applied biosystems

UKB_WCSGAX: UK Biobank 500K Samples Processing by the Affymetrix[1] Research Services Laboratory

June, 2017

[1] Now part of Thermo Fisher Scientific

ThermoFisher SCIENTIFIC

Contents

Overview

The Applied Biosystems™ Microarray Research Services Laboratory (formerly the Affymetrix™ Research Services Laboratory [ARSL]) located in Santa Clara, CA has processed approximately 500,000 samples for the UK Biobank and UKBiLEVE projects. This document describes the processing of samples from receipt to delivery of genotypes. Details of analysis for these projects are given in another document (1).

Facility

Bioinformatics Services (BiS) personnel fulfill the data analysis and project management functions for services projects. Lab personnel perform incoming sample quality control steps and Applied Biosystems™ Axiom™ assay biochemistry workflow. Pre-whole genome amplification (WGA) and post-WGA activities are performed in separate areas to minimize PCR contamination risk.

Pre-WGA: The Pre-WGA Room is a positive pressure environment with limited personnel access and dedicated garments and equipment. All refrigerators and freezers in this space are real-time monitored for any temperature deviations. Samples are tracked by in-house developed laboratory information management system (LIMS). Non-amplified template (genomic DNA samples) received in 96-well microtiter plates undergo accessioning and LIMS registration, quantification, and generation of normalized daughter sample plates. Such daughter sample plates are then genotyped on the Applied Biosystems™ Axiom™ platform in the Axiom main laboratory. Original gDNA sample plates received from UKBiobank never leave the Pre-WGA lab other than for return shipment to the customer or disposal.

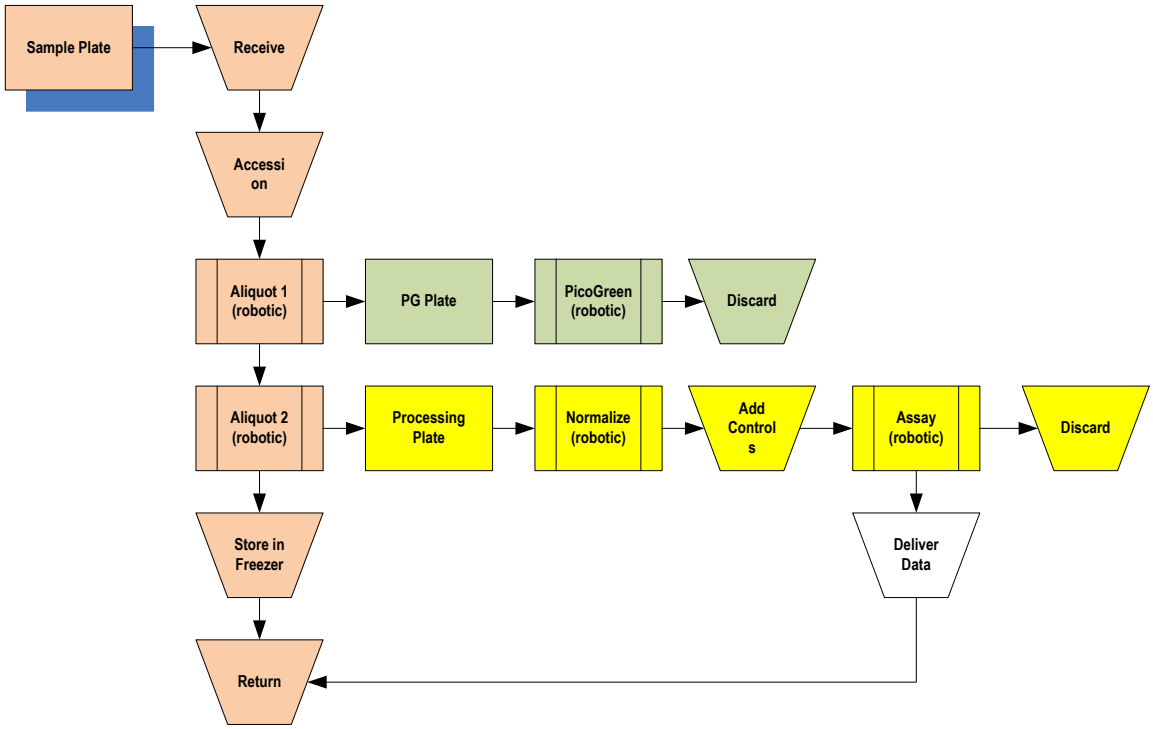
Main Axiom lab: The main Axiom lab is where normalized genomic DNA sample plates are processed to genotype data. This is a negative pressure environment with dedicated garments and equipment. Normalized sample plates are tracked via LIMS through all steps of sample processing in this space.

