

G-test: comparison of groups of more SAGE libraries. The loglikelihood-ratio test (dubbed G-test in agreement with Sokal and Rohlf, Biometry) that we propose to use for the comparison of SAGE libraries originates from the analysis of frequencies. It offers the overall test (H_0 : all libraries are equal) as well as a straightforward way to continue when the H_0 is rejected and the tag proportions in the libraries are not all equal. The latter procedure can be adapted to fit the design of a particular SAGE experiment. In the paper (Schaaf, Ruijter et al., 2005) we demonstrate the use of the G-statistic in a comparison of subsets of libraries. In the comparison of subsets of libraries that is demonstrated in the paper, the Gwithin and Gpooled statistics provide insights into the variation within and between groups of libraries and are intrinsic to the data set used. Although the application described in the paper only compares one standard and one test set of libraries, it should be emphasized that the G-test procedure allows for the definition of multiple test sets. The G-test procedure assumes the sampling or within-library variation to be binomially distributed. The use of the G-test on contingency tables of libraries and specific and non-specific tags counts, can be envisioned as an extension of the use of the Chi-squared test for the comparison of two SAGE libraries. The latter test was shown to lead to similar conclusions as other test proposed for that purpose (Ruijter et al. 2002) and to have the best power for detecting differences when the number of specific tags are small. For further details on the calculation and testing of G-statistics: see the Appendix of Schaaf, Ruijter et al (2005).

FAQs

- **I tried to import a data into G-test: the error message was "invalid variant type conversion".**

This error message indicates that the program is trying to convert a text that is not a number into a number. This most often happens because: 1: you gave the wrong input range or 2: the decimal separator in Excel is different from the one in Windows. G-test works with the decimal separator in Windows, so you have to set Excel to use the system separator (in Excel: options menu - international Tab in the options dialog).

- **Q.**
A.

References

- [Schaaf, Ruijter et al.](#) Full transcriptome analysis of rhabdomyosarcoma, normal, and fetal skeletal muscle: statistical comparison of multiple SAGE libraries. FASEB J. 2005;19(3):404-6.
- [Schaaf et al.](#) Statistical comparison of two or more SAGE libraries: one tag at a time. Methods Mol Biol. 2008;387:151-68.
- [Ruijter et al.](#) Statistical evaluation of SAGE libraries: consequences for experimental design. Physiol Genomics. 2002:37-44.

Version History

G_test version 2.1, March 2012

- Updated version number in About box.
- Checked and extended maximum library size that could be handled.

G_test version 2.0, January 2012

- Changed data processing to calculations per block of tags.
- Number of tags per block will depend on available computer memory.
- This version of G-test should be able to handle any number of libraries and tags.

G_test version 1.26, January 2012

- Changed data exchange with Excel to prevent "call was rejected by receiver" error.
- This version of G-test can handle up to 70,000 tags.

G_test version 1.25, May 2011

- Improved handling of decimal separator.

G_test version 1.24, January 2011

- Added check for updated version (uses Windows Registry).

G_test version 1.23, September 2004

- First release with paper in FASEB Journal.