

# Analysis of FFPE DNA Data in CNAG 2.0 A Manual

## Table of Contents:

<b>I.</b>	<b>Background</b>	<b>P.2</b>
<b>II.</b>	<b>Installation and Setup</b>	
	a. Download/Install CNAG 2.0	P.3
	b. Setup	P.4
<b>III.</b>	<b>Extract Mapping 500K FFPE Data</b>	<b>P.7</b>
<b>IV.</b>	<b>Analyze Mapping 500K FFPE Data</b>	
	a. Pair-wise	P.9
	b. Unpaired	P.11
<b>V.</b>	<b>Determine Fragment Size Filter for FFPE Data</b>	<b>P.13</b>
<b>VI.</b>	<b>View Data</b>	
	a. In CNAG	P.19
	b. In IGB	P.23
	c. As a .txt file	P.25
	d. View Regions of interest in the UCSC browser	P.26

## **I. Background**

CNAG (Copy Number Analyzer for GeneChip) is a software tool freely available to academic researchers for the analysis of copy number (CN) using Affymetrix GeneChip® Mapping arrays. This tool can be used in the analysis of formalin-fixed, paraffin-embedded (FFPE) samples for CN changes. The Mapping 500K arrays should be used when analyzing FFPE samples.

The following files are required by CNAG for data analysis:

1. CEL files, which provide intensity values for the oligo probes
2. CHP files, which provide genotyping data per SNP
3. Library files for the Mapping 250K Nsp and the Mapping 250K Sty arrays - these can be downloaded from the Affymetrix website ([www.affymetrix.com](http://www.affymetrix.com))

Non-FFPE references are preferred for analysis of FFPE DNA data. If FFPE references are used, it will be impossible to determine a fragment size cutoff according to the method described in this manual.

## **II. Installation and Setup**

### **II.A. Download/Install CNAG 2.0**

1. To download CNAG, visit the website: <http://www.genome.umin.jp/>
  - a. Click on the CNAG icon on the left.
  - b. Click on the "go to download page" hyperlink.
  - c. Click "I accept" at the bottom of the page.
  - d. Click on the "Program" hyperlink under Version 2.0.
  - e. Save the zip file to your desktop (~ 35 Mb).
2. To install CNAG:
  - a. Unzip the "CNAG2.zip" file.
  - b. Double click on the "SetUp.exe" icon in the "CNAG2" folder, and follow directions.
3. To open/start CNAG:
  - a. The CNAG program is automatically placed in the following folder:  
C:/Program Files/GenomeLab/CNAG ver2.0
  - b. Start CNAG by double clicking on the "CNAG.exe" icon in this folder, or by finding it in the "All Programs" section of the Start menu.

## II.B. Setup CNAG 2.0

1. To set up CNAG, first create a new folder on your computer called "CNAG2.0". Within this folder, create three new folders (Fig 1):
  - a. "Array Data"
  - b. "Results"
  - c. "Output"

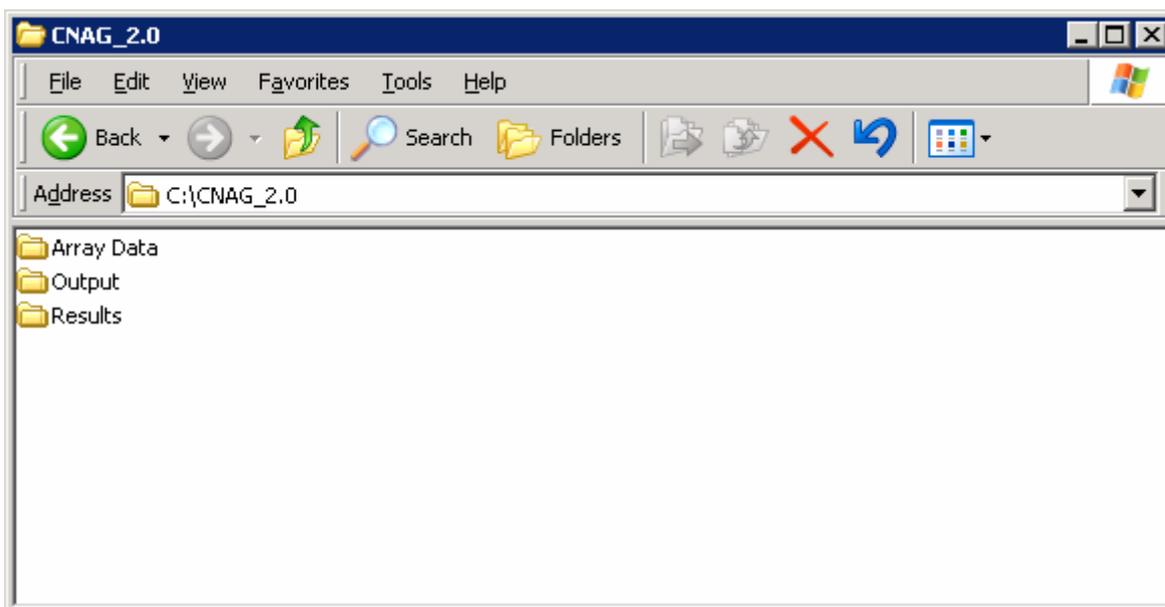


Figure 1

2. In CNAG, go to the Parameters menu item, and select "Default Setting" (Fig 2).

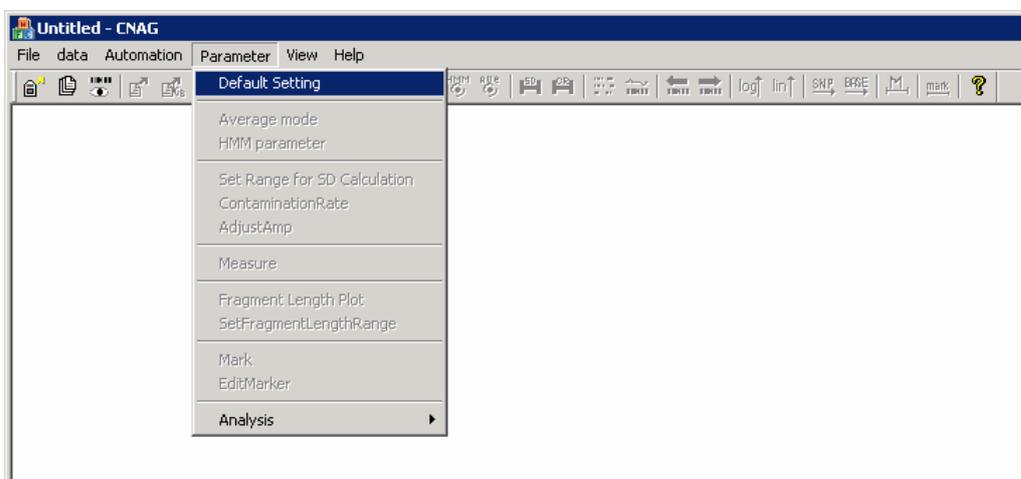


Figure 2

3. Within the Default Setting Window, there are five folders that need to be identified (Fig 3). These folders can be set up as follows:
  - a. CEL\_CHP files
    - i. If data was analyzed using GTYPE on your local computer, the CEL and CHP files will be at the following location:  
C:/Program Files/Affymetrix/GeneChip/Affy\_Data/Data/
    - ii. If data was analyzed on a server, the data should be on that server.
  - b. Library files
    - i. These should be in the following location:  
C:/Program Files/Affymetrix/GeneChip/Affy\_Data/Library/
    - ii. If these are not on the computer, they can be downloaded for installation at the following site:  
<http://www.affymetrix.com/support/technical/libraryfilesmain.affx>
  - c. Array Data files (for storing extracted data)
    - i. Select the new folder created in the CNAG2.0 folder titled "Array Data".
  - d. Result files (for storing analyzed CN data)
    - i. Select the new folder created in the CNAG2.0 folder titled "Results".
  - e. Output files (for exporting analyzed data as .txt or IGB files)
    - i. Select the new folder created in the CNAG2.0 folder titled "Output".

