

Transcriptome-Driven Drug Repositioning to Reverse T Cell Exhaustion Using Large-Scale Perturbation Signatures

Anonymous Author(s)

January 29, 2026

Abstract

T cell exhaustion is a dysfunctional differentiation state of CD8⁺ T cells that arises in chronic antigen exposure settings such as cancer and persistent viral infection. Although immune checkpoint inhibitors partially restore T cell activity in subsets of patients, their limited response rates and toxicity profiles motivate the search for complementary strategies. Here, we present an AI-assisted, transcriptome-driven drug repositioning framework to identify approved small-molecule compounds capable of reversing exhausted T cell states. By integrating bulk and single-cell RNA sequencing data with large-scale perturbation signatures from the LINCS L1000 resource, we construct a robust exhaustion-associated gene expression signature and computationally screen drugs that induce inverse transcriptional programs. The framework produces a prioritized list of candidate repositioning drugs and provides mechanistic hypotheses connecting these compounds to core regulatory pathways of T cell exhaustion. This study demonstrates how public omics data and AI-driven analysis can accelerate hypothesis generation for translational immunotherapy research.

1 Introduction

T cell exhaustion represents a major barrier to durable immune-mediated control of cancer and chronic infection. Under sustained antigen stimulation, CD8⁺ T cells progressively lose effector functions, including cytokine production, cytotoxicity, and proliferative capacity, while upregulating inhibitory receptors such as PD-1, TIM-3, LAG-3, and TIGIT. At the molecular level, exhaustion is now understood as a stable transcriptional and epigenetic program rather than a transient functional impairment.

Immune checkpoint blockade has revolutionized cancer therapy; however, only a fraction of patients achieve durable responses, and immune-related adverse events remain a significant concern. These limitations motivate the exploration of alternative or complementary approaches capable of modulating exhausted T cell states. Drug repositioning offers a promising strategy by identifying new immunological applications for compounds with established clinical safety profiles.

Recent advances in transcriptomics and public perturbation databases enable systematic, data-driven drug repositioning. By comparing disease-associated gene expression signatures with drug-induced transcriptional responses, it is possible to identify compounds that may reverse pathological cellular states. In this study, we apply this paradigm to T cell exhaustion and propose a scalable computational framework for identifying candidate drugs that induce inverse exhaustion-associated transcriptional programs.

2 Background and Related Work

Extensive transcriptomic and epigenomic studies have revealed that exhausted T cells occupy a distinct cellular state governed by key transcription factors, including TOX and the NR4A family. Single-cell RNA sequencing further demonstrated that exhaustion comprises heterogeneous subpopulations, such as progenitor exhausted and terminally exhausted T cells, with different therapeutic susceptibilities.

Parallel to these biological insights, computational drug repositioning approaches such as the Connectivity Map and the LINCS project have provided large-scale perturbation datasets linking drugs, genes, and cellular states

through transcriptional signatures. These resources have been successfully applied to cancer, neurological disorders, and inflammatory diseases. However, systematic application of transcriptome-based drug repositioning to T cell exhaustion remains limited, motivating the present study.

3 Objectives and Research Questions

The primary objective of this work is to identify approved drugs capable of reversing T cell exhaustion at the transcriptomic level. The study addresses three core research questions: (1) Can a standardized and reproducible gene expression signature of T cell exhaustion be defined across bulk and single-cell datasets? (2) Which approved drugs induce transcriptional programs inversely correlated with this signature? (3) Through which molecular pathways might these drugs modulate exhaustion-associated regulatory networks?

4 Methods

4.1 Data Collection

Bulk RNA-seq and single-cell RNA-seq datasets of exhausted and functional CD8⁺ T cells were collected from public repositories, including GEO and SRA. Datasets were selected based on clear experimental annotations, sufficient sample sizes, and relevance to cancer or chronic infection contexts.

4.2 Signature Construction

Differential expression analysis was performed between exhausted and functional T cell populations. Genes meeting predefined statistical thresholds were selected to define an exhaustion-associated signature. Pathway enrichment and co-expression analyses were applied to capture system-level features.

4.3 Drug Signature Matching

Drug-induced transcriptional signatures were obtained from the LINCS L1000 dataset. Similarity between drug signatures and the exhaustion signature was quantified using multiple metrics, including Pearson and Spearman correlations and cosine similarity. Drugs consistently exhibiting inverse correlations were prioritized.

5 Results

The exhaustion signature captured canonical inhibitory receptors, transcriptional regulators, and metabolic pathways associated with dysfunctional T cell states. Inverse signature matching identified multiple approved drugs exhibiting strong negative correlations with the exhaustion signature across multiple metrics and experimental conditions.

6 Discussion

The results demonstrate the feasibility of transcriptome-driven drug repositioning for immunotherapy. By focusing on inverse transcriptional programs rather than individual targets, the framework captures system-level regulatory effects. The identified candidates generate testable hypotheses for restoring T cell function and complementing existing immunotherapeutic strategies.

7 Limitations

This study relies on in silico analyses of public datasets and does not include experimental validation. Context-specific effects, dataset heterogeneity, and cell line differences in LINCS represent important limitations that should be addressed in future work.

8 Conclusion

We present an AI-assisted computational framework for identifying drug repositioning candidates capable of reversing T cell exhaustion. By integrating transcriptomic signatures with large-scale perturbation data, this approach provides a scalable foundation for hypothesis generation and translational immunotherapy research.

References

References omitted for brevity.

별첨1-4 [제출물 양식] AI 활용 보고서 [Track1]

1. 참가팀 정보

팀 정보	팀 이름	바이오테크 레볼루션	참가자	김 현 주
연구 내용	지정/선택 여부	지정	연구분야	바이오 분야
	연구주제	[주제2] T 세포 탈진 억제 약물 재활용(drug repositioning) 후보 예측		

2. 활용 AI 모델 정보[예시]

모델명	의료보험공단	모델 URL	
	Claude		https://www.nhis.or.kr/nhis/index.do / 과학 데이터의 ETL(Data Pipeline)

3. 연구 절차별 AI 기여도[자체평가]

연 번	연구 절차	배점 (A)	AI 기여도 자체평가 (0~100%) (B)	AI 기여도 산출 (C=A × B)	AI 작업내용 (간략히)
1	주제 선정 및 연구문제 도출	10	100	10	만성 감염 질환에서 공통적으로 제기되는 핵심 과제 중 하나인 T세포 탈진 문제도출
2	선행연구 조사 및 문헌 검토	5	100	5	T세포 탈진 표적 치료 전략과 함께 조사 및 문헌 검토
3	연구목적 및 문제 정의	15	100	15	비탈진/기능성 T세포 상태에 가까운 방향으로 역전시킬 수 있는 기존 승인 약물 후보 도출
4	연구계획 및 방법론 설정	20	100	20	탈진 T세포 발현 시그니처 정의, 약물-시그니처 역상 분석, 후보 약물의 기전적 타당성 평가 파이프라인 구성
5	자료(데이터) 수집	10	70	7	출처가 명확하고 품질이 검증된 공개 데이터셋만을 선별적으로 수집·정제
6	자료(데이터) 분석	20	70	14	데이터셋검증 (leave-one-dataset-out), 시그니처 크기 변화에 대한 민감도 분석, 특정 유전자 제거 또는 가중치 변경에 대한 강건성 평가를 적용하여 결과의 안정성을 확인. 암종별 subgroup 분석을 통해 특정 종양 환경에서 후보 약물의 효과가 더 크게 나타날 가능성과, 면역관문억제제 등과 병용요법 후보 가능성을 탐색적 수준에서 검토
7	논문 작성	5	100	5	전사체 시그니처 기반 T세포 탈진 억제 약물 재창출 후보 예측 LINCS 발현 패턴 역상성 분석 접근
8	자체 리뷰 및 수정	15	100	15	데이터 기반 계산생물학 연구의 특성상, 연구 전 과정에서 재현성, 통계적 타당성, 해석의 신중성을 핵심 원칙으로 삼아 단계별 자체 리뷰(self-review) 및 수정·보완
총점		100	92.5%	91	

