

Expanding AI-Driven Drug Discovery: Virtual Screening of Natural Compounds for Enhanced Cultured Meat Production

Hyunseung Kong*, Inyoung Kim**

*Student, Interdisciplinary Program in Bioinformatics, Seoul National University, Seoul, Korea

**Research Professor, Artificial Intelligence Institute of Seoul National University (AIIS), Seoul, Korea

[Abstract]

In this study, we propose a methodology for deriving natural additives that expedite cultured meat production through a combination of bioinformatics and artificial intelligence-based approaches in drug development. Leveraging a bioinformatics approach, we identify differentially expressed genes that are genetically associated with muscle and adipose tissue growth in breast cancer cells. Subsequently, employing an artificial intelligence-based drug development approach, we construct a model for predicting protein and compound interactions, and conduct virtual screenings of two natural product databases (COCONUT and the Korean Traditional Medicine Database). This enables us to identify 22 candidates of natural compounds that bind to proteins responsible for inhibiting muscle and adipose tissue growth.

▶ **Key words:** Cultured meat, artificial intelligence, natural products, bioinformatics, drug discovery

[요 약]

본 연구에서는 생물정보학과 인공지능 기반의 신약 개발 접근법을 활용하여 배양육 생산을 가속화하기 위한 천연물 첨가제를 도출하는 방법을 제안한다. 생물정보학적 접근으로 지방암 세포 대비 정상 세포에서 높게 발현되는 차등 발현 유전자들을 선별하고, 이를 근육 조직 및 지방 조직 성장을 저해하는 유전자 후보로 정의하였다. 인공지능 기반의 신약 개발 접근법을 이용하여 단백질과 화합물의 결합 예측 모델을 구성하였고, 두 개의 천연물 데이터베이스(COCONUT, 한국한의학진흥원 천연물 물질 데이터베이스)를 가상 스크리닝하여 근육 조직과 지방 조직의 성장을 저해하는 단백질들과 잠재적으로 결합하는 천연물 후보 22종을 도출하였다.

▶ **주제어:** 배양육, 인공지능, 천연물, 생물정보학, 신약 개발

-
- First Author: Hyunseung Kong, Corresponding Author: Inyoung Kim
 - *Hyunseung Kong (hskong@snu.ac.kr), Interdisciplinary Program in Bioinformatics, Seoul National University
 - **Inyoung Kim (iykim@snu.ac.kr), Artificial Intelligence Institute of Seoul National University (AIIS)
 - Received: 2023. 06. 02, Revised: 2023. 06. 13, Accepted: 2023. 06. 21.

I. Introduction

계속되는 인구 증가와 기후 급변화로 인해, 현재 식량 위기가 사회적인 문제로 불거지고 있다. 이에, 식량 위기 문제를 해결하기 위한 해결책들이 여러 방향으로 연구 및 제안되고 있으며, 그중 육류를 대체하기 위한 수단으로 더 적은 에너지로 육류를 생산할 수 있는 배양육이 대안으로 부상하고 있다. 배양육 생산은 동물 세포를 배양하여 육류를 생산하여 기존의 고기 생산에 따르는 환경, 윤리 및 자원 관련 문제를 해결할 수 있는 접근방식으로, 근육세포와 지방세포를 배양하여 인공적으로 고기를 만드는 과정이다 [1]. 이러한 배양육 기술은 동물 학대 없이 고기를 생산할 수 있는 잠재력을 가지고 있으며, 친환경적인 생산 방법으로 알려져 있다.

배양육은 일반적으로 다음의 과정을 거쳐 생산된다. 1) 동물 세포가 필요로 하는 영양소와 환경 조건을 최적화한 배양액에서 근육세포와 지방세포가 각기 배양된다[2]. 이러한 배양액에는 세포의 증식과 성장을 지원하기 위해 필수적인 영양소와 성장인자(growth factor) 등이 포함된다. 2) 세포는 이러한 배양액에서 적절한 온도와 산소 공급 등의 환경 조건에서 증식하고 분화한다. 3) 기존 고기와 유사한 식감을 가지도록 근육세포와 지방세포 등을 조합한다.

배양육에 사용되는 동물 유래 세포로는 동물에서 직접 채취하고 유한한 분열능을 가지는 세포인 1차 세포(primary cell)와 유전자 조작이나 돌연변이를 통해 분열에 제한이 없는 세포주(cell line)가 고려될 수 있으나, 일반적으로 배양육 생산에는 1차 세포가 이용된다. 배양 과정에 동물 유래 세포로 세포주가 이용된다면 지속적으로 세포를 채취할 필요가 없어 배양이 용이하지만 암세포의 특징을 지니기 때문에 안전상의 이유로 거의 사용되지 않는다. 하지만, 1차 세포는 유한한 생존 기간과 분열 능력을 가지고 있어 지속적인 대량 생산에는 제한이 있어 배양육 생산 비용 증가의 원인이 되고 있다[3]. 이에 따라 최근에는 1차 세포의 단점을 보완하기 위해 성장과 분열을 가속화 하기 위한 연구들이 이루어지고 있다[4].

