

김가경 (Kim, Ka-Kyung)

검체에서 임상 판단까지, 유전체 분석 설계와 운영

바이오인포매틱스 / 헬스케어 AI / 임상 유전체 분석

Email kakyung.kim@gmail.com

Homepage kakyungkim.github.io

LinkedIn linkedin.com/in/kakyungkim

GitHub github.com/kakyungkim

■ 요약

- 분자과학 박사, SCI(E) 20편 이상. 유전체·임상 데이터 기반 연구·분석과 프로젝트 리딩 약 17년
- 멀티오믹스(NGS, 전사체, 단일세포) 전처리, 정규화, 통합 분석 및 ML/DL 예측·생성 모델 설계·검증
- LLM·AI Agent 기반 분석 자동화 설계·구현 (VeriVar citation-grounded RAG, AutoBioX 다중 에이전트 연구 자동화)
- Oncology, 정밀의료, 임상시험(Ph1-3) 데이터 분석·해석 전문성; 임상 데이터에서 분석에서 리포트, 서비스까지 연결하는 실행력
- 임상 유전체 분석 경험과 LLM, RAG, AI Agent 구현, 모델 평가·검증 설계 역량을 함께 보유. 도메인 이해, AI 구현, 임상 등급 검증 체계를 연결하는 것이 강점

■ 핵심 역량

BIOINFORMATICS & OMICS

멀티 플랫폼 NGS(WGS/WES/Targeted-seq, RNA-seq, scRNA-seq), HLA/Immune profiling, eQTL/GWAS/CNV, pathway 분석

CLINICAL & LEADERSHIP

임상시험(Ph1-3) 분석 전략 및 멀티오믹스 분석 서비스 로드맵 수립, K-MASTER/K-DNA 국책과제 총괄, 팀·부서 리딩, GCLP/CAP/ISO27001 대응

AI/ML & LLM

병리 foundation model embedding 기반 분자 표현형 예측과 바이오마커 예측(PyTorch/scikit-learn); LLM, RAG, AI Agent로 변이 해석과 임상 리포트 자동화

DATA-TO-PRODUCT & 확장성

임상 데이터에서 분석에서 소프트웨어, 사업화까지 연결; 데이터 종류에 매이지 않는 통합, 모델링, 자동화 역량으로 생체신호·헬스케어 데이터에도 확장

■ 주요 프로젝트

세포 이미지 판독 보조 AI 모델 개발·진단 업무 적용 (2026, 싸이토젠)

- 판독자 판정 기록을 재구성해 측정값 기반과 이미지 딥러닝 두 계열로 분류 모델 개발; 검체 단위 교차검증으로 데이터 누수 차단(누수 시 kappa가 0.910으로 과대평가되던 것을 0.704로 정정)
- 확신도 기반 선별로 사람이 우선 확인할 대상을 70퍼센트 줄이고, GPU 학습과 사내 CPU 추론을 분리해 배포; 판독 보조 기술백서 한/영 작성

AI 활용 기준 수립과 사내 데이터 도구 4종, 통합 포털 구축 (2026, 싸이토젠)

- AI 도구 3종을 같은 조건으로 비교해 업무별 활용 기준을 정하고, 판독과 실험, 영업 부서 담당자가 실제로 쓰는 도구를 각 부서와 협의해 구현
- 검사 현황판, 리포트 생성기, 분석 대시보드, 문헌 조사 도구를 기획·개발해 사내망 포털에 통합하고 서버 운영을 담당; 분석 대시보드에는 답변과 근거 문헌을 함께 확인하는 질의응답(RAG)을 연동

ADMET과 약동학 분할 검증, 후보 선별 평가 (2026, 개인 검증 실험)