Process overview

1) Applied Biosystems Microarray Research Services Laboratory provides blank barcoded 96-well plates and seals to UK Biobank
2) UK Biobank plates DNA samples and prepares the plates for return shipment
3) Receipt and accessioning of the DNA sample plates upon return to the Applied Biosystems Microarray Research Services Laboratory
4) DNA concentration readings by the Applied Biosystems Microarray Research Services Laboratory
5) Normalization of DNA to working concentration
6) Processing of samples using the Axiom assay
7) Reprocessing as necessary
8) Data analysis
9) Data delivery

Safe sample handling

To minimize the possibility of contamination, degradation, loss, and mistracking, we have automated most steps in the process to minimize handling of customer plates. The process is detailed in the flowchart below, with steps utilizing robotics noted.



Sample handling workflow

All gDNA samples (original and normalized aliquots) are stored in temperature-monitored freezers, inside restricted access laboratory areas. Samples are tracked during the entire Axiom assay process using proprietary Laboratory Information Management System (LIMS) software.

Services Process Flow

Sample Receipt

UK Biobank shipped DNA sample plates to the Applied Biosystems Microarray Research Services Laboratory on an ongoing basis with the schedule set and distributed by the UKBiobank Program Manager. Sample plates were shipped on dry ice using shipping boxes provided by the Applied Biosystems Microarray Research Services Laboratory. Upon shipment, the UKBiobank Clinical Trial Service Unit (CTSU) uploaded the manifest file to a secure cloud location accessible only by CTSU and a select group of Applied Biosystems employees involved in project management.

Prior to shipment to the UK Biobank, plate barcodes were associated with the project in the LIMS database. Upon receipt of the plates, Applied Biosystems

Microarray Research Services Laboratory personnel scanned plate barcodes into the LIMS interface accession screen, causing the underlying database to be updated with the receipt date for the plate. Any anomalies in plate appearance were noted at this time and such notes were communicated to the UK Biobank Program Manager.

DNA quantification

Any plates received in good condition proceeded directly to quantification. Quantification was performed by the Applied Biosystems Microarray Research Services Laboratory personnel using Invitrogen™ PicoGreen™-based quantification method. The results of this quantification were uploaded to a secure cloud location and an email notification was sent. For the purpose of this study, a DNA concentration of ≥2ng/uL was used as a threshold. Plates with ≥80% of samples having a concentration higher than the threshold were considered to be passing. In practice, plates with more than a few samples below this threshold were flagged for review by UK Biobank staff and processing of such plates was undertaken only upon agreement by both parties. In some cases, replacement sample plates were provided.

Sample normalization

Plates passing the quantification step were then used to make normalized working plates. Using the concentrations determined, samples were transferred and diluted as appropriate to the working concentration for the Axiom assay. During the normalization step, two control gDNA samples were added to each plate. When plating samples, customers are instructed to leave wells A12 and E12 blank for this purpose.

Sample processing using the Axiom assay

For an overview of the Axiom assay, please see Axiom 2.0 Assay Automated Workflow (2).

Reprocessing

In cases where sample quality was regarded by an experienced analyst to be inadequate, samples were reprocessed. This could be done either for a whole processing plate, in which case the assay would be redone from the normalized DNA, or for a subset of samples in a plate. For the latter case, samples would be combined using automation to create a new barcoded 96-well plate of normalized DNA for processing. The transfer was tracked in LIMS such that the original DNA well location was preserved.

Data analysis

Data analysis was done according to the best practices guide in the Axiom Genotyping Solution Data Analysis Guide (3). Any exceptions or additions are detailed in the Affymetrix UKB_WCSGAX Genotype Data Generation note (1). Note that for each sample plate, two well-characterized control samples derived from cell lines were included. These controls (HG00097 and HG00264) allowed for checking of control concordance to 1000 Genomes data as well as sample

reproducibility. This information, combined with sample data quality for the rest of the plate, was used to make decisions about reprocessing samples. Note that in some cases data from a previous experiment with a given sample might have been included in the final delivery if the reprocessing data was inferior. Processing plate identities are included in the sample table delivered with the genotypes.
Data was delivered in batches for which genotype clustering was done together. These batches were created by gathering the equivalent of approximately 50 processing plates or about 4800 samples (including about 100 positive control samples). Array image scans were chosen for a specific batch because data quality was deemed adequate or reprocessing had already been attempted. For each batch, when deliverables were created by one analyst, another analyst would review the work carefully. An additional check ("final review") was done upon completion of the peer review before data was made available for download by UK Biobank staff.