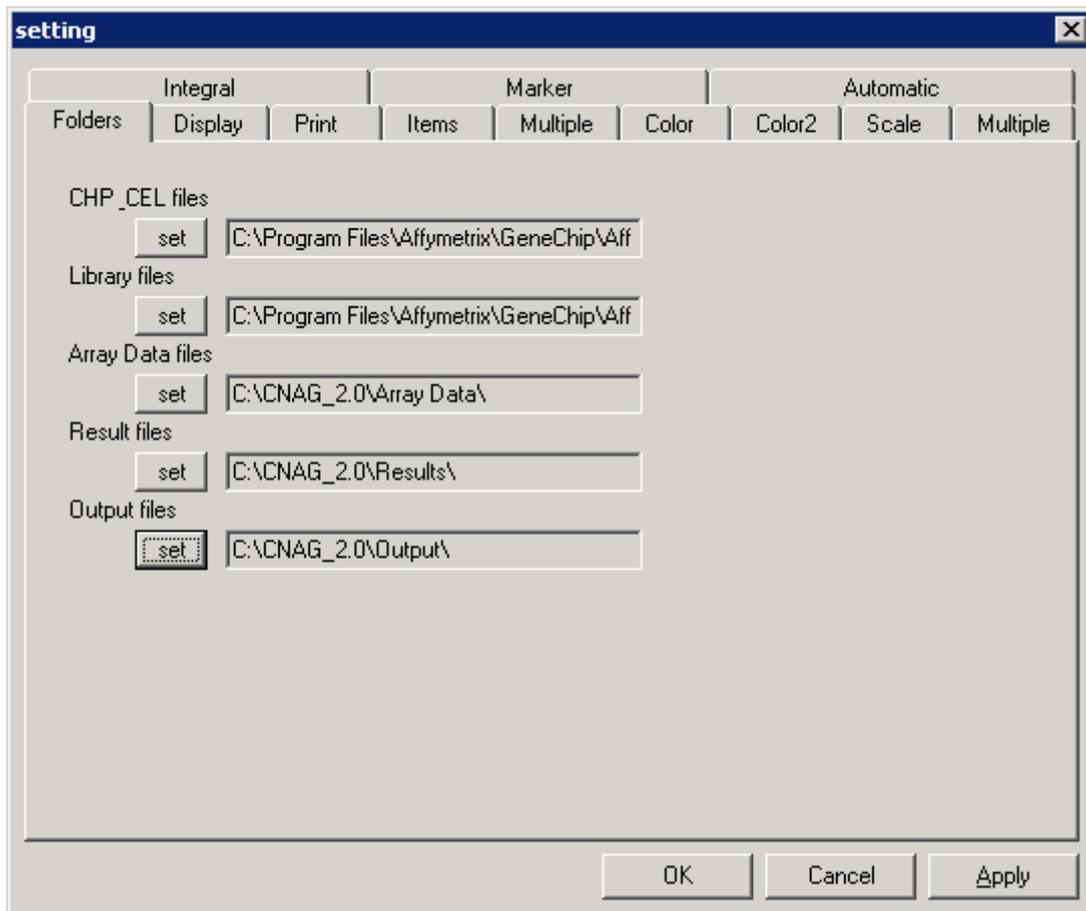


Figure 3

### III. Extract Mapping 500K FFPE Data

1. The data needs to be extracted from the CEL and CHP files in a form that CNAG will be able to use. During this step, CNAG will create a file with a .CFH extension and this file will later be used for copy number analysis. The .CFH files will be stored in the "Array Data" folder unless otherwise specified in the Default Settings (see Section II.B.3).
2. Go to the Data menu item and select "Extract Data" (Fig 4).

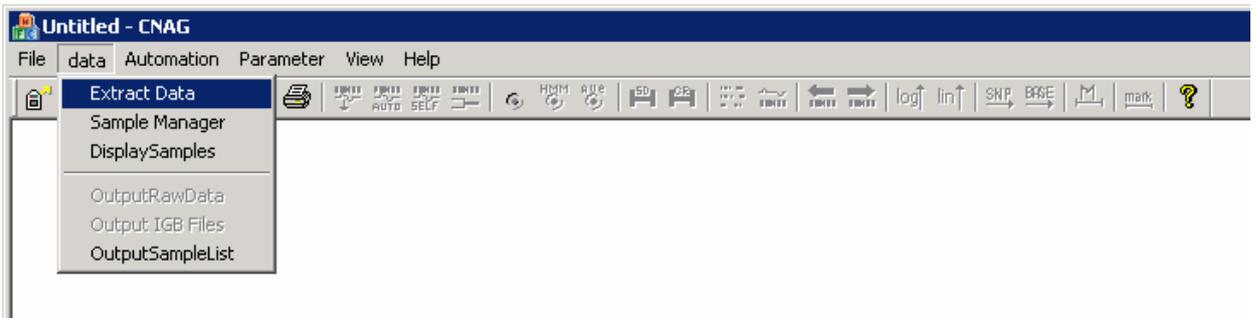


Figure 4

3. In the Extract Data window (Fig 5), there will be a section on the left that displays all CHP files in the folder specified as "CEL\_CHP files" in the Default Settings (Section II.B.3).

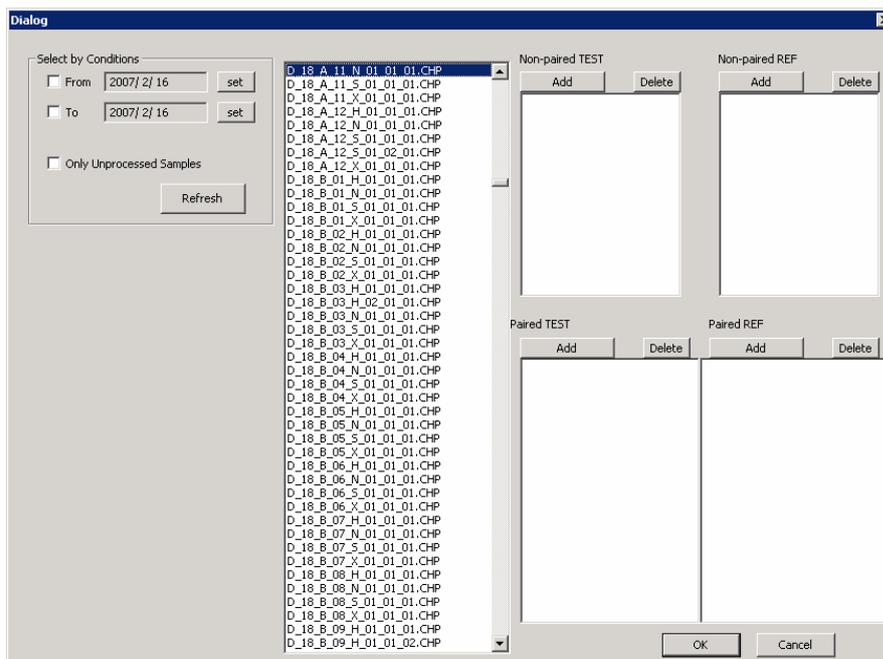


Figure 5

4. Select the CHP file for a given sample, and click on the "Add" icon above one of the four boxes on the right, depending on which box the sample is targeted for. The four boxes on the right are should be selected as follows:
  - a. For samples that will be compared to paired references (for example, tumor/normal pairs, parent/child pairs), add the CHP file of the test sample to the lower left box (labeled "Paired TEST") and the reference sample to the lower right box (labeled "Paired REF").
    - i. **When multiple pairs are extracted at the same time, each pair must be aligned horizontally with its matched pair.**
    - ii. When performing CN analysis, you will have the option to analyze these test samples with its paired reference or with unmatched references.
  - b. Place test samples that will be compared to unmatched references in the top left window (labeled "Non-paired TEST")
    - i. When performing CN analysis, you will only have the option to analyze these test samples with unpaired references.
    - ii. The unpaired references that will be used for CN analysis will be selected in a later step. This sample will not necessarily have to be compared to the references placed in the top right window (labeled "Non-paired REF") at the same time of data extraction.
    - iii. It is possible to extract data for an unpaired test sample even if data for a reference sample is not getting extracted at the same time.
  - c. Place samples that will be used as unmatched references in the top right window (labeled "Non-paired REF").
    - i. These samples will be made available as unmatched references for any unpaired analysis, not just for the analysis of those test samples extracted at this same time.
    - ii. It is possible to extract data for reference samples even if data for test samples is not getting extracted at the same time.
    - iii. As more unmatched reference samples are extracted into the "Array Data" folder, more references samples will become

available for unmatched references during CN analysis,  
potentially lending to decreased noise in CN predictions.

