AnalysisNote

affymetrix
Biology for a better world

Analysis workflow for UK Biobank Axiom® Array

Overview

UK Biobank Axiom® Array is a powerful tool for translational research in the fields of epidemiology, human disease, and population genetics. Designed by leading researchers for use by UK Biobank, the highly informative content categories include markers corresponding to observed or expected rare alternate alleles of potentially significant phenotypic interest as well as markers in complex regions of the genome. This advanced design requires custom analysis steps to gain full value from the array.

Axiom® Genotyping Solution Data Analysis Guide (P/N 702961) provides detailed information for analyzing Axiom® arrays. For best results users should follow the steps of the Best Practices Genotyping Analysis Workflow. This Analysis Note provides additional instructions for customized analysis options and unique markers specific to UK Biobank Axiom Array.

Multi-allelic markers

UK Biobank Axiom Array content includes a set of 2,881 markers corresponding to 1,360 multi-allelic loci (each having multiple pairs of A/B alleles for the same chromosomal position). Some of these markers correspond to observed or expected rare alternate alleles which were added to the array because of their potential phenotypic impact. For a number of loci in the 'Rare variants in cancer predisposition genes' and 'Rare variants in cardiac disease predisposition genes' categories it is important to know the exact number of each possible A,C,G,T allele. Thus, specific probes were added to the array for each of these alleles.

Multi-allelic markers require additional calculations for interpretation and the development of new genotype-calling algorithms is necessary to analyze them. The current Axiom analysis software does not support multi-allelic marker genotyping and these markers are excluded from the standard analysis option; however, an option for genotyping these markers is provided for users who would like to develop their own methods to interpret the calls.

SNP rs429358 and rs7412 in the ApoE gene

UK Biobank Axiom Array interrogates two challenging SNPs in the ApoE gene (rs429358 and rs7412). These SNPs are important in the study of Alzheimer's disease, coronary heart disease, Rheumatoid Arthritis as well as other conditions. Due to high GC content in the flanking regions, genotyping these SNPs reliably requires a variation from the standard genotyping method. Therefore, the probe set for rs429358 marker is removed from both the standard and optional marker lists and a supplemental option is provided to genotype this one marker separately. SNP specific priors (see *Axiom® Genotyping Solution Data Analysis Guide*, Chapter 2: What is a SNP Cluster Plot for AxiomGT1 Genotypes for more details) have been included for both rs429358 and rs7412.

Analysis options

The customized UK Biobank Axiom Array workflow requires use of the .r3 version of analysis package, UK Biobank Axiom Array, r3 for genotyping rs429358 and rs7412. This package is available for download from within Genotyping Console™ 4.2 (GTC) or from the Technical Documentation tab of the Axiom Biobank Genotyping Array product page, see the "Additional Information" section below for more information. Earlier versions of the analysis library files (.r1 and .r2) do not support the complete custom analysis workflow.

As described in the *Axiom® Genotyping Solution Data Analysis Guide*, the Best Practices Genotyping Analysis Workflow Steps 1-7 can be performed using either GTC or Affymetrix Power Tools (APT). The Best Practices Genotyping Analysis Workflow Step 8, SNP QC, requires the use of APT or SNPolisher. For detailed information on these software packages refer to the *Affymetrix® Genotyping Console 4.2 User Manual* (P/N 702982), *Axiom® Genotyping Solution Data Analysis Guide* (P/N 702961), APT Manual: apt-probeset-genotype (1.16.1), and the *SNPolisher User Guide* (Version 1.5 or greater).

There are two options for executing Best Practices Genotyping Step 7 for UK Biobank Axiom® Array. Option one is the standard analysis option which produces genotype calls for the bi-allelic markers only. The genotypes for these markers are called using the Axiom genotyping algorithm (AxiomGT1) and are thus supported. This option is enabled by selecting the "Bi-allelic markers" list, Figure 1 and Table 1.

