

Information for zenodo project:

## Single-trait GWAS for reproductive fitness in *Drosophila melanogaster*

William P. Gilks, 28th June 2017, University of Sussex, UK, wpgilks@gmail.com

### Synopsis

The shell script, `code_drive_basic_gwas.sh`, downloads input data files from the internet, drives Plink to select LD-independent SNPs, and then perform a genome-wide association test against female and male fitness, separately. Plink is also used to assign functions and gene names to SNPs. Bash/Unix code is used for formatting/compatibility adjustments, and also to add NCBI-dbSNP IDs to results. The shell script starts an R script that calculates expected p-values and FDR values based on independent SNPs, plots diagnostic graphs, and outputs data to file. Explanations are provided in the code at each stage. A single R script for installing the packages used is provided. Remember to check software and package versions.

### File descriptions

File name	Description
<code>code_basic_gwas_postanalysis.R</code>	FDR calculations, graphing of results
<code>code_drive_basic_gwas.sh</code>	Drive GWAS in Plink1.9, & R post-analysis.
<code>code_install_packages_basic_gwas.R</code>	Required R packages
<code>data_fdr_basic_gwas.txt</code>	FDR q-values, independent & all SNPs.
<code>data_full_basic_gwas.txt</code>	GWAS results, all SNPs, with gene/function etc.
<code>data_results_summary_basic_gwas.txt</code>	Numerical summary.
<code>data_sig_gene_Female.txt</code>	Female significant genes
<code>data_sig_gene_Male.txt</code>	Male significant genes
<code>data_sig_snp_Female.txt</code>	Female significant SNPs
<code>data_sig_snp_Male.txt</code>	Male significant SNPs
<code>data_tagging_snps.txt</code>	List of the LD-independent SNPs
<code>data_top_snps_basic_gwas.txt</code>	Best results, each sex, all & independent SNPs
<code>gene_positions_plink.txt.mod</code>	Gene positions reference file
<code>log_drive_basic_gwas.txt</code>	Log for driver script
<code>log_md5_checksums_lhm_gwas.txt</code>	MD5 file IDs
<code>log_r_basic_gwas_postanalysis.txt</code>	R log
<code>log_sh_basic_gwas_postanalysis.txt</code>	R log
<code>plots_fdr_gwas_lhm.png</code>	p-values, q-values & cumulative counts
<code>plots_manhattan_qq_gwas_lhm.png</code>	Manhattan and QQ-plots of GWAS results

## Column headings

Column headings for the GWAS results files (e.g. data\_full\_basic\_gwas.txt). All the upper-case names are derived from the GWAS results file generated by Plink (more more info see website). Generation of additional information available at Zenodo community, Sussex Drosophila GWAS (see links).

Column name	Description
SNP	In-house unique ID
Phenotype	Which trait the result is for
CHR	Chromosome arm name, dm6 assembly
BP	Position on chromosome, in base-pairs
NMISS	Number of non-missing genotypes
BETA	Regression coefficient, aka effect-size
SE	Standard error
L95	Lower 95% credible interval
U95	Upper 95% CI
STAT	Coefficient t-statistic
P	p-value of the association, unadjusted.
gene	Gene names, UCSC Genome Browser RefGene track
snp_funct	SNP function, predicted by SNPEff
rs_id	NCBI-dbSNP unique ID
A1	'Allele 1'
A2	'Allele 2'
MAF	Minor allele frequency
NCHROBS	Number of chromosome pairs tested
tagging_snp	Is the SNP a tagging/independent SNP?

## Useful links

Name	URL
Morrow lab	<a href="http://www.sussex.ac.uk/lifesci/morrowlab/">http://www.sussex.ac.uk/lifesci/morrowlab/</a>
Sussex Drosophila GWAS	<a href="https://zenodo.org/communities/sussex_drosophila_gwas/">https://zenodo.org/communities/sussex_drosophila_gwas/</a>
Plink1	<a href="http://zzz.bwh.harvard.edu/plink/index.shtml">http://zzz.bwh.harvard.edu/plink/index.shtml</a>
Plink2	<a href="https://www.cog-genomics.org/plink2">https://www.cog-genomics.org/plink2</a>
R-project	<a href="https://www.r-project.org/">https://www.r-project.org/</a>
fdrtool	<a href="http://strimmerlab.org/software/fdrtool/">http://strimmerlab.org/software/fdrtool/</a>
Genotype data publication	<a href="https://f1000research.com/articles/5-2644/v3">https://f1000research.com/articles/5-2644/v3</a>