

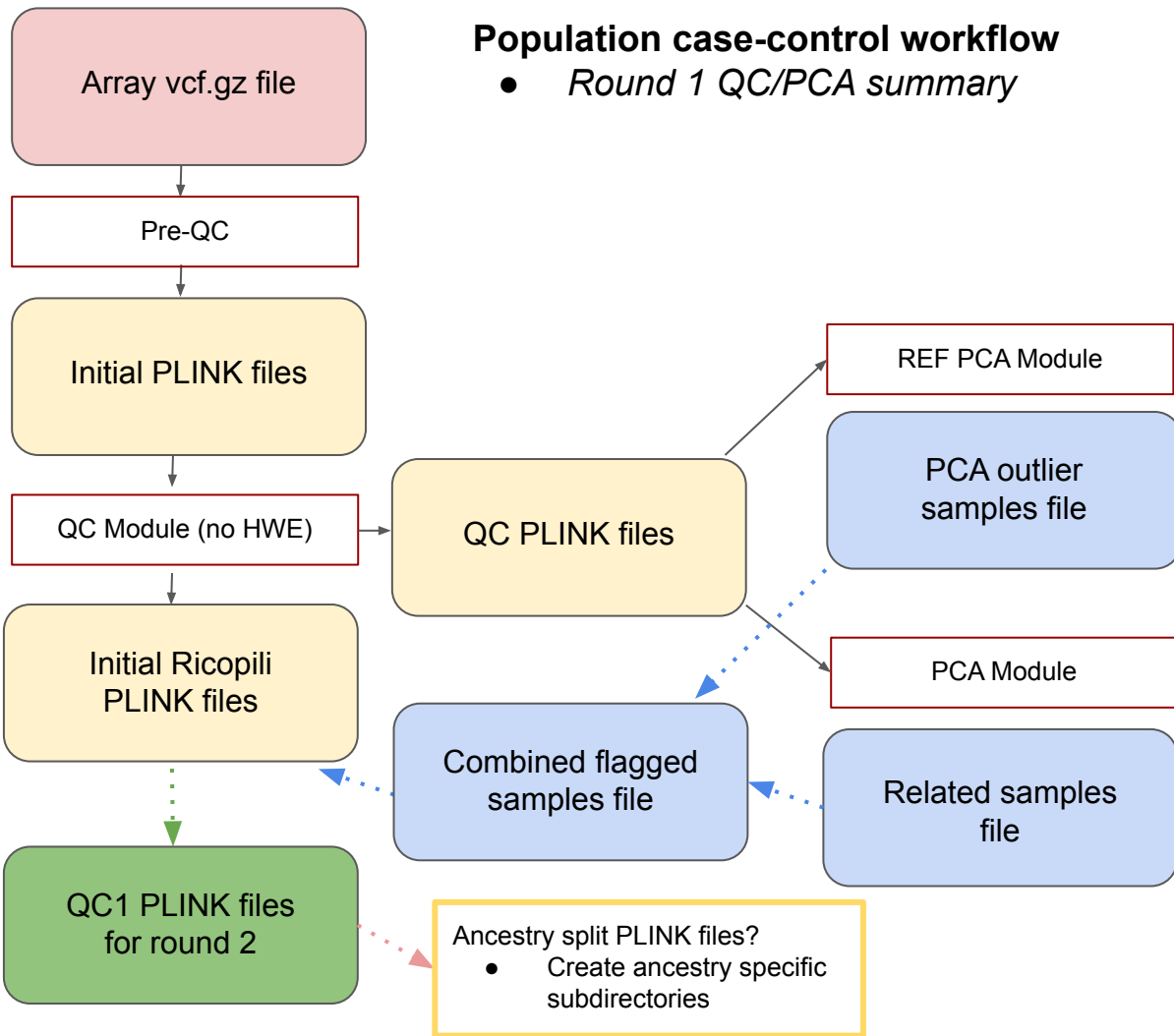
# Stage 1 Analysis

## QC workflow

Ricopili + helper R scripts  
2022

## Population case-control workflow

- *Round 1 QC/PCA summary*



### Pre-QC steps

- Convert to PLINK files
- Match VCF/phenotype file IDs
- Update .fam file
- Add reported sex / affection status
- Incorporate pedigree info (if provided)

### QC Module steps

- Create qc1 subdirectory
- Soft-link relevant PLINK files
- **Run preimp\_dir module (--hwe 1e-100)**
- Copy pdf report to readable directory

### REF PCA Module steps

- Within qc1/, create refpca/ subdirectory
- Soft-link relevant PLINK files
- Run pcaer module
- **Run PC\_filter\_casecon.R**
- Copy pdf report to readable directory