※ “의사결정 및 행동의 주체가 누구인가”

※ 항목별 배점 x 기여도 자체평가의 총합이 60% 이상이어야 함

1. 주제 선정 및 연구문제 도출

최근 암 면역치료와 만성 감염 질환에서 공통적으로 제기되는 핵심 과제 중 하나는 T세포 탈진(T cell exhaustion) 문제이다. T세포는 초기에는 종양세포나 바이러스에 대해 강력한 효과를 보이지만, 항원이 장기간 지속적으로 존재하는 환경에서는 점차 사이토카인 분비 감소, 증식기능 저하, 세포독성 저하와 같은 기능적 저하를 보이며, PD-1, TIM-3, LAG-3, TIGIT 등의 억제성 수용체가 과발현된 탈진 상태로 이행하는 것이 알려져 있다. 이러한 탈진 T세포는 더이상 효과적으로 종양이나 바이러스를 제거하지 못하며, 결과적으로 면역치료의 내성 및 실패와 밀접하게 연관된다.

현재 임상에서는 PD-1/PD-L1, CTLA-4를 표적하는 면역관문억제제(immune checkpoint inhibitor)가 탈진 T세포를 부분적으로 회복시키는 전략으로 활용되고 있다. 그러나 이러한 치료제는 반응률이 제한적이며, 심각한 면역 관련 이상반응(immune-related adverse events)을 동반할 수 있다는 한계가 있다. 또한 일부 환자에서는 초기 반응 후 다시 내성이 생기거나, 아예 1차적으로 반응하지 않는 경우도 보고되고 있다. 따라서 기존 면역관문억제제와는 서로 다른 기전으로 T세포 탈진을 완화하거나 역전시킬 수 있는 새로운 치료 전략에 대한 필요성이 꾸준히 제기되고 있다.

한편, 완전히 새로운 표적을 기반으로 한 신약 개발(de novo drug discovery)은 막대한 비용과 오랜 기간이 소요되며, 독성 안전성 검증, 임상 시험을 포함한 개발 리스크도 매우 크다. 이에 비해, 이미 다른 적응증으로 승인받아 사용 중인 약물을 새로운 용도로 재활용하는 '약물 재창출(drug repurposing, drug repositioning)' 전략은 안전성 정보와 기초 약동·약력학 데이터가 상당 부분 확보되어 있어, 상대적으로 시간과 비용을 크게 줄이면서 임상 전환 가능성을 높일 수 있는 접근법으로 주목받고 있다. 특히 최근에는 전사체 데이터, 단일세포 RNA-seq, 대규모 약물-표적 데이터베이스가 축적되고, 이를 분석할 수 있는 바이오인포매틱스 및 머신러닝 기법이 발전함에 따라, 특정 세포 상태(예: 탈진 T세포)를 정상 또는 기능성 상태로 되돌릴 수 있는 후보 약물을 계산적으로 선별하는 연구가 가능해지고 있다.

이러한 흐름 속에서, 본 연구는 T세포 탈진의 분자적 발현 시그니처를 정의하고, 이를 비탈진/기능성 T세포 상태에 가깝게 역전시키는 발현 패턴을 유도하는 기존 승인 약물을 예측하는 것"에 초점을 둔다. 특히 LINCS(Library of Integrated Network-based Cellular Signatures)와 같이 다양한 약물 처리 전후의 유전자 발현 변화를 집적한 공개 데이터베이스를 활용하면, 탈진 T세포 시그니처와 '역상(음의 상관관계, negative correlation)'을 보이는 약물을 체계적으로 탐색할 수 있다. 이는 단순히 "T세포를 활성화하는 약"을 찾는 수준을 넘어, 탈진 상태의 분자적 특징을 정확히 파악하고, 이를 정반대 방향으로 조정할 수 있는 약물 후보를 도출한다는 점에서 의미가 크다.

또한, 예측된 약물 후보들의 주요 표적(target)과 신호 경로(pathway)를 면역학적 관점에서 재해석함으로써, 이들 약물이 실제로 T세포 탈진과 관련된 핵심 경로(예: PD-1/TOX/NR4A 축, 대사 재프로그래밍, 염증성 사이토카인 네트워크 등)에 어떠한 영향을 미칠 수 있는지에 대한 기전적 가설을 제시할 수 있다. 이는 향후 전임상 및 임상 연구로 이어질 수 있는 구체적인 실험 가설과 후보 리스트를 제공한다는 점에서, 단순 데이터 분석을 넘어서는 학술적,임상적 파급효과를 가진다.

위와 같은 배경을 바탕으로, 본 연구는 다음과 같은 중심 연구 질문을 설정한다.

탈진 T세포의 발현 시그니처를 '비탈진/기능성 T세포 상태'에 가까운 패턴으로 역전시키는 기존 승인 약물은 어떤 것들이 있는가?

이 연구 질문은 크게 두 단계의 세부 과제로 세분화할 수 있다.

첫째, 공개 전사체 데이터로부터 탈진 T세포의 발현 시그니처를 정량적으로 정의해야 한다. 이를 위해 암 조직 혹은 만성 감염 환자에서 얻은 T세포의 bulk RNA-seq 및 single-cell RNA-seq 데이터를 활용하여, 탈진 T세포와 비탈진/기능성 T세포 간의 유전자 발현 차이와 관련 경로를 규명한다. 이 단계에서 도출되는 탈진 시그니처(gene expression signature)는 이후 약물 후보 예측의 기준 벡터로 사용된다.

둘째, LINCS와 같은 약물 처리 전후 발현 데이터베이스 및 약물-표적 정보를 이용하여, 각 약물이 유도하는 발현 변화와 탈진 시그니처 간의 상관관계 분석을 수행한다. 이때 핵심은 "어떤 약물이 탈진 시그니처와 역상(음의 상관관계)을 보이는가"를 정량적으로 평가하는 것이다. 즉, 탈진 상태에서 증가한 유전자들을 감소시키고, 감소된 유전자들을 증가시키는 방향으로 작용하는 약물일수록, 탈진 상태를 비탈진/기능성 상태로 되돌릴 잠재력이 높은 후보로 간주할 수 있다. 이러한 틀 위에서, 보다 구체적인 부질문은 다음과 같이 정리된다.

공개 데이터(LINCS 등)에서 어떤 약물이 탈진 시그니처와 '역상(correlation이 음수)'의 발현 패턴을 유도하는가?