A second option is provided to genotype all markers on the array including the unsupported multi-allelic markers. This option is enabled by selecting the "Bi-allelic Plus Unsupported Multiallelic markers" list, Figure 1 and Table 1. Note: The AxiomGT1 algorithm is not designed to handle multi-allelic markers; therefore, the genotype calls from the multi-allelic markers in this option are not supported by Affymetrix. Users of this option should exclude output for probe sets not in the bi-allelic marker list for any routine downstream analysis. New algorithms must be developed by the user to call these multi-allelic markers.

A complete list of markers is included in the annotation file (Axiom_UKB_WCSG.na34.annot.db). The identity of probe sets associated with multi-allelic markers is also provided.

Note: There are 267 bi-allelic markers whose "best" probe set (selected by the SNPolisher Classification step) may change between the Bi-allelic markers and Bi-allelic Plus Unsupported Multiallelic markers options.

Complete instructions for executing the Best Practices Workflow are detailed in the *Axiom® Genotyping Solution Data Analysis Guide* (P/N 702961) for GTC (Chapter 7) and APT (Chapter 8). Genotyping fewer than 96 unique individuals requires the use of SNP specific priors; select the appropriate options in Table 1, section "LessThan96", when genotyping fewer than 96 unique samples. Select the appropriate options in Table 1, section "96orMore", when genotyping 96 or more unique samples.

Figure 1. Genotyping Analysis Options in GTC. Analysis configuration options available for genotyping various marker lists in GTC. Marker lists are available for QC genotyping (Step1) and sample genotyping (Step2). Sample genotyping includes options for Bi-allelic markers only, Bi-allelic Plus Unsupported Multiallelic markers, and Supplemental Analysis (for genotyping rs429358).
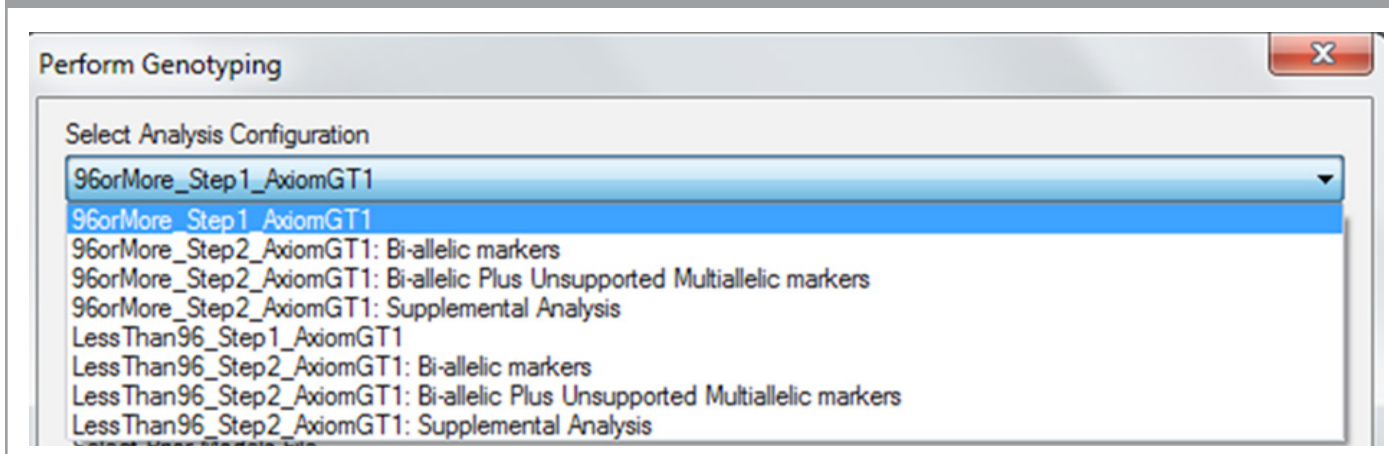
Best Practices Genotyping Analysis Workflow for UK Biobank Axiom® Array

The steps listed below can be applied to analysis performed in GTC or APT. Refer to the *Axiom® Genotyping Solution Data Analysis Guide* (P/N 702961); see Chapter 3: Best Practices Genotyping Analysis Workflow for a description of each step. An example workflow is provided in Figure 2. Table 1 lists the files used by each software package. Ensure the correct sample size option is also selected. There are no changes from the instructions provided in the *Axiom® Genotyping Solution Data Analysis Guide* (P/N 702961) for Steps 1 – 6.

