PG-Find Software

User Guide

FOR USE WITH:
PG-SEQ™ KITS

Compatible with Illumina® platforms

PG-Find™ Software User Guide

1. Overview

The PG-Find™ software is designed to assist with the identification of chromosome copy number changes in the human genome, in the context of Preimplantation Genetic Testing (PGT) using Next Generation Sequencing (NGS). Utilizing the PG-Seq™ kit technology, a single or small number of cells or the equivalent of amount of DNA undergo whole genome amplification followed by PCR-indexing to prepare the sample for sequencing. Following Illumina® sequencing, the copy number status of the samples is analyzed and visualized with the PG-Find™ software.

2. Revision History

Version	Date	Description
v1.0	September 2019	Release of PG-Find™ Software v2.0
v2.0	July 2021	Release of PG-Find™ Software v3.0
v3.0	January 2024	Rebranded to Revvity

This User Guide is intended as a manual for the PG-Find™ Software v3.0.

3. Input Requirements

The PG-Find™ software allows direct loading of BAM file types. For best results, FASTQ files should be aligned to the hg19 reference genome with BWA-MEM. Apart from alignment, no other bioinformatic processing is required.

Note that the software will not process the following:

- Samples that have not been generated using the PG-Seq™ Kit assay
- Unaligned FASTQ
- Files with less than 50,000 reads
- Corrupted BAM files

4. Computer Requirements

Access to the PG-Find™ software is included as part of each kit in the range of PG-Seq™ Kits.

The PG-Find™ software requires installation onto a computer with the following minimum hardware specifications:

- Operating system: Windows 64-bit
- Memory: Minimum 4GB (recommended 8GB)
- Hard-drive: Minimum 500GB

5. Download, Installation & Activation

5.1 PG-Find™ Reference Builder

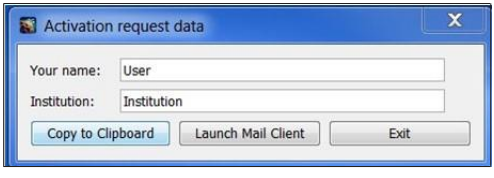
1. Download the PG-Find™ Reference Builder setup application by visiting the following link: installers.biodiscovery.com/PG-Find/Windows64/VM/PG-Find.exe
2. Double-click on the setup application icon to start the installation process.
3. Click run when prompted, then follow the installation prompts.

No Activation is required for the PG-Find™ Reference Builder.

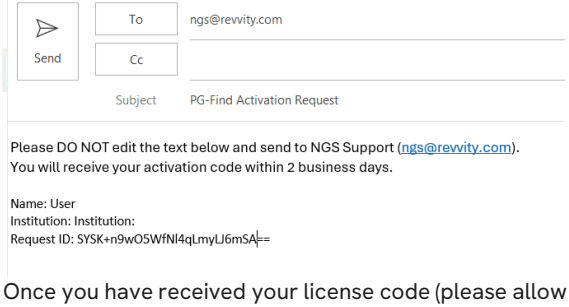
5.2 PG-Find™ Software

1. Download the PG-Find™ software setup application by visiting the following link: installers.biodiscovery.com/PG-Find/Windows64/VM/PG-Find Ref Builder.exe
2. Double-click on the setup application icon to start the installation process.
3. 3. Click run when prompted, then follow the installation prompts.
4. 4. Once installed, launch the PG-Find™ software.
5. Click Request activation key via email.

6. Complete the Activation request data form.

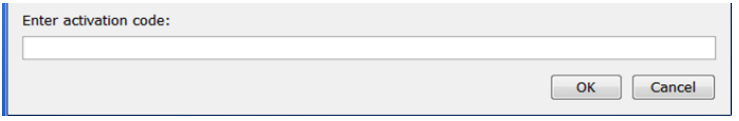


7. Click Launch Mail Client. This should open your email client with a pre-filled request (if an email client is not launched, then click Copy to Clipboard and paste the pre-filled request in an email). Send the request email to NGS@revvity.com.



8. Once you have received your license code (please allow 1-2 business days), click Activate now.

9. Copy and paste your license code into the Enter activation code field then click OK.



NOTE: Please contact Revvity at NGS@revvity.com if you are experiencing issues with installation or activation of the PG-Find™ software.

6. Algorithm Basics

The PG-Find™ software offers users two different analysis algorithm options to determine the copy number status of a sample. Once the data is loaded and depending on the algorithm chosen, the software executes the following internal processes before the result visualization is displayed.

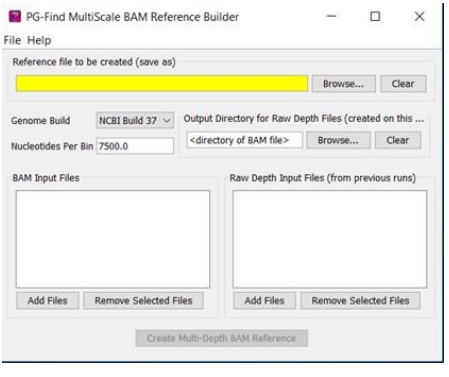
Algorithm Option:	Self Reference Algorithm	Multi-Scale Reference Algorithm
Filtering	A read filtering process is performed to remove reads containing anomalous, un- structured and highly repetitive sequences.	
Binning	The PG-Find™ software divides the genome of an individual sample into windows/regions (bins) of a fixed length, as specified by the user. This parameter is called the "Target Bin Width" and can be adjusted in the settings menu. The number of reads within each bin are then calculated and a systematic correction is applied to correct for any effects of GC content or bias.	The PG-Find™ software divides the genome of samples into windows/regions (bins) using an adjustable dynamic binning method that regulates the bin width based on the amount of genome coverage. The bin width is determined and can be adjusted during the creation of a reference file in the PG-Find Reference builder, using the parameter "Target Nucleotides per bin". The number of reads for a particular sample within each bin are then calculated and a systematic correction is applied to correct for any effects of GC content or bias.
Normalization	After binning, the read counts for each bin are normalized and assigned a copy number value based on the median read count across all bins of the autosome chromosomes.	After binning, the read counts for each bin are normalized according to the total number of reads. The normalized read count is then compared to the reference read profile and assigned a copy number value.
Segmentation	After normalization, a statistical algorithm is used to separate the genome into clusters containing bins with similar copy number values. The median copy number value of the clusters is used to plot a segment line and the position of this segment line is used to determine whether there is a copy number gain or loss involved. The degree of difference between the clusters and the	