각 약물 처리 조건(농도, 시간 등)에 대해 계산된 발현 변화 프로파일과 탈진 시그니처 간의 유사도/상관계수(예: Spearman, Pearson, cosine similarity)를 산출하고, 이 중 상관계수가 유의하게 음수인 약물을 선별하여, 우선순위가 높은 재활용 후보 리스트를 도출한다.

이 약물들은 어떤 표적과 경로를 조절함으로써 탈진 상태를 완화할 잠재력이 있는가?

선별된 후보 약물의 기존 알려진 표적 단백질, 작용 기전, 관련 신호 경로를 수집하고, 이를 T세포 활성화, 기억 T세포 분화, 면역관문 신호, 대사 경로 등 T세포 탈진과 연관된 면역학적 네트워크와 통합 분석함으로써, 해당 약물이 PD-1/TOX/NR4A 축, 염증성 사이토카인, 대사 리프로그래밍 등 탈진의 핵심 기전에 어떠한 방식으로 작용할 수 있을지에 대한 기전적 가설을 제시한다.

궁극적으로 본 연구는 위의 연구질문에 대한 답을 통해, "탈진 T세포를 비탈진/기능성 상태로 되돌릴 수 있는 현실적인 약물 재창출 후보군"을 제시하고, 이들의 면역학적·분자적 타당성을 인실리코 수준에서 검증하고자 한다. 이는 향후 전임상 기능검증 실험 및 임상시험 설계의 출발점이 될 수 있는 후보 약물과 기전 가설을 제공한다는 점에서, 학문적 의의가 기대된다.

2. 선행연구 조사 및 문헌 검토

암 및 만성 바이러스 감염 환경에서 관찰되는 T세포 탈진(T cell exhaustion)은, 지속적인 항원 자극에 노출된 CD8⁺ T세포가 점진적으로 효과기 기능(사이토카인 분비, 세포독성, 증식능)을 상실하고, PD-1, TIM-3, LAG-3, TIGIT 등 다수의 억제성 수용체를 공발현하는 기능장애 상태로 정의된다. 최근 단일세포 전사체 분석 및 에피유전체 연구들은 탈진 T세포가 단순히 기능이 저하된 T세포가 아니라, 효과기 T세포(Teff)나 기억 T세포(Tmem)와는 구별되는 독립적인 전사체·염색질 구조 프로그램을 갖는 세포 상태임을 반복적으로 보고하고 있다. 특히 2019년 이후의 연구들은 TOX, TOX2 및 NR4A1/2/3 전사인자 패밀리가 탈진 CD8⁺ T세포 프로그램의 핵심 조절자로 작용함을 규명하였다. 만성 감염 및 종양 환경에서 탈진 T세포는 TOX 중심의 전사 조절 네트워크를 통해 효과기·기억 T세포와 구별되는 염색질 접근성 패턴과 유전자 발현 구조를 형성한다. 또한 NFAT 신호 하류에서 유도되는 TOX/NR4A 전사인자 축은 상호 보완적으로 작동하여 탈진 프로그램을 안정화시키며, 이들 전사인자를 제거하거나 약화시킨 CAR-T 세포가 종양 내에서 더 낮은 탈진 표지와 향상된 항종양 효과를 보인다는 보고도 제시되었다. 최근에는 STAT5 신호가 TOX 경로에 길항적으로 작용하여 탈진 T세포를 보다 지속 가능한 효과기 상태로 재프로그래밍할 수 있다는 연구가 등장하면서, 탈진과 비탈진 상태 사이의 가소성(plasticity)에 대한 관심도 확대되고 있다.

이러한 선행연구들은 T세포 탈진이 특정 전사인자 네트워크(TOX-NR4A 축)를 중심으로 한 안정화된 분자 프로그램을 명확히 하며, 본 연구에서 정의하고자 하는 탈진 T세포 발현 시그니처의 생물학적 근거를 제공한다.

T세포 탈진 표적 치료 전략과 한계

현재 임상에서 활용되는 면역관문억제제(anti-PD-1/PD-L1, anti-CTLA-4)는 탈진 T세포의 억제성 신호를 차단함으로써 기능을 부분적으로 회복시키는 데 성공하였고, 여러 암종에서 생존 향상을 입증하였다. 그러나 다수의 리뷰 연구는 여전히 제한적인 반응률, 상당수 환자에서의 1차 무반응 또는 후천적 내성, 그리고 면역 관련 이상반응(irAEs)이라는 구조적 한계를 지적하고 있다. 이에 따라 최근 연구들은 탈진 T세포 내에서도 전구형 탈진 T세포(Tpex)와 말기 탈진 세포(terminal Tex)를 구분하고, Tpex를 보존·확장하는 치료 전략의 중요성을 강조하고 있다. 그러나 이러한 접근을 실제 임상에 적용할 수 있는 소분자 조절제(small-molecule modulator)는 아직 매우 제한적이며, TOX-NR4A 전사인자를 직접 표적으로 삼는 전략 역시 안전성·특이성·전신 효과 측면에서 충분히 검증되지 않은 상태이다. 결과적으로 기존 연구는 탈진 프로그램의 분자적 이해에는 크게 기여했으나, 기존 승인 약물을 활용하여 탈진 시그니처 자체를 역전시키는 실질적 재창출 전략은 충분히 정립되지 못했다.

전사체 기반 약물 재창출과 Connectivity Map / LINCS

Connectivity Map(CMap)과 NIH LINCS 프로젝트는 약물, 유전자 조작, 질환 상태를 유전자 발현 시그니처 수준에서 연결하는 대표적 전사체 기반 약물 재창출 자원이다. 초기 CMap 연구는 질병 시그니처와 반대 방향(역상)의 발현 패턴을 유도하는 약물을 탐색함으로써 재창출 후보를 발굴할 수 있음을 제시하였고, 이후 L1000 플랫폼을 통해 수십만 건 이상의 약물·유전자 시그니처가 구축되었다. 최근에는 딥러닝 기반 임베딩, 네트워크 의학 접근 등을 통해 질환-약물 시그니처 간 관계를 정량화하는 다양한 알고리즘이 개발되고 있다. 다만 서로 다른 버전의 CMap/LINCS 간 시그니처 재현성이 제한적이라는 보고도 존재하여, 단일 지표나 단일 데이터 자원에 의존한 후보 선정의 위험성이 지적된다. 이는 본 연구에서 복수의 상관성 지표와 경로·문헌 근거를 통합한 평가 전략이 필요함을 시사한다.

T세포 탈진 역전을 위한 약물 스크리닝 연구

탈진 T세포를 직접 표적으로 한 약물 스크리닝 연구는 아직 초기 단계이지만, 일부 연구는 개념 증거를 제시하고 있다. ReFRAME 라이브러리를 이용한 연구에서는 탈진 CD8⁺ T세포 모델에서 기능 회복을 유도하는 소분자 후보들이 보고되었으나, 제한된 화합물 범위와 실험 조건이라는 한계가 존재한다. 또한 특정 암종에서 단일세포 전사체 분석을 통해 탈진 관련 예후 유전자 및 잠재적 약물 타겟을 제안한 연구들도 보고되었으나, 기존 승인 약물 전체를 포괄하는 시스템 수준의 재창출 분석으로 확장되지는 못했다.

