

Mendelian Randomization을 이용한

# 질병 타겟발굴 및 검증

"유전역학 이론인 Mendelian Randomization에 기반하여, 공개된 GWAS 및 eQTL 빅데이터를 통합 분석해 보고, 신약 개발 타겟을 발굴 및 검증하는 프로젝트 진행"



Prof. Kim Sangsoo



## 지원 요건

R/Bioconductor 중급 이상  
Linux bash 쉘 숙련 필요



## 커리큘럼

프로젝트 수행을 위한 교육과 멘토링 구성  
(교육) 방법론 및 도구 교육  
(멘토링) 팀별 타겟발굴 및 검증 프로젝트 멘토링

\*타겟발굴 역량강화를 위한 LAIDD교육 40시간 수강필요



## 운영 방식

총 14회\* 온·오프라인 교육 운영

(교육) 토요일 오전 오프라인\*\* 진행

(멘토링) 토요일 오전 오프라인\*\* 진행

\*회별 2시간 이상 소요

\*\*참여인원 협의 후 온라인 진행 병행

## Work

송실대 의생명시스템학부 교수

송실대 교무처장 겸 대학교육혁신원장

## Experience

한국유전체학회 회장

한국생명정보학회 부회장

국가유전체정보센터장

한국생명공학연구원 책임연구원

LG화학 바이오텍연구소 책임연구원

## Education

Iowa State Univ. 물리화학 박사

온라인 지원하기 

# 멘토 교육과정

프로젝트  
교육

01. **SNP 및 GWAS 이론 교육**  
-1~2주차 교육 수행

프로젝트  
멘토링

02. **타겟 질환 선정 및 데이터 확보**  
-3~4주차 교육 수행  
-GWAS Catalog에서 팀별 타겟 질환 선정 및 Summary 데이터 확보

프로젝트  
교육

03. **Mendelian Randomization 교육**  
-5주차 교육 수행

프로젝트  
멘토링

04. **eQTL 개념 교육**  
-6주차 교육 수행

프로젝트  
멘토링

05. **Druggable eQTL 데이터 확보**  
-7~9주차 교육 수행  
-팀별 타겟에 대한 Druggable eQTL 데이터 확보

프로젝트  
멘토링

06. **Mendelian Randomization 기반 타겟발굴**  
-10~14주차 교육 수행  
\*교육과정의 세부 내용과 일정은 변경 될 수 있음.



## 수강 전 추천 과목



R프로그래밍



생물정보학개론



Genomic  
Analysis



분자생물학



**LAIDD 무료 제공**

\*수강 전 추천 과목을 수강 시 선발과정에 유리

## 역량강화 교육

**"타겟발굴 및 검증 역량 강화를 위한  
관련 LAIDD 강의 40시간 이수"**

세포신호전달개론

분자생물학 기초

약물 기전해석 및  
약물재창출

시스템생물학

Exome시퀀싱 기반  
신약타겟발굴

인공지능을 활용한  
EMR 데이터 분석

Genomic  
Analysis

전장유전체  
변이분석 이해

\*배정과목 필수 이수



Prof. Kim Hwajong

Work

강원대학교 컴퓨터공학과 교수  
KAIST IT융합연구소 겸직교수  
(주)데이터사이언스랩 대표

Experience

인공지능신약개발지원센터장  
강원대학교 데이터분석센터장

Metoring

기업 대상 AI/DX 집중교육 수행  
-LG전자, LG화학, LS그룹, 하림그룹 등

Education

KAIST 전기및전자과 데이터통신 박사

연합학습(Federated Learning)을 활용한  
약물-표적 상호작용(DTI) 예측

"연합학습의 동작원리와 구현기술을 배우고, 이를 이용하여 분산환경에서 약물-표적 상호작용(DTI) 등의 머신러닝 모델 구현 기법을 배운 후 응용 과제를 수행 "



