Operation - EnFocus FSHD Analysis - 30321. To obtain information on data coverage targets, which can vary depending on the analysis, see Data Collection Guidelines – 30173. To obtain information on how Bionano determines quality control of raw data, please refer to Bionano Access Dashboard Guidelines – 30304 and Molecule Quality Report Guidelines – 30223.
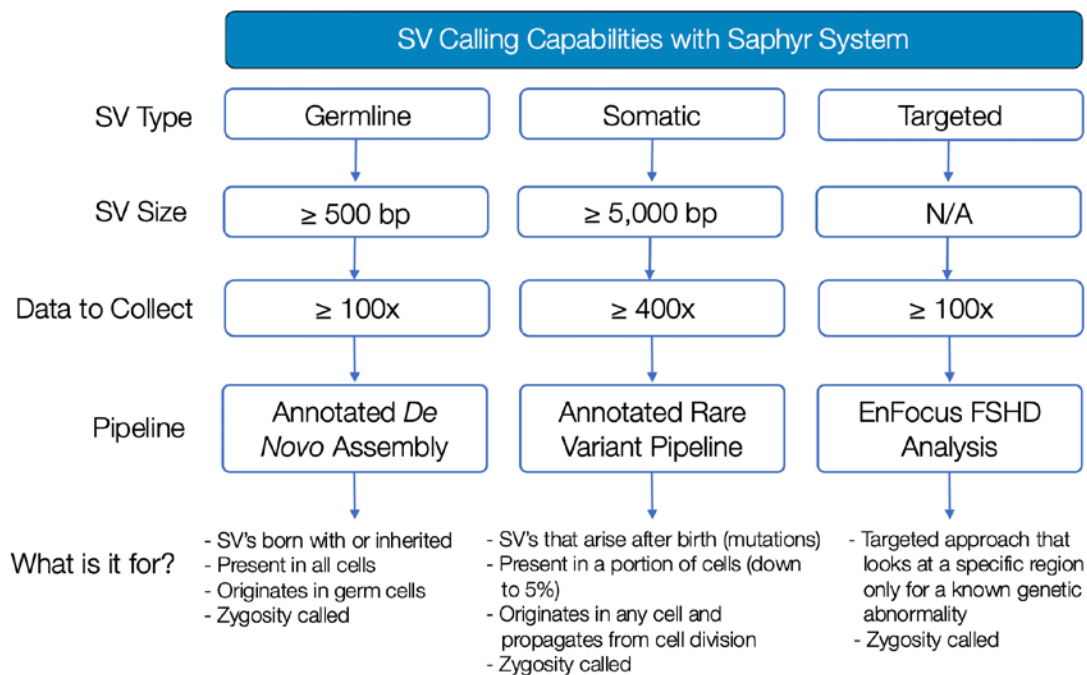
- Zygosity called

Figure 1- Schematic representation of the Bionano structural variant detection workflows. Variant Analysis Pipeline- includes first constructing a de novo map and then comparing the resulting map with a reference which is highly sensitive in detecting

Workflow Overview

This workflow discusses how to filter and narrow your search for high confidence SVs that were identified using the Rare Variant Pipeline (RVP). Once DNA is isolated and labeled, Bionano chips are loaded and run on the Saphyr instrument. Molecules and labels imaged on the Saphyr instrument are digitized and saved as in a BNX format (raw data). This serves as input for downstream analyses. RVP detects SVs by the "split-read" analysis, where initial molecule alignment and molecule extension refinement, allows identification and detection of somatic variants, and identifies variants that are 5,000 bp or longer in size. For more information on variant annotation Pipeline and relevant file formats, read SV Annotation Pipeline file or BNX , XMAP and SMAP file format specification sheet.

To access the results from your RVP analysis, the user begins by logging into the web server location which contains your data. Once logged in, the user navigates to the project which contains the data for the RVP experiment of interest. RVP can be used for a single sample, duo (matched tumor-normal for example) or trio (parents-child for example) analysis. For any individual experiment, clicking on the sample data will take the user to a circos plot which has a global overview of the mapped genome.

The user may then perform filtering of the rare variants and identify genes of interest that might been affected by a rare variant event. Additional filtering and masking of low confidence calls may be performed using a BED file. Once the dataset has been filtered appropriately the user can visually explore the structural variants, create a report and then export the data for further bioinformatic analysis. For more details on the thresholds and experimental designs refer to the Theory of Operations Variant Annotation Pipeline document.
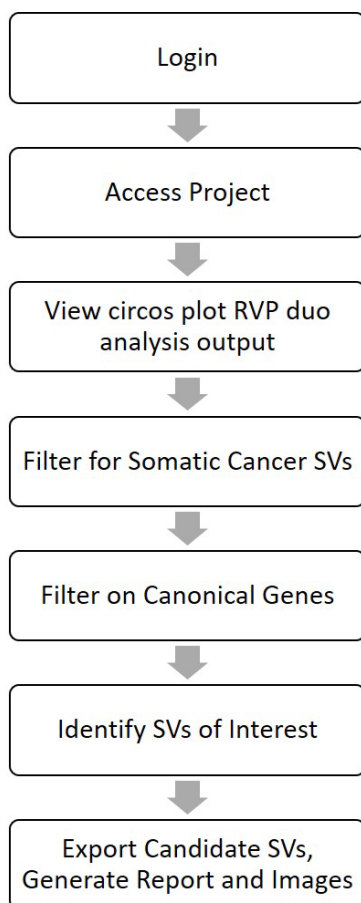

Figure 2- Bioinformatic workflow for post RVP data analysis in Bioanaoaccess

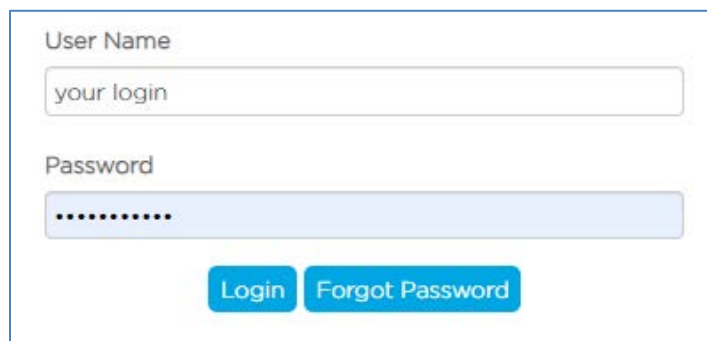# Accessing Bionano Data and Filtering with Recommended Settings

A. GOAL OF THE TUTORIAL

This quick start guide directs users on how to view annotated rare variants, which are detected from Bionano's genome mapping data. It can be used to evaluate your own data or familiarize yourself with the software using a pre-loaded demo dataset. The demo dataset is provided by Bionano to help users get familiarized with the Bionano Access software interface that enables analysis, visualization, and data export. To walk through the analysis, follow the numerical prompts in this guide. Text outside of the numerical steps is provided to facilitate learning. This guide assumes all necessary analyses pipelines have been run on your samples, and so this tutorial will guide you through selecting the proper data, and performing filtering based on our current guidance before exporting the data for presentations or further analysis outside of Bionano Access. The filter sets used here represent our current recommended settings but may be adjusted to fit the needs of a project. In the hands-on section of the tutorial, we will walk you through the analysis of tumor (HCC2281) and non-tumor (HCC2281BL) cell lines established from the same patient with breast cancer. The goal is to identify rare somatic variants in HCC2281 breast carcinoma cell line when compare its matched normal cell lines.

To obtain information on data coverage targets, which can vary depending on the analysis, see Data Collection Guidelines – 30173. To obtain information on how Bionano determines quality control of raw data, please refer to Bionano Access Dashboard Guidelines – 30304 and Molecule Quality Report Guidelines – 30223. To obtain information on ways by which bioinformatic analysis of annotated structural variants are recommended, see Data Analysis Guidelines.

B. ENTERING BIONANO ACCESS

1. Using a web browser (Google Chrome is recommended), navigate to the web address of your Bionano Access data (e.g.: https://us.bionanoaccess.com). Users with an installed Saphyr System will utilize an Access Installation hosted on their local Bionano Access Server. Enter your User Name and Password, then press "Login" (Figure 3). To access the demo dataset or user's sample data generated by the Bionano Commercial Services Lab, a user will need login credentials as provided by a Bionano representative.

Figure 3– Login to Bionano Access

2. To open your sample *data* or demo project *data*, click on "Projects" on the home page (Figure 4*).* A "Project" is the central repository for all objects such as experiments and any bioinformatic analysis that is done.

Figure 4– Bionano Access login screen.

3. In the "Projects" view, click on the hyperlinked project name corresponding to your submitted project, or "BionanoAccess_somatic" to follow along with this demo dataset (Figure 5). This will open the list of data types (Objects) within that Project (Figure 6).

Figure 5 – Bionano Access Home page - contains data repository in "Projects".

The Project Data page (Figure 6) will list the various data types for a given sample or analysis. For rare variant with variant annotation pipeline, your project may contain some or all the following file "Types."

a) **Molecules** – the raw, unassembled, molecules data (also known as a .bnx filetype). Molecules file is generated by the image analysis software. It contains single-molecule optical mapping data generated on Bionano Saphyr instrument and is the first data type produced. May or may not be present, not essential for visualization.

b) *De Novo* **Assembly** - the assembled consensus genome maps (also known as a .cmap filetype). *De novo* assembly file contains the assembled consensus genome maps which are then compared against a reference assembly (e.g. hg19 or hg38) to produce un-annotated SV calls. May or may not be present.

c) **Rare Variant Analysis –** The results from the rare variant annotation pipeline are output to a text file with/without annotation for each SV detected for each sample.

d) **Annotated Rare Variant Analysis-**This file is unique to a duo analysis. The important thing about this filetype is that it is 1) annotated 2) compared to the internal control database. For duo analysis, the calls are then compared to each other and contains the rare variants identified in the cancer sample compared to its matched control samples.

