(안건: LDpred로 정녕 한번에 multiancestry PRS를 한번에 만들어도 되는가?

비슷하게 *multianestry*를 때려박아 *PRS*를 만든 *ref paper*가 없다면 어떻게 진행하면 좋을까*?*

- multipred / prs는 secondary analyses so by other methods provided by review paper, we could reduce the accuracy issue.

*GWAS summary stats*이 대부분 *EA*에서 비롯된다는걸 감안해서 *LD reference panel/*소프트웨어는 *LDpred*가 최선인가*?*

- LDpred / LDpred2 최선 .. but idk why finding correlations with chr takes so long, especially first several chr =2 etc

*multipred*방법*(Márquez-Luna et al 2017, genetic epidemiology)* 은 좋아보이는데 왜 상용화가 많이 안됐나.. 등*)*

- our LPpred risk prediction method which analyzing sumstats in conjunction with LD info from a reference panel is more accurate in European populations than the informed LD-pruning + P value thresholding approach applied here: we did not employ LDpred due to the complexities of admixture-LD in analyses of admixed populations that explicitly model LD, but extending LDpred to handle these complexities could further improve accuracy
- Adding an ancestry predictor only improves prediction when we use a stringent p-value threshold to build PRS
- did not incorporate data from x chromosome > heritability
- focused on common variants, no rare variants

Difficulty in analyzing genetic data from diverse populations / constructing individual-level scores from a cross-ancestry meta-analysis

Problem:

>GWAS study focuses only on EA
>Genetic risk prediction attenuates with an increasing divergence between the discovery and target populations
>polygenic risk scores (PRSs) based on Eurocentric GWASs are not equally predictive when applied to non-European populations

Methodological Considerations

while analyzing samples from multiple populations:

1) empirically assign samples to major continental and/or admixed populations using genome-wide data, analyze each population separately, and conduct cross- ancestry meta-analysis (stratified meta-analysis approach);

2) analyze samples from multiple populations together, most commonly with a mixed model (joint mixed-model approach).
    - analysis step: quality control to reference alignment in imputation, association model, and the suitability of results for secondary analyses.

Genotyping technology

1) population-specific arrays. Multi-ancestry arrays, such as the Multi-Ethnic Global Array (MEGA), Global Screening Array (GSA), and the H3Africa array
2) to sequence whole genomes; low-depth sequencing

Quality Control can be applied..

Applying standard QC without adjustment for population structure> erroneous removal of too many variants and samples from minority subgroups and admixed samples > reduce statistical power.

1) stratifying the cohort into major populations prior to filtering (the stratified meta-analysis approach)
2) adjusting the QC measure to allow for varying allele frequencies

Inferring Population Structure

- use for description and QC and later discuss methods for controlling for population structure
- common tool PCA - can be computed <u>within the cohort</u> or from an external reference

Imputation and Population Reference Panels

- Imputation accuracy depends on reference panels (better coverage of haplotypes from the population of the genotyped cohort > greater number of well-imputed variants for GWASs, especially among lower-frequency variants)
- Joint imputation - necessary to consider imputation quality separately within subsets of individuals even if the samples are jointly imputed, since imputation accuracy for a variant may vary widely across individuals of different ancestries.

Genome-wide association

- proper control of population stratification to ensure that observed associations reflect genetic effects of each locus rather than correlations with ancestry.
- increase power of joint analysis - increased sample size, control variance explained by the genetic relatedness between individuals
- whether to stratify samples into major population groups or to analyze the full cohort jointly (assuming imputation was also done jointly).

Meta-analysis of GWAS Summary Statistics

- due to differences in genetic background + environment > cross-cohort heterogeneity > difference in marginal effect size

- So we should use random effects or trans-ancestral meta-analysis model

Fine-Mapping > 저희는 하나요**? x**

- refines GWAS loci to a smaller set of likely causal variants to facilitate interpretation and follow-up studies
- Combining samples across ancestries has an advantage for fine-mapping:
    - for casual variants: assuming that many causal variants are shared across populations LD patterns that differ across populations can improve the resolution
    - Non-causal variants: if LD is different, tagging the causal variants have marginally different effects across populations> allowing the causal variant to be distinguished from non-causal variants. >improve the resolution of fine-mapping
- Problem: presumption that the causal variants and their effect sizes are identical across populations > solution: The Probabilistic Annotation INTegretOR (PAIN- TOR) method

PRS in a diverse population

- increasing genetic ‘‘distance’’ between the discovery and target datasets, there is often attenuation of polygenic predictive value.
- how to construct? Use of trans-ancestry meta-analytic results to weight alleles, MultiPred
- follow up research: what is multipred? combines PRSs based on European training data with PRSs based on training data from the target population
- 내가 생각한 것: 저는 prs 점수 형성 방법은 secondary analysis 라 볼 수 있고 이 외에도 데이터 분할, 분석 과정이 더 중요할 수 있겠다는 생각
- 질문: 저희가 최종적으로 넣을 데이터는 multi-ethnic? compatibility with combining? multipred 적용 가능?

Heritability and Genetic Correlation

- problem: difficulty in estimating SNP heritability and genetic correlation in multiancestry
- solution:
    - methods relying on relatedness estimation require estimation methods robust to population structure
    - methods modeling LD require either ancestry-matched reference panels or individual-level data for LD calculations. Ancestry-matched reference panels, along with the large GWAS sample sizes required for robust estimation using these methods, maybe especially challenging to acquire for studies in under-represented or admixed groups.
    - local ancestry tracts in admixed population samples can be leveraged to estimate heritability, genetic-effect, genetic-impact correlations of observed variants can be estimated using Popcorn if LD information is available and the two populations are relatively homogeneous.

-

Rare Variant Association Analysis

- problem: limited power to identify trait associations of individual rare variants
- solution: aggregation methods such as burden tests, variance-component tests, and hybrid tests have been developed to test the combined effect of several variants.
- 저희는 rare variant 도 보나요? no > but not that necessary

Non-genetic Contributors to Trait Variability

- measurement of environmental, social, cultural factors difficult
- Self-reported measures of diversity / GxE
- need better understanding and measurement of causal environmental risk factors

Areas in need of development

- More methodological development is needed before mixed models or other strategies for joint GWASs of a diverse cohort can be confidently recommended as robust.
- more method development in control inflation for collapsing and variance-component methods to better identify trait associations of rare variants

Response: consider three main domains: (1) researcher participation, (2) data resources, and (3) analytic methods > 에대해 우리가 self-reflection할 때 적용가능하다 생각했음

design model

european ancestry

- gps + genotype
- european -> african (lassosum)
- UK biobank - ldpred - gps - ukb