## IV. Analyze Mapping 500K FFPE Data

### IV.A. Pair-wise Data Analysis

1. In the Data menu, select "Sample Manager" (Fig 6).

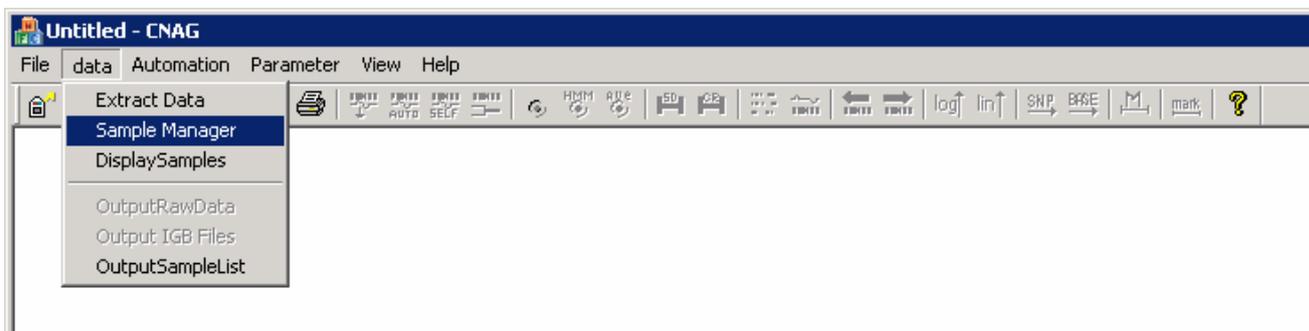


Figure 6

2. In the middle column of the Sample Manager there is a list of all of the data files that have been extracted into the Array Data folder (Fig 7).

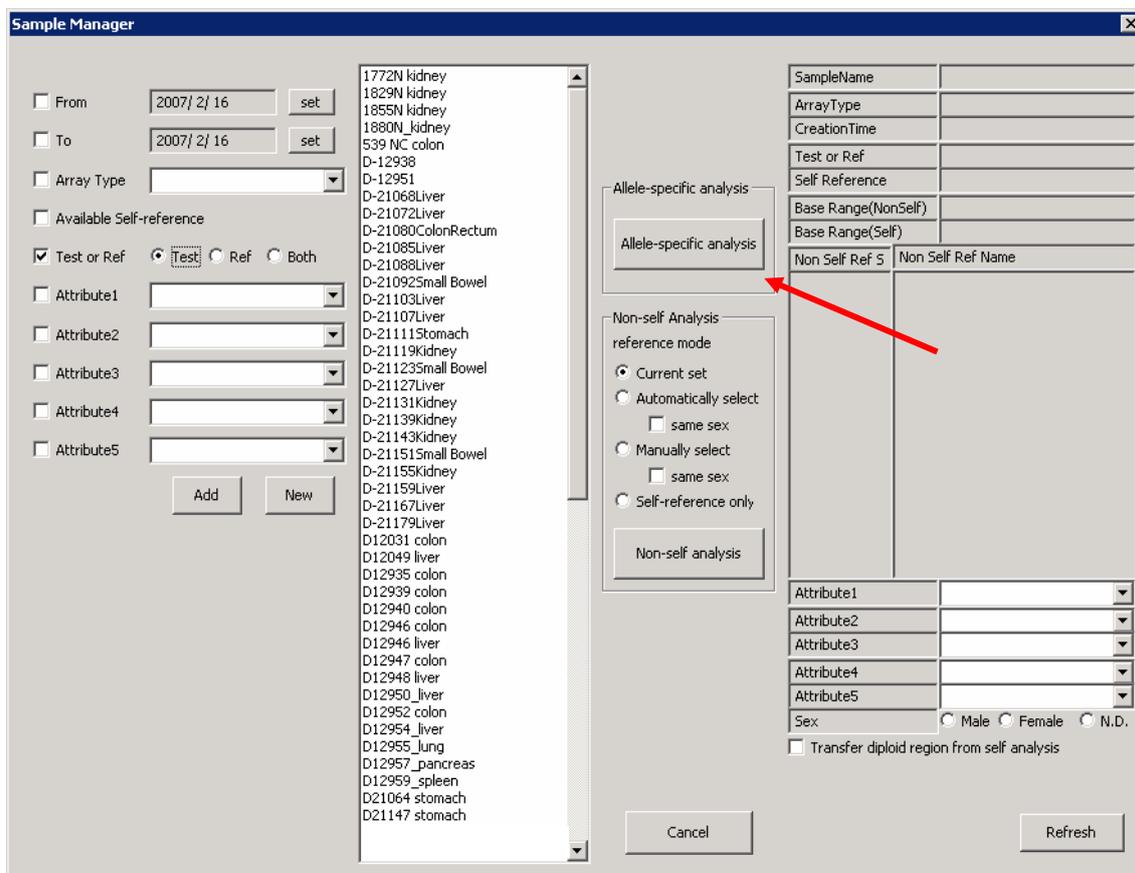


Figure 7

3. Click on the sample that you would like to analyze.
4. Click on the "Allele-specific analysis" icon (see arrow, Fig 7). This will result in the CN analysis of your test sample with its paired reference sample. If this icon is gray, then the selected sample was not extracted with a paired reference. In this case, re-extract as paired test/reference samples or perform an un-paired analysis.
5. CNAG will place the results of this paired analysis in the "Results" folder specified in the Default Settings. CN results from paired analyses will be given a .CFS extension.
6. CNAG will automatically display the results in a whole-genome view once CN analysis is complete.

**Note:** **When** comparing FFPE samples with non-FFPE, standard references are necessary to determine the proper fragment size cutoff.

## IV.B. Unpaired Data Analysis

1. In the Data menu, select "SampleManager" (Fig 6).
2. In the middle column there is a list of all of the data files that have been extracted into the Array Data folder (Fig 8).
3. Select the sample that you would like to analyze by clicking on it.

