prs 를 활용한 논문들을 읽으면서, prs를 활용하는 방법, 예측방법, 한계에 대해 생각해보고, 우리의 input variable, output variable, predictor method, code architecture와 비교해봄. 복잡한 주제였어서 논문을 많이 읽고, 데이터에 대한 이해, 데이터의 활용 방법과 deep learning models 그리고 데이터와 그 활용/분석 단계에서의 한계에 대해 전반적으로 이해하는 한 주였다. Code 를 아직 활용하지 못했지만, 다음주부터는 code를 돌리는 작업할 것 같다.

주말동안 준비해야 할 것

1. 논문 두개 읽기
2. Pytorch tutorial 기본
3. Code 보고 전반적 이해

논문1. The personal and clinical utility of polygenic risk scores

ABSTRACT

대부분의 질병들은 아주 복잡한 유전 아키택처로 진행되고 만들어지기에 생각만큼 개인의 dna를 활용해서 질병을 예측하는 것은 실질적인 어려움이 있다. 그럼에도 불구하고 왜 polygenic risk profiling을 해야하는 지 설명하는 논문

- overview of the genetic architecture of common adult-onset diseases
- describe how genetic risk factors can be combined to produce PRSs
- review recent studies that have demonstrated the utility of PRSs for disease risk stratification as well as their implications for early disease detection, prevention, therapeutic intervention and/or life planning
- limitations of PRSs and the remaining barriers in utilization

Componets of disease risk

- Genetic susceptibility - quantified by heretability
- Environmental exposure
- Lifestyle factors

What is heritability?

- quantitative genetics perspective: the proportion of phenotypic variation in a population that can be explained by genetic variation

Heritability explained in a population vs individual disease risk

인구에 대한 genetic risk stratification 과 heritability 가 연관은 되어있지만 개인의 질병 가능성 취약성과 직접적 연관은 없을 수 있음.

예를 들어, In other words, although the total heritability explained by BRCA1 and BRCA2 variants is low, BRCA1 and BRCA2 testing can identify a subset of individuals whose absolute risk of disease is significantly higher than that of the average individual in the general population.

Genetic architecture of common diseases- monogenic(rare high-risk) vs polygenic

-a continuum of common low-risk to rare high-risk genetic variants

Polygenic disease susceptibility

-f many common adult-onset diseases is mediated by numerous common (MAF>5%) and low-frequency (MAF >0.5% and <5%) genetic variants captured through genome-wide genotyping and/or imputation
-omnigenic37 model of inheritance복잡한 형질에는 거의 모든 유전자가 영향을 미친다는 이론
Ex. coronary artery disease- common variants with small effect size
Ex. large-scale comprehensive GWAS for breast cancer found that 41% of familial relative risk of breast cancer can be explained by genetic variants captured by genotyping and imputation, again with no low-frequency variants of moderate effect size detected despite sufficient power to detect such associations4
statistical modelling and empirical results of comprehensive genomic studies reinforce the conclusion that the genetic architecture for many common adult-onset diseases is composed of a familial form, responsible for 1–10% of disease incidence, linked to highly penetrant rare variants within a small set of genes known to drive familial disease, and a nonfamilial form of disease that is mostly driven by an amalgamation of common variants of small effect distributed throughout the genome, in combination with a smaller contribution from rare variants of moderate effect in genes known to cause familial disease.

statistical modelling and empirical results of comprehensive genomic studies reinforce the conclusion that the genetic architecture for many common adult-onset diseases is composed of a familial form, responsible for 1–10% of disease incidence, linked to highly penetrant rare variants within a small set of genes known to drive familial disease, and a nonfamilial form of disease that is mostly driven by an amalgamation of common variants of small effect distributed throughout the genome, in combination with a smaller contribution from rare variants of moderate effect in genes known to cause familial disease.

PRS is most commonly calculated as a weighted sum of the number of risk alleles carried by an individual, where the risk alleles and their weights are defined by the loci and their measured effects as detected by GWAS.4
utility of GWAS-based genetic risk estimates has been assessed, perhaps inappropriately, on the basis of their ability to comprehensively discriminate between diseased and non-diseased individuals — usually quantified by the area under the curve (AUC) of a receiver operating characteristic curve, a plot of the true positive rate (sensitivity) versus false-positive rate (specificity).

PRS : risk stratification 이 아니라 realistic and practical goal is the identification of a subset of individuals at elevated risk of disease on the basis of genetic factors in combination with clinical risk factors.

The utility of prs risk score

PRS-informed therapeutic intervention (the part that PRS can play in the selection of interventions to treat or prevent disease)
PRS-informed disease screening (the role that PRS can have in the decision to initiate and the interpretation of disease screens)
PRS-informed life planning (the personal utility that PRSs can provide, even in the absence of preventive actions).

there may be differences in disease presentation, severity and available therapeutic interventions for disease for familial versus polygenic genetic susceptibility7,85.

Polygenic risk interpretation 의 차이,large-scale prospective studies examining the clinical utility of PRSs should be conducted. (simplicity)

논문2. Developing and evaluating polygenic risk prediction models for stratified disease prevention
ABSTRACT
Risk 예측 모델을 설계, 평가, 적용하는 방법들에 대한 정리
Case study를 통한 primary and secondary disease prevention
앞으로의 기회, 난제