Collection of tips for Ricopili QC pipeline

Gio Pan. 16.11.2017

1. Filters for SNP MAF inclusion/exclusion should be symmetrical, e.g. to create a filter that only includes SNPs with MAF>.01 the proper condition is (MAF>0.01 AND MAF<0.99).
   Example: *awk '($5<0.02 || $5>0.98 || $13>0.01 || $14>0.01) NR>1 {print $1}' bip_XXX_eur_gp-qc.detres > bip_XXX_eur_gp-qc.snps.excl* would create an exclusion list of SNPs with MAF<.02

2. After PCA analysis, *menv.mds_cov file should be used with the PLINK --keep command . This files takes is already a deduplicated version of the group of individuals. So please don't use it with --remove the outliers, but rather --keep the non-outliers (otherwise the possible related partners are still in)

3. As such the *preimp_dir* module will take care of naming convention of the plink files and also standardized the IIds in the fam file.
   Standard naming convention of Ricopili
   Files - [dis]_[stdy1]_[pop]_[user_init]-qc.
   Fam IIDs- [phe]_[dis]_[stdy1]_[pop]_[user_init]_[geno-platform]*orginalFamIId
   You must standardize the files and IIds if you want to skip preimp module and want to go to the pcaer/imputation step directly. This can be done by using *id_tager_2* script (found in the ricopili distribution) as follows.
   Usually you add a short name for genotyping platform after user initials (ui) e.g.
   id_tager_2 --nn bip_stud1_eur_ui_A6.0 --cn bip_stud1_eur_ui --create plink.fam

   --nn : string; to rename the fam ids.
   --cn : string; to rename the plink files.

4. Usage of the *names file when starting preimp module: For bipolar disorder ("bip") and 2 studies, named "stud1" and "stud2", in the bip.names file the nicknames should simply be stud1 and stud2, not bip_stud1 and bip_stud2, because in the latter case the "bip" prefix will be duplicated.
5. When used on multiple cohorts, "pcaer" module will project all studies on the same space of all common alleles (common among all studies). For the QC

steps it might be more helpful to run PCA on each study separately, so as to have higher precision/resolution.