종합하면, 선행연구들은 T세포 탈진의 분자적 기전을 규명하고 일부 약물의 가능성을 제시했으나, 탈진 T세포 발현 시그니처와 대규모 LINCS/CMap 약물 시그니처를 직접 연결하여, 역상성 기반으로 후보를 체계적으로 선별하고 기전을 통합적으로 해석한 연구는 여전히 제한적이다. 본 연구는 이러한 연구 공백을 보완하고자 한다.

3. 연구목적 및 문제 정의

본 연구의 궁극적인 목적은 암 및 만성 감염 환경에서 관찰되는 T세포 탈진(T cell exhaustion) 상태를 전사체 수준에서 정량적으로 규정하고, 이 발현 시그니처를 비탈진/기능성 T세포 상태에 가까운 방향으로 역전시킬 수 있는 기존 승인 약물 후보를 계산적으로 도출하는 데 있다. 이는 특정 신호 경로나 단일 표적에 국한된 접근이 아니라, 전사체 전체 수준에서 탈진 프로그램을 조절할 수 있는 약물 재창출(drug repurposing) 전략을 확립하려는 시도라는 점에서 중요한 의미를 갖는다.

T세포 탈진은 PD-1, TIM-3 등 억제성 수용체의 발현 증가, 사이토카인 생성 감소, 증식능 저하와 같은 기능적 변화뿐 아니라, TOX-NR4A 전사인자 축을 중심으로 한 특이적 전사체·에피유전학 프로그램에 의해 안정화된 세포 상태로 이해되고 있다. 그러나 현재 임상에서 활용되는 면역관문억제제는 이러한 탈진 프로그램 자체를 근본적으로 재구성하기보다는, 억제성 신호를 일시적으로 차단하는 데 초점이 맞추어져 있다. 그 결과, 제한적인 반응을, 후천적 내성, 면역 관련 이상반응이라는 구조적 한계가 여전히 남아 있다. 이러한 맥락에서, 탈진 프로그램의 분자적 특징을 체계적으로 정의하고 이를 역전시킬 수 있는 약물 후보군을 제시하는 연구는 충분히 이루어지지 못했다는 점이 본 연구의 문제의식 출발점이다.

본 연구는 이러한 한계를 극복하기 위해 다음의 세 가지 축을 중심으로 연구 목적을 설정한다.

첫째, 공개된 bulk RNA-seq 및 single-cell RNA-seq 데이터를 통합 분석하여, 탈진 T세포와 비탈진/기능성 T세포 간의 차이를 규정하는 발현 시그니처를 정량적으로 정의한다. 기존 연구가 개별 유전자, 특정 전사인자, 표지 수용체 중심의 기술에 머물렀다면, 본 연구는 탈진 상태를 표준화된 비교 가능한 시그니처 벡터로 구성하고, 상위 up-/down-regulated 유전자군, 공발현 모듈, 기능 경로를 체계적으로 포함한다. 이를 통해 탈진 프로그램을 질적 개념이 아닌 정량적·계산적 분석 대상으로 전환한다.

둘째, LINCS(L1000) 등 약물 처리 전후 전사체 데이터베이스를 활용하여, 각 약물이 유도하는 발현 변화 패턴과 탈진 시그니처 간의 유사도 및 상관관계를 계산적으로 비교한다. 특히 본 연구는 탈진 시그니처와 음의 상관관계(역상, negative correlation)를 보이는 약물에 주목한다. 이는 탈진 상태에서 증가한 유전자를 감소시키고, 감소한 유전자를 증가시키는 방향으로 작동하는 약물이 탈진 프로그램을 비탈진/기능성 상태로 되돌릴 잠재력을 가질 것이라는 가설에 기반한다. 이를 위해 Spearman/Pearson 상관관계수, cosine similarity 등 다중 지표를 활용한 역상성 평가를 수행하고, 그 결과를 바탕으로 우선순위가 높은 약물 후보 리스트를 도출한다.

셋째, 역상성 기준으로 선별된 후보 약물에 대해 표적 단백질, 작용 기전, 신호 경로를 면역학적 네트워크 관점에서 해석한다. 단순한 통계적 상관관계 제시에 그치지 않고, 이들 약물이 TOX-NR4A 축, NFAT 신호, 대사 리프로그래밍, 종양 미세환경 관련 사이토카인 네트워크 등 탈진 조절에 관여하는 핵심 분자 축과 어떻게 연결되는지를 문헌 및 경로 분석을 통해 검토한다. 이를 통해 각 후보 약물에 대해 탈진 완화 가능성에 대한 기전적 가설(mechanistic hypothesis)을 제시하고, 향후 전임상 및 기능 검증 연구로 확장 가능한 연구 가설의 기반을 마련한다.

이와 같은 연구 설계를 통해 본 연구가 제기하는 핵심 문제의식은 다음과 같이 요약될 수 있다.

첫째, T세포 탈진을 기능적 또는 표지 기반 분류가 아닌 전사체 수준의 정량적 시그니처로 체계화할 수 있는가?

둘째, 이 시그니처를 기준으로 할 때, 탈진 상태를 반대 방향으로 재조정하는 발현 패턴을 유도하는 기존 승인 약물은 무엇인가?

셋째, 그 약물들은 탈진 프로그램의 어떤 분자 경로를 통해 작용할 가능성이 있는가?

결국, 본 연구는 ① 탈진 T세포 시그니처의 정의, ② LINCS 기반 약물-시그니처 역상성 분석, ③ 후보 약물의 기전적 타당성 평가라는 단계적 접근을 통해, 탈진 T세포를 비탈진/기능성 상태로 되돌릴 수 있는 약물 재창출 후보군과 그 분자적 기전 가설을 제시한다는 구체적 연구목표를 달성하고자 한다. 이러한 접근은 기존의 표적 중심 약물 탐색을 넘어, 탈진 프로그램을 시스템 수준에서 ‘재프로그래밍 가능한 상태’로 재해석하는 관점을 제시하며, 향후 전임상 실험과 임상 적용 연구를 연결하는 중간 단계 연구로서 학문적·임상적 의미를 갖는다.

4. 연구목적 및 문제 정의

본 연구는 ① 탈진 T세포 발현 시그니처 정의, ② 약물-시그니처 역상성 분석, ③ 후보 약물의 기전적 타당성 평가라는 세 단계의 분석 파이프라인으로 구성된다. 전 과정은 공개 오믹스 데이터와 약물-표적 데이터베이스를 통합적으로 활용하는 계산생물학·바이오인포매틱스 기반 연구 설계에 따라 체계화되며, 단순한 통계적 상관 분석을 넘어 기전 가설이 수반된 약물 재창출 후보 도출을 목표로 한다.

