



The document below identifies the software application classes and associated specifications that apply to the use of the Affymetrix GeneChip-compatible™ logo. The use of such logo is governed by the terms of the Affymetrix GeneChip-compatible™ Applications Agreement. You may only use the Affymetrix GeneChip-compatible™ logo if you have executed such an agreement with Affymetrix or have received explicit written authorization from Affymetrix, Inc. to use such logo. Please contact compatible@affymetrix.com for more information on the Affymetrix GeneChip-compatible™ Applications Program.

Compatibility Overview

As detailed in Section 1 below, the GeneChip-compatible™ Specifications center around four basic criteria:

- ARRAY Information - Reading of native Affymetrix data files (CEL/CHP)
- ANNOTATION Information - Supporting Affymetrix annotations provided via NetAffx via an automated retrieval mechanism
- ATTRIBUTE Information - Reading sample attribute information stored in XML files from DTT archives or ARR files from AGCC and link these attributes to associated CEL/CHP files through embedded GUIDs
- DATA EXCHANGE - Writing verified native Affymetrix CHP files if probe level data (CEL files) are input thus providing a data interchange mechanism. Known probe condensation methods should be compared against reference implementation. Applications that focus on upstream management functions must register samples into the Affymetrix core software

Summary of Changes from June 2006 Compatible Specification

- Removed Drug Development Category due to ambiguous definition
- In Definitions, added requirement of CHP file source identification
- Removed references to timeline which have already expired
- In Section I(3b) – AGCC Sample Attribute Information, added the requirement of linking ARR and CEL/CHP files through GUIDs by 30 June 07
- In Section II(1) - Gene Expression, clarified that it is a requirement to parse exon arrays analyzed at the gene level
- In Section II(9) – Exon Expression, added a note regarding the availability of exon-specific analysis files in the NetAffx SDK
- Throughout the document added more hotlinks to application pages, white papers, and sample data sets
- Clarified CHP parsing for II(7) Data Management and II(8) LIMS

- Added requirement to Section II(1) gene expression, Section II(2) General Data Analysis, and Section II(9) exon expression to handle PGF/CLF library files and automated library file downloads by 01 Aug 2007.
- For Section II(10) Pathway/Network Analysis, identifiers associated with newly launched arrays should be supported no less than 90 days following availability on NetAffx.

Application Class	File Reading	Annotation Specification	Sample Attributes Specification	File Writing
Gene Expression	Read CHP files	NetAffx Support	Read XML & ARR files	Write CHP, if CEL is input.
Regulation Analysis	Read CHP files	NetAffx Support	Read XML & ARR files	Write CHP, if CEL is input.
General Purpose Data Analysis	Read CHP files	NetAffx Support	Read XML & ARR files	Write CHP, if CEL is input.
SNP Analysis	Read CHP files	NetAffx Support	Read XML & ARR files	Write CHP, if CEL is input.
Chromosomal Copy Number	Read CNT files	NetAffx Support	Read XML & ARR files	None, currently
DNA Sequence Analysis	Read FASTA files	NetAffx Support	Read XML & ARR files	None
Pathway/Network Analysis	None	NetAffx Support	Read XML & ARR files	None
Expression Data Management	Read CHP files	NetAffx Support	Read XML & ARR files	Register Samples
SNP Data Management	Read CHP files	NetAffx Support	Read XML & ARR files	Register Samples
Laboratory Information Management	Read CHP files	None	Read XML & ARR files	Register Samples
Exon Expression / Alternative Splicing	Read TXT & CHP files	NetAffx Support	Read XML & ARR files	Write CHP, if CEL is input.

Table 1 - Summary of Compatible Specifications

Specifications Date: 05 February 2007 v1.0

Definitions

GCOS-compliant – Data files are GCOS-compliant if they can be parsed using the [GCOS-compliant ANSI C++ file parsers](#). CHP files from GeneChip-compatible applications should be written using AGCC formats, not XDA formats. If XDA CHP files are desired as the output data, GCOS-compliant CHP file writers are available within the [Fusion SDK](#). Sample attribute information should utilize [MAGE DTD XML formats](#) with the [DTT v1.1 specifications](#). Data files (DAT, CEL, & CHP) should NOT be imported into the GCOS process database. GCOS compliance of a file is determined only by file parsing compatibility. The following Affymetrix applications generate GCOS-compliant data: GCOS versions 1.x, GDAS v1 to v3, GSEQv4, GTYPEv4, GTRANS v1, TASv1.0, CCNT v1, CNAT v2 to v3, and BAT v1.

