



# SNP 6.0 Probe Level Access Tool User Manual

<b>INTRODUCTION .....</b>	<b>2</b>
ABOUT THIS MANUAL .....	2
TECHNICAL SUPPORT.....	2
<b>WORKING WITH SNP 6.0 PROBE LEVEL ACCESS TOOL .....</b>	<b>3</b>
INSTALLATION INSTRUCTIONS.....	3
UN-INSTALLATION INSTRUCTIONS.....	3
UPDATES AND GENERAL INFORMATION.....	4
BASIC WORKFLOW IN SNP 6.0 PROBE LEVEL ACCESS TOOL .....	4
TOOLBOX.....	5
STATUS MESSAGES PANE.....	5
<b>USER PROFILES .....</b>	<b>5</b>
CREATE/SELECT A USER PROFILE .....	6
DELETE A USER PROFILE.....	7
<b>LIBRARY AND SNP ANNOTATION FILES.....</b>	<b>8</b>
SETTING UP THE LIBRARY PATH .....	8
DOWNLOAD LIBRARY FILES .....	9
MANUALLY COPY LIBRARY FILES .....	10
DOWNLOAD ANNOTATION FILES.....	11
MANUALLY COPY ANNOTATION FILES .....	12
SETTING UP ANNOTATION FILES .....	12
<b>DATA EXTRACTION AND ANNOTATION MERGING .....</b>	<b>13</b>
DATA EXTRACTION.....	13
TWO WORKFLOWS: SNP AND CN PROBES OR CNV REGIONS ONLY .....	15
ANNOTATION MERGING (OPTIONAL).....	19
COMMAND-LINE SCRIPTS.....	20

## Introduction

The SNP 6.0 Probe Level Access Tool (SPLAT) extracts probe intensities for collections of Genome-Wide SNP Array 6.0 (CEL) files. The software tool writes output files containing normalized probe intensities and genomic position information for over 1.8 million genetic markers (both SNP and Copy Number probes) on the SNP Array 6.0 in standard tabular form. The tool also extracts summarized CN estimates for known Copy Number regions extracted from the Toronto DGV database (<http://projects.tcag.ca/variation/>). These regions currently number approximately 5000 and the annotation file containing these regions will be regularly updated on NetAffx.

The following sections provide information [About this Manual](#) and [Technical Support](#).

## About this Manual

This manual presents information about SPLAT in the following chapters and appendices:

[Working with the SNP 6.0 Probe Level Access Tool:](#) Describes how to install and configure the software including setting up user profiles and installing/downloading library and annotation files.

[User Profiles:](#) Describes how to create, select, and delete user profiles.

[Library and SNP Annotation Files:](#) Describes how to set up the library path and download library and SNP annotation files.

[Data Extraction and Annotation Merging:](#) Describes how to extract probe level data and merge results with annotations as well as run the command line algorithm.

## Technical Support

Support for SPLAT is handled through the Affymetrix Developer Network. Specifically, questions, problems, feature requests, etc. can be raised through [devnet@affymetrix.com](mailto:devnet@affymetrix.com). For more general information on support for DevNet tools see the Software Support Policy ([http://www.affymetrix.com/support/technical/software\\_support\\_policy.affx](http://www.affymetrix.com/support/technical/software_support_policy.affx)).

## Working with SNP 6.0 Probe Level Access Tool

SPLAT is a stand-alone application. It can be installed on computers that have GeneChip® Operating System (GCOS) software, Affymetrix GeneChip® Command Console™ (AGCC) software, Affymetrix Genotyping Console, or none of these software programs.

The minimum hardware recommendations are:

- Memory (RAM): 2 GB
- Hard drive: 40 GB\*
- Processor: 2.0 GHz Intel Pentium or higher

\*The larger file sizes associated with Genome-Wide SNP 6.0 Arrays data should be taken into account when calculating the necessary free space requirement.


The following sections describe [Installation Instructions](#), [Un-installation](#), [Updates and General Information](#), [Basic Workflow in SNP 6.0 Probe Level Access Tool](#), [Toolbox](#), and the [Status Messages Pane](#).

### Installation Instructions

To install the software:

1. Download the software from the DevNet website:  
<http://www.affymetrix.com/support/developer/tools/devnettools.affx>, follow the SPLAT link.
2. Unzip the downloaded software package.
3. Double-click on SplatSetup.exe to install the software.
4. Follow the directions provided by the installer.

To launch SPLAT go to Start/Programs/Affymetrix/SNP 6.0 Probe Level Access Tool.

 **Note:** The setup process installs the required Microsoft components, which includes the .NET 2.0 framework and VC++ runtime libraries.

### Un-installation Instructions

To un-install the software:

1. Go to Start/Settings/Control Panel.
2. Select Add or Remove Programs.
3. Migrate to SNP 6.0 Probe Level Access Tool in the programs list.
4. Click on the Change/Remove button and following the on screen directions.

## Updates and General Information

New information about this and other Affymetrix software will be made available to customers through the Update Button on the main tool bar. There are 3 different options: Updates Available, No New Updates, or Updates (Offline).

When updated information is available, click on the green Updates Available button on the main tool bar and a web browser will be launched indicating what new information is available.



When there are no new updates available, the following button will be displayed on the main tool bar. Clicking on the button will launch a web browser showing the current informational messages.

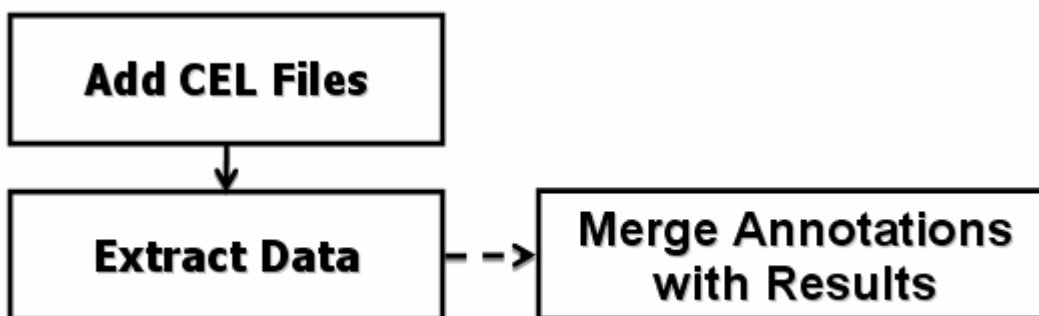


If the computer is offline, the software will be unable to determine if there are any updates available and the Updates button will indicate the offline status.



## Basic Workflow in SNP 6.0 Probe Level Access Tool

The figure below outlines the basic workflow in SPLAT.