On highlighting each file type, different operations become visible under the options pane on the right.
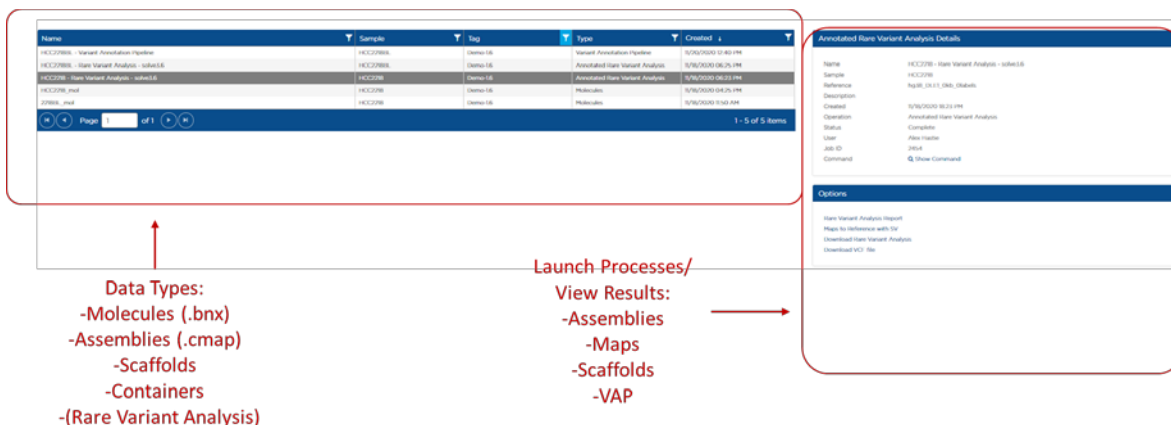
Figure 6– Project Data page. This contains a list of object(s) and list of operations (options) that can be performed with the selected object.

C.   VARIANT ANNOTATION PIPELINE

In this tutorial, you will find the project contains at least three different files-

1. HCC2218- Rare Variant Analysis - solve3.6
2. HCC2218BL - Rare Variant Analysis - solve3.6
3. HCC2218BL-Variant Annotation Pipeline

The rare variant calls from each sample (HCC2218and HCC2218BL) are saved as separate files. The third file includes the dual annotation file which includes all rare variants identified in the samples compared to an internal database of more than 200 normal apparently healthy individuals. In this tutorial, the dual analysis file is used to identify rare somatic variants.

Figure 7-Selecting the Variant Annotation Pipeline Data Type (file Type). Each column lists the auto-generated sample name, given sample

To view a sample, click on the sample name. Depending upon your experimental design, the content may be different. To identify the files used to generate the dual analysis, click and highlight "HCC2218BL-Variant Annotation Pipeline" as shown in Figure 8. Check the "Name" under "Container Details". For more information see Variant Annotation Theory of Operation – 30190.
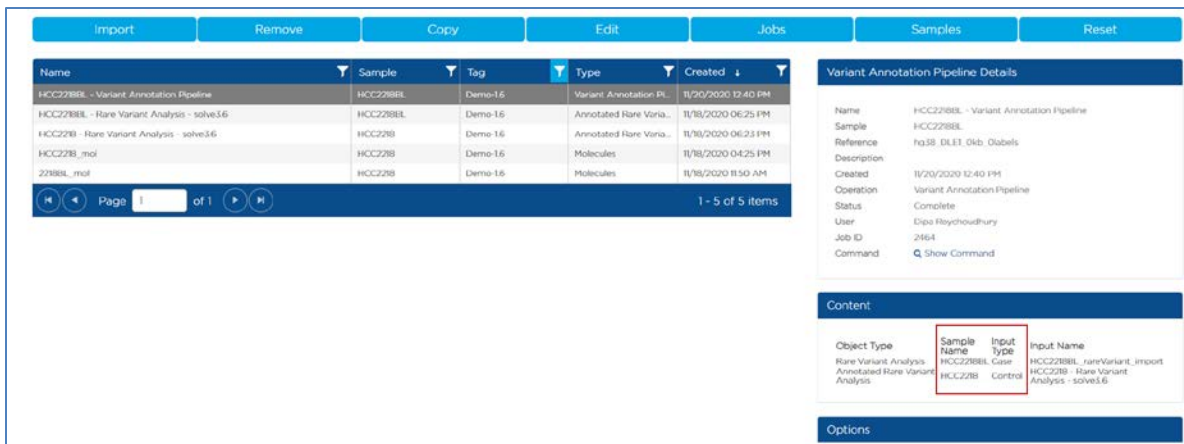
Figure 8 ─ Dual Analysis File-The content lists the samples compared in this analysis.
The first operation under "Options" is "View variant annotation results". This launches the circos plot.

4. Click on "View variant annotation results" to launch circos plot.

D.   CIRCOS PLOT VIEW

You can view one sample or file type at a time in the circos plot view. In a duo analysis, all SVs detected in normal vs tumor sample that is compared against the control database can be viewed on a circos plot. Using the filer settings as discussed below, you can display SVs unique to tumor or normal. The control database

In Bionano Access, the Circos Plot is displayed by default for human assemblies and variant annotations. The Circos Plot is displayed as the circular plot on the left, with "Filter Criteria", "SV Filter Confidence scores" "BED Files", and "Samples" on the right. To zoom on the circos plot, rotate your mouse-wheel forward (scroll it down). The user can also click and drag the plot around using the mouse.
Note: For more mouse and keypad short cuts to maneuver the circos plot, click on the "?" icon which is the last icon on the top panel.
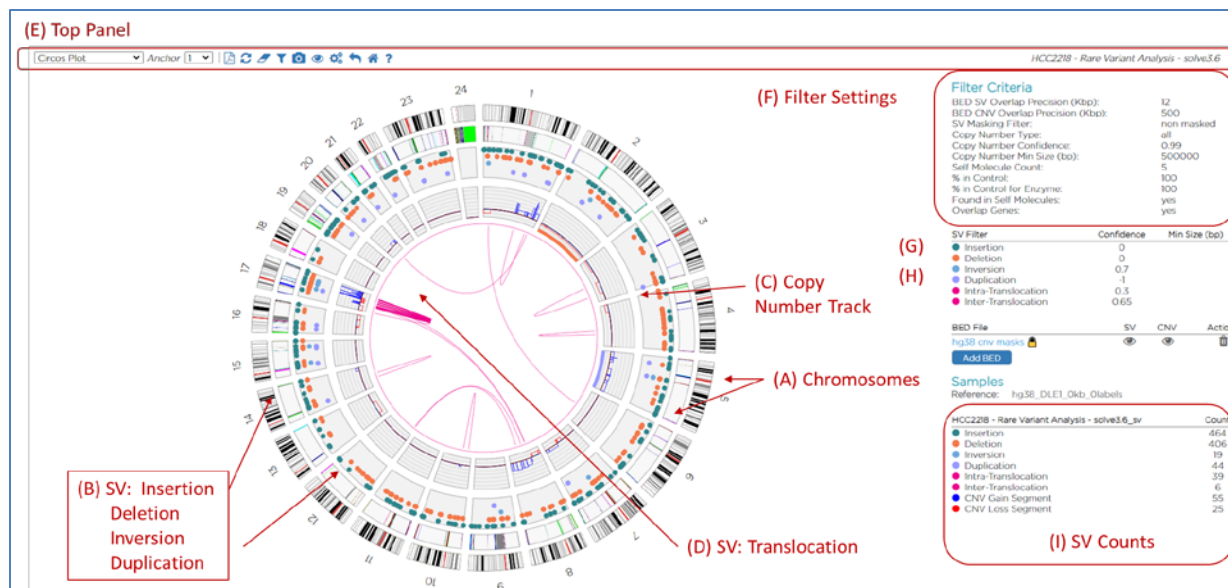
Figure 9 — Circos plot view includes a top panel with filtering options (E), options to add BED files (G), Sample Name (H), summary SV counts (I), and the central circos plot with the different tracks (A-D)

In a circos plot (Figure 9), each chromosome is positioned end-to-end-in a large circle. Each track represents a different type of SV or annotation result. Clicking on the cytoband, SV, or copy number track, will open the genome browser and show the user a detailed view of Bionano consensus genome map for that locus. The user can navigate back to this circos plot from other views by clicking on the circle icon - ○ - on the Top Panel (E) toolbar. The rings of the circos plot are described from the outermost track to the inner most track:

- The outermost numerical track (A) corresponds to the chromosome number, with cytoband information shown in the black-and-white banding pattern. Currently, Chromosome 23 corresponds to X and Chromosome 24 to Y.

- The next track (B) includes the detected SVs as represented by color-coded dots. A user can hover over the dot to view info about a given SV call.

