AFFYMETRIX® GENOTYPING CONSOLE WORKFLOW

Introduction

Affymetrix Genotyping Console is a new genotyping analysis software package designed to streamline whole-genome genotyping analysis and quality control for collections of Genome-Wide SNP Array 5.0 and 6.0 CEL files.

Genotyping Console implements a novel genotype-calling algorithm called Birdseed, an evolution of the RLMM genotype-calling algorithm (Rabbee and Speed, 2006). It performs a multiple-chip analysis to estimate a signal intensity for each allele of each SNP, fitting probe-specific effects to increase precision (like the BRLMM-P algorithm developed for the Genome-Wide Human SNP Array 5.0). It then makes genotype calls by fitting a Gaussian mixture model in the two-dimensional A-signal vs. B-signal space, using SNP-specific models to increase performance. In addition to the Birdseed algorithm, Genotyping Console supports the BRLMM-P algorithm for the SNP Array 5.0.

Genotyping Console displays metrics and annotation information in standard tabular form, to evaluate the data quality for a given array or SNP. Scatter plots and line graphs give you the power to quickly identify features of interest in your data set. Numerous data and visualization export features make it easy to share results with other applications and users.

Genotyping Console Workflow Overview

Create Workspace → Create Dataset → Add Data → Perform Genotyping → Export Genotypes

Add Data → Review CEL Data Quality

Perform Genotyping → Review Genotype Data Quality

Create Workspace

A workspace contains data sets, data files and SNP lists available within a single session. The intent is to contain related data (by PI, study, etc.). The workspace is stored in an XML file with .gtc_workspace extension.

Create Data Sets

A data set is a collection of sample attribute files (XML/ARR), intensity files (CEL) and/or genotyping files (CHP). One or more data sets can be added to a workspace. All data in a data set must be of a given array type.

Add Data

Data files (ARR/XML/CEL/CHP) may be added at any time to a data set. Options include automatic search and addition of sample or intensity files (e.g., select ARR and CEL files are automatically loaded) and automatic QC. The data are shown in a table. Line graphs may be generated of the intensity QC results.

Perform Genotyping

Genotyping is the process of specifying a list of CEL files and invoking the BRLMM-P (Genome-Wide SNP Array 5.0) or Birdseed (Genome-Wide SNP Array 6.0) algorithm. The CHP summary statistics (e.g., computed gender) are displayed in a table after the analysis completes. Line graphs may be generated for the CHP summary statistics.

Default View

	File	computed_gender	call_rate
1	NA12875_GW6_C.birdseed.chp	female	99.86
2	NA12878_GW6_C.birdseed.chp	female	99.81
3	NA12891_GW6_C.birdseed.chp	male	99.92
4	NA12892_GW6_C.birdseed.chp	female	99.86
5	NA18500_GW6_Y.birdseed.chp	male	99.53
6	NA18501_GW6_Y.birdseed.chp	male	99.8

Figure 3: CHP summary table. Standard table operations include selecting which columns to display, sort and search, and export to clipboard or file. A line graph can be drawn from this table.

Review CEL Data Quality

Genotyping Console provides new streamlined quality control. Samples are automatically sorted by QC Call Rate threshold (QC Call Rate Threshold ≥ 86). Genotyping Console offers graphical display of QC metrics across samples to identify outliers and includes 72 signature SNPs for tracking sample identity.

Default View

	File	Bounds	QC Call Rate
1	NA19240_GW6_Y.CEL	In	98.08
2	NA06985_GW6_C.CEL	In	97.88
3	NA06991_GW6_C.CEL	In	98.21
4	NA06993_GW6_C.CEL	In	93.98
5	NA06994_GW6_C.CEL	In	98.28
6	NA07000_GW6_C.CEL	In	97.75

Figure 1: Intensity QC table. Standard table operations include selecting which columns to display, sort and search, and export to clipboard or file. A line graph can be drawn from this table.