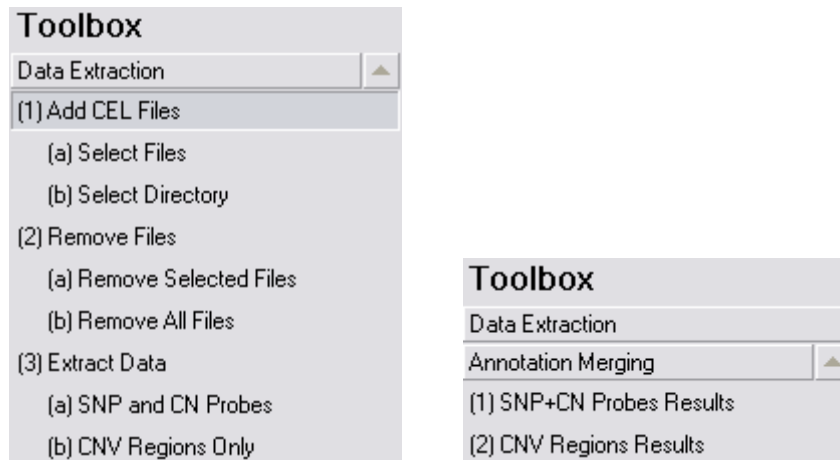


The workflow begins by [Adding CEL Files](#). SPLAT supports analysis of the Genome Wide SNP 6.0 array only. After extracting the probe level intensity information from the CEL files, the resulting data can be [Merged](#)

[with Annotations](#). SPLAT displays the analysis steps in a [Toolbox](#). The analysis steps are ordered by the Basic Workflow.

## Toolbox

Commands in SPLAT can be accessed using the Toolbox. There are two sections: Data Extraction and Annotation Merging.



The Data Extraction section contains the steps for running the algorithm and generating probe intensities. These include adding/removing CEL files to be analyzed and extracting data. The Annotation Merging section contains the steps for merging the results with annotation information.

## Status Messages Pane

The Status Messages pane displays all status and algorithm progress information.

```
Status Messages
5/23/2007 7:39:47 PM : Copy number data file loaded and indexed
5/23/2007 7:39:47 PM : Creating output file
5/23/2007 7:39:47 PM : Copy number output file created
5/23/2007 7:39:47 PM : Merge completed
5/23/2007 7:39:47 PM : The data has been merged to E:\Software\Plan 10\SPLAT_Testing\CNVRegions\CNVRegionsOnly.cnv.summary+annot.txt
```

## User Profiles

A User Profile stores a user's preferences for custom analysis settings, table and graph viewing options, and other application settings. Security by profiles is not provided by the application, it is simply a means of storing application parameters. The following sections describe how to [Create/Select a User Profile](#) or [Delete a User Profile](#).

## Create/Select a User Profile

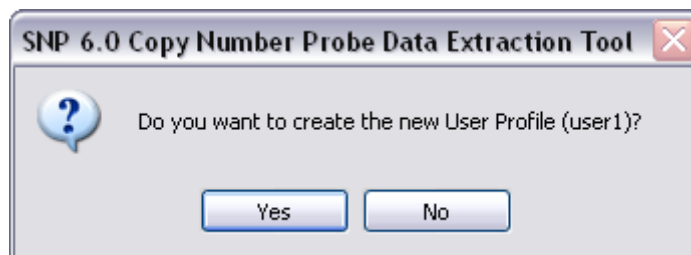
A user profile holds algorithm parameters, report thresholds and other application parameters selected by a user. Security is not provided by the application.

To create a new User Profile:

1. Start Genotyping Console by double-clicking on its shortcut on the Desktop, or from the Start Menu, in Programs/Affymetrix/Genotyping Console. Genotyping Console opens with the User Profile window displayed. Type in a name for your profile and click OK.



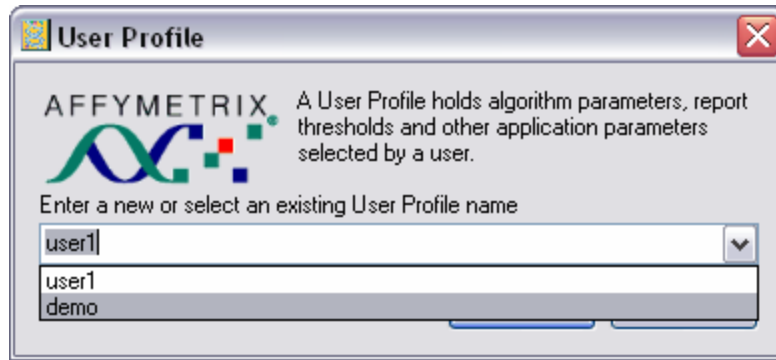
2. The software will prompt you to create the new profile.



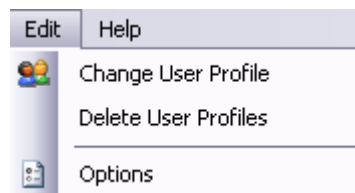
After setting up a User Profile, the software will prompt you to select a Library File path. See [Setting Up the Library Path](#).

To select an existing Profile:

1. To select a previously generated profile, use the drop-down menu on the User Profile window.



1. To change profiles, select Edit/Change User Profile



2. Enter a new profile name or select a previously generated profile from the drop-down box (see above).

## Delete a User Profile

The list of previously created profiles is found in the drop-down menu on the Profile Information window. To remove profiles no longer needed:

1. Select Edit/Delete User Profile.
2. Select the User Profile to be deleted and select OK.



3. The selected User Profile, and all parameter files associated with the profile, will be removed. To add a new User Profile, see [Create/Select a User Profile](#).

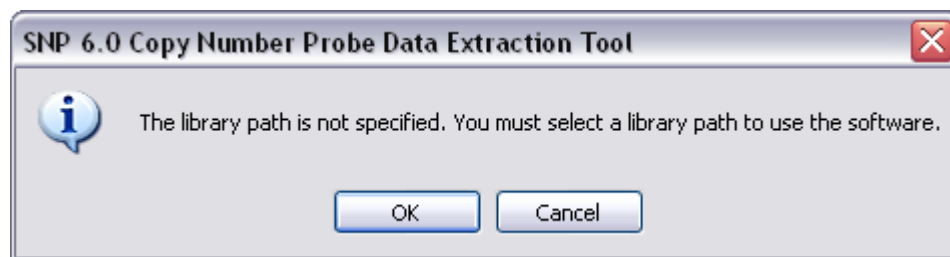
## Library and SNP Annotation Files


SPLAT requires information stored in library and annotation files to analyze the CEL files generated by GCOS or Affymetrix GeneChip® Command Console™ (AGCC) software. These files are available from NetAffx and can be downloaded within SPLAT. SPLAT downloads only those library files it requires from NetAffx for analysis, but these are not registered with GCOS or Command Console and are not sufficient to scan arrays.