- The next track (C) shows copy number calls. For the copy number track, the baseline state of autosomes in a diploid genome is a copy number of 2. If a region shows a copy number gain, the line will move outward from 2 (colored blue), reflecting an increase in the local copy number state. If the region has a loss, it will move inward (colored red). See Introduction to Copy Number Analysis – 30120.

- The innermost track (D) contains information about translocations, both inter-chromosomal and intra-chromosomal (intra-chromosomal translocations are defined by an abnormal fusion within one chromosome with breakpoints at least 5 Mbps apart). The lines (colored magenta) going from one chromosome to another are drawn to show chromosomal rearrangements.

The top panel (E) of Figure 9 includes "Anchor" (which corresponds to the chromosome number).

Note: Chromosome X=map 23, Y=map 24, and changing numbers using the drop-down takes the user to the genome bowser view for any selected chromosome. "Range" (which corresponds to coordinate ranges on the genome) allows the user to view SVs of interest by specifying genomic coordinates (refer to Section E).

By default, when using Access for the first time, all SVs detected can be visualized in the Circos Plot. The "Filter Criteria" (F) lists the default settings used to detect rare variants in the analysis pipeline. Once changed, they will remain at that setting going forward until they are changed again or reset to default.

Note: In Bionanoaccess, "Masking" refers to genomic sequence that has been scanned for either N-base gaps, sub-centromeric and sub-telomeric regions that are prone to generating putative false positive translocation breakpoint calls. Apart from these, regions with putative SVs such as translocation breakpoint can be masked during the alignment step in Solve. Access allows users to view these annotations and regions during the SV filtering steps. In this tutorial, the default settings were used to look for SVs making option.

E.   VISUAL COMPARISON OF CIRCOS PLOT (OPTIONAL)

5.   To compare the SVs detected in each sample, highlight the cancer sample HCC2218 (Figure 10).

6.   Under the "Options" section on the right, click on the second option "Maps to Reference with SV" (Figure 10). This will launch the circos plot.

Figure 10 – Launching the circos plot for HCC2218
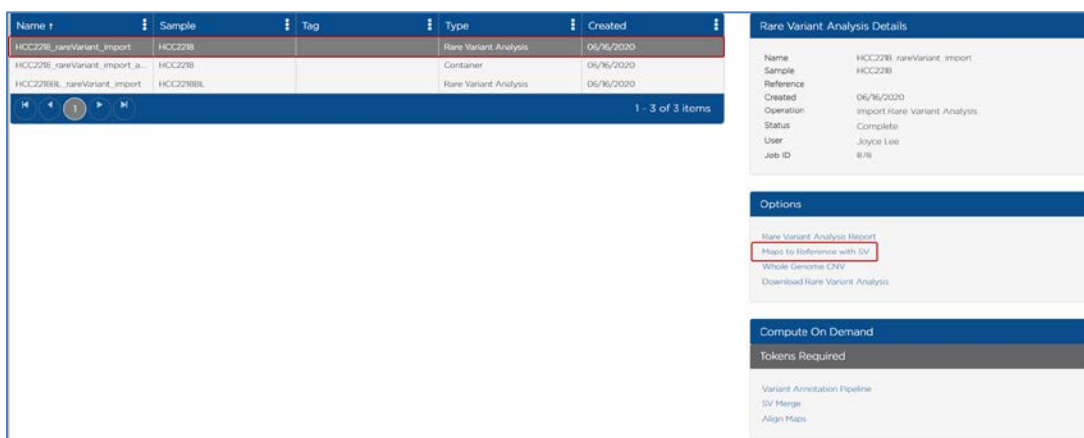
7.   Similarly, to view the circos plot for the normal sample (HCC2218BL), first click on the back icon on the top panel. This will take you back to the project page.
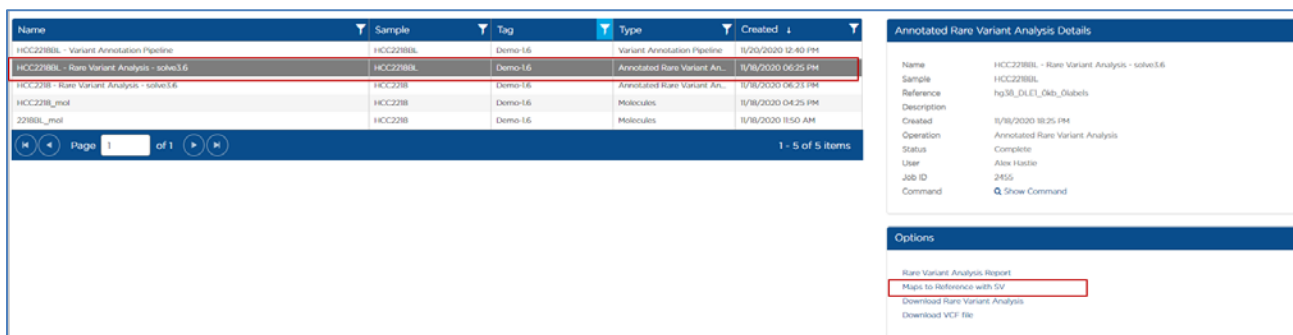

Figure 11 – Launching the circos plot for HCC2218BL

8.   Click on the sample HCC2218BL to highlight it. From the "Options" section, click on "Maps to Reference with SV" to launch circos plot (Figure 11).

HCC2218 - Rare Variant Analysis - solve3.6_sv	Count
● Insertion	464
● Deletion	406
● Inversion	19
● Duplication	44
● Intra-Translocation	39
● Inter-Translocation	6
● CNV Gain Segment	55
● CNV Loss Segment	25

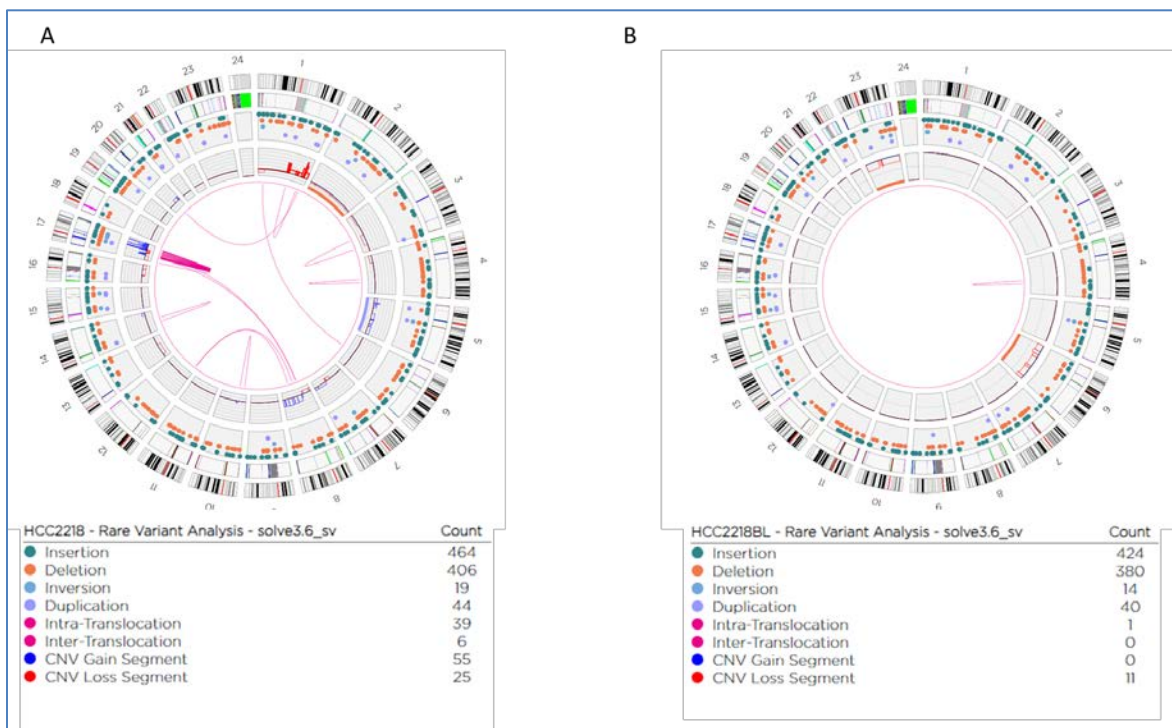| HCC2218BL - Rare Variant Analysis - solve3.6_sv	Count
● Insertion	424
● Deletion	380
● Inversion	14
● Duplication	40
● Intra-Translocation	1
● Inter-Translocation	0
● CNV Gain Segment	0
● CNV Loss Segment	11

Figure 12 — Comparison of the circos plots from two samples with Tumor cell line (A) and Normal cell line (B)

For ease of comparison, we have pasted the two circos plots side by side" as shown in Figure 12. Note the difference in the nature of the distribution of rearrangements between the two cell lines depicted. The whole genome profile on the left (A) has 147 translocation rearrangements that are heavily clustered on Chromosome 17. In contrast, no such rearrangements were found in the normal (B) cell line present in the profile on the right-hand side.

Note: Users can customize the enzyme colors, SV colors, maps (reference, genome) colors, minimum and maximum map height value, molecule and copy number height value, and others.
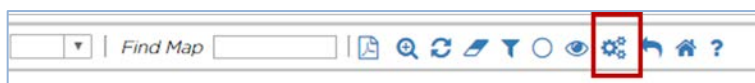

Figure 13 — Setting Options-This screenshot is from the top panel of the Circos plot. This includes Anchor and Range to navigate across the chromosomes, "Molecules" "Confidence" and "Find Map" to customize the visualization area. The settings icon in the red box customization of the circos plot.