	determination of whether they are different enough to be segmented into separate segment lines is controlled by the "Significance Threshold" parameter. The algorithm will look at the distribution of the bin profiles in a region and compare this to the distribution of the neighboring clusters (to the left and right). This distribution of bin profiles will generate a statistical p-value. If this value is less than the Significance Threshold (classified as significantly different), then the segments are divided. If the p-value is greater than the Significance Threshold (classified as not significantly different), then the segments will be joined. At the completion of the segmentation process, the entire genome is represented as a series of segments with each assigned a copy number value based on the median read count. If the segment copy number value is outside of the user-defined thresh- olds, a copy number gain or loss will be called.
Smoothing	Once segmentation has been performed and the segment lines for each region of the genome are set, smoothing is performed. In the PG-Find™ software, smoothing of the grey data points towards the segment line is performed to assist with visualization and the ease of interpretation. The raw data (unsmoothed) can also be viewed to assist with an assessment of the result quality.

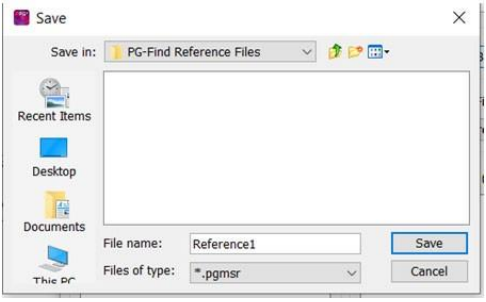
7. Creating a Reference

To enable use of the Multi-Scale Reference analysis algorithm, a reference file is required and can be generated using the PG-Find™ Reference Builder.

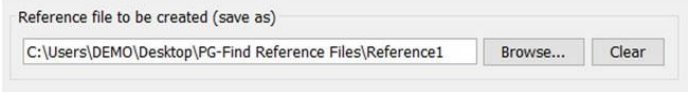
1. Launch the PG-Find™ Reference Builder.



2. Select a location to save the reference file by selecting Browse... under the Reference file to be created (save as) heading.
3. Create a folder or navigate to an existing folder where you wish to save the reference file.
4. Enter a name for the reference file in the File name: field.



5. Press Save. This will populate the Reference file to be created (save as) field with the location and name of the reference file. For example:

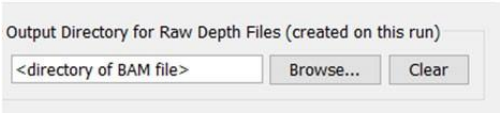


6. In the field Nucleotides Per Bin, enter a value to specify what bin width the software will use when dividing the genome into regions (bins).

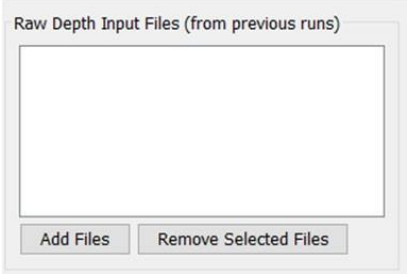
The value for the number of Nucleotides Per Bin can be calculated by determining how many reads per bin you wish to use and multiplying this number by the sequencing read length. For example, The PG-Seq™ 2.0 kit recommends using a read length of 75bp. To segment the genome into bins containing approximately 100 reads each, you would multiply 100 reads by the sequencing read length of 75bp = 7500 nucleotides per bin. Increasing or decreasing the number of nucleotides per bin value will influence the genomic distance between each bin along with the amount of noise in the analysis (quality score).

7. Under the BAM Input Files field, select Add Files then navigate to and select the .BAM files you wish to include in your reference file. Please note:
    a. A minimum of 10 sample BAM files should be used to generate a reliable reference.
    b. Samples included in reference generation should be of the same sample type and have the same pre-processing performed.
    c. The gender of the sample files used to generate a reference should match (the default PG-Find settings assume the reference is generated from male samples with 1 copy of chrX and chrY). If the gender is unknown, the PG-Find™ Self-Reference algorithm can be used to determine the gender before reference generation.
    d. Test samples containing copy number variants can be included in the sample set for reference generation, as long as multiple samples do not contain the same abnormality.
8. Select Create Multi-Depth BAM Reference

NOTE: During reference generation, an intermediary binary file is created for each .BAM file, termed the “Raw Depth File”, with the extension .MSD. By default, this file will be saved in the same directory as the .BAM files selected. You can change the output directory of these files by selecting Browse… under the field Output Directory for Raw Depth Files.

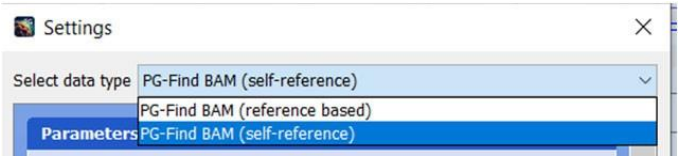


Previously generated Raw Depth Files can be used in the creation of new references and offers processing time savings. If all .MSD files are stored in a single directory it can aid in the ease of finding previous samples to include in a reference file. To include Raw Depth Files in the reference generation, select the Add Files button under the field Raw Depth Input Files (from previous runs).



