The Journal of Open Source Software

qqman: an R package for visualizing GWAS results using Q-Q and manhattan plots

Stephen D. Turner[1, 2]

1 Department of Public Health Sciences, University of Virginia School of Medicine, Charlottesville Virginia 22908 **2** Bioinformatics Core, University of Virginia School of Medicine, Charlottesville Virginia 22908

DOI: 10.21105/joss.00731

Software

- Review
- Repository
- Archive

Submitted: 25 April 2018
Published: 19 May 2018

Summary

Genome-wide association studies (GWAS) have been successful in identifying thousands of trait and disease-associated single nucleotide polymorphisms (SNPs). The primary result of a GWAS analysis is a list of SNPs, their associated chromosomal position, and a P-value representing the statistical significance of the association. A commonly used method used to visualize GWAS results is the "manhattan plot" – a plot of the $-log_{10}(P)$ of the association statistic on the y-axis versus the chromosomal position of the SNP on the x-axis. Another commonly used results diagnostic plot is the quantile-quantile ("Q-Q") plot. Q-Q plots display the observed association P-value for all SNPs on the y-axis versus the expected uniform distribution of P-values under the null hypothesis of no association on the x-axis.

One of the most commonly used software packages for manipulating and analyzing GWAS data is PLINK (Purcell et al. (2007)). `qqman` is an R package that allows for quick and flexible generation of publication-ready Q-Q and manhattan plots directly from PLINK results files. The `qqman` package is a user-friendly tool to visualize results from GWAS experiments using Q-Q and manhattan plots. The `manhattan()` function in the `qqman` package takes a data frame with columns containing the chromosome number, chromosomal position, P-value, and optionally the SNP name. By default, `manhattan()` looks for column names corresponding to those output by the `plink --assoc` command, namely, "CHR," "BP," "P," and "SNP," although different column names can be specified by the user. Thresholds for suggestive and genome-wide significance are drawn, and users also have the ability to highlight/annotate SNPs of interest. Finally, the `qq()` function can be used to generate a Q-Q plot to visualize the distribution of association P-values. An example of the plots produced by `qqman` is shown in Figure 1.

These graphics can be created in other software, such as the standalone desktop software Haploview (Barrett et al. (2004)), or for focused regions using the web-based application LocusZoom (Pruim et al. (2010)). Conversely, `qqman` is distributed as an R package with no other dependencies that can be easily integrated into existing R-based scripted workflows to further enable automated reproducible research. Furthermore, users can take advantage of R's very granular control of graphical output, enabling a high degree of customizability in creating high-resolution, publication-ready figures. The `qqman` package ships with example data and a detailed vignette illustrating its usage and further features not described here. The package is available on GitHub under the GNU General Public License at https://github.com/stephenturner/qqman and on the Comprehensive R Archive Network (CRAN) at https://cran.r-project.org/package=qqman.

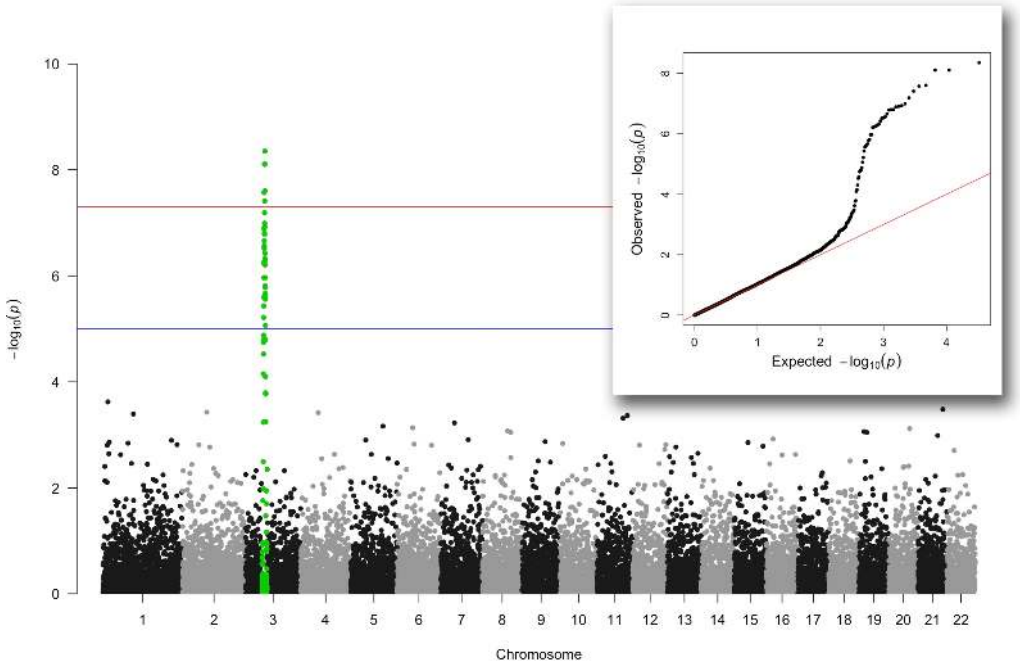

Figure 1: Manhattan plot highlighting SNPs of interest on chromosome 3, with Q-Q plot showing substantial deviation from the diagonal (inset).

References

Barrett, Jeffrey C, B Fry, JDMJ Maller, and Mark J Daly. 2004. "Haploview: Analysis and Visualization of Ld and Haplotype Maps." *Bioinformatics* 21 (2). Oxford University Press:263–65. https://doi.org/10.1093/bioinformatics/bth457.

Pruim, Randall J, Ryan P Welch, Serena Sanna, Tanya M Teslovich, Peter S Chines, Terry P Gliedt, Michael Boehnke, Gonçalo R Abecasis, and Cristen J Willer. 2010. "LocusZoom: Regional Visualization of Genome-Wide Association Scan Results." *Bioinformatics* 26 (18). Oxford University Press:2336–7. https://doi.org/10.1093/bioinformatics/btq419.

Purcell, Shaun, Benjamin Neale, Kathe Todd-Brown, Lori Thomas, Manuel AR Ferreira, David Bender, Julian Maller, et al. 2007. "PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses." *The American Journal of Human Genetics* 81 (3). Elsevier:559–75. https://doi.org/10.1086/519795.