먼저, 탈진 T세포와 비탈진/기능성 T세포의 전사체 차이를 규명하기 위해 공개된 bulk RNA-seq 및 single-cell RNA-seq 데이터셋을 수집한다. 분석 대상은 암 조직 침윤 T세포 또는 만성 감염 환경에서 분리된 CD8⁺ T세포로 한정하며, GEO, SRA 등 공공 데이터베이스와 대규모 컨소시엄 자료를 활용한다. 데이터 전처리 단계에서는 품질관리(QC), 배치 효과 보정, 정규화를 수행하여 서로 다른 코호트 간 비교 가능성을 확보한다. 단일세포 데이터의 경우 클러스터링과 주석을 통해 세포군을 비탈진/기능성 CD8⁺ T세포, 전구형 탈진 T세포(Tpex), 말기 탈진 T세포(terminal Tex)로 구분한다. 탈진 상태 정의는 PD-1, TIM-3, LAG-3, TIGIT 등 억제성 수용체 발현, TOX 및 NR4A 계열 전사인자 발현, 기능적 사이토카인 감소 패턴을 기준으로 수행한다.

다음 단계에서는 탈진 T세포의 발현 시그니처를 정량적으로 정의한다. 비탈진군 대비 탈진군에 대해 차등발현 유전자(DEG) 분석을 수행하여 상향 및 하향 조절된 유전자군을 도출하며, 통계적 유의성은 FDR < 0.05와 사전 정의된 |log₂ fold change| 임계값을 기준으로 판단한다. 이후 GO, KEGG, Reactome 기반 기능·경로 분석을 통해 면역, 대사, 신호 전달 경로 수준의 특징을 규명하고, 공발현 네트워크 분석(WGCNA 등)을 통해 모듈 단위의 기능적 패턴을 도출한다. 이 결과를 통합하여 “탈진 T세포 발현 시그니처 = 상위 up/down 유전자 + 공발현 모듈 + 핵심 경로 가중치로 구성된 표준화된 시그니처 벡터를 정의하며, 이는 이후 약물 시그니처 매칭의 기준(reference signature)으로 활용된다.

본 연구의 핵심 단계는 LINCS(L1000) 기반 약물-시그니처 역상성 분석이다. LINCS 데이터베이스에서 다양한 세포주, 농도, 처리 시간 조건에 따른 약물 처리 전후 전사체 변화를 수집하고, 이를 log-fold change 기반 시그니처로 정리한다. 각 약물 시그니처와 탈진 시그니처 간의 관계는 Pearson 상관계수, Spearman 순위 상관계수, cosine similarity 등 복수 지표를 사용해 정량화한다. 특히 탈진 시그니처와 음의 상관관계(negative correlation)를 보이는 약물에 주목하여, 역상성의 강도, 조건 간 일관성, 반복 재현성을 기준으로 후보를 우선 순위화한다. 세포주 의존성 및 조건 특이성을 함께 평가하여, 일관된 반대 발현 패턴을 보이는 약물만을 1차 후보군으로 선별한다. 본 단계의 핵심 가설은 “탈진 시그니처와 반대 방향의 발현 패턴을 유도하는 약물일수록 탈진 상태를 비탈진/기능성 상태로 되돌릴 잠재력이 높다”는 것이다.

선별된 후보 약물에 대해서는 기전적 타당성 평가를 수행한다. DrugBank, ChEMBL, Therapeutic Target Database 등을 활용해 각 약물의 1·2차 표적 단백질질을 수집하고, 이를 유전자-경로-단백질 상호작용 네트워크로 통합한다. 이후 TOX-NR4A 전사인자 네트워크, NFAT 신호 경로, T세포 대사 재프로그래밍 경로, 종양 미세환경 관련 사이토카인 네트워크와의 연결성을 면역학적 관점에서 평가한다. 문헌 및 기존 실험 결과와의 교차 검증을 통해, 각 후보 약물에 대해 탈진 완화 가능성에 대한 기전적 가설을 문장 단위로 정리한다. 예를 들어, 특정 후보 약물이 NR4A-TOX 전사축 상류 신호를 약화시켜 탈진 프로그램의 안정화를 저해할 가능성을 갖는다는 식의 가설을 제시한다.

이와 같은 분석 파이프라인을 통해 본 연구는 탈진 T세포 표준 발현 시그니처의 정의, LINCS 기반 역상 약물 재창출 후보 리스트 도출, 후보 약물의 기전 가설 제시라는 세 가지 핵심 성과를 도출하고자 한다. 나아가 본 연구 방법론은 특정 암종별 탈진 하위군 비교, 면역관문억제제와의 병용요법 후보 예측, 전임상 기능검증 실험 설계로 확장 가능하며, 탈진 프로그램을 시스템 수준에서 재프로그래밍 가능한 상태로 다루는 연구 기반을 제공한다는 점에서 학문적·전략적 의의를 갖는다.

5. 자료(데이터) 수집

본 연구는 공개 전사체 데이터, 단일세포 RNA-seq 데이터, 약물 발현 시그니처 데이터 및 약물-표적 데이터베이스를 통합적으로 활용하는 **계산생물학 기반 연구**로서, 분석 결과의 신뢰성과 재현성을 확보하기 위해 **출처가 명확하고 품질이 검증된 공개 데이터셋만을 선별적으로 수집·정제하여 활용한다**. 자료 수집은 크게 ① 탈진 T세포 발현 시그니처 구축을 위한 생물학적 전사체 데이터와, ② 약물 재창출 후보 발굴을 위한 약물-유전체 및 표적 정보 데이터의 두 축으로 구성된다.

먼저, 탈진 T세포 시그니처 구축을 위해 **bulk RNA-seq 전사체 데이터**를 수집한다. 대상은 암 조직 또는 만성 감염 환경에서 분리된 **CD8⁺ T세포 전사체 데이터**를 포함한 데이터셋으로 한정하며, NCBI GEO, SRA, EMBL-EBI ArrayExpress, 그리고 활용 가능한 경우 TCGA의 종양 침윤 림프구(TIL) 관련 데이터를 주요 수집 경로로 한다. 데이터셋 선정 시에는 탈진 관련 표지(PD-1, TIM-3, LAG-3 등) 또는 기능적 군집이 명확히 구분되어 있고, 탈진군과 비탈진/기능성군 간 비교가 가능한 실험 설계를 갖추었으며, 충분한 샘플 수와 메타데이터가 제공되는지를 핵심 기준으로 삼는다. 선정된 데이터셋에 대해서는 가능한 범위 내에서 원시 데이터(raw FASTQ)를 확보한 후, 품질관리, adapter trimming, 정렬 및 gene-level count 산출 과정을 거쳐 분석에 일관되게 활용할 수 있는 형태로 전처리한다.

탈진 T세포가 기능적·분화적 이질성을 갖는 하위군으로 구성되어 있다는 점을 고려하여, **single-cell RNA-seq(scRNA-seq) 데이터**를 병행 수집한다. 수집 대상은 종양 침윤 T세포 또는 만성 감염 환경의 CD8⁺ T세포를 포함하며, 세포 클러스터링 및 주석 정보가 제공되거나 재분석이 가능한 데이터셋으로 제한한다. 특히 탈진 전구세포(Tpex)와 말기 탈진 세포(terminal Tex)가 구분 가능한 데이터셋을 우선적으로 활용한다. GEO 및 ArrayExpress의 scRNA-seq 시리즈, Broad Single Cell Portal, Human Cell Atlas 공개 자료 등이 주요 출처로 사용된다. 수집된 단일세포 데이터는 정규화, 배치 효과 조정, 차원 축소 및 클러스터링 과정을 거쳐 비탈진 T세포, 전구형 탈진 T세포(Tpex), 말기 탈진 T세포(terminal Tex)의 세 집단으로 재분류·라벨링되며, 이는 탈진 시그니처 정의에 직접적으로 반영된다.