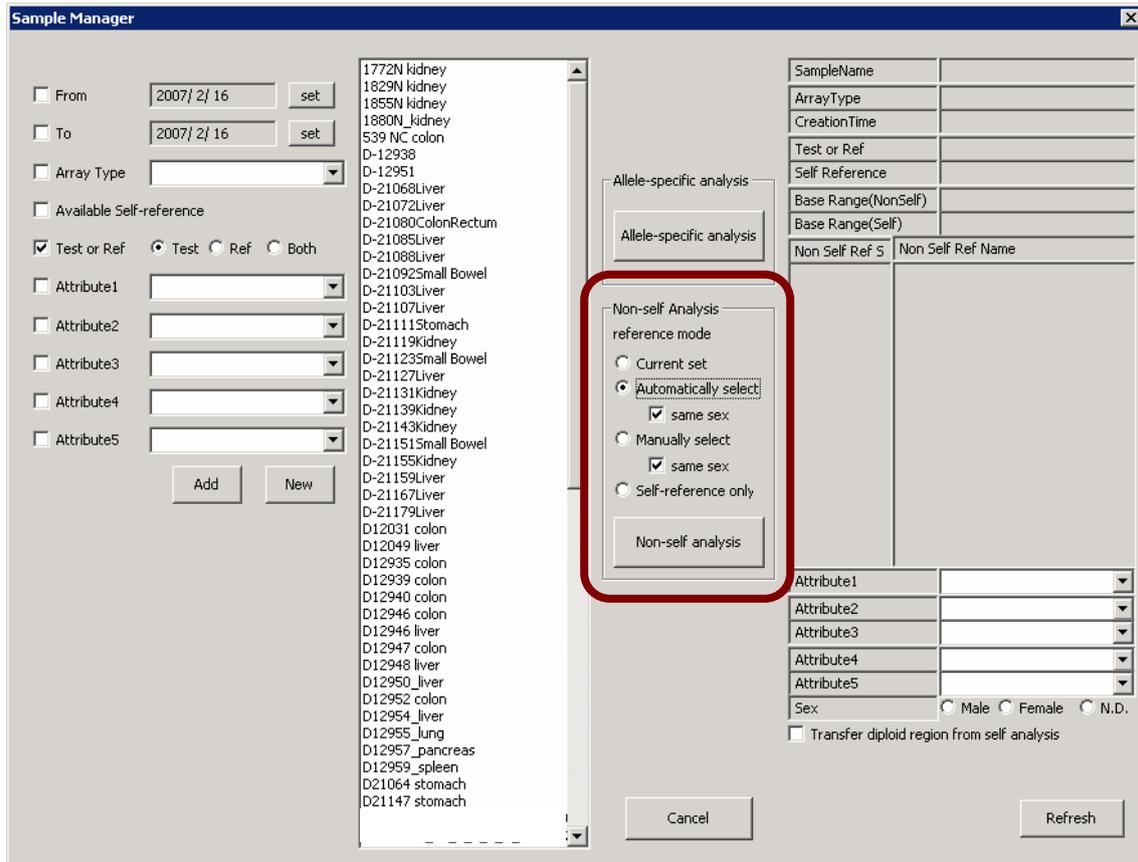


Figure 8

4. For unpaired analysis, there are several choices, all found within the box titled "Non-self-Analysis" (see brown box, Fig 8). **If CN predictions of the X chromosome will be used, check the "same sex" box and choose only female references. By analyzing against an all-female reference set, the X chromosome CN results can be easily interpreted. When mixed gender reference sets are used, interpretation of the X chromosome results may be difficult.**
5. Choose between the four options by clicking on the radio button beside the option. After the selection has been made, click on the "Non-self analysis"

- icon to begin analysis. The four unpaired analysis options are as follows:
- a. Current set: If re-analysis is desired using the same reference samples that were used in the prior analysis of the same sample, choose this reference mode. Note, this will only be an option if the sample has already been analyzed.
  - b. Automatically select: CNAG will compare the test sample to each reference data sample extracted to the Array Data folder to determine which comparison provides the lowest standard deviation. It will also compare the test sample to combinations of reference samples. CNAG will then automatically select the reference sample or set of reference samples that result in the lowest standard deviation.
    - i. To identify which samples were chosen as the reference, click on the "Info" icon at the top of the CNAG window once the data is presented (Fig 17).
    - ii. This is often the method of choice for unpaired analyses.
  - c. Manually select: CNAG will provide a list of all reference samples whose data have been extracted to the Array Data folder, after the "Non-self analysis" icon has been selected to begin analysis. Select the checkboxes next to the desired references. Ideally, a minimum of five references should be selected.
  - d. Self-reference only: This option is only available if the test sample was extracted with a paired reference sample. This will result in CN analysis of the test sample using its paired reference, but will not include allele-specific analysis because the analysis will be performed as if it is unpaired.
6. CNAG will place the results of this unpaired analysis in the "Results" folder specified in the Default Settings. CN results from unpaired analyses will be given a .CFN extension.
  7. CNAG will automatically display the results in a whole-genome view once CN analysis is complete.

**Note:** When comparing FFPE samples with non-FFPE, standard references are necessary to determine the proper fragment size cutoff.

## V. Determine Fragment Size Filter for FFPE Data

1. Note, once analysis is done, it is possible to quickly switch between the display of various analyzed datasets. See section VI.A (View Data in CNAG) for details.
2. With the data displayed for the FFPE sample of choice, select “Fragment Length Plot” under the Parameter menu item (Fig 9).

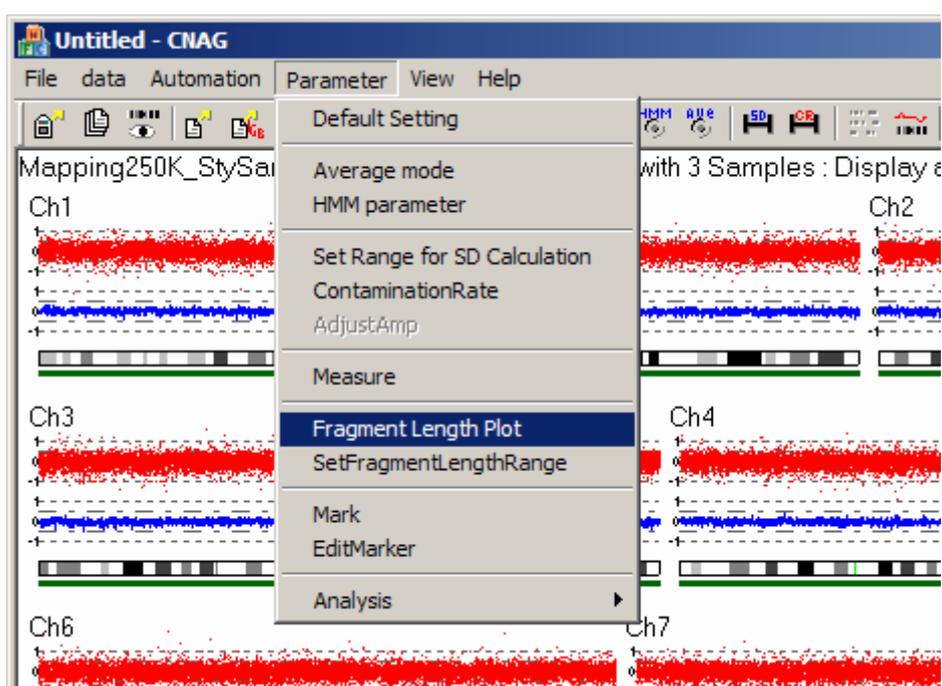
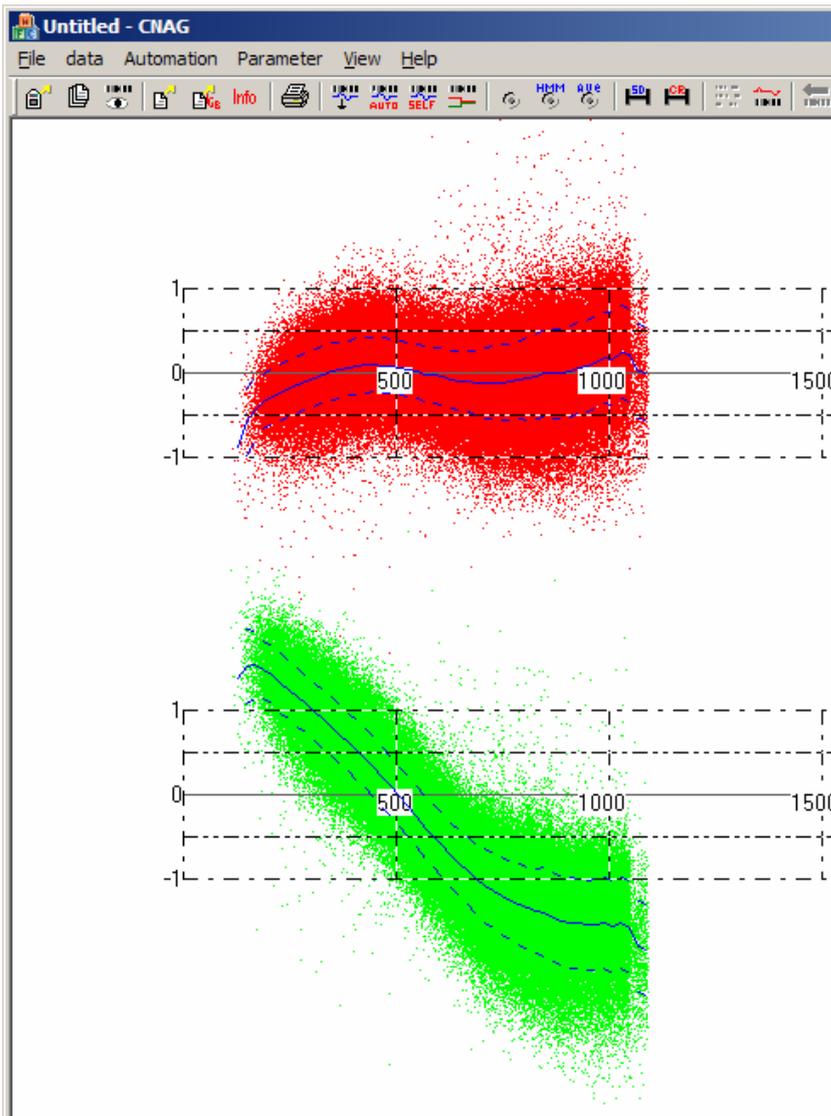


