

with Omics Foundation Model

Bridging Life Sciences and AI

SilicoPharm

Development and Application of the Omics Foundation Model

Development of a Generative AI for Life Science Experimental Results and
Its Application to Neurological Disease Drug Development

2025. 09. 02.

K-Global@SiliconValley

The background features a person in a white lab coat, likely a scientist or researcher, with their hands clasped. Overlaid on this is a stylized, isometric illustration of four server racks. The racks are light blue with dark blue horizontal lines and small colored indicators (red, green, blue). They are connected by a network of thin, glowing blue lines that form a grid and loop patterns across the scene. In the top right and bottom left corners, there are small, stylized icons of desktop computers with monitors and keyboards, also connected to the network lines. The overall color palette is dominated by light blues, dark blues, and greys, with a soft, ethereal glow from the network lines.

This document is prepared and shared solely for the designated recipients. It may contain important information protected by law, including the Unfair Competition Prevention and Trade Secret Protection Act. Any disclosure, copying, distribution, or secondary use of this document or any actions based on its contents require confirmation from SilicoPharm, Inc. Any secondary use without such confirmation is strictly prohibited. If you have received this document in error, please inform us and delete it immediately.

Contents

I . SilicoPharm, Inc.

II . Problem and Background

2.1 The Core of Modern Life Science – Omics

2.2 Challenges in Acquiring, Producing, and Analyzing Omics Data

III . Solution

3.1 Virtual Experimental Data Generation Using an Omics Foundation Model

3.2 Application Cases: Validation in Tumors and Neurological Disorders



Contents

IV. Scale-up

4.1 Market Characteristics and Current Landscape

4.1.1 A Life Science AI Market Where Conservatism and Innovation Coexist

4.1.2 Foundation Models and Analytics with Standardization and Generalizability

4.1.3 Markets Addressing Unmet Medical Needs such as ALS

4.2 Perspective & Milestone

V. Team

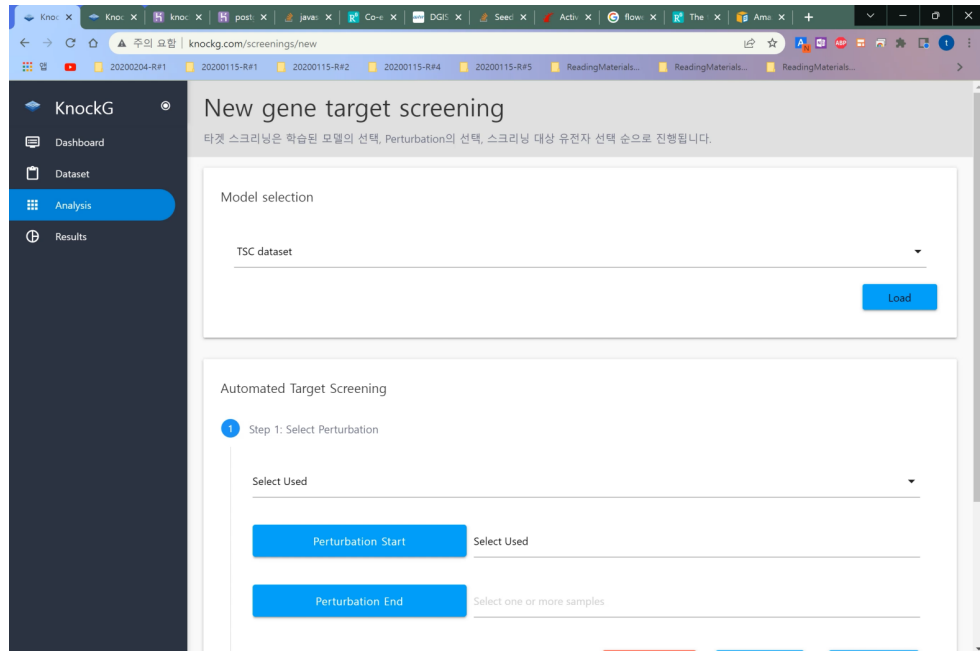
VI. Appendix



SilicoPharm, Inc.

AI company developing generative AI solutions that enable researchers to virtually simulate experimental results

- Core Technology: An AI technology that generates omics data to virtually produce experimental outcomes.
- Business Model: Primarily S/W licensing of the AI solution, with selective in-house development of drug candidates for technology transfer



Value

Omics generative solution

- (Research Productivity) virtual data without physical experiments
- (Exploratory Research) data unattainable by experimental methods

