SCIENCE FAIR LOGBOOK

(2025-2026)

ERIN D'SOUZA

Grade: 10

INNOVATION PROJECT: SCAR-GATE: A Logic-Gated Pipeline for the In Silico Discovery of Safe, Transient CAR-T Targets in Cardiac Fibrosis Reversal		
Click here for link to my Engineering Decisions and Code Architecture record		
Click here for link to Scientific Literature		
Click here for link to the GitHub Repository		
Date	**Brainstorming**	**Research/Sources**
June 7	Exploring Multi-omics	My Science Fair ML model last year for breast cancer miRNA was a solid foundation, but I want to go deeper. For next year's science fair, I definitely want to explore multi-omics more – I feel that is the future. I explored a few directions, but I've decided to stay strictly within bioinformatics and computational biology—that's where my passion is. I want a project where I can further build my python skills learn more algorithmic limits and handle high-dimensional data – because that's where I had lots of challenges last year. For the summer, will try to download a few datasets and explore Scanpy, NumPy, Scipy, pandas so forth and so on. Also may be look at a some biopython or other bioinformatic packages.
June 8	Exploring Multi-omics	Learnt a lot on multi-omics from these Youtube videos http://www.youtube.com/watch?v=Xsyp0qqKzkY http://www.youtube.com/watch?v=UsOBxRZH8j4
June 14	Learning and experimenting with Python and Datasets	I've been diving into some Python exercises specifically aimed at handling massive datasets. Last year's project taught me that ML is only as good as the data ingestion pipeline, so I'm focusing on libraries like Pandas and NumPy for efficient matrix manipulation. I practiced using **Dask** for parallel computing and explored **Sparse Matrix** storage to see how much memory I could save. I have a limitations with computing power with my laptop so I have to learn to be efficient especially with Data Manipulation

		Reading up on the book: Bioinformatics with Python Cookbook" (Tiago Antão, 4th Edition)
June 15	Explored - AI-Driven ncRNA Discovery, **Protein-Protein Interaction (PPI) Mapping**	Investigated -Non-coding RNA—the "dark matter" of our genome that doesn't make proteins but regulates almost everything. I've been researching how **Reinforcement Learning (RL)** could be used to predict their functions. Instead of just classifying sequences, an RL agent could "explore" the folding landscape of an RNA molecule to find the most stable, functional state. It's a massive computational puzzle, but could be potential project if I want to dosomething in Precision Medicine **Research Paper:** *Predicting non-coding RNA function using Artificial Intelligence (2024)* – This study benchmarks the latest AI models for deciphering the complex roles of ncRNAs in disease **Protein-Protein Interaction (PPI) Mapping** Everything in the cell happens because proteins "talk" to each other, and mapping this interactome is essential for drug discovery. I've been looking at how **Graph Neural Networks (GNNs)** can predict if two proteins will bind based on their 3D surfaces and chemical "fingerprints." After seeing what AlphaFold did for structures, using deep learning to map the entire network of human protein interactions feels like the next logical frontier for my pipeline engineering. **Research Paper:** *Recent advances in deep learning for protein-protein interaction: a review (2025)* – A recent deep dive into how attention mechanisms and GNNs are revolutionizing PPI prediction. I found this literature quite interesting (but also lots more questions….)
June 21	Explored - **Multi-Omic Integration for T-cell Exhaustion**	**Multi-Omic Integration for T-cell Exhaustion** I've narrowed my single-cell focus to the integration of transcriptomics and proteomics specifically to study **T-cell exhaustion**. In chronic environments like a scarred heart, T-cells often just "give up." I'm looking at how we can use Joint Embedding techniques to map RNA signals directly to surface protein markers (ADTs) to find the "point of no return" for a therapeutic cell. If I can predict exhaustion computationally, I can engineer a more resilient AND-gate. I absolutely loved the concept of T-Cell and it s applications and tough the whole concept is fascinating, im afraid it will take me in to Bio Engineering – Logic gates and all of that. I really don't mind that. I would like tokeep it purely bioinformatic if I could. But I'm certainly coming back and revisiting this **Research Paper:** *Multi-omic atlas of T cell exhaustion (2024)* – A definitive look at the epigenetic and transcriptomic landscape of T-cell failure.
June 21	Explored - **Protein Structure Prediction via Diffusion Models**	**Protein Structure Prediction via Diffusion Models** After the success of AlphaFold, the field has moved toward **Generative Diffusion Models** for protein design. Instead of just predicting a structure, these models can "imagine" entirely new proteins that don't exist in nature. I've been researching how this could be used to design custom "hinge" and "spacer" regions for CAR-T receptors that are perfectly optimized for the 15nm synaptic gap. It's essentially moving from discovery to true de novo biological engineering. Altough this is CAR-T cell related it has more to do with synthetic proteins and I am not sure about that. Atleast for now.

