Array Analysis CLI (ACLI) Release Notes

v2.1.0

January 2023

Introduction

Array Analysis CLI (ACLI) provides accurate, comprehensive, and efficient local analysis of microarray data. The command-line interface makes it easy for power users to have granular control and flexibility to support large scale microarray genomic studies.

Highlights include:

ACCURATE PGX CNV CALLING

With a few simple commands, accurately determine hard to discern pharmacogenomic (PGx) genes like *CYP2D6*. The software is optimized to give accurate copy number variation (CNV) results for Infinium™ PGx catalog arrays.

COMPREHENSIVE OUTPUTS

Variant call file (VCF) outputs for both genotyping and CNV make further incorporation in multi-omic studies simple and easy.

EFFICIENT GENOTYPING ANALYSIS

Single nucleotide variant (SNV) genotyping analysis can be efficiently obtained from any Infinium microarray saving time for large experiments.

These Release Notes detail the key features and changes to software components for the release of Array Analysis CLI (ACLI) v2.1.0. For information on how to use the system, see the *Array Analysis CLI v2.1 Product Documentation*. ACLI is an efficient solution for array genotyping and PGx CNV analysis, including features such as:

- Convert raw iScan data (IDAT) to genotype calls (GTC)
- Genotype call files can be converted to bedgraph or VCF file formats
- Create copy number calls for key PGx genes
- Train copy number model for a set of samples

RELEASE v2.1.0 HIGHLIGHTS

- Improved PGx CNV calling accuracy for Infinium PGx catalog arrays.
- Ability to filter SNV VCF to a limited marker list.
- Implementation of multi-threading for parallel processing to obtain genotyping results faster.

NEW FEATURES IN DETAIL

- Genotype Calling
  - Added “num-threads” option for the “genotype call” subcommand. This new feature allows for parallel processing so genotyping results can be obtained faster.
- PGx CNV Calling
  - Improved CNV calling accuracy for the following key PGx genes:
    - CYP2D6, CYP2A6, CYP2E1, GSTM1, GSTT1 and UGT2B17
  - CNV calling support for the following products:
    - Infinium Global Diversity Array with enhanced PGx
    - Infinium Global Screening Array version 4 with enhanced PGx
    - Infinium Global Clinical Research Array with enhanced PGx
  - Improved CNV calling accuracy for saliva sample type.
  - Lowered CNV calling sample minimum to 24 pre-QC and 22 post-QC
    - 96 samples batches are still recommended for optimal CNV performance
  - Removed reference genome file absolute path display in CNV VCF
  - Added verification check for EGT and BPM against every GTC
- Copy number training
  - Supports CNV training for EX and LCG platforms through a platform flag
- VCF Conversion
  - New options for “expand-identifiers” and “unsquash-duplicates” were added to “genotype gtc-to-vcf” subcommand
    - “Expand-identifiers” allows all Infinium assay names to be entered in VCF ID field for variants with multiple assays on the array
    - “Unsquash-duplicates” generates unique VCF records for duplicate assays
- QC Improvements
  - TGA control is included in Genotype Summary File for easy review of the success of the Targeted Genome Amplification step for PGx.
  - The more common GenCall score is reported instead of Phred scaled GenCall score in the SNV VCF.
  - Column header text in Genotype Summary File is updated from “Gender Estimate” to “Sex Estimate”.

KNOWN ISSUES

- Corrupt or invalid GTC files will abort with an error instead of skipping. The corrupt or invalid GTC files will need to be removed before proceeding.
- In the gtc-to-vcf subcommand a mismatch between BPM and CSV manifests will not cause the command to abort with an error. The mismatch will need to be addressed before proceeding.
- For gtc-to-vcf, multi-allelic variants designed with multiple assays might not always collapse into one variant correctly and be reported as two separate variants instead. This is especially the case when one allele is an Insertion Deletion (indel) and the other is a Single Nucleotide Variation (SNV).
- Some indel variants missing from SNV VCF due to mapping issue between the designed indels and the reference genome.
- Manifest names greater than 80 characters will cause failure when converting IDATs to GTCs.
- CNV header indicates that up to copy number 4 may be reported; however, the CNV VCF does support reporting of copy number 5.

KNOWN LIMITATIONS

- Running multiple instances in parallel on macOS will result in the following error - 'The process cannot access the file because it is being used by another process.'

Open known issues from previous releases can be found in Array Analysis CLI Release Notes v.1.0.1