- TDC ADMET 12개 과제와 약동학 3개 과제에서 무작위, 골격, 구조 분리 분할을 비교; 지문+LightGBM, MPNN, ChemBERTa 미세조정을 같은 조건으로 검증
- 구조 분리 분할에서 성능 지표가 12개 과제 전부 하락(하락 중앙값 0.098, 시드 5회); 시드를 15회로 늘린 선별 평가에서는 상위 10퍼센트 적용률이 8개 과제 전부 하락하고 분포용적(VDss)은 0.358에서 0.176으로 내려감($p < 0.001$). 초기 결론 셋을 철회하고 재검증 과정을 공개, 분할 감사 절차는 화학에 국한되지 않는 패키지로 정리

VeriVar: LLM 기반 체세포 변이 해석 파이프라인 (개인 프로젝트, 공개 준비 중)

- citation-grounded retrieval과 faithfulness 계약으로 근거 없는 인용을 통제; ClinGen/CGC/VICC 2022 발암성과 AMP/ASCO/CAP 2017 임상 tier를 함께 판정하고 규칙, LLM, 무검색 baseline을 나란히 실행
- 기술: Python, LangChain, RAG, 지식그래프, TCGA PanCancer-CIViC 공개 데이터 기반 오프라인 재현. 2026.01 ABDD iTalk 발표

AutoBioX: 여러 AI 에이전트로 생물정보 분석을 자동화하는 사이드 프로젝트 (프로젝트 리더, PseudoCon 2026 발표)

- SpatialPathoAgent: H&E whole-slide image를 병리 foundation model(UNI, CONCH 등) embedding으로 변환해 ER/PR/HER2, PAM50 등 분자 표현형 예측 (A100 GPU 파이프라인)
- 자료 준비, 예측, 검증, 근거 조사를 나눈 다중 에이전트 워크플로; 데이터 누수와 baseline, 생물학적 타당성 Critic 체크리스트에 작성자-검토자 분리 게이트 적용

MCP 서버와 에이전트 구현 (2025~2026, 해커톤 산출물)

- PharmaSignal: 규제 공시와 임상시험 등록 자료를 모아 신약 안전성과 경쟁 신호를 정리하는 에이전트. 데이터 민감도에 따라 추론 경로를 나눠 반출 제한 자료가 상용 API로 가지 않게 설계 (Agent Forge AI Hackathon Seoul, 2026.08, github.com/kakyungkim/pharmasignal)
- ChillMCP: FastMCP로 MCP 서버를 구현해 모델이 호출하는 도구 12개와 대화형 프롬프트 5종, 상태 관리와 테스트 클라이언트를 작성하고 MIT 라이선스로 공개 (SKT AI Summit Hackathon 2025 사전 미션, github.com/kakyungkim/chillmcp)

Multi-omics 기반 TME 바이오마커 예측 (2025 Cheiron ResearchThon, Creative Academic Impact Award)

- Liquid biopsy 멀티오믹스 통합 및 feature engineering으로 종양 미세환경(TME) 예측 AI 모델 설계
- 개 유선암 WES/WGS/WTS를 사람 유방암과 비교한 종간 발암 시그니처 연구 (Sci Data 2019 제1저자, Nat Commun 2020 공저자)
- 위암 PDX 유전체·전사체 통합, 소아 천식 노코딩 eQTL 등 다양한 질환 오믹스 분석 수행

Vital STORM (AI Agent Hackathon) 및 임상 NGS 자동화

- Instruct.KR x Sonic 2025 해커톤: Parser STORM + Predictor STORM 이중 구조, LLM+RAG로 국가건강검진 매뉴얼 학습 및 수치 판정·2차 검진 권고 자동화
- 마크로젠 시퀀싱 데이터 실시간 트리거, 임상 변이 해석·리포트 생성 워크플로 구축
- BRCA1/2 변이 분석 소프트웨어 개발·성능검증 및 CE-IVD 수준 문서화

■ 학력 사항

2021.03 ~ 2023.08 **한국방송통신대학교** 학사: 보건환경학과, 경영학과 복수 전공

2007.03 ~ 2011.02 **성균관대학교** 박사: 분자의과학, 삼성생명과학연구소 인간유전체학연구실 연구 수행
(Thesis: HapMap 데이터를 이용한 한국인의 결측치 유전자형 추정의 인구유전학적 분석)