9. To change the size of the SVs detected in the circos plot, click on the "settings" icon in the top panel as shown in Figure 13. Scroll down to find "Circos plot SV size" and change this to 2. Change "Circos plot translocation width" size to 1 (Figure 14). This increases size of variants in the SV track and the translocations in the innermost track. Click "Save"

Note: Changed settings parameters are saved under your username and across datasets, then retained when you open Bionano Access later. The user may select "Reset Options" to return to original default filters (Figure 14).
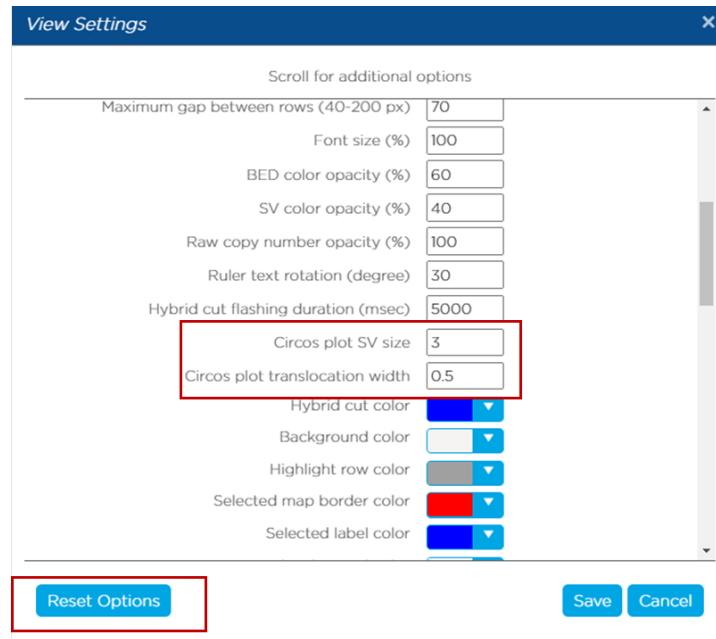
Figure 14 — Viewing SV events. This screenshot lists the customizable options that allows better visualization of the tracks. The user may select "Reset Options" to return to default settings

F. IDENTIFY SOMATIC VARIATIONS IN TUMOR-NORMAL COMPARISON

In general, SV filters can be customized to identify structural variants that are *de novo* or proband-specific in family studies or somatic SVs in cancer datasets. It can also be used to identify rare variants found infrequently in the control database within single individuals, or variants that overlap with annotated genes. In this tutorial, we are filtering for cancer-specific somatic variants. The dual analysis for RVP filters only those cancer sample specific variations. To remove germline calls, Access filters can be adjusted to identify tumor-specific somatic variations from dual annotation file called "HCC2218_rareVariant_import_annotationDual file".

10. Click on either the back arrow or the home page to navigate back to the "Project Browser" (Figure 15).

Figure 15 — Navigate back to the Project data page to find the container file " HCC2218_rareVariant_import_annotationDual file "

11. Follow instructions as described earlier to launch the circos plot page for "HCC2218BL-Variant Annotation Pipeline"

Filter Criteria

BED SV Overlap Precision (Kbp):	12
BED CNV Overlap Precision (Kbp):	500
SV Masking Filter:	non masked
Copy Number Type:	all
Copy Number Confidence:	0.99
Copy Number Min Size (bp):	500000
Self Molecule Count:	5
% in Control:	100
% in Control for Enzyme:	100
SV Chimeric Score:	pass
Found in Control Sample Assembly:	no
Found in Control Sample Molecules:	no
Found in Self Molecules:	yes
Overlap Genes:	yes

SV Filter	Confidence	Min Size (bp)
● Insertion	0	
● Deletion	0	
● Inversion	0.7	
● Duplication	-1	
● Intra-Translocation	0.3	
● Inter-Translocation	0.65	

Figure 16 — Filter Criteria for SVs are filtered on user defined criteria. By default, stringent confidence score settings are used to detect high confidence SVs.

Tip: To ensure none of the additional filtering options are active, check the filter criteria listed above (Figure 16) and click on "Reset Filters" before you choose the recommended filters.
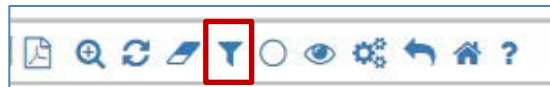

Figure 17 — Filter Icon for SVs

12. To choose filtering options, click "SV Filter" funnel icon (Figure 17).

13. The filter dialog window opens showing the "Filter by SV Type" options (Figure 18).

In general, this wizard allows the user to set rules for identifying variants relevant to their data based on confidence of a call and the length of the detected SV. The confidence score of a detected SV is defined as the probability of the SV being a true positive call. The positive predictive value (PPV) is a score that ranges between 0 to 1. The recommended value is set to 0 for insertions, deletions, and translocation calls. Paired Inversion calls are determined by the number of aligned labels within the inverted region and are assigned 0.01, 0.90, 0.98, and 0.99 for 2, 3, 4, and 5+ aligned labels respectively.

The filter dialog pop-up window contains four filter tabs-

- The "Filter by SV Type" tab allows basic filtering of types of SV using size cutoff and "recommended" or "all" confidence cutoffs. "recommended" confidence cutoffs balance false positive sand false negatives when calling SVs,

- Second tab allows filtering for the General SVs when utilizing the masking BED file. It will allow you to select which chromosomes to display, the precision of the BED file overlap with called SVs and how you want the masking to occur (either by showing the areas that are masked or unmasked),

- Third tab is for controlling the Variant Annotation filtering and will allow you to search for SVs present in a given percentage of the samples in the control database as well as determining whether you want SVs shows that are present in your control sample, and
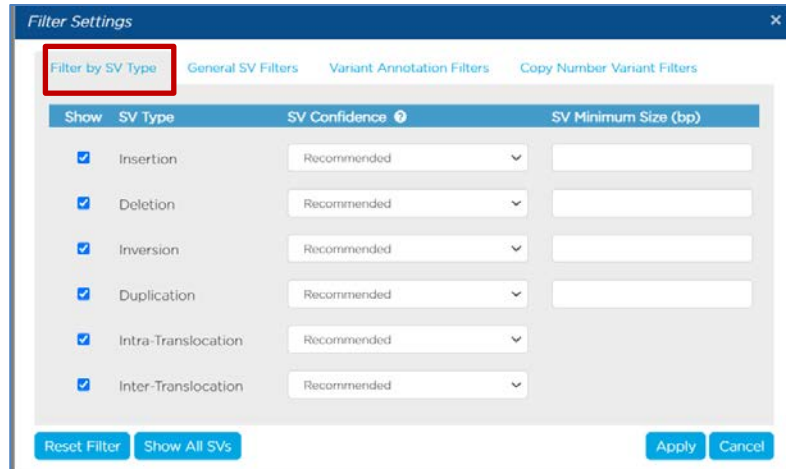
- Fourth tab is for filtering copy number variants.

Figure 18 – Filter Settings, SV Type- Filter settings include four different tabs as described below.

14. For this demo dataset, choose the default or "Recommended" setting for SV Confidence and leave blank for SV Minimum Size (bp). In short, make no changes for the Filter by SV Type tab. Then click on "General SV Filters" (Figure 18).
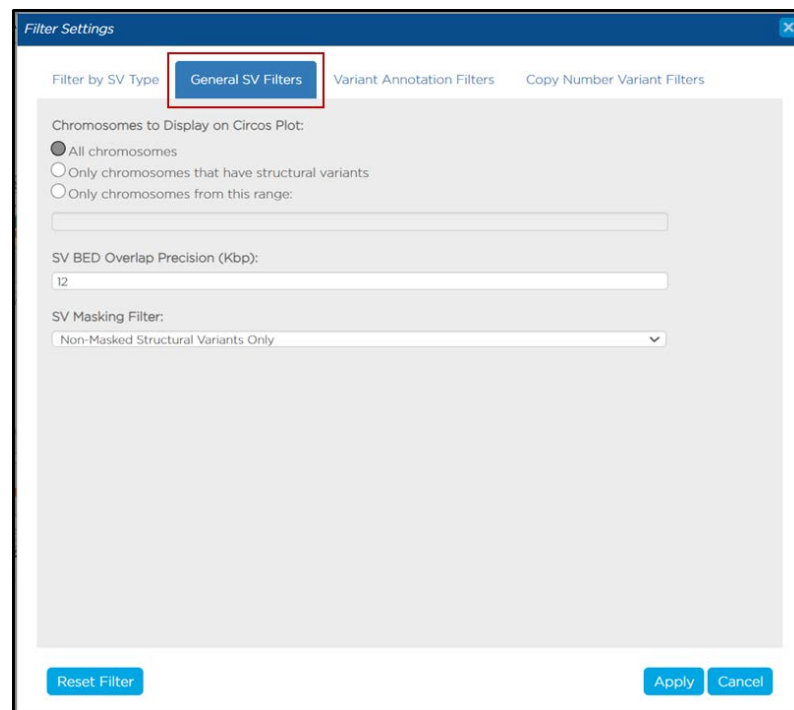

Figure 19 – General SV Filters.

