Quality control using Plink:

Autosome and PAR1, PAR2 regions:

- SNP quality control
  SNPs with imputation R2<0.3 will be removed:
  zcat {input.file1}.info | awk -F '\t' '$7<=0.3 {print $0}' > { snpsR2.file2}
  plink2 --pfile {input.file1} –exclude {snpsR2.file2} --make-pgen --out {output}
  SNPs with MAF<0.01, genotyping rate and missingness rate below 0.02 will be removed:
  plink2 --pfile {input.file1} --maf 0.01 --geno 0.02 --mind 0.02 --make-pgen --out {output}
  Estimation of of Hardy-Weinberg equilibrium exact test statistics and remove of signf. SNPs:
  plink2 --pfile {input.file1} --hardy --out {output}
  cat {input.file1}.hardy | awk -F'\t' '{ if ($10<=5e-06){print $2, $3, $4, $10}}' > {output}
  plink2 --pfile {input.file1} --exclude {hwe.snps} --make-pgen --out {output}
- Sample quality control
  Estimation of independent SNPs
  plink2 --pfile {input.file1} --indep-pairwise 1500 150 0.2 --extract {input.file2} --out {output}
  plink2 --pfile {input.file1} --extract {input.file2}.prune.in --make-bed --out {output}
  Sex-check analysis:
  plink2 --bfile {input.file1} --check-sex --out {output}
  Relationship analysis:
  plink2 --bfile {input.file1} --remove {incorrect_sex.file2} --king-cutoff 0.025 --out {output} --make-king-table --make-bed
  Estimation of heterozygosity rate:
  plink --bfile {input.file1} --remove {related.file2} --het --out {output}
  PCA calculation on cleaned samples:
  plink2 --bfile {input.file1} --remove {heterozygous.file2} --pca 20 --make-bed --out {output}

X chromosome:

- SNP quality control
  SNPs with imputation R2<0.3 will be removed:
  plink2 --vcf {input.file1} --split-par 'b37' --extract-if-info "INFO >= 0.3" --make-pgen --out {output}
  Only SNPs in non-PAR region will be extracted for QC:
  plink2 --pfile {input.file1} --split-par 'b37' --chr X --make-pgen --update-sex {input.file2} --out {output}
  Estimation of Hardy-Weinberg equilibrium exact test statistics and remove of signf. (p<1e-04) SNPs:
  plink2 --pfile {input.file1} --make-pgen --keep {input.file2} --hardy --out {output}
  cat {input.file1}.hardy.x | awk -F'\t' '{ if ($14<=1e-04){print $2, $3, $4, $14}}' > {hwe.snps}

SNPs with MAF<0.01 in male and female separately:
plink2 --pfile {input.file1} --max-maf 0.01 --filter-females --exclude {hwe.snps} --make-pgen --out {output}
plink2 --pfile {input.file1} --max-maf 0.01 --filter-males --exclude {hwe.snps} --make-pgen --out {output}

Separate GWAS in ADNI and EMIF:

Extract dosage from imputed vcf files with autosome regions:
plink2 --vcf {input.file1} dosage=DS --keep {qc_samples.file2} --update-sex {sex.file3} --make-pgen --out {output}
Extract dosage from imputed vcf files with X-chromosome and PAR1, PAR2 regions
plink2 --vcf {input.file1} dosage=DS --split-par 'b37' --make-pgen --out {output}
Clean up QC SNPs and QC samples in all files:
plink2 --pfile {input.file1} --extract {qc_snps.file2} --keep {qc_samples.file3} --make-pgen --out {output}
Perform associations analysis separate in autosome and separate for X-chromosome and PAR1/PAR2 region:
plink2 --pfile {input.file1} --glm --no-psam-pheno --pheno {input.file2} --covar {input.file3} --covar-col-nums {covar_col_nums_PC_sex_MCI_ststus} --out {output}

This was followed by meta-analysis of ADNI and EMIF datasets using an inverse variance weighting model, like implemented in METAL software.