Technical Note

Affymetrix®
Revolutionize life

Analysis workflows for Axiom® Genome-Wide Human Origins 1 Array

Overview

Axiom® Genome-Wide Human Origins 1 Array is a powerful tool for human population geneticists to learn about human history, migration, and natural selection. Designed specifically for studies of human population and evolutionary genetics, the array includes well-documented SNPs selected using a simple and clean ascertainment strategy in samples of known ancestry. This avoids the confounding biases that can result from use of GWAS arrays in population genetics studies and permits evolutionary hypotheses to be studied in a straightforward and quantitative way, thus enabling valuable inferences about human history.

The Axiom Genome-Wide Human Origins 1 Array design strategy is described by Keinan A., *et al.*, *Nature Genetics* **39**(10):1251-5 (2007)[1], in which SNPs are discovered by comparing two chromosomes from the same individual of known ancestry and then genotyped in a larger panel of samples from the same population to reduce ascertainment bias.

Axiom Human Origins 1 Array contains over 629,000 SNPs from modern human populations. Among these variants, 542,399 SNPs were discovered using the strategy referred to above for 11 populations (San Bushmen, Yoruba, Mbuti Pygmies, French, Sardinian, Han, Cambodian, Mongolian, Karitiana, Papuan, and Bougainville). An additional 87,044 SNPs were selected from the Axiom® Genomic Database and include mitochondrial SNPs, Y chromosome SNPs, and SNPs present in the Affymetrix® Genome-Wide Human SNP Array 6.0. The SNPs are collected into 13 different panels. The first 12 panels contain tens of thousands of SNPs per population, which can be used for allele frequency spectrum analysis in each population separately or as a whole. The 13th panel is based on alignment of chimpanzee, Denisova, and San Bushmen genomes and also contains SNPs from chromosome X for X-autosome comparisons.

Data generated from Axiom Human Origins 1 Array can be analyzed using Genotyping Console™ (GTC) Software in two different genotyping workflows:

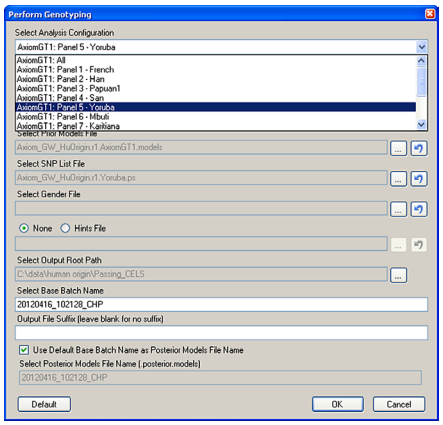
- **Pre-analysis filtering:** Genotype a subset of markers using one of the predefined marker lists for your population of interest.
- **Post-analysis filtering:** Genotype all of the markers on the array at once, and restrict analysis to the population-specific marker data by filtering the results. This method requires less processing time, making it faster to compare and contrast the results from multiple populations. However, care must be taken to ensure that only the markers for a relevant population are included in the analysis by filtering the results using the population-specific marker lists.

Clustering protocols

Pre-analysis filtering:

1. Highlight the CEL files that passed QC.
2. Select the population-specific marker list in GTC Software analysis configuration to produce genotype calls for that panel (Figure 1).

Figure 1: Choose the population-specific analysis configuration from the drop down menu.

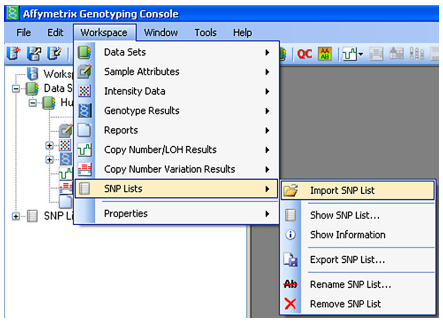


3. To export the population-specific data after the analysis is completed, select the CHP files of interest, and then select "Export Genotyping Results."
4. To obtain the genotypes for a different population, a second analysis run will need to be performed by selecting the second population of interest.

Post-analysis filtering:

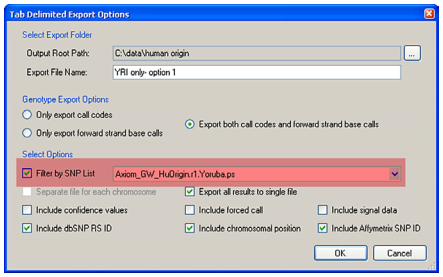
1. Select all of the CEL files that passed QC.
2. Choose the "Axiom GT1 All" option in the GTC Software analysis configuration (Figure 1).
3. Choose the "Import SNP list" option under "Workspace-SNP Lists" in GTC, and select the SNP list for the population of interest. (Figure 2).

Figure 2: Browse to the folder containing the population-specific SNP lists, and upload them to the GTC Software.



4. After the analysis is completed, filter and export the predefined marker lists.
5. To export the population-specific data, select the CHP files of interest, and select "Export Genotyping Results" (Figure 3).

Figure 3: Check the "Filter by SNP List" option, and select the SNP list corresponding to the appropriate population.



Exporting the data to PLINK

Select the "Export Genotyping Results for PLINK" option.

Note: PLINK requires specific sample information to function, which should be provided using the pedigree template in GTC Software. Detailed information about running analyses, importing and filtering using SNP lists, and exporting data from GTC Software is found in the *Genotyping Console™ User Guide*.

Additional information

For more information about Axiom® Genome-Wide Human Origins 1 Array and Genotyping Console™ Software, please consult the following resources:

Axiom® Genome-Wide Human Origins 1 Array Data Sheet, P/N DNA01155 Rev. 1

A SNP Array for Human population genetic studies, P/N DNA01075 Rev. 1

Genotyping Console™ 4.1 User Manual, P/N 702982 Rev. 1

Axiom® 2.0 Genotyping Solution Data Analysis Guide, P/N 702961 Rev. 1

Harvard HGDP-CEPH Genotypes for Population Genetics Analyses FLAT FILES SUPPLEMENT 10. (http://www.cephb.fr/en/hgdp/)

References

1. Keinan A., *et al*. Measurement of the human allele frequency spectrum demonstrates greater genetic drift in East Asians than in Europeans. *Nature Genetics* **39**:1251-5 (2007).

Affymetrix, Inc. Tel: +1-888-362-2447 ■ **Affymetrix UK Ltd.** Tel: +44-(0)-1628-552550 ■ **Affymetrix Japan K.K.** Tel: +81-(0)3-6430-4020
Panomics Solutions Tel: +1-877-726-6642 panomics.affymetrix.com ■ **USB Products** Tel: +1-800-321-9322 usb.affymetrix.com

www.affymetrix.com Please visit our website for international distributor contact information.

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

P/N DNA01149 Rev. 1