Ka-Kyung Kim, Ph.D.

Genomic analysis from specimen to clinical decision

Chief Researcher, Diagnosis Division, Cytogen · Seoul, Korea

Email kakyung.kim@gmail.com
Homepage kakyungkim.github.io
LinkedIn linkedin.com/in/kakyungkim
GitHub github.com/kakyungkim
Scholar scholar.google.com/citations?user=jSjoINAAAAAJ

Summary

I have worked with genome, transcriptome, single-cell, and clinical multi-omics data since 2004, about seventeen years of professional work outside the doctoral period. The analysis carries through to reports and services that hospitals and the business actually use.

Focus Areas

- Multi-platform NGS and multi-omics analysis, including scRNA-seq
- Clinical trial data analysis (Ph1-3) and regulatory frameworks (GCLP/CAP/ISO)
- End-to-end pipeline development with automation and QC standardization
- AI/LLM-driven variant interpretation and analysis workflows (PyTorch, scikit-learn, LangChain)

Professional Experience

2025-present — **Cytogen** Chief Researcher, Diagnosis Division

Develops and deploys AI for diagnostic work: a cell image review assistant, four in-house data tools on one internal portal, and shared criteria for AI tool use. Also leads CTC diagnostics, scRNA-seq pipeline development, clinical report automation, and GCLP quality systems; first to define the diagnostic service portfolio and its pricing, with the market research behind it

2021-2024 — **Theragen Bio** Director · Clinical Analysis Team / Bioinformatics

Led Ph1-3 oncology trial analysis, obtained new GCLP accreditation and designed its data management scheme, analysis lead for the national bio-big-data pilot; oversaw a large-scale WGS delivery pipeline run on AWS to meet fast-turnaround customer demand

2018-2021 — **Macrogen** Deputy GM · Business Development / Clinical Analysis

Led K-MASTER genomic analysis, BRCA software with CE-IVD documentation, and liquid-biopsy commercialization strategy; assessed the feasibility of licensing in an overseas genomic analysis service and set analysis-service strategy by segment (Business Development, 2021)

2015-2018 — **Yonsei University** Postdoctoral Researcher · Biomedical Systems Informatics

Multi-omics disease variant discovery and non-coding eQTL translational research across gastric cancer PDX, canine mammary tumors, pediatric allergy, and neurological disease

2012-2015 — **SAIHST, Sungkyunkwan University** Postdoctoral Researcher · Molecular Informatics and Genomics

Built the miRTCat database; NRF Research Fellow studying disease variants at non-canonical microRNA target sites

2007-2012 — **Samsung Biomedical Research Institute** Researcher · Human Genomics (2007-2011 concurrent with the Ph.D.)

Korean HapMap and KARE national cohort analysis, pharmacogenomics, and genotype imputation for the Korean population

2004-2007 — **Korea National Institute of Health** Research Technician · Center for Genome Science

Disease-associated variant discovery in the Korean population; built and operated the SNP database

Selected Programs

Cytogen	**CTC-based liquid biopsy diagnostics** Built and commercialized the end-to-end operating system for circulating-tumor-cell diagnostics: assay concept, pipeline, clinical report, and customer technical support.
Theragen Bio	**Immuno-oncology clinical trial analysis, Ph1-3** Led oncology trial analysis for three pharmaceutical companies.
Theragen Bio	**Large-scale WGS delivery on AWS** Oversaw an AWS-based cloud pipeline for high-volume, fast-turnaround WGS delivery to pharma clients.
Macrogen	**K-MASTER precision oncology genomics** Led genomic analysis for a national precision-oncology platform.
Macrogen	**BRCA1/2 variant analysis software, CE-IVD documentation** Directed development of software spanning variant calling, clinical interpretation, and automated report generation, together with CE-IVD-grade documentation and performance evaluation.
2021-2024	**Quality and regulatory systems** Obtained new GCLP accreditation for a clinical specimen analysis laboratory, and established recurring certification cycles: CAP proficiency testing twice a year and ISO27001 annually.

Built

In-house · Vision	**Cell image review assistant** A classification model for CTC image review trained on 1,531 labeled cells, with specimen-level cross-validation to block data leakage (an inflated kappa of 0.910 was corrected to 0.704 before release). Confidence-based triage cut the cells needing human confirmation by about 70%. Trained on GPU and deployed for CPU inference in-house.
In-house · Tooling	**AI usage criteria, four data tools, and an internal portal** Compared three AI tools under the same conditions and set criteria for which task each fits, then built the tools that review, laboratory, and sales staff actually use, in consultation with each team. A test status board, a report generator, an analysis dashboard, and a literature search tool run on one internal portal, with retrieval-based question answering (RAG) that shows the answer together with its source documents.
In-house · LLM	**LLM-based genomic Q&A and clinical report automation** An LLM system that answers from test results and source documents (RAG), together with clinical report automation.
Personal · Validation	**ADMET split audit** A benchmarking study on how the train/test split changes reported performance. Compared random, scaffold, and structure-disjoint splits across 12 TDC ADMET tasks plus 3 pharmacokinetic tasks, on fingerprint+LightGBM, a message-passing neural network, and a fine-tuned ChemBERTa. Under the structure-disjoint split all 12 tasks lost performance (median drop 0.098, 5 seeds); with 15 seeds the top-10% hit rate fell in all 8 tasks evaluated, volume of distribution from 0.358 to 0.176 ($p<0.001$). Three initial conclusions were retracted after adding seeds, and the correction was published openly.
Personal · LLM	**VeriVar** A somatic variant interpretation pipeline. Repository release in preparation.
Multi-agent research	**AutoBioX** A multi-agent research harness that splits data preparation, prediction, verification, evidence search, and manuscript drafting.

