

MANUAL: snp5-geno-qc (SNP5-1.0)

Contents

- [Introduction.](#)
- [Quick Start - getting up and running.](#)
- [Support.](#)
- [The Report File - explanation of contents.](#)
- [Program Options - command line options.](#)

Introduction

snp5-geno-qc is a program for making statistical assessments of experimental qualities from Affymetrix SNP microarrays using the model based algorithm DM.

As DM is a single sample model based algorithm it processes one CEL file a time in a multiple CEL file batch. Since DM requires MM existence snp5-geno-qc only operates on SNPs tiled with mismatches, therefore, snp5-geno-qc works on all SNPs on Mapping500K, but only works on a small subset of SNPs on Python tiled for QC purposes.

Quick Start

We illustrate the most basic way to run snp5-geno-qc with an example. This example runs an analysis under the default parameter settings to generate a report file containing QC statistics for each CEL file with each specified method (or same method with different parameters). The requirements are:

- A collection of CEL files to process (one or many). For this example the CEL files specified by a CEL file list file with full pathes to each of the CEL files.
- The CDF file corresponding to the array type of the CELs. For this example the CDF file is called 'Mapping250K_Sty.cdf'. It can be downloaded from <http://www.affymetrix.com/support/technical/libraryfilesmain.affx>
- A file name with full path specified which you want the results to be written
- A snp5-geno-qc configuration file with methods and associated parameters specified

```
snp5-geno-qc \  
  --cdf-file Mapping250K_Sty.cdf \  
  --qca-file Mapping250K_Sty.qca \  
  --qcc-file Mapping250K_Sty.qcc \  
  --cel-files cel_file_list.txt \  
  --out-file results.txt
```

Support

Support for snp5-geno-qc is handled through the Affymetrix Developer Network. Specifically, questions, problems, feature requests, and other inquiries should be made through the Developer Network email address, devnet@affymetrix.com. snp5-geno-qc is not supported through the Affymetrix call center, Field Application Specialists, or the standard Affymetrix Technical support channels.

If you encounter an issue please make sure to collect the following information and report the problem to devnet@affymetrix.com

- Program output - cut and paste everything that the program reports to the screen.
- The specific command used.
- Type of machine (operating system, amount of memory).

The Report File

snp5-geno-qc creates a report file with file name specified by -o or --out-file option. The report file contains QC stats to measure the experimental quality for each CEL file. The format of the file is tab-delimited text with a header followed by a column header, then a line for each CEL file analyzed and a column for each method specified in the configuration file. The header contains information of which CDF file is used and what kind of methods are used for generating QC stats, and parameters associated with each method. The column entries are:

1. The CEL file name.
2. The stats for each method, variable number of columns equal to number of methods in the configuration file

Options:

The explanation of options is divided into two - the first section deals with options which deal with input, output and other options which all still yield the default estimates of geno QC stats. The second deals with options which alter aspects of running the program.

The following options do not affect the analysis or genotype calls produced.

```
Input:
-c, --cdf-file File defining probe sets. Can be downloaded
                from
www.affymetrix.com/support/technical/libraryfilesmain.affx
--cel-files Text file specifying full or relative path
            of CEL files to process, one per line with
"cel_files" as column header.
-a, --qca-file File defining the configurations for
                the QC analysis.
-q, --qcc-file File defining the SNP lists for the
```

QC analysis.

Output:

-o, --out-file File name with full path into which results
will be written
-r, --rpt-out Flag to indicate a rpt file should be written out
for each cel file (experimental)

Other:

-h, --help This help message.
-f, --force Force to not check chip type consistency