GeneChip® Command Console (AGCC)-Compliant – Data files are AGCC-compliant if they can be parsed using the [Fusion File SDK](#) and conform to the [AGCC File Format Specifications](#). AGCC-compliant CHP files should be written using CHP writing component of the Fusion File SDK. Sample attribute information should utilize ARRAY file formats within the Fusion SDK. [Sample AGCC-compliant data](#) is currently available. The following applications generate AGCC-compliant CHP files: TASv1.1.x, Expression Console v1.x, and BAT v2.

CHP file verification - CHP files must be verified as “parsable by the Fusion SDK” by use of [CHP File Verification Tools](#). Data integrity is of utmost importance to Affymetrix; therefore, two additional tests are required. First, CHP files written using RMA, MASv5, and PLIER signal estimation algorithms will be compared against CHP files from reference implementations of the respective algorithms in Expression Console v1 using the CHP to CHP comparison tool. Second, regardless of the algorithm used to generate the signal estimates or genotypes, data exported into a CHP file will be compared against data within the internal data model of the GeneChip-compatible application by comparing data exported into ASCII text files against data exported into a CHP files using the CHP to Text comparison tool. Failure to ensure that data is accurately exported to a CHP file as well as failure to ensure that CHP files can be used as a data exchange mechanism between GeneChip-compatible applications is grounds for immediate dismissal of the offending party from the GeneChip-compatible Applications Program.

GeneChip Command Console (AGCC) – [GeneChip Command Console](#) (formerly project Calvin) is a current Affymetrix software development project which will deliver the next generation core software providing instrument control, file creation, and limited data management & distribution functionality. AGCC will coexist with GCOS and is scheduled for commercial release in July 2007. A beta release is available for download.

File Source Identification – CHP files written in AGCC formats should have the authoring application, version number, and company information stored in the header so the source of the file can be identified. More information can be found in the [CHP writing white paper](#) available.

I. General Specifications

The GeneChip-compatible specification generally consists of the following four components: File Reading, NetAffx Annotation Support, Sample Attribute Information, and File Writing. Should conflicts arise between the general specifications (Section I) and application class specific specifications (Section II), then the application class specific specifications shall have precedence.

1. *File Reading* – Effectively immediately, a compatible application must be able to read both GCOS-compliant and AGCC-compliant native [Affymetrix file formats](#) associated with a given application class. See application class for specific files required to read. Use of the open source [Affymetrix Fusion File Parsers](#) is the preferred method to meet file reading compatibility specifications. Array quality report information, formerly stored in txt RPT files, has been moved to header information within AGCC-compliant CHP files. For application classes which are required to read CHP files, header information, especially array quality information, is essential to parse and present to the end user for consideration within the analysis.
2. *NetAffx Annotation Support* – Effectively immediately, a compatible application must provide automated tools to ensure updated content is available within the GeneChip-compatible application. Such functionality is easily accomplished through incorporation of the [NetAffx SDK](#) subroutines or use of the XML index. Specific implementation of automated update of NetAffx Annotation information is left to the discretion of each individual software provider, although the general requirement for automated update of annotation information is not optional. Creating URL hyperlinks to individual NetAffx content at the probe set level is not sufficient to satisfy this requirement. Affymetrix considers the annotation information associated with the array to be a critical piece of analysis information which should be incorporated into the analysis workflow at the genomic level.
3. *Reading Sample Attribute Information* – Effectively immediately, a compatible application must: (a) read sample attribute information stored in a Data Transfer Tool (DTT) v1.1 generated XML file. [DTT XML file format documentation](#) is currently available with the [DTT SDK](#). The DTT v1.1 XML file represents sample and experiment associated data from the GCOS process database, and (b) also read the XML-based ARRAY/.ARR files generated by the AGCC software. [Sample .ARR files](#) and [ARR file format documentation](#) are available. By 30 June 2007, a compatible application should associate ARR-sample attribute information with any relevant array data (stored in CEL or CHP files) that will aid the end user

in properly associating sample attribute and microarray data. This association should occur through GUID linking.