In the "General SV Filters" tab (Figure 19), select "All Chromosomes" button. Leave "SV BED Overlap Precision" to 12 kbp to identify SVs overlapping a region of interest when using a BED annotation file to filter data. Select "Non-Masked Structural Variants Only" from the dropdown box under the "SV Masking Filter".
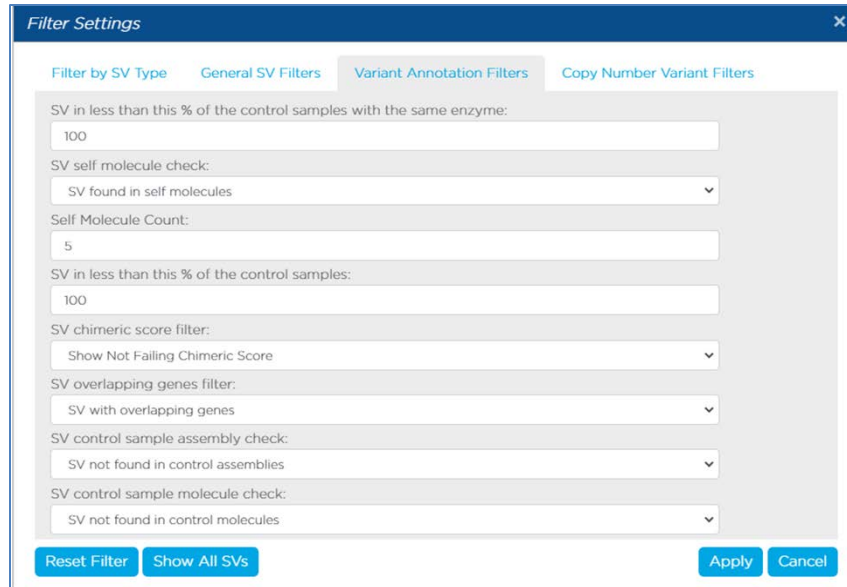
Figure 20 — RVP Variant annotation filtering options. The filters in the variant annotation tab have eight different options that can be grouped in six different categories. (A) Check against somatic variants detected in control database of healthy individuals, (B) Check for Spurious alignment, (C) Check for against one control sample (for duo analysis) or two parental control samples (for trio analysis. Not shown), (D) Check for sample-specific assembly or molecule, (E) Check for overlapping genes-

15. To filter for somatic variations unique to the cancer sample make the following changes in the "Variant Annotation Filters" tab–

   I.   SV in less than this % of the Bionano control samples: **100**

   II. SV in less than this % of the Bionano control samples with the same enzyme: **100**

   III. SV chimeric score filter: **Show Not Failing Chimeric Score**

   IV. SV control sample assembly check: **SV not found in control assemblies.**

   V.   SV control sample molecule check: **SV not found in control molecules.**

   VI. SV self molecule check: **SV found in self molecules.**

   VII. SV overlapping genes filter: **SV with overlapping genes.**

   VIII. Self Molecule Count: **5**

 The first option filters for somatic variants detected in control database of healthy individuals. By setting these values to "100", the software will display structural variants which are present in any of the individuals in the control database. The control database is not exhaustive and does not include all variants. But the control database may help to identify pathogenic SV that could be present in a control sample, who may not be "healthy" at the time of sample collection. The SV chimeric score filter ensures that regions on the Bionano maps from which SVs are detected have sufficient molecule support to create a ~~very~~ confident consensus genome map. The next set of options shows filters to remove any variants detected in the matched normal sample. This is only applicable for a paired sample or a trio analysis. The "SV self molecule check" and "Self Molecule Count" options show filters to include either molecules that align only to the cancer sample (HCC2218) or limit the number of supporting molecules, respectively. Higher number of molecules supporting a call, greater is the confidence of that SV call. The option "SV overlapping genes filter" helps identify the biological importance of the calls by identifying genic information. For more details on any of the settings, please see Theory of Operation

Structural Variant Calling – 30110. For more details, see the Bionano Solve Theory of Operation: Variant Annotation Pipeline (document #30190) for guidance on setting these parameters.

16. After clicking "Apply" you will return to the Circos Plot. Confirm that the "SV counts" lists three insertions, 36 deletions, 9 inversions, 9 duplications, 64 Intra-Translocation, 17 Inter-Translocation, 55 copy number gains and 25 copy number losses (Figure 21). These are represented by different colors on the SV track of the Circos Plot. Notice, the SVs detected after the filter is applied is lower than the SVs initially detected in sample HCC2218 (Figure 12 A).

HCC2218 - Variant Annotation Pipeline_sv	Count
Insertion	3
Deletion	36
Inversion	9
Duplication	9
Intra-Translocation	64
Inter-Translocation	17
CNV Gain Segment	55
CNV Loss Segment	25

Figure 21 – Structural Variants. Color coded by the different types.

There are many ways to identify a set of rare variants that might be of interest to a user. Detailed study is therefore needed to identify rare variants that are biologically important. One way to make sense of your data is to identify relevant publicly available information for the region, type of SV or gene of interest. In the next section we will describe how to apply a BED file for identify conserved genic regions in the overlapping genomic coordinates.

Practice Session – Identifying Potentially Pathogenic SVs, VAP Filter

This section can be used to facilitate learning for how to find a specific SV of interest. You can choose to apply basic filtering to identify potential somatic variants that are rare among phenotypically "normal" individual. You can then use masking BED file to identify SVs specific to a gene of interest. ATCC describes HCC2218 and HCC2218BL cell lines are patient derived lymphoblastoid cell lines that are highly positive for expression of Her2/neu and p53. HCC2218BL, is an EBV transformed cell line from the same patient. Dependency Map (DepMap) portal lists large number of mutations and gene fusion events for HCC2218. The portal also reported around 500 SNPs and 36 inter and intra-chromosomal translocation events. Of all the genes known to have a translocation, INTS2 genes was reported to have six different intra-chromosomal rearrangements. COSMIC database lists 2581 mutations and one gene fusion for HCC2218. We will also check the cancer cell line (HCC2218) to identify INTS2 CNV gain as reported by DepMap and Cosmic.

As a proof of principle, we will apply simple filtering criteria to identify the ZNF652-INTS2 translocation in the cancer (HCC2218) sample. To identify the genomic coordinates of the SVs impacting them, you need to add annotations to the SV calls by providing overlapping gene information. This same filtering options can be applied to identify other variants rare among phenotypically "normal" individuals that may overlaps with annotated genes or is a potential false positive call.

G. FILTER BASED ON GENE ANNOTATIONS

To visualize genic regions on a chromosome, users can import annotations from public resources such as the UCSC Browser or use a pre-loaded BED file available in the software. A BED file (.bed) is a tab-delimited text file that defines a feature track. Canonical or known genes dataset is constructed by identifying a single transcript or isoform for a gene that many have many reported isoforms. For ease of analysis, only the longest isoform is chosen for a gene and its chromosomal coordinates are included in the BED file. Because the chromosomal coordinates are reported in a BED file, genes can be visualized as tracks in a circos plot or genome map. Click here to read on how canonical genes are identified. In this tutorial, known genes corresponding to human build hg38 ("hg38 known

canonical.BED") is used for analysis since raw data was aligned using hg38 genome. This BED can be easily customized and has been manually curated to include recent genes.

Adding BED files for Filtering

17. To add a BED file, click the "Add BED" button on the Circos plot display page as shown in the figure (Figure 22), this will prompt the user to select from the numerous reference genome-specific BED files that can be easily added to any analysis.
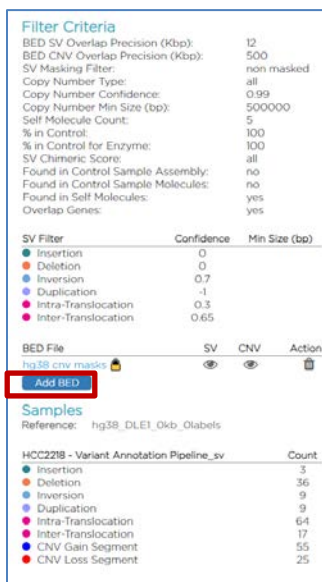

Figure 22 — Filtering settings and BED (masking) file.

"hg38 known Canonical" will also be used to filter structural variants to just those that overlap known genes and view them in the Circos plot or the Genome Browser.

18. In the BED Configuration window, select the "hg38 Known canonical" file from the dropdown as shown in Figure 23.
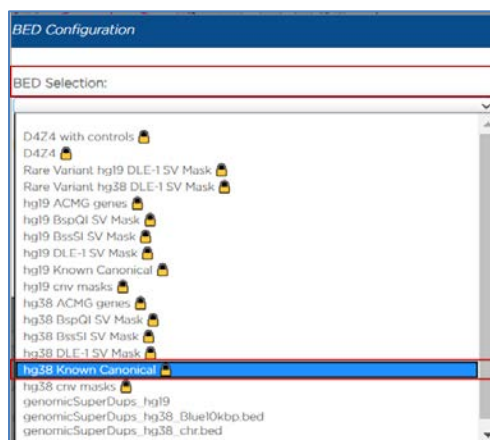

Figure 23 — Selecting hg38 Known Canonical

19. Choose the button for "Show structural variants that overlap BED Regions" as highlighted (Figure 24) and then press "Ok".
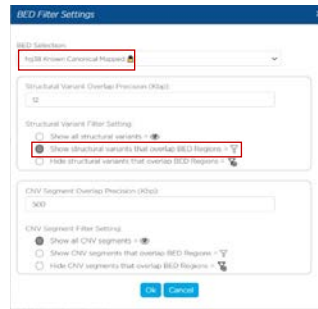
Figure 24 ─ Selecting a BED file. Preloaded bed files can be used filter out structural variation in non-specific regions.