The following sections describe how to [Set the Library Path](#), [Download Library Files](#), [Manually Copy Library Files](#), [Download Annotation Files](#), [Manually Copy Annotation Files](#), and [Setting up Annotation Files](#).

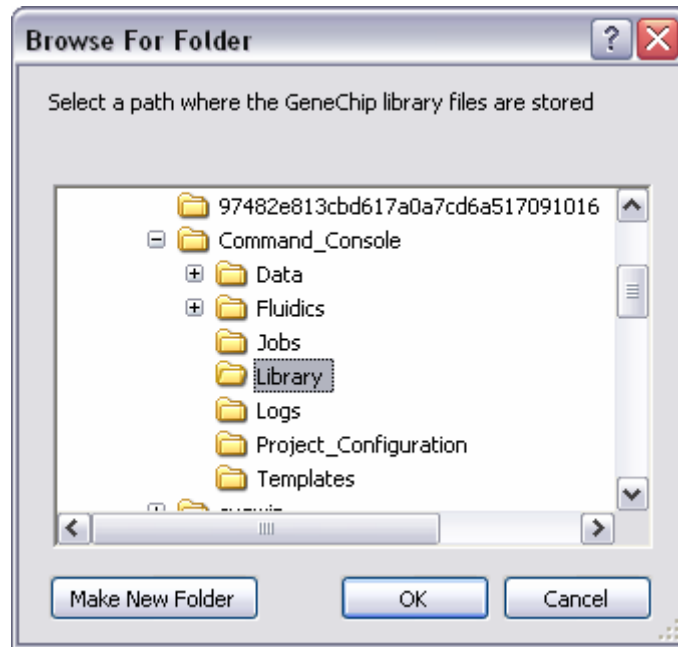
### Setting up the Library Path

SPLAT will prompt you to select a location for the library path.



If you wish to change an existing Library Path, either click on the Options shortcut  on the main toolbar, or select Edit/Options.

Browse to the folder which contains the library files or create a new folder for your library files. Make sure all library files are copied to this folder for use in SPLAT.



**Note:** You can select any location for the library files folder. However, once you direct the software to the folder location; do not place any library files in a subfolder. SPLAT cannot find library files in a subfolder!

**Note:** If the Affymetrix GeneChip® Operating System software (GCOS) is installed on your system, Affymetrix recommends that you do NOT select the GCOS library file directory as the library file directory for Genotyping Console, to avoid confusion.


**Note:** The selected library path is indicated in the bottom left corner of the application.

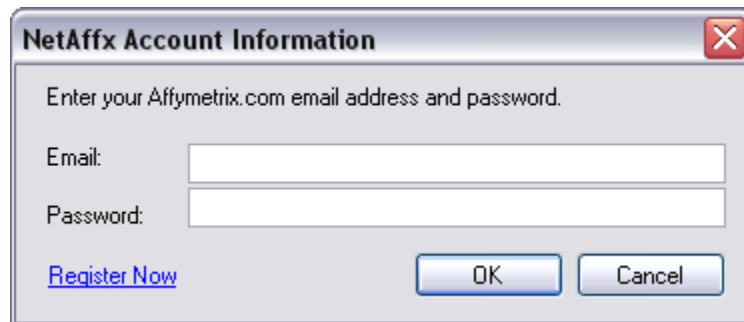
Library path: C:\Program Files\Affymetrix\Library

**Note:** GCOS users must use DTT v1.1, using the Flat File option, to transfer files to be analyzed by Genotyping Console software from the GCOS database to an independent folder. More detailed instructions can be found at [www.affymetrix.com](http://www.affymetrix.com); then go to Support/Technical/Tutorial/GCOS.

## Download Library Files

Library files can be downloaded from NetAffx™:

1. Click on the Download Library Files shortcut  on the main toolbar, or select File/Download Library Files. A window opens requesting your account information for NetAffx.

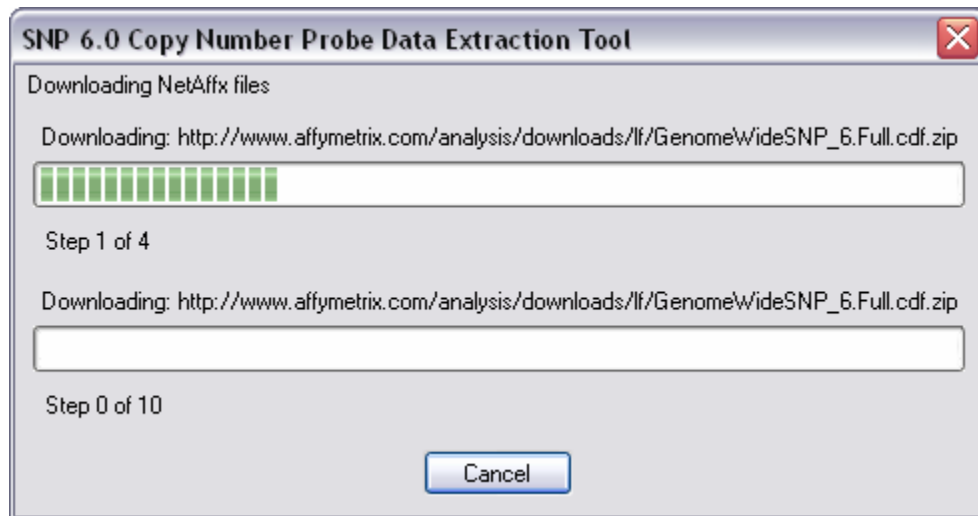


The dialog box is titled "NetAffx Account Information" and contains the following elements:

- Text: "Enter your Affymetrix.com email address and password."
- Label: "Email:" followed by a text input field.
- Label: "Password:" followed by a text input field.
- Link: "[Register Now](#)"
- Buttons: "OK" and "Cancel"


If you do not have a NetAffx account, click on the [Register Now](#) button which will launch [www.affymetrix.com](http://www.affymetrix.com) and follow the instructions to set up an account.

2. Enter your registered email address and password.
3. A window will display the progress of the download.



The window is titled "SNP 6.0 Copy Number Probe Data Extraction Tool" and displays the following information:

- Text: "Downloading NetAffx files"
- Text: "Downloading: [http://www.affymetrix.com/analysis/downloads/lf/GenomeWideSNP\\_6.Full.cdf.zip](http://www.affymetrix.com/analysis/downloads/lf/GenomeWideSNP_6.Full.cdf.zip)"
- Progress bar: A green progress bar showing approximately 25% completion.
- Text: "Step 1 of 4"
- Text: "Downloading: [http://www.affymetrix.com/analysis/downloads/lf/GenomeWideSNP\\_6.Full.cdf.zip](http://www.affymetrix.com/analysis/downloads/lf/GenomeWideSNP_6.Full.cdf.zip)"
- Text: "Step 0 of 10"
- Button: "Cancel"

 **Note:** Downloading of the library files may take several minutes or more based on connection speed, as these files are large.