지원 요건

Python 프로그램 초급 이상 필수

-pandas,numpy,matplotlib

머신러닝 동작원리 및 패키지 사용 경험 권장

-scikit-learn



커리큘럼

프로젝트 수행을 위한 교육과 멘토링 구성

(교육) 연합학습 및 DTI예측 모델 구현방법 교육

(멘토링) DTI 예측 모델 활용을 주제로한 프로젝트 멘토링

\*역량강화를 위한 맞춤형 LAIDD교육 40시간 수강 필요



운영 방식

총 10회\* 이상의 온·오프라인 교육 운영

(교육) 예제 프로그래밍 실습 및 질의 응답 참여

(멘토링) 응용 과제 수행 및 멘토링 (개인 및 팀별)

\*회별 3시간 이상 소요

\*\*참여인원 협의 후, 온라인/오프라인 교육 진행 협의

온라인 지원하기

# 멘토 교육과정

프로젝트  
교육

## 01. 연합학습 모델 이해 및 구현

- 연합학습 개념 소개
- 연합학습 기본 예제 실습
- 클라이언트 데이터 불균형에 대한 학습 결과 비교

프로젝트  
교육

## 02. 분자 독성 예측 모델 구현

- 분자 독성 예측 모델 구현
- 개별모델, 통합모델, 연합학습의 성능 비교

프로젝트  
교육

## 03. 전사인자 결합 예측 모델 구현

- 민감한 개인정보인 유전체 데이터를 연합학습으로 외부 이동 없이 모델에 적용
- 전사인자 결합 예측 모델 구현 실습 후 연합학습에 의한 성능 향상에 확인

프로젝트  
교육

## 04. 약물-표적 상호작용 예측 모델 구현

- 실험을 통해 제작 된 약물 표적 상호작용 데이터 소개
- DTI 모델 및 Pytorch 소개
- 각 클라이언트(기관)가 보유한 데이터량이 다른 시나리오 실습
- 개별적인 학습과 연합학습의 성능 비교

프로젝트  
멘토링

## 05. 프로젝트 주제 선정

- 교육 내용을 기반으로한 응용프로젝트 주제 선정
- 응용프로젝트는 개인별 참여를 기본으로하나, 필요 시 팀단위 응용과제 수행 가능

프로젝트  
멘토링

## 06. 프로젝트 수행 및 멘토링

- 프로젝트 수행 내용에 대한 멘토링 진행
- \*교육과정의 세부 내용과 일정은 변경 될 수 있음.



## 수강 전 추천 과목

기계학습



파이썬프로그래밍



LAIDD 무료 제공

\*수강 전 추천 과목을 수강 시 선발과정에 유리

## 역량강화 교육

"DTI 및 연합학습 역량 강화를 위한  
관련 LAIDD 강의 40시간 멘티별 맞춤 배정"



AI



Python



Chem



Data

\*배정과목 필수 이수



Dr. Hyunjin Shin

Work

녹십자 목암생명과학연구소 부소장

Experience

Takeda Phameceutical

-Principal Scientist

Dana-Farber Cancer Institue

-Research Fellow

Education

(PhD)The University of Texas at Austin

-Electrical & Computer Engineering

오믹스 데이터 및 약물 데이터를 이용한

약물반응예측

"유전체 및 전사체 데이터를 포함한 오믹스 데이터와 약물 데이터의 구조와 형태를 이해하고 머신러닝 및 인공지능 모델을 적용하여, 효능 및 독성 마커 발굴 및 예측 모델을 개발하는 프로젝트를 진행."



지원 조건

오믹스 및 약물 데이터에 대한 기본 이해  
기계학습 및 인공지능에 대한 기초  
Python / R Programming 기술 및 경험



커리큘럼

프로젝트 수행을 위한 교육과 멘토링 구성  
(교육) 오믹스, 약물 데이터의 이해 및 머신러닝 원리/활용 교육  
(멘토링) 약물반응 예측 모델링에 필요한 데이터 이해/처리, 프로젝트 디자인, 결과 해석 및 방향 설정  
\*생물정보학 역량강화를 위한 LAIDD교육 40시간 수강필요



운영 방식

총 10회 이상의 온·오프라인 병행 운영  
(교육) 금요일 오전 오프라인  
(멘토링) 금요일 오전 오프라인  
\*회별 3시간 이상 소요  
\*\*협의에 따른 온라인 진행 가능