Collabo



...
Researcher & Pharma

BM

Solution sales

AI model licensing

Innovative joint research

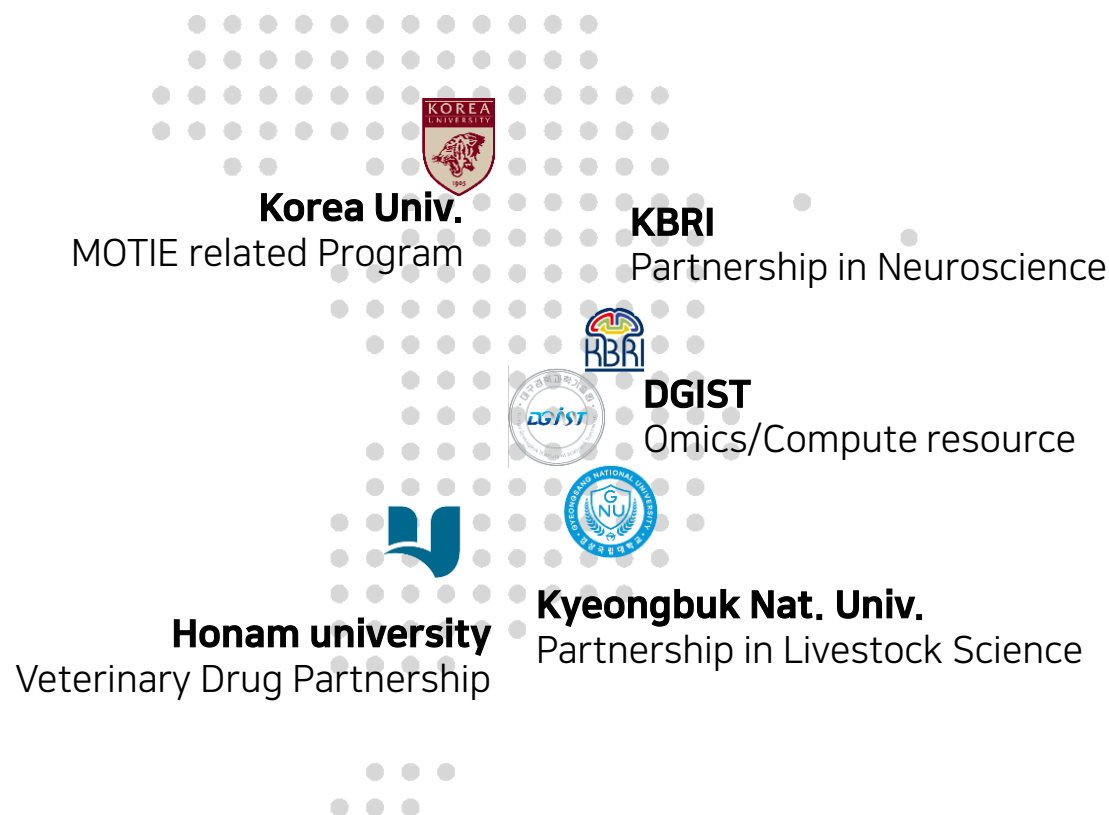
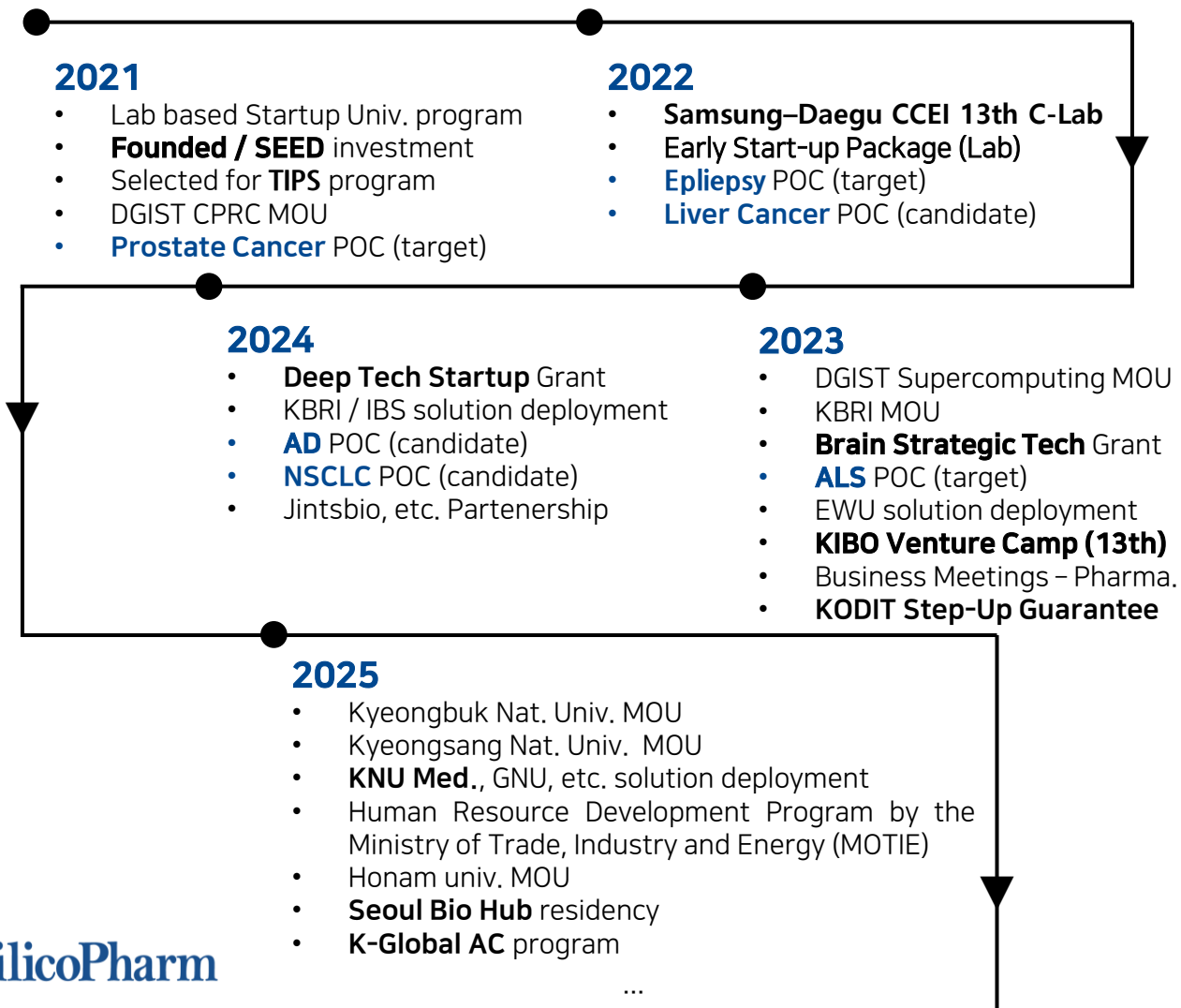
Equity in research outcomes