4. *File Writing* – Depending on specific application requirements a compatible application must support: (a) CHP WRITING - effective immediately, if CEL files are read and estimates of genotypes or signal estimations at any level are generated within the compatible application, then the compatible application should support both i) the reading of AGCC-compliant CEL files and ii) writing of AGCC-compliant CHP files using formats for the appropriate application class. The resulting CHP file should have identifying information of the authoring application, its version number, and company. The user interface supporting CHP file writing should be clearly available within the FILE menu and be labeled as “Export to CHP file” or provide a similar level of clear and simple access by written agreement with Affymetrix. Please consult the [CHP file writing white paper](#) for more information. CHP files written by a compatible application should undergo the CHP file verification process outlined in the definitions section above; and/or (b) SAMPLE REGISTRATION - effective immediately, a compatible application must register a sample in GCOS by writing an XML file according to DTTv1.1 specifications and registering that file in GCOS using DTT v1.1. Detailed documentation is available with the [DTT SDK](#) . An alternative method for GCOS-compliant sample registration is available through the [GCOS SDK](#) using the [Workflow Action COM Object](#). Sample registration into GCOS needs to be supported until the official GCOS end of life as determined by Affymetrix, (currently targeted for 2008). Further, a compatible application should write an [AGCC-compliant ARRAY file](#) according to Fusion file specifications and place that file into the user-specified AGCC project folder.

II. Detailed Application Class Specifications

1. **Gene Expression** – A compatible application within this application class must:
 - read CHP files as described in Section I(1) and specifically must parse gene-level CHP files from [Expression Console](#)
 - [3' IVT parsing white paper](#)
 - [exon parsing white paper](#)
 - [sample AGCC-formatted 3' IVT data](#)
 - [sample gene-level exon data](#)
 - support NetAffx annotations as described in Section I(2),
 - read sample attribute information as described in Section I (3a and 3b),
 - if CEL files are read, write AGCC-compliant CHP files as described in Section I (4a).
 - Sample Registration Section I(4b) is not applicable.
 - By 01 August 2007, if a compatible application parses data from a CEL file, it must support [PGF/CLF-based library files](#) and provide a mechanism

for automated retrieval of any necessary analysis files (e.g., library files), such as the NetAffx SDK.

- 2. General Purpose Data Analysis** - A compatible application within the General Purpose Data Analysis application class must:
 - read CHP files as described in Section I(1),
 - support NetAffx annotations as described in Section I(2),
 - read sample attribute information as described in Section I(3a and 3b),
 - if CEL files are read, write AGCC-compliant CHP files as described in Section I(4a).
 - Sample Registration Section I(4b) is not applicable.
 - By 01 August 2007, if a compatible application parses data from a CEL file, it must support [PGF/CLF-based library files](#) and provide a mechanism for automated retrieval of any necessary analysis files (e.g., library files), such as the NetAffx SDK.

- 3. Transcriptional Regulation** – A compatible application within the Transcriptional Regulation application class must:
 - read GCOS-compliant BAR file and AGCC-compliant CHP files as described in Section I(1) from [Tiling Analysis Software v1.1](#) (TAS),
 - support NetAffx annotations as described in Section I(2),
 - and read sample attribute information as described in Section I(3a and 3b),
 - if CEL files are read and probe set summaries are created, write AGCC-compliant tiling formatted CHP files as described in Section I(4a). Data ideally should be written to AGCC-formatted CHP files and not BAR files.
 - Sample Registration Section I(4b) is not applicable.