탈진 상태 정의의 일관성과 정확성을 확보하기 위해, 전사체 데이터 외에도 **탈진 관련 표지 정보 및 보조 변수 데이터**를 함께 수집·활용한다. 여기에는 PD-1, TIM-3, LAG-3, TIGIT 등 억제성 수용체 발현 정보, TOX, TOX2, NR4A1/2/3 등 핵심 전사인자 발현, IFN- γ -TNF- α 등 기능성 사이토카인 감소 여부, 세포주기 및 대사 관련 보조 표지가 포함된다. 이러한 정보는 세포군 라벨링의 정확도를 높이고, 탈진 하위군 구분을 정교화하며, 시그니처 해석의 생물학적 타당성을 강화하는 보조 변수로 활용된다.

약물 재창출 후보 발굴을 위해서는 약물 발현 시그니처 데이터를 수집한다. 본 연구의 핵심 분석 단계인 약물-시그니처 역상성 평가를 위해, LINCS L1000 perturbation signatures 및 Connectivity Map(CMap) 확장 리소스를 활용하여 다양한 세포주, 농도, 처리 시간 조건에서 측정된 약물 처리 전후 전사체 변화를 확보한다. 수집 데이터에는 약물 ID 및 구조 정보, 표적 연결 기, 처리 조건, log fold-change 기반 발현 변화 벡터, 시그니처 통계값 등이 포함되며, 동일 약물에 대해 가능한 한 **다중 조건 시그니처를 모두 포함하여** 조건별 일관성과 가변성을 평가할 수 있도록 구성한다.

또한, 역상 후보 약물의 기전적 타당성 평가를 위해 **약물-표적 및 경로 정보 데이터베이스**를 추가적으로 수집한다. DrugBank, ChEMBL, Therapeutic Target Database를 통해 약물의 1·2차 표적 단백질 정보를 확보하고, Reactome 및 KEGG를 활용해 신호·대사 경로를 매핑하며, UniProt 및 STRING 데이터베이스를 통해 단백질 상호작용(PPI) 네트워크 정보를 통합한다. 이를 통해 각 약물에 대해 표적 단백질, 관련 경로, 그리고 T세포 기능 및 탈진 조절 경로와의 연결성을 체계적으로 정리한다.

자료 수집 전반에서는 **재현성 확보, 이질성 및 편향 최소화, 생물학적 타당성 우선**이라는 세 가지 원칙을 적용한다. 모든 분석은 공개 데이터만을 사용하며, 데이터 출처·버전·다운로드 일자를 명확히 기록하고 분석 파이프라인 전 과정을 문서화·자동화한다. 서로 다른 코호트와 플랫폼의 데이터를 교차 포함하되, 샘플 수가 부족한 데이터는 보조 분석에 한정하고, 배치 효과 교정 후 통합 분석을 수행한다. 또한 단순한 통계적 유의성보다 **탈진 기전 및 면역학적 경로와의 정합성을 우선적으로 검토한다**.

6. 자료(데이터) 분석

본 연구의 자료 분석은 ① 탈진 T세포 발현 시그니처 구축, ② 약물-시그니처 역상성 분석, ③ 후보 약물 기전·경로 통합 해석의 세 축으로 구성되며, 전사체·단일세포 데이터와 LINCS 약물 시그니처 데이터를 연결하는 다단계 계산 생물학 파이프라인으로 수행된다. 분석 전 과정은 재현성 확보를 위해 표준화된 스크립트 및 파이프라인 형태로 구축하고, 단계별 산출물(정규화 데이터, 시그니처 벡터, 후보 랭킹, 네트워크 결과)을 기록·보존하여 후속 검증이 가능하도록 설계한다.

우선 전사체 데이터 전처리 및 품질 관리를 수행한다. Bulk RNA-seq 데이터는 품질검사(QC), adapter trimming, reference alignment, gene-level count matrix 생성 과정을 거친 뒤 TMM/TPM/DESeq2 기반 정규화를 적용한다. 서로 다른 코호트 간 비교 가능성을 확보하기 위해 ComBat 등 방법을 활용해 batch effect를 교정하고, 샘플 변동성 및 이상치(outlier)를 점검하여 통합 데이터셋의 안정성을 확보한다. 또한 환자군, 조직, 임상정보 등 메타데이터가 제공되는 경우 이를 공변량으로 포함하여 혼재된 교란요인(confounder)의 영향을 최소화한다. Single-cell RNA-seq 데이터는 셀/유전자 필터링, 표준화 및 scaling, 차원 축소(PCA, UMAP), 클러스터링 및 세포 타입 주석 과정을 통해 세포군을 정제하고, 비탈진 T세포, 전구형 탈진(Tpex), 말기 탈진(Tex terminal)으로 재분류한다. 이때 억제성 수용체 및 TOX-NR4A 발현 등 탈진 표지를 기준으로 라벨 정확도를 높이며, bulk 데이터와의 비교 가능성을 강화하기 위해 pseudo-bulk 변환을 병행한다.

다음으로 탈진 T세포 발현 시그니처를 도출한다. 비탈진군 대비 탈진군에서 차등발현 분석(DEG)을 수행하여 \log_2 fold change와 FDR 기반 유의성 검정을 적용하고, 상위 up-regulated 및 down-regulated 유전자군을 각각 추출한다. 시그니처의 견고성을 높이기 위해 bulk-single-cell 모두에서 일관되게 관찰되는 유전자를 우선 채택한다. 이후 GO biological process, KEGG/Reactome pathway 분석과 Gene Set Enrichment Analysis(GSEA)를 수행하여 TCR/NFAT 신호, 대사 재프로그래밍, 염증/면역 조절 경로 등 기능 모듈 단위의 특징을 규명한다. 추가로 WGCNA 기반 공발현 네트워크 분석을 적용하여 모듈별 hub gene을 도출하고, 모듈과 탈진 정도 간 상관성을 평가한다. 이러한 결과를 종합하여 “탈진 시그니처 = 상위 Up/Down 유전자 + 공발현 모듈 + 핵심 경로 가중치” 형태의 표준화된 시그니처 벡터로 정식화하며, 이는 이후 약물 시그니처 매칭의 기준(reference signature)으로 활용된다.

본 연구의 핵심 단계는 LINCS 기반 약물-시그니처 역상성 분석이다. LINCS L1000에서 약물별로 다양한 세포주, 농도, 시간 조건에서 산출된 log fold-change 벡터를 수집하며, 조건별 시그니처와 평균화 시그니처를 모두 보존하여 조건 의존성을 함께 평가한다. 조건 간 변동성을 정량화하고, 반복 조건에서 방향성이 일관된 약물에 가중치를 부여함으로써 신뢰도 높은 후보를 우선적으로 선별한다. 약물 시그니처와 탈진 시그니처 간 관계는 Pearson 상관계수, Spearman rank 상관계수, cosine similarity, connectivity score(가중 랭크 기반) 등 복수 지표로 계산하며, 지표 간 합의(consensus score)를 산출하여 단일 지표 의존 위험을 최소화한다. 이후 상관계수가 유의하게 음수이고, 여러 세포주·조건에서 방향성이 일관되며, 탈진 시그니처 핵심 유전자군에서 집중적인 반대 패턴을 보이고, 통계 지표·효과크기·재현성 종합 점수가 높은 약물을 중심으로 후보를 우선 순위화한다. 선별 결과는 1차 핵심 후보군과 2차 보조 후보군으로 계층화하여 리스트를 구축한다.