Screenshot of solution We aim to develop web-based(www.knockg.com) easy-to-use solution for researchers

Research Grant and collaborations Founding by graduate students of DGIST's Brain Science program in 2021, SilicoPharm has pursued diverse research projects using its software platform.

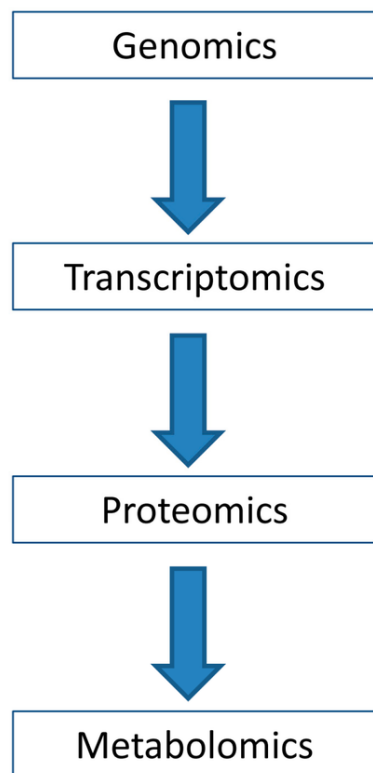
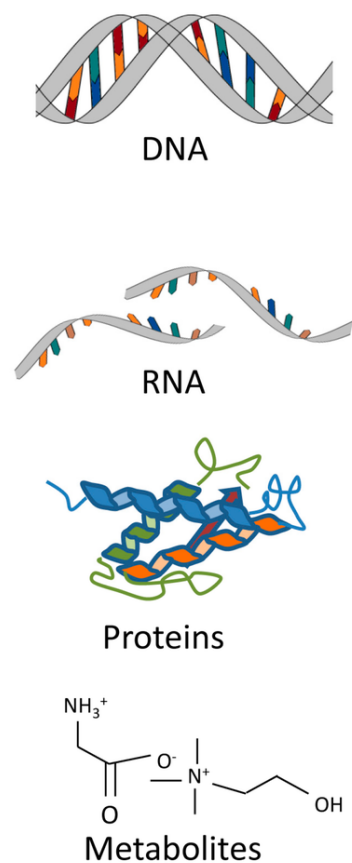
SilicoPharm, Inc.

Founded in July 2021 by Ph.D. students in Brain Science at DGIST, now led by the co-founders after completing their degrees



The Core of Modern Life Science – Omics

-Omics is a terminology for wholistic approach in life science, produce few thousands by tens thousand information



	A	B	C	D	E	F	G	H
1	gene	U937cell_Monocytesgenotype	U937cell_N	U937cell_N	U937cell_N	U937cell_N	U937cell_N	U937cell_N
2	ENSG00000149531.15	7.784208506	3.535688	7.034182	5.369058	7.356506	7.126199	6.288241
3	ENSG00000141759.15	36.21779575	30.31863	30.74413	28.93852	28.06538	24.21635	24.08808
4	ENSG00000015133.20	6.049473639	2.474681	2.680479	2.44967	5.700953	4.477414	5.391512
5	ENSG00000242797.3	0.293700026	0.4145	0.427591	0.734337	0	0.866369	0.448725
6	ENSG00000203791.15	1.860495533	1.717456	1.533199	2.106472	2.51619	1.691323	2.681631
7	ENSG00000164663.15	2.148227007	1.687264	1.921859	1.718786	2.3629	3.049075	3.120406
8	ENSG00000106459.15	8.281820215	11.05827	9.530329	9.274757	12.17999	9.654999	10.60751
9	ENSG00000254470.3	128.4139827	98.35276	104.4698	124.6536	105.4694	107.4766	108.739
10	ENSG00000175322.12	0.051756095	0.06087	0.062792	0.012941	0.109112	0.152672	0.065896
11	ENSG00000250988.7	3.879741083	4.380396	4.142178	4.656243	3.271708	3.051901	3.754147
12	ENSG00000090615.16	24.36556974	26.51655	25.46799	29.08228	26.0905	21.39706	24.27025
13	ENSG00000265393.1	0	0	0.586567	0.604416	0.637039	0	0
14	ENSG00000249459.11	0.11910692	0.149419	0.192672	0.119121	0.125551	0.078077	0.202195
15	ENSG00000097021.20	108.9042214	87.1993	73.9792	77.86468	53.89134	48.77864	47.67518
16	ENSG00000226822.4	0.294405527	0.36933	0.428618	0.637954	0.362055	0.289483	0.399825
17	ENSG00000198010.13	0.013501776	0.025407	0.039314	0.013503	0.014232	0.013276	0.068762
18	ENSG00000165527.8	82.14917518	86.13742	95.19059	98.96067	124.3934	112.5	120.3559
19	ENSG00000108506.13	6.01416083	3.876323	5.170794	6.393766	6.813743	6.728479	6.825157
20	ENSG00000252274.1	2.982019971	1.402846	2.894302	1.491187	1.571673	0	1.518677
21	ENSG00000000000.16	100.2206066	202.0552	107.0066	212.1067	164.1302	100.1502	172.7070