8. Settings

The settings control which analysis algorithm is used and define parameters used in the analysis and visualization. There are different settings available for the Self-Reference algorithm compared to the Reference-Based algorithm. To view and change the settings for each, select the Data type dropdown box at the top of the screen.



The settings can be adjusted as desired based on the following explanation of the metrics:

<table>
<tr>
<td>Parameters (Self Reference only)</td>
<td>Target Bin Width</td>
<td>
The target size of bins or sections of the chromosome along the genome in base pairs. It is recommended that the bin size is based on the desired minimum resolution but limited by the number of reads per bin. As a guide, it is recommended to use a bin width that is 10% of your desired minimum resolution. E.g. for 5Mb minimum resolution, use a target bin width of 0.5Mn (500,000bp). The table below displays a guide for setting target bin width according to desired CNV size resolution and number of mapped reads available.
<table>
<tr><th>Target Resolution</th><th>Number of Mapped Reads</th><th>Target Bin Width</th></tr>
<tr><td>Whole chromosome</td><td>150,000</td><td>500,000-2,000,000</td></tr>
<tr><td>10Mb</td><td>250,000</td><td>500,000-1,000,000</td></tr>
<tr><td>5Mb</td><td>500,000</td><td>500,000</td></tr>
<tr><td>2Mb</td><td>1,000,000</td><td>200,000</td></tr>
<tr><td>1Mb</td><td>3,000,000</td><td>100,000</td></tr>
<tr><td>0.5Mb</td><td>12,000,000</td><td>50,000</td></tr>
<tr><td>0.1Mb</td><td>24,000,000</td><td>10,000</td></tr>
</table>
This table has been generated as a guide. Please note the attainable resolution is dependent on individual sample data quality. These data were generated from in silico testing using PG-Seq™ 2.0 Kit and PG-Seq™ Rapid Kit for Illumina data and should be validated for use in individual laboratories.
</td>
</tr>
<tr>
<td>Recenter Probes</td>
<td>Type</td>
<td>
Sometimes an entire set of data points will be seen to shift up or down from the baseline. The default recentering option (Median) uses the median value of the autosome chromosomes and places this at a copy number ratio of 2. Other centering options include specifying known diploid regions or performing no centering (none).

To use the diploid regions re-centering method, Diploid Regions needs to be selected in the Type dropdown of the Recenter Probes parameter. In addition, a column called Diploid Regions needs to be present in the Data Set table and the regions to use for re-centering need to be specified as values of this column. To perform no centering, select the option “None”.
</td>
</tr>
</table>

Analysis	Significance Threshold	The significance threshold is used to adjust the sensitivity of the segmentation algorithm. Using a less stringent (larger) value for this threshold reduces the difference required for two clusters to be separated. A more stringent (smaller) value increases the difference required for two clusters to be separated. A less stringent value is more likely to cause false positive results whereas a more stringent value is more likely to cause false negative results. In other words, less stringent sig. thresholds can be set to improve detection of copy number events, but caution will be needed when interpreting the results as the likelihood of false positives will increase.
	Max Contiguous Probe Spacing (Kbp	Here, the user can specify the maximum spacing allowed between adjacent probes before a segment line is divided. The default setting of 50000 prevents the segment line from separating into 2 segments if travelling over the centromere.
	Min number of probes per segment	Here the user can set the minimum number of probes required in order to create a new segment line. A value of 3 is recommended.
	High Gain	The value for this setting will indicate the threshold for 2 or more copy gains.
	Gain	The value for this setting will indicate the threshold for a single copy gain.
	Loss	The value for this setting will indicate the threshold for a hemizygous loss (1 copy deleted).
	Big Loss	The value for this setting will indicate the threshold for a homozygous loss (2 copies deleted).
Robust Variance Sample QC Calculation	Percent outliers to remove	This QC step calculates the probe to probe variance and this single parameter is used to remove from calculation of the variance, the extreme outliers that one would expect to be due to copy number breakpoints. It is meant to measure how much successive probes differ from each other on average. The score is displayed in the Quality column in the Data Set tab. It is computed by first ordering by magnitude the difference between adjacent probes and then re- moving a percentage of the probes that fall at the top and bottom of the list. The percentage to remove is set by the user via the Settings window (Percent outliers to remove). If the setting is 3.0 (for 3%), then half of this percentage of probes (1.5%) are removed from the top of the list and the other half, from the bottom. The default value is 3% but can be changed individually for each data type for which QC calculation is available.

Select Chromosomes	Type	This setting allows you to select only particular chromosomes for result reporting and display. By default, all chromosomes are visualized however the option “select chromosomes” can be chosen to allow separate chromosomes to be unselected.

The default settings are recommended and are optimized for detecting copy number events of 10Mb or larger. The settings can be reverted to default by selecting the Default button.

NOTE: After changing settings, any samples already loaded in the project will need to be reanalysed with the new settings. See Reanalysing Data section for instructions.

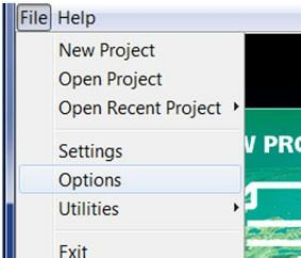
NOTE: When loading new samples into an established project, they will be automatically processed with the currently active settings. Settings do not load with projects, so be conscious of this when switching between projects that may use different settings.

9. Display Options

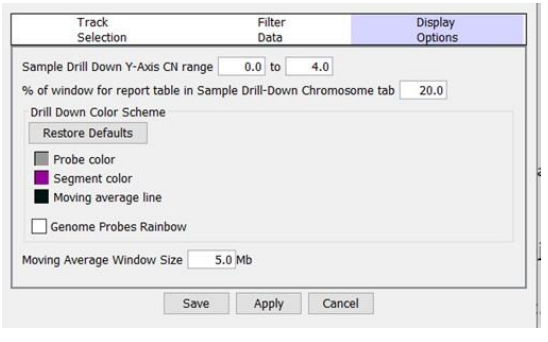
The display options control the aesthetics of the Whole Genome and Chromosome copy number variation (CNV) plots. Display options can also be used to filter the data to exclude visualization of copy number events based on the size or number of bins.