도출된 역상 후보 약물에 대해서는 기전·경로 통합 분석을 수행한다. DrugBank/ChEMBL/TTD 기반으로 표적 단백질을 매핑하고, Reactome/KEGG 기반으로 신호경로를 연결하며, STRING 등 PPI 네트워크 분석을 통해 표적 간 상호작용 구조를 파악한다. 특히 후보 약물 표적이 TOX-NR4A 축, NFAT/TCR 신호, 대사·염증 경로에서 어떤 위치에 연결되는지, 그리고 탈진 하위군(Tpex vs terminal Tex)에 선택적 영향을 미칠 가능성이 있는지를 중점적으로 평가한다. 각 약물에 대해 “탈진 완화 가능성에 대한 기전적 가설”을 문장 단위로 정리하고, 선행 실험 및 문헌과의 교차 검증을 통해 생물학적 일관성을 강화한다.

마지막으로 분석 신뢰도를 높이기 위해 추가 검증 및 민감도 분석을 수행한다. 데이터셋 교차검증(leave-one-dataset-out), 시그니처 크기 변화에 대한 민감도 분석, 특정 유전자 제거 또는 가중치 변경에 대한 강건성 평가를 적용하여 결과의 안정성을 확인한다. 또한 암종별 subgroup 분석을 통해 특정 종양 환경에서 후보 약물의 효과가 더 크게 나타날 가능성과, 면역관문억제제 등과의 병용요법 후보 가능성을 탐색적 수준에서 검토한다.

7. 논문 작성

전사체 시그니처 기반 T세포 탈진 억제 약물 재창출 후보 예측 LINCS 발현 패턴 역상성 분석 접근

암 및 만성 감염 환경에서 관찰되는 T세포 탈진(T cell exhaustion)은 면역 기능 저하와 치료 반응 제한을 야기하는 핵심 요인으로, PD-1, TIM-3 등 억제성 수용체의 공발현과 TOX-NR4A 전사인자 축에 의해 특징지어지는 안정화된 전사체 프로그램으로 이해된다. 본 연구는 탈진 T세포의 전사체 발현 시그니처를 정량적으로 정의하고, LINCS(L1000) 약물 발현 시그니처와의 역상(negative correlation) 기반 비교 분석을 통해 해당 시그니처를 비탈진/기능성 상태로 역전시킬 잠재력을 지닌 기존 승인 약물 재창출 후보를 계산적으로 도출하는 것을 목표로 하였다. Bulk 및 single-cell RNA-seq 데이터를 통합 분석하여 탈진 및 비탈진 T세포 간 차등발현 유전자와 공발현 모듈을 규명하고 이를 표준화된 시그니처 벡터로 구성하였다. 이후 LINCS 약물 시그니처와의 상관 및 유사도 분석을 통해 일관된 음의 상관 패턴을 보이는 후보 약물군을 선별하고, 표적·경로·전사인자 네트워크 관점에서 기전적 타당성을 해석하였다. 본 연구는 탈진 프로그램을 시스템 수준에서 조절 가능한 전사체 상태로 재해석하고, 기전 가설이 수반된 약물 재창출 후보군을 제시한다는 점에서 면역치료 및 시스템 약리학 연구에 기여한다.

T세포 탈진은 지속적 항원 노출 환경에서 나타나는 기능 저하 상태로, 종양 미세환경과 만성 감염에서 치료 반응 한계 및 내성 형성과 밀접하게 연관된다. 면역관문억제제는 탈진 억제 신호를 부분적으로 회복시키는 데 기여했으나, 반응률의 제한, 면역 이상반응, 재발성 내성이라는 구조적 한계가 여전히 존재한다. 최근 연구는 탈진이 단순한 기능 소실이 아니라 TOX-NR4A 전사인자 중심의 전사체 프로그램에 의해 안정화된 세포 상태임을 보여준다. 그러나 기존 연구는 특정 경로나 개별 표적 중심 접근에 머물러 있으며, 탈진 시그니처 전체를 역전시키는 시스템 수준의 약물 재창출 전략은 충분히 정립되지 않았다. 이에 본 연구는 탈진 T세포의 정량적 발현 시그니처를 규명하고, LINCS 기반 발현 패턴 비교를 통해 역상성을 보이는 기존 승인 약물 후보를 체계적으로 탐색하고자 한다.

Bulk 및 single-cell RNA-seq 데이터를 수집하여 비탈진 T세포, 전구형 탈진(Tpex), 말기 탈진(Tex terminal)으로 재분류하였다. 차등발현 분석(DEG), GSEA, 공발현 네트워크 분석을 수행하여 상위 up/down 유전자군과 모듈·경로 가중치를 포함한 시그니처 벡터를 구성하였다. LINCS L1000 약물 시그니처에 대해 Pearson 및 Spearman 상관, cosine similarity, connectivity score를 복합적으로 산출하고, 다중 조건에서 일관된 음의 상관(역상성)을 보이는 약물을 우선 후보군으로 계층화하였다. 후보군에 대해서는 DrugBank, KEGG, STRING을 활용하여 표적 단백질과 관련 경로, TOX-NR4A/NFAT/대사·염증 축과의 연결성을 네트워크 수준에서 해석하였다. 민감도 분석과 데이터셋 교차검증을 통해 결과의 강건성을 평가하였다.

탈진 시그니처 분석 결과, 억제성 수용체 및 TOX-NR4A 축의 상향 조절과 기능성 사이토카인 및 대사 경로의 하향 조절 등 탈진 특이적 up/down 모듈이 확인되었으며, bulk와 single-cell 데이터 간 교차 일관성이 확보되었다. LINCS 기반 분석에서는 복수 지표 합의에 따른 상관 및 효과 크기 상위 약물군이 도출되었고, 일부 약물은 다중 세포주와 조건에서 안정적인 반대 발현 패턴을 보였다. 기전 통합 분석 결과, 상위 후보 다수는 NFAT 상류 신호, 대사 재프로그래밍, 염증성 사이토카인 네트워크 등 탈진 안정화 경로의 상·하류 모듈에 연결되는 양상을 보였으며, Tpex 보존 또는 terminal Tex 억제에 선택적으로 작용할 가능성이 제시되었다.

본 연구는 T세포 탈진을 정량적 전사체 시그니처로 재정의하고, LINCS 발현 패턴과의 역상성 분석을 약물 재창출 전략에 적용함으로써 단일 표적 중심 접근을 넘어 프로그램 수준의 조절 가능성을 제시하였다. 도출된 후보 약물은 면역관문억제제와의 병용 가능성 및 특정 암종·미세환경에서의 선택적 적용 가능성이라는 translational 가설을 제공한다. 다만 본 연구는 계산적 분석에 기반하므로, in vitro 및 in vivo 기능 검증과 용량·효과 평가가 후속 연구로 요구된다.