## Manually Copy Library Files

For computers that are not connected to the internet and therefore cannot take advantage of the library file download option, it is possible to manually copy the necessary files.

1. Create a folder on the computer for the library files.
2. Copy the necessary files to the library file folder.
3. For Genome-Wide SNP 6.0 Arrays, the following files are required:
  - a. GenomeWideSNP\_6.cdf
  - b. GenomeWideSNP\_6.na22.dgv-cnvMay07.mps
4. Do not create subdirectories within the library file folder. SPLAT does not look at subdirectories.

## Download Annotation Files

Annotation files can be downloaded from NetAffx™:

1. Click on the Download SNP Annotation Files shortcut  on the main toolbar, or select File/Download SNP Annotation Files. A window opens requesting your account information for NetAffx.

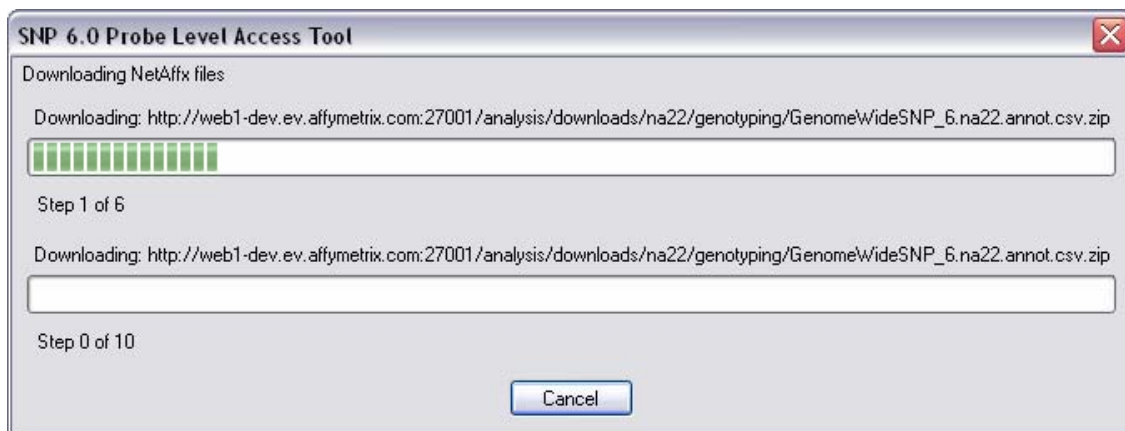


The dialog box is titled "NetAffx Account Information" and contains the following elements:

- Text: "Enter your Affymetrix.com email address and password."
- Label: "Email:" followed by a text input field.
- Label: "Password:" followed by a text input field.
- Link: "[Register Now](#)"
- Buttons: "OK" and "Cancel".


If you do not have a NetAffx account, click on the [Register Now](#) button which will launch [www.affymetrix.com](http://www.affymetrix.com) and follow the instructions to set up an account.

2. Enter your registered email address and password.
3. A window results which will monitor the progress of the download.



The dialog box is titled "SNP 6.0 Probe Level Access Tool" and contains the following elements:

- Text: "Downloading NetAffx files"
- Text: "Downloading: [http://web1-dev.ev.affymetrix.com:27001/analysis/downloads/na22/genotyping/GenomeWideSNP\\_6.na22.annot.csv.zip](http://web1-dev.ev.affymetrix.com:27001/analysis/downloads/na22/genotyping/GenomeWideSNP_6.na22.annot.csv.zip)"
- Progress bar: A green progress bar showing approximately 10% completion.
- Text: "Step 1 of 6"
- Text: "Downloading: [http://web1-dev.ev.affymetrix.com:27001/analysis/downloads/na22/genotyping/GenomeWideSNP\\_6.na22.annot.csv.zip](http://web1-dev.ev.affymetrix.com:27001/analysis/downloads/na22/genotyping/GenomeWideSNP_6.na22.annot.csv.zip)"
- Text: "Step 0 of 10"
- Button: "Cancel"

 **Note:** Downloading of the annotation files may take several minutes based on connection speed. These files are extremely large, so please be patient.

## Manually Copy Annotation Files

For computers that are not connected to the internet and therefore cannot take advantage of the annotation file download option, it is possible to manually copy the necessary files.

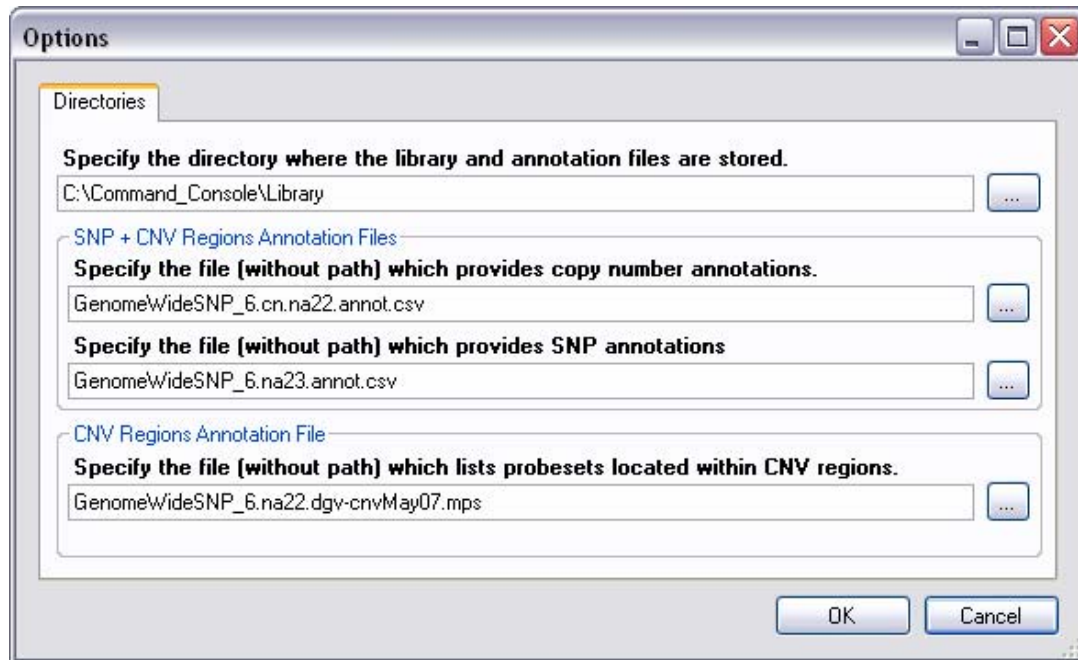
1. Create a folder on the computer to hold the library and annotations files for SPLAT.
2. Copy the \*.csv file(s) to the library file folder. See [Setting up Annotation Files](#) for more information on the csv files.
3. Do not create subdirectories within the library file folder. SPLAT does not look at subdirectories.