컴퓨터과학의 발전과 함께 의학 분야에서는 생물정보학(bioinformatics)과 인공지능 기반 신약 개발(AI-driven drug discovery)이 융합되어 특정 질병의 치료하기 위한 치료제를 도출하려는 시도가 있다. 생물정보학적 접근으로 정상 세포와 특정 질병에 걸린 세포 사이의 차등 발현 유전자(differentially expressed genes, DEGs)를 도출하여 질병 특이적 단백질이 식별될 수 있다 [5]. 그리고 기계학습을 이용한 접근으로 단백질과 화합물

의 결합에 대한 약물-타겟 상호작용(drug-target Interaction)에 대한 분류모델을 만들어 잠재적인 약물 후보(potential hits)들을 도출할 수 있다[6,7,8,9].

본 논문에서는 생물정보학과 인공지능 기반 신약 개발의 접근법을 이용하여 배양육 생산을 가속화 하는 천연물 첨가제(natural additives)를 도출하는 방법을 제안한다. 본 연구의 목적은 근육 조직과 지방 조직의 분열을 저해하는 유전자들을 선별하고, 선별된 유전자들을 억제하여 세포 분열을 가속화하는 천연물을 가상 스크리닝하는 것이다. 지방 조직과 근육 조직에서 세포주기 진행을 억제하는 유전자를 찾기 위해 유방암(breast cancer)에서 낮은 발현량을 보이는 차등 발현 유전자들을 선별하였다. 그 후 딥러닝 기반의 화합물-단백질 상호작용 모델을 이용하여 천연물 데이터베이스에서 타겟 단백질들과 결합할 수 있는 천연물을 가상 스크리닝하였다. 천연물은 오랜 기간 식품으로 사용되며 안전성이 일정 부분 입증되었기 때문에 인공적으로 합성된 화합물에 비해 상대적으로 소비자들의 거부감이 적을 것으로 예상된다.

본 논문의 구성은 다음과 같다. 2장에서는 관련 연구와 이론적 배경에 대해 설명한다. 3장에서는 본 논문에서 사용된 데이터 세트와 제안하는 파이프라인으로 도출된 천연물들에 대해 설명하고, 4장에서는 본 논문의 결론과 한계점을 다룬다.

II. Preliminaries

배양육은 세포 배양을 통해 생성된 고기라는 뜻으로, 2013년 등장한 최초의 배양육은 동물에서 얻은 1차 세포를 사용하여 제조되었다[1, 2]. 하지만 1차 세포는 배양이 어렵고 분열 횟수가 한정되어 지속적으로 동물에서 세포를 채취해야 하기 때문에[3], 최근에는 자연 불멸화 세포(spontaneously immortalized cell)와 같이 오랜 기간 분열할 수 있는 세포가 이용되고 있다. 자연 불멸화 세포는 유전자 돌연변이나 유전자 조작 없이 자연적으로 분열 횟수가 제한되지 않고 계속해서 증식할 수 있는 세포로, 동물 세포의 한계를 극복할 수 있는 가능성을 보여준다. 동물 세포의 경우 분열을 위해 성장인자를 필요로 한다. 때문에 세포 배양 시 소태아혈청(fetal bovine serum, FBS)이나 성장인자를 이용한다[4].

여성의 유방은 지방과 근육으로 이루어진 기관으로 주로 임신 및 수유 과정에서 중요한 역할을 한다[10]. 유방암은 유방 조직에서 발생하는 암으로, 여성에서 가장 흔하게

