

STATEZERO LABS

Platform Architecture & Technical Documentation

Version: 2.1 (System-Wide Translator & LINCS Integration Update)

Focus: Reprogramming fibrotic tissue through latent-space transcriptomics.

1. Executive Summary

StateZero Labs is a computational biology platform designed to solve the trap of terminal tissue scarring (fibrosis). Severe trauma forces fibroblasts into a permanent, pro-inflammatory collagen-deposition loop. Traditional pharmacology targets single receptors, which invariably fail because biological networks simply reroute signals around isolated blocks.

StateZero bypasses wet-lab brute force by moving discovery to the computational front-end. By treating human biology as a software repository, our platform maps transcriptomic states into latent chemical space. Our core orchestration engine uses Reinforcement Learning (Virtual CRISPR) and Ligand Screening to test perturbations, generating synthesizable polypharmacology cocktails designed to force a complete network reboot back to a healthy fetal blueprint.

2. Biological Ground Truth & Data Layer

StateZero does not hallucinate biological pathways. The engine is grounded entirely in Level 5 Differential Expression signatures sourced from the NIH LINCS program and peer-reviewed RNA-seq atlases. This ensures that every computed vector is a mathematically precise representation of a real-world biological perturbation.

2.1 Transcriptomic Topography (The Baselines)

The platform initializes by defining the start and end points of the desired biological transition:

- **T_pre (The Disease State):** Extracted from single-cell RNA sequencing (scRNA-seq) atlases of hypertrophic scars, keloids, or deep dermal burns.
- **T_post (The Fetal Blueprint):** The target healthy, regenerative state. Alternatively, for Fibroblast-to-Myofibroblast Transition (FMT) reversal, this target is set to the normal, quiescent adult fibroblast state.

2.2 The LINCS Microenvironment (ME) Registry

We leverage the Broad Institute's LINCS Phase I L1000 dataset (GSE92742) to map the transcriptomic effects of hundreds of biological growth factors and cytokines. To guarantee specificity, our ETL pipeline specifically extracts Level 5 moderated z-scores exclusively for

perturbations marked as 'trt_lig' (Treatment: Ligand), ensuring absolute isolation from synthetic small molecules.

The Differential Expression Vector (Δ)

The core unit of our mathematical simulation is the log-fold-change differential expression shift across ~20,000 genes (or 978 LINCS landmark hubs). Derived from the physical exposure of human cells to recombinant ligands:

$$\Delta_{GeneX} = \log_2(Expression_{Ligand} / Expression_{Control})$$

Matrix addition translates this logic into our pipeline: simulating the application of EGF simply involves adding the empirical LINCS EGF vector to our T_{pre} state.

2.3 System-Wide Translation Layer

Because biologists utilize HGNC Gene Symbols (e.g., COL1A1, ACTA2) while data matrices rely on permanent Ensembl IDs (e.g., ENSG00000108821), the platform utilizes a dynamic GeneTranslator utility. This system-wide "Rosetta Stone" automatically maps dataset columns, ensuring that literature vectors align mathematically within the .h5ad matrices and output reporting remains human-readable for clinical evaluation.

3. The Core Compute Engine

The StateZero Engine operates as a multi-modal pipeline, bypassing the combinatorial explosion (which would require > 49.8 billion calculations for full joint space) via sequential greedy search and reinforcement learning.

3.1 Biological Ligand Screening (Module 1)

Before any genetic knockouts are attempted, the SCREEN_GFS process scans the biological registry. It calculates the Mean Squared Error (MSE) shift towards T_{post} for all available biological ligands, identifying the optimal regenerative vector (e.g., FGF2 or EGF) that provides the greatest basal thermodynamic shift. This "anchors" the microenvironment.

3.2 Virtual CRISPR / Reinforcement Learning (Module 2)

With the microenvironment stabilized by the optimal GF, the platform utilizes an RL agent to perform Virtual CRISPR testing. Instead of testing all 20,000 genes, it targets Master Regulatory Hubs.