Since the visualization is interactive, the circos plot will now show add an extra annotation track. The figure now shows the known canonical genes on the second, track right next to the cytobands. In this track, each vertical line represents a gene (Figure 25).
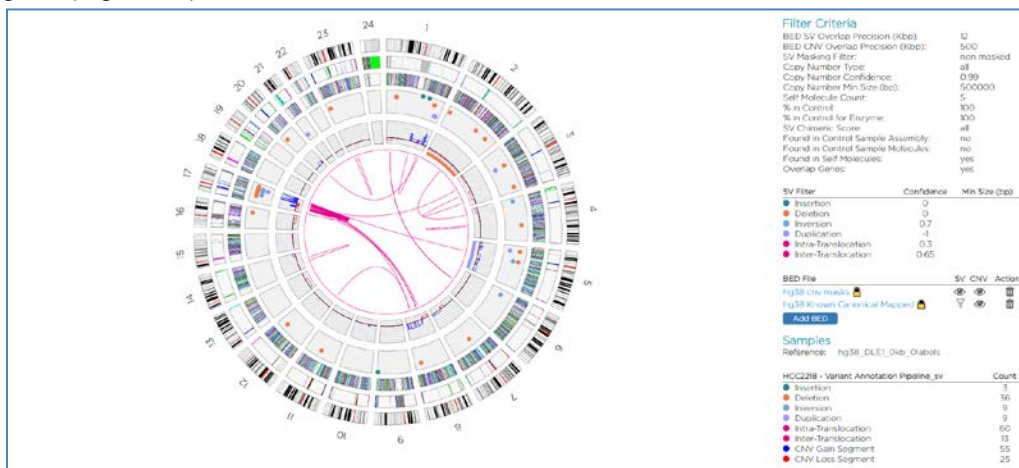

Figure 25 ─ Rare variants overlapping known genes- Circos plot with BED track.

The dataset has now been filtered according to Bionano's recommended filter settings and a BED masking file has been applied to the dataset so that only SVs that overlap with known genes will be visible (Figure 25). Confirm that the "SV counts" as shown in the figure (Figure 25). These are represented by different colors on the SV track of the Circos Plot. Note the fewer number of rare variants are identified upon implementing the filters.
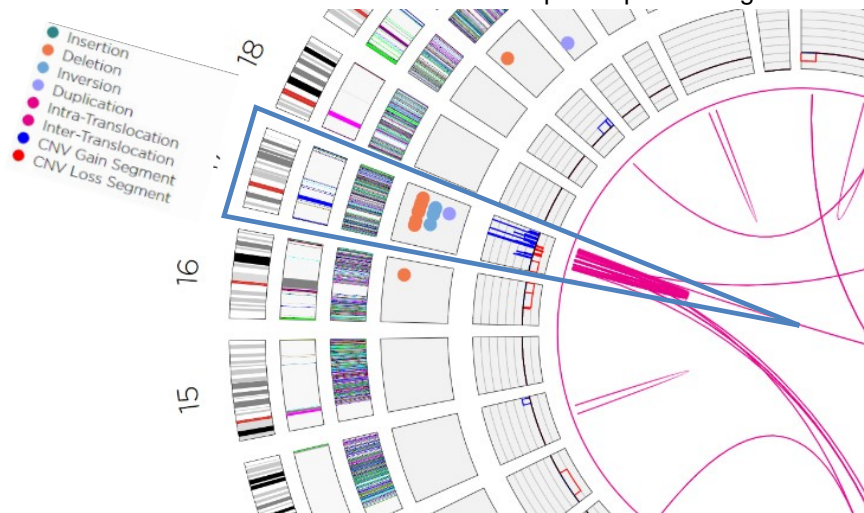

Figure 26 ─ Circos plot with Variant Annotation filters applied. Zoom in and mouse over SVs on chromosome 17.

We will now confirm whether we can find the SVs on chromosomes 17 in the regions specified by COSMIC database.

20. Zoom in on the circos plot region for chromosome 17 using the mouse wheel or track pad (Figure 26).

21. Hover your mouse cursor over the SV track for chromosome 17, which also has a large copy number change associated with that region.

22. Click over the deletion on chr 17. Confirm that the position is roughly 41,109,560-67,483,840 according to the Bionano data. (Figure 26).

23. Click on the intrachromosomal translocation to open the "Genome Map View" (Figure 27).You could also choose "Genome Browser" and Chr 17 for Anchor from the dropdown as shown in the figure.
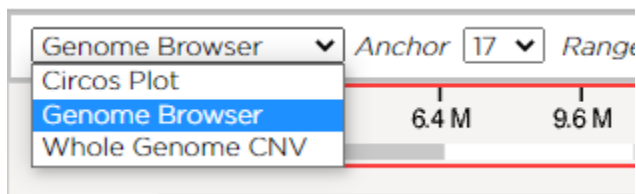

Figure 27Dropdown from the top panel allows navigation to genome browser view. "Anchor" represents chromosome number

24. This will launch the Genome browser view. This genome map view shows an intrachromosomal translocation event on chromosome 17.

Figure 28 – "Genome Map View" visualization

This zoomed out genome map view shows the chromosome 17. The top panel of the browser is consistent with the circos plot view. The cytoband is the top track. The "Anchor" and "Range" shows the precise coordinates on top reference that is being viewed. In the central region, we can view the SV in more details.

The small red box in this track indicates the location of the intrachromosomal translocation. The second track includes the SVs detected (red box) which in this example is an intrachromosomal translocation. Known genes from hg 38 are added in the view as a reference. The vertical lines on the reference map and assembled genome map

are the recognition sequence of the labeling enzyme and that connections between maps (grey) lines) are alignments. If the labels on the assembled genome maps matches with the reference map, then no structural variations are reported in a region of interest. In this example, by comparing the enzyme marks on Bionano map (blue) and reference genome locations (green), you can see there are alignment marks from the assembled genome that matches marks of chromosome 17 The abnormal alignment marks (pink lines) show the region of the map that skips contiguous alignment to the reference map suggesting a translocation event. Using your mouse, draw a box on the cytoband track around 41 Mb to 44 M region to zoom in. in the tabular section of the view, click on the header of each column to sort it. Use the mouse to scroll zoom in or zoom out. Please refer to SV Annotation Pipeline File Format Specification Sheet – 30168 for more details about each column.

25. Use the scroll bar at the base of the table to navigate to the "Overlap Genes" column. This lists the overlapping genes and nearest non-overlap genes. The first few rows in the table lists the SVs detected in this region.

26. Click on the header column to sort the overlapping gene names. This will list all the SV events for a fused gene ZNF652 - INTS2.
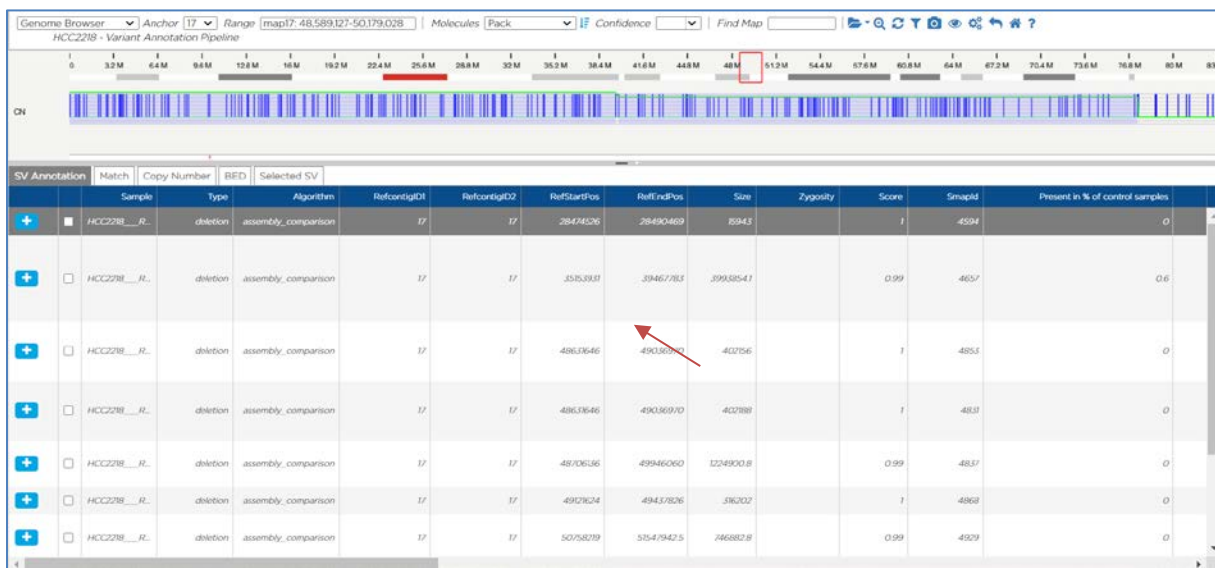

Figure 29 – Table at the bottom of the Genome Map lists the SVs for genes in the region.