2002.09 ~ 2004.08 **고려대학교** 석사: 생명과학
(Thesis: Immunolocalization of Neurosubstances in Larval Brains of the Wax Moth Galleria mellonella)

1996.03 ~ 2000.02 **동국대학교** 학사: 컴퓨터공학과

■ 경력 (약 17년, 박사과정 기간 제외)

2025.03 ~ 현재 **싸이토젠**

- 수석연구원 / 팀장 - 진단서비스팀 (2025.03 ~ 현재)

세포 이미지 판독 보조 모델 개발과 사내 배포, 사내 데이터 도구 4종과 통합 포털 구축, AI 도구 활용 기준 수립, 순환종양세포 기반 임상 서비스 운영 체계 구축 및 상용화 기반 마련, 단일세포(scRNA-Seq) 분석 파이프라인 내재화 및 품질 기준 수립, 분석-실험 통합 QC 체계 구축 및 임상 리포트 표준화

2021.10 ~ 2024.06 **테라젠바이오**

- 부장 - 임상분석팀 (2022.11 ~ 2024.06)
- 부장 - 생명정보부 / NGS 사업본부 (2021.10 ~ 2022.11)

제약사 임상시험(Ph1-3) 분석 총괄 및 다기관 프로젝트 운영, GCLP 기반 분석 SOP 및 규제 대응 체계 구축 (식약처 대응 포함), 멀티오믹스 분석 서비스 고도화 및 자동화 시스템 구축; AWS 기반 대량 WGS 클라우드 처리 체계 총괄

2018.10 ~ 2021.10 마크로젠

- 차장 - 사업개발 / 사업개발팀 (2021.01 ~ 2021.10)
- 부서장 - 임상분석 / 임상분석1부 (2020.01 ~ 2020.12)
- 임상개발/암유전체팀장 (2019.01 ~ 2019.12) - 임상분석 / 임상분석부

임상 유전체 분석 조직 운영 및 자동화 파이프라인 구축, 해외 유전체 분석 서비스의 도입 타당성 평가와 분야별 서비스 전략 수립 (사업개발팀, 2021)

2015.03 ~ 2018.10 연세대학교

- 박사후연구원 / 의생명시스템정보학교실

NGS 멀티오믹스로 질환 메커니즘을 분석하고, eQTL과 WGS 가족 데이터로 바이오마커를 탐색

2012.03 ~ 2015.02 삼성융합의과학원

- 박사후연구원 / 분자정보유전체학연구실

마이크로RNA 표적 데이터베이스(miRTCat)를 구축하고 질병 연관 변이를 통합 분석. 한국연구재단 리서치펠로우 수행

2011.02 ~ 2012.03 삼성생명과학연구소

- 박사후연구원 / 인간유전체학 연구실 (2007.03~2011.02 박사과정은 학력)

Korean HapMap과 KARE 코호트를 분석하고, 유전 질환과 약물 반응 관련 변이를 연구

2004.09 ~ 2007.02 국립보건연구원

- 기술연구원 / 유전체센터

한국인 질환 관련 변이를 발굴하고 SNP 변이 데이터베이스와 관리 시스템을 구축

■ 기술 역량

BIOINFORMATICS & DATA ANALYSIS

멀티 플랫폼 NGS (WGS/WES/Targeted, RNA-seq, scRNA-seq); Seurat, Scanpy, DESeq2, EdgeR, GSEA; 멀티오믹스 통합, eQTL 분석, variant 해석, 임상 QC 체계 구축

AI / LLM 활용

LangChain 기반 LLM 파이프라인, RAG, AI Agent 설계·구현; variant 해석 및 건강검진 결과 해석 자동화

CLINICAL & QUALITY

GCLP, CAP, ISO 기반 품질 시스템 운영 및 규제 대응, 임상 리포트 작성

PIPELINE & AUTOMATION

Snakemake-Docker 기반 파이프라인 설계·자동화; AWS 기반 클라우드 처리; 시퀀서 산출 데이터를 자동 감지해 분석까지 잇는 실시간 트리거 워크플로 구축