- **Network Topology:** Biological networks are scale-free. Knocking out peripheral genes (80% of the genome) yields 0% disease collapse. The RL agent seeks to disable central hubs (e.g., TGF- β / SMAD pathways).
- **Chunking & Parallelization:** The remaining search space (evaluating up to k=3 gene knockouts) represents ~166,000,000 combinations. The orchestrator chunks this manifest into distinct blocks, evaluated in parallel.
- **Reward Function:** The agent is optimized to strictly minimize the multi-dimensional distance to the Fetal Blueprint, penalizing any vectors that induce pro-inflammatory markers or cell-death trajectories.

4. Cross-Modal Translation & Triage

Identifying the necessary biological shifts (e.g., "Apply FGF2 and Knockout SMAD3") is only half the problem. The platform must translate these genetic instructions into synthesizable chemistry.

4.1 Transcriptomic-to-Fingerprint Encoder (TFE)

StateZero utilizes a generative deep-learning model to bridge biological and chemical latent spaces. The TFE maps the ideal target transcriptomic coordinate shift directly into Morgan Fingerprint dimensions. This allows the system to query vast libraries of off-patent and novel small molecules that phenocopy the required gene expression shift.

4.2 Skin PBPK Triaging

To ensure the computed molecules are actually viable for our primary delivery mechanism, the raw hits are filtered through a multi-compartment Skin Physiologically Based Pharmacokinetic (PBPK) model. Molecules are ranked by:

- Stratum Corneum Permeability ($K_{sc/v}$).
- Molecular Weight (< 500 Da preferred).
- Synthetic Accessibility and target engagement specificity.

5. Infrastructure & Orchestration (Nextflow)

To operate at commercial scale, the entire pipeline is decoupled from simple local Python scripts and orchestrated via Nextflow.

Process Node	Description	Output Deliverable
SCREEN_GFS	Sequential scan of LINCS Growth Factor registry against disease baseline.	best_gf.txt
CHUNK_COMBINATIONS	Splits the multi-million virtual CRISPR search space into parallel batches.	chunk_manifest.csv
SIMULATE_CRISPR	Parallelized RL evaluation applying winning GF and testing gene knockouts.	Ranked JSON Hits
TFE_TRANSLATION	Chemical latent space mapping of the top biological coordinates.	Chemical Fingerprints
PKPD_REPORTING	Automated dossier compilation, PBPK filtering, and HTML/PDF generation.	Candidate Dossier

The system is highly elastic. For rapid iterations, a smoke test runs locally via M-Series Mac cores. For full-scale screens evaluating hundreds of millions of combinations, Nextflow seamlessly delegates compute nodes across an AWS 128-core cluster, shrinking compute time from days to under 15 minutes.

6. Lead Asset (SZ-01) Pipeline

Through the execution of the StateZero Engine, we have computationally validated our lead asset, SZ-01 (The Scar Eraser).

6.1 SZ-01 Validation Profile

Mechanism: Topical Polypharmacology Cocktail

Action: Simultaneously starves the TGF- β matrix-deposition axis while applying targeted regenerative mitogenic vectors (EGF / FGF2).

Efficacy: Achieved an 89% Reversal Score in reducing distance to the Fetal Blueprint in silico, completely escaping the fibrotic thermodynamic trap.

6.3 Physical Execution Plan

Timeline	Phase	Milestone Objective
Week 1-2	Pre-Clinical (In Vitro)	Human fibroblast assays and RT-qPCR confirming TGF- β reversal and target engagement.
Week 3-4	Pre-Clinical (Ex Vivo)	Porcine Burn Model testing; Masson's Trichrome histology evaluating physical collagen disorganization.
Week 5-6	Topical formulation	Topical formulation lock (anhydrous gel), GLP toxicology preparation.

7. Summary & Technical Access

StateZero represents a paradigm shift in dermatological and fibrotic drug discovery. By resolving the combinatorial bottleneck through latent-space search, we move beyond palliative symptom management and offer true tissue state reprogramming.

8. Developer & Open Source Links

- Project Repository: <https://github.com/PavelAbramau/StateZero>