To change the Display Options:

1. Click File > Options.



2. Under the Display Options tab, change the options as desired then click Save.



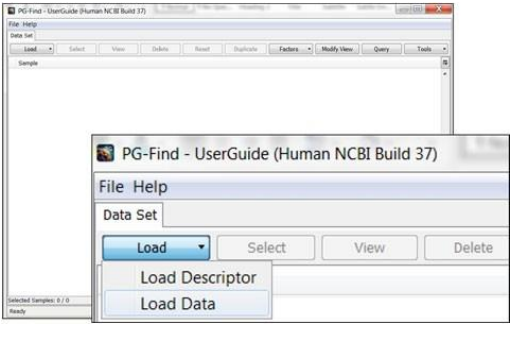
10. Launching the Software and Creating a Project

1. Launch the PG-Find™ Software.
2. On the Home screen, click New Project.
3. Enter a unique project name in the Project Name field then click Create.

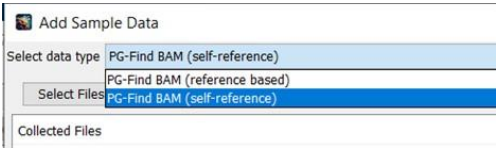
11. Loading Data

After a project has been created or loaded, sample data can be added to the project. To add samples to a project:

1. On the Data Set tab click Load then Load Data.

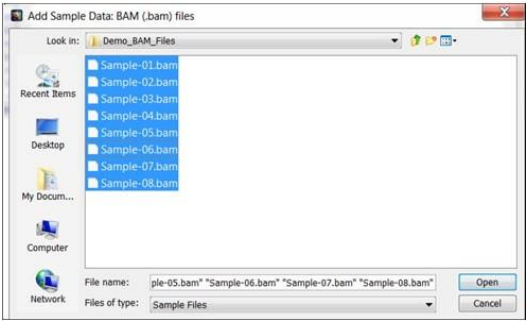


2. Select from either the self-reference or the reference-based algorithm by selecting the desired option in the Select data type field.



3. [Referenced based algorithm only] Press the Select button next to the Reference: field. Navigate to and select the reference file you wish to use for the analysis of your sample set.

4. Press Select Files and navigate to the location of your BAM files.
5. Select the BAM files you wish to analyze then click Open.

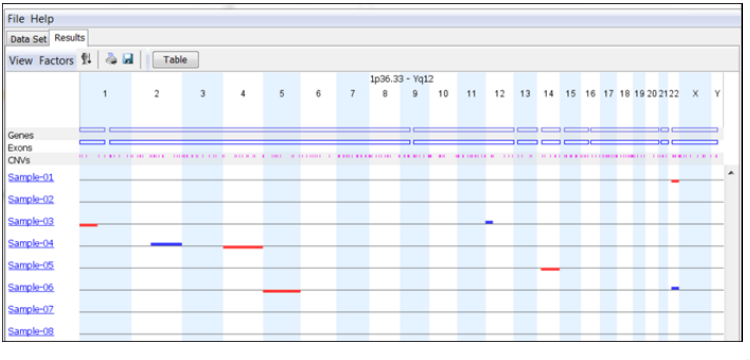


6. Click Done to begin processing the selected files.

12. Analyzing Data

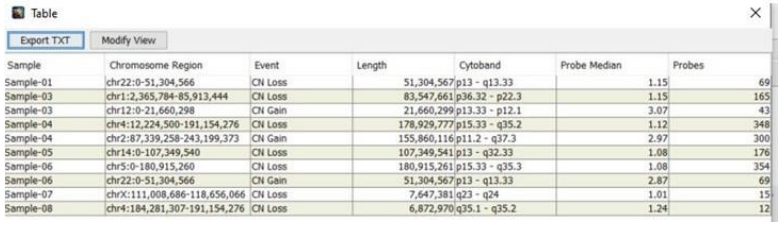
12.1 Results Tab

Once processing is complete, an overview of the sample results will appear in the "Results" tab.



This tab shows the results of each individual sample in an overview page. Sample names appear in rows on the left side of the page, each with a hyperlink to see further information. Each chromosome is listed in columns. A red line indicates a copy number loss while a blue line indicates a copy number gain.

A summary document displaying each copy number event detected can be viewed and exported by selecting the Table button.