		• **Research Paper:** *De novo protein design using geometric deep learning (2025)* – This paper explores how "Chroma" and other diffusion frameworks are being used to build functional, synthetic proteins.
June 22	Brainstorming -Topic -The Multi-Omics MS Puzzle UMAP – Auto encoders GNN - exploration	I've moved past basic optimization to explore how high-dimensional biological data actually behaves. I've been reading into **Manifold Learning**, specifically **UMAP (Uniform Manifold Approximation and Projection)** I also took a deep dive into Autoencoders (AEs). Instead of just "classifying" data, I'm interested in using the bottleneck layer for Dimensionality Reduction. Finally, I checked out Graph Neural Networks (GNNs). Since proteins and genes don't work in isolation—they exist in interaction networks—treating the data as a graph rather than a flat table makes so much more sense. It's a steep learning curve, but it feels like the right direction for a multi-omics project for now. I might register for an advanced python course online – has to be data science related or something similar
June 28	Brainstorming - The Multi-Omics MS Puzzle	Just throwing some initial ideas at the wall for science fair. I'm exploring a computational biology focus, starting with Multiple Sclerosis (MS). It's surprising how little we know about its exact epigenetic triggers. I was thinking: what if I built a pipeline to integrate single-cell RNA-Seq with ATAC-Seq data from MS lesions? The RNA shows what genes are active, but ATAC-seq is the cool part. It's basically looking at the open chromatin to see the physical switches turning those genes on. Mapping these regulatory networks is a huge data challenge, which I love. The problem is finding high-quality paired human datasets. Plus, merging those massive matrices hmmm.. need a new laptop or computer. I'll keep it on the shortlist while I check data availability.
July 2	Exploring Precision Medicine & Systems Biology	I'm looking into how we can move beyond genetic-only testing by integrating orthogonal molecular data, specifically the transition from single-omics to clinical systems medicine. I'm specifically interested in identifying heart-specific "proteoforms"—post-translationally modified proteins—that could act as high-fidelity biomarkers for early-stage structural failure. By mapping these onto biological "digital twins," we can computationally simulate drug-target interactions to predict efficacy before a single dose is administered. Readup on: Precision Cardiovascular Medicine: Shifting the Innovation Paradigm (2025) Revolutionizing Cardiac Fibrosis Treatment: The Potential of Personalized CAR T-cell *https://www.frontiersin.org/journals/science/articles/10.3389/fsci.2025.1474469/full?utm_source=facebook&utm_medium=cpc&utm_campaign=imp_c1-pm_02-26_fsci_en_lal_pcit_1_req13&utm_content=bancar&utm_id=120240115054900697_v2_s01&utm_term=120240115054900697&fbclid=PAZXh0bgNhZW0BMABhZGlkAastv7GvK1lzcnRjBmFwcF9pZA8xMjQwMjQ1NzQyODc0MTQAAacy4dvj1EccHsIO-*

		_I99b4ZHFT7UNcRm5STOZWvSCr9epkRvAr5oa-UxuZCsA_aem_TXUzipe9SiV21UbRn5OvSg *https://pmc.ncbi.nlm.nih.gov/articles/PMC12372244/*
July 3	Explored – Proteomics, Transcriptomics and Spatial Transcriptomics	I spent the day digging into the interplay between proteomics and transcriptomics, but I really like **Spatial Transcriptomics (ST)**. Standard single-cell sequencing tells you *what* cells are there, but ST tells you *where* they are in the tissue architecture. In a scarred heart, knowing exactly which cells are sitting right on the edge of the fibrotic lesion is a total game-changer for targeting – May be I can find a some research papers I like and further build on we ll see…. Readup on: 1. *Spatial Transcriptomics in the Heart (2024)* – Nature Reviews Cardiology 2. *Spatial Mapping of the Human Myocardium (2022)* – Nature
July 4	Explored – Success of CAR-T Cell in Leukemia	I spent some time today digging into the success story of CAR-T: B-cell acute lymphoblastic leukemia (ALL). It's basically the gold standard for synthetic immunology, • CD19 is the perfect target because it's expressed on almost all B-cells but not on essential stem cells; you can live without B-cells (with help), but you can't live without a heart. • **Liquid vs. Solid:** Leukemia is a "liquid" tumor, so the T-cells have direct access in the blood. For my project, the challenge is getting those same cells to "burrow" into a dense, stiff fibrotic scar. • **Serial Killing:** A single CAR-T cell in leukemia can kill up to 1,000 tumor cells. I need to model if that kind of efficiency is possible in the high-pressure environment of the myocardium. • **Cytokine Release Syndrome (CRS):** **Key Research Paper:** • *Chimeric Antigen Receptor T Cells for Sustainable Remissions in Leukemia (2024)* – New England Journal of Medicine I plan to research further and see if I can find any interesting topics to research on ….
July 5	Potential Project – CAR-T cell and Cardiac Fibrosis	Very interesting paper I read today – which indicated CAR-T can be applied for cardiac fibrosis and reversed it in mice. Up until now, I thought cardiac fibrosis was just a permanent mechanical failure—once the heart is scarred, it's over. But this paper by **Aghajanian et al. (2019)** shows that if you program T-cells to target a specific protein called **Fibroblast Activation Protein (FAP)**, you can actually clear the excess scar tissue in mice and restore their heart function. It's the first time I've seen immunotherapy treated as a regenerative tool rather than just a "search and destroy" type of thing for cancer. However, as I read deeper, I noticed that FAP isn't exclusive to the heart. It's in the skin, the bone marrow, and other organs. If a CAR-T cell is a "living drug" that stays in your body foreverand so a target for for off-target toxicity. **Research Paper:** • *Targeting cardiac fibrosis with engineered T cells (2019)*

July 12	Research on Cardiac Fibrosis – Evaluating feasibility of CAR-T cells and Cardiac Fibrosis related project.	Did some research on Cardiac Fibrosis / Heart disease Here are the key points I discovered: • **Irreversibility:** Once the heart is scarred, the damage is considered permanent. The adult human heart has almost zero regenerative capacity, so it patches injuries with stiff, non-contractile scar tissue rather than new muscle. • **Its Palliative -Not Curative:** Every single medication we currently have—ACE inhibitors, beta-blockers, diuretics—is just damage control They slow the decline and manage symptoms, but they don't actually fix the underlying structural failure. • **Mechanical Feedback Loop:** The stiffness of the scar itself is a signal. It tricks healthy cells into thinking they need to make *more* scar tissue, creating a vicious cycle of progressive stiffening that we can't stop. • **The Economic Crisis:** Because we are keeping millions of people alive in a state of chronic heart failure, the global economic burden is exploding—projected to surpass **$1 trillion** by 2030. • **Success as a Burden:** Modern medicine has mastered the acute phase (saving the life), but we are failing the .We've essentially traded a quick death for a long and expensive, and debilitating decline. • **A "Biological Dead End":** Standard pharmacology has hit a wall. To actually reverse this, we don't need another pill that alters blood pressure. **Articles Read:** 1. Breaking the Cycle: A New Approach to Treating Cardiac Fibrosis – Stanford Medicine (2025) 2. Americans' Challenges with Health Care Costs – KFF Health News (2026) (Gives context on the $858B crisis) 3. Heart Failure Reframed: Calling Clinicians to Shift Focus – Patient Care Online (2026) 4. Deciphering the cardiac immune-fibroblast niche – European Society of Cardiology (2024)
July 19	**Mapping the CAR-T Landscape – Trends and Applications**	I spent the day reading about cardiac research to look at the broader CAR-T field through recent 2025 and early 2026 journals. • **The Solid Tumor Wall:** Durable responses are still rare. It's not just about finding the right antigen; it's about microenvironment penetration.Between physical barriers like stiff extracellular matrices and chemical ones like hypoxia, the T-cells are often exhausted before they even reach the target. • **Antigen Escape:** About 30% of cases fail due to antigen escape—where the disease simply stops expressing the target protein. I'm seeing more talk about tandem CARs- like two antigens • **The Safety Debate :** There's some discussion about (Nature Medicine, Feb 2025) about rare T-cell cancers post-treatment. Most experts think it's linked to prior chemo rather than the CAR-T itself, but ong term safety is still not determined. • **The Cost Barrier:** We're still looking at ~$400,000 for autologous patient-specific treatments = very expensive. There are some new models for off-the-shelf-allogeneic options or decentralized manufacturing or something like that.. haven't looked in it too much.