Major omics domain(left) and example of omics data(right) -Omics such as genomics, transcriptomics, proteomics, metabolomics produce high-dimensional data – Single-cell transcriptome which includes few thousand gene expression lv of each tens thousand cells.

Challenges in Acquiring, Producing, and Analyzing Omics Data

Omics enable comprehensive insights, three barriers continue to drive unmet medical needs and conservatism in market

Acquisition of clinical samples

Obtaining patient-derived samples is a problem that cannot be solved simply with time and cost. In certain diseases, such as neurological disorders, it may be fundamentally impossible to acquire clinical samples

Condition-specific generation

At both the sample acquisition stage (e.g., patient genotype) and during processing of acquired samples (e.g., target knockdown, drug administration), countless conditions exist. It is infeasible to test all of them through real experiments.

High-dimensional data analysis

Omics data consist of thousands of gene expression values for each of tens of thousands of cells, making them impossible to interpret or utilize without mathematical and computational methods.

Virtual Experimental Data Generation Using an Omics Foundation Model

Omics foundation model enables generation of unseen/artificial omics data – like other foundation models

Model Name	Input	Output	Usage
GPT (ChatGPT)	Natural language data	Natural language data (Generated)	Text generation, translation, question answering, etc.
Alphafold	Amino acids sequence	Protein Structure	Drug development, functional protein design, etc.
KnockG	Experimental condition	Omics data (Generated)	Drug development, precision medicine, diagnostics, and therapeutics, etc.

Virtual Experimental Data Generation Using an Omics Foundation Model

We have developed its AI solution as a web-based application (KnockG) - any researcher can easily access and utilize it

The screenshot shows the KnockG web application interface. The browser address bar displays 'knockg.com/screenings/new'. The left sidebar contains navigation links: Dashboard, Dataset, Analysis (highlighted), and Results. The main content area is titled 'New gene target screening' and includes a sub-header '타겟 스크리닝은 학습된 모델의 선택, Perturbation의 선택, 스크리닝 대상 유전자 선택 순으로 진행됩니다.' Below this, there is a 'Model selection' section with a dropdown menu set to 'TSC dataset' and a 'Load' button. The 'Automated Target Screening' section shows 'Step 1: Select Perturbation' with a 'Select Used' dropdown. Below this are two buttons: 'Perturbation Start' and 'Perturbation End', each followed by a 'Select one or more samples' dropdown.

**생명과학 연구자를 위한
오믹스 데이터 생성 인공지능 서비스**

무료버전 체험 및 선사용 후 결제 가능

실험 없이 실험 결과에 해당하는
오믹스 데이터를 예측·생성하여 제공

- 타겟 녹다운 및 활성화, 약물 처리 실험 등
- 30분 내 2만개의 데이터 생성
- 타겟의 문헌, 약물, 특허 정보 연동

생성 인공지능을 활용하여
연구과정을 혁신하세요

실리코팜은 생명과학과 인공지능을 연결하는 기업으로,
연구자가 손쉽게 사용할 수 있는 인공지능 솔루션으로
생명과학 연구를 가속화합니다.



lite.knockg.com

무료 버전 솔루션



www.knockg.com

상세조정 및 신규모델 학습

핵심 기술 및 서비스

생성 인공지능 기반의 타겟 녹다운 시뮬레이션으로 연구자가 필요한 솔루션 제공

시계열 데이터 생성

t_0 , t_1 시점의 데이터로 t_2 , t_3 의 중간 시점
데이터나 t_0 , t_1 등등의 미래시점 데이터 생성
이 가능합니다.

타겟 녹다운/활성화 데이터 생성

특정 유전자/단백질 타겟이 녹다운되거나
활성화 되었을 때의 데이터를 생성하여 다양
한 분석에 활용할 수 있습니다.

신규 타겟 및 후보물질 탐색

생성 데이터 간 비교/분석을 통해 특정 질환이
나 생명현상의 신규 타겟 탐색이 가능하며, 타
겟의 문헌, 약물, 특허 정보가 제공됩니다.

