

Analyzing NGS Data with NextGENe® Software AutoRun Tool

November 2016

Megan McCluskey, Kevin LeVan, John McGuigan, Shouyong Ni, CS Jonathan Liu

Introduction

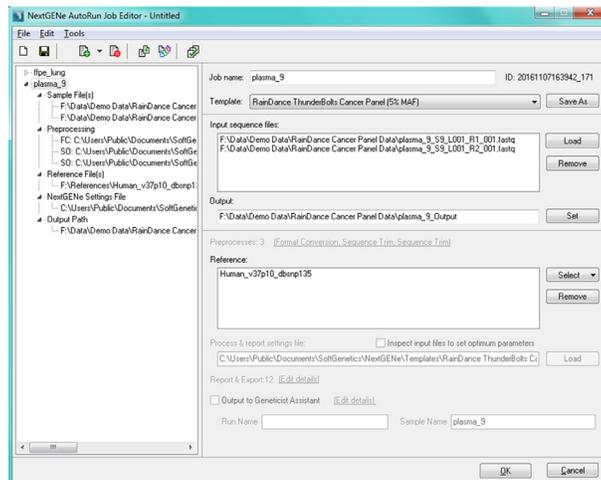
Next Generation Sequencing Technologies allow for the sequencing of numerous samples in short time frames. The use of barcoding or multiplexing techniques increases the number of samples that can be processed on each machine run. Because of this high output, setting up analysis parameters for the output samples can be time consuming and tedious.

NextGENe Software includes a specialized Pipeline Tool, NextGENe AutoRun, which provides enhanced, streamlined batch processing. This tool allows users to quickly set up multiple jobs to be analyzed consecutively. Multiple samples can then be analyzed unattended. Jobs can include processing of raw sample files from start to finish - any preprocessing steps such as removal of duplicates, adaptor or primer trimming, quality filtering, or barcode sorting, can be included as part of the job and post-processing, such as exporting reports or exporting to Geneticist Assistant can also be included in the job file. Templates can also be created to save all the combined settings to be loaded in a single step.

Procedure

Create Job File

1. Open NextGENe Auto Run Tool by going to Tools > NextGENe AutoRun.
2. When NextGENe Auto Run opens, go to Tools > Job Editor.
3. Optionally, to use an existing template, select the applicable template from the Template dropdown menu (skip Steps 6-10).
4. Under Input sequence files, click Load to load sample files. Multiple samples can be loaded at once. An individual job for each sample can be created automatically later (Step 11).
5. Specify directory and folder name for output files.
6. If any pre-processing steps are needed, click Add next to Preprocessing. If preprocessing steps are not to be included for the job, skip to Step 9.
7. Click Format Conversion, Barcode Sorting or Sequence Operation to load the appropriate pre-saved settings file. Repeat to load settings files for all required preprocessing steps.
8. Click OK to close the Preprocessing steps dialog.
9. Specify Reference File(s).
 - a. Click Select to select a preloaded reference OR
 - b. Click the arrow next to Select and choose Add Sequence files to load fasta or GenBank format reference files.
10. Specify analysis settings by loading a pre-saved *.ini file.
 - a. The analysis settings file can include post-processing options. In the Job Editor you can add, modify, or remove post-processing settings.
 - b. Click Edit Details next to Report and Export to make any changes to the post-processing settings.
 - c. To automatically output the NextGENe results for each sample to Geneticist Assistant, select Output to Geneticist Assistant, and enter a Run Name. Click Edit details to modify the global Geneticist Assistant settings if needed.
11. If files for multiple samples were loaded in Step 4, click Edit > Group Samples into Jobs.
 - a. You can group your sample files into individual jobs based on sections or characters in the file name, or simply based on the order of files loaded. When grouping based on sections or characters in the file name, the Job names and output folder names are created based on the defined section or characters.
 - b. Select to group sample files based on Sections, Fixed Positions or Order. The left side of the Group Jobs dialog shows the sample files loaded and the right side shows how the files will be grouped into jobs based on the current selections.
12. Optionally, if you wish to save the settings used for this job file to a template for later use, next to the Template dropdown menu, click Save and enter a name for your template.
13. When finished, click File > Save Job File to save the job file.



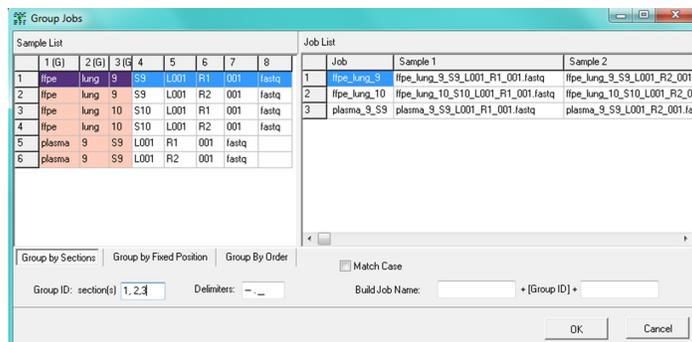


Figure 2: Automatically group sample files into individual jobs based on file names. Output folders can also be named according to the specified section(s) of the file names.

Processing Job Files

1. Click on the Settings icon () in NextGENe AutoRun to specify the folder where the job file(s) are saved under Job file detecting directory.
2. To begin processing the job file(s), click on the Detect icon () or go to File > Detect.

Discussion

NextGENe's AutoRun Tool allows users to quickly set up job files for batch processing by specifying sample files, reference files and settings for each analysis step. Frequently used settings can be saved to a template to facilitate streamlined creation of job files. NextGENe AutoRun scans the job files to check for the availability of all files needed for processing a job. If all files are available, NextGENe AutoRun launches NextGENe to begin processing the job queue, consecutively processing each job. If some files are not yet available, NextGENe AutoRun will periodically re-check for the files at the time interval specified in the NextGENe AutoRun settings. Additional job files can be created to automatically add jobs to the queue while NextGENe is currently processing another set of jobs. NextGENe will automatically begin the next list of jobs that are queued once the first list is complete.

NextGENe can be used to analyze next generation sequencing data from Illumina MiSeq, HiSeq and NextSeq systems, Roche/ 454 GS FLX, FLX Titanium and Junior, Thermo Fisher Ion PGM, Ion Proton, Ion S5 and S5 XL platforms. Multiple application types are available including SNP/Indel Analysis, RNA-Seq, de novo assembly, and CNV analysis. Results are displayed in the interactive NextGENe Viewer, providing a detailed visualization not found in other commercial programs like Lasergene's SeqMan Pro, CLC Bio & DNASTAR's NGEN or in open-source tools like TopHat, Bowtie & BWA.

Acknowledgement

We would like to thank GeneDx for their suggestions in the development of the NextGENe AutoRun Tool.