Step 1: Group samples plates into batches

Step 2: Generate sample Dish QC (DQC) values

Step 3: QC the samples based on DQC values. Genotyping is performed on passing samples.

Step 4: Generate sample QC call rates (Step1.AxiomGT1), see Table 1 for file names.

Step 5: QC samples based on QC call rate

Step 6: QC the plates based on sample pass rate and average QC call rate of passing samples

Step 7: Genotype passing samples and plates.

A marker list must be selected to perform sample genotyping. It is recommended to select the "Bi-allelic markers" list unless you have developed advanced algorithms to handle the multi-allelic markers in the "Bi-allelic Plus Unsupported Multiallelic markers" list. Select one of the following two marker list options; see Table 1 for file names:

   A. Bi-allelic markers (recommended)
   B. Bi-allelic Plus Unsupported Multiallelic markers (requires user developed advanced algorithm)

The genotype output includes all markers in the selected option, except ApoE SNP rs429358. To genotype this SNP, select the appropriate Supplemental Analysis file, Table 1. The genotype output from Supplemental Analysis includes only ApoE SNP rs429358.

Manually append the output files from Supplemental Analysis to the corresponding output files from the selected marker list (Bi-allelic markers or Bi-Allelic Plus Unsupported Multiallelic markers) to produce genotypes for all selected markers. This must be done for each of the following output files: AxiomGT1.calls.txt, AxiomGT1.confidences.txt, AxiomGT1.snp-posteriors.txt, and AxiomGT1.summary.txt. To append the files, remove the header lines (those that begin with "#") from the appropriate output file produced by the Supplemental Analysis step. After header removal, append the file to the file with same name produced by the Step 7 genotyping (Bi-allelic markers or Bi-allelic Plus Unsupported Multiallelic markers).

Step 8: Execute SNP QC on the appended output file including all markers. Refer to the *Axiom® Genotyping Solution Data Analysis Guide* (P/N 702961) for details on SNP QC when genotyping in GTC (Chapter 7) and APT (Chapter 8).

Figure 2 Example analysis workflow for UK Biobank Axiom® Array. Boxes enclose GTC and APT genotyping steps (file names required for each Best Practices Workflow step are detailed in Table 1), circles enclose output genotypes, and curved arrows indicate output files to be appended before executing SNP QC.
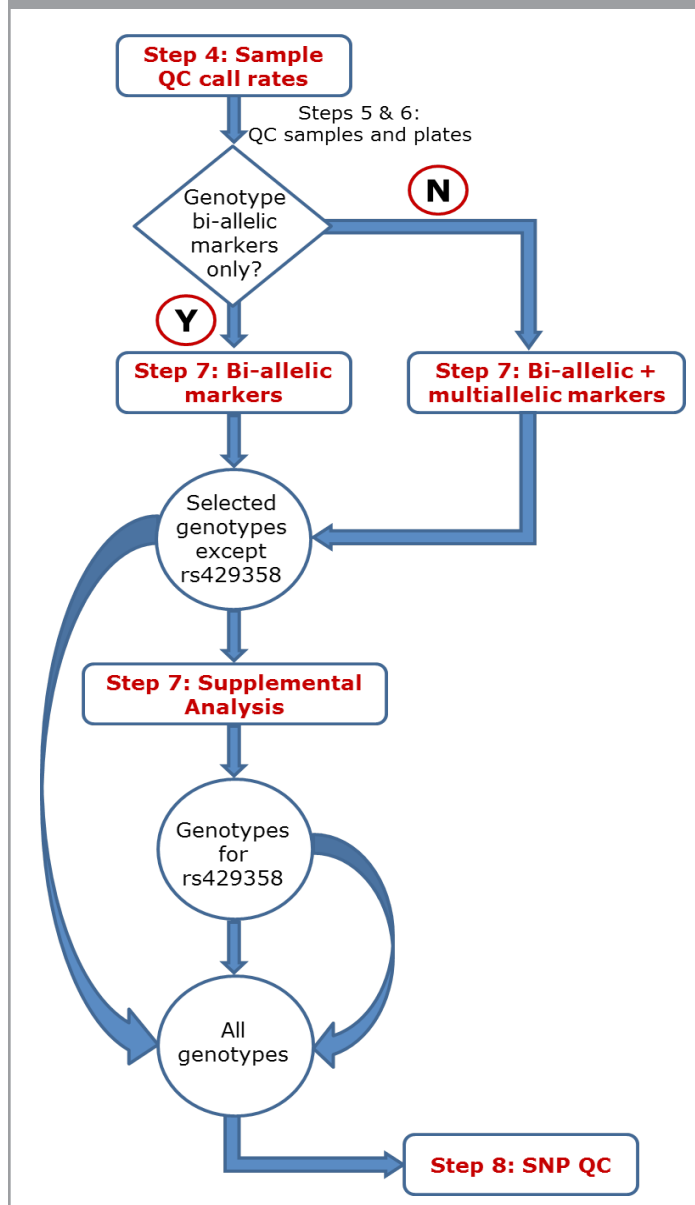
Table 1. Files used in GTC and APT for QC and sample genotyping.