Table

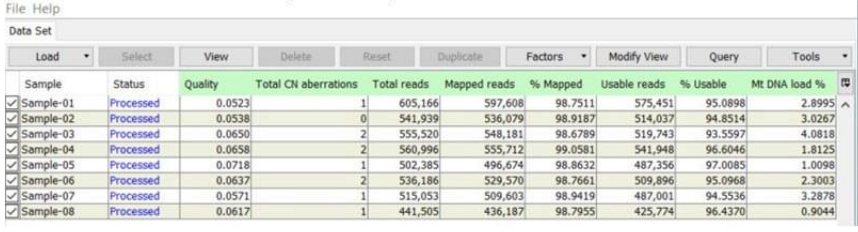
Export TXT | Modify View

Sample	Chromosome Region	Event	Length	Cytoband	Probe Median	Probes
Sample-01	chr22:0-51,304,566	CN Loss	51,304,567	p13 - q13.33	1.15	69
Sample-03	chr1:2,365,784-85,913,444	CN Loss	83,547,661	p36.32 - p22.3	1.15	165
Sample-03	chr12:0-21,660,298	CN Gain	21,660,299	p13.33 - p12.1	3.07	43
Sample-04	chr4:12,224,500-191,154,276	CN Loss	178,929,777	p15.33 - q35.2	1.12	348
Sample-04	chr2:87,339,258-243,199,373	CN Gain	155,860,116	p11.2 - q37.3	2.97	300
Sample-05	chr14:0-107,349,540	CN Loss	107,349,541	p13 - q32.33	1.08	176
Sample-06	chr5:0-180,915,260	CN Loss	180,915,261	p15.33 - q35.3	1.08	354
Sample-06	chr22:0-51,304,566	CN Gain	51,304,567	p13 - q13.33	2.87	69
Sample-07	chrX:111,008,686-118,656,066	CN Loss	7,647,381	q23 - q24	1.01	15
Sample-08	chr4:184,281,307-191,154,276	CN Loss	6,872,970	q35.1 - q35.2	1.24	12

This table can be exported by selecting the Export TXT button.

12.2 Data Set tab

The Data Set tab shows additional information about each sample including the quality and sample metrics from sequencing and alignment. For information on each metric, please see the table “Quality and Sample Metrics”.



File Help

Data Set

Load | Select | View | Delete | Reset | Duplicate | Factors | Modify View | Query | Tools

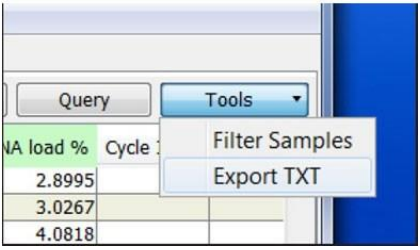
Sample	Status	Quality	Total CN aberrations	Total reads	Mapped reads	% Mapped	Usable reads	% Usable	Mt DNA load %
Sample-01	Processed	0.0523	1	605,166	597,608	98.7511	575,451	95.0898	2.8995
Sample-02	Processed	0.0538	0	541,939	536,079	98.9187	514,037	94.8514	3.0267
Sample-03	Processed	0.0650	2	555,520	548,181	98.6789	519,743	93.5597	4.0818
Sample-04	Processed	0.0658	2	560,996	555,712	99.0581	541,948	96.6046	1.8125
Sample-05	Processed	0.0718	1	502,385	496,674	98.8632	487,356	97.0085	1.0098
Sample-06	Processed	0.0637	2	536,186	529,570	98.7661	509,896	95.0968	2.3003
Sample-07	Processed	0.0571	1	515,053	509,603	98.9419	487,001	94.5536	3.2878
Sample-08	Processed	0.0617	1	441,505	436,187	98.7955	425,774	96.4370	0.9044

Quality and Sample Metrics

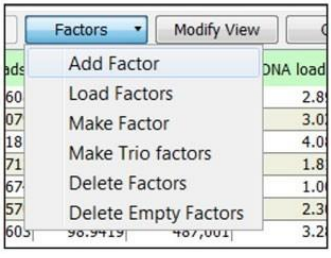
Sample	File name of the BAM file that was loaded.
Status	The processing status of the sample. The “processed” link can be selected to see what settings each sample was processed with.
Quality	The bin to bin variance in each sample before data smoothing, analogous to standard deviation. A lower quality score indicates more reliable data while a higher quality score indicates less reliable data. The quality score will vary based on the bin width selected and the number of usable reads per sample and per bin. An acceptable quality score should be determined and validated in each laboratory. A high quality score can indicate low quality DNA or issues during the WGA or library preparation.
Total CN aberrations	The total number of copy gains or copy losses detected by the software for each sample. A copy gain will be called if the copy number ratio of a chromosome or chromosome segment is greater than the user specified “Gain” value in the settings. Similarly, a copy loss will be called if the copy number ratio of a chromosome or chromosome segment is less than the user specified “Loss” value in the settings.
Total reads	The total number of reads assigned to a particular sample after de-multiplexing.
% mapped reads	The percentage of reads which successfully align to hg19. Lower values can indicate non-human contamination in your

	final library DNA.
Usable reads	The number of reads available for copy number analysis after PG-Find™ Software filtering. Reads with poor mapping quality, duplicate reads (if marked), secondary alignments and reads aligning to anomalous, unstructured and highly repetitive sequence are filtered from the analysis.
% usable reads	The percentage of usable reads compared to the total number of reads.
Mt DNA load %	The percentage of mitochondrial DNA reads compared to the total number of autosomal mapped reads.
Reference [Applicable only for samples analyzed with the reference-based algorithm]	The location and name of the reference file used to analyze the sample.
Control Gender [Applicable only for samples analyzed with the reference-based algorithm]	The gender of the samples used to generate the reference file. The software uses this information to calculate the copy number value of chromosome X and Y. If this field is blank, it assumes the gender of the reference is Male (46,XY). If the gender of the samples used to generate the reference file is female (46,XX), the value in this field should be changed to “Female” or “F”. Samples will require re-processing if this field is changed.
Gender	For samples analyzed with the self-reference algorithm, this column displays the calculated gender of the sample. For more details on how the gender estimation is performed, see the Gender Estimation section of this manual. The contents of the Gender column determine where the High Gain, Gain, Loss and Big Loss threshold lines will be placed for chrX and chrY. If the value in this column is blank, “Male” or “M”, the threshold lines for chrX and chrY will be placed at the user designated values for the autosome chromosomes minus 1 copy. If the value in this column is “Female” or “F”, the threshold lines for chrX and chrY will be placed at the user designated values for the autosome chromosomes. For samples analyzed with the self-reference algorithm, the threshold lines will automat- ically change based on the gender the software determined the sample to be. For samples analyzed with the reference based algorithm, the threshold lines are set as default to the “Male” threshold lines. The Gender column has no impact on the processing of the sample and the calculation of the chrX and chrY copy number values, only the location of the threshold lines.