PROGRAMMING

Python, R 기반 데이터 처리·분석; PyTorch, scikit-learn; Linux/Shell scripting

■ 외국어 및 해외 경험

- 영어 OPIc IM2, 영문 논문과 기술 문서 독해 및 작성
- 미국 어학연수 (1년), 국제 학회(ASHG, ISMB/ECCB) 참석

■ 자격 및 교육

- 정보처리기사 (한국산업인력공단, 2006.06), 리눅스마스터 2급 (한국정보통신진흥협회, 2006.11)
- 머신러닝-데이터 분석 (Coursera, Johns Hopkins Univ., KDT 아이펠 데이터사이언티스트 과정(2024)): ICU 환자 임상 데이터 분석, OSINT/GEOINT 기반 공간정보 분석, LangChain-RAG 기반 챗봇 서비스 개발 등 팀 프로젝트 수행
- 임상·규제 교육 (GMP, GLP, 임상시험 설계, EU-MDR/IVDR)
- AI 기반 신약개발 데이터 분석 및 SaMD 모델 구축 (한국보건복지인재원·한국표준협회, 2025), 바이오헬스 AX 최고전략과정 (한국보건복지인재원, 2026.08~09)

■ 발표

- 2026.11 GIW ISCB-Asia 2026 구두 발표 채택: 유전자별 multiome RNA velocity 파라미터의 신뢰도 지도 (BIOINFO 2026 / GIW XXXV / ISCB-Asia, 연세대. 287편 중 22편 채택)
- 2026.07 AutoBioX 다중 에이전트 파이프라인 (PseudoCon 2026, 발표)
- 2026.01 종양 NGS 임상 리포트 자동화: 이중 모드 LLM 파이프라인 (ABDD Summit 2026 at Stanford, i-Talk 및 포스터)
- 2025.11 단일세포 QC 표준화와 scRNA-seq×면역형광 파이프라인 (Future of Medicine Symposium, 라이트닝 토크)
- 2024.10 포스터 발표 (ABDD Summit, Best Poster Presentation 수상)
- 2022.04 암: 만병의 황제의 역사 (제1회 K-BioX Global Mentoring EXPO, 초청 강연)

■ 수상 및 활동

- 2025 Cheiron ResearchThon Creative Academic Impact Award (K-BioX)
- 2024 ABDD Summit Best Poster Presentation Award (Merit Awards)
- 2021 의료정보 분석 전문가 심화과정 장려상 (서울아산병원)
- 2019 Korea Clinical Datathon 대상 (국가임상시험지원재단)
- 2011 한국분자세포생물학회 우수 포스터상

■ 논문 및 기고

SCI(E) 논문 20편 이상 | GOOGLE SCHOLAR

scholar.google.com/citations?user=jSjolNAAAAAJ 대표 논문: Kim TM et al. (2020) Nat Commun 11:3616 / Kim KK et al. (2019) Sci Data 6:147 (제1저자) / Kim KW et al. (2022) Allergy 77:218-229 / Ann SJ†·Kim KK† et al. (2018) PLoS One 13:e0193649 (공동 제1저자) / Kim KK et al. (2013) Bioinformatics 29:1898 (miRTCat)

학회 초록

GIW ISCB-Asia 2026 초록 2건 채택(구두 1건, 포스터 1건). 포스터는 AutoBioX 공동 연구 "Predictability Is Not Substitutability: A cost-of-substitution framework for H&E-based molecular prediction across 5 cancers"

BRIC VIEW 기고

클라우드 기반 유전체 정보 관리(2017) / 알레르기 질환 치료제 발굴 동향(2018) / 단일세포 분석 기반 신약개발과 환자선별(2025) / 바이오 빅데이터의 통합과 확장(2026) / 생성형 AI의 바이오·신약개발 혁신 동향(2026)