## Setting up Annotation Files

There are several different annotation files for the Genome Wide SNP 6.0 array:

- GenomeWideSNP\_6.cn.na22.annot.csv – This file contains the copy number probe annotations.
- GenomeWideSNP\_6.na23.annot.csv – This file contains the SNP probe annotations.
- GenomeWideSNP\_6.na22.dgv-cnvMay07.mps - This file lists the probesets located within CNV regions from the Toronto DGV database (<http://projects.tcag.ca/variation/>).

Using the Options shortcut  on the main toolbar or by select Edit/Options, the following dialog opens:



Specify the location of the annotation files listed above.

**Note:** These annotation files are required in order to merge the results with annotations.

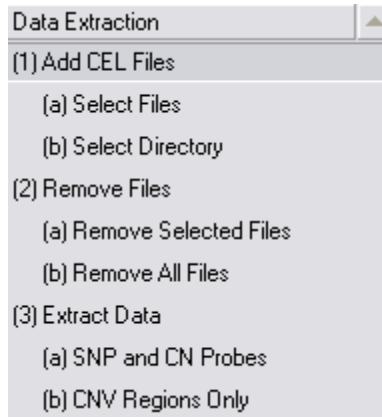
## Data Extraction and Annotation Merging

The following sections describe [Data Extraction](#), [Two Workflows: SNP and CN Probes or CNV Regions Only](#), [Annotation Merging](#), and [Command-line Scripts](#).

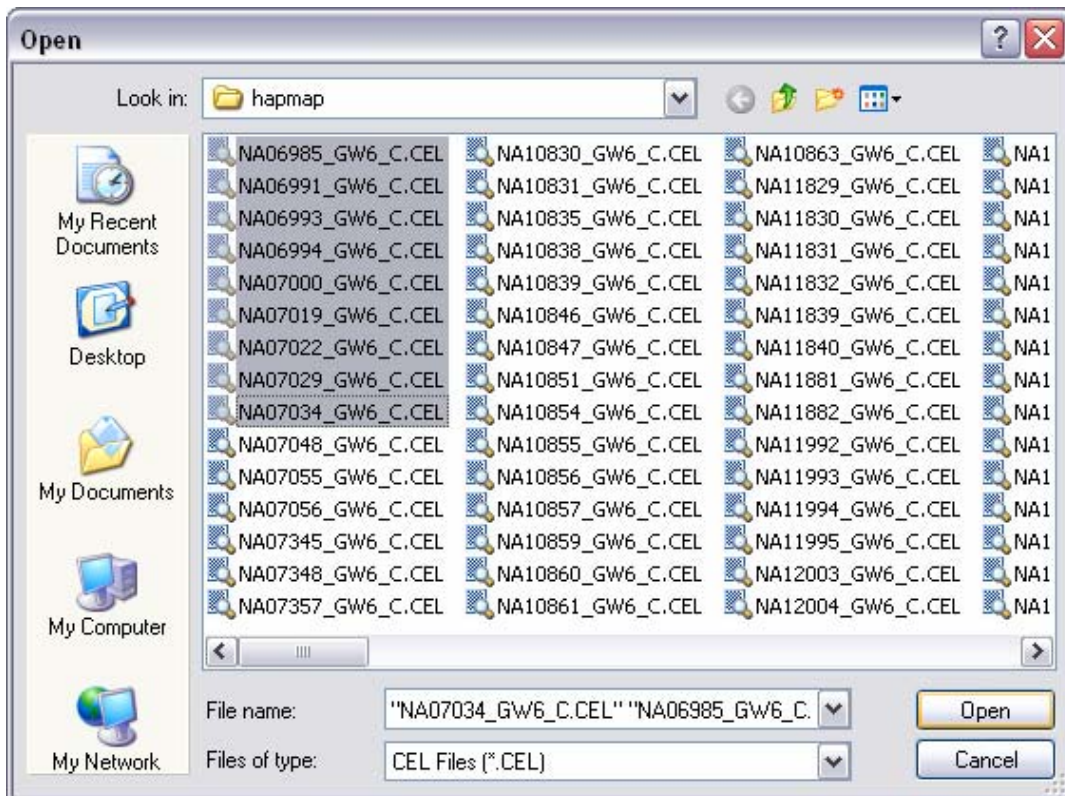
### Data Extraction

Follow the steps below to Add and Remove CEL files.

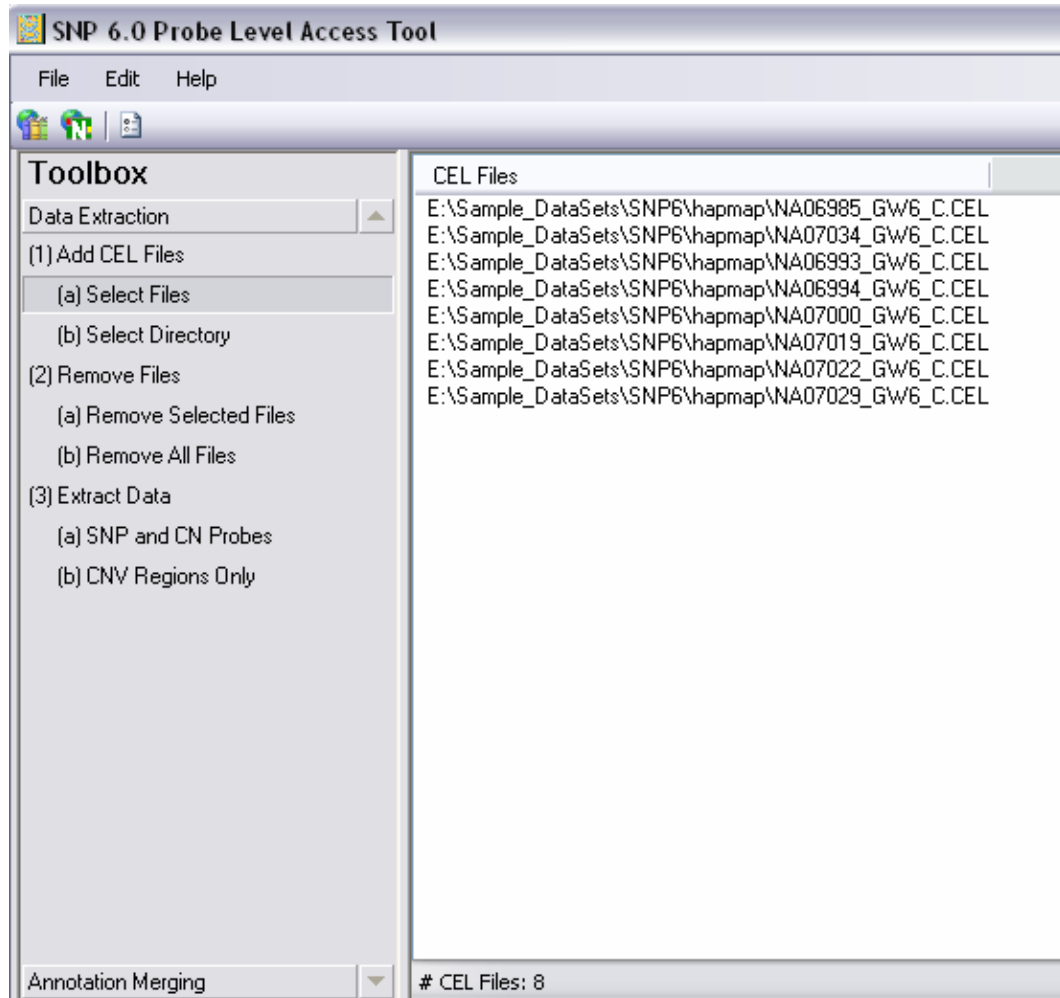
1. After launching the application and selecting/creating a user profile, use the Workflow Toolbox to guide you through the analysis steps.



