ABSOLUTE.summarize (v2) BETA

(/modules/docs/ABSOLUTE.summarize/2)

> This module is currently in beta release. The module and/or documentation may be incomplete.

Summarizes the results from multiple ABSOLUTE runs so an analyst can manually select solutions.

Author: Scott Carter, Matthew Meyerson, Gad Getz

Contact:

- For ABSOLUTE questions:
  - Use the Biostars forum <https://www.biostars.org/t/absolute/ (https://www.biostars.org/t/absolute/)>
  - Use the CGA discussion and help forum (http://www.broadinstitute.org/cancer/cga/cga_forums), especially for help with data interpretation
- The module is provided for academic non-commercial research purposes only. Other parties interested in using ABSOLUTE should contact the authors at <absolute-help@broadinstitute.org (mailto:absolute-help@broadinstitute.org?subject=ABSOLUTE%20question--%5Bbe%20specific%20here%5D)>.
- For GenePattern site questions, contact gp-help@broadinstitute.org (mailto:gp-help@broadinstitute.org?subject=ABSOLUTE)

Algorithm Version: ABSOLUTE 1.0.6

Summary

The ABSOLUTE.summarize module takes two or more RData files produced by ABSOLUTE (/modules/docs/ABSOLUTE) and summarizes them in a format that facilitates manual selection of alternative solutions. Because ABSOLUTE mathematically models putative solutions, often the highest scoring model is not the best solution. ABSOLUTE.summarize produces the following three results files, the first two of which are provided in turn to ABSOLUTE.review (http://www.broadinstitute.org/modules/docs/ABSOLUTE.review/2) for finalized solutions.

1. A tab-delimited plain text file with metrics for the top ranked solution for each sample. Genome doubling events are called in this file with 0 = no doublings, 1 = one doubling, and 2 = two doublings.
   - For use in ABSOLUTE.review, to override top ranked solutions, manually add a new first column to this file and input the solution number from the original ABSOLUTE rankings.
   - To continue with top ranked solutions for all the samples, provide the unmodified file directly to ABSOLUTE.review.

2. An Rdata file containing all of the proposed solutions previously plotted by ABSOLUTE.
3. Optionally for reference, a PDF file recapitulating the four types of plots generated by ABSOLUTE for the top three ranked solutions for each sample.

For descriptions of summarize output files and tips on selecting solutions, see the Analyzing ABSOLUTE Data (http://www.broadinstitute.org/analyzing-absolute-data) page. For background information on ABSOLUTE, example data and links to other resources, see the ABSOLUTE module documentation (http://www.broadinstitute.org/modules/docs/ABSOLUTE/).

References

Carter SL, Cibulskis K, Helman E, McKenna A, Shen H, Zack T, Laird PW, Onofrio RC, Winckler W, Weir BA, Beroukhim R, Pellman D, Levine DA, Lander ES, Meyerson M, Getz G. Absolute quantification of somatic DNA alterations in human cancer. *Nat Biotechnol*. 2012;30(5):413-21. (abstract and link to PDF (http://www.nature.com/nbt/journal/v30/n5/abs/nbt.2203.html))

Parameters

Name	Description
collection name *	A descriptive name for this collection of samples used for display.
absolute files *	A group of ABSOLUTE RData output files from individual sample runs to be summarized together. • Leave the batch parameter unchecked.
copy number type *	The copy number type to assess. This should match the parameter used in ABSOLUTE. • allelic (default) • total
plot modes *	Set this to FALSE to disable plotting of the purity/ploidy modes. Default: TRUE

* - required

Input Files

At least two ABSOLUTE sample results are needed to run ABSOLUTE.summarize.

1. Two or more **RData files** produced by ABSOLUTE module runs.
   Each file represents an individual sample. Samples provided together constitute a collection.

Output Files

1. <collection.name>.PP-calls_tab.txt

- A tab-delimited table detailing the top ranked called results for each sample. Reported information includes array, sample, call status, purity, ploidy, Genome doublings, delta, Coverage for 80% power, Cancer DNA fraction, Subclonal genome fraction, tau, and E_CR. For genome doubling events, 0 denotes no doublings, 1 denotes one doubling, and 2 denotes two doublings.

2. <collection.name>.PP-modes.data.RData
    - A saved object named segobj.list contains all information used to generate the other output files. Also known as the modes file.
3. <collection.name>.PP-modes.plots.pdf
    - If *plot modes* is set to TRUE, this optional file is produced. The file recapitulates the four types of plots generated by ABSOLUTE for the top three ranked solutions for each sample.
4. *.ABSOLUTE_UNCALLED_PLOT.pdf
    - A plot for every uncalled result.

Requirements

The module runs only on GenePattern 3.4.2 or above and requires R2.15 with the following packages, each of which will automatically download and install when the module is installed:

- numDeriv_2012.9-1
- getopt_1.17
- optparse_0.9.5

Please install R2.15.3 instead of R2.15.2 before installing the module. The GenePattern team has confirmed test data reproducibility for this module using R2.15.3 compared to R2.15.2 and can only provide limited support for other versions. The GenePattern team recommends R2.15.3, which fixes significant bugs in R2.15.2, and which must be installed and configured independently as discussed in *Using Different Versions of R* (http://www.broadinstitute.org/cancer/software/genepattern/administrators-guide#using-different-versions-of-r) and *Using the R Installer Plug-in (http://www.broadinstitute.org/cancer/software/genepattern/administrators-guide#using-the-r-installer-plug-in)*. These sections also provide information on patch level fixes that are necessary when additional installations of R are made and considerations for those who use R outside of GenePattern.

Platform Dependencies

Task Type:
SNP Analysis

CPU Type:
any

Operating System:
any

Language:
R2.15

Version Comments

Version	Release Date	Description
1.4	2015-10-13	Updated to make use of the R package installer.
1	2013-06-30	Initial version.