특장점

사용성	통합성	낮은 학습 데이터
Web 어플리케이션 형태로 누구나 손쉽게 사용	생성 데이터를 실제 실험결과 1:1 비교/검증	저비용으로 시작, 회귀/차이 신경망 적용가능
타겟 연구에서 실제 도입가능한 생성 인공지능 솔루션		

세 가지 특장점을 바탕으로 중앙, 뇌 신경, 회귀질환
관련 다양한 연구 협력을 진행하고 있습니다.

사용방법

Lite.KnockG	KnockG
회원가입/로그인	회원가입/로그인
생성 모델 선택	생성 모델 선택
타겟 선택	타겟 선택
타겟 녹다운/활성화	타겟 녹다운/활성화
타겟 선택	타겟 선택
타겟 녹다운/활성화	타겟 녹다운/활성화
Pseudo-blotting 분석	Pseudo-blotting 분석
Pseudo-cytometry	Pseudo-cytometry

무료로 체험/분석 후 결제 가능합니다

실리코팜 주식회사 대구광역시 달성군 DGIST 산학협력관

Email contact@silicopharm.com

Website www.silicopharm.com / www.knockg.com

Application Cases: Validation in Tumors and Neurological Disorders

We have completed data generation and experimental validation in liver cancer, epilepsy, Alzheimer's disease, ALS, and NSCLC

Disease Name	Data Generation & Derivation	Efficacy Verification	Results & Current Status
Liver Cancer	Generate knock-down data ↓ New target /drug candidate	K-medihub Preclinical Center	- New target / repurposed cardiovascular drug - Own cell-based experiments → Project suspended due to newly arising issues
Epilepsy		Zefit	- New target / repurposed drug - CRO efficacy evaluation confirmed effectiveness
Alzheimer's Disease			- New target / repurposed cardiovascular drug - Efficacy evaluation confirmed one valid candidate → Exploring related technology transfer
ALS (Lou Gehrig's Disease)	Generate condition specific Knock-down data ↓ Disease mechanisms and therapeutic discovery	Korea Brain Research Institute	- Seven new targets identified - Based on effective model results, 12 candidate substances s elected → Exploring related technology transfer
Non-Small Cell Lung Cancer (NSCLC)	Generate drug combination data ↓ Drug which enhance efficacy of the primary drug	K-medihub Preclinical Center	- Candidate identified for new target-based treatment for erlotinib-resistant patients - Multiple target knockdown simulations identified one valid tr eatment regimen candidate → Exploring related technology transfer

A Life Science AI Market Where Conservatism and Innovation Coexist

Ultimately, the key is not performance scores, but the ability to generate data that users can tangibly perceive and value.

Metric	AlphaFold	Insilico Medicine	KnockG (H1 2025)	KnockG (H2 2025)
Reproducibility & Scalability	High Numerous validation cases; 1:1 comparison possible with actual results	High Dozens of clinical pipelines	Medium Several lab-level validations, 1:1 verification	High Clinical pipeline derivation, etc.
Third-Party Usability	High Partial code released and public database available	Low No solution available with code release or usability	High Web-based solution provided	High Web-based solution provided
Effectiveness & Benefit	High Visually understandable results; scalability from tens to hundreds of existing processes	Low Effective only for pharmaceutical companies, not for academic researchers	Low Limited academic outputs / automation	High Numerous publications; improved automation and diverse visualization features
Pioneering Use Cases	High Awarded Nobel Prize in 2024	High Lead compound/Phase 1-2 candidate derivation requires 9-18 months	Low Limited institutional POC	High Collaborations with pharma and research institutions

Foundation Models and Analytics with Standardization and Generalizability

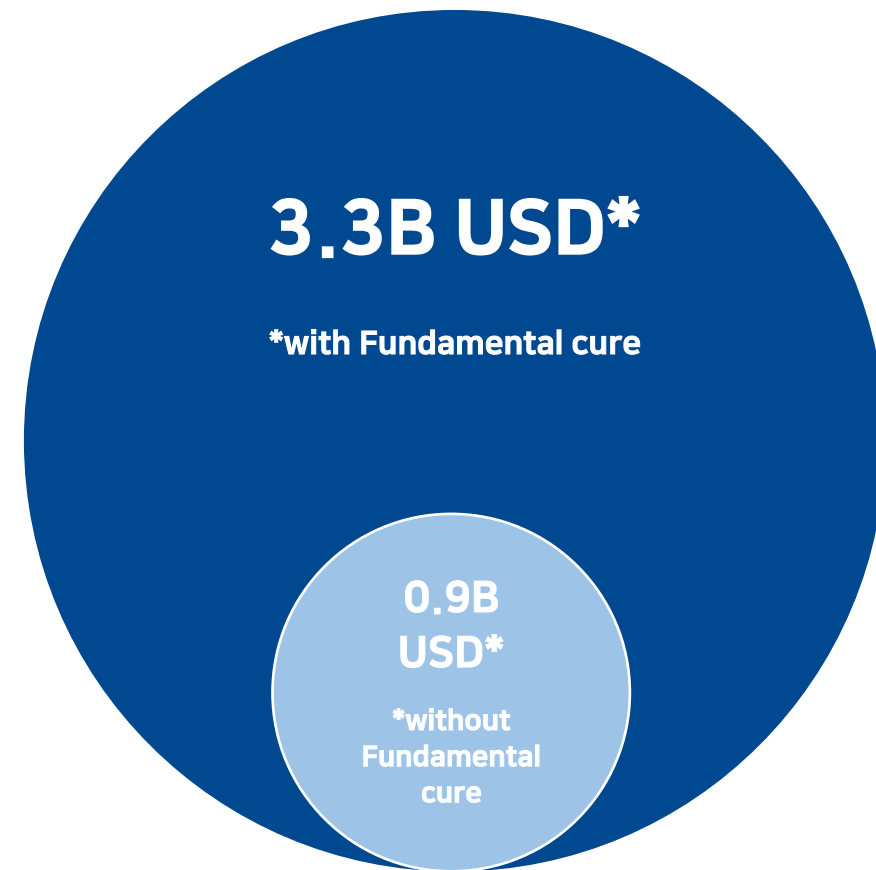
Standardized and generalizable foundation model accelerate diversification and application across various domains