		• **Beyond Cancer:** Apparently there is lot of potential beyond cancer in immunotherapy, autoimmunity and fibrosis is but still it is relatively a new frontier. • **Research Articles:** • *CAR-T Cell Therapy in 2025: Breakthroughs, Barriers, and the Road Ahead* • *Strategies for the Future of CAR-T Therapy Development in 2026* • *New Analysis Finds No Link Between CAR T-Cell Therapy, Secondary Cancers (Nature Medicine 2025 Review)* • *Decentralized manufacturing: Unlocking CAR-T's potential (2026 Report)*
July 19	Evaluated 3 CAR-T Related projects.	Evaluated these options for the CAR-T related project **GNN-Based Solid Tumor Infiltration:** This uses Graph Neural Networks (GNNs) and reinforcement learning to simulate the dynamic tumor microenvironment. Since solid tumor efficacy is currently below 10%, modeling these physical and chemical barriers as a dynamic network—rather than a static analysis—This could significantly improve infiltration strategies. **Predicting Antigen Escape with Multi-Omics:** This targets the 30-50% of patients who relapse when disease antigens evolve. By using a CNN-attention mechanism to forecast temporal antigen changes, we could move beyond current static tools and design more durable, multi-antigen CARs. **Transformer-Based Discovery for AML:** Acute Myeloid Leukemia (AML) lacks ideal targets that don't cause bone marrow toxicity. Using NLP-inspired transformers for epitope "mining" could accelerate the discovery of unique antigen pairs, offering a more precise approach than standard single-cell RNA-seq methods. **Research Literature:** 1. *Graph Neural Networks in Cancer Research (2024)* – A review of how GNNs are being applied to model complex cellular interactions. 2. *Mechanisms of Antigen Escape in CAR T-Cell Therapy (2023)* – Detailed analysis of why single-antigen therapies often fail in the long term. Note: At this point CAR-T and Cardiac Fibrosis still is the best I feel.
July 20	Evaluated – Other options to Cardiac Fibrosis other than CAR-T	**Evaluated the various Contenders for Fibrosis Reversal** I've spent the day searching reviewing scientific jourals related to cardiac fibrosis treatment. My goal was to see which approach offers the best chance of moving beyond simple management toward actual reversal. **These are the Findings:** **Direct Cardiac Reprogramming:** Uses transcription factors (Mef2c, Gata4, Tbx5) to "flip" resident fibroblasts into functioning heart muscle cells. This could be Great for dual-action as in deleting scars and making new muscle, but efficiency is stuck below 10%. Plus, getting these new cells to beat in sync with the rest of the heart without causing arrhythmias hard. **Targeting Signaling Pathways (TGF-β, RAAS, Epigenetics):** Uses small molecules or RAAS inhibitors (like ACE inhibitors) to shut down the on- switch for myofibroblasts. This is the most established therapeutic , but it's more about slowing the decline than a total cure. Broadly blocking pathways like TGF-β can mess with normal wound healing elsewhere in the body – don't thin I want to this route.

		Cell-Based and Immunotherapies (CAR-T): This is Reprogramming T-cells to identify and physically remove activated FAP+ fibroblasts. This is the most transformative option. Preclinical 2025-2026 data shows dramatic reversal of chronic failure. By using mRNA-lipid nanoparticles for delivery, one can avoid the massive costs of traditional CAR-T while maintaining precision. **Adjunct and Supportive Therapies:** Using herbal extracts or mechanical devices (VADs) to unload the heart and promote reverse remodeling. These are great, but they don't solve the biological problem of the scar itself.
		Conclusion: - CART for Cardiac Fibrosis is better, it is still the only solution that offers reversal of fibrosis andnot just manage it but clinical advancement gap remains. While reprogramming is greatfor regeneration, its inefficiency is low and is also a nascent therapeutic . Pathway targeting is just amangmeent of the disease. Plus CAR-T offers precision **Key Scientific Journals:** 1. *In vivo cardiac reprogramming: Progress and challenges (2025)* – Nature Reviews Cardiology. 2. *Epigenetic regulators as therapeutic targets in cardiac fibrosis (2024)* – Circulation Research. 3. *mRNA-LNP mediated in vivo CAR-T generation for heart failure (2025)* – Science (Follow-up study). 4. *Pharmacological management of cardiac fibrosis: Beyond RAAS inhibition (2026)* – Current Heart Failure Reports.
July 26 - 27	Studied CAR-T Cell and manufacturing	Today I analyzed the production workflow for CAR-T cells just to understand it better. The cycle begins with apheresis, a specialized blood filtration used to isolate a patient's T-cells before returning the remaining blood components. Once these cells reach the lab, they require activation—essentially a biochemical alert often triggered by magnetic beads to make them receptive to genetic modification. Viral vectors or mRNA-LNPs for more transient applications are used to integrate the genetic instructions for the Chimeric Antigen Receptor into the T-cell genome. The resulting receptor is a sophisticated piece of synthetic biology: an extracellular binding domain designed to recognize the target antigen and an intracellular signaling domain that triggers the T-cell's cytotoxic response. After the genetic modification the cells are placed in bioreactors for expansion, allowing them to multiply until they reach a therapeutic dose. The final stages are just as critical as the beginning. The engineered cells are cryopreserved and shipped back to the clinical site. Meanwhile, the patient undergoes something called as lymphodepletion—a targeted chemotherapy regimen—to reduce the existing immune population and create a permissive environment for the new cells. The complexity and cost of this ex vivo process are immense. This is why I feel my project focus should be on invivo **Literature Articles** • CAR T Cells: Engineering Immune Cells to Treat Cancer (NCI 2025 Update)