Personal · MCP

MCP servers and agents

PharmaSignal, an agent that gathers regulatory filings and clinical trial registrations to organize drug safety and competitive signals, routing inference by data sensitivity so restricted material never reaches a commercial API (Agent Forge AI Hackathon Seoul, 2026.08; github.com/kakyungkim/pharmasignal). ChillMCP, an MCP server on FastMCP with 12 model-callable tools, 5 interactive prompts, state handling, and a test client, released under MIT (SKT AI Summit Hackathon 2025 pre-task; github.com/kakyungkim/chillmcp).

AI agent · Hackathon

Vital STORM

An agent that grades key health-screening values on a five-level scale and routes to follow-up examinations, answering from four national health-screening manuals through retrieval (RAG).

Open source · Agent harness

econ-radar

An agent harness that assembles and publishes a daily economic digest weighted toward pharma, bio, and AI.

Selected Publications

2020 Cross-species oncogenic signatures of breast cancer in canine mammary tumors
Kim TM, … Kim KK, … Kim S. Nature Communications 11:3616

2019 Whole-exome and whole-transcriptome sequencing of canine mammary gland tumors First author
Kim KK, Seung BJ, … Kim S. Scientific Data 6(1):147

2018 Palmitate and minimally-modified low-density lipoprotein cooperatively promote inflammatory responses in macrophages Co-first author
Ann SJ†, Kim KK†, … Miller YI. PLoS One 13:e0193649

2013 miRTCat: a comprehensive map of human and mouse microRNA target sites including non-canonical nucleation bulges First author
Kim KK, Ham J, Chi SW. Bioinformatics 29(15):1898-1899

The full list of 20+ papers is on Google Scholar. Conference abstracts: two abstracts were accepted at GIW ISCB-Asia 2026, one oral and one poster; the poster is joint AutoBioX work, "Predictability Is Not Substitutability: A cost-of-substitution framework for H&E-based molecular prediction across 5 cancers".

Talks

2026.11 **A reliability map for per-gene multiome RNA velocity parameters in single-cell kinetics**
BIOINFO 2026 · GIW XXXV · ISCB-Asia, Seoul · oral presentation (accepted; 22 of 287 abstracts)

2026.07 **The AutoBioX multi-agent pipeline**
PseudoCon 2026 · talk

2026.01 **Automated Clinical Reporting for Oncology NGS: Dual-Mode LLM Pipeline**
ABDD Summit 2026 at Stanford · i-Talk and poster

2025.11 **Single-cell QC standardization and an scRNA-seq × immunofluorescence pipeline**
Future of Medicine Symposium: At the Crossroads of Computational Pathology and Spatial Biology · lightning talk

2024.10 **Poster presentation, ABDD Summit**
Awarded Best Poster Presentation

2022.04 **Cancer: a history of the emperor of all maladies**
1st K-BioX Global Mentoring EXPO · invited talk

Writing

2026	**Generative AI in biotech and drug discovery: technology, cases, and ecosystem outlook** BRIC View 2026-T24 A trend report separating how strong the evidence is at each stage, from protein design through the clinic.
2026	**Integration and expansion of bio big data: data, AI, and industry** BRIC View 2026-T11
2025	**Single-cell analysis in drug discovery and patient selection** BRIC View 2025-T29
2018	**Discovery trends and outlook for allergic disease therapeutics** BRIC View 2018-T16
2017	**Cloud-based management and services for genomic information** BRIC View 2017-T14

Awards

2025	Cheiron ResearchThon, Creative Academic Impact Award · liquid-biopsy multi-omics model for tumor microenvironment prediction
2024	ABDD Summit, Best Poster Presentation Award · Merit Awards
2021	Medical Informatics Advanced Course, Encouragement Award · Asan Medical Center
2019	Korea Clinical Datathon, Grand Prize · sepsis prediction, KoNECT
2011	KSMCB Annual Meeting, Outstanding Poster Award

Education

2021-2023	**Korea National Open University** B.S. Health & Environment, Business (double major)
2007-2011	**Sungkyunkwan University** Ph.D. Molecular Medicine
2002-2004	**Korea University** M.S. Life Science
1996-2000	**Dongguk University** B.S. Computer Science

Certifications and Languages

- Engineer Information Processing (HRD Korea, 2006); Linux Master Level 2 (KAIT, 2006)
- English: OPIc IM2. Reads and writes technical documents and papers in English; one year of language study in the United States
- Machine learning and data analysis: Coursera (Johns Hopkins Univ.), KDT AIFFEL Data Scientist course (2024)
- AI-driven drug discovery data analysis and SaMD model development (KOHI · KSA, 2025); Biohealth AX Executive Strategy Program (KOHI, Aug-Sep 2026)

Updated September 2026