온라인 지원하기

# 멘토 교육과정

프로젝트  
교육

## 01. 유전체 및 전사체 다중 오믹스 데이터와 약물데이터 이해

- Genomics and Transcriptomics Data
- Drug Screening Profiles

프로젝트  
교육

## 02. 머신러닝의 기본 이해

- Supervised Learning
- Performance Evaluation

프로젝트  
교육

## 03. 머신러닝 활용

- Training / Validation / Testing
- Key points to remember in the real-world application

프로젝트  
멘토링

## 04. 신약개발에서 약물반응 예측의 중요성

- Translational research for drug development
- Patient Stratification

프로젝트  
멘토링

## 05. 데이터 전처리 및 통합

- Public Databases
- Featurization

프로젝트  
멘토링

## 06. 정확한 약물반응 예측을 위한 머신러닝 활용

- Model design
  - Prediction results comparison
- \*교육과정의 세부 내용과 일정은 변경 될 수 있음.

## 수강 전 추천 과목

기계학습

전사체데이터분석

화학정보학개론

LAIDD 무료 제공

\*수강 전 추천 과목을 수강 시 선발과정에 유리

## 역량강화 교육

"생물정보학 및 인공지능 역량 강화를 위한  
관련 LAIDD 강의 40시간 배정"

AI

Python

NGS

Bioinformatics

\*배정과목 필수 이수



Prof. Lee Juyong

## Work

서울대학교 약학대학 제약학과 교수

## Experience

강원대학교 화학과 교수

미국 국립보건원(NIH) 연구원

고등과학원 계산과학부 연구원

## Subject

계산화학 및 계산생물학 기반

인공지능 신약개발

## Education

서울대학교 자연과학대학 화학부 박사

그래프 신경망을 활용한

# 분자 가상 스크리닝

"단백질-리간드 도킹 계산 및 그래프 신경망 기초를 학습하여, 실제 신약 타겟의 단백질-리간드 가상 스크리닝을 진행하여 후보 물질을 찾는 프로젝트를 진행"



### 지원 요건

단백질 및 단백질 구조에 대한 기초 지식  
계산화학을 이해하기 위한 기초 화학 지식  
Python 프로그래밍 기초 지식(클래스 활용)



### 커리큘럼

프로젝트 수행을 위한 교육과 멘토링 구성  
(교육) 가상 스크리닝 방법론 및 도구 교육  
(멘토링) 신약타겟 후보 물질 발굴 프로젝트 멘토링

\*화학정보학 역량강화를 위한 LAIDD교육 40시간 수강필요



### 운영 방식

총 10회 이상\*의 온·오프라인 병행 운영

(교육) 교육생 협의 후, 온라인 위주\*\* 진행

(멘토링) 팀별 협의 후, 온라인 위주\*\* 진행

\*회별 3시간 이상 소요

\*\*필요 시 토요일 오전 오프라인 교육 및 멘토링 진행

온라인 지원하기

# 멘토 교육과정

프로젝트  
교육

## 01. 단백질-리간드 도킹 계산

-Autodock-GPU 도킹 계산 프로그램 실습

프로젝트  
교육

## 02. 화학정보학 및 RDKit 교육

-분자 물성 계산 및 유사도 계산 실습

프로젝트  
교육

## 03. 머신러닝 기초 교육

-머신러닝 및 Tree-based 예측 방법 개념  
-딥러닝 기본 개념

프로젝트  
멘토링

## 04. ADME/Tox 예측 모델 개발

-ChEMBL의 분자 물성 데이터 다운로드 및 예측 모델 개발

프로젝트  
멘토링

## 05. 리간드 활성화도 예측 모델 개발

-리간드 활성 데이터 수집 및 예측 모델 개발

프로젝트  
멘토링

## 06. 단백질-리간드 기반 가상 스크리닝 수행

-실제 Medicinal Chemistry 라이브러리 기반 스크리닝 계산 수행

\*교육과정의 세부 내용과 일정은 변경 될 수 있음.