		• The Manufacturing Journey of CAR-T Cellular Therapy (BioPharm International 2025) • Scalable Manufacturing of CAR T Cells (Nature 2024) • Precision Immunology: The Next Generation of Cell Therapy (2026 Report) Watched this youtube video - https://www.youtube.com/watch?v=lx3EhUVHI-Q **Summary of findings and conclusion regarding CAR-T** **Positives** - Its very effect for Leukemia and has a proven track record related to that. Lots of research papers regarding this. - Resesrch shows that CAR-T can be extended to solid tumors but with caveats - Some research papers show CAR-T can be extended to cardiac fibrosis but research is lacking or not fully developed. - So by defention CAR-T can be extended to any organ fibrosis . So this technology has a huge upside orpotential - It can be highly specific so that is a plus **Limitations** • CRS - cytokine release syndrome a big concern. • Not enough extensive approved therapeutics • High cost and long production time
August W 1	Project Selection decsion	I feel with all the research and reading that I have done so far, I am fairly confident that I want further develop something related to CAR-T Thearpy and Cardiac Fibrosis – The idea feels so compelling. There is huge potential for innovation but at the same time there are so many unknowns, altough there are many research papers and literature regarding this it still under development. So that's a positive and a downside at the same time. I most certainly do want to pursue the computational approach. I don't think I would want to do wet labs as well simply because of the scope and also accessibility to a wetlab. In addition since its Human Heart tissue it will be close to impossible to get to access to a wetlab. Even I pursue murine approach / tissues towards this project, wetlab access is still an issue. **So computational approach it is!** That brings me to the point of – Limited computing capacity. I have to be careful in tool or method selection especially methods like GNN or 3D require high computing power. Second I cannot base the research on massive datasets. Which I have noticed can get really large.
August 1-31	Project – Iteration 1.0 (Bare bones)	I am considering a computation pipeline. Possible end to end pipeline. Use tools like - Scanpy / Seurat v5 possibly for preprocessing, clustering, and UMAP… things like that. IMPORTANT - Logic Gate and Transient mRNA LPN is coming up in literature a lot… I need to investigate that further. Use GNN – for safety and risk prediction – since CRS is a huge issue – Note: need to find literature on that.

		Insilico – is the way to go. That's simply the best approach that will work for now. Spatial Transcriptomics will play a huge role and immunotherapy in general Validating End to End pipeline - https://pmc.ncbi.nlm.nih.gov/articles/PMC5429012/ Insilico - https://www.zeclinics.com/blog/differences-between-in-vitro-in-vivo-and-in-silico-assays-in-preclinical-research/ Language I will use is python – Its simply the best for ML and phase based approach again just standard practice for pipelines – It just makes it easier in general for transparency, building the code, testing, reproducibility , debugging so forth and so on. Use ScRNA dataset -may be ATAC-seq (e.g., GSE155882 mouse TAC, GSE185265 recovery, GSE168742/GSE270788 human post-MI) or something like that but I am not sure yet.
	Project Iteration 1.1	**Model current Iteration: thoughts, ideas and main points** - CART-T Therapy → safety first → because of CRS - Transient mRNA delivery (lots of research supporting this) - How to predict scar reversal ? without side effects - Use Scanpy or Seurat → process single-cell RNA-seq + spatial data - Use MOFA (Multi-Omics Factor Analysis) to integrate multi-omics data - Multi-omics target discovery → finds best FAP + second antigen pairs - Spatial-aware ranking → checks targets are only in scarred heart regions - Transient optimization → tests different "on" times and doses - Benchmarking → compares predictions to real mouse experiments (FAP-CAR, CD248-CAR results) **Key Biological Concepts** Cardiac fibrosis = excess scar tissue in heart → causes heart failure Activated fibroblasts (myofibroblasts) make the scar FAP = is the main marker on bad fibroblasts (but not perfect alone) Seems like second antigen is necessary – atleast in cardiac fibrosis case Transient CAR-T via mRNA-LNP (lasts only ~2–5 days) → avoids long-term damage to healing Logic gating (AND) → extra safety so healthy cells are not killed Should I consider Alpha Fold ? – can be computationally very demanding Note: Need to study Logic Gating further. Same wit CAR-T kinetics **Possible Pipeline flow** 1. Download & analyze public heart datasets → find best target pairs 2. Design transient AND-gated CAR constructs 3. Simulate CAR-T behavior + fibroblast death + scar shrinking 4. Predict safety (CRS risk) + efficacy (how much fibrosis reverses) 5. Compare to published mouse results → show if realistic Scientific Literature:

		- CAR T cells produced in vivo to treat cardiac injury: Rurik, Epstein et al - https://www.science.org/doi/10.1126/science.abm0594 - Logic-gating for extra safety (FAP AND second antigen) - https://www.nature.com/articles/s41586-023-05778-2 - Computational modeling of cardiac fibroblasts – - https://pmc.ncbi.nlm.nih.gov/articles/PMC4846515/
	Logic Gate	**Understanding Logic Gating:** Did a bunch of reading , google search and othe resources related to logic gating Here is a summar of my findings: • **Boolean AND-Gate Logic:** The T-cell only activates if *both* Target A and Target B are present on the same cell—it prevents the T-cell from killing healthy tissue that might only express one of the antigens. • **Combinatorial Antigen Recognition:** By requiring two keys to unlock the cell's "kill" command, the mathematical probability of off-target toxicity drops exponentially. • **Inhibitory CARs (iCARs) / NOT-Gates:** These provide a "dominant-negative" signal. If a T-cell hits a healthy cell with a " Shield antigen, the iCAR sends an off-switch signal that overrides any activation. • **Split-CAR Architectures:** This is an engineering trick where the signaling domains are split between two different receptors. Neither can work alone; they have to physically come together on the target cell to form a functional synapse. • **SynNotch Receptors:** A "Multi-Stage" gate where the first antigen doesn't trigger a kill, but instead triggers the *expression* of a second CAR. It's a sequential logic gate that adds an extra layer of verification. • **The 15nm Synaptic Gap:** Logic gates aren't just code; they are physical. If the antigens are too tall or too short, the "AND-gate" can't physically form the immunological synapse, causing the circuit to fail regardless of the genetic logic. • **Tonic Signaling & Leakage:** Even "OFF" gates can have leakage—low-level basal firing that eventually exhausts the T-cell. Engineering a "tight" gate with a high signal-to-noise ratio is a major optimization challenge. • **Affinity Tuning:** You can't just pick any two antigens. You have to tune the binding affinity so that the T-cell *needs* the cumulative signal of both to cross the activation threshold. • **Logic Gate "Veto" Thresholds:** In my pipeline, I'm setting a strict 1% expression limit in the brain as a "Veto." If a pair shows up even slightly in vital tissue, the logic gate is considered unsafe and the pair is discarded. • **mRNA Decay as a Temporal Gate:** Using mRNA instead of viral DNA acts as a "time gate." The logic only stays active for a few days before the antigen naturally deletes itself, preventing long-term systemic risks. **Research & Engineering Links:** 1. **Design of Chimeric Antigen Receptors with Programmable Logic (2023)** – *Nature Biotechnology*: A deep dive into how to build "smart" T-cells. 2. **Universal and Programmable CAR T Cells (2024)** – *PubMed Central*: Explores the modular "SUPRA" CAR system for complex logic. 3. **Synthetic Notch Receptors as Custom Logic Gates (2025 Review)** – *Cell*: The latest on multi-stage cellular decision-making. 4. **The Physics of the Immunological Synapse (2025)** – *Frontiers in Immunology*: Essential for understanding the physical constraints of logic gating.

	Project Iteration 2.0	**I feel the computational pipeline is taking shape much better now** **Main Concepts** • **Cardiac fibrosis → driven by heterogeneous profibrotic fibroblasts** • **Logic-gated CAR-T → AND logic (two antigens required) for higher specificity** • Safety Margin Score → custom metric balancing efficacy (fibroblast coverage) vs. safety (sparing healthy/vital tissues) • Multi-omics atlas → single-cell + spatial + ATAC-seq to map fibroblast subpopulations • Transient CAR-T → mRNA/LNP delivery for short-lived effect (added safety layer) **Core Tools & Libraries** - Scanpy (single-cell analysis, clustering, preprocessing) - MOFA+ (multi-omics integration) - SciPy (entropy calculations) - Mesa (minimal agent-based modeling) - Human Protein Atlas / GTEx / UniProt (surface filtering & normal-tissue benchmarking) - Biopython / Ensembl BioMart (orthology mapping) **Pipeline Flow (3 Phases)** 1. **Phase 1 – Data Atlas Construction (primary focus)** o Curate public scRNA-seq, scATAC-seq, spatial datasets o Preprocess + batch correction + clustering (Scanpy) o Multi-omics integration (MOFA+) o Orthology mapping (mouse → human) o Filter surface antigens (UniProt/HPA) o Benchmark vs. healthy tissues (GTEx/Tabula) **Phase 2 – Logic Search & Pair Ranking (primary focus)** • Enumerate pairs from top ~50 surface antigens • Compute refined Safety Margin Score (efficacy × specificity − penalty for healthy/vital organs) • Incorporate coverage & sparing metrics (inspired by LogiCAR/Kwon et al.) • Rank with random forest + bootstrap uncertainties **Phase 3 – Minimal Stress Test & Benchmarking (minimal / illustrative only)** • Simple agent-based model (Mesa) with basic killing rules + transience decay • Simulate pathogenic fibroblast depletion vs. healthy cell sparing • Benchmark against known toxic single-antigen controls • Monte Carlo + sensitivity analysis for uncertainty **Defining the Scope of the Project:** After getting feedback, I decided to define the scope of the project to protize phase1 and 2 which is mainly antigen discovery and ranking , Phase3 - validation with stress tesing and benchmarking thorough MESA agent. Phase 3 gets minimum focus phase 1 and 2 is the key for safety and success of the project
	Project Iteration 2.1	**Exploring and including the concept of Surface Proteins in the pipeline** : Surface proteins are basically proteins embedded in or attached to a cell's plasma membrane, making them accessible from the extracellular space.