Batch size	Step	GTC	APT
< 96 samples	Step 4	LessThan96_Step1_AxiomGT1	Axiom_UKB_WCSG_LessThan96_Step1.r3.apt-probeset-genotype.AxiomGT1
	Step 7: Bi-allelic markers only	LessThan96_Step2_AxiomGT1: Bi-allelic markers	Axiom_UKB_WCSG_LessThan96_Step2_Bi-allelic.r3.apt-probeset-genotype.AxiomGT1.xml
	Step 7: Bi-allelic + multiallelic markers	LessThan96_Step2_AxiomGT1: Bi-allelic Plus Unsupported Multiallelic markers	Axiom_UKB_WCSG_LessThan96_Step2_Bi-allelicPlusUnsupportedMultiallelic.r3.apt-probeset-genotype.AxiomGT1.xml
	Step 7: Supplemental Analysis	LessThan96_Step2_AxiomGT1: Supplemental Analysis	Axiom_UKB_WCSG_LessThan96_Step2_Supplemental_Analysis.r3.AxiomGT1.xml
≥ 96 samples	Step 4	96orMore_Step1_AxiomGT1	Axiom_UKB_WCSG_96orMore_Step1.r3.apt-probeset-genotype.AxiomGT1
	Step 7: Bi-allelic markers only	96orMore_Step2_AxiomGT1: Bi-allelic markers	Axiom_UKB_WCSG_96orMore_Step2_Bi-allelic.r3.apt-probeset-genotype.AxiomGT1.xml
	Step 7: Bi-allelic + multiallelic markers	96orMore_Step2_AxiomGT1: Bi-allelic Plus Unsupported Multiallelic markers	Axiom_UKB_WCSG_96orMore_Step2_Bi-allelicPlusUnsupportedMultiallelic.r3.apt-probeset-genotype.AxiomGT1.xml
	Step 7: Supplemental Analysis	96orMore_Step2_AxiomGT1: Supplemental Analysis	Axiom_UKB_WCSG_96orMore_Step2_Supplemental_Analysis.r3.AxiomGT1.xml

Support

Users should contact their local Affymetrix Field Application Specialist or send an email to Support@affymetrix.com.

Additional information

For more information about UK Biobank Axiom® Array and Axiom Genotyping Solution data analysis, please consult the following resources:

- UK Biobank Axiom® Array Data Sheet, P/N GGNO03529
- Genotyping Console™ 4.2 User Manual, P/N 702982
- Axiom® Genotyping Solution Data Analysis Guide, P/N 702961
- APT Manual: apt-probeset-genotype
- Analysis library files are available on the Axiom® Biobank Genotyping Arrays product page, Technical Documentation tab:
  - UK Biobank Axiom® Array Analysis Files, r3
  - UK Biobank Axiom® Array, Annotation Converter, r3