Figure 9

3. The data will now be displayed as copy number versus fragment size plots for that specific data set (Fig 10). Each dot in the plot represents a single SNP. The plots do not have labels, so here is a description for understanding the new viewer:
  - a. **Y-axis:** Copy number (if numbers vary from 0 to 4) or  $\log_2$  ratio (if numbers vary from -1 to 1). To switch between copy number and  $\log_2$  ratio, click on the “log” or “lin” icons at the top of the CNAG window (Fig 17).
  - b. **X-axis:** SNP Fragment Size. For the Mapping 500K array, the SNP fragment size will range from 100 to ~ 1100bp.
  - c. **GREEN:** raw data, before computed compensation against fragment size and GC content.

- d. **RED**: adjusted data, after compensation (by quadratic regression) to correct for biases due to fragment size and GC content.
- e. **BLUE Line**: mean copy number or mean  $\log_2$  ratio.
- f. **BLUE Dotted Line**: standard deviation for copy number or  $\log_2$  ratio predictions.

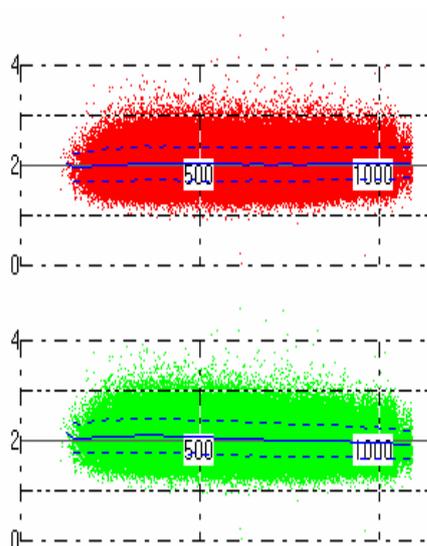


**Figure 10**

- g. In Fig 10, note that for FFPE samples, which tend to be degraded, contaminated, and of sub-optimal quality, there is an artifactual correlation between  $\log_2$  ratio and fragment size for the raw data (green dots). This is because larger fragments are poorly amplified

during PCR of low quality FFPE samples, resulting in artificially low probe intensities and consequently false predictions of low copy numbers. This effect of fragment size on CN prediction can only be seen if the reference is of high quality, standard DNA.

- i. In comparison, Fig 11 provides an example of a fragment size plot for a typical non-FFPE sample.



**Figure 11**

- h. Correction of fragment size bias (red dots) removes some of the effect of fragment size on CN, but for many samples, such as this one, this *in silico* correction is not sufficient for removing the entire bias.
  - i. If the mean CN (blue line) on the adjusted CN plot (red dots) was stable across fragment sizes, no fragment size exclusion would be necessary. This is the case for higher quality FFPE samples. For this sample, because CN varies along the X axis, fragment size exclusion is also necessary.
4. To choose the proper fragment size cutoff, it will be necessary to test a few cutoffs and watch as the blue line (mean copy number) in the red plot (after correction) indicates that copy number is no longer fragment size dependent **for the remaining SNPs retained in the analysis**.
    - a. The goal is to find the maximum fragment size cutoff that can be used without fragment size artifactually influencing CN estimation.
    - b. The maximum fragment size cutoff will enable the maximum

number of SNPs to be retained in the analysis.

5. Select "SetFragmentLengthRange" under the Parameters menu item while viewing the copy number versus fragment size plots (Fig 12).