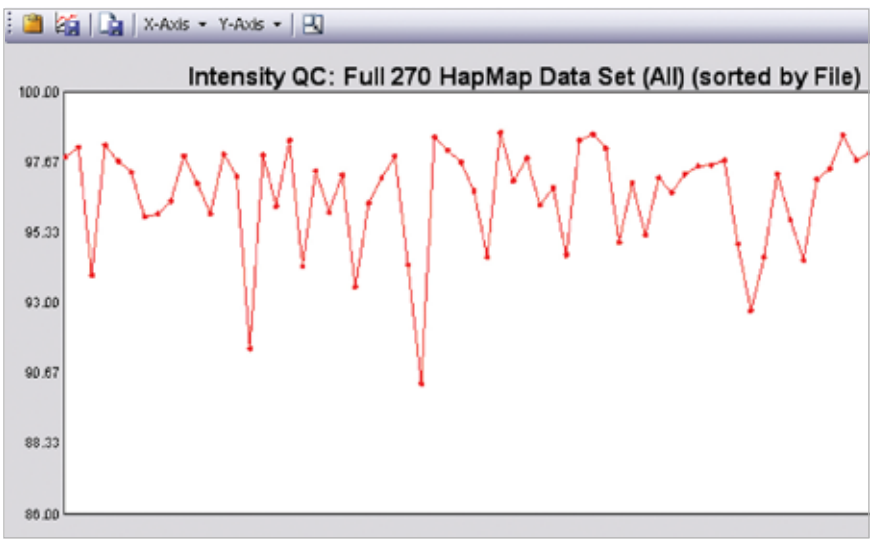

Figure 2: Intensity QC line graph. A user can select the desired X-axis category to plot, and also one or more Y-axis categories. The image or underlying signal data may be saved to file.

Review Genotype Data Quality

Genotyping Console provides investigation of the genotyping results, graphical display of genotyping call rates across samples, display of SNP-level summary statistics, flexible SNP filtering, and visual inspection of SNPs in the genotyping clustering space.

Default View

	SNPID	SNP Call Rate	SNP %AA
1	SNP_A-2131660	100	33.7037
2	SNP_A-1967418	100	3.703704
3	SNP_A-1969580	99.63	0
4	SNP_A-4263484	99.63	23.33333
5	SNP_A-1978185	100	76.2963
6	SNP_A-4264431	100	51.85185

Figure 4: SNP summary table. Includes basic SNP summary statistics, as well as NetAffx annotations. Standard table operations are available.

For many genotyping applications, poorly performing SNPs can lead to an increase in false positives and a decrease in power. Such under-performing SNPs can be caused by systematic or sporadic errors that occur due to stochastic, sample or experimental factors. Prior to downstream analysis, it is prudent to apply some SNP-filtering criteria to remove SNPs that are not performing ideally in the data set in question. Studies on multiple data sets have shown that SNPs with a lower per-SNP call rate tend to have a higher error rate, and disproportionately contribute to the overall error rate in the experiment.

Some common filters used will:
- Remove SNPs with a significantly low per-SNP call rate
- Remove SNPs out of Hardy-Weinberg (HW) equilibrium in controls
- Remove SNPs with significantly different call rates in cases and controls
- Remove SNPs with Mendelian errors

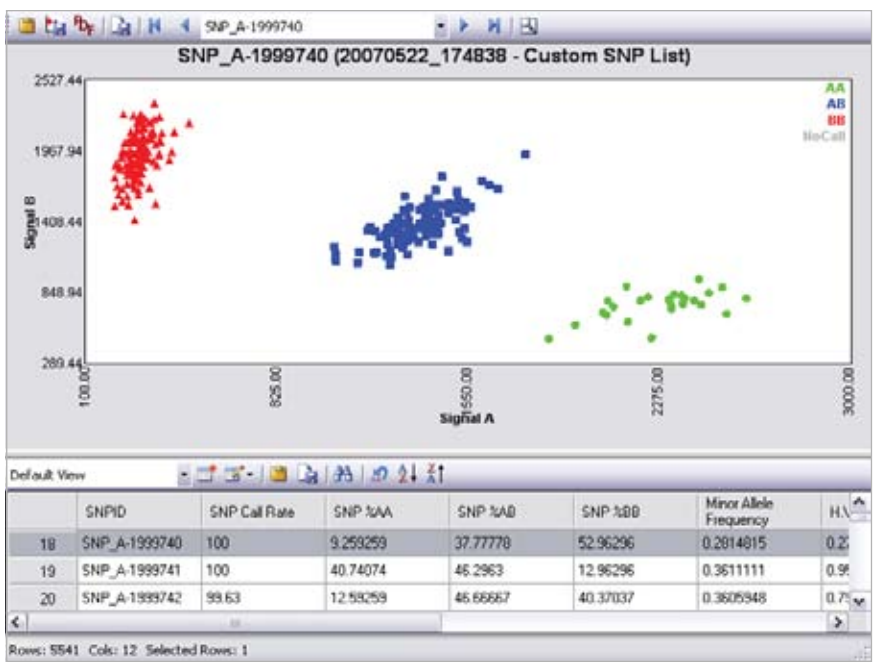

Figure 5: SNP cluster graph. The image or underlying signal data may be saved to file, and the series of selected plots may be saved to a single PDF. The graph and table are linked, and standard table operations are available.