27. Click on the first row to automatically navigate into this region with the SV overlapping the gene ZNF652 or INTS2.
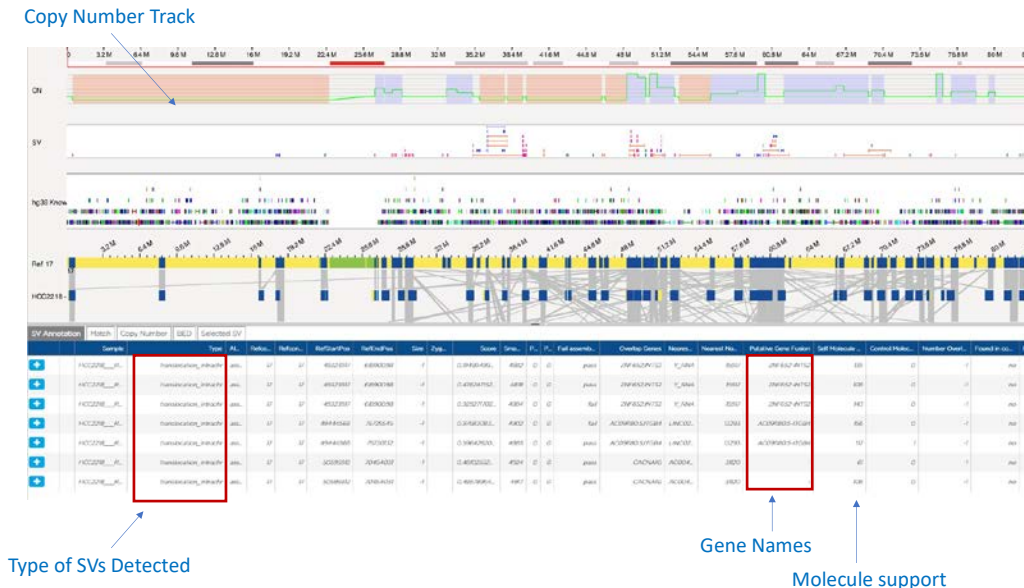
Figure 30 ─ ZNF652 gene showing several structural variations in this region.

The table at the lower half lists all the SVs, CNVs and their associated annotations as shown in Figure 30. The highlighted row lists "translocation_intrachromosomal" under the "Type" column. The next few columns provide information about the algorithm used to call the SV, contig information and breakpoint coordinate information for the reference map, etc.

28. Zoom in to the region map17: 49,100,615-49,570,725 to view the ZNF652 gene. This region shows several structural variants including translocations events in the region as well as copy number change. When regions located on the same chromosome fuse, they are called an intrachromosomal rearrangement. These intrachromosomal rearrangements are well known characteristic of malignancies. Click here to read more on translocation calls detected using Bionano optical mapping techniques in 30190 Theory of Operation. Additionally, SV track and the table on the lower half also suggest two large deletion events spanning the region.

29. Scroll to the right to find the column "Self Molecule Count". For the list of translocation events the self-molecule count is listed in the range of 135 to 143, suggesting that the breakpoints for this translocation event was supported by molecules that line up to the breakpoint coordinate. While hovering over the table, you can press

the left and right arrow keys to navigate to the additional cells or use the trackpad to drag the table left or right (Figure 30)
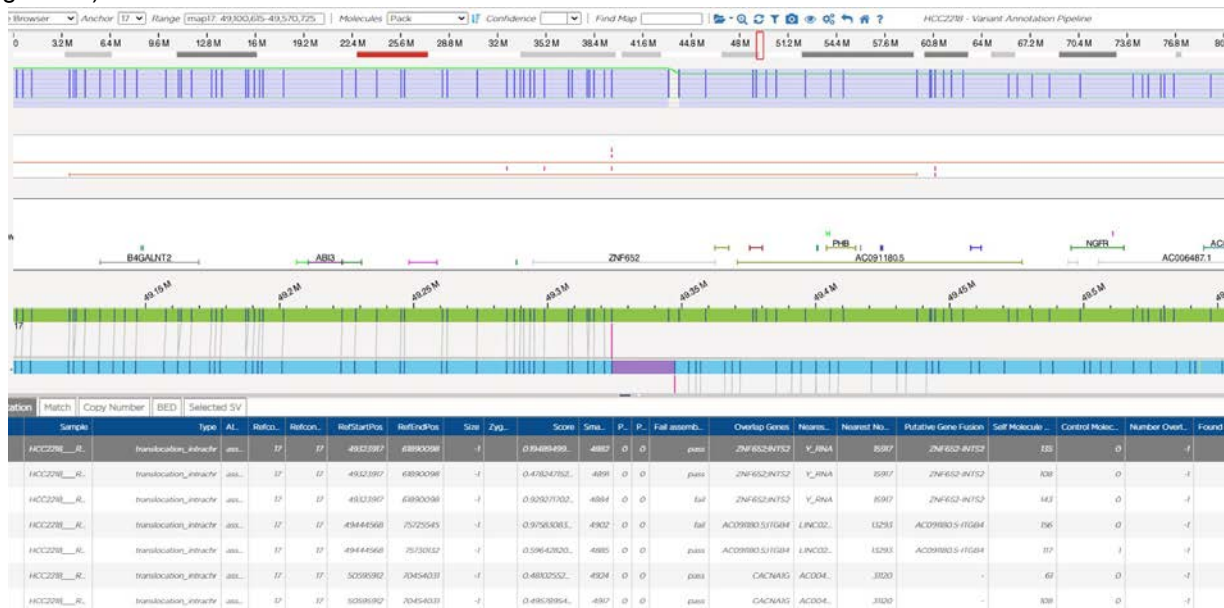

Figure 31- Genome view with the copy number track at the top.

30. Click on the "+" sign on the left of the row that has the intrachromosomal translocation SV of interest. This will add it to the "Selected SV's list, and the SV report you will print out later.

31. Right-click on the assembled genome map (light blue) shows options to show the molecule pileup (Show Molecules) used to find the variants (Figure 31).
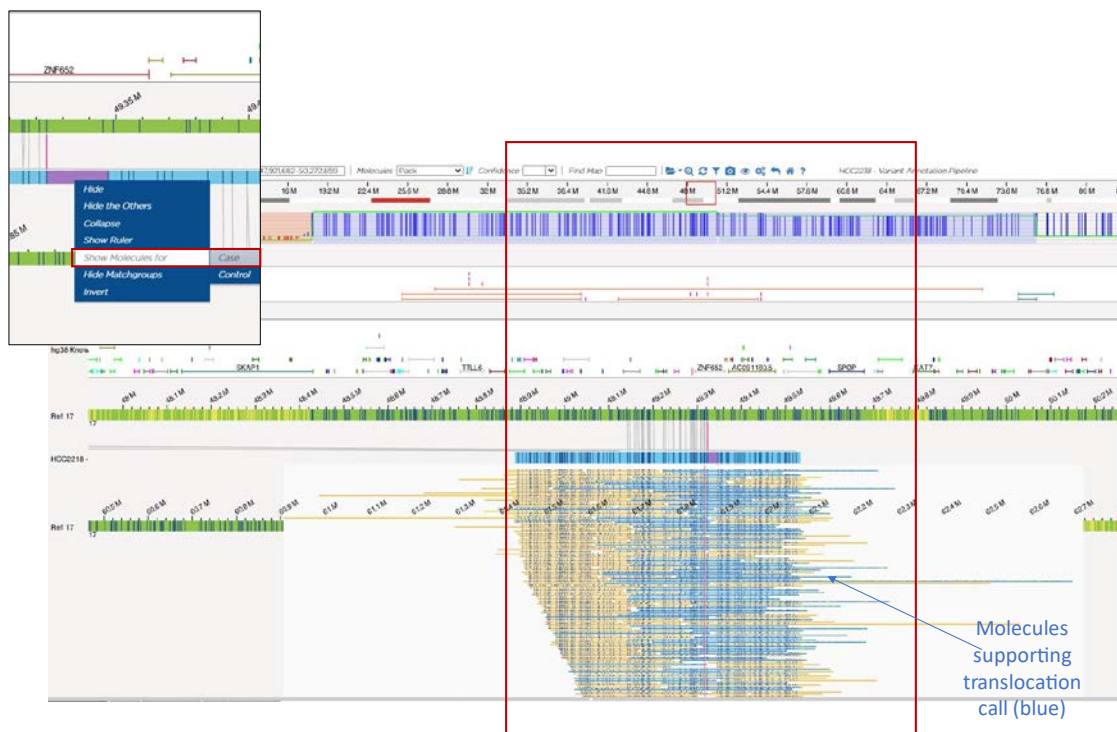

Figure 32- Right click options to add molecules or ruler to the genome map view

Thus, optical mapping confirms the translocation call. In addition, optical mapping also reports a large deletion (color coded orange), copy number increase (in the copy number track) as well as other translocations in this region. To view the any other SV call, click on the variant and it will launch the optical map supporting the call and the molecules supporting . Similar filtering criteria can also be applied to your own dataset. The circos plot and genome map navigation will be applicable to all datasets analyzed in access. In this figure, the CNV track is shown with the SV track, gene information track and the assembled Bionano map with the reference map. To find more information about copy number calls, click on the "copy number" tab. For more information on how CNVs are calculated click here (Theory of Operations - Copy Number Variation - 30210)

