Validation of Illumina array-based genotyping for the All of Us Research Program

MICHAEL DaSILVA[1], ALYSSA MACBETH[1], STEVEN HARRISON[1], BETTY WOOLF[1], WENDY BRODEUR[1], MAEGAN HARDEN[1], GINA VICENTE[1], SARAH BABCHUCK[1], KATELYN FLOWERS[1], KUNSANG GYALTSEN[1], BRIAN GRANGER[1], GEORGE GRANT[1], HEIDI REHM[1], STACEY GABRIEL[1], NIALL LENNON[1]

[1]*Broad Institute of MIT and Harvard, 320 Charles Street, Cambridge, MA 02141*

Introduction

The Genomics Platform at the Broad Institute has established College of American Pathologists (CAP) / Clinical Laboratory Improvement Amendments (CLIA) compliant genotyping capabilities at scale, which will be utilized for the All of Us Research Project (AoURP).

AoURP aims to generate sequencing and genotyping data from 1 million or more research participants across the U.S. Medically actionable results from a predefined set of genes (AoU Medically Actionable Panel, AoUMAP) will be returned to participants after orthogonal validation in a clinical validation laboratory.

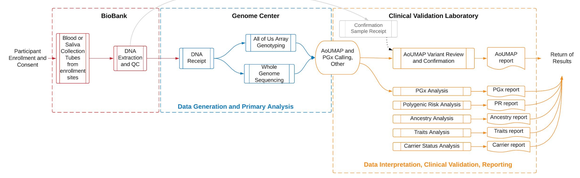

Figure 1. High level AoU sample and data flow (proposed)

As one of three Genome Centers selected by the AoURP program, the Broad Institute Genomics Platform has built the operational and clinical infrastructure required to genotype >300,000 AoURP participants over the next 4-5 years.

We performed an analytical validation study to assess the accuracy and precision of the AoU Array. A combination of reference samples, PGx cell lines, and previously tested clinical samples were used to verify that the array is suitable for its intended use as part of the AoURP.

About the AoU Array

- >1.8 million markers using the Illumina Infinium platform
- Designed to cover content including:
  - >28,000 variants known to be pathogenic and likely pathogenic sites (as defined by ClinVar) across the AoUMAP (https://www.ncbi.nlm.nih.gov/clinvar/)
  - >29,000 variants with pharmacogenomic (PGx) relevance
  - >14,000 common disease variants identified from GWAS
  - >18,000 HLA variants to enable low resolution HLA calling
  - >428,000 variants identified in a diverse panel of gnomAD exomes (http://gnomad.broadinstitute.org/)

**These numbers are not for the commercial Global Diversity Array*

Onboarding the AoU Array into our LIMS

1. SNP manifest file (.bpm and csv) and cluster file (.egt) provided by Illumina (based on data from >1200 individuals).
2. Cluster file manually reviewed and edited by Kim Doheny @ CIDR, based on SNP and sample criteria.
3. Files implemented into our internal Laboratory Information Management System (LIMS), associated to AoU chip.

Clinical Array Lab and Analysis Workflow

Each lab step is run on automation and messaged to our LIMS, to capture metadata and maintain chain of custody. Once chips are scanned, raw scanner files and metadata from LIMS are pushed to the cloud-based Analysis Workflow. Data is reviewed and signed off before delivery.

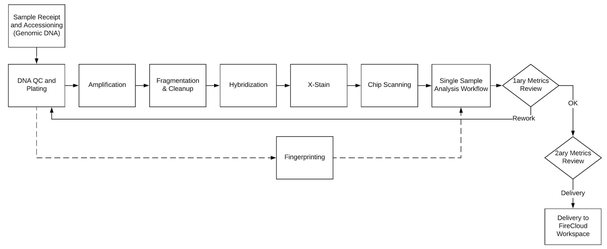

Figure 2. High level AoU sample and data flow

The pipeline takes scanned data files as input and utilizes Illumina-supplied and internally developed methods, run in the cloud, to produce annotated VCF files and quality control metrics.

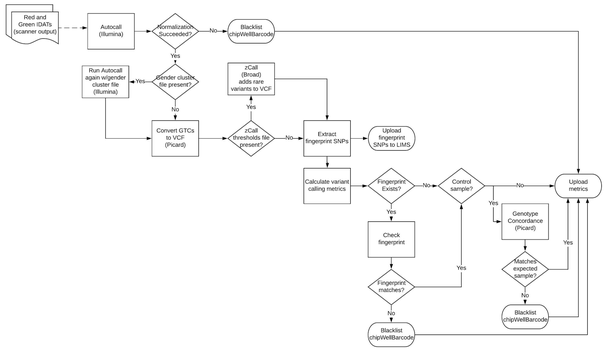

Figure 3. Detailed overview of analysis workflow for array sample

Formalization of CAP/CLIA quality standards in our genotyping process

Quality monitoring and quality management have been and are paramount in our research setting. To become CAP and CLIA compliant we focused on the following areas:

- Version Control
  - Lab and automation protocols
- Establishing equipment equivalency
  - Across sample prep automation and detectors
- Equipment maintenance procedures
  - Scheduled PMs, breakdown policies
- Personnel
  - Training plans, experience and qualification documentation
- Documentation of standard operating procedures
  - Lab, LIMS, Pipeline
- Monitoring
  - Data quality, TAT, equipment performance, deviation tracking

AoU Array Clinical Validation Plan

The study design compared known results: reference samples, previously tested clinical samples with known pathogenic variants, and matched pairs of DNA derived from blood and saliva from the same patient with PCR-free whole genome sequencing data for comparison.

