

# 난소암 항암치료제 데이터 연구 계획서

참석자: 문정섭(2020-21324), 김민준(2013-13430), 백대현(2014-10451)

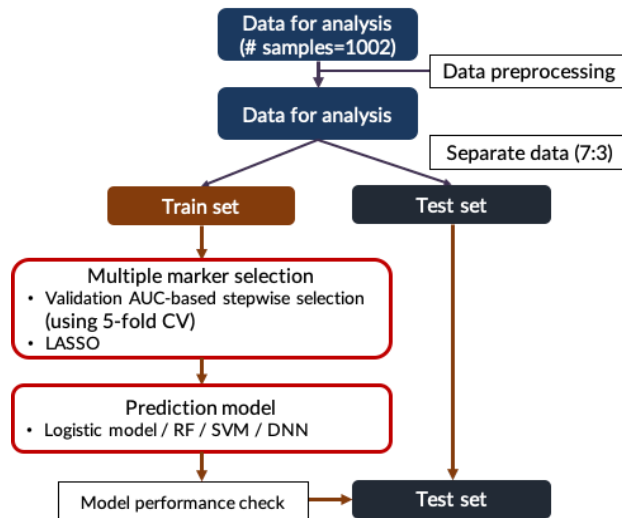
## 1. 연구목적:

- High grade Serous Ovarian Cancer(이하 HGSOC) 환자의 임상 자료를 기반으로 항암제 반응을 예측하기 위한 마커를 찾고, 그 마커들을 통해 예측하기 위한 통계 모델 개발.

## 2. 환자 임상자료 요약:

- sample수: 1002명
  - 세부사항: 서울대학교병원(568명), 서울아산병원(246명), 신촌세브란스병원(188명)
- 예측변수: 46개
  - Continuous 변수 10개
    - Age, No of family member with breast/gynecologic cancer upto 2nd degree, Height, Weight, BMI, CA125, Hb, No of harvested/positive LNs
  - Categorical 변수 36개
    - Parity, Menopausal\_state, Hypertension, Diabetes, Dyslipidemia, Personal\_history\_of\_breast\_cancer, Total cycles of 1st regimen, Familial\_history\_of\_breast/gynecologic\_cancer, Origin, Grade, FIGO2014, PLN/PALN\_status, NAC, Cycle\_NAC, Cycle\_1st\_regimen,

## 3. 연구 방법



- Train과 test를 위해 sample을 7:3 비율로 나눔.
- 결측치 중 continuous변수는 평균값으로, categorical변수는 제외하여 진행.
- Stepwise selection과 Lasso방법을 통해 마커 선택.
- 선택된 마커들을 이용하여 다양한 prediction model 학습.
- 학습된 model을 test set을 이용하여 확인.