32. To view the whole genome copy number view, use the dropdown on the top panel to navigate to the WGCNV view as shown in the figure 33.
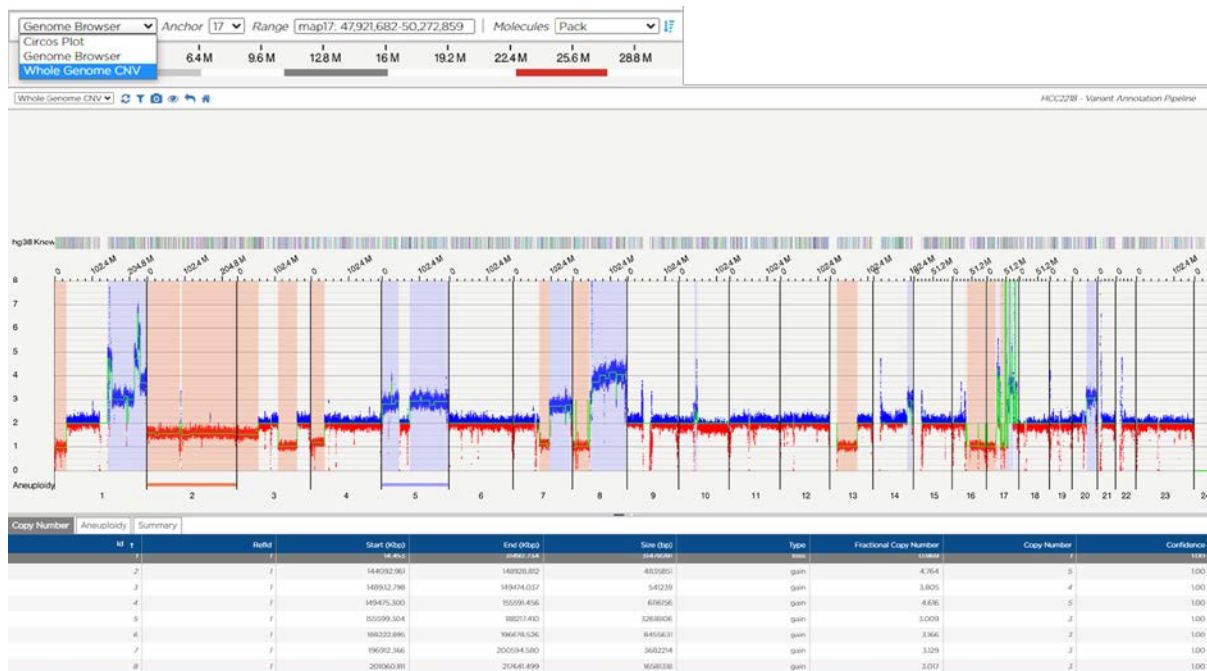

Figure 33-Whole Genome Copy Number View. The individual label copy number data points are plotted in blue or red. The segmented copy number line is green, and the copy number variant segments identified are shaded according to the variant type. Aneuploidy information is also marked on the bottom. Zoom in/out to see the scale of chromosome.

Optical mapping also shows aneuploidy and dramatic copy number changes in the genome.

33. Now that you know how to find the SVs and get the information about the calls, we will instruct on how to create a report and export your entire dataset as a shareable file.

Report Generation, Exporting Data, and Creating Images

H.   REPORT GENERATION AND DATA EXPORT

Data export can be useful record keeping or for tertiary analysis. To learn more about the different data files, read the format specification sheet. Results from the rare variant annotation pipeline are output to a text file with annotation for each SV detected for each sample. Exporting data from Access allows the user to share their findings with others or for record keeping. Access users can create a pdf report with analysis details, as well as scalable vector graphics (SVG) or screen captures for the genome map views or circos plots. These figures are also a part

of the pdf generated from Access. The image in the report will reflect what is projected on the screen when you select the SV in the data grid by default.

34. To generate a PDF SV annotation report containing the SVs that we have added, make sure you are on the "Circos plot view" and then select the Generate Report icon (pdf icon in Figure 34 and the pdf dialog box will pop up (Figure 34).

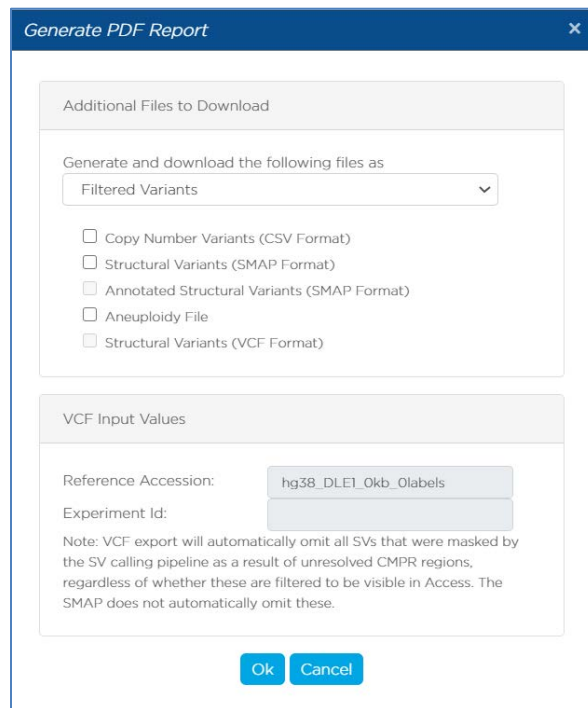*Figure 34 – Generating a PDF report. The icon in the green box launches the PDF generating dialog box*

Figure 35 – Generating a pdf report.

35. To download all the filtered variants, check the box according to the file format you desire (or check all of them), ensure "Filtered Variants" is selected in the dropdown box and then click "Ok" (Figure 35). This command will save a zipped file in the default folder (downloads or desktop).

This file is called "HCC2218_rareVariant_import_annotationDual_879_6_30_2020_2_33_18.zip". This folder will contain a PDF report, a VCF file and two CSV files of the selected variant (Figure 36). Tip: To download the variants detected in each sample, choose either "HCC2218" or "HCC2218BL" that are listed as "Rare Variant Analysis" under the column "Type". For each of these files, click on "Download Rare Variant Analysis".

Note: To generate a close-up of the Circos plot, zoom in to the desired location before selecting Generate Report.

Name	Type	Compressed size	Password protected	Size
HCC2218_rareVariant_import_annotationDual_879_6_30_2020_2_...	PDF File	953 KB	No	953 KB
HCC2218_rareVariant_import_annotationDual_879_6_30_2020_2_...	SMAP File	65 KB	No	65 KB
HCC2218_rareVariant_import_annotationDual_879_CNV_6_30_20...	Microsoft Excel Comma Separated Values File	1 KB	No	1 KB
HCC2218_rareVariant_import_annotationDual_879_CSV_SV_6_30...	Microsoft Excel Comma Separated Values File	60 KB	No	60 KB

Figure 36 ─ Rare variant downloaded folder showing the different files downloaded as a part of the report

36. To view the files, Unzip the folder. Double-click on PDF to read the file.

This report includes a screen shot of the circos plot complete with the legend, a screen shot of each structural variant, details on each structural variant, and general information about the dataset and any comments that were included during the analysis. To obtain more information about the SMAP file, refer to 30041 - SMAP Specification Sheet.

Note: Another option to export files or the whole project for the sample analyzed is to download the SV annotation files. To do this, first click on "Home" icon or the "Back" icon to go back to the project data page (Figure 37).

Figure 37 - Relaunching the project. The back icon in the red box relaunched the project browser

I. CREATING IMAGES

Bionano Access is equipped with a tool to generate publication quality images in Jpeg file format in access version 1.6. Preserving graphical quality is essential in publishing a reported finding.

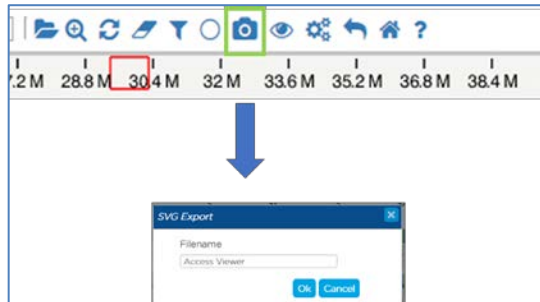44. To capture an image of a structural variant, navigate to the SV of interest.

Figure 38 - Generating images of regions of interest. Inset shows camera icon that launches the SVG export wizard.

45. Click on the camera icon displayed in the genome browser (Figure 39). A preview icon is also available to allow the user to view the image that was captured.

46. This launches an interactive "AccessViewer.svg" file from the local download. This image can be further edited where it can scale to any dimension without losing quality.

This brings us to the end of the hands-on section of the tutorial. It contains the basic skills to view, filter and export your data. Additional information can be found in the Resources section.

Resources and Information

For more about the software and data analysis, click on the help button (question mark icon) on the home page. This opens the support page

Click on the video tutorials for short videos to learn about specific analysis steps-
http://www.bnxinstall.com/Videos/TutorialIndex.html

- For Structural Variants and Rare Variant Pipeline analysis, refer to
Bionano Solve Theory of Operation: Structural Variant Calling

- For Variant Annotation Pipeline, refer to
Bionano Solve Theory of Operation: Variant Annotation Pipeline

- For Copy Number Variations, refer to
Introduction to Copy Number AnalysisGuidelines for Running Bionano Solve Pipeline on the Command Line

- For hybrid scaffolding pipeline details, refer to
Bionano Solve Theory of Operation: Hybrid scaffold

- When using command line, refer to
Guidelines for Running Bionano Solve Pipeline on the Command Line

You can also find the latest versions of these documents on Bionano's Data Analysis Documentation page.

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**

Bionano Genomics, Inc.
9540 Towne Centre Drive, Suite 100
San Diego, CA 92121