# Evoker: a visualization tool for genotype intensity data

James A. Morris [1], Joshua C. Randall [2],
Julian B. Maller [2] and Jeffrey C. Barrett [1],
[1]Wellcome Trust Sanger Institute,
Hinxton, Cambridge, CB10 1HH, UK
[2]Wellcome Trust Centre for Human Genetics,
University of Oxford, Roosevelt Drive, Oxford, OX3 7BN, UK

Email: jm20@sanger.ac.uk
Web site: `http://www.sanger.ac.uk/resources/software/evoker`
Source code URL: `http://sourceforge.net/projects/evoker/develop`
License: MIT

April 13, 2010

Genome-wide association studies (GWAS) are now widely used in complex human disease genetics, these approaches produce huge volumes of data, which has created a real need for user friendly tools for data quality control and analysis of GWAS datasets. One critical aspect of quality control in a GWAS is evaluating genotype cluster plots to verify sensible genotype calling in putatively associated SNPs. The normalized intensity files from which genotype cluster plots are generated are extremely large and unwieldy in the default formats from SNP chip providers (uncompressed text-format intensities from a GWAS of 10,000 individuals would be hundreds of gigabytes). Extracting subsets of data and plotting hundreds of SNPs of interest is typically a tedious procedure requiring some computational sophistication. Therefore, we have developed Evoker, a Java program which supports two simple and compact binary data formats and is designed to make genotype cluster plot inspection a highly efficient process.

Important features of the Evoker program include remote connection to data sources, which means users do not need to have local copies of the large genotype and intensity files saving greatly on local disk space. The program only needs to transfer the data for the SNPs of interest, so Evoker is able to remain responsive even when dealing with very large datasets. Users are able to load lists of SNPs of interest, such as those showing evidence of association. The user can then quickly view, asses and make a decision on the quality of the cluster plot for each SNP, with the decision recorded in a separate file. Evoker can also be used to visualize the impact that excluding samples (such as those with borderline QC results) has on structure of the clusters.

The main Evoker program has been written in Java and so will work on any platform with Java 1.5 or later installed. The Evoker software also includes easy to use scripts for the generation of binary format files.

Evoker is an easy to use tool for visualizing genotype cluster plots, and provides a solution to the computational and storage problems related to working with such large datasets.