		They function as receptors, transporters, adhesion molecules, and immune recognition markers (antigens). Because of their location, they are prime targets for immunotherapies like CAR-T cells, antibodies. they are overexpressed on diseased cells (e.g., cancer or activated fibroblasts in fibrosis) while ideally absent or low on healthy tissues**.** **Bausch-Fluck et al. (2023) — "Cell surfaceome: The frontier of target discovery for immunotherapy" (Nature Reviews Immunology)** **https://www.pnas.org/doi/10.1073/pnas.1808790115** **RCTD (Robust Cell Type Decomposition)** - A computational method (R package) that analyzes spatial transcriptomics data (e.g., Visium, Slide-seq). Used to map where specific antigens (surface markers) are expressed spatially in cardiac tissue — e.g., are they enriched in fibrotic scar regions vs. healthy myocardium? Helps confirm tissue-specific localization. **UniProt / Human Protein Atlas (HPA)** Two major public databases used together to validate whether a gene/protein is actually a surface protein. **Changes to PHASE -2** Find the best 2-antigen pairs (AND-gate logic) → Take top ~50 antigens from Phase 1. → Make all possible pairs (~2,500). → Score each pair using "Safety Margin Score": • High score if both antigens are co-expressed a lot in bad fibroblasts • Low score if either is expressed in healthy heart or dangerous organs • Bonus for late-stage specificity and low dropout bias → Rank pairs → pick top 10–20 safest & most specific combinations. → Output: ranked list of best antigen pairs + visualizations. **The issue of - RNA-protein concordance:** RNA and protein levels often show only moderate correlation (typically 0.4–0.6 Pearson in bulk data, lower in single cells), with ~20–50% discordance in complex tissues like cardiac fibroblasts due to post-transcriptional regulation, degradation rates, and technical noise (e.g., scRNA-seq dropouts). This limits RNA as a direct proxy for surface protein abundance in CAR-T target selection. Scientific Literature: "Integrating transcriptomics and proteomics to understand complex diseases" Ronin et al., Nature Reviews Immunology (2023) **Changes made:** Added a qualitative negative control (single-antigen FAP scenario) in Phase 3, calibrated to published preclinical outcomes (e.g., Rurik et al., 2022). **The rest of the piplelne mostly remain the same except this change in** **Logic Search & Ranking →** Enumerate pairs; compute refined Safety Margin Score (entropy, coverage, sigmoidal non-linear threshold, temporal/spatial penalties); rank top 10–20 AND-gate pairs.

	Project 2.2 Iteration	I learnt more about fibroblast today **Fibroblast Characteristics:** Cardiac fibroblasts are the most abundant non-myocyte cell type in the heart. Upon injury (e.g., MI, hypertension), they rapidly activate into myofibroblasts. The distinction is critical for CAR-T design: targeting late-stage-enriched antigens (e.g., via combinatorial AND-gates) aims to clear chronic profibrotic cells while sparing early reparative ones to avoid interfering with initial healing or causing rupture risk. **Scientific Literature:** Integration mapping of cardiac fibroblast single-cell transcriptomes elucidates cellular principles of fibrosis in diverse pathologies" https://www.science.org/doi/10.1126/sciadv.adk8501 **Pipeline Flow** **Phase 1: Data Atlas Construction** • **This srage b**uilds integrated multi-omics map of cardiac fibroblasts (fibrotic vs healthy); identifies surface antigens on profibrotic and subpopulations. • **Main tools & methods**: GEO datasets curation, Scanpy (normalization, Harmony batch correction, Leiden clustering), , MOFA+ integration, RCTD spatial deconvolution, Biopython/BioMart orthology, UniProt/HPA surface filter, CITE-seq/HPA protein validation, GTEx normal-tissue benchmarking, temporal/etiology stratification. • **Key concepts**: Fibroblast heterogeneity mapping, surface antigen filtering, RNA-protein concordance mitigation, human > mouse data priority, go/no-go thresholds (multi-omics ≥0.6, surface ≥0.7). • **Phase 2: Logic Search and Antigen Pair Ranking** • This stage generates and ranks AND-gate antigen pairs for optimal specificity and safety. • Main tools & methods: itertools pair enumeration, sklearn MinMaxScaler & Random Forest ranking, SciPy entropy & sigmoidal threshold, custom Safety Margin Score (coverage + entropy – healthy penalties – microenvironment proxies), **Phase 3: Stress Test and Benchmarking** • **What it does**: Performs minimal, hypothesis-generating simulation to compare relative safety/efficacy of top pairs vs single targets. • **Main tools & methods**: Mesa agent-based model (ABM), exponential transience decay, Monte Carlo uncertainty (n=1000), negative control (FAP-alone), simple killing rules. Main concepts reviewed: **Co-expression scoring:** Measures how strongly two antigens are simultaneously expressed on the same profibrotic fibroblast cells (target coverage >0.6 norm). **Non-linear gating:** Uses sigmoidal/logistic threshold to model AND-gate biology—requires high expression of both antigens for strong activation, penalizing partial/low co-expression (mimics real CAR-T leak risks).

		Microenvironment proxy penalties (HIF1A/TGFB1): Subtracts score points when antigens co-express with hypoxia (HIF1A) or TGF-β pathway markers in suppressive fibrotic niches, reducing ranking of less specific candidates. **Coverage vs safety trade-off:** Balances high profibrotic fibroblast hit rate (>90% desired) against low healthy/off-target expression (<10% ideal), visualized on Pareto frontier plots to select optimal pairs. Per-donor variability check: Calculates IQR of the Safety Margin Score across different patient/donor datasets to ensure pair performance is consistent (low IQR <0.2 preferred). **Key Scientific Literature Reviewed today:** Patrick et al. (2024) – Single-cell transcriptomics of cardiac fibroblast heterogeneity and fibrosis principles (supports Phase 1 atlas building). https://www.science.org/doi/10.1126/sciadv.adk8501 **Rurik et al. (2022)** – Preclinical CAR-T for cardiac fibrosis (transient anti-FAP approach, key biological rationale). https://www.science.org/doi/10.1126/science.abm0594 **Madan et al. (2025)** – LogiCAR designer: Single-cell-guided computational pipeline for logic-gated antigen combinations (AND/OR/NOT) in tumors, directly parallels Phases 1-2 scoring/ranking. https://www.biorxiv.org/content/10.1101/2025.03.19.644074v1 (preprint; published version in PMC: https://pmc.ncbi.nlm.nih.gov/articles/PMC12191104/) **Testa et al. (2025)** – SCAN-ACT: Computational pipeline using single-cell transcriptomics for adoptive T-cell target discovery and logic-gated pairs. https://pubmed.ncbi.nlm.nih.gov/40814001/ **Kwon et al. (related context, 2024/2025 pipelines)** – Computational frameworks for CAR-T antigen discovery and safety scoring (often cited in logic-gate oncology-to-non-oncology adaptations). Search for exact match in Nature Biotechnology series; representative: https://www.nature.com/articles/s41587-024-02345-6 (analogous antigen prioritization). **Patrick et al. (2024)** – Single-cell transcriptomics of cardiac fibroblast heterogeneity and fibrosis principles (supports Phase 1 atlas building). https://www.science.org/doi/10.1126/sciadv.adk8501
	Project 2.3 Iteration	**This is the updated version of the pipline** **Phase 1 – Data Atlas Construction Purpose**: Build comprehensive map of surface antigens on profibrotic vs healthy cardiac fibroblasts using public multi-omics data. **Main concepts & steps** • Dataset curation & hierarchy (human > mouse) • Preprocessing (normalization, batch correction, zero-inflation mitigation) • Clustering & integration (latent factors) • Spatial deconvolution • Surface antigen filtering (membrane localization) • Orthogonal protein validation (precision/recall) • Temporal stratification (early vs late activation) • Donor heterogeneity checks (etiology-specific subsets, IQR) • Normal-tissue safety benchmarking

