



LD Clumping

Typically, genetic associations are clustered together in loci that consist of many associated variants in one genomic region. In many cases, a locus only includes one independently associated, or causal, variant, while the associations of the other variants are due to the fact that they are genetically linked to the causal variant.

This genetic linkage is expressed as linkage disequilibrium (LD), a measure of the extent of correlation between any two alleles. LD is usually expressed as r^2 , which is calculated using a formula that takes into account the frequency at which the alleles are found together on a single chromosome. An r^2 value of 1 indicates that the alleles are completely correlated—that is, they are always inherited together—while an r^2 value of 0 indicates that the alleles are in linkage equilibrium, inherited completely independently of each other.

LD clumping reports the most significant genetic associations in a region in terms of a smaller number of “clumps” of genetically linked SNPs. This can help the researcher to assess how many independent loci are associated with a given trait.

LD clumping in the Knowledge Portals is performed using [PLINK](#). The parameters are set as follows:

```
PLINK_P1 = 5e-8
PLINK_P2 = 1e-2
PLINK_R2 = 0.2
PLINK_KB = 250
```

If there are no associations for a trait (across the whole genome) that have genome-wide significance or better, the P1 parameter (the significance threshold p-value for index SNPs) is increased in 10-fold increments until there are at least 50 clumps for that trait. The value of P1 is never set to be higher than P2, the secondary significance threshold.

LD information for clumping is derived from the [1000 Genomes](#) project. Rare variants that are not represented in 1000 Genomes and do not fall within the boundaries of existing clumps are appended to the results as single-variant clumps.

We are happy to provide help in evaluating these results; please [contact us](#).