Data Delivery

Data was delivered using a secure cloud location and an email was sent to notify UK Biobank/WTCHG staff of data availability. Files delivered are described below.
• **UKB_WCSGAX_Report.pdf**: summary data analysis report for this project.
• **1 each of the following files for each batch**:
• **UKB_WCSGAX_<batch number>_Sample_Table.cvs**: sample summary report for all samples.
• **AxiomGT1.calls.txt**: file containing tables with genotype calls for all samples and all probe sets.
• **AxiomGT1.calls.NMA.txt**: file containing tables with updated genotype calls only for rare variants.
• **AxiomGT1.calls.rs429358.txt**: 1 files(s) containing tables with genotype calls for all samples and one ApoE probeset for which special genotyping parameters were applied.
• **AxiomGT1.confidences.txt**: file containing tables with confidences of the genotype calls for all samples and all probesets.
• **AxiomGT1.confidences.rs429358.txt**: 1 file(s) containing tables with confidences of the genotype calls for all samples and one ApoE probeset for which special genotyping parameters were applied.
• **AxiomGT1.snp-posteriors.txt**: file containing tables with locations of the cluster centers, etc., for each probeset.
• **AxiomGT1.summary.txt:** file containing tables with summarized A and B intensities for all samples and all probesets. These can be used for SNP plotting, using the SNPolisher package.
• **AxiomGT1.probabilites.txt**: matrix file containing comma-delimited probabilities for each genotype (AA, AB, BB) for each probeset and sample.
• **Ps.performance.txt**: matrix file with performance statistics (CallRate, Allele Counts, FLD, etc) and classification into one of seven classes: PolyHighResolution, MonoHighResolution, OffTargetVariant, NoMinorHom, CallRateBelowThreshold, Hemizygous, and Other.
• **converted.ps**: file with a suggested filtered SNP list. For human, it includes SNPs in PolyHighResolution.ps, MonoHighResolution.ps, NoMinorHom.ps, and

Hemizygous.ps. *It is highly recommended to exclude probesets that are not in the converted.ps list from downstream analysis.*
• **PolyHighResolution.ps**: file containing SNPs with good cluster resolution and at least two examples of the minor allele.
• **MonoHighResolution.ps**: file containing SNPs with good cluster placement but less than two examples of the minor allele.
• **OffTargetVariant.ps**: file containing off-target variant SNPs. Can be called by OTV genotyping.
• **NoMinorHom.ps**: file containing SNPs that are missing examples of the minor homozygous genotype.
• **CallRateBelowThreshold.ps**: file containing SNPs that pass all SNP QC thresholds except for call rate.
• **Hemizygous.ps**: file containing Y, W, and MT SNPs.
• **Other.ps**: file containing SNPs with one or more cluster properties below threshold.
• **AxiomGT1.cnv.baf.final.txt**: file containing B allele frequencies.
• **AxiomGT1.cnv.log2ratio.final.txt**: file containing Log2 ratios for copy number analysis.
• ***.CHP files**: Axiom™ GT1 genotypes file named as: [Sample Identifier].AxiomGT1.chp (1 file for each customer and Affymetrix control sample genotyped in a given batch)
• **raw_data folder**: 1 set of raw data files named as: [Sample Identifier].[File Type] for each customer and Affymetrix sample. Each set of files contains 2 files for a given array: *.CEL and *.ARR.
• **Axiom Reference Files**: standard reference files for Axiom™ including library files, annotation files, file format description, etc.

References

1. Affymetrix Genotype Data Generation. [Online]
http://biobank.ctsu.ox.ac.uk/crystal/refer.cgi?id=155582.

2. Axiom 2.0 Assay Automated Workflow. [Online]
http://tools.thermofisher.com/content/sfs/manuals/702963_4_Final_Ax_96FX_UG.pdf.

3. Axiom Genotyping Solution Data Analysis Guide . [Online]
http://tools.thermofisher.com/content/sfs/manuals/axiom_genotyping_solution_analysis_guide.pdf.