Table 1. Validation Sample Sets

Sample Set	Detail
NA12878 input titration	0ng, 20ng, 40ng, 60ng, and 80ng gDNA input
Reference Samples - Inter-Run Repeatability	NA12878 (HapMap) and RM8392 trio (family of Ashkenazi Jewish ancestry), processed by 2 Operators across 3 batches
Reference Samples - Intra-Run Repeatability	4 replicates of NA12878, processed by a single operator
Clinical Samples	82 samples with previously-characterized variants in AoUMAP genes at locations covered by AoU array
Matched Blood and Saliva Pairs	6 pairs of DNA derived from blood and saliva from the same patient
Saliva-derived DNA input titration	0ng, 20ng, 40ng, 60ng, and 80ng saliva-derived gDNA input

Analytical validity of the array assessed the following:

Table 2. Summary of AoU Validation Plan

Analytical Test	To Assess:
AutoCall call rate	quality performance in array
Analytical Sensitivity (TP/TP+FN)	sensitivity for SNVs to NIST NA12878 over targeted sites in AoUMAP
Analytical Specificity (TN/TN+FP)	specificity for SNVs to NIST NA12878 over targeted sites in AoUMAP
Analytical Sensitivity: Concordance with Previously Identified SNV Variants in Clinical Samples	concordance with known pathogenic or likely pathogenic SNV calls in targeted sites in AoUMAP from previously characterized samples
Analytical Sensitivity: Concordance with Previously Identified InDel Variants in Clinical Samples	concordance with known pathogenic or likely pathogenic InDel calls in targeted sites in AoUMAP from previously characterized samples
Intra-run precision	repeatability of the same sample run under the same conditions; concordance in targeted sites in AoUMAP to NIST NA12878
Inter-run precision	reproducibility of the same sample run under varying conditions; concordance in targeted sites in AoUMAP to NIST NA12878 and RM 8392 trio
HapMap DNA Input Limit of Detection: DNA derived from blood	acceptable genomic inputs to assess SNVs and InDels to NIST NA12878 over targeted sites in AoUMAP
DNA Input Limit of Detection: DNA derived from saliva	acceptable genomic inputs from DNA Derived From Saliva
Sample Identity QC	concordance with orthogonal Fingerprinting SNV data from same stock sample

FN= false normal FP = false positive TN = true normal TP = true positive

AoU Array Clinical Validation Results

Array quality performance exceeded acceptance criteria, including in target sites in the AoUMAP (>99.8% AoUMAP average call rate).

Validation samples met passing thresholds across nearly all analytical tests. In the AoUMAP region, 28 of 28 SNVs were correctly called across 21 genes. 19 of 23 InDels were correctly called across 14 genes.

Table 3. Analytical Results of AoU Validation

Test	Acceptance criteria:	Metric Average	Percentage of samples passing test
AutoCall call rate	98% call rate threshold	*99.62%	*100%
Analytical Sensitivity (TP/TP+FN)	≥95% analytical sensitivity	99.77%	100%
Analytical Specificity (TN/TN+FP)	≥95% analytical sensitivity	98.91%	100%
Analytical Sensitivity: Concordance with Previously Identified SNV Variants in Clinical Samples	≥90% concordance		100%
Analytical Sensitivity: Concordance with Previously Identified InDel Variants in Clinical Samples	≥90% concordance		**86.4%
Intra-run precision	≥95% concordance across all replicates	99.86%	100%
Inter-run precision	≥95% concordance across all replicates	98.61%	100%
HapMap DNA Input Limit of Detection: DNA derived from blood	95% concordance	99.88%	*100%
DNA Input Limit of Detection: DNA derived from saliva	95% concordance: saliva vs blood from same individual	99.94%	*100%
Sample Identity QC	>0 Log Odds Ratio (LOD)	33.08	100%

Indicates test PASSED

*0ng input samples excluded as these were expected to fail
**InDel detection is a known limitation of arrays

Conclusion

The AoU array and associated processes are suitable for the intended use of generating high quality genotyping data for researchers and return of results applications. The validation has demonstrated accurate calling of targets sites across the AoUMAP region.

Based on the performance in this validation study, our Laboratory Director has approved the AoU array for clinical use.

We have established an end-to-end CAP/CLIA workflow with the capacity to meet AoURP scale, while running other concurrent, large-scale projects.

Figure 4. Image of Illumina array, or chip

Acknowledgments

We thank the All of Us working groups for input on the validation design and Kim Doheny of Johns Hopkins specifically for curation of the manifest content.

Data used in this poster was generated at the Broad Institute, for more information please visit: http://genomics.broadinstitute.org/