## 수강 전 추천 과목

화학정보학개론



RDKit 기초



LAIDD 무료 제공

\*수강 전 추천 과목을 수강 시 선발과정에 유리

## 역량강화 교육

"화학정보학 및 인공지능 역량 강화를 위한  
관련 LAIDD 강의 40시간 배정"

AI

Python

Chem  
informatics

Data

\*배정과목 필수 이수

## 다중 오믹스 데이터 통합분석을 통한 암 치료 약물 타겟 발굴

"전사체-단백체-인산화단백체-단일세포 전사체 데이터 통합분석 능력을 배양하고, 암 데이터에 적용하여 암치료 약물 타겟 발굴하는 프로젝트를 진행 "



Prof. DAEHEE Hwang

### Work

서울대학교 자연과학대학 생명과학부 교수

### Experience

디지스트 뉴바이올로지 교수  
포스텍(Postech) 조/부교수

### Subject

시스템생물학 기반 연구수행

### Education

매사추세츠공과대학(MIT) 화학공학과 박사



### 지원 요건

NGS 시퀀싱 데이터 분석 기초

-유전체, 전사체, ChIP seq 중 하나 이상 경험

기본 통계 및 다변량통계분석 기초

-통계가설 검증, 주성분분석, 군집화 등

프로그래밍 경험

-Matlab/R/Python 중 1



### 커리큘럼

프로젝트 수행을 위한 교육과 멘토링 구성

(교육) 개별 오믹스 데이터 수집 및 다중 오믹스 통합분석  
(멘토링) 단일세포 전사체 데이터 통합을 통한 타겟발굴

\*생물정보학 역량강화를 위한 LAIDD교육 40시간 수강필요



### 운영 방식

총 10회 이상의 온·오프라인 병행 운영

(교육) 토요일 오전 오프라인 진행

(멘토링) 토요일 오전 오프라인 진행

\*회별 3시간 이상 소요

\*\*협의를 따른 온라인 진행 가능

온라인 지원하기

# 멘토 교육과정

프로젝트  
멘토링

## 01. 타겟 질환 선정 및 데이터 확보

- 타겟 질환 선정 (e.g., 췌장암)
- 해당 질환에 대한 메타 정보 확보 (나이, 기본 EMR, 암진단 시기, Histological Subtype, 수술여부, 전이여부, 생존기간, 화학약물 종류 등)

프로젝트  
교육

## 02. 전사체 데이터 수집 및 분석

- 타겟 질환에 대한 공개 데이터 수집(ICPC, CPTAC)
- 전사체 데이터 분석(alignment, quantification, normalization, etc)

프로젝트  
교육

## 03. 단백질체 데이터 통합 분석

- 단백체 데이터(ICPC, CPTAC) 수집 (Global Proteome 및 Phosphoproteome)
- 단백체 데이터 분석(peptide identification, quantification, normalization, etc)

프로젝트  
교육

## 04. 다중오믹스 통합 분석

- 개별오믹스 (전사체, 단백질체, 인산화 단백질체) 기반 환자 클러스터링
- 다중 오믹스 통합 클러스터링
- 분자 시그니처 및 세포 신호경로 동정

프로젝트  
멘토링

## 05. 단일세포 전사체 데이터 통합

- 예후가 좋지않은 군집(Subtype)의 분자 시그니처 동정
- 시그니처 기반으로 단일세포 전사체 데이터를 통합하여 타겟세포에서 발현하는 타겟분자 제시

프로젝트  
멘토링

## 06. 데이터 분석을 통한 타겟 발굴

- 네트워크 분석 및 단일세포 전사체 기반 타겟 후보군 선별
  - 선정된 타겟 분자에 대한 특허, 논문 서치를 통한 신규 타겟 선별
- \*교육과정의 세부 내용과 일정은 변경 될 수 있음.



## 수강 전 추천 과목

R프로그래밍



시스템생물학



LAIDD 무료 제공

\*수강 전 추천 과목을 수강 시 선발과정에 유리

## 역량강화 교육

"생물정보학 및 인공지능 역량 강화를 위한  
관련 LAIDD 강의 40시간 배정"

AI

Python

NGS

Bioinformatics

\*배정과목 필수 이수