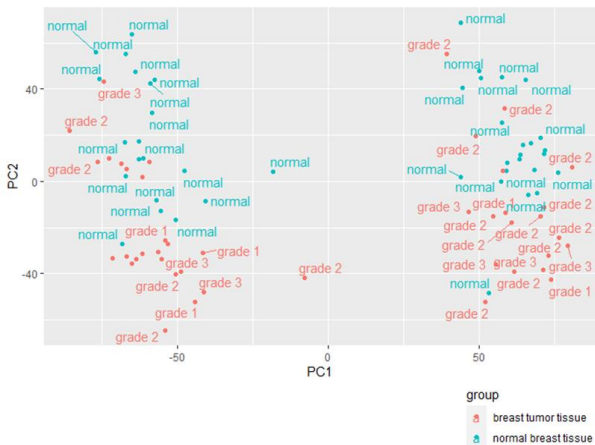


Fig. 1. PCA analysis of the breast tissue samples

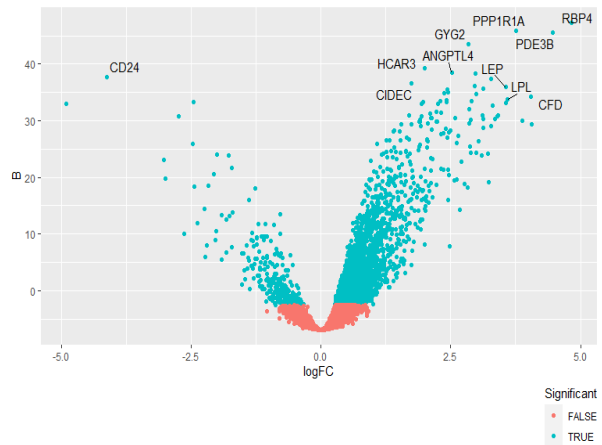


Fig. 2. Volcano plot of the DEGs

발생하는 종류의 암으로 알려져 있다[11]. 유방암은 주로 유방 조직에서 발생하고, 유방 조직은 일부 근육과 유방선, 지방 조직으로 구성된다. 즉, 정상 조직과 유방암 조직의 차등 발현 유전자를 비교함으로써 근육 조직과 지방 조직의 증식과 관련된 유전자들을 도출할 수 있다.

생물정보학은 생물학과 컴퓨터과학의 융합 분야로서, 생물학적인 데이터를 수집, 저장, 관리, 분석하는 기술과 방법을 연구하는 학문이다. 이를 통해 유전자, 단백질, 대사 네트워크 등 다양한 생물학적인 정보를 이해하고, 생물학적인 현상과 질병 등에 대한 통찰력을 얻을 수 있다. 생물정보학에서 중요한 주제 중 하나는 유전자 발현 데이터를 분석하여 차등 발현 유전자를 도출하는 것이다. 차등 발현 유전자는 특정 조건 또는 집단 간에 유전자 발현량의 차이를 보이는 유전자들을 의미한다. 이를 통해 특정 질병의 발생 메커니즘을 이해하거나, 약물의 효능을 평가하는 등의 응용이 가능하다. 본 논문의 분석에 사용한 Limma(Linear Models for Microarray and RNA-Seq Data) 패키지[5]는 고차원 생물학 데이터를 분석하는 데 사용되는 R 패키지이다. Limma는 특히 마이크로어레이(microarray)와 RNA-Seq 데이터를 분석하는 데에 특화되어 있다. Limma를 사용하면 다른 조건이나 그룹 간에 유전자 발현의 차이를 비교하는 차등 발현 유전자 분석을 수행할 수 있다. 차등 발현 유전자 분석은 바이오마커 발견, 질병 메커니즘 연구, 약물 효과 평가 등 다양한 분야에 이용될 수 있다.

인공지능의 발전은 신약 개발 분야에 많은 영감을 주었다. 기존의 신약 개발 과정은 매우 복잡하고 시간과 비용이 많이 소요되는 작업이었다. 인공지능을 통해 연구자들은 축적된 실험 데이터들을 이용하여 화합물의 효능과 독성의 숨겨진 패턴을 학습하여 신약 개발을 가속화할 수 있

게 되었다. 신약 개발의 초기 단계는 타겟 단백질에 결합하여 기능을 저해하는 분자를 찾는 초기 후보물질 발굴(hit discovery) 단계이다. 이 과정에서 단백질과 화합물은 그래프[7], 문자열[8], 이미지, 특징 벡터[9] 등으로 표현될 수 있다. 예를 들면, 단백질은 20종류의 아미노산의 조합으로 표현되는 문자열로 간주할 수 있고, 각 아미노산 잔기의 거리는 인접행렬 이미지로 간주될 수 있다. 이러한 다양한 표현 방법을 이용하여 단백질과 화합물 사이의 상호작용을 예측하는 인공지능 모델들이 개발되고 있다. 예를 들면, 단백질-화합물 상호작용 예측 모델(protein-compound interaction prediction model)은 입력으로 단백질과 화합물의 표현을 받아 그들 간의 상호작용 가능성을 예측한다. 이때, 단백질과 화합물을 문자열이나 이미지로 가정하여 자연어처리(natural language process, NLP), 컴퓨터 비전(computer vision, CV)에 사용되는 딥러닝 기술들을 이용할 수 있다. 단백질과 화합물의 상호작용은 결합 여부만을 추론하는 이진분류(binary classification) 모델로 구성하거나, 해리상수(Kd), pIC50, pEC50 등의 실험값들을 추론하는 회귀(regression) 모델로 구성한다[12]. O. Kadioglu et al.[13]은 단백질-화합물 상호작용 예측 모델을 이용한 가상 스크리닝과 MD(molecular dynamics) simulation을 통해 잠재적으로 SARS-CoV-2를 억제할 수 있는 천연물을 도출하였다. L. A. Machado et al.[14]는 단백질-화합물 상호작용 예측 모델을 이용한 가상 스크리닝을 통해 잠재적으로 HIV-1 integrase를 억제할 수 있는 천연물을 도출하였다. 이렇게 단백질-화합물 상호작용 예측 모델을 통한 천연물 가상 스크리닝이 다양하게 적용되고 있지만, 배양육 생산 분야에 적용된 사례는 없다.