Export Genotypes

CHP file data (genotype calls) may be directly imported into GeneChip®-compatible™ applications that provide comprehensive solutions for genotyping applications such as association studies and linkage. CHP file data may also be exported to a text file for use in other downstream analyses. This includes the probe set name, call, confidence, genotype call and signal data (contrast/strength for BRLMM-P and signal A/B for Birdseed). Options to sort and separate the data by chromosome or export a subset based on a SNP list are available.

Association Analysis and Data Management Solutions Using GeneChip®-compatible™ Tools

Genotypes exported from Genotyping Console can be imported into other software programs for further analysis. The Affymetrix GeneChip-compatible catalog contains solutions for both SNP analysis and data management.

SNP analysis applications provide data analysis solutions for Affymetrix genotyping arrays used for genome-wide linkage or association studies. Applications currently available include:

- Partek® Genomics Suite™
- JMP® Genomics (SAS®)
- Helix Tree® (Golden Helix)
- Exemplar (Sapio Science)
- GenSense (Inforsense)
- Syllego (Rosetta Biosoftware)

AFFYMETRIX GeneChip™ Compatible

SNP data management applications provide centralized storage of data generated by the Affymetrix genotyping arrays coupled with storage of the necessary family or co-variant sample attribute data. Applications currently available include Biocomputing Platforms' BC|Gene and BC|SNPMax and Progeny Software's Progeny Lab 6.

Association Analysis Using Academic Tools

Genotypes exported from Genotyping Console can also be imported into various academic programs for use in association analyses. Applications currently available for whole-genome association analysis include Haploview and Plink.

Genotyping Console File Type Glossary

File Type	Definition
ARR	Command Console sample file; Contains sample attributes
XML	GCOS experiment/sample attribute file (generated by DTT)
CEL	Intensity file (GCOS or AGCC)
GQC	QC results file (contains QC call rate, gender, signature SNP calls)
GTC_WORKSPACE	Workspace file, contains references to ARR/XML/CEL/GQC/ CHP/BIN files and SNP lists
SUMMARY.BIN	SNP statistics file

Genotyping Table/Graph Glossary

Table/Graph	Contents
Sample Attributes	XML/ARR contents
Intensity QC	QC metrics, In/Out Bounds, File Date, #CHP/CEL, XML/ARR
Signature Genotypes	DM Calls for predefined set of 72 SNPs; These SNPs can be used to verify a sample's identity by comparing the genotype calls to a different technology or other reference
CHP Summary	CHP level summary statistics (gender, call rate, % calls, other algorithm metrics)
SNP Summary	SNP level statistics (call rate, % calls, HW p-value) and annotations
Cluster Graph	Graph of contrast vs. strength (BRLMM-P) or signal A vs. B (Birdseed); Includes the SNP summary table
SNP List	The list of SNPs and their annotations
Annotation File	The SNP annotations in a CSV file

AFFYMETRIX, INC.
3420 Central Expressway
Santa Clara, CA 95051 USA
Tel: 1-888-362-2447
(1-888-DNA-CHIP)
Fax: 1-408-731-5441
sales@affymetrix.com
support@affymetrix.com
www.affymetrix.com

AFFYMETRIX, UK Ltd.
Voyager, Mercury Park,
Wycombe Lane, Wooburn Green,
High Wycombe HP10 0HH
United Kingdom
Tel: +44 (0) 1628 552550
Fax: +44 (0) 1628 552585
saleseurope@affymetrix.com
supporteurope@affymetrix.com

AFFYMETRIX JAPAN K.K.
Affymetrix Japan K.K.
Mita NN Bldg 16 Floor
4-1-23 Shiba
Minato-ku, Tokyo 108-0014
Tel: *81-(0)3-5730-8100
Fax: *81-(0)3-5730-8101
salesjapan@affymetrix.com
supportjapan@affymetrix.com