		Main tools/methods Scanpy, Harmony, MOFA+, RCTD, MAGIC, Biopython/BioMart, GTEx/Tabula/HPA APIs, Leiden clustering, Wilcoxon tests **Phase 2 – Logic Search & Pair Ranking Purpose**: Find and rank best AND-gate antigen pairs for high specificity + safety. **Main concepts & steps** • Pair enumeration (~2,500 combinations) • Expression normalization & entropy calculation • Non-linear sigmoidal gating model • Microenvironmental proxy penalties (hypoxia, TGF-β) • Safety Margin Score (target coverage – healthy/off-target penalties) • Coverage vs safety Pareto frontier • Random forest ranking + feature importance • Donor-level variability reporting (IQR) • Orthogonal dataset sanity-check (hold-out re-scoring) **Main tools/methods** sklearn (MinMaxScaler, RandomForest, permutation importance), SciPy (Shannon entropy, logistic), itertools, bootstrapping **Phase 3 – Stress Test & Benchmarking Purpose**: Perform minimal in silico simulation to compare relative performance of top pairs (hypothesis-generating only). **Main concepts & steps** • Agent-based modeling of CAR-T – fibroblast interactions • Transient CAR-T decay kinetics • Pathogenic fibroblast depletion vs healthy cell sparing • Negative control (FAP single-target) • Monte Carlo uncertainty analysis • Relative benchmarking against single-antigen baselines **Main tools/methods** Mesa (ABM framework), SciPy (Monte Carlo, exponential decay), basic rule-based killing proportional to expression These were the features that were newly adopted into this iteration: **What was newly adopted / emphasized in P5A** • Clearer go/no-go thresholds listed per phase • Stronger focus on microenvironment proxy penalties (HIF1A/TGFB1) in Phase 2 • Explicit orthogonal dataset re-scoring (hold-out sanity-check) in Phase 2 • Pareto frontier / coverage-safety trade-off plots highlighted • Permutation feature importance from random forest added • Per-donor IQR variability reporting strengthened • Protein validation raised to explicit precision/recall computation + threshold raise if recall <0.8
	Gaps in Project 2.3	Read through some literature today and researched limitations and found quite a few gaps in the pipeline. The pipeline needs to reevaluated. Here are the gaps and limitations: • scRNA-seq data (Phase 1): Biased toward abundant cells (monocytes/macrophages), low coverage (5–20%), misses rare fibroblasts, few human fibrosis datasets → incomplete antigen map. • RNA-protein discordance (Phase 1): RNA levels mismatch protein by 20–30% → surface antigen predictions unreliable. • Donor/temporal heterogeneity (Phases 1–2): 15–30% variability across patients and time → pairs may not generalize.

		• Modeling simplicity (Phase 3): Basic ABM ignores TME suppression, poor CAR-T penetration in dense fibrosis, non-cancer context → limited predictive power. • Antigen discovery (Phase 2): No true fibrosis-specific pairs yet → higher off-tumor risk, antigen escape, heterogeneity problems. Scientific literature reviewed: Single-cell transcriptomics reveals cell-type-specific diversification in human heart failure (Koenig et al., 2022): https://www.nature.com/articles/s44161-022-00028-6 Single-Cell RNA Sequencing in Cardiovascular Development, Disease and Medicine (Paik et al., 2020): https://pubmed.ncbi.nlm.nih.gov/32231331 Mathematical models and computational approaches in CAR-T therapeutics (Putignano et al., 2025): https://www.frontiersin.org/journals/immunology/articles/10.3389/fimmu.2025.1581210/full CAR-T cell therapy in hepatocellular carcinoma: from mechanistic insights to clinical translation (Elahi et al., 2025): https://www.sciencedirect.com/science/article/pii/S0305737225001689 Mathematical modeling insights into improving CAR T cell therapy for solid tumors with bystander effects (Kara et al., 2024): https://www.nature.com/articles/s41540-024-00435-4
	Explore various concepts related to pipeline	I went back to the drawing board, did more research, considered a few concepts but ultimately ended up dropping those ideas **1. Graph Neural Networks (GNNs)** **These are** AI models designed to analyze complex webs of data, like predicting how different proteins, cells, and cytokines interact in a massive network. But the issue with GNNs is that it requires massive amounts of highly specific, labeled training data to work. Since large-scale, perfect datasets for cardiac fibrosis cytokine release syndrome (CRS) don't exist publicly, the model would likely just "hallucinate" or over-fit, giving me unreliable risk predictions. **AlphaFold & 3D Protein Design** This is advanced deep learning tools that predict the exact 3D physical structure of a protein from its genetic sequence. Unless I find a better way to integrate I am not considering it **now is** It is computationally heavy and essentially overkill for this phase. I have limited computational capacity. Predicting the 3D shape of the CAR-T receptor doesn't matter if I haven't securely identified the right biological "lock" (the target antigen) on the fibroblast first. **Mouse-Based Target Discovery** Using mouse single-cell RNA sequencing (scRNA-seq) data as the primary starting point to find targets, which is the standard approach in early preclinical biology. **Why I dropped it is** Mice and humans have about a 20% divergence in their genetics. Furthermore, post-translational modifications (like glycosylation) can hide an antigen in humans even if it works in a mouse. I pivoted to an exclusive human-data-first approach to avoid this translation trap. **Complex Whole-Body Simulations**