The data contained within the Data Set tab can be exported by selecting Tools > Export TXT.



Additional columns with user editable fields can be added by selecting Factors > Add Factor.



Factors can be shown or hidden from view by selecting the Modify View button.

12.3 Reanalyzing Data

The same sample can be analyzed using different settings to gain further insight into the data. To reanalyze an existing sample with new settings, perform the following:

1. Ensure the desired settings are set.
2. On the Data Set tab, highlight the samples for reanalysis.
3. Click Reset.
4. Click View.

12.4 Saving Images

In many locations within the PG-Find™ software you can save an image by using the save image icon or via a Save button.

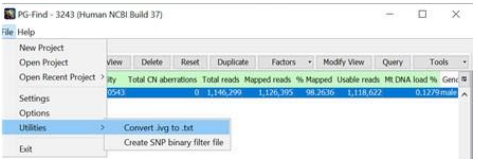
When saving an image, the Save dialog box opens and requests you to specify a location for the file. Alternatively, you can select the copy to clipboard button to allow you to copy and paste the image. Image saving options, including the resolution and file type, can be modified in the Save dialog box. When the save button is selected, the sample name is automatically added to the bottom of the image.

12.5 Exporting Bin Values

The value for each data point for each sample (in logR) can be exported to a text file. To

export these values:

1. Ensure the samples of interest are selected in the Data Set tab
2. Go to File > Utilities > Convert .ivg to .txt.



3. Select a folder to store the output (one folder will be created per sample) and click "Open".
4. After export, within each sample folder will be a probes.txt file with the bins and logR values. To convert logR values to copy number values, use the formula: 2*(2^logR value).

12.6 Sample Drill Down

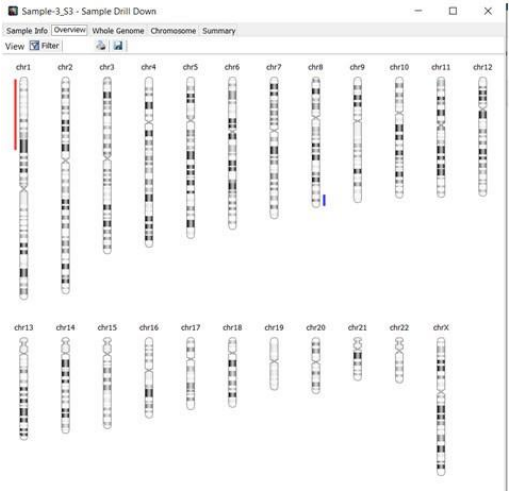
The drill down menu for each sample can be accessed by selecting the sample hyperlink in the Results tab.

a. Sample Info:

The Sample Info tab displays the same information seen in the "Data Set" tab. Additional sample notes can be added by selecting the Edit button at the bottom of the page.

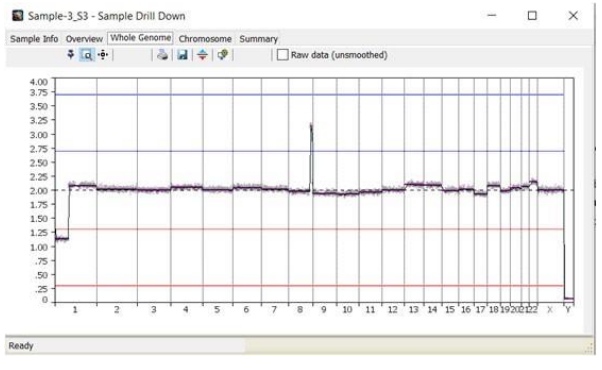
b. Overview

The "Overview" tab displays an ideogram with copy number gains highlighted in blue and copy number losses highlighted in red.



c. Whole Genome

The “Whole Genome” tab displays a smoothed graphical representation of the chromosome profile. Chromosome numbers are listed on the x-axis and a copy number ratio is listed on the y-axis. Each grey data point or bin represents a region on the genome. The pink segment line represents the median copy number value for a cluster of bins and the black line is the moving average. The red and blue lines indicate the threshold values and determine the calling of copy number events. When the segment line is above the lower blue threshold line it is called as one additional copy of that region, when it is above the upper blue line it is called as two additional copies. Similarly, when the segment is below the upper red line it represents one copy loss, whereas when it is below the lower red line it is two copy losses. The locations of the threshold lines can be adjusted in the settings menu.

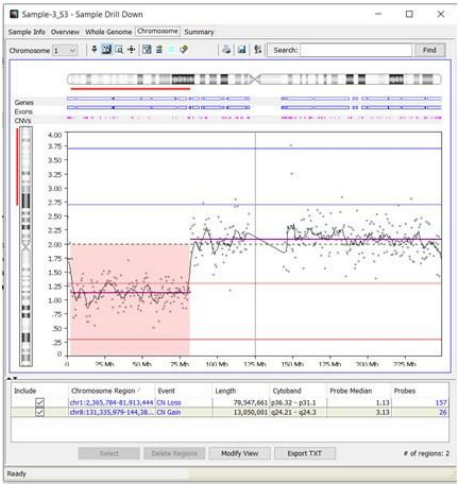


In the Whole Genome tab, the position of the grey bins are smoothed towards the segment line. To view the unsmoothed data, check the Raw data (unsmoothed) box above the image.

d. Chromosome Tab

The Chromosome tab allows you to view each chromosome in more detail and displays the raw unsmoothed data. Different chromosomes can be selected by using the

dropdown menu in the top left of the screen. As well as providing a zoomed image of each chromosome, all copy number events across all chromosomes for the sample listed in the table at the bottom. Each event has additional details such as the length and cytoband position.



A summary of the events, including settings used, can be exported using the Export TXT button.

The region displayed in the “Chromosome Region” field is a hyperlink that will show a list of common genes and CNVs published from that region if clicked.

