**Packages:**

<http://www.bios.unc.edu/research/genomic_software/Matrix_eQTL/>

Shabalin, A.A. Matrix eQTL: Ultra fast eQTL analysis via large matrix operations. *Bioinformatics* 28, no. 10 (2012): 1353-1358.

**Methods Descriptions:**

Fairfax, B. P., Makino, S., Radhakrishnan, J., Plant, K., Leslie, S., Dilthey, A., ... & Knight, J. C. (2012). Genetics of gene expression in primary immune cells identifies cell type–specific master regulators and roles of HLA alleles. *Nature genetics*, *44*(5), 502.

The quality control filtering for the genotyping data and association analysis was performed using PLINK[63](https://www.nature.com/articles/ng.2205#ref63) and Haploview[64](https://www.nature.com/articles/ng.2205#ref64), with imputation performed using Impute2 (ref. [65](https://www.nature.com/articles/ng.2205#ref65)). For *cis* associations, permutation analysis (*n* = 1,000) was performed by switching the phenotype labels. This number corresponds to an approximation of the number of SNPs tested per probe. The distribution of minimum *P* values in each permutation was used to identify significance thresholds. For *cis*associations, a permutation *P* value of 0.001 was used as the significance threshold for both monocytes and B cells. All significance values presented are based on the linear model unless otherwise stated. In addition to linear analysis, we performed Spearman rank analysis for all *cis* associations and eQTLs not observed to be additionally significant to permutation *P* value < 0.001 using this analysis were excluded. For *trans* associations, Wald tests were used to identify the genome-wide associations, and significance thresholds were determined by Bonferroni correction of commonly accepted significance levels. To ensure *trans* associations were only called if robust expression was detected, associations with maximum normalized expression of <6 were excluded. Due to power considerations, only *cis* and *trans* associations to SNPs and genes on autosomes are reported. Statistics were analyzed using R and appropriate packages. Graphs were generated using ggplot2 (ref. [66](https://www.nature.com/articles/ng.2205#ref66)), and local association plots were generated with Locus Zoom[67](https://www.nature.com/articles/ng.2205#ref67). HLA allele imputation was performed using HLA\*IMP[36](https://www.nature.com/articles/ng.2205#ref36),[37](https://www.nature.com/articles/ng.2205#ref37).

Christie, N., Myburg, A. A., Joubert, F., Murray, S. L., Carstens, M., Lin, Y. C., ... & Wighard, S. S. (2017). Systems genetics reveals a transcriptional network associated with susceptibility in the maize–grey leaf spot pathosystem. *The Plant Journal*, *89*(4), 746-763.

An eQTL data analysis pipeline (Figure S5, Methods S1) was

developed in Python (http://www.python.org) to analyse microarray-

based gene expression profiles for 19 281 reporters in the

maize leaf samples across 100 RILs of the CML444 9 SC Malawi

population. It was implemented in the online data analysis platform

Galaxy (http://galaxyproject.org) (Giardine et al., 2005),

which provides a user-friendly web-based interface for commandline

tools. The pipeline, available as three consecutive workflows

via the Toolshed (http://toolshed.g2.bx.psu.edu), does the following:

(i) determines the likelihood ratio (LR) threshold that corrects

for genome-wide markers using permutation tests (Churchill and

Doerge, 1994) via QTL Cartographer’s Zmapqtl module (Basten

et al., 1994) and multiple e-traits (in this study, the 95th percentiles

for 105 randomly chosen e-traits were determined by 1000 permutations

each, and the average LR of the 95th percentiles was taken

as the ‘estimated experiment-wise threshold’ (conversion

LOD = 0.217 9 LR) (Table S16); (ii) maps eQTLs, using QTL Cartographer

(Basten et al., 1994) [with the following parameters:

walking speed of 2 cM; composite interval mapping; forward

regression and backward elimination (P-value = 0.1); estimated

experiment-wise LOD threshold = 2.8], as independent parallel

tasks using different nodes on a compute cluster; and (iii) classifies

eQTLs as cis or trans (in this study eQTLs closer than 6.25 cM

– half the average size of an eQTL – to their linked gene were

called cis-eQTLs) in order to identify significant trans-eQTL hotspots

(by first calculating the genome-wide frequency of eQTLs

and then normalizing for local gene density). R (R Core Team,

2014) was employed for statistical analysis and data visualisation

(mainly in workflow iii).