Model	Data generation and application
GPT (ChatGPT)	Generate text that maintains contextual coherence to respond to user queries, serving as Chatbot
	Draft text content tailored to various domains, such as marketing, advertising, articles, and website content
	Translate user-input text into other languages (language translation)
Alphafold	Generate protein structural data to facilitate research on protein structure and function
	Generate protein-drug binding structure data to predict protein-drug interactions and study protein function
KnockG	Generate and compare target knockdown data to discover novel disease targets or candidate compounds
	... Organ- or tissue-specific data to predict potential drug toxicity
	... Quantitative omics data on dose-response relationships to determine optimal drug dosages
	... cohort-specific omics data responsive to a given drug to identify optimal clinical trial cohorts before initiation

Markets Addressing Unmet Medical Needs such as ALS

We aim to address previously untapped unmet medical needs and realize potential markets

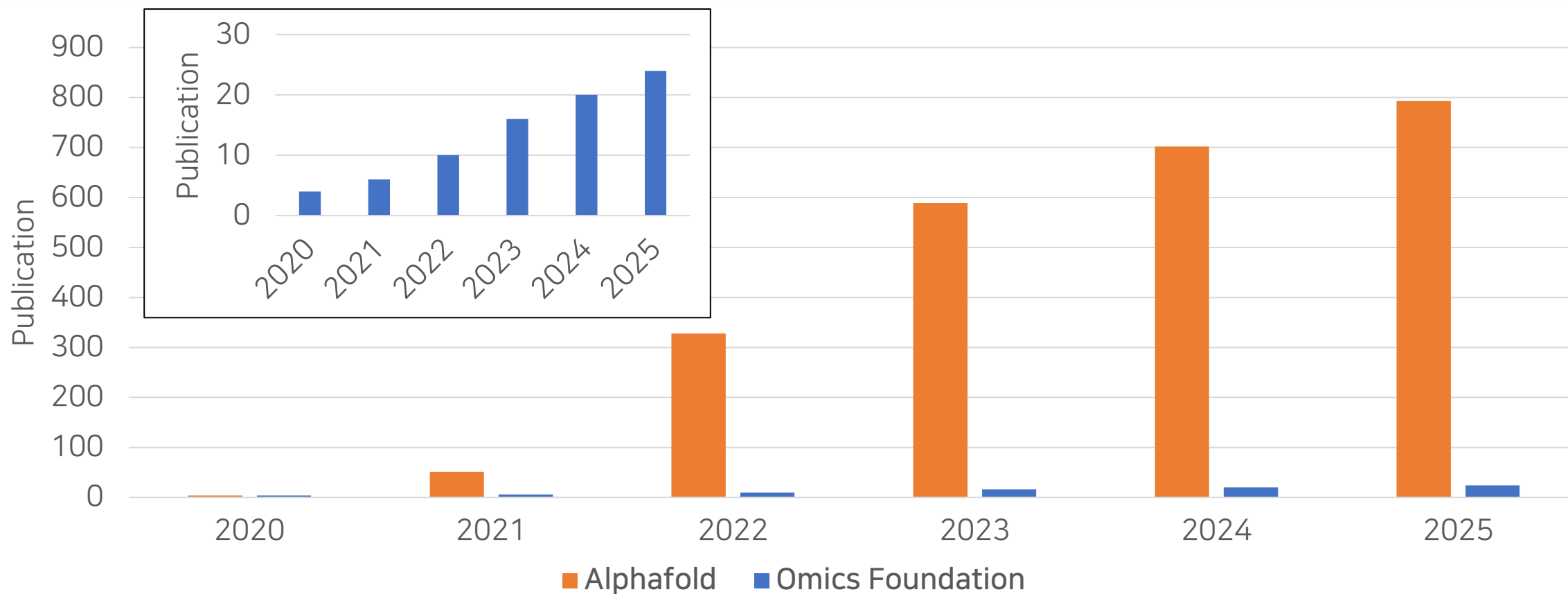
Item	Details
Target Disease	ALS (Amyotrophic Lateral Sclerosis) and motor neuron diseases
Collaborating Institutions	Korea Brain Research Institute, Kyungpook National University School of Medicine
Problem & Background	The cause of ALS remains unclear, and there is currently no fundamental treatment. Patients rely on a few drugs that can only alleviate certain symptoms, and as the disease progresses, it leads to severe social costs due to the loss of motor function.
Solution	Utilizing an omics foundation model to generate condition-specific virtual data (e.g., motor neuron transcriptome/protein datasets), which cannot be obtained through actual experiments, and conducting fundamental research on the pathogenesis of ALS using the generated data.
Outcomes & Current Status	- Selected and conducted as a national project (Brain Industry Strategic Technology Development Program) - New causative genes of ALS identified, with subsequent research in progress