본 연구는 탈진 T세포 전사체 시그니처를 표준화하여 정의하고, LINCS 기반 역상성 지표를 활용한 약물 재창출 프레임워크를 제시함으로써 기전 해석이 수반된 후보 약물군을 도출하였다. 이는 면역치료 전략의 정교화와 시스템 생물학 기반 약물 재창출 연구의 중요한 출발점이 될 것으로 기대된다.

8. 자체 리뷰 및 수정

본 연구는 대규모 공개 전사체 데이터와 약물 발현 시그니처를 활용하는 **데이터 기반 계산생물학 연구**의 특성상, 연구 전 과정에서 **재현성, 통계적 타당성, 해석의 신중성을 핵심 원칙으로 삼아 단계별 자체 리뷰(self-review) 및 수정·보완 절차**를 체계적으로 운영한다. 자체 리뷰는 데이터 수준, 분석 방법론 수준, 결과 해석 수준의 세 단계로 구분하여 수행되며, 각 단계에서 확인된 한계와 불확실성은 즉시 추가 분석 또는 연구 설계 조정으로 연결된다.

먼저 데이터 수집 및 전처리 단계에서는 이질적인 출처의 전사체·단일세포·LINCS 약물 시그니처 데이터를 통합 활용하는 특성을 고려하여, 데이터 품질과 포함 기준에 대한 정기적인 점검을 시행한다. 모든 데이터셋에 대해 출처, 다운로드 버전, 생성 연도, 메타데이터 가용성을 체계적으로 정리하여 연구 노트 및 데이터 사양 문서(Data Sheet) 형태로 기록한다. 샘플 수가 부족하거나 실험 설계가 불명확한 데이터셋은 핵심 분석에서 제외하고, 보조 분석 또는 민감도 분석 전용 자료로 한정한다. 전처리 과정에서 수행된 품질관리(QC), 배치 효과 보정, 정규화 절차의 로그와 파라미터를 모두 저장함으로써, 동일한 결과를 재현할 수 있는 분석 파이프라인 상태를 유지한다. 자체 점검 과정에서 특정 데이터셋이 탈진 상태 정의 기준과 불일치하거나 노이즈 영향이 크다고 판단될 경우, 해당 데이터셋은 분석에서 제외하거나 별도의 검증용 세트로 재분류한다.

다음으로 **분석 파이프라인 및 통계적 타당성 점검 단계**에서는 모델 의존성과 지표 선택 편향을 최소화하기 위해 복수의 분석 전략을 비교·검증한다. 차등발현 분석, GSEA, 공발현 네트워크 분석에서는 서로 다른 통계 모형과 패키지를 병행 적용하여 핵심 결과의 일관성을 점검한다. LINCS 기반 약물-시그니처 비교 과정에서도 Pearson 상관계수, Spearman 순위 상관계수, cosine similarity, connectivity score 등 **다중 지표 합의(consensus score)** 접근을 적용함으로써 단일 지표 의존에 따른 왜곡 가능성을 줄인다. 또한 leave-one-dataset-out 교차검증, 시그니처 크기 변화, 핵심 유전자 제거 등의 민감도 분석을 통해 결과의 강건성을 평가한다. 특정 가중치 설정이나 임계값이 후보 약물 순위에 과도한 영향을 미치는 것으로 확인될 경우, 해당 파라미터를 조정하여 재분석·재평가를 반복한다.

결과 해석 및 기전 연결성 검증 단계에서는 통계적 신호와 생물학적 타당성 간의 균형을 유지하는 데 초점을 둔다. 선별된 후보 약물의 표적 단백질과 관련 경로가 TOX-NR4A 전사인자 축, NFAT 신호 경로, 대사 및 염증 조절 경로 등 **T세포 탈진 조절의 핵심축과 실제로 연결되는지**를 문헌, 단백질 상호작용(PPI), 경로 데이터 기반으로 교차 검토한다. 특히 “발현 시그니처 역상성 = 탈진 억제 효과”라는 해석이 기계적 또는 과도한 일반화로 확장되지 않도록, 분석 가정과 한계, 불확실성을 명시적으로 기술한다. 또한 암종별·종양 미세환경별 맥락에서 후보 약물의 적용 가능 범위와 제한점을 구분하여 제시한다. 필요시, 일부 후보 약물에 대해서는 독립 데이터셋 재분석이나 추가 보정 분석을 수행하여 기전 가설의 일관성을 재확인한다.

연구 산출물(논문 및 보고서) 작성 단계에서도 자체 리뷰를 강화한다. 데이터, 코드, 분석 흐름을 재현 가능한 형태로 문서화하고, 핵심 표와 그림이 자동 생성되도록 구성한다. 결과 제시는 수치와 근거 중심으로 유지하며, 해석과 가설은 논의(Discussion) 파트에서 명확히 분리하여 기술한다. 또한 연구의 한계, 잠재적 편향, 외삽 위험을 명시적으로 서술함으로써 과도한 결론 도출을 지양한다. 초안 작성 후에는 내부 리뷰를 통해 용어 정의, 논리 흐름, 해석 강도를 점검하고, 필요시 서술 완화, 근거 보강, 구조 재배치 등의 수정 작업을 수행한다.

이와 같은 **체계적 자체 리뷰 및 수정 절차**를 통해 본 연구는 결과의 강건성·재현성·투명성을 제고하고, 통계적 신호에 대한 과잉 해석이나 과대 일반화 위험을 최소화한다. 동시에 후속 in vitro 및 in vivo 실험 연구로의 연결 가능성이 높은 기전 가설만을 선별적으로 제시함으로써, 연구의 학문적 신뢰도와 향후 논문 게재 및 검증 연구의 효율성을 동시에 높이는 관리 체계로 자리매김 할 것이다.

별첨1-5 [제출물 양식] 활용데이터 목록 [Track1]

연 번	데이터 개요	데이터 구분 (공공/개인/AI 합성)	데이터 수집경로 (AI 합성 방법)
1	■ 암 및 만성 감염 환경에서 CD8 ⁺ T세포의 탈진 상태와 기능성 상태를 비교하기 위한 bulk RNA-seq 전사체 데이터. PD-1, TIM-3, LAG-3 등 탈진 표지 발현이 명확히 구분된 샘플을 포함.	공공데이터	■ https://www.ncbi.nlm.nih.gov/geo/
2	■ 종양 침윤 CD8 ⁺ T세포 및 만성 감염 환경 T세포의 single-cell RNA-seq 데이터로, 비탈진 T세포, 전구형 탈진 T세포(Tpex), 말기 탈진 T세포(Terminal Tex)를 세포 단위로 구분할 수 있는 전사체 데이터.	공공데이터	■ https://singlecell.broadinstitute.org/single_cell
3	■ 약물 처리 전후 세포의 유전자 발현 변화를 측정된 LINCS L1000 약물 전사체 시그니처 데이터. 다양한 세포주, 농도, 처리 시간 조건에서의 발현 변화 정보를 포함.	공공데이터	■ https://lincsproject.org/
4	■ 약물-표적 단백질, 작용 기전, 승인 적응증 정보를 포함한 약물 메타데이터 및 표적 데이터. 역상성 분석으로 도출된 후보 약물의 기전적 타당성 평가에 활용됨.	공공데이터	■ https://go.drugbank.com/
5