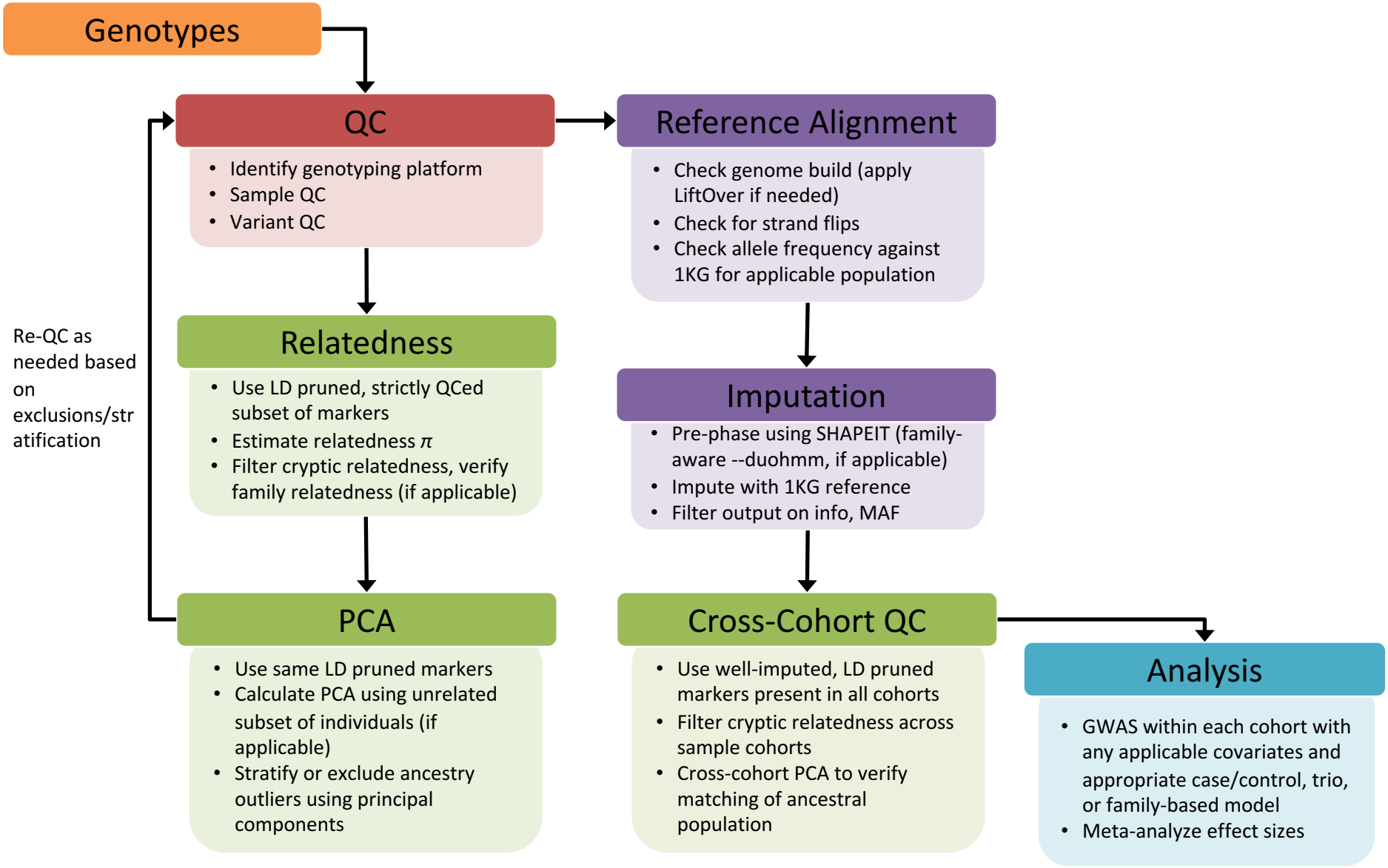
PGC Ricopili workflow



- Steps are color-coded according to applicable module of ricopili (see https://sites.google.com/a/broadinstitute.org/ricopili/). Additional details and access to the ricopili user group are available on that site. Code is also available from the github repository: https://github.com/Nealelab/ricopili/.

- Compared to standard GWAS meta-analysis, this pipeline using full genotype data (1) ensures uniform QC/imputation/analysis procedure for all cohorts, (2) standardizes build/imputation reference matching to simplify matching variants for meta-analysis, and (3) allows QC of cross-cohort relatedness and population ancestry matching.

- Additional details:
  Sample QC includes missingness rate, heterozygosity, and sex check. Variant QC includes missingness rate, differential case/control missingness, and Hardy-Weinberg equilibrium. For trio/family samples, mendelian error rates are also checked. Additional cohort-specific QC may also be applied (e.g. for site/batch effects).

- The “strictly QCed” subset used for PCA/relatedness is: MAF > 5%, HWE p > 1e-4, missingness < 2%, no strand ambiguous SNPs, no MHC region, no chr8 inversion region, and LD pruned to r2 < .2.