Addressing unmet medical needs (ALS therapeutics) In collaboration with the Korea Brain Research Institute, we are conducting research that was previously impossible by conventional methods. Through this, we aim to target the fundamental therapeutic market for ALS, which has remained only a potential market until now.

Perspective & Milestone

There has been no product in the omics foundation model that truly achieved explosive growth, unlike Alphafold



An NCBI-based analysis of yearly publications Unlike AlphaFold which experienced explosive growth and became a standard, omics foundation models—despite continuous publications—have not yet demonstrated such momentum.

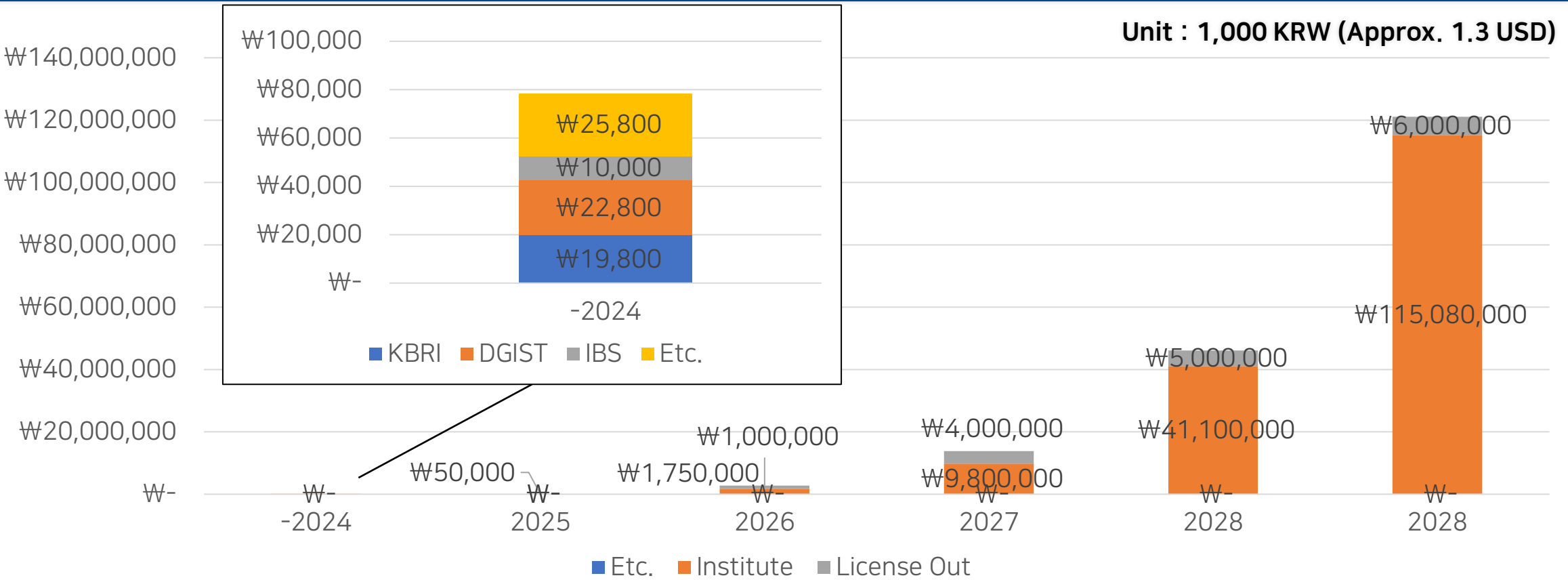
Perspective & Milestone

Targeting explosive growth within 3 months after the free version launch in late 2025

	Sep. 2021	1H 2025	1H 2026	-2027
Key Goal	-	Proof of concept & initial business	Scalable structure & Explosive user growth via free version	Market leadership in omics foundation models
Milestone	• SEED Funding • Tech validation • TIPS program	• Lab-level tests, 3rd-party validation • Product development, init users acquiring • Pro foundation model (0.5B-1B scale)	• Pre-Series A funding • Global entry (US pharma partners) • mid-scale foundation model (5B-10B) • 3 months rapid user growth	• Top-tier partnerships • Clinical approvals • Large foundation model
Valuation	KRW 2B (~\$2M)	KRW 15B (~\$15M)	KRW 60B (~\$60M)	KRW 240B (~\$240M)
Notes	-	• ~KRW 1.6B funding secured • >12 public validations • Early revenue • Partnerships (KBRI, Kyungpook Univ., etc.) • Exclusive datasets & Validation results ...	• Target 25 institutional clients • <u>Global support projects (K-Global AC)*</u> • <u>Compute resource partners (Oracle)*</u> • Complement metric (6 publications, etc.) • Scalable KPIs achieved ...	-

Perspective & Milestone

Starting in 2026 with solution deployment to 25 institutions, we are planning for rapid, large-scale growth



Revenue projection Based on projected adoption and the expected growth curve of foundation models, we anticipate entering a phase of explosive growth starting in 2026.

