

## File S1

### Mixed Model Permutation Test

When (cryptic) relatedness or population structure is present in a sample, then naïve permutation test that randomizes the phenotype values can result in inflated type-1 error (Churchill & Doerge, 2008). To address this concern we employ a permutation scheme that preserves an estimated phenotypic covariance structure as estimated using a mixed model. The idea, which is inspired by (Müller et al., 2011), is to apply a transformation to the phenotypes so that they become (approximately) independent, permute them, and then transform them back. We can show that under the mixed model assumptions, this transformation is the Cholesky decomposed inverse phenotypic covariance matrix, as estimated from using a mixed model. Hence, we transform the phenotypes as follows:

$$Y^* = \text{cholesky}(V^{-1})'Y,$$

where  $Y$  denotes the phenotype vector and the  $V$  the estimated phenotypic covariance matrix. Under the model,  $\text{Var}(Y^*) = I$ , which allows us to permute those values, and then apply the inverse transformation to obtain permuted phenotypes that preserve the estimated structure as follows:

$$Y_{perm} = \text{cholesky}(V)'Y^*_{perm}.$$

Interestingly, this approach is similar to the approach of (Aulchenko et al., 2007), where they permuted the residuals after regressing out the genomic BLUP. The difference is that we do not attempt to remove the effects of family and population structure (as inferred by a mixed models) but instead apply a transformation that preserves the (estimated) phenotypic covariance structure. Finally, in the context of mixed model association mapping it is possible to perform the permutation test very efficiently by applying this transformation to the genotypes as well. Then the least square estimate using these transformed quantities (phenotypes and genotypes) is (trivially) identical to the generalized least square estimate as obtained from EMMAX (Kang et al., 2010). For obtaining a 5% genome-wide significance threshold we performed 500 permutations and redid the genome-wide association using the EMMAX algorithm. This permutation test is implemented in the mixmogam software (Segura et al., 2012).

#### References

- Aulchenko, Y. S., de Koning, D.-J., and Haley, C., 2007 Genomewide Rapid Association Using Mixed Model and Regression: A Fast and Simple Method For Genomewide Pedigree-Based Quantitative Trait Loci Association Analysis *Genetics* **177**: 577--585
- Churchill, G. A. and Doerge, R. W., 2008 Naïve application of permutation testing leads to inflated type I error rates. *Genetics* **178**: 609--610
- Kang, H. M. M., Sul, J. H. H., Service, S. K., et al., 2010 Variance component model to account for sample structure in genome-wide association studies. *Nature genetics* **42**: 348--354
- Müller, B. U., Stich, B., and Piepho, H.-P. P., 2011 A general method for controlling the genome-wide type I error rate in linkage and association mapping experiments in plants. *Heredity* **106**: 825--831
- Segura, V., Vilhjalmsón, B. J., Platt, A., et al., 2012 An efficient multi-locus mixed-model approach for genome-wide association studies in structured populations. *Nature genetics* **44**: 825--830