2. To add CEL files to SPLAT, either click on the Select files or Select directory option. In the Select files option, migrate to the location of your CEL files, choose the appropriate files, and select OK. In the Select directory option, migrate to the directory which contains the CEL files. By selecting the directory, all CEL files within the directory will be added.



3. The selected files will be displayed in the main window.



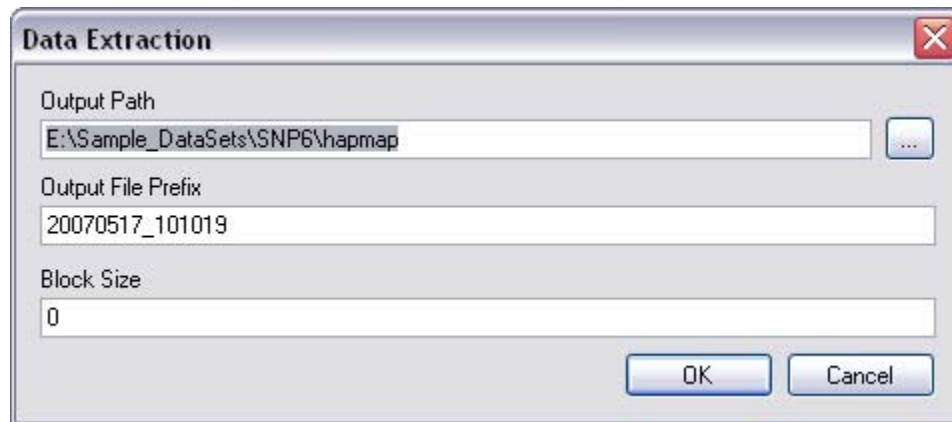
4. To remove a file or set of files, highlight the files in the main window and select Remove Selected Files from the Toolbox. To remove all files select Remove All Files from the Toolbox.

### Two Workflows: SNP and CN Probes or CNV Regions Only


To extract the probe level intensity values, there are two available workflows: SNP and CN Probes and CNV Regions Only. The first workflow (SNP and CN Probes) will analyze all SNP and CN probes on the array and return a result for all 1.8 million genetic markers. The second workflow (CNV Regions Only) will analyze only those CN probes which are located within CNV regions as defined by the Toronto DGV database (<http://projects.tcag.ca/variation/>). The known copy number regions are specified in a meta-probeset file available from NetAffx (GenomeWideSNP\_6.na22.dgv-cnvMay07.mps).

Follow the steps below for both workflows in SPLAT:

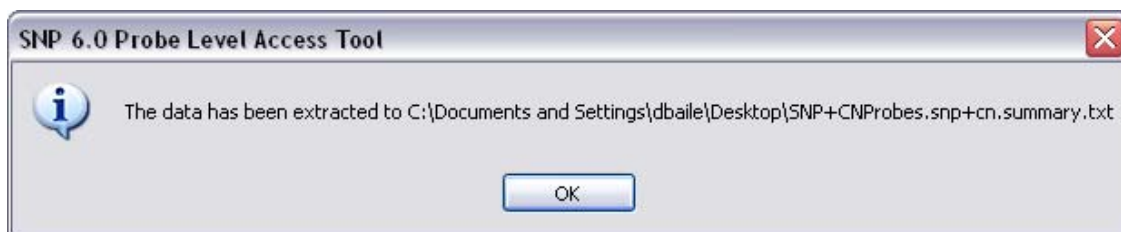
1. Select either SNP and CNV Regions or CNV Regions Only. The following dialog appears.



2. In the Data Extraction dialog:
  - a. Select a location for the analysis output by either selecting a folder or manually enter the full path to the folder
  - b. Select a prefix for the output file name. By default the output file prefix name is generated automatically using the time date stamp. This name is assigned to output file with a \*snp+cn.summary.txt or \*.cnv.summary.txt extension.
  - c. Enter a Block Size. Block Size refers to the number of probesets to process at once. This parameter is useful when memory is limited. If set to 0 (default), SPLAT attempts to determine the available RAM and set it appropriately. Refer to the apt-probeset-summarize manual for more information on adjusting this parameter.

 **Note: The default batch name includes the date and time; therefore, it is unique for each run.**

3. A dialog will appear once the analysis is complete.



4. The output text file names will be based upon the user defined Output File Prefix (see step #1 above) and the workflow selected (either SNP and CN Probes or CNV Regions only). For example:
  - 20070517\_101019.snp+cn.summary.txt – For the SNP and CN Probe workflow
  - 20070517\_101019.cnv.summary.txt – For the CNV Regions only workflow