Team Status

Key Executives & Team



Sangho Ji (CEO)

2018-2025 Ph.D. in Brain Science, DGIST
2017-2018 Post-Master Researcher at CPRC
2016-2018 M.S. in Medical Science, KNU School of Medicine;

Key Experience & Skills

- Life science and protein property experiments with data analysis
- mass spectrometer operation and data processing;
- molecular dynamics simulation



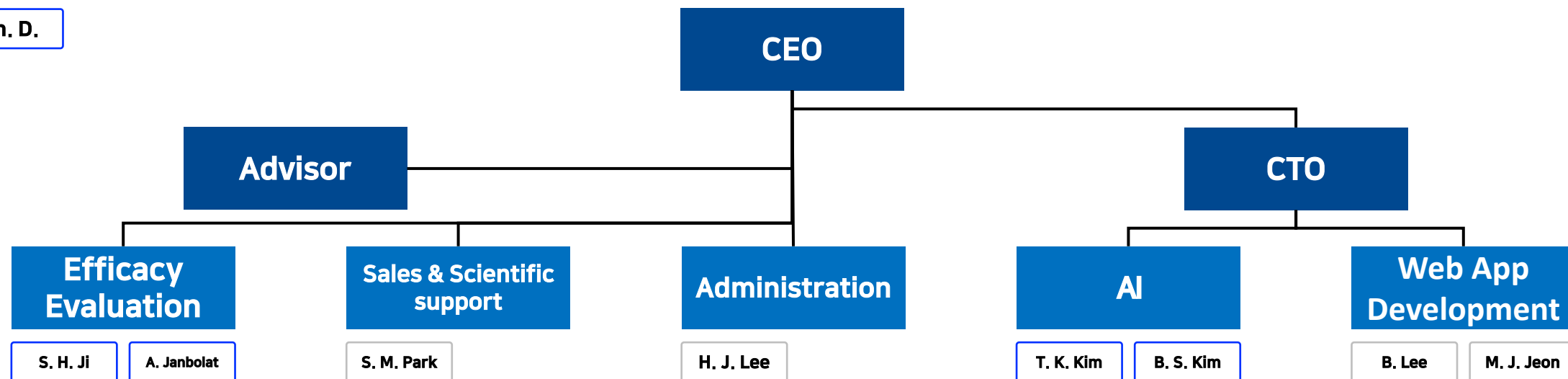
Taehyeong Kim (CTO)

2020-2025 Integrated M.S./Ph.D. program in Brain Science, DGIST
2016-2018 Leader of "LikeLion" program at JBNU
2015-2019 B.S. in Biological Sciences, JBNU

Key Experience & Skills

- Life science and omics data analysis;
- leveraging AI in life science research;
- Web application development and server/database

Ph. D.



Team Status

Advisors



Wookyung Yu



2016-Present Assistant/Associate Professor of Brain Science, DGIST
 2017-Present Deputy Director (joint appointment) of Core Protein Resources Center
 2016-Present Deputy Director (joint) of Supercomputing and Big Data Center ...
 2012-2016 Postdoctoral Fellow at University of Chicago

11 first-author SCI papers

Computational techniques for transcriptome and proteomics analysis



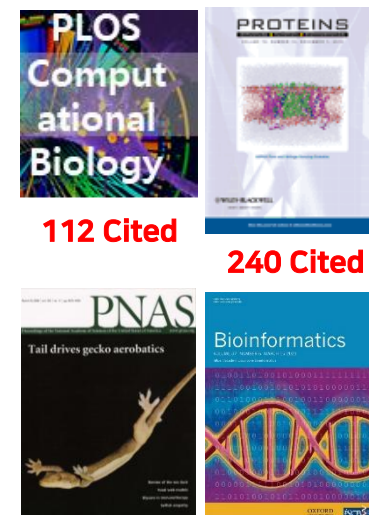
Mookyung Cheon



2016-Present Senior Researcher, Korea Brain Research Institute (KBRI)
 2015-2016 Research Professor, DGIST Brain & Cognitive Science
 2013-2014 Research Professor, DGIST New Biology
 2009-2012 Research Professor, Pusan National University
 2007-2009 Postdoctoral Fellow, North Carolina State University
 2005-2007 Postdoctoral Fellow, University of Cambridge
 2003-2005 Postdoctoral Fellow, Pusan National University

17 first-author SCI papers

Machine learning and computational analysis techniques



112 Cited

240 Cited