Zhang 2011: Integrating pathway analysis and genetics of gene expression for genome-wide association study of basal cell carcinoma

We estimated non-genetic contributions in gene expression measures using principal component analysis (Leek and Storey [2007](https://link.springer.com/article/10.1007/s00439-011-1107-5#CR31); Stegle et al. [2008](https://link.springer.com/article/10.1007/s00439-011-1107-5#CR48)). Top principal components were included in the eQTL regression model as covariates. The number of principal components used was chosen to maximize the number of genome-wide significant cis eQTLs. Association analysis was applied with the FASTASSOC option implemented in MERLIN (Abecasis et al. [2002](https://link.springer.com/article/10.1007/s00439-011-1107-5#CR1); Chen and Abecasis [2007](https://link.springer.com/article/10.1007/s00439-011-1107-5#CR11)).

Puig-Oliveras 2016 Expression-based GWAS identifies variants, gene interactions and key regulators affecting intramuscular fatty acid content and composition in porcine meat <https://www.nature.com/articles/srep31803#methods>

### Gene-expression association analysis

An eGWAS was also performed using the genotypes of BC1\_LD animals and the expression values from muscle. A mixed model was employed implemented on Qxpak 5.0[59](https://www.nature.com/articles/srep31803#ref59):

https://media.nature.com/full/nature-assets/srep/2016/160818/srep31803/images/srep31803-m1.gif

in which yijkl was the kth individual record, sex (two levels) and batch (five levels) were fixed effects, λk was a −1, 0, +1 indicator variable depending on the kth individual genotype for the lth SNP, al represents the additive effect associated with the lth SNP, uk is the infinitesimal genetic effect treated as random and distributed as N(**0**, **A**σu2), where **A**is the numerator of the pedigree-based relationship matrix and eijkl the residual.

The same mixed model was applied to perform the association analyses with the *ACSM5* (rs331702081) and *IGF2* (IGF2-intron3-G3072A) polymorphisms and the *ACSM5* and *IGF2* mRNA expressions, respectively.

To correct for multiple testing, the false discovery rate (FDR) was calculated with the q-value library of the R package setting the threshold at q-value ≤ 0.05[58](https://www.nature.com/articles/srep31803#ref58),[60](https://www.nature.com/articles/srep31803#ref60). Given that *cis*-acting eQTLs tend to have larger magnitudes of effect on gene expression than do *trans*-acting eQTLs[61](https://www.nature.com/articles/srep31803#ref61), and several SNPs are affected by strong LD, a more restrictive threshold to avoid false positives was applied at p-value ≤ 0.001 for those eGWAS showing a *cis*-eQTL (see Q-Q plot in [Supplementary Fig. S4](https://www.nature.com/articles/srep31803#s1)).

The SNPs identified were classified as *cis* when they were located within 1 Mb from the gene analysed and as *trans* when they were located elsewhere in the genome. The number of significant SNPs belonging to the same interval was considered among associated SNPs less than 10 Mb apart.

QTLs  were  mapped  using  a  linear  model  implemented  in  Matrix  eQTL71,  and   FDR  was   estimated   by   permutations   as   follows: … For   transcript   ratio   QTLs,   we   permuted   ratios   of   all   transcripts   of   randomly  selected  1000  genes  3000  times  and  calculated  a  genome-­‐wide  p-­‐value   limit  based  on  the  median  of  the  most  stringent  transcript  per  gene.  For  gene  and   repeat  eQTLs,  we  permuted  randomly  selected  1000  genes  and  500  repeats,  and   used  their  median  as  a  genome-­‐wide  p-­‐value  limit