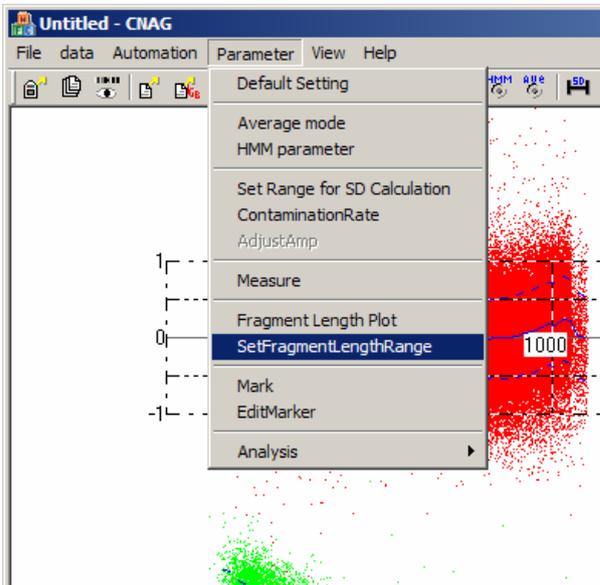


Figure 12

6. This will provide a box in which you can type in your preferred fragment size range (Fig 13).

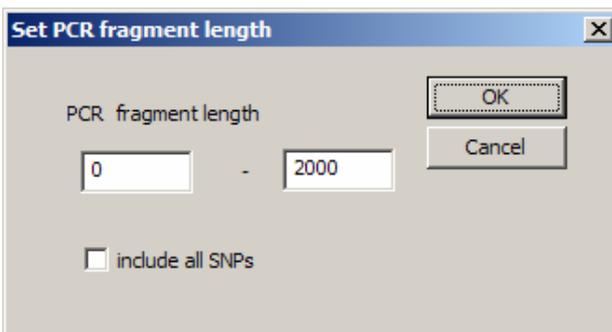
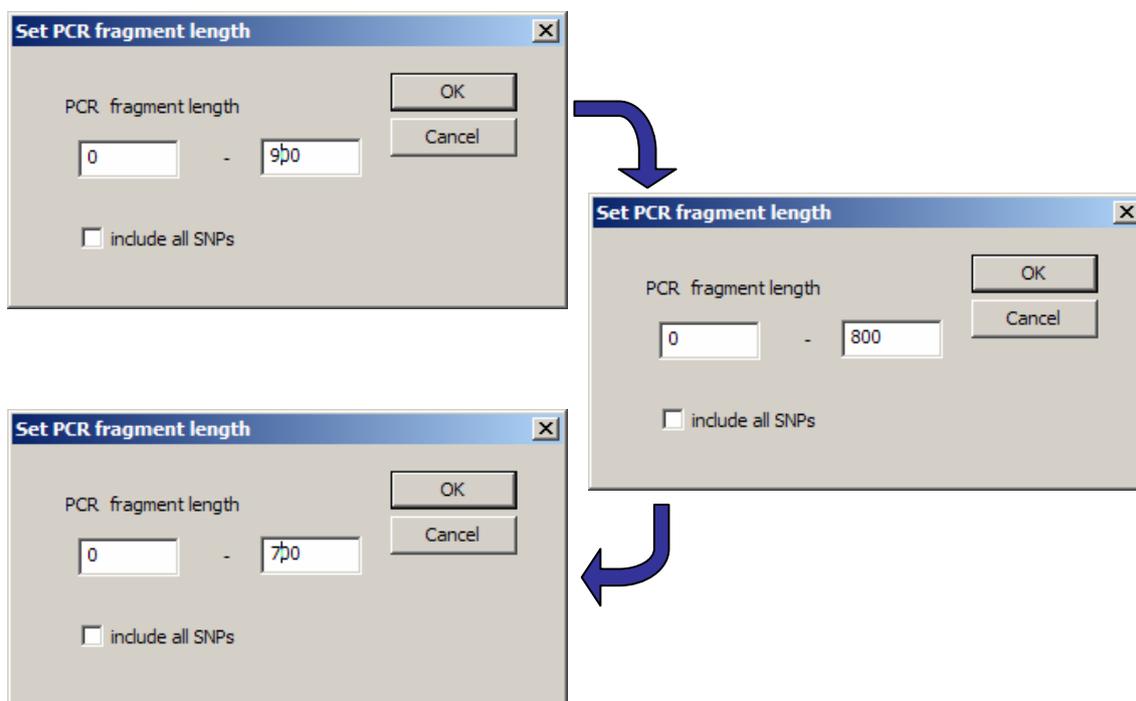


Figure 13

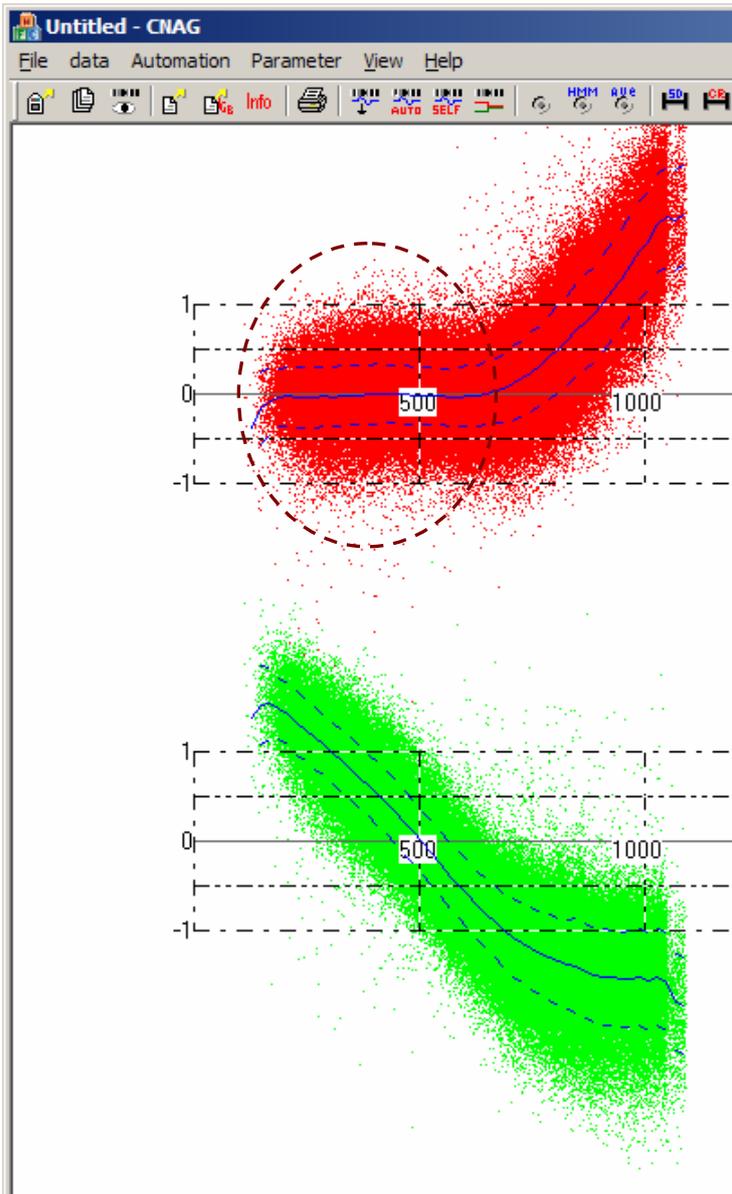
7. Test various sizes, until you see that the mean copy number is consistent

across retained fragment sizes (Fig 14).



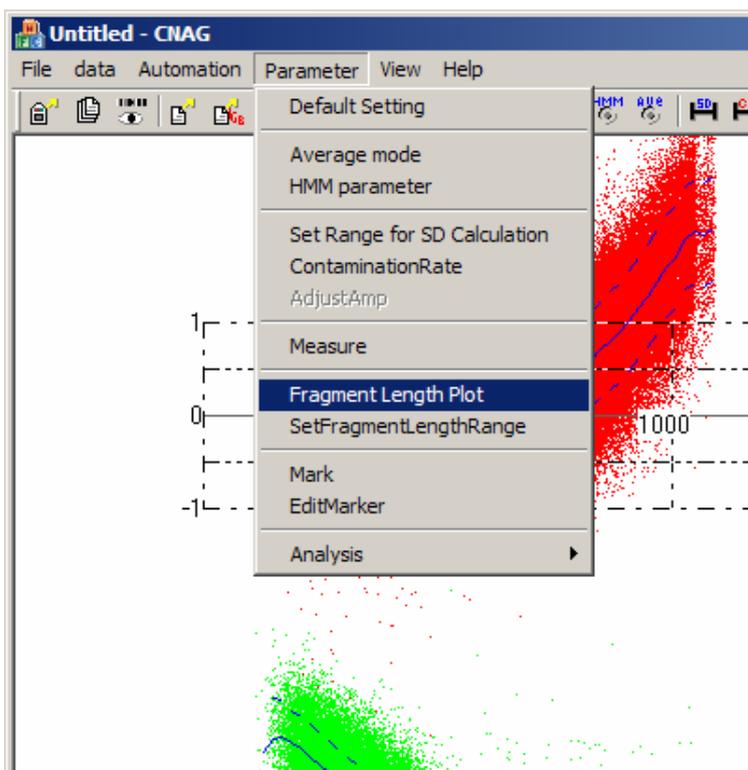
**Figure 14**

- a. Each time a new fragment size range is selected, CNAG will re-compute the data and re-apply the correction using this fragment size range.
- b. Select the fragment size limit that results in a consistent CN across retained fragment sizes. All SNPs will be displayed in the plot, but only those within the range specified need to present a stable CN prediction (Fig 15).



**Figure 15:** For SNPs on fragment sizes less than 700 bp, the  $\log_2$  ratio (similar to CN) is consistent across retained fragment sizes after adjustment (red dots, area within the circle). Therefore, 700 bp is a reasonable fragment size cutoff for this sample.

- Having selected and applied the correct fragment size cutoff for a given sample, return to the original CNAG viewer in which CN is displayed against genomic position for the entire genome or a single chromosome. This is done by again selecting "Fragment Length Plot" under the Parameter menu item (Fig 16).



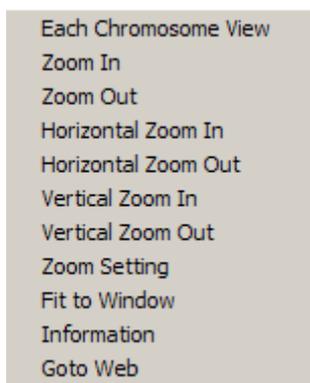
**Figure 16**

- Note that unless you re-set the fragment length range, CNAG will always remember your last choice for fragment restriction and maintain that choice. If you close CNAG and re-open and re-display the data for that sample, it will still be displayed with the same fragment size cutoff that was last selected.

## **VI. View Data**

### **VI.A. In CNAG**

1. Immediately after analysis, the data is displayed in the CNAG viewer.
2. To change the parameters for this display, it is possible to right-click on the window. This will give you the options for viewing an individual chromosome, zooming, gaining information about the references and samples, and linking to the web (Fig 17).



**Figure 17**

- a. The y-axis in the CNAG viewer can be changed from log-scale ( $\log_2$  ratio) to linear scale (copy number) by clicking on the icons at the top of the window, listed as "log" and "lin" (Fig 18).



**Figure 18**

- b. The display of average copy number (blue line) can be adjusted. By default, the line represents an average of 10 SNPs. To change this, click on the icon at the top of the window that says "Ave" with a bulls-eye on it, and enter a number between 1 and 50. Larger values will increase the smoothing of the signal.
- c. The default display shows copy number with consecutive SNPs lined up as though they are equidistant ("SNP distance"). To change this to show CN against true genomic position, click on the icon at the top of the window that says "base". To return to SNP distance instead of

genomic distance, click on the icon that says "SNP".

3. Once multiple samples have been analyzed, it is possible to switch between the displays of these different samples by selecting the "DisplaySamples" option under the data menu item (Fig 19).

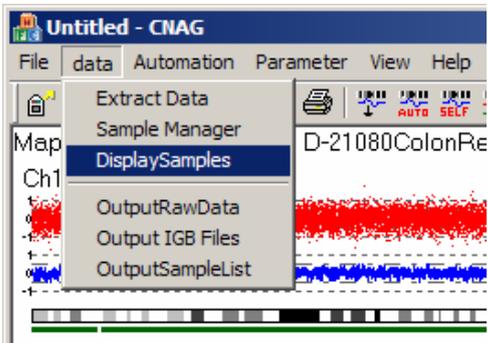


Figure 19

4. A dialog box will pop up to allow you to select samples to display (Fig 20)

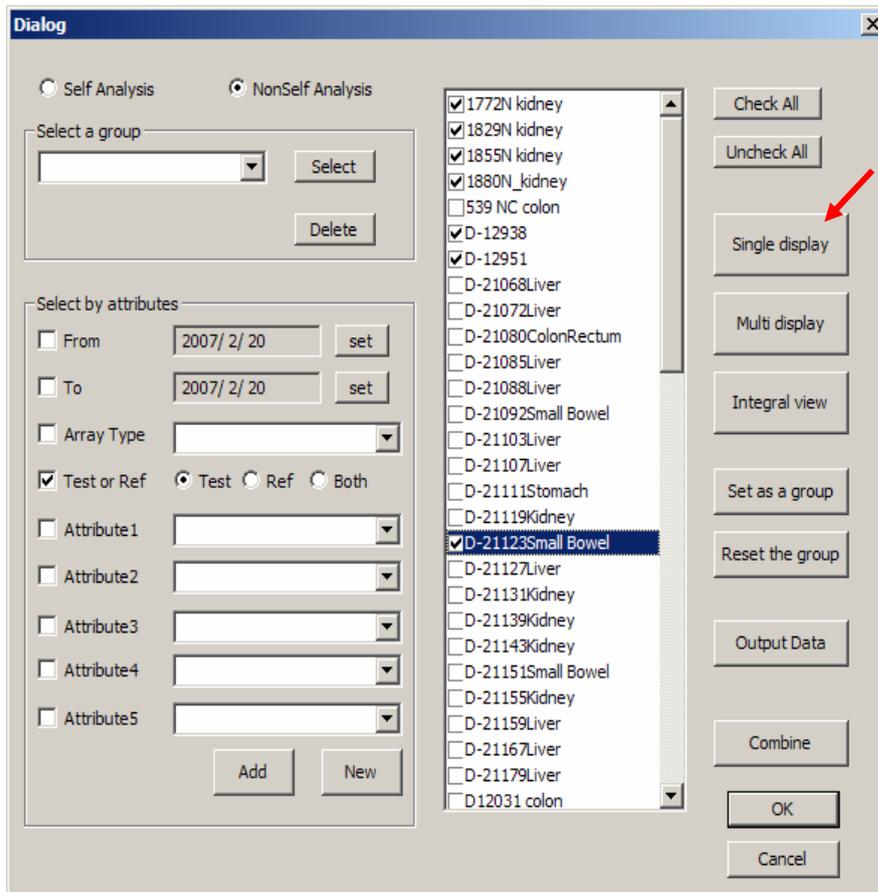


Figure 20

5. If the samples you wish to display are unpaired, select the “NonselfAnalysis” radio button at the top left of the “DisplaySamples” dialog window. If the samples are from paired analysis, select the “SelfAnalysis” radio button.
6. In the white box there will be a list of all samples that have been analyzed using either paired or unpaired analyses. Select the check boxes for any samples you wish to display.
7. Select “Single Display” (see red arrow, Fig 20).
8. After selecting “Single Display”, a new window will pop up the lists each of the samples that were selected in the dialog box (Fig 21).



**Figure 21**

9. To switch the view to different samples, simply click on the name of the sample you want to see. To view another sample, click on the data for another sample.
10. Each sample will be displayed using the same parameters that were defined for the previous sample, e.g. the same zoom, the same chromosome, the same plot-type.
11. Each sample will be displayed with the same fragment size restriction that had been defined for **that specific sample** the last time it was viewed.

## VI.B. In IGB

1. To view the data in the Integrated Genome Browser (IGB), export the data as an IGB file by selecting the "Output IGB Files" option under the data menu item (Fig 22).

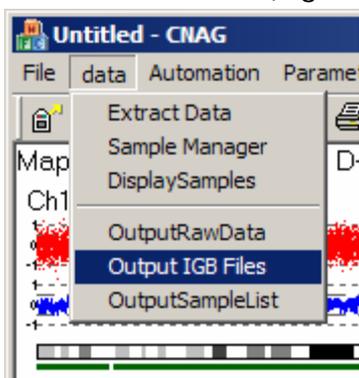


Figure 22

2. A window will pop up that allows you to choose what type of data should be placed in the IGB file (Fig 23). A separate IGB file will be created for each data type.

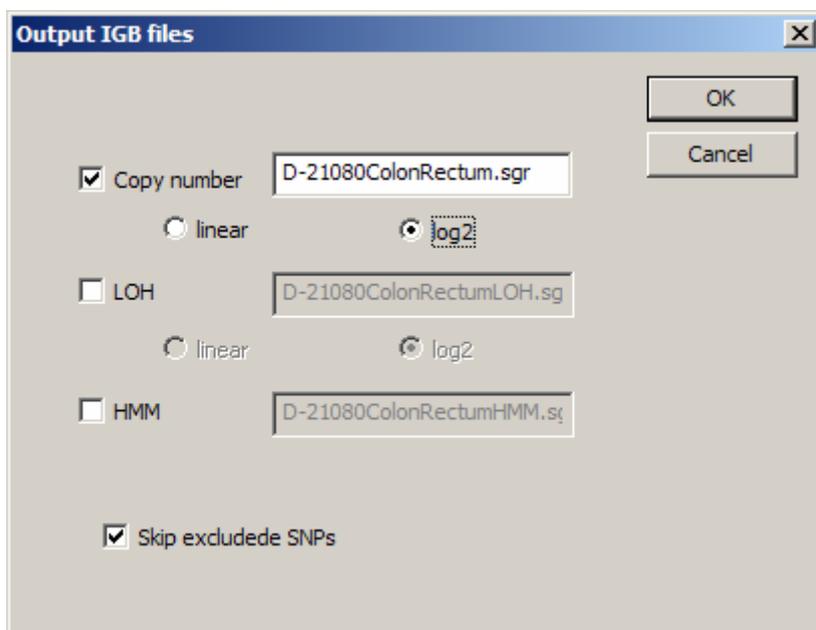


Figure 23

3. Note that if a fragment size filter was applied, it is important to check the box next to "Skip excluded SNPs" to remove the filtered SNPs from the export.

4. The IGB files will be placed in the folder assigned in section II.3.B as "Output".
5. IGB files are similar to text files, but have an ".sgr" extension on them.
6. To view the IGB file, go to the IGB browser. If you do not have this program, download it for free at the following website:  
[http://www.affymetrix.com/support/developer/tools/download\\_igb.affx](http://www.affymetrix.com/support/developer/tools/download_igb.affx)
7. Within IGB, open the file by finding it in the Output folder.

### VI.C. As a .txt file

1. The data can also be viewed as a text file. To create this file, select "OutputRawData" under the data menu item (Fig 24).

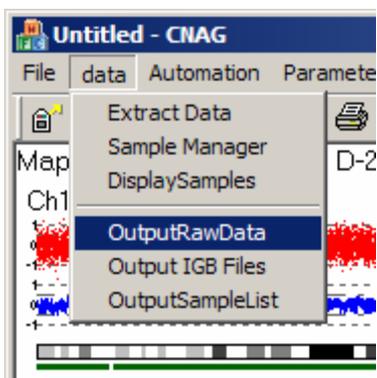


Figure 24

2. The raw data will be placed in the folder assigned in section II.3.B as "Output".
3. **Note that at this time, all SNPs are exported into the raw data file. The data is post-correction (i.e., represented by the red dots in the earlier plots). The data for the larger fragment SNPs that were excluded during analysis are exported into this data file and must be filtered out in another program.**

#### VII.D. View regions of interest in the UCSC browser

1. If there is a specific region of interest, it is possible to view it in the UCSC browser.
2. Select the red "M" icon at the top of the CNAG window (see red arrow, Fig 25). Then highlight the region of interest by clicking on the mouse at one end of the region, holding the mouse button down while dragging over the region, and releasing at the other end of the desired location. The highlighted region will be contained in a black box (as demonstrated in Fig 25).

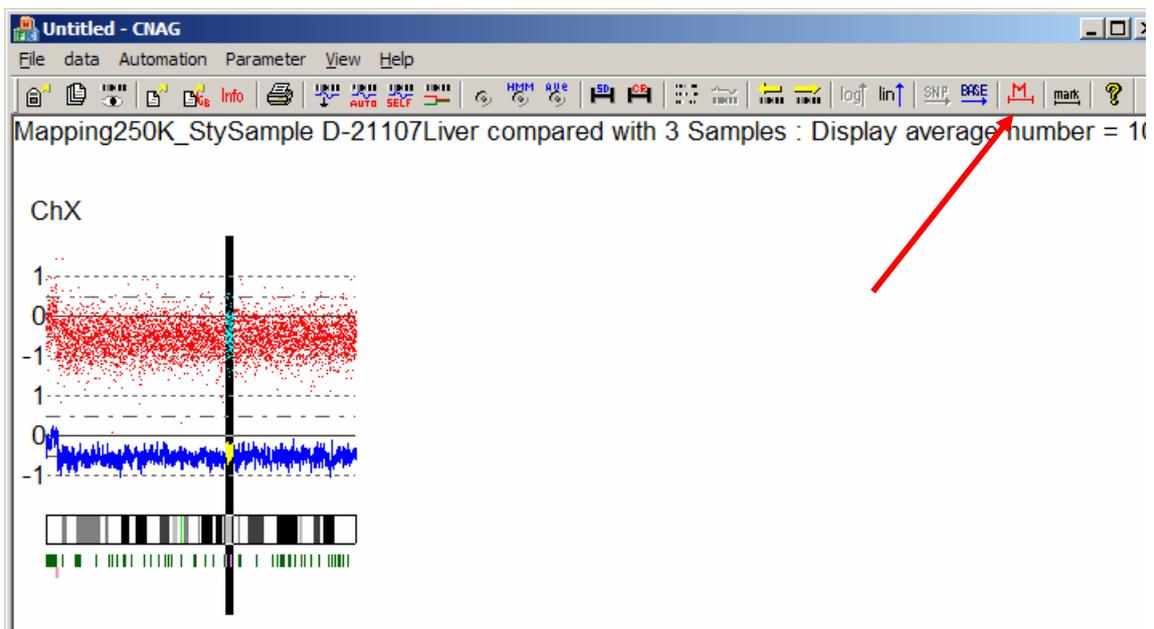


Figure 25

3. Once the area is selected, a new window will pop up that describes the location and size of this region. To view this specific region in the UCSC browser, select "Go to Web" within this pop up window (Fig 26).

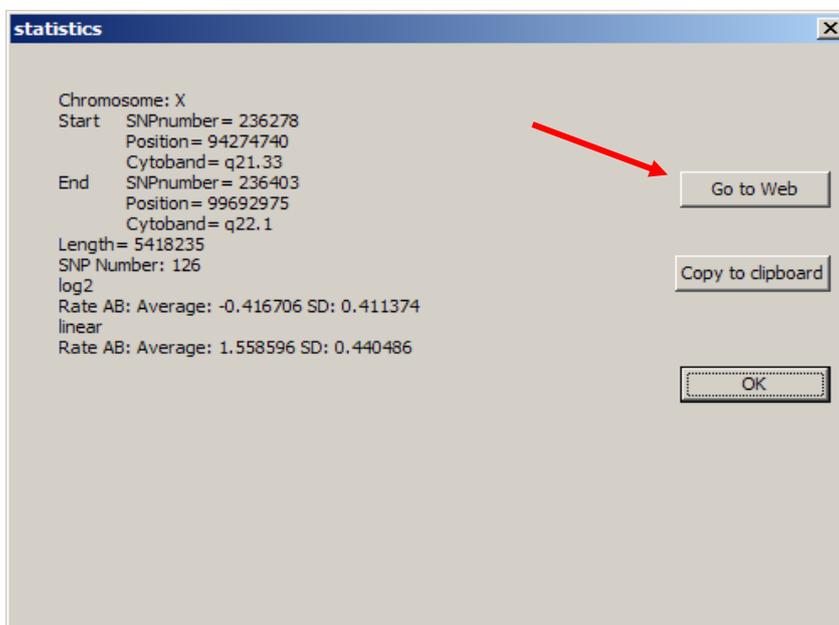


Figure 26

For more information about CNAG:

Nannya Y, Sanada M, Nakazaki K, Hosoya N, Wang L, Hangaishi A, Kurokawa M, Chiba S, Bailey DK, Kennedy GC, & Ogawa S. *Cancer Research* 65(14). July 15, 2005.

For more information about application of FFPE DNA samples to the Affymetrix Mapping 500K arrays:

Jacobs S, Thompson ER, Nannya Y, Yamamoto G, Pillai R, Ogawa S, Bailey DK, & Campbell IG. *Cancer Research* 67(6). March 15, 2007.