The event type displayed in the “Event” field is a hyperlink that will zoom to the event if clicked. The “Length” field displays the length of the copy number event region in bases.

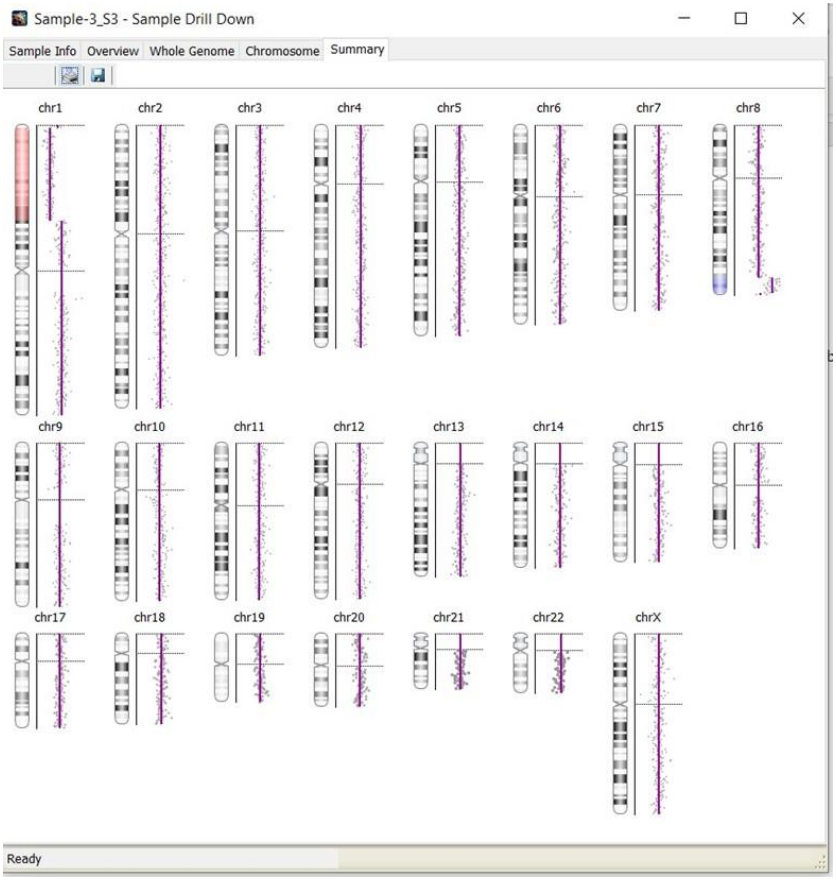
The “Cytoband” field displays the cytobands involved in the copy number event.

The “Probe Median” field displays the median copy number of the probes in the copy number event.

The number of probes in the copy number event region are displayed under the “Probes” field, which is also a hyperlink that if clicked will show a table containing the start and end point of each probe in the event, the CN ratio and the segment line value.

e. Summary Tab

The “Summary” tab displays an ideogram similar to the “Overview” tab with copy number gains highlighted in blue and copy number losses in red. This tab also includes a simplified graphical representation of the genome profile.

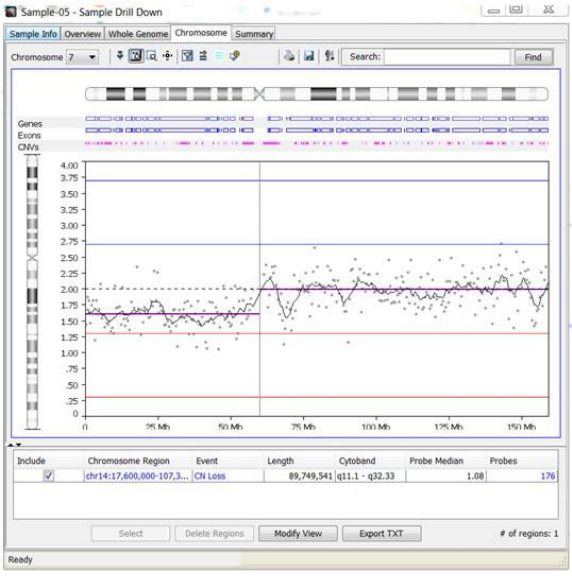


13. Manual Addition of Copy Number Calls

Regions of each chromosome can be manually called as a copy gain or loss using the "Add/Join Region" function in the Chromosome tab.

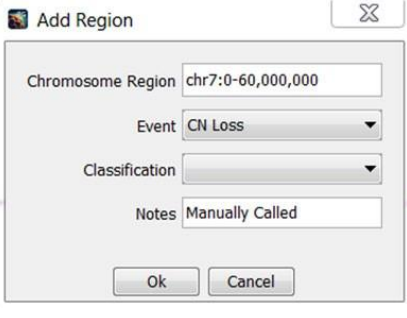
To manually call a copy number gain or loss:

1. Zoom into the location of the region or enter the region coordinates into the Search field.

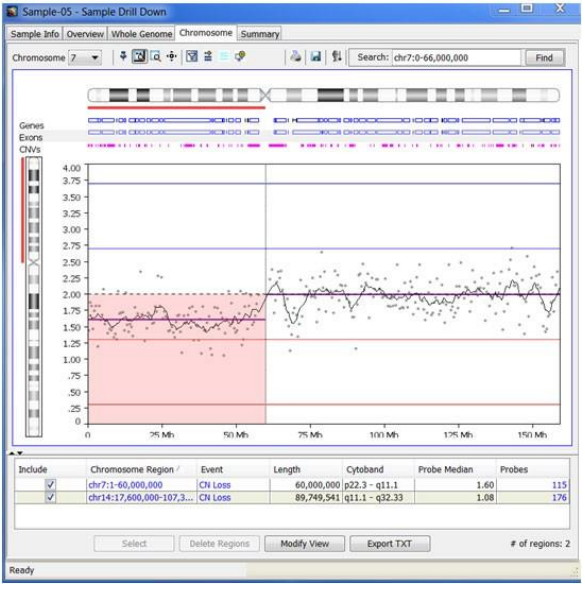


2. Select the Add/Join Region icon.

3. Confirm the Chromosome Region is correct.
4. In the Event dropdown, select the relevant copy number event.



5. Click Ok.



The selected region has now been assigned as a Copy Number Loss, with details present in the results table. To remove the manual call, highlight the event in the table then select Delete Regions.

