ABSTRACT

Background: Large-scale multiplexed SNP genotyping has been made possible by the ongoing development of high-density DNA oligonucleotide arrays coupled with new molecular biology assays. The data analysis processing of these arrays requires automated methods and workflows and thus Axiom® genotyping best practices have been developed and applied to the human genome (http://www.affymetrix.com/support/downloads/manuals/axiom_best_practice_supplement_user_guide.pdf). Plant and animal genomes however present additional challenges due to the complex organization of polyploid genomes, high levels of species-specific nucleotide and structural diversity, and the use of relatively novel sequencing-derived content that has not been previously validated with independent genotyping methods.

Results: We describe the Axiom® data analysis workflow that combines the best practices for human genotyping with new steps and software that are specifically targeted to the complexities associated with plant and animal genotyping. Genotyping calls are produced for humans, plants, and animal genomes with AxiomGT1, which is an automated clustering algorithm used in Axiom® Genotyping Console™ Software that features adaptive, dynamic clustering. Following genotyping, “SNPolisher™”, a newly developed R package for Axiom® Genotyping Arrays, uses AxiomGT1 output to (1) produce SNP Quality Control (QC) metrics; (2) divide markers into six classes that are relevant to the genomic complexities of plants and animals; (3) provide cluster visualization; and (4) execute “Off-target variant (OTV) genotyping”. In summary, this workflow returns the maximum number of high-quality genotype calls along with automatic division of polymorphic, high-resolution markers from those markers that can benefit from further examination based on knowledge of the genetics of the sample under study.

Best practice workflow

Plant and animal genomes have long been known to be much more complex than human genomes, which pose computational challenges in the genotyping process. We adapted the best practice Axiom® workflow for human genotyping with two new steps executed by “SNPolisher™”, shown in Figure 1.

SNP QC metrics generation and SNP classification

The steps of post-processing genotypes (Step 8 in Figure 1) are as follows:
2. SNPs/probe sets are classified based on SNP QC metrics into six categories: Polymorphic high-resolution, “Monomorphic high-resolution”, “Off-target variant”, “Call rate below threshold”, “No minor homoyzogote”, and “Other”.

Note: These two functions are achieved in the SNPolisher™ R package.

Figure 2: Typical patterns of six SNP/probe set categories.

Genotype cluster visualization

To visualize SNPs, SNPolisher can draw genotype cluster plots, as shown in Figure 3 with color legends. Both prior and posterior information of genotype clusters are represented by ovals. If reference genotypes are provided, they can be plotted in parallel as a comparison. Moreover, a list of specified samples can be highlighted by green points on the cluster plots.

Off-target variant genotyping (Step 9 in Figure 1)

Off-target variants (Didion, et al. 2012) can be detected and genotyped by SNPolisher. Figure 4 shows an example of OTV clusters before and after OTV genotyping.

Figure 4: Genotype cluster plots of a SNP before and after OTV genotyping.

| Step 1: Group samples into batches. For each batch, perform the following |
| Step 2: Generate sample “DQC” values |
| Step 3: QC samples based on DQC values |
| Step 4: Generate sample QC call rates |
| Step 5: QC samples based on QC call rate |
| Step 6: QC the plates offline (Excel or script) |
| Step 7: Genotype passing samples & plates using all SNPs and generic priors |
| Step 8: SNP QC and Classification (SNPolisher R package) |
| Step 9: (optional) “OTV genotyping” (SNPolisher R package) |

Legend

- Step completed in Affymetrix Power Tools or Genotyping Console
- Step completed with Excel or R script in Windows or Linux environment

Introduction to SNPolisher™ R Package

SNPolisher is an R package specifically designed to post-process genotyping results by Affymetrix’ Axiom® Genotyping Arrays. It can calculate the QC metrics for each SNP/probe set to determine its quality and classify SNPs/probe sets into six major categories, “Polymorphic high-resolution”, “Monomorphic high-resolution”, “Off-target variant”, “Call rate below threshold”, “No minor homozygote”, and “Other”. It can select the best probe set to represent a SNP if multiple probe sets exist for a specific SNP. It can also generate the cluster plot visualization for each SNP/probe set. SNPolisher can also detect OTV clusters and OTV-genotype those SNPs to produce AA, AB, BB, and OTV genotype clusters.