The output file (<userdefined>.snp+cn.summary.txt or <userdefined>.cnv.summary.txt) contains algorithm parameter options in the header and log<sub>2</sub> transformed signal estimates arranged in a table with the Probeset ID in rows and CEL file names in columns. Note that the output file will also contain other control probesets (including some control SNPs).

```

#%guid=0000027696-1179953334-0000001855-0000025694-0000023497
#%affymetrix-algorithm-param-apt-engine=ProbesetSummarizeEngine
#%affymetrix-algorithm-param-apt-prog=SNP 6.0 Probe Level Access Tool
#%affymetrix-algorithm-param-apt-cmd=
#%affymetrix-algorithm-param-apt-exec-guid=0000027696-1179953005-0000008951-0000016241
#%affymetrix-algorithm-param-apt-analysis-guid=0000027696-1179953190-0000028442-0000001
#%affymetrix-algorithm-param-apt-time-str=Wed May 23 13:43:25 2007
#%affymetrix-algorithm-param-apt-version=1.0.2699.16810
#%affymetrix-algorithm-param-apt-cvs-id=SNP 6.0 Probe Level Access Tool-1.0.2699.16810
#%affymetrix-algorithm-param-apt-opt-do-ps-names=false
#%affymetrix-algorithm-param-apt-opt-precision=1
#%affymetrix-algorithm-param-apt-opt-cdf-mode=true
#%affymetrix-algorithm-param-apt-opt-pgf-mode=false
#%affymetrix-algorithm-param-apt-opt-force=false
#%affymetrix-algorithm-param-apt-opt-block-size=0
#%affymetrix-algorithm-param-apt-opt-cdf-file=C:\Command_Console\Library\GenomeWideSNP
#%affymetrix-algorithm-param-apt-opt-spf-file=
#%affymetrix-algorithm-param-apt-opt-pgf-file=
#%affymetrix-algorithm-param-apt-opt-clf-file=
#%affymetrix-algorithm-param-apt-opt-bgp-file=
#%affymetrix-algorithm-param-apt-opt-ps-list-file=
#%affymetrix-algorithm-param-apt-opt-meta-ps-file=
#%affymetrix-algorithm-param-apt-opt-precomp-feature-eff=
#%affymetrix-algorithm-param-apt-opt-out-dir=C:\Documents and Settings\dbaile\Desktop
#%affymetrix-algorithm-param-apt-opt-cc-chp-out-dir=
#%affymetrix-algorithm-param-apt-opt-xda-chp-out-dir=
#%affymetrix-algorithm-param-apt-opt-sketch-in-file=
#%affymetrix-algorithm-param-apt-opt-chip-type=GenomeWideSNP_6
#%affymetrix-algorithm-param-apt-opt-probe-count=6892960
#%affymetrix-algorithm-param-apt-opt-cel-1=E:\Sample_DataSets\SNP6\hapmap\NA06985_GW6_
#%affymetrix-algorithm-param-apt-opt-pgf-kill-list=
#%affymetrix-algorithm-param-apt-opt-qc-groups-file=
#%affymetrix-algorithm-param-apt-opt-analysis-name=SNP+CNProbes.snp+cn
#%affymetrix-algorithm-param-apt-opt-analysis-spec=quant-norm.sketch=50000.bioc=false.
#%affymetrix-algorithm-param-quantification-name=med-polish
#%affymetrix-algorithm-param-quantification-version=1.0
#%affymetrix-algorithm-param-quantification-scale=log2
#%affymetrix-algorithm-param-quantification-type=signal
#####
probeset_id      NA06985_GW6_C.CEL
AFFX-5Q-123      9.1
AFFX-5Q-456      8.5
AFFX-5Q-789      10.1
AFFX-5Q-ABC      10.4
AFR_A02_SB       7.1
AFR_A04_SB       7.5
AFR_A06_SB       7.4
AFR_A08_SB       7.0
AFR_A10_SB       7.7
AFR_A12_SB       7.6
AFR_A14_SB       7.7
AFR_A16_SB       7.5
AFR_A18_SB       7.1
AFR_A20_SB       7.2
AFR_A22_SB       8.4
AFR_A24_SB       8.5
AFR_A26_SB       8.8
AFR_A28_SB       7.4
AFR_A30_SB       9.2
AFR_A32_SB       7.7
AFR_A34_SB       10.2
AFR_A36_SB       10.4

```

In addition to the <userdefined>.snp+cn.summary.txt file, a <userdefined>.snp+cn.report.txt will also be generated. This file contains algorithm metrics typically used in expression studies. These metrics have not been validated in these workflow and may not be applicable. For more information on the contents of this file see the apt-probeset-summarize manual (<http://www.affymetrix.com/support/developer/powertools/index.affx>).

In the CNV Regions Only workflow, an additional file is generated (<userdefined>.cnv.pca-select.report.txt). This file contains the list of probesets that were selected and used by pca-selection. For more information on this file and algorithm option, see the apt-probeset-

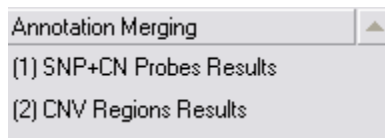
summarize manual

(<http://www.affymetrix.com/support/developer/powertools/index.affx>).

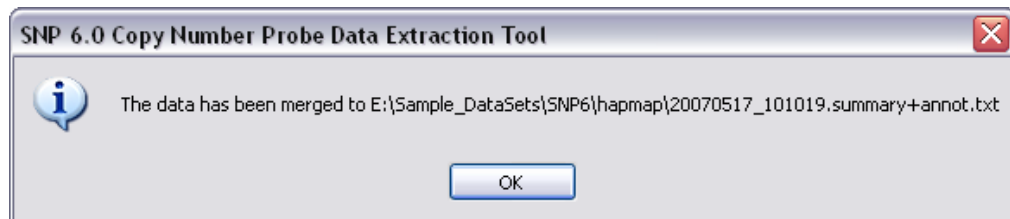
## Annotation Merging (Optional)

Follow the steps below to merge annotations with the extracted signal data.

1. After completing the Data Extraction step, select the Annotation Merging section of the Toolbox.



2. Select the Merge Annotations with Results option.
3. Select the location of the data file from the extraction step.
4. A dialog will appear once the analysis is complete.