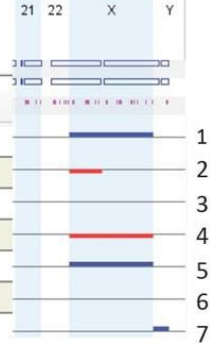
14. Gender Estimation [Self-Reference Algorithm Only]

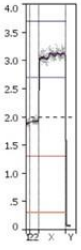
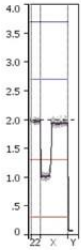
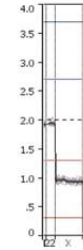
The PG-Find™ software has an inbuilt gender estimation algorithm with automatic threshold adjustment.

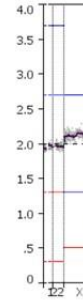
When the software detects a Y-chromosome copy number ratio above a set threshold, the sample is designated as male and the expected copy number thresholds for both the X and Y chromosome are adjusted to allow for automatic copy number calling.

See below for examples of possible sex chromosome results and the associated gender estimation and copy number calling.

Sample Number	Gender	Gender Karyotype
1	female	XXX
2	female	XX (delXp22.33p11.22)
3	male	XY
4	female	X
5	male	XXY
6	female	XX
7	male	XYY



Sample 1	Sample 2	Sample 3	Sample 4
XXX	delXp22.33p11.22	XY	X0
			
In this sample no chromosome Y is detected. There is no change to the sex chromosome threshold lines and the sample is estimated to be Female. Because the bins of chromosome X are above the copy gain threshold line (blue line), a gain of chromosome X has been called. The X chromosome also has a copy number value of 3 which indicates XXX.	In this sample no chromosome Y is detected. There is no change to the sex chromosome threshold lines and the sample is estimated to be Female. Because the bins of part of chromosome X are below the copy loss threshold line (red line), a loss of Xp22.33p11.22 has been called.	In this sample chromosome Y has been detected. The copy gain and loss threshold lines are adjusted and the sample is estimated to be Male. The X and Y chromosomes sit between the blue and red threshold lines at 1 copy each, indicating that this sample is XY.	In this sample no chromosome Y is detected. There is no change to the sex chromosome threshold lines and the sample is estimated to be Female. Because the bins of chromosome X are below the copy loss threshold line (red line), a loss of chromosome X has been called. The X chromosome has a copy number value of 1 which indicates X0.

Sample 5	Sample 6	Sample 7
XXY	XX	XYY
		
In this sample chromosome Y has been detected. The copy gain and loss threshold lines are adjusted and the sample is estimated to be Male. Because the bins of chromosome X are above the copy gain threshold line (blue line), a gain of chromosome X has been called. The X chromosome has a copy number value of 2 and the Y chromosome has a copy number value of 1, indicating this sample is XXY.	In this sample no chromosome Y is detected. There is no change to the sex chromosome threshold lines and the sample is estimated to be Female. The copy number value of the X chromosome is 2, and the Y chromosome is 0, indicating this sample is XX.	In this sample chromosome Y has been detected. The copy gain and loss threshold lines are adjusted and the sample is estimated to be Male. Because the bins of the Y chromosome are above the copy gain threshold (blue line), a gain of chromosome Y has been called. The X chromosome has a copy number value of 1 and the Y chromosome has a copy number value of 2, indicating this sample is XYY.

If the gender estimation is incorrect, which is especially possible in cases where the gender is not clearly XX or XY, the user can manually correct the value in the Gender field then reprocess the sample. Acceptable values for the Gender field include M, F, Male, Female, male, female, unknown. If the Gender value is deleted and left as blank, the sample will be reset, and the gender will be re-estimated during processing.

15. Recommendations and Cautions

- We recommend complementing calls made by the software with visual inspection of the copy number events. In some cases, visual inspection of the probes can provide

improved resolution of the segment (e.g. around breakpoints) called by the software.

- The acrocentric chromosomes (chromosomes 13-15 and 21-22) are likely to have no probes on the p-arms. For these chromosomes, caution needs to be taken when interpreting copy number events involving the p-arm; if a copy number event is called at the start of the q-arm, the event will be called through the p-arm also, as no probes exist on the p-arm to inform the algorithm to the contrary. In theory, such a copy number event may signal a segment on the q-arm only, or a segment on the q-arm that bridges the centromere and involves any part of the p-arm from the q-arm end.
- When loading new samples into an established project, they will be automatically processed with the currently active settings. Settings do not load with projects, so be conscious of this when switching between projects that may use different settings.
- There are two methods for selecting samples; by marking the checkbox next to a sample name and by highlighting a sample row by clicking the sample name. The checkbox selection is only used for the “View” button. For other buttons such as “Delete” and “Reset”, the samples need to be highlighted. Multiple samples can be highlighted by click and dragging the mouse, and multiple sections can be used by holding down the “Ctrl” key. Highlighted samples can have the checkbox marked by clicking the “Select” button.

16. Software Demo

For a software demonstration or technical support, please contact NGS@revvity.com

www.revvity.com

revvity

Rewity, Inc.
940 Winter Street
Waltham, MA 02451 USA

(BOO) 762-4000
www.rewity.com

For a complete listing of our global offices, visit www.revvity.com