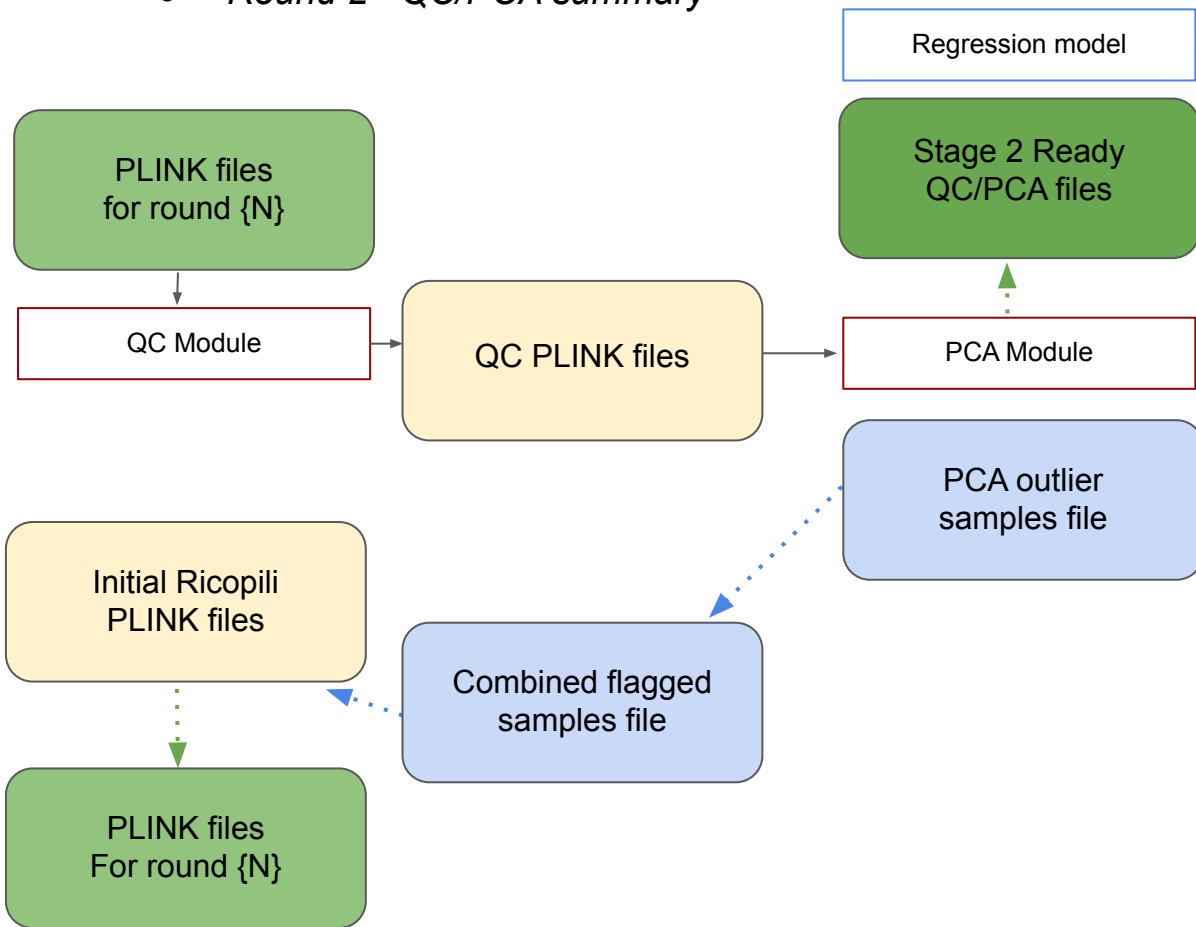
### PCA Module steps

- Within qc1/, create pca/ subdirectory
- Soft-link relevant PLINK files
- Run pcaer module
- **Run related\_samples\_filter.R**
- Copy pdf report to readable directory

- Create qc2/ subdir and write PLINK files

## Population Case-control workflow

- *Round 2+ QC/PCA summary*



### QC Module steps

- Create qc{N} subdirectory
- Soft-link relevant PLINK files
- Run preimp\_dir module
- Copy pdf report to readable directory

### PCA Module steps

- Within qc/, create pca/ subdirectory
- Soft-link relevant PLINK files
- Run pcaer module
- **Run PC\_filter\_casecon.R**
- Copy pdf report to readable directory

### Regression model

- Logistic regression model in PLINK
- Include 1st 10 PCs
- Examine QQ plot and median lambda

### Round 2+ considerations

- Stage 2 ready:
  - No more samples to remove from PCA
  - Logistic regression model w/PCs looks good
- Additional QC/PCA round
  - More samples to filter from PCA