GWAS Activity

BST 281 (Spring 2019)

Part 0: Background

The website *easyGWAS* (https://easygwas.ethz.ch/) provides tutorials on performing GWAS analyses along with public data and results from existing GWAS studies. *easyGWAS* was published as an application note here:

Grimm DG, Roqueiro D, Salomé PA, Kleeberger S, Greshake B, Zhu W, Liu C, Lippert C, Stegle O, Schölkopf B, Weigel D. easyGWAS: a cloud-based platform for comparing the results of genome-wide association studies. The Plant Cell. 2017 Jan 1;29(1):5-19. [https://doi.org/10.1105/tpc.16.00551]

The examples on the website are focused on the model plant Arabidopsis thaliana, but the methods they employ generalize to other systems. You can read more about Arabidopsis here:

https://en.wikipedia.org/wiki/Arabidopsis_thaliana

- *What properties of the arabidopsis genome make it a useful model system for exploring genetic association studies?*

We will now explore three different contributions of the *easyGWAS* website. Note that all of these sections are further detailed/explained in the website's own "Tutorials" section linked off of the main landing page.

Part 1: Data Exploration

The first useful contribution of easyGWAS is its role as a public data repository. Click on the "Public Data Repository" button on the landing page to explore this portion of the site.

- *What other species (beyond arabidopsis) have data represented in easyGWAS?*

Select a species and dataset to review its properties. You can download the raw data underlying the dataset using the "Download Manager" link at the right.

- *What are the formats of the raw data files you downloaded? What biological information is represented within these files?*

Part 2: Study Exploration

The second useful contribution of *easyGWAS* is the ability to review previous GWAS results in a common, graphical framework. Select "Public GWAS" from the main menu, then select a study

to explore. Each study is divided into “experiments,” which are GWAS analyses of individual phenotypes measured for the model system. Pick a study and experiment to explore (for the next part of this section, it will be useful to pick an experiment/phenotype whose meaning is clear, or which you can clarify online).

- *What visualization technique is used to display the results of an individual experiment?*
- *Why are there multiple copies of this visualization shown?*
- *Do any SNPs pass the Bonferroni threshold for association with your phenotype?*

Click on an individual SNP to view its detailed analysis results (ideally one that was significantly or maximally associated with your phenotype).

- *What were the genotypes for this SNP?*
- *Did the SNP fall within a protein-coding region? If not, can you identify the nearest protein-coding region?*
- *Does the function of the protein-coding gene identified above relate to the phenotype under investigation? (You will likely need to perform an online search for the name/code for the gene to learn more about it.)*

Part 3: Performing a GWAS analysis

The final useful contribution of *easyGWAS* is its ability to perform new GWAS analyses using existing genotype and phenotype data (of the sort explored in the first section). Note that you need to create a (free) *easyGWAS* account to make use of this section. If you do not wish to create an account, you can work with a classmate or skip this section.

Follow the link under “Perform a GWAS” on the main landing page to begin. Next, choose a model organism, a relevant dataset (genotype + phenotype information), and a set of gene annotations for the genome of your model organism.

- *Are gene annotations required to perform a GWAS analysis? What role do they have in the analysis you are performing?*

Click “Continue” to select a phenotype. The text box here will autocomplete options as you type. As above, pick a phenotype that is easy to understand to facilitate interpretation of associated SNPs and their nearby genes.

The next page will show you a distribution of your phenotype of interest, provides the option to transform the phenotype, and tests for deviation from normality.

- *Why is having a normally distributed phenotype desirable?*

- *(Tricky) Why are many biological phenotypes normally distributed?*

Pick a transform for your phenotype of interest that produces the most normal-looking distribution (or the largest *P*-value in the test for deviation from normality).

Click through the remaining pages of the analysis setup. You will be given the option to control for one or more principal components (PCs) in the underlying SNP data.

- *What is a principal component?*
- *What is the purposes of controlling for principal component values in a GWAS?*

When given the option, initially select all available SNPs (then “Continue”). You can then filter SNPs based on their minor allele frequency. Pick a value of at least 1%.

- *What effect does increasing the minor allele frequency cutoff have on a GWAS?*

Stick with the default association method (EMMAX) and continue forward to review and submit your study. You can review the progress of the running study under the “My History” and “My Experiments” pages linked on the left-hand side of the page.

When your experiment completes, inspect the results. Follow the same sequence of steps we used in Part 2 when inspecting the results of a publicly available GWAS study.

- *Did you identify any SNPs that were significantly associated with your phenotype?*
- *Were nearby genes plausibly functionally related to your phenotype?*