The output file (<userdefined>.snp+cn.summary+annot.txt or <userdefined>.cnv.summary+annot.txt) contains algorithm parameter options in the header and log<sub>2</sub> transformed signal estimates arranged in a table with the Probeset ID in rows, CEL file names in columns, and Chromosome, start, stop position columns. Note that the output file will also contain other control probesets (including some control SNPs).

```

#####
#%guid=0000027696-1179954075-0000029085-0000027812-0000006441
#%affymetrix-algorithm-param-apt-engine=ProbesetSummarizeEngine
#%affymetrix-algorithm-param-apt-prog=SNP 6.0 Probe Level Access Tool
#%affymetrix-algorithm-param-apt-cmd=
#%affymetrix-algorithm-param-apt-exec-guid=0000027696-1179953811-0000007545-0000017241
#%affymetrix-algorithm-param-apt-analysis-guid=0000027696-1179953928-0000014854-000001
#%affymetrix-algorithm-param-apt-time-str=Wed May 23 13:56:51 2007
#%affymetrix-algorithm-param-apt-version=1.0.2699.16810
#%affymetrix-algorithm-param-apt-cvs-id=SNP 6.0 Probe Level Access Tool-1.0.2699.16810
#%affymetrix-algorithm-param-apt-opt-do-ps-names=false
#%affymetrix-algorithm-param-apt-opt-precision=1
#%affymetrix-algorithm-param-apt-opt-cdf-mode=true
#%affymetrix-algorithm-param-apt-opt-pgf-mode=false
#%affymetrix-algorithm-param-apt-opt-force=false
#%affymetrix-algorithm-param-apt-opt-block-size=0
#%affymetrix-algorithm-param-apt-opt-cdf-file=C:\Command_Console\Library\GenomeWideSNP
#%affymetrix-algorithm-param-apt-opt-spf-file=
#%affymetrix-algorithm-param-apt-opt-pgf-file=
#%affymetrix-algorithm-param-apt-opt-clf-file=
#%affymetrix-algorithm-param-apt-opt-bgp-file=
#%affymetrix-algorithm-param-apt-opt-ps-list-file=
#%affymetrix-algorithm-param-apt-opt-meta-ps-file=C:\Command_Console\Library\GenomeWid
#%affymetrix-algorithm-param-apt-opt-precomp-feature-eff=
#%affymetrix-algorithm-param-apt-opt-out-dir=C:\Documents and Settings\dbaile\Desktop
#%affymetrix-algorithm-param-apt-opt-cc-chp-out-dir=
#%affymetrix-algorithm-param-apt-opt-xda-chp-out-dir=
#%affymetrix-algorithm-param-apt-opt-sketch-in-file=
#%affymetrix-algorithm-param-apt-opt-chip-type=GenomeWideSNP_6
#%affymetrix-algorithm-param-apt-opt-probe-count=6892960
#%affymetrix-algorithm-param-apt-opt-cel-1=E:\Sample_DataSets\SNP6\hapmap\NA06985_GW6_
#%affymetrix-algorithm-param-apt-opt-pgf-kill-list=
#%affymetrix-algorithm-param-apt-opt-qc-groups-file=
#%affymetrix-algorithm-param-apt-opt-analysis-name=CNVRegionsOnly.cnv
#%affymetrix-algorithm-param-apt-opt-analysis-spec=rma-bg.quant-norm.sketch=50000.bioc
#%affymetrix-algorithm-param-quantification-name=med-polish
#%affymetrix-algorithm-param-quantification-version=1.0
#%affymetrix-algorithm-param-quantification-scale=log2
#%affymetrix-algorithm-param-quantification-type=signal
probeset_id      NA06985_GW6_C.CEL      Chromosome      Start      Stop
Variation_3274   10.1      1      377      988150
Variation_2041   10.1      1      17845     167209
Variation_0675   10.1      1      17845     167209
Variation_0991   10.7      1      45257     76166
Variation_1534   11.1      1      55871     68941
Variation_2294   10.1      1      77822     1008567
Variation_4189   7.7       1      307737    415038
Variation_4190   8.8       1      529118    646613
Variation_4191   10.5      1      569355    733174
Variation_0001   9.7       1      1083805   1283805
Variation_5452   9.3       1      1142956   1147823
Variation_3275   9.5       1      1189183   1309735
Variation_3276   9.9       1      1314392   2014198
Variation_2295   9.9       1      1452629   1683586
Variation_2042   9.6       1      1483131   1637358
Variation_0002   9.8       1      1609516   1720123
Variation_4192   10.1      1      2062347   2242269
Variation_4193   10.1      1      2145626   2314237
Variation_4194   9.9       1      3117044   3216520
Variation_4195   10.0      1      3142453   3323676
Variation_4196   9.9       1      3207806   3373233
Variation_3278   9.8       1      3278798   3538713
#####

```

## Command-line Scripts

The following section describes how to run the [SNP and CN Probe workflow](#) using Affymetrix Power Tools (APT):

The default analysis runs quantile normalization (quant-norm) across all chips. For the CN Probes no further summarization is done after the normalization. For the SNP probes, several additional steps are performed. The option “pm-sum” adds the sum of each pair of probes (SNP probes are spatially arranged in pairs where each A-allele probe is paired with a B-allele probe). The option “med-polish” is performed on the log sums to produce a summary value per SNP. The “expr.genotype=true” modifier ensures that genotyping probesets will be

included in the analysis. For SNPs, there are 2 alleles that each will have identical values as they both report on the sum of A and B probes, the modifier "allele.a=true" will output only the first allele id to avoid writing redundant rows in the output.

The default Unix command is:

```
apt-probeset-summarize
--cdf-file
    lib_directory_name/GenomeWideSNP_6.cdf \
--analysis
    quant-norm.sketch=50000,pm-sum,med-
polish,expr.genotype=true.allele-a=true \
--out-dir
    output_directory_name \
celfile_directory
    /*.CEL
```

🟡 Note: The output file will also contain other control probesets (including some control SNPs) that can be removed (in UNIX) using the following commands:

```
grep SNP quant-norm.pm-sum.med-polish.expr.summary.txt | grep -v
AFFX > snp-and-cn-only.txt
grep CN quant-norm.pm-sum.med-polish.expr.summary.txt >> snp-and-
cn-only.txt
```

🟡 Note: The library\_directory\_name and celfile\_directory are placeholders. Replace these with the actual directories names in your system.

The following section describes how to run the **CNV Regions Only workflow** using Affymetrix Power Tools (APT):

The default analysis runs quantile normalization across all chips and med-polish on probes selected as most informative, with no background correction. The known copy number regions are specified in a meta-probeset file available from NetAffx (GenomeWideSNP\_6.na22.dgv-cnvMay07.mps). The meta-probeset file contains only CN probes located with CNV regions as defined by the Toronto DGV database. "pca-select" is a modification to the med-polish algorithm that uses PCA to pick probes that are near the first principal component to use in the summarization. This is useful for allowing for the fact that the regions in the public database have noisy boundaries and are including some probes that are not part of the CNV. The pca-select will look for a subset of probes that behave consistently and that best pick up the variance across the samples.

The default Unix command is:

```
apt-probeset-summarize
--cdf-file
    lib_directory_name/GenomeWideSNP_6.cdf \
--analysis
    quant-norm.sketch=50000,pm-only,med-
polish,pca-select
```

```
--meta-probesets  
    lib_directory_name/GenomeWideSNP_6.na22.dgv-  
cnMay07.mps \  
--out-dir      output_directory_name \  
celfile_directory /*.CEL
```

📌 Note: The library\_directory\_name and celfile\_directory are placeholders. Replace these with the actual directories names in your system.

For more information on the algorithm, see Affymetrix Power Tools (<http://www.affymetrix.com/support/developer/powertools/index.affx>).