- 4. SNP Analysis** – A compatible application within the SNP analysis application class must:
 - read CHP files as described in Section I(1) from [BRLMM Analysis Tool](#),
 - (Note: Future versions of this specification will include detailed information for integrating with the targeted genotyping platform currently served by the [GTGS application](#).)
 - [White paper on parsing genotyping CHP files](#)
 - support NetAffx annotations as described in Section I(2),
 - read sample attribute information as described in Section I(3a and 3b),
 - if CEL files are read, write AGCC-compliant CHP files as described in Section I(4a). Sample Registration Section I(4b) is not applicable.

- 5. Chromosomal Copy Number** – A compatible application within the Chromosomal Copy Number application class must:

- read CNT files from [CNAT 4](#) or if primary copy number estimates are generated read CEL and/or CHP files as described in Section I(1) as appropriate for a given algorithm,
- support NetAffx annotations as described in Section I(2),
- and read sample attribute information as described in Section I(3a and 3b),
- Currently copy number results are written to a CNT file in ASCII format. Sample Registration I(4b) is not applicable.

6. DNA Sequence Analysis – A compatible application within the DNA Sequence Analysis application class must:

- read FASTA files,
- support NetAffx annotations as described in Section I(2),
- read sample attribute information as described in Section I(3a and 3b).
- File Writing Section I(4) is not applicable.

7. SNP Data Management OR Expression Data Management– A compatible application within either of these application classes must:

- read CHP files as described in Section I(1),
 - full array data need not be parsed from a CHP file, but it is necessary to extract header information to enable quality control measures as well as file lineage
- support NetAffx annotations as described in Section I(2),
- read sample attribute information as described in Section I(3a and 3b),
- if CEL files are read, write AGCC-compliant CHP files as described in Section I(4a),
- register samples to GCOS and AGCC as described in Section I(4b).
 - Registering samples to GCOS is a requirement through a minimum of Dec 2007

8. Laboratory Information Management – A compatible application within the Laboratory Information Management application class must:

- read CHP files as described in Section I(1), (Note that future versions of this specification may extend reading requirements within the LIMS category to include various log and report files that are available within the AGCC system.)
 - full array data need not be parsed from a CHP file, but it is necessary to extract header information to enable quality control measures as well as file lineage
- NetAffx Support Section I(2) is not applicable
- read sample attribute information as described in Section I(3a and 3b),
- CHP File Writing Section I(4a) is not applicable,
- register samples to both GCOS and AGCC as described in Section I(4b).

- Registering samples to GCOS is a requirement through a minimum of Dec 2007

9. Exon Expression/Alternative Splicing – Effective immediately, a compatible application within the Exon Expression application class must:

- read AGCC-compliant CHP files from [Expression Console v1](#) described in Section I(1) ensuring that content can be parsed at both the probe set/exon and transcript cluster/gene levels and full array data content can be efficiently manipulated,
 - [parsing exon CHP files white paper](#)
 - [AGCC formatted CHP files from Human Exon arrays](#)
- support NetAffx annotations as described in Section I(2) ensuring that exon content is available at both the probe set and transcript unit levels as well as for all available NetAffx-supported genome builds,
 - the NetAffx SDK also hosts a [collection of files](#) which can be used specifically for exon analysis.
- read sample attribute information as described in Section I(3a and 3b),
- if CEL files are read and signal estimates are generated, write AGCC-compliant CHP files as described in Section I(4a). Data should be able to be summarized at both the gene/transcript cluster and exon/probe set levels.
 - [Writing CHP files white paper](#)
- Sample Registration Section I(4b) is not applicable.
- By 01 August 2007, if a compatible application parses data from a CEL file, it must support PGF/CLF-based library files and provide a mechanism for automated retrieval of any necessary analysis files (e.g., library files), such as the NetAffx SDK.

10. Pathway/Network – A compatible application within the Pathway/Network Analysis application class must:

- no requirement for file reading Section I(1),
- Support input of Affymetrix probe set identifiers, have incorporated (in an automated fashion to ensure that updated content is available) and utilized NetAffx annotation content within the GeneChip-compatible application's knowledge base and user interface,
- For newly launched arrays, Affymetrix identifiers should be allowed to be input into a compatible application no later than 90 days following the availability of array content on NetAffx.
- read sample attribute information as described in Section I(3a and 3b),
- no requirement for writing Affymetrix files in Section I(4).