		This is basically building digital "twins" to simulate how the CAR-T therapy travels through the bloodstream, biodistributes across all organs, and triggers systemic immune responses. I dropped this one because true multi-organ pharmacokinetics requires far too many assumptions for an *in silico* model built on limited public data. I scaled this back to a minimal, agent-based model focused strictly on the local cardiac fibrosis environment so my results stay scientifically grounded. **NOT-Gates & CRISPR Engineering** **This is d**esigning synthetic biology circuits where a T-cell is programmed to shut off (NOT-gate) if it sees a healthy cell marker, often requiring complex CRISPR-Cas9 genome editing. This one also goes in favour of simplicity - I need to keep the computational search space constrained and transparent. Focusing purely on dual-target "AND-gate" logic (requiring two specific markers to activate) is enough to drastically improve safety without overcomplicating the design. **Literature:** *Frontiers in Bioengineering and Biotechnology* (2022) discusses the limitations of data sparsity in training network-based models for complex diseases. (https://doi.org/10.3389/fbioe.2022.846906) Jumper, J., et al. (2021). Highly accurate protein structure prediction with AlphaFold. *Nature*, 596(7873), 583-589. (https://doi.org/10.1038/s41586-021-03819-2) *Nature Reviews Genetics* (2022) on cross-species divergence, and Trost et al. (2016) regarding epitope masking and glycosylation discrepancies. (https://doi.org/10.1038/s41576-022-00466-9) *Cell Systems* (2024) reviews the boundaries and computational limits of multi-scale dynamical modeling in translational medicine. Kwon et al. (2023). LogiCAR: A framework for discovering logic-gated CAR T targets. *bioRxiv*. (https://doi.org/10.1101/2023.05.12.540581)
Sept 1 – 30	Datasets Evaluated	**Reviewed Datatsets:** I spent the first week of Sept, organizing my data strategy for Phase 1. I listed out every major dataset I considered, weighed the pros and cons, and cut the ones that didn't meet my strict "Pan-Ventricular" and high-fidelity criteria. Here is the breakdown of my thought process: **The Safety Reference (Finding the Baseline). M**y first task was figuring out how to define Healthy"so the CAR-T doesn't kill normal cells. **Dataset Considered: GTEx V8 (Bulk RNA-seq) -I**t's huge, covers all organs, and is super lightweight. I could process it in seconds because it just gives an average"expression score for a whole tissue. *T*he flaw with this tough is signal dilution -> dropped , too risky. I need single-cell resolution. **Dataset Considered: Tucker et al. (2020) -It's a** standard, single-cell healthy heart atlas. The flaw tough is It's getting outdated. It only has about 280k nuclei. I dropped this one too **Dataset Considered: Litviňuková et al. (2020) (Heart Cell Atlas) -** This is the current gold standard. It has over 500,000 cells with way better resolution on rare immune and endothelial subsets. I will be keeping this one - This is my official Safety Veto dataset.

		2. The Discovery Set - Next, I needed to see what happens the exact moment a heart attack causes fibroblasts to turn bad. **Dataset Considered: GSE234714 (hEHT Model) -** It uses Engineered Heart Tissue in a lab. You get perfect time-control (Day 1, Day 2, etc.) without messy patient differences. Fibroblasts grown on plastic undergo shock and express markers that don't actually exist in real human hearts (Tallquist, 2017). **DROPPED since it has a h**igh risk of false positives. I must stick to in vivo human data. **Dataset Considered: Kuppe et al. (2022) -** The best spatial/temporal map of a real human heart attack. It lets me trace the exact trajectory of a resting fibroblast turning into a pathogenic myofibroblast.I am **KEEPING.** This is my core Discovery engine. **The Validation Sets - which Proves it works long-term -** Finding a target in an acute heart attack is great, but the therapy needs to work for chronic heart failure too. **Dataset Considered: He et al. (2025) (Meta-Atlas) T**his has a massive statistical power. It aggregates millions of cells from dozens of studies. The issue with this tough is sample duplication. Meta-atlases often reuse patients from older studies. If I train on this, I might double-count patients, which messes up my confidence intervals. Plus, merging that much data causes massive batch effects that smooth over the rare antigens I'm looking for. **DROPPED.** Quality over quantity. **Dataset Considered -Rao et al. (2021) -** Covers both Ischemic (heart attack scars) and Non-Ischemic (genetic) failure.The *Flaw:* Redundancy. If I integrate too many datasets, the batch correction algorithms wash out the subtle biological signals. I decided it's better to use two highly specific, separate datasets rather than one noisy, blended one – DROPPED **Datasets Considered: Wang et al. (2020) & Koenig et al. (2022) -** *The Appeal:* Wang gives me pure Chronic Ischemic failure. Koenig gives me pure Chronic Genetic failure (DCM). **KEPT.** By requiring a target to show up in Kuppe AND Wang AND Koenig, I mathematically prove the target is universal ("Pan-Ventricular") across all major causes of heart failure.
October 8	Reevaluating the Scoring Engine for Phase-1	Reviewing the literature revealed a fatal flaw - the "Drowning Effect." If I just add raw values, a target that is incredibly stable in the heart but dangerously expressed in the brain could still score high. Its massive stability number would mathematically mask the terrible safety score. In cardiac CAR-T, an off-target attack is lethal; safety cannot be averaged out. Current Focus: Designing a continuous scoring metric Thoughts & Progress: I drafted a theoretical additive score: (Specificity + Stability + Discordance) to move away from binary filtering and eventually rank candidate antigens. **References:** 1. Macaluso, M., et al. (2021). "Targeting Cancer with CAR T-Cells: New Combinations and Prioritization Strategies." *Journal of Clinical Medicine*, 10(4), 766. 2. Choi, B. D., et al. (2019). "CAR-T cells secreting BiTEs circumvent antigen escape without detectable toxicity." *Nature Biotechnology*, 37(9), 1049-1058.

Oct 11 - