affymetrix

User Manual

Axiom CNV Summary Tool

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

P/N 703216 Rev 1

Trademarks

Limited License Notice

Patents

Copyright

Introduction

This document provides the general guidelines on how to use the **Axiom CNV Summary Tool**.

The **Axiom CNV Summary Tool** generates input files for BioDiscovery Nexus using Axiom data.

The included **Axiom CNV Viewer** allows you to view the data generated by the Axiom CNV Summary Tool. To use the Viewer, see *Viewing Data in the Axiom CNV Viewer on page 11*

Prerequisites

Before you can use the Axiom CNV Tools you must have:

- An appropriate Annotation File (*annot.db) downloaded from the Affymetrix NetAffx website. (User Account and Password required)
- Good quality genotying data processed through the Affymetrix Genotyping Console™ (GTC) or Affymetrix Power Tools™ (APT) application.

Recommended System Requirements

64-bit Operating System	Speed	Memory (RAM)	Available Disk Space	Web Browser
Microsoft Windows 7 Professional	2.83 GHz Pentium Quad Core Processor	16 GB	150 GB HD + Data Storage	IE 8.0 and above
Microsoft Windows 8 Professional	2.83 GHz Pentium Quad Core Processor	16 GB	150 GB HD + Data Storage	IE 8.0 and above

Installing Axiom CNV Summary Tool

To install the Axiom CNV Summary Tool:

1. Go to **www.affymetrix.com** and navigate to the following location:

   Home > Products > Microarray Solutions > Instruments and Software > Software >
2. Locate and download the zipped ***Axiom CNV Summary Tool*** software package.
3. Unzip the file, then double-click **AxiomCNVSetup.exe** to install it.
4. Follow the installer's instructions.

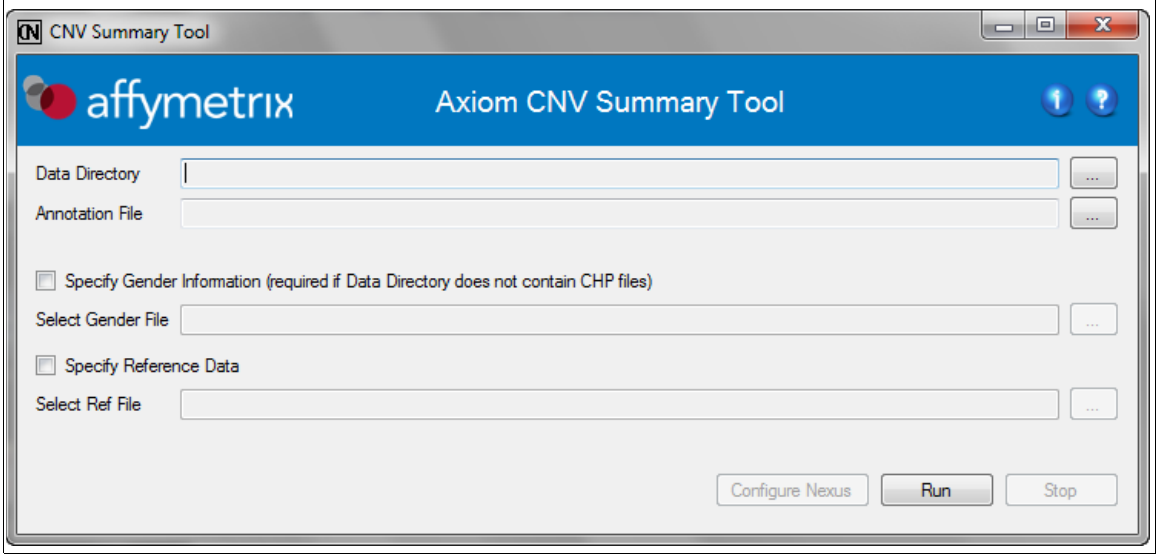
Starting Axiom CNV Summary Tool

To start the CNV Summary Tool:

1. Click **Start** -> **All Programs** -> **Affymetrix** -> **Axiom CNV Tools**.

2. Locate and click on **Axiom CNV Summary Tool**.
   The application opens. (Figure 1.1)

Figure 1.1 CNV Summary Tool - Main window



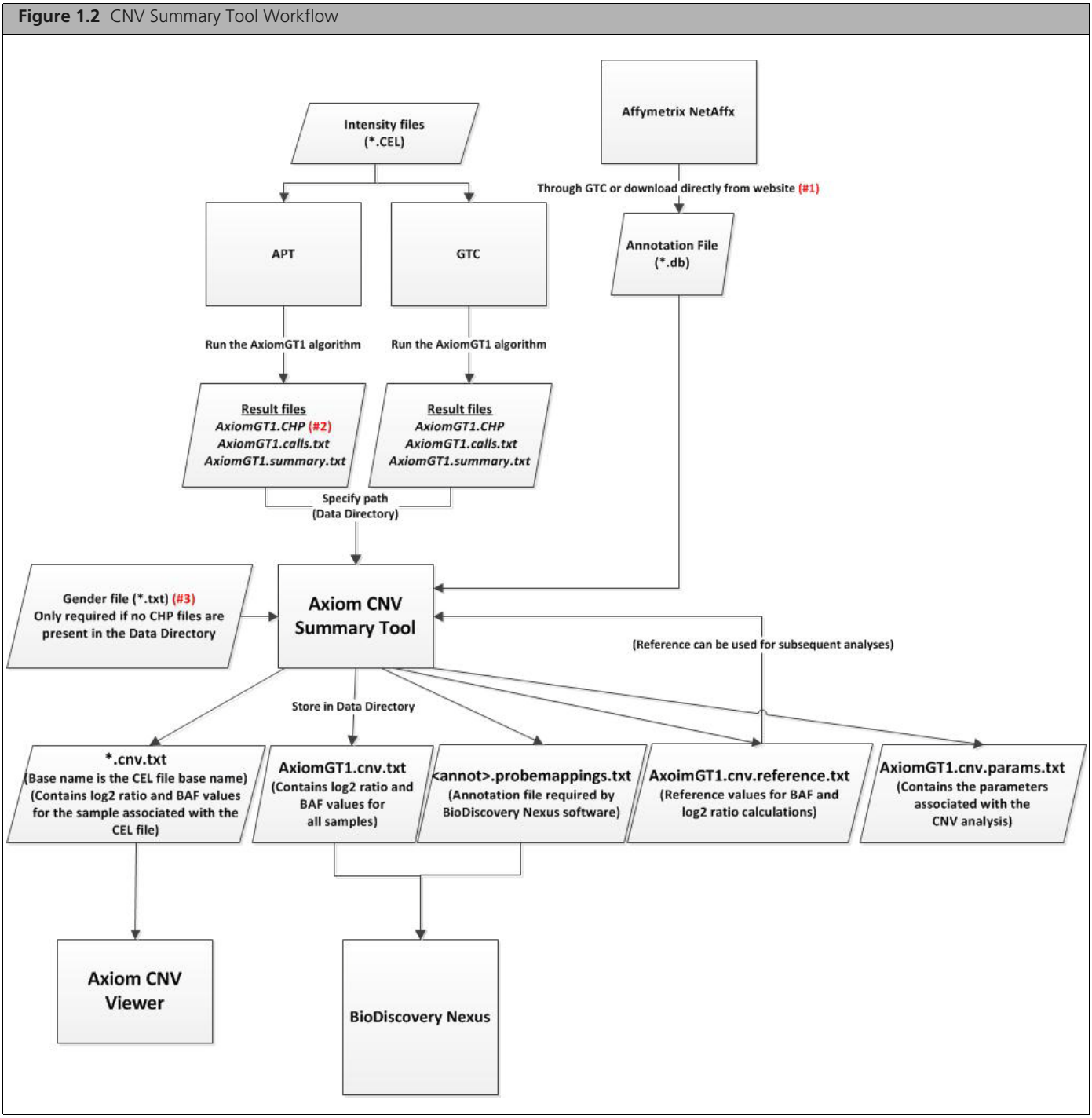
Overview of Workflow and Useful Tips

Overview of Workflow

Please review the diagram below. It shows the data analysis workflow starting with GTC or APT and ending with data for use with BioDiscovery Nexus.

The file types that are required and generated by the Axiom CNV Summary Tool are also noted in this diagram. (Figure 1.2)

Figure 1.2 CNV Summary Tool Workflow



Useful Tips

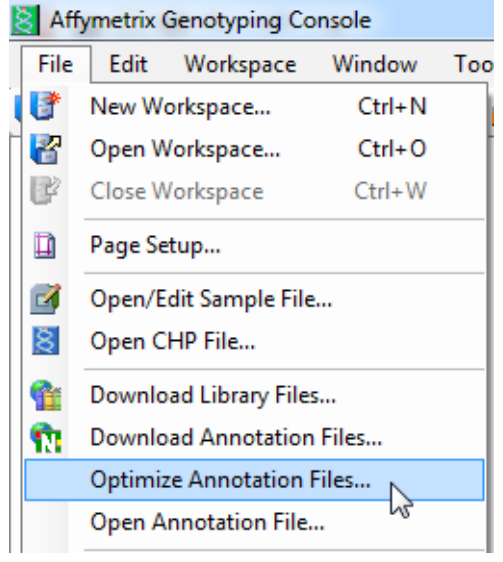
Tip #1

When you download an annotation file using GTC, the annotation file is automatically indexed for optimum processing.

If your annotation file was not downloaded from GTC, you can optimize it for faster processing using the GTC software, as follows:

Click **File** -> **Optimize Annotation File...** (Figure 1.3)

Figure 1.3 GTC File menu



Tip #2

To create CHP files in ATP, use the parameter: ***--cc-chp-output***

Tip #3

You can create your own gender.txt file by using the information in the *AxiomGT1.report.txt* file. This file is located in your *Genotyping Results* folder.

The second column in the report.txt file is *computed_gender*. (Figure 1.4)

Select, then copy the gender information you want to use from the report.txt file, then paste it in a new *gender.txt* file.

See Figure 2 on page 9 for gender.txt file format requirements, then follow steps; *To include gender values for all samples: on page 9*.

Figure 1.4 Axiom GT1.report.txt file example

AxiomGT1.report.txt

```
#%affymetrix-algorithm-param-gender_method_used=cn-probe-ch
#%affymetrix-application-meta-data-info-posterior-file=2013
#%affymetrix-application-meta-data-info-configuration-file=
#%affymetrix-application-meta-data-info-Reagent Version=2
#%guid=00006917-231b-45a7-3a2d-003e48006a62
cel_files   computed_gender   call_rate   total_call_rate   het
NA18523_200ng_Exome319_24hr_20120509_scan2_F08.CEL   female
NA18858_200ng_Exome319_24hr_20120509_scan2_A10.CEL   female
NA18859_200ng_Exome319_24hr_20120509_scan2_G11.CEL   male
NA18860_200ng_Exome319_24hr_20120509_scan2_H09.CEL   male
NA18870_200ng_Exome319_24hr_20120509_scan2_G01.CEL   female
NA18871_200ng_Exome319_24hr_20120509_scan2_F04.CEL   male
NA18914_200ng_Exome319_24hr_20120509_scan2_E11.CEL   male
NA19092_200ng_Exome319_24hr_20120509_scan2_H08.CEL   male
NA19102_200ng_Exome319_24hr_20120509_scan2_G08.CEL   female
NA19127_200ng_Exome319_24hr_20120509_scan2_E12.CEL   female
NA19131_200ng_Exome319_24hr_20120509_scan2_D10.CEL   female
NA19132_200ng_Exome319_24hr_20120509_scan2_B12.CEL   female
NA19154_200ng_Exome319_24hr_20120509_scan2_D05.CEL   male
NA19159_200ng_Exome319_24hr_20120509_scan2_E08.CEL   female
NA19161_200ng_Exome319_24hr_20120509_scan2_F10.CEL   male
NA19171_200ng_Exome319_24hr_20120509_scan2_G09.CEL   male
NA19201_200ng_Exome319_24hr_20120509_scan2_E01.CEL   female
NA19205_200ng_Exome319_24hr_20120509_scan2_G06.CEL   male
NA19206_200ng_Exome319_24hr_20120509_scan2_C09.CEL   female
NA19211_200ng_Exome319_24hr_20120509_scan2_F05.CEL   male
```

Axiom CNV Summary Tool

Before you can run the Axiom CNV Summary Tool, you must first process the CEL files through one of the following software applications:

- Affymetrix Genotyping Console (GTC) v4.1.4 (or higher)
- Affymetrix Power Tools (APT)

GTC Users

Do the following:

1. Run the AxiomGT1 algorithm through GTC (using your *.CEL Intensity files) as you normally would.

GTC produces the following Result files:

- CHP files
- AxiomGT1.calls.txt
- AxiomGT1.summary.txt

APT Users

Do the following:

1. Run the AxiomGT1 algorithm through APT (using your *.CEL Intensity files) as you normally would.

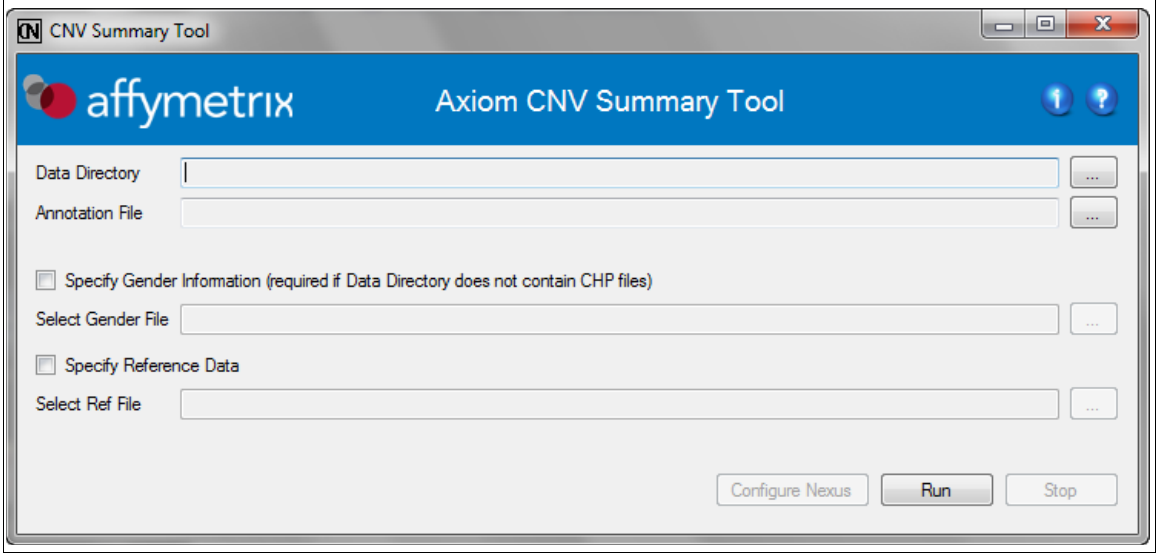
APT produces the following Result files:

- AxiomGT1.calls.txt
- AxiomGT1.summary.txt
- CHP files (APT's ***--cc-chp-output*** option must be enabled)

Using the Axiom CNV Summary Tool

From the Axiom CNV Summary Tool window: (Figure 1)

Figure 1 CNV Summary Tool Main window



1. Click the **Data Directory** Browse button.
2. Navigate to the folder that contains your input data files (*AxiomGT1.CHP, AxiomGT1.calls.txt, and AxiomGT1.summary.txt files), then single click, Ctrl click, Shift click, or Ctrl-A (to select multiple files).
3. Click **OK**.

   The Data Directory path is now populated.
4. Click the **Annotation File** Browse button.
5. Navigate to the folder that contains the annotation file you downloaded earlier from www.affymetrix.com.

NOTE: Annotation files are array specific. If you are running an analysis for a specific array, make sure you use the appropriate annotation file.

Annotation files for *Axiom myDesign* arrays are provided directly to you from Affymetrix (for each custom array designed).

When you download an annotation file using GTC, the annotation file is automatically indexed for optimum processing. If your annotation file was not downloaded from GTC, see *Tip #1 on page 5*.

6. Click to select the annotation file, then click **OK**.

   The Annotation File path is now populated

Select Gender File.

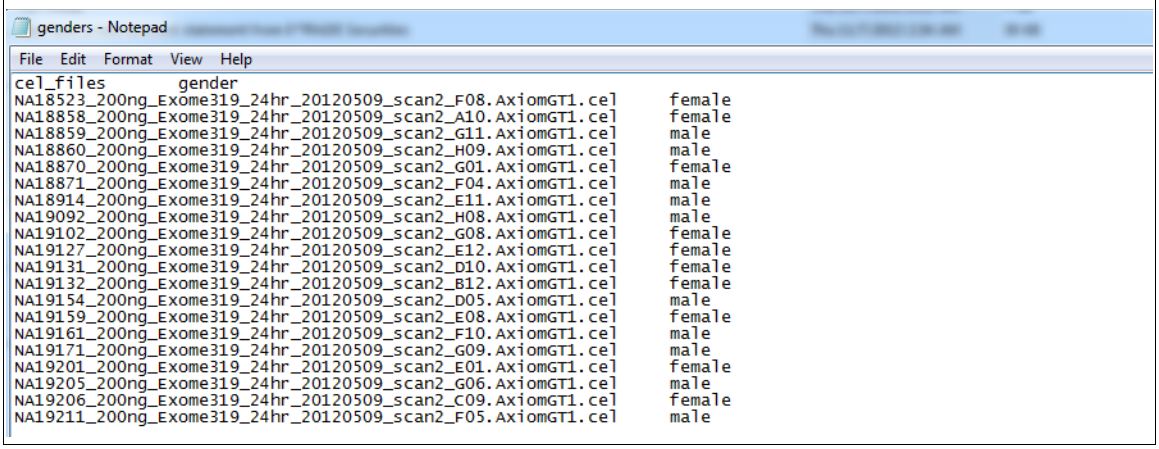
IMPORTANT: If there are no CHP files present in your Data Directory folder, you MUST include gender values for all samples. If you are using CHP files, skip this *Select Gender File* option.

To include gender values for all samples:

1. Click the ***Specify Gender information*** checkbox.
2. Click the **Gender File** Browse button.
3. Navigate to the folder that contains your Gender files.

NOTE: Your gender file must be a tab-delimited text file with 2 columns. Its first column header must be *cel_files.* The second column header must be *gender*, as shown in Figure 2.

Figure 2 Gender Text File contents example.



The Gender column (far right) (Figure 2) is not case-sensitive.

- For *Female* type: **Female, female, F,** or **2**
- For *Male* type: **Male, male, M,** or **1**
- To specify an *unknown* gender type: **unknown** or **0**

  If no gender was specified or the gender was specified other than the required naming conventions stated above, the gender entry will be treated as *unknown*.

4. Click to select the *gender.txt* file you want to use, then click **OK**.

   The Gender File path is now populated.

Reference File (Optional)

The choice of samples to be used as a reference is critical for accurate CNV detection because the log2ratio at a marker is computed by dividing the intensity of the marker by the median intensity of that marker in the chosen reference set, in log space. The reference set, therefore, should represent the normal copy number state for each marker. One approach is to create the reference based on the individuals genotyped on the plate, provided that for each marker the vast majority of individuals on the plate are expected to have normal copy number states. Another approach is to create a separate reference based

on individuals expected to have normal copy number states genotyped on different plates. If the latter approach is chosen the number of samples used for the reference should be as large as possible, preferably at least 100. The analysis can be carried out with any number of samples but will be less accurate for smaller reference sets.

Do the following to specify reference data:

1. Click the ***Specify Reference Data*** checkbox.
2. Click the **Select Ref File** Browse button.
3. Navigate to the folder that contains your reference data file.
4. Click to select the file you want, then click **OK**.

   The Select Ref File path is now populated.

Running the Axiom CNV Summary Tool

1. After your Axiom CNV Summary Tool data paths are set, click **Run**.

   A green progress bar appears. Processing time varies depending on the amount of data you are processing, the number of SNPs on your array, and your system's hardware specifications.

   After the data has been successfully processed, a message appears.
2. Click **OK** to acknowledge the message.

Retrieving the Axiom CNV Summary Tool Data

The following files are produced and are stored in the **Data Directory** folder you assigned earlier:

- ***.cnv.txt** - Base name is the CEL file base name. (This file contains log2 ratio and BAF values for the sample associated with the CEL file.)
- **AxiomGT1.cnv.txt** - Contains log2 ratio and BAF values for all samples.
- **<annot>.probemappings.txt -** Required by BioDiscovery's Nexus software.
- **AxoimGT1.cnv.reference.txt** - Reference values for BAF and log2 ratio calculations.
- **AxiomGT1.cnv.params.txt** - Contains the parameters associated with the CNV analysis.

Ways to Use the Axiom CNV Summary Tool Data

Subsequent Analyses

Use the newly generated **AxoimGT1.cnv.reference.txt** for additional analysis.

1. Click the **Specify Reference Data** checkbox.
2. Click the Select Ref File Browse button, navigate to your Data Directory folder, then click to select the file: **AxoimGT1.cnv.reference.txt**
3. After your Axiom CNV Summary Tool data paths are set, click **Run**.

A green progress bar appears. Allow time for your data to process.

After the data has been successfully processed, a message appears.

4. Click **OK** to acknowledge the message.

Viewing Data in the Axiom CNV Viewer

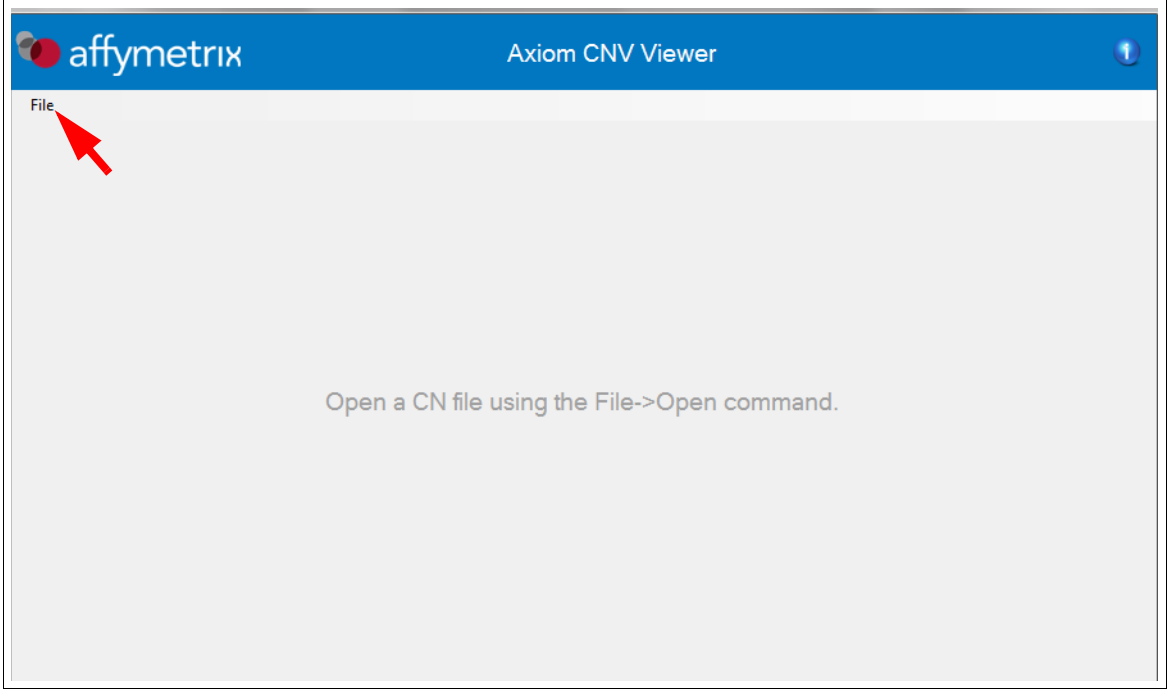
View the newly generated ***.cnv.txt** for additional analysis in the included Axiom CNV Viewer.

To start the Axiom CNV Viewer:

1. Click **Start** -> **All Programs** -> **Affymetrix** -> **Axiom CNV Tools**.
2. Locate and click on **Axiom CNV Viewer**.

   The application opens. (Figure 3)

