Male pattern baldness summary statistics

10th December 2018

This file contains summary statistics for the male pattern baldness (MPB) genome-wide association study conducted in Yap *et al*. [1].

Data

We used individual-level data from the UK Biobank filtering for European males (n=205,327). We included 18,065,087 autosomal and 1,064,602 X-chromosome (1,024,430 non-PAR and 40,172 PAR1) SNPs.

The MPB phenotypes were adjusted for age, assessment centre, ethnicity and 40 principal components calculated using the UKB European population.

Full QC details are provided in Yap *et al.* [1].

Software

The analysis was conducted using BOLT-LMMv2.3 [2], (using the –allowX command) using LD-pruned autosomal SNPs to calibrate the analysis (--modelSNPs command). We used the 1000 Genomes LD scores file and hg19 genetic map file provided with the BOLT-LMM software.

Output

We provide summary statistics (tab-delimited) for a total of 19,123,863 SNPs. Each row corresponds to an individual SNP. The columns of the output are direct from the BOLT-LMMv2.3 software [2], and are as follows:

SNP:	rsID or positional SNP identifier
CHR:	chromosome
BP:	base pair position
GENPOS:	genetic position
ALLELE1:	effect allele
ALLELE0:	other allele
A1FREQ:	allele frequency of ALLELE1
F_MISS:	fraction of individuals with missing genotype at this SNP
BETA:	effect size (BOLT-LMM approximation to infinitesimal mixed model)
SE:	standard error of the effect size
P_BOLT_LMM_INF:	p-value from the infinitesimal mixed model association test
P_BOLT_LMM:	p-value from the non-infinitesimal mixed model association test

Please note that males are haploid for the X-chromosome. BOLT-LMM encodes males as 0/2, whereas the actual number of alleles is 0/1. Hence, the per-allele effect size is half the reported effect size in the summary statistics output.

References

[1] Yap CX, *et al.* (2018). Dissection of genetic variation and evidence for pleiotropy in male pattern baldness. *Nature Communications*. 2018 (accepted).
[2] Loh PR, *et al.* (2015). Efficient Bayesian mixed-model analysis increases association power in large cohorts. *Nature Genetics*. **47**, 284-290 (2015).

Please refer any queries to Peter Visscher (p.visscher@uq.edu.au).