

강의개요

DNA Foundation Model 이론 및 실습

생물학 텍스트와 DNA 서열을 컴퓨터는 어떻게 이해할 수 있을까? 수억 편의 논문과 수십억 염기쌍의 게놈 데이터에서 의미 있는 패턴을 어떻게 찾아낼 수 있을까? 자연어 처리에서 발전한 언어 모델 기술은 생물정보학 연구의 방식을 근본적으로 변화시키고 있다.

본 강의에서는 텍스트와 DNA 서열 분석을 위한 AI 언어 모델의 핵심 원리를 다룬다. 단어를 벡터로 변환하는 Embedding 기법, 문맥을 이해하는 Attention 메커니즘, DNA 서열 학습하는 Foundation Model의 개념을 설명하며, 이를 DNA 유전체 변이 분석에 적용한 최신 모델들을 소개한다. 이를 통해 유전 변이 효과를 예측하고, 장거리 조절 상호작용을 분석할 수 있는 실무 능력을 갖추는 것을 목표로 한다.

강의는 다음의 내용을 포함한다:

- 언어 모델의 기본 원리와 생물학 텍스트 처리
- 대규모 사전학습 모델과 전이학습 방법론
- DNA 서열을 위한 언어 모델 설계
- 장거리 게놈 분석을 위한 차세대 아키텍처

*참고강의교재:

AI for Genomic Science (chaek.org) <https://chaek.org/books/ai-for-genomic-science>

*교육생준비물:

노트북 (인터넷 연결 필요 - 구글 코랩 실습을 위한 구글 계정)

* 강의 난이도: 중급

* 강의: 안준용 교수 (고려대학교 바이오시스템의과학부)

Curriculum Vitae

Speaker Name: **Joon-Yong An, Ph.D.**



► Personal Info

Name Joon-Yong An
Title Associate Professor
Affiliation Korea University

► Contact Information

145 Anam-ro, Seongbuk-gu, Seoul, South Korea
Email joonan30@korea.ac.kr
Phone Number 02-3290-5646

Research interest : Whole genome sequencing, Single cell RNA sequencing, and neurodevelopmental disorders

Educational Experience

2010 B.S. in Molecular Biotechnology, Konkuk University
2011 M.S. in Molecular Biology, University of Queensland (Australia)
2016 Ph.D. in Neuroscience, University of Queensland (Australia)

Professional Experience

2015-2019 Postdoctoral Fellow , University of California, San Francisco
2019-2022 Assistant Professor, Korea University
2022- Associate Professor, Korea University

Selected Publications (3 maximum)

1. Kim SW*, Lee H*, (...), Kim EJ**, Werling DM**, Yoo HJ**, An JY**, Whole genome sequencing analysis identifies sex differences of familial pattern contributing to phenotypic diversity in autism. Genome Medicine, 2024
2. Song KJ*, Choi S*, Kim K*, Hwang HS*, (...), Na S**, Jang SJ**, An JY**, Kim KP**, Proteogenomic Analysis of a Korean Cohort Reveals Lung Cancer Subtypes Predictive of Metastasis, Chromosome Instability, and Tumor Microenvironment, Nature Communications, 2024
3. Kim Y*, Jeong M*, (...), An JY**, CWAS-Plus: Estimating category-wide association of rare noncoding variation from whole-genome sequencing data with cell-type-specific functional data, Briefings in Bioinformatics, 2024