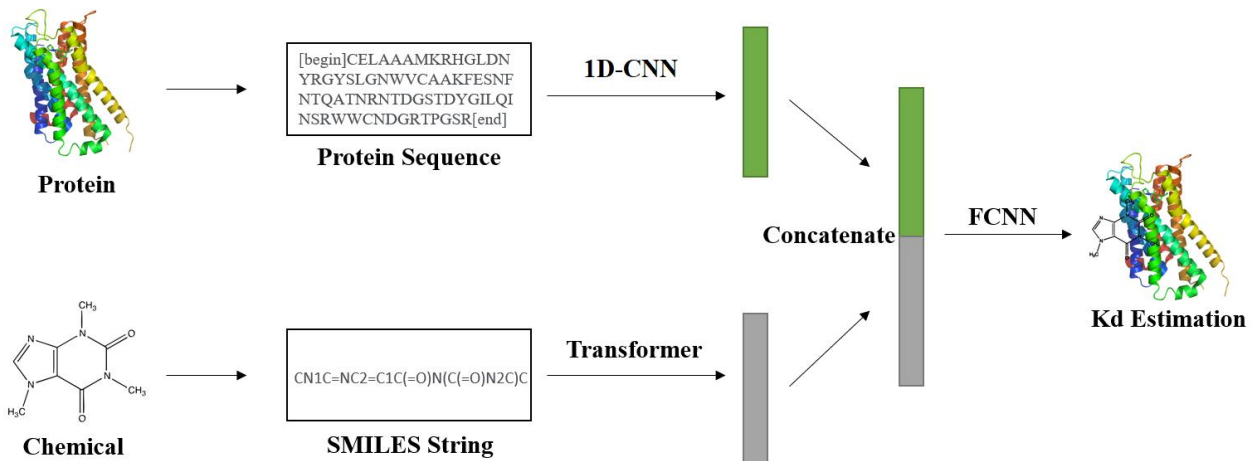


Fig. 3. The structure of the proposed model

III. Experiment and Results

1. Datasets

본 논문에서는 인간의 정상 유방 조직과 유방암 조직의 차등 발현 유전자를 도출하기 위해 GEO 데이터베이스의 GSE15852[15]를 이용하였다. GSE15852는 유방 종양과 정상조직 간의 유전자 발현 차이를 확인하기 위해 수행된 연구에서 생산된 데이터이며 Affymetrix genechip U133A를 사용하여 43개의 종양 샘플과 43개의 정상조직 샘플의 유전자 발현량 데이터를 포함한다. 이 연구는 말레이시아의 다민족 유방암 환자(말레이시아인, 중국인, 인도인)를 대상으로 수행되었다.

단백질-화합물 상호작용 예측 모델 학습을 위해 사용한 데이터세트는 BindingDB[16]이다. BindingDB는 약물과 단백질, 핵산 등의 생체 분자 간의 결합에 대한 실험 결과와 관련 정보를 제공하는 데이터세트이다. BindingDB에는 수천 개의 약물과 수백만 개의 단백질 간의 상호작용 데이터를 포함한다. 이 데이터는 실험 결과로부터 얻은 약물-단백질 결합 상태, 결합 상수, 억제제 등의 정보를 담고 있다. 본 논문에서 사용한 데이터베이스 버전은 202305버전으로 1,164,246개 화합물과 9,023개 단백질의 2,707,699개 측정값을 포함한다.

천연물을 도출하기 위해 COCONUT 데이터베이스[17]와 한국한의약진흥원 천연물 물질 데이터베이스[18]를 이용하였다. COCONUT 데이터베이스는 천연물 화합물의 구조, 화학식, 화학 클래스, 생물 활성 등 다양한 정보를 제공하는 데이터베이스로 2012년부터 개발되었으며 SMILES 형식으로 화합물의 구조를 제공한다. 한국한의약진흥원 천연물 물질 데이터베이스는 한국한의약진흥원에

서 생산 및 도출된 실험정보 기반의 정형데이터로, 990개의 생약에서 추출된 화합물들에 대한 정보를 제공한다.

2. Differential gene expression analysis

본 논문에서는 R의 Bioconductor Limma 패키지[5]를 사용하여 정상조직과 유방암 조직 데이터(GSE15852)의 차등 발현 유전자를 분석하였다. Figure 1은 43개의 정상조직들과 43개의 유방암 조직들의 주성분 분석(Principal Component Analysis, PCA) 결과를 보여준다. 유방암 조직의 유전자 발현 패턴은 암의 진행단계에 따라 명확히 구분되지는 않지만, 정상 조직과 유방암 조직의 유전자 발현 패턴은 크게 네 개의 클러스터(상단 정상 조직 클러스터 2개, 하단 유방암 조직 클러스터 2개)로 구분되는 것을 보인다. Figure 2는 정상 조직 대비 유방암 조직에서 낮게 발현된 유전자를 식별하기 위해 최소 4배($p < 0.01$, $|\log FC| > 2$)로 하향 조절된 82개 유전자에 대한 volcano plot이고, 변화 절대치가 큰 상위 10개의 유전자를 표시하였다. 유방암 조직 대비 정상 조직에서 높게 발현되는 유전자를 선별하기 위해 유방암 조직 대비 정상 조직에서 발현량이 높은 10개 유전자($\log FC > 2$)들을 선별하였다. Table 1은 이 유전자들의 심볼과 $\log FC$, adj.P.value를 보여준다.

Table 1. TOP 10 genes of DEGs

Symbol	logFC	adj.P.Val
RBP4	4.83	2.96E-21
PDE3B	4.47	5.67E-21
HBB	4.06	4.99E-15
CFD	4.05	1.07E-16
SAA1	3.88	3.08E-15
PPP1R1A	3.76	5.67E-21
LPL	3.59	1.64E-16
SORBS1	3.57	2.20E-16
LEP	3.56	2.83E-17
DPT	3.41	1.42E-15

Table 2. Hyperparameters of the proposed model

Hyperparameter	Value
Epochs	100
Learning rate	0.001
Batch size	128
CLS hidden dimension	1,024
	1,024
	512
CNN filters	32
	64
	96
CNN kernels	4
	8
	12

3. Model training

단백질-화합물 결합력 예측 모델의 학습을 위해 BindingDB[16] 데이터베이스가 이용되었다. 본 논문에서 사용한 모델은 단백질과 화합물이 주어질 때 둘 사이의 결합력을 의미하는 해리상수(Kd)를 추론하는 회귀(regression) 모델이며, 학습과 평가에 2,707,669개의 측정값이 이용되었다. 데이터셋을 구성하는 과정에서 모델 학습의 용이성을 위해 로그 스케일로 정규화하였기 때문

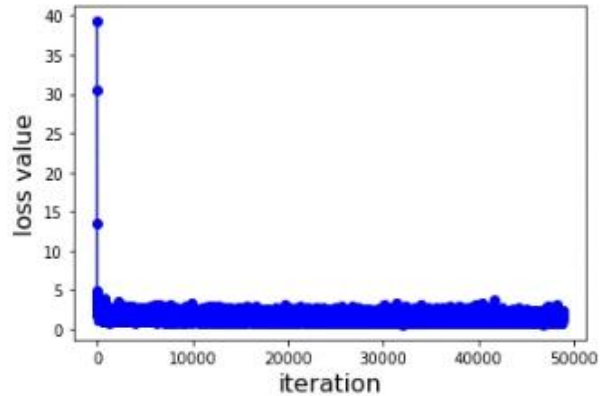


Fig. 4. Training Result of the model

에 모델의 학습에는 Kd가 아닌 pKd가 이용된다. 본 논문에 이용된 모델은 Kexin Huang의 연구에서 BindingDB를 이용한 단백질-화합물 결합 예측 모델의 구조[6]를 이용하였다. 학습, 검증, 테스트를 위해 데이터베이스는 각각 70%:10%:20%의 비율로 분리되었다. 화합물은 SMILES 형태로 표현되어 transformer layer에서, 단백질은 아미노산 서열 형태로 표현되어 1D-CNN(1D Convolutional Neural Networks) 레이어에서 벡터화되었고, FCNN(Fully Connected Neural Network)를 거쳐 최종적으로 해리상수(pKd)가 추론된다. Figure 3은 모델의 구조이고, Table 2는 제안하는 모델의 세부적인 하이퍼파라미터이다. CLS hidden dimension은 화합물 트랜스포머의 하이퍼파라미터이고, CNN filters와 CNN kernels는 단백질 1D-CNN의 하이퍼파라미터이다. 본 모델의 활성화 함수(activation function)는 ReLU(Rectified Linear Unit)를, 손실함수(loss function)는 평균제곱오차(Mean Squared Error, MSE)를 이용하였다. Figure 4는 제안하는 모델의 100 에포크, 50,000

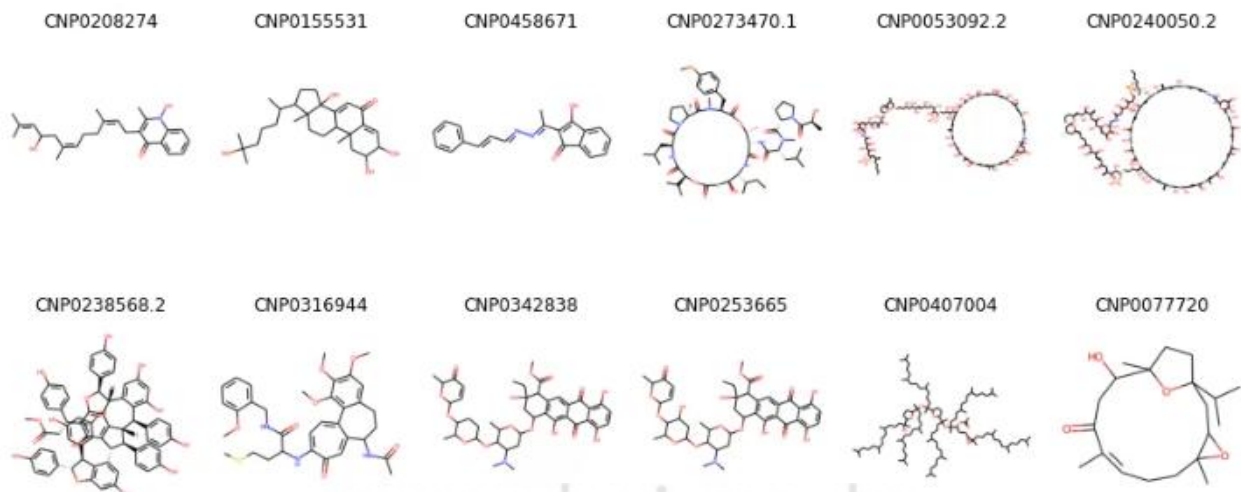


Fig. 5. Active compounds from the COCONUT database

Table 3. Detail of active compounds from the COCONUT database

COCONUT ID	Name	Species
CNP0208274	Aurachin Re	<i>Rhodococcus erythropolis</i>
CNP0155531	Hyousterone B	<i>Synicum adareanum</i>
CNP0273470.1	Tamandarin A	<i>Brazilian ascidian</i>
CNP0342838	Cinerubin R	<i>Streptomyces eurythermus</i>

Table 4. Detail of active compounds from the Korea Institute of Oriental Medicine Promotion natural substance database

ID	Name	Botanical Medicine Latin name
0227-11-1-I-0000	Fraxinellone	Dictamni Radicis Cortex
0479-11-1-I-0000	Tomentosin	Inulae flos
1263-14-1-O-0000	Isorhamnetin-3-O-rutinoside	Silver Magnolia
0111-11-1-I-0000	Britanin	Inula britannica
0490-11-1-I-0000	Typhaneoside	Typhae Pollen
1234-14-1-O-0000	Farnesiferol B	Ferulae Resina
0841-12-1-I-0000	Paeoniflorigenone	Moutan Cortex Radicis
1221-14-1-O-0000	Desmosterol	Sea hare
0624-12-1-I-0000	Atractylon	Atractylodes Rhizome
1131-13-1-I-0000	Sinomenine N-oxide	Sinomeni Caulis et Rhizoma

학습 iteration (500 iteration/epoch) 학습 과정의 validation loss를 보여준다. 모델은 학습 초기에 수렴하였으며 학습 완료 후, MSE 1.499, Pearson Correlation 0.693, Concordance Index 0.757을 달성하였다.

4. Virtual screening

GSE15852 데이터세트에서 도출된 인간의 유방 조직과 유방 종양 조직의 10개 차등 발현 유전자 단백질을 저해하는 천연물을 찾기 위해 COCONUT 데이터베이스[17]의 895,557개의 천연물과 한국한약진흥원 천연물 물질 데이터베이스[18]의 990개 생약 천연물을 가상 스크리닝 하였다. 각 유전자의 단백질 서열은 Uniport의 canonical sequence를 이용하였다. 결합 여부를 추론하기 위해 BindingDB로 학습된 단백질-화합물 결합력 예측 모델이 이용되었다. 가상 스크리닝을 위해 차등 발현 유전자 분석을 통해 도출된 10개 단백질과 데이터베이스의 모든 화합물에 대해 모델의 결합력(pKd)을 각각 추론하였고, 10개 단백질에서 모두 결합력이 상위 10% 이내인 화합물을 선별하여 천연물을 도출하였다.

Figure 5는 COCONUT 데이터베이스 가상 스크리닝 결과 도출된 12개의 천연물이다. 도출된 12개 천연물 중 명확히 식별된 천연물은 4개(Aurachin Re, Hyousterone B, Tamandarin A, Cinerubin R)이다. Table 3은 도출된 4개 천연물들의 이름과 기원생물을 보여준다. Aurachin Re와 Cinerubin R은 박테리아 유래의 천연물로 안전상의 이유로 배양육의 첨가제로 이용되기 어려울 것으로 생각된다. Hyousterone B은 해양무척추동물인 멧

게류에 속하는 *Synicum adareanum* 유래의 천연물이고, Tamandarin A은 멧게나 미더덕이 대표적인 해초강 피낭동물에 속하는 *Brazilian ascidian* 유래의 천연물이다.

Figure 6는 한국한약진흥원 천연물 물질 데이터베이스 가상 스크리닝 결과 도출된 10개의 천연물이다. 도출된 물질들은 백선피(Dictamni Radicis Cortex), 선복화(Inulae flos), 후박나무(Silver Magnolia), 포황(Typhae Pollen), 아위(Ferulae Resina), 목단피(Moutan Cortex Radicis), 군소(Sea hare), 백출(Atractylodes Rhizome), 방기(Sinomeni Caulis et Rhizoma)에서 유래된 천연물들이다. 이 중 2개의 천연물(Tomentosin, Britanin)이 공통적으로 선복화에서 유래된 것을 확인하였다. 이 결과는 선복화 추출물이 잠재적으로 배양육 생산 가속화를 위한 첨가제로 기능할 수 있음을 암시한다.

IV. Conclusions

본 논문에서는 배양육 생산을 가속화하기 위한 천연물 첨가제를 도출하는 방법을 제안하였다. GSE15852 데이터 세트를 이용하여 유방암 조직의 차등 발현 유전자를 도출하고, 이를 통해 근육조직과 지방조직의 성장과 관련된 유전자를 타겟 단백질로 선별하였다. 그 후 BindingDB 데이터베이스를 이용하여 단백질-화합물 결합력 예측 모델을 학습시키고, 이를 이용하여 천연물 데이터베이스에서 타겟 단백질들과 결합 가능한 천연물들을 도출하였다.

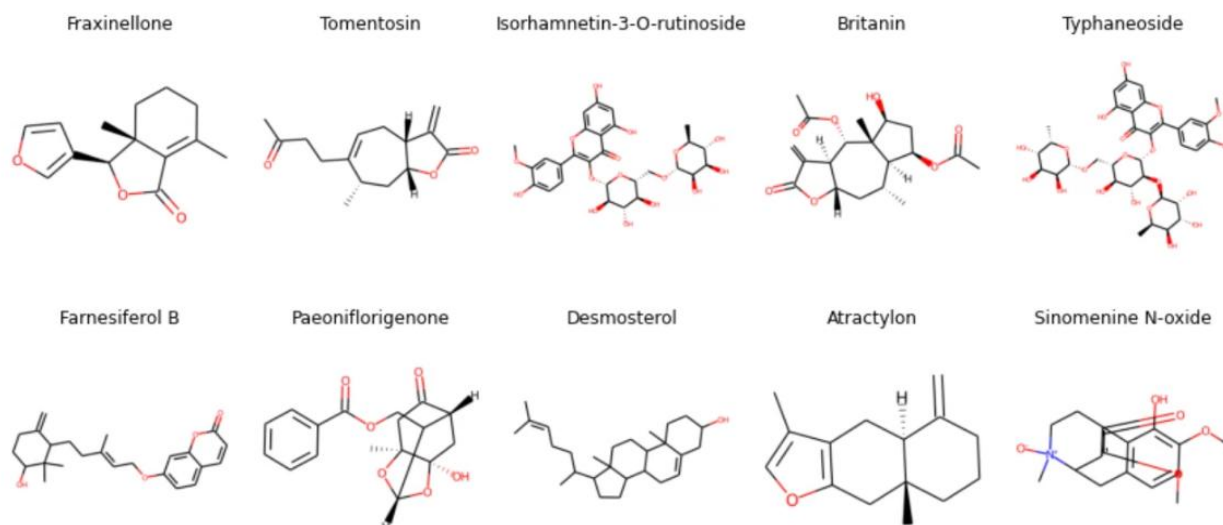


Fig. 6. Active compounds from the Korea Institute of Oriental Medicine Promotion natural substance database

COCONUT 데이터베이스 가상 스크리닝 결과 12개의 천연물이 도출되었으며, 이 중 4개의 천연물은 명확히 식별된 것으로 확인되었다. 한국한약진흥원 천연물 물질 데이터베이스 가상 스크리닝 결과 10개의 식물성 천연물이 도출되었으며, 이 중 2개의 천연물(Tomentosin, Britanin)이 공통적으로 선복화에서 유래된 것을 확인하였고, 이 결과는 선복화 추출물이 잠재적인 배양육 생산 가속화를 위한 첨가제로 기능할 수 있음을 암시한다.

본 연구는 배양육 생산을 가속화하기 위한 천연물 첨가제 도출에 관한 초기 연구이며, 실제 실험을 통해 검증되어야 한다. 또한, 천연물의 안전성, 효능, 생산성 등을 고려하여 실제 응용에 적합한 첨가제를 선별하는 과정이 필요하다. 또한, 본 연구의 타겟 유전자 도출을 위해 이용된 데이터는 인간 유전체로, 실제 배양육 생산에 적용하기 위해서는 해당 동물의 유전체 데이터가 이용되어야만 한다. 이를 위해 더욱 정교한 모델의 도입과 추가적인 실험 및 검증이 요구된다.

ACKNOWLEDGEMENT

This work was supported by Samyang Igeon (以建) Scholarship Foundation.

REFERENCES

- [1] S. Ng and M. Kurisawa, "Integrating biomaterials and food biopolymers for cultured meat production," *Acta Biomaterialia*, vol. 124, pp. 108–129, Apr. 2021. doi:10.1016/j.actbio.2021.01.017
- [2] M. Pallaoro et al., "Towards a more realistic in vitro meat: The cross talk between adipose and Muscle Cells," *International Journal of Molecular Sciences*, vol. 24, no. 7, p. 6630, 2023. doi:10.3390/ijms24076630
- [3] K. Choi et al., "Muscle stem cell isolation and in vitro culture for meat production: A Methodological Review," *Comprehensive Reviews in Food Science and Food Safety*, vol. 20, no. 1, pp. 429–457, Nov. 2020. doi:10.1111/1541-4337.12661
- [4] M. J. Post et al., "Scientific, sustainability and regulatory challenges of cultured meat," *Nature Food*, vol. 1, no. 7, pp. 403–415, Jul. 2020. doi:10.1038/s43016-020-0112-z
- [5] M. E. Ritchie et al., "Limma powers differential expression analyses for RNA-sequencing and Microarray Studies," *Nucleic Acids Research*, vol. 43, no. 7, p. e47, Apr. 2015. doi:10.1093/nar/gkv007
- [6] K. Huang et al., "DeepPurpose: A Deep Learning Library for drug-target interaction prediction," *Bioinformatics*, vol. 36, no. 22–23, pp. 5545–5547, Dec. 2020. doi:10.1093/bioinformatics/btaa1005
- [7] J. Lim et al., "Predicting drug-target interaction using a novel graph neural network with 3D structure-embedded graph representation," *Journal of Chemical Information and Modeling*, vol. 59, no. 9, pp. 3981–3988, Aug. 2019. doi:10.1021/acs.jcim.9b00387
- [8] H. Öztürk, A. Özgür, P. Schwaller, T. Laino, and E. Ozkirimli, "Exploring chemical space using natural language processing methodologies for Drug Discovery," *Drug Discovery Today*, vol. 25, no. 4, pp. 689–705, Apr. 2020. doi:10.1016/j.drudis.2020.01.020
- [9] H. Kong, I. Kim, and B.-T. Zhang, "Robust deep learning architecture for herg-related cardiotoxicity prediction on

small-scale dataset,” *KIISE Transactions on Computing Practices*, vol. 28, no. 6, pp. 339–347, Jun. 2022. doi:10.5626/ktcp.2022.28.6.339

- [10] A. Alex, E. Bhandary, and K. P. McGuire, “Anatomy and physiology of the breast during pregnancy and lactation,” *Advances in Experimental Medicine and Biology*, pp. 3–7, Aug. 2020. doi:10.1007/978-3-030-41596-9_1
- [11] M. Akram, M. Iqbal, M. Daniyal, and A. U. Khan, “Awareness and current knowledge of breast cancer,” *Biological Research*, vol. 50, no. 1, Oct. 2017. doi:10.1186/s40659-017-0140-9
- [12] H. Öztürk, A. Özgür, and E. Ozkirimli, “DeepDTA: Deep Drug–target binding affinity prediction,” *Bioinformatics*, vol. 34, no. 17, pp. i821–i829, Sep. 2018. doi:10.1093/bioinformatics/bty593
- [13] O. Kadioglu, M. Saeed, H. J. Greten, and T. Efferth, “Identification of novel compounds against three targets of SARS COV-2 coronavirus by combined virtual screening and supervised machine learning,” *Computers in Biology and Medicine*, vol. 133, 104359, Mar. 2021. doi:10.1016/j.compbiomed.2021.104359
- [14] L. A. Machado, E. Krempser, and A. C. Guimarães, “A machine learning-based virtual screening for natural compounds capable of inhibiting the HIV-1 integrase,” *Frontiers in Drug Discovery*, vol. 2, Oct. 2022. doi:10.3389/fddsv.2022.954911
- [15] I. B. Pau Ni et al., “Gene expression patterns distinguish breast carcinomas from normal breast tissues: The Malaysian context,” *Pathology - Research and Practice*, vol. 206, no. 4, pp. 223–228, Apr. 2010. doi:10.1016/j.prp.2009.11.006
- [16] T. Liu, Y. Lin, X. Wen, R. N. Jorissen, and M. K. Gilson, “BindingDB: A web-accessible database of experimentally determined protein-ligand binding affinities,” *Nucleic Acids Research*, vol. 35, no. 35(Database issue), pp. 198–201, Jan. 2007. doi:10.1093/nar/gkl999
- [17] M. Sorokina, P. Merseburger, K. Rajan, M. A. Yirik, and C. Steinbeck, “Coconut online: Collection of open natural products database,” *Journal of Cheminformatics*, vol. 13, no. 1, Jan. 2021. doi:10.1186/s13321-020-00478-9
- [18] Korea Institute of Oriental Medicine Promotion Status of natural substance DB, <https://www.data.go.kr/data/15109115/fileData.do>

Authors



Hyunseung Kong received the B.S. in systems biology in 2016 from Yonsei University, Seoul, Korea and M.S. in artificial intelligence in 2021 from IU Internationale Hochschule, Erfurt, Germany. He is currently pursuing

Ph.D. degree in Bioinformatics at Seoul National University, Seoul, Korea. His research interests include bioinformatics, drug design and artificial intelligence.



Inyoung Kim received the B.S., M.S. and Ph.D. degrees in Computer Science and Engineering from Seoul National University, Korea, in 1998, 2001 and 2020, respectively. She worked as a senior research engineer at

Samsung Electronics from 2001 to 2007. Dr. Kim joined the faculty of the Artificial Intelligence Institute of Seoul National University(AIIS) at Seoul National University in 2021. She is currently a Research Professor in the AIIS. Her research interests lie in the fields of artificial intelligence and bioinformatics.