The script is taking the data from Phenoscanner (this includes proxies with an r2>06) based on the 77 variants from the paper by Locke and excludes all of the variants that are associated with BMI from the GIANT consortium and any other study.

***pleio() function***

* The pleio() function takes a subset of the data based on: a) whether to include proxies or not; b) a specified r2 value; and c) a p-value for the genetic associations. Note that when the data was extracted it was based on an r2>0.6 and the p-value>0.05.
* The pleio() function then loops round each of the 77 variants and considers each study in turn. It ranks the genetic association based on the year of publication and either the r2 value or the p-value.
* Each variant within a given study may have multiple entries as either they have proxy SNPs or the study has more than one trait, or both. To make sure we have a record of all of the traits the variant is associated with, all of the unique traits within the study are extracted and recorded under the column name ‘Traits’.
* The first entry (i.e. the variant from the most recent publication that either has the largest r2 value or p-value is then extracted. If the r2 value is used to rank the associations it is likely that the variant extracted will not be a proxy.
* Note that when we are interested in what traits the variants are associated at the GWS level with we probably want to rank the variants based on the r2 value. However, when we want to investigate whether the variant is associated with traits that may be risk factors of the outcome and/or confounders of the exposure-outcome association then we probably want to rank the associations based on the p-value as we don’t want to miss anything as this may invalidate the MR analysis.
* The resulting data frame then consists of the all the variants (either proxy or not) that are associated (at the specified significance level) with at least one of the traits for a given study. Where the variant is associated with multiple traits these are recorded in the Traits variable. Note that each of the 77 variants may have multiple rows for each study proxies of the variant may be associated with the trait.

***summary\_traits() function***

* This function creates a table with all of the unique values of var\_group and then provides a list of the factors in var\_list that are associated with each unique value of var\_group.
* Perhaps the most relevant combination would be var\_group=rsID and var\_list=Traits. Therefore, we obtain a table with all of the 77 variants with a list of all the traits the variants are associated with.
* Note we could consider this the other way round.
* This function is considered on all the non-adiposity traits when var\_group=rsID and Trait.
* var\_group=rsID is considered twice, once with non-adiposity traits and again with just the adiposity traits. The information is then combined and saved in ‘gws\_rsid\_combined’.

The above was first considered with p.value=GWS, and then at the nominal level. Note that since a large number of the traits will be associated with the variants, we have only considered whether known risk factors of asthma or potential confounders are associated with the variants. We loop through each trait from the relevant consortium and then record: a)consortium; b) no of variants associated with the trait; c) minimum p-value; and d) the average p-value; and e) a list of the rs numbers of the variants that are associated with the trait.