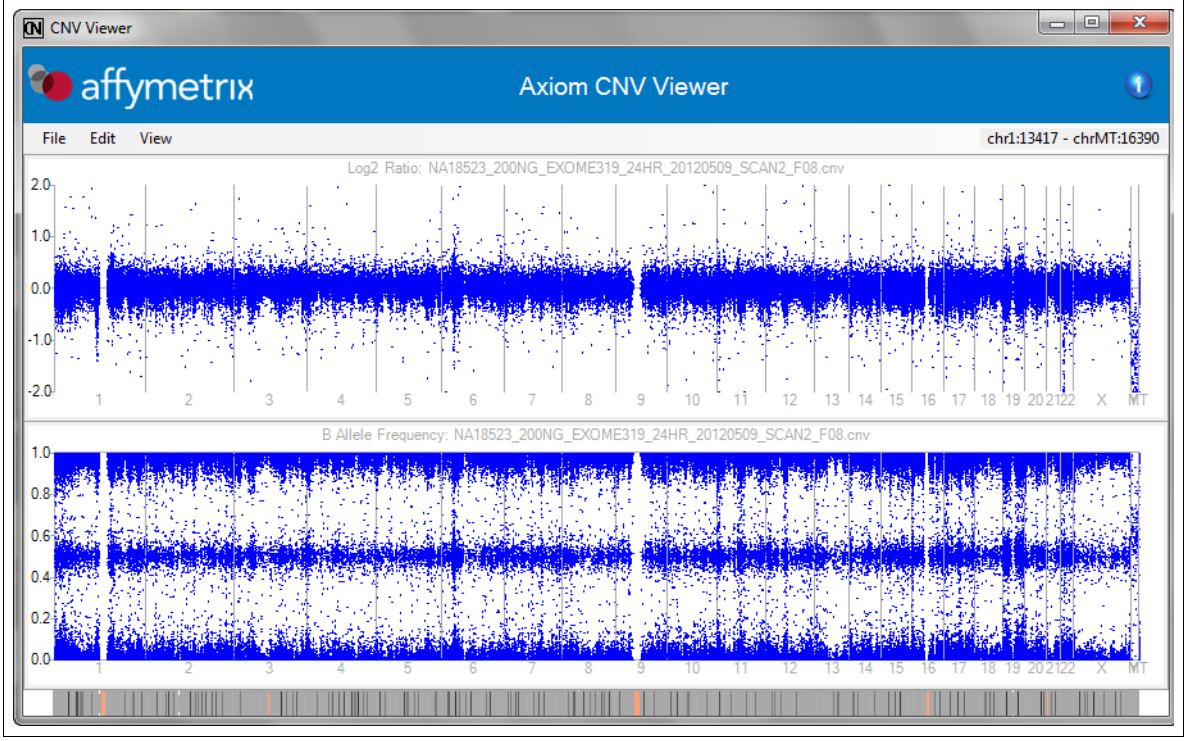
Figure 3 CNV Viewer - Startup window



3. Click **File** -> **Open**.
4. Navigate to your Data Directory folder, then select the ***.cnv.txt** file(s) you want to view.
5. Click **OK**.

The Viewer displays your data. (Figure 4)

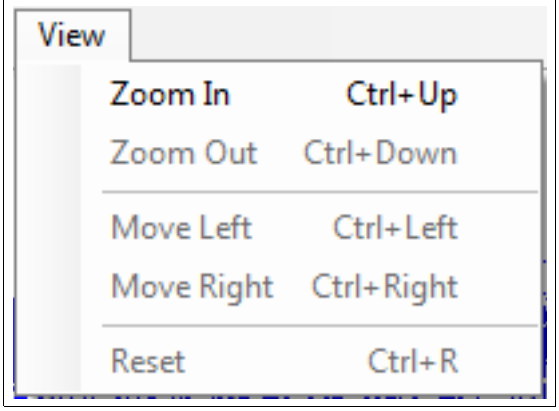
Figure 4 CNV Summary Tool - Main window populated



To customize the display view:

1. Click **View**, then click to select one of the following viewing options or use the equivalent keyboard commands shown. (Figure 5)
2. Repeat the viewing command as needed to reach the desired view.

Figure 5 CNV Summary Tool - View drop-down menu



To reset your customized view back to the default whole genome view:

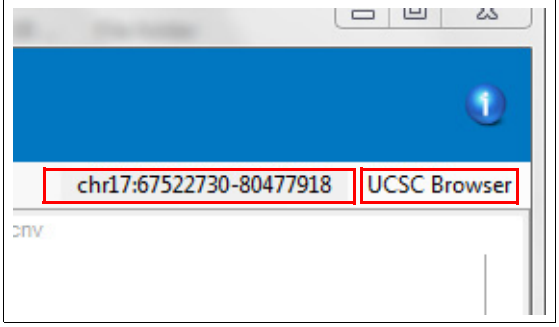
1. Click **Reset**.

To use the CNV Viewer to investigate your copy number changes:

Option #1

1. Use the *Zoom In* command or click, then drag your mouse cursor across a region of interest.
2. Once the Viewer has zoomed into a chromosome, a *UCSC Browser* button appears (upper right corner). (Figure 6)

Figure 6 UCSC Browser link and chr text filed.



3. Click on the **UCSC Browser** button.
   The UCSC website page appears (Figure 7) and displays the current region based on the chromosome positions listed in the *chr* text box. (Figure 6)

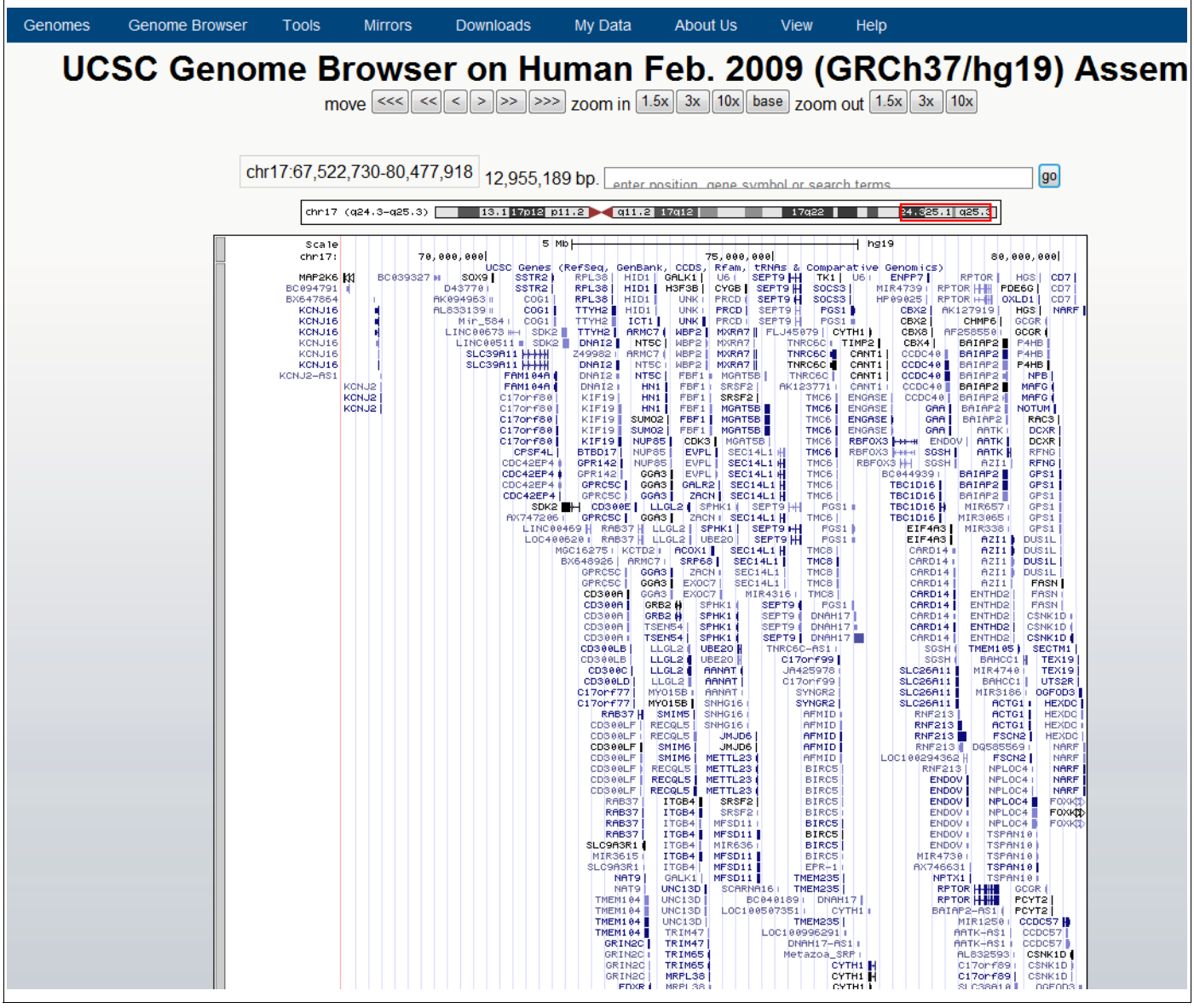
Option #2

1. Click inside the *chr* text box (Figure 6), then manually enter your chromosome positions. You must use one of the following formats:
   - **chr17:67522730-80477918**
   - **chr17:67,522,730-80,477,918**
2. Press **Enter.**

NOTE: The region displayed in the CNV Viewer may be smaller than the chromosome positions you entered, because the CNV Viewer auto-adjusts your start and stop positions next to the nearest available start and stop markers.

3. (Optional) Click on the **UCSC Browser** button.
   The UCSC website page appears (Figure 7) and displays the current region based on the chromosome positions listed in the *chr* text box. (Figure 6)

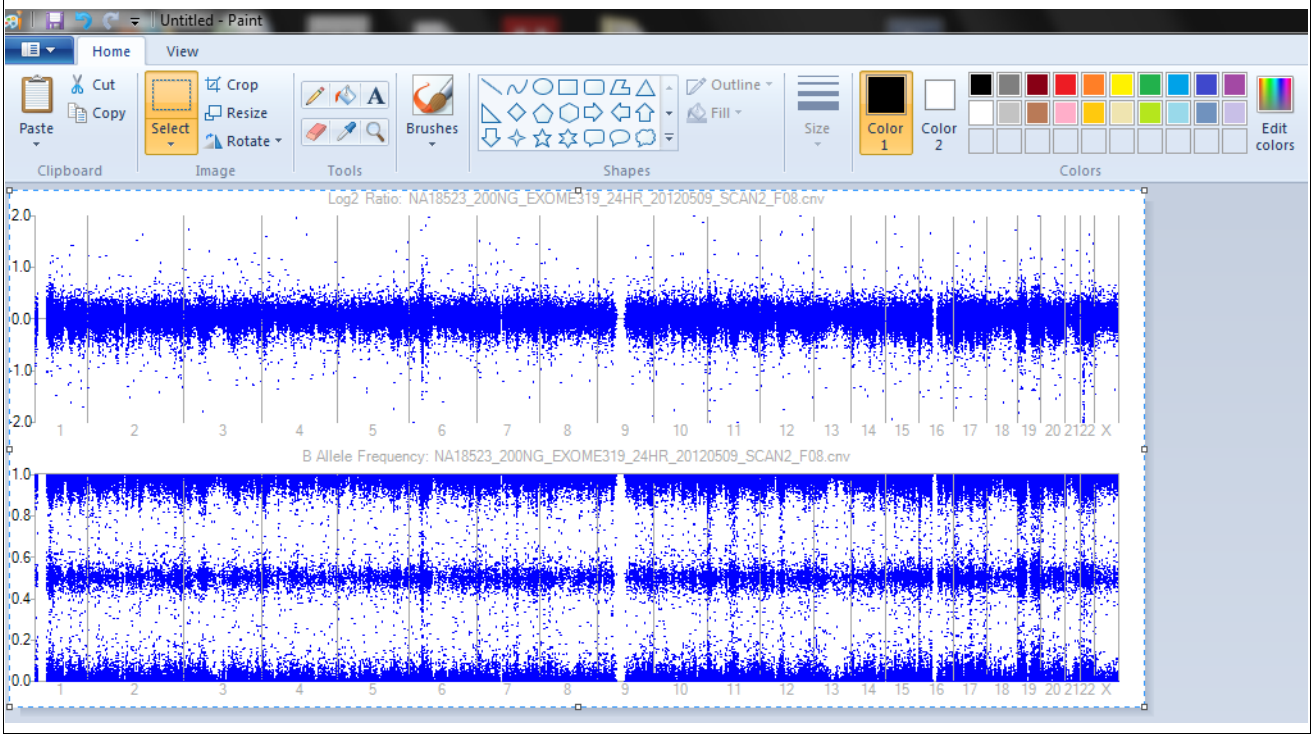
Figure 7 UCSC website page



To copy the current view to your Clipboard:

1. Click **Edit** -> **Copy to clipboard**.
2. Use the paste command (**Ctrl-V**) to copy the current view into another software application, such as MS Paint. (Figure 8)

Figure 8 Paste to Clipboard - MS Paint Example



Further Copy Number Analysis Using BioDiscovery's Nexus Software

Use the newly generated **AxiomGT1.cnv.txt** and **<annot>.probemappings.txt** with BioDiscovery's Nexus software to perform copy number analysis.

Do the following to configure Axiom CNV Summary Tool output data to work with BioDiscovery's Nexus software:

1. Click the CNV Summary Tool's Configure Nexus button (bottom right).

NOTE: If Nexus is not detected on your system, the *Configure Nexus* button is disabled.

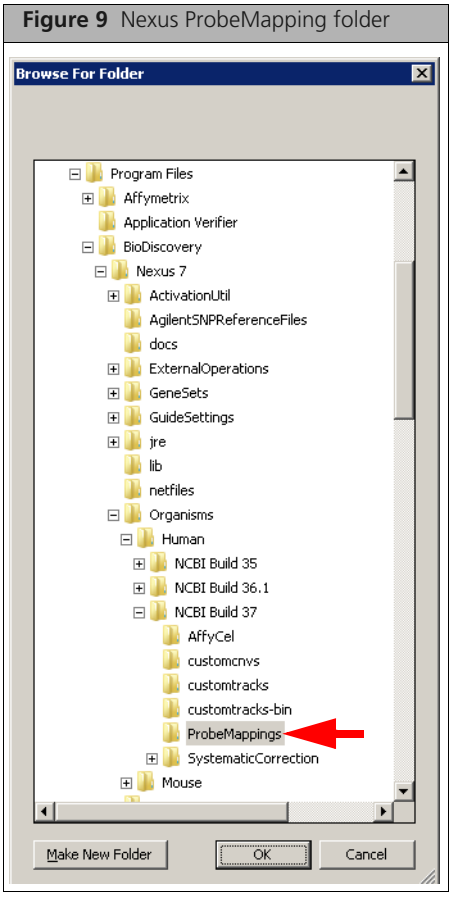
If multiple versions of Nexus are detected, a drop-down menu appears. Use this menu to select the appropriate version of Nexus. This menu does not appear if only one version of Nexus is detected.

A file window appears.

2. Click to select the *probe mapping .txt* file. This file resides in your master Data Directory folder you setup earlier. See Step 1 on page 8.
3. Click **Open**.

   An Explorer window appears. (Figure 9)
4. Navigate Nexus's *ProbeMappings* directory/folder, then click **OK**.

Figure 9 Nexus ProbeMapping folder



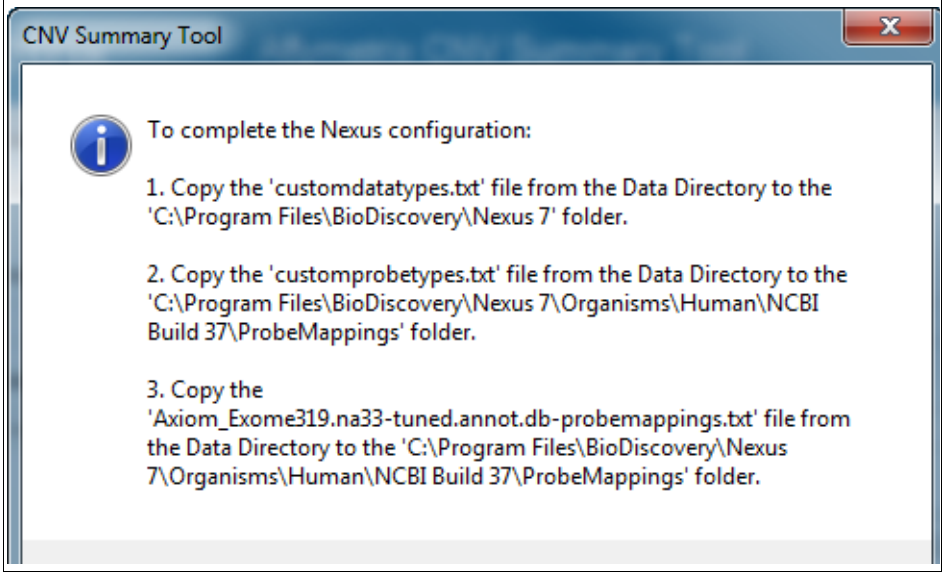
The message *Configuration Complete* appears.

5. Click **OK**.

IMPORTANT: You are responsible for knowing the location of Nexus's *ProbeMappings* folder. If you are unsure of its location, contact BioDiscovery. In most cases, the Nexus Probe Mapping folder resides here: *C:\Program Files\BioDiscovery\NexusX*

If you do not have access (Administrator Privileges) to some of your computer's folders, a message with file copying instructions appears. (Figure 10) Follow the 3 steps shown to configure the Nexus software manually.

Figure 10 To complete the installation message



6. Use the Nexus software as you normally would.

   If you want to perform a GC Correction in Nexus, see *Performing GC Correction in Nexus on page 18*

Performing GC Correction in Nexus

IMPORTANT: You must first download an appropriate BED file from affymetrix.com.

To download a BED file from affymetrix.com:

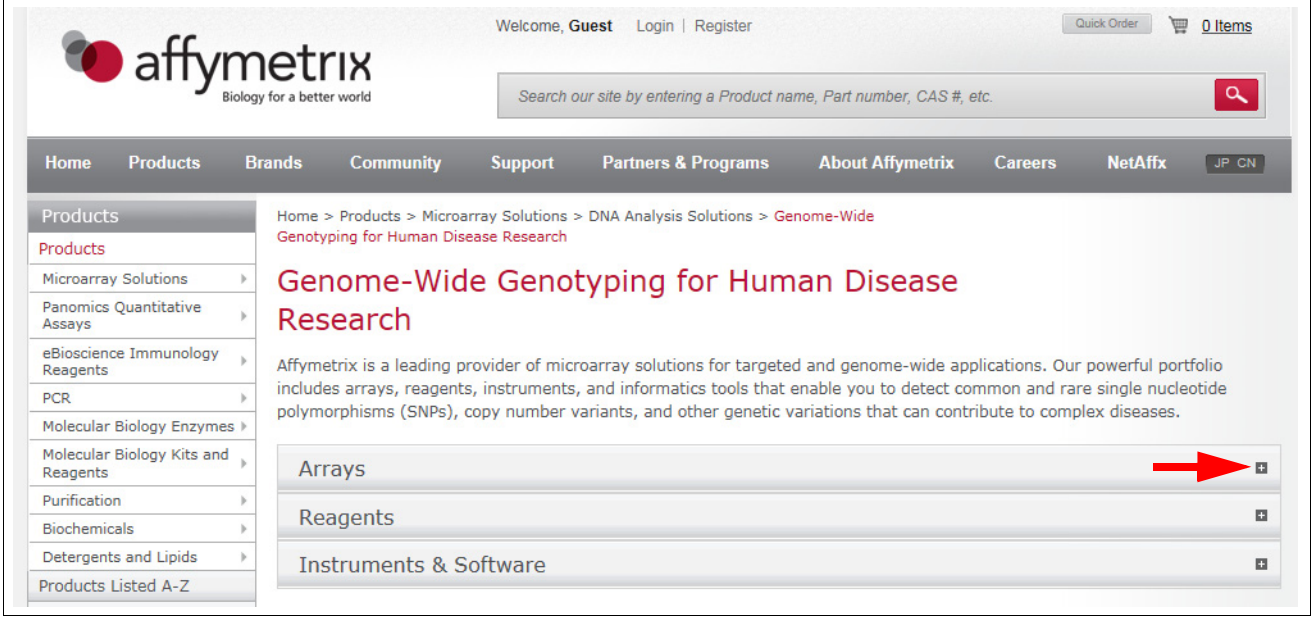
1. Go to *www.affymetrix.com*.
2. Login as you normally would or click **Register**, then follow the on-screen instructions.
3. Click **Products** -> **Products** (top left).

   The Products page appears.
4. Click **Microarray Solutions** (left pane).

   The Microarray Solutions pane appears.
5. Under the **DNA Analysis Solutions** header, click to choose the option you want. Example: *Genome-Wide Genotyping for Human Disease Research*.

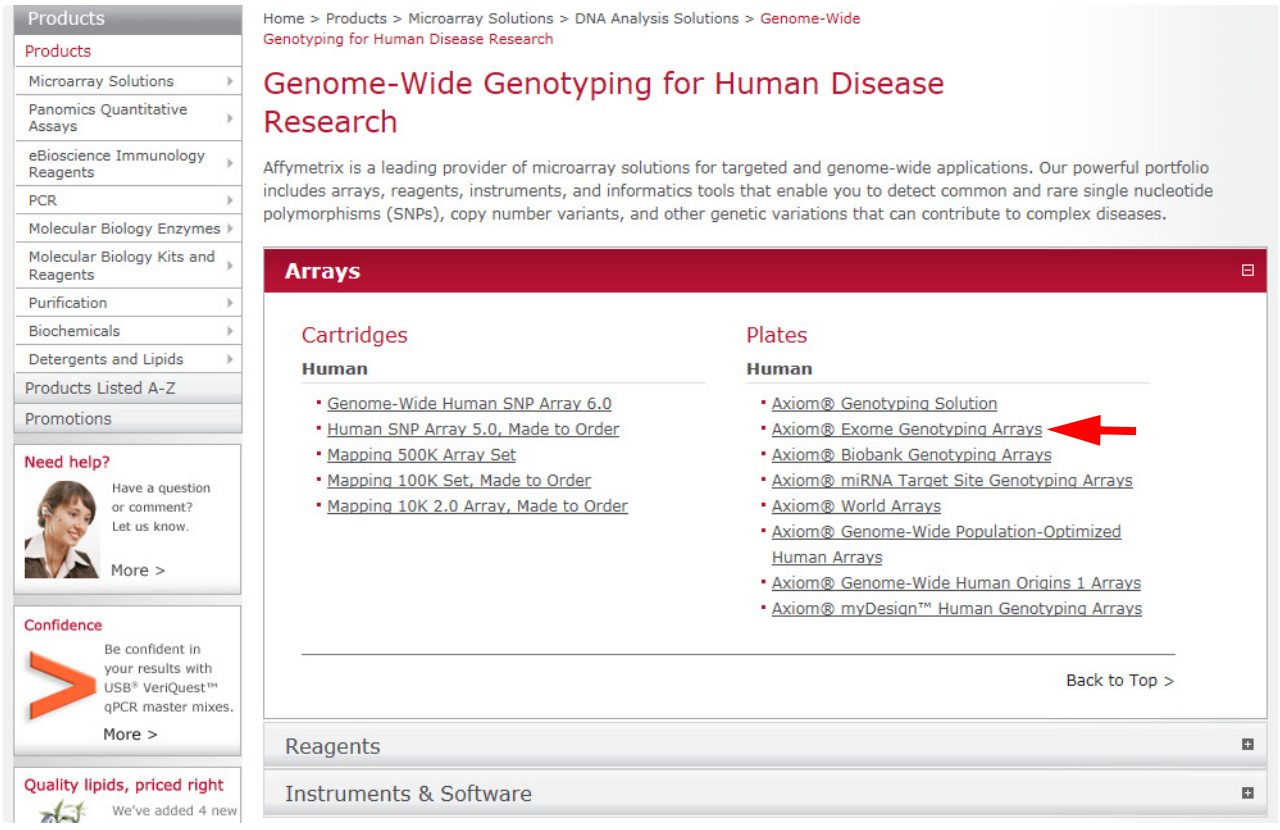
   For the *Genome-Wide Genotyping for Human Disease Research* example, 3 options appear.
6. Click the *Arrays* adjacent **[+]** button. (Figure 3.1)

Figure 3.1 Arrays menu



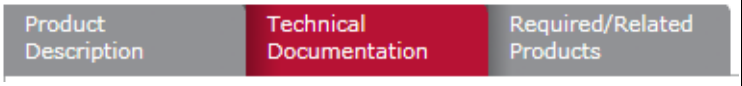
For the *Genome-Wide Genotyping for Human Disease Research* example, the following page appears. (Figure 3.2)

Figure 3.2 Arrays menu



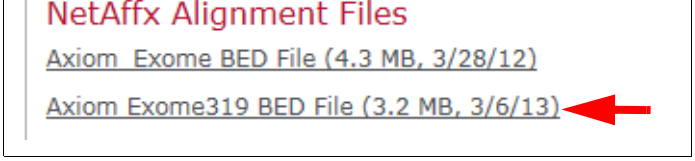
7. For this example, click the **Axiom® Exome Genotyping Arrays**. (Figure 3.2)
   The *Axiom® Exome Genotyping Arrays* page appears.
8. Click on the **Technical Documentation** tab. (Figure 3.3)\

Figure 3.3 Technical Documentation tab.



9. Scroll down and locate *NetAffx Alignment Files*, then (for this example) click on **Axiom Exome319 BED File**. (Figure 3.4)

Figure 3.4 Technical Documentation tab.



A Windows Explorer window appears.

NOTE: BED files for *Axiom myDesign* arrays are provided directly to you from Affymetrix (for each custom array designed).

10. Save the zip file to a convenient location.

Do the following to submit your BED file to BioDiscovery for GC Correction:

1. Extract the downloaded BED.zip file, then contact BioDiscovery and tell them you need a GC Correction file created from a BED file.

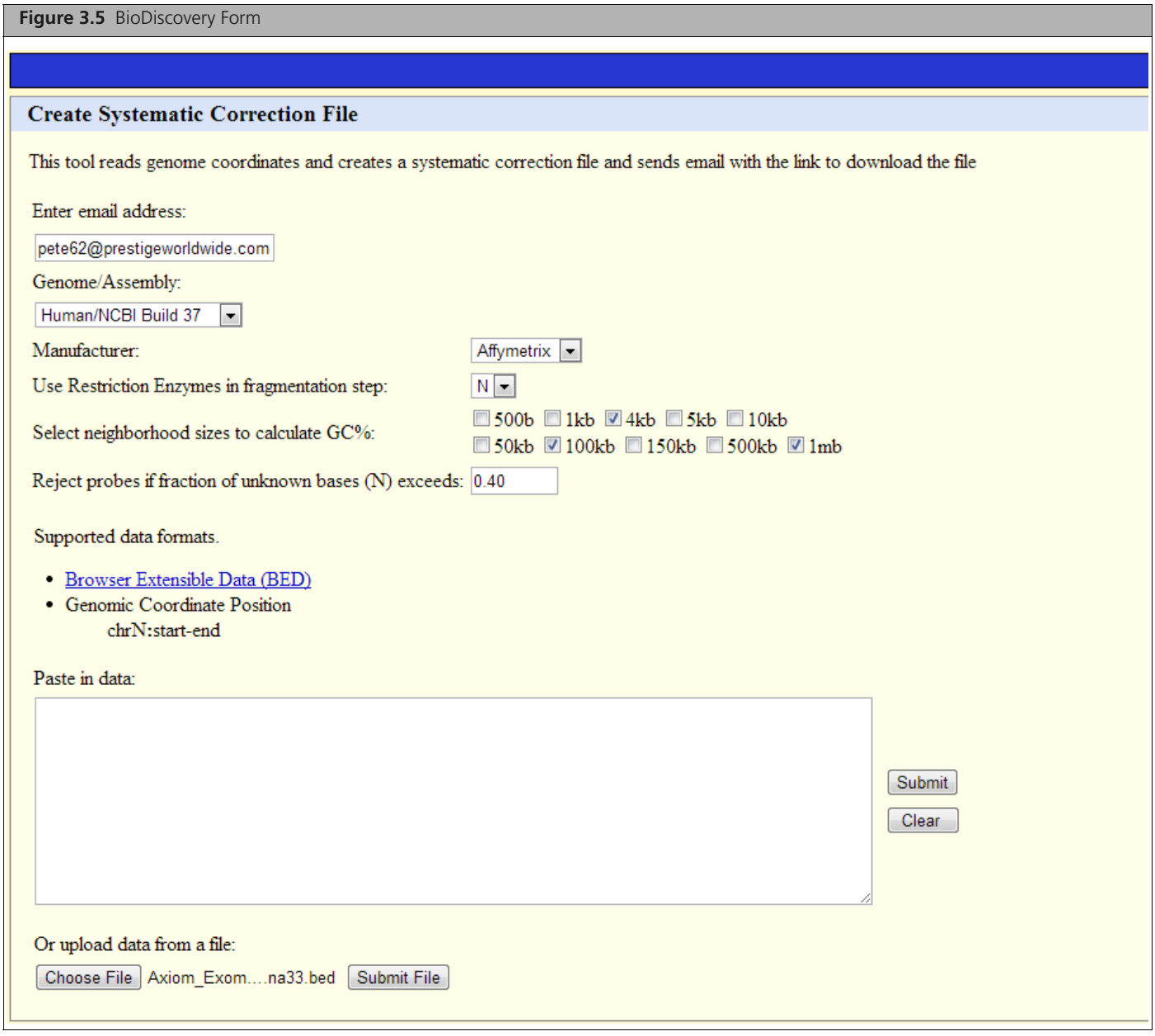
   BioDiscovery will respond with an email containing a hyperlink.

2. Click on the hyperlink provided by BioDiscovery.

   The following form appears. (Figure 3.5)

3. Complete the form, click **Browse** to upload your unzipped BED file, then click **Submit File**.

Figure 3.5 BioDiscovery Form

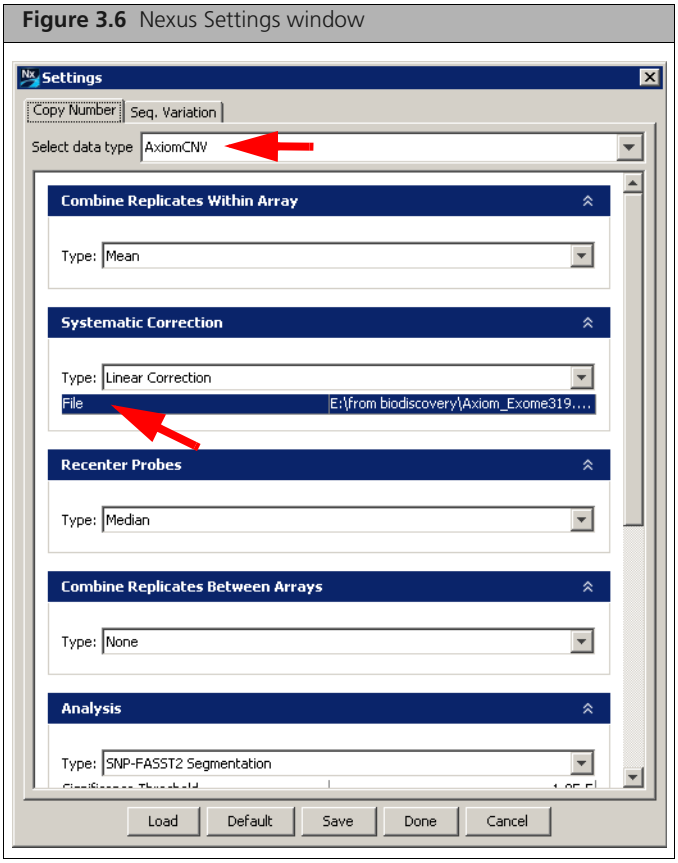


NOTE: **BioDiscovery will email you a second hyperlink to download the GC Correction file for use with their Nexus software.**

4. Click on the hyperlink to download/save the GC Corrected BED file. Make sure you save the file to a convenient location.
5. Open the *BioDiscovery Nexus* application as you normally would.
6. Click **Settings**.

The following window appears. (Figure 3.6)

Figure 3.6 Nexus Settings window



7. From the *Select data type* drop-down menu, click **AxiomCNV**. (Figure 3.6)
8. From the *Systematic Correction* drop-down menu, select your GC Correction Type.
9. Click the File banner (Figure 3.6), then select your GC Correction file.
10. Use the other applicable drop-down menu selections to complete the Settings form, then click **Done**.
11. Use the Nexus software as you normally would.

Calculations of log2 Ratios and B allele Frequencies

Affymetrix Axiom® Arrays, designed to detect genome-wide associations with SNPs and indels, can also detect copy number variations. The CNV detection method is based on computed log2 ratios and B allele frequencies (BAFs) at individual markers across the genome.

The log2 ratios and B allele frequencies are calculated as follows:

Computing log2 Ratio

Log2 ratios are computed at each marker site as the sum of the A and B allele intensities for the marker normalized by the median intensity of that marker in phenotypically normal individuals, assumed to represent the normal copy number state at that marker site.

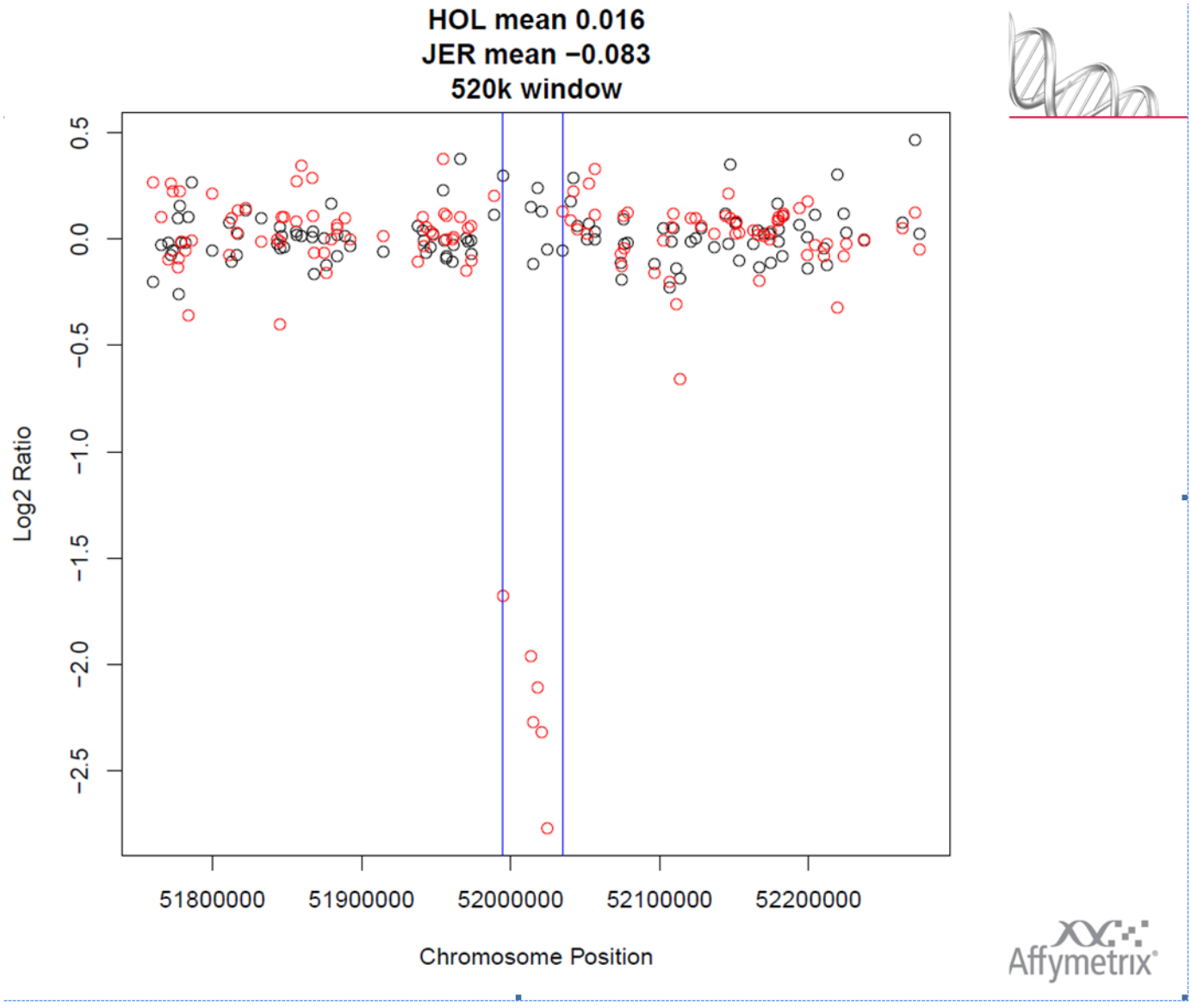
a) Total Intensity, $T = Intensity_A + Intensity_B$

b) Across a reference set of samples for each marker determine a Reference value R = median (T_1, T_2, … T_N). If most individuals are expected have normal copy number states, one approach is to create the reference based on the N (~96) individuals genotyped on the plate. For the X chromosome, median values are taken over only Female samples. For the Y chromosome, median values are taken over only Male samples.

c) For each marker for each individual sample, $\log_2 ratio = \log_2(T) - \log_2(R)$

The log2 ratio computed above may be ordered genomically and inspected visually for regions diverge from zero. Figure 1 (below) shows a region with significant deletion in the genome of Jersey cattle relative to Holstein. The intensities were produced by genotyping with the Axiom® Genome-Wide BOS 1 Bovine array by Rincon et al [1]. (Figure 1)

Figure 1 log2 ratio values for copy number loss from Rincon, et al.



B allele Frequencies (BAF)

The BAF at a marker is the ratio of the B allele intensity to the sum of the allele intensities. The raw BAFs are standardized based on mean BAFs for homozygous and heterozygous genotype calls for each marker and further scaled to be within [0,1]. (Figure 2)

Figure 2 BAF marker equations

$$\text{Raw BAF } (\lambda) = \frac{Intensity_B}{Intensity_A + Intensity_B}$$

Standardized BAF =

$$\frac{\lambda - \lambda_{AB}}{|\lambda_{AA} - \lambda_{AB}|} \text{ when } \lambda < \lambda_{AB}$$

$$\frac{\lambda - \lambda_{AB}}{|\lambda_{BB} - \lambda_{AB}|} \text{ when } \lambda \geq \lambda_{AB}$$

where

λ_{AA} = mean(λ | AA)

λ_{AB} = mean(λ | AB)

λ_{BB} = mean(λ | BB)

For markers in the non-PAR regions of chromosome X, mean values are taken over only Female samples. For markers in the PAR regions of chromosome X, mean values are taken over all samples. For markers on the Y Chromosome, mean values are taken over only Male samples.

When there are no samples for a specific genotype at a marker site, then mean values must be interpolated.

If only λ_{AA} is missing, then $\lambda_{AA} = \lambda_{AB} - (\lambda_{BB} - \lambda_{AB})$

If only λ_{AB} is missing, then $\lambda_{AB} = (\lambda_{AA} + \lambda_{BB})/2$

If only λ_{BB} is missing, then $\lambda_{BB} = \lambda_{AB} + (\lambda_{AB} - \lambda_{AA})$

If two mean values are missing, then the missing value is set to the median of the corresponding value of all probesets:

λ_{AA} = median(λ_{AA})

λ_{AB} = median(λ_{AB})

λ_{BB} = median(λ_{BB})

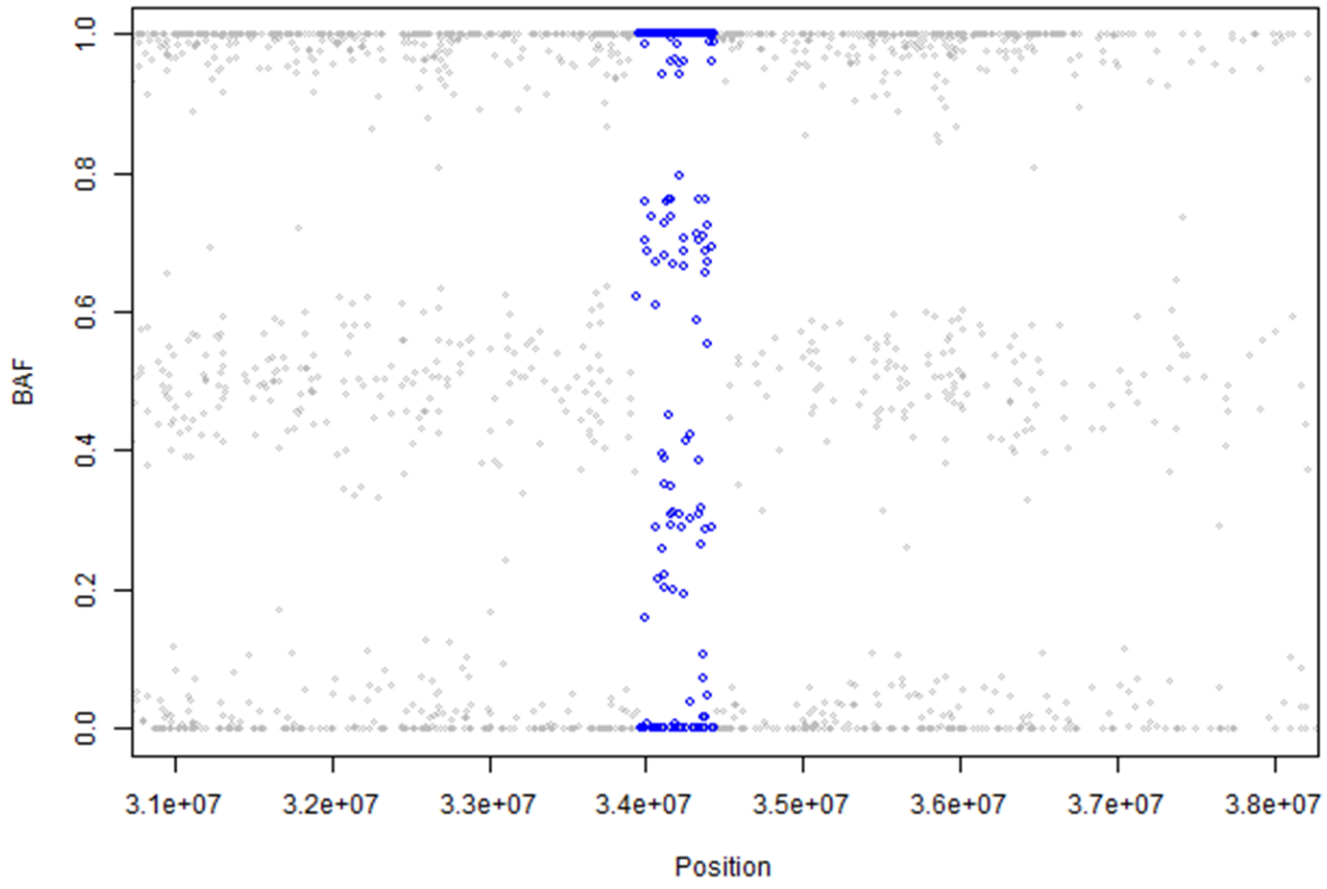
Scaled BAF = 0.5 * (Standardized BAF + 1.0)

The Scaled BAFs are truncated to be within [0,1]

The BAFs computed above may be ordered genomically and inspected visually for regions diverge from the expected pattern of 3 bands at 0, 0.5 and 1.0.

Figure 3 shows a region with a copy number 3 in the genome of a YRI HapMap sample. The intensities were produced by genotyping with Axiom® Genome Wide PanAFR array.

Figure 3 BAFs showing a band split characteristic of a copy number gain (shown in blue).



References

1.Rincon G, Weber K, Van Eenennaam A, et al. Hot topic: Performance of bovine high-density genotyping platforms in Holsteins and Jerseys. J Dairy Sci. 94:6116-6121 (2011).