ABSOLUTE.review (v2)

BETA (/modules/docs/ABSOLUTE.review/2)

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

Finalizes solutions of absolute copy number and multiplicity for a collection of tumor samples. Used after ABSOLUTE and ABSOLUTE.summarize.

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.review module finalizes selected solutions originally modeled by ABSOLUTE (http://www.broadinstitute.org/modules/docs/ABSOLUTE). It takes the original or a modified version of the calls output file (*.PP-calls_tab.txt) and the modes file (*.PP-modes.data.RData) from ABSOLUTE.summarize (http://www.broadinstitute.org/modules/docs/ABSOLUTE.summarize/2) and provides finalized results for each sample in the collection.

You **manually review** solutions from ABSOLUTE results and annotate the *calls table file* from ABSOLUTE.summarize with the desired solutions. To continue with top ranked solutions for all the samples, provide the unmodified file directly to ABSOLUTE.review.

- Start the annotation by inserting a new first column into the *.PP-calls_tab.txt file from the ABSOLUTE.summarize module.

- For each sample row, type the number corresponding to the ranking of the desired solution into the first column. Rankings correspond to the order in which solutions are provided in the original ABSOLUTE analysis result <sample.name>.ABSOLUTE_plot.pdf file.
    - To keep the default solution, leave the cell blank or enter 1.
- If you use a spreadsheet program, e.g. Excel, be sure to save the update as a tab-delimited plain text file.

For 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
reviewed pp calls file *	The annotated calls file either directly or modified from an ABSOLUTE.summarize module run.
analyst id *	A display name of the analyst who selected alternate solutions for the calls file for record keeping.
modes file *	The <collection.name>.PP-modes.data.RData file produced by ABSOLUTE.summarize. Contains data for all the plotted solutions of the ABSOLUTE analysis.
collection name *	A descriptive name for this collection of samples. Can vary from the collection name used in the ABSOLUTE.summarize module run.
copy number type *	The copy number type to assess. This should match the parameter used in ABSOLUTE. • allelic (default) • total

* - required

Input Files

1. A single **calls table file** in tab-delimited plain text format, either directly from ABSOLUTE.summarize or modified.
    - This is the tab-delimited plain text file named <collection.name>.PP-calls_tab.txt from ABSOLUTE.summarize that lists the top ranking solution for each sample.

- To override selecting the top ranking solution, insert a new first column into the calls file, and for each sample row, type the number corresponding to the ranking of the desired solution from the original ABSOLUTE analysis result <sample.name>.ABSOLUTE_plot.pdf file.

2. A corresponding **modes file** in RData format from ABSOLUTE.summarize called <collection.name>.PP-modes.data.RData

Output Files

The collection name corresponds to that supplied to ABSOLUTE.review. The <sample.name> field corresponds to that supplied in the original ABSOLUTE module run.

1. <collection.name>.<analyst.id>.ABSOLUTE.table.txt
    - A finalized version of the *calls table file* reflecting the selected solutions.
2. <collection.name>.called.ABSOLUTE.plots.pdf
    - A PDF file with 3–4 types of plots for each of the selected solutions as described in ABSOLUTE module documentation. The fourth type of plot is given only for samples with mutation data.
3. For each sample, <sample.name>.ABSOLUTE.<analyst.id>.called.RData
    - A finalized RData version of the ABSOLUTE output for the selected solution.
4. A segmentation file <sample.name>.segtab.txt containing additional annotations from the original input. For the example dataset, these columns are as follows.
    - sample, Chromosome, Start.bp, End.bp, n_probes, length, seg_sigma, W, copy_ratio, modal_cn, expected_cn, subclonal, cancer_cell_frac, ccf_ci95_low, ccf_ci95_high, and hz.
5. If supplied mutation data, then a modified <sample.name>_ABS_MAF.txt containing additional annotations from the original input. For the example dataset, these columns are as follows.
    - mut_seg_ix, q_hat, HS_q_hat_1, HS_q_hat_2, Pr_somatic_clonal, Pr_germline, Pr_subclonal, Pr_subclonal_wt0, Pr_wt0, Pr_ge2, Pr_cryptic_SCNA, LL, modal_q_s, Pr_somatic, subclonal.ix, subclonal_wt0.ix, clonal.ix, wt0.ix, clonal_het.ix, ge2.ix, homozygous.ix, cell_mult, old_cancer_cell_frac, cancer_cell_frac, ccf_CI95_low, ccf_CI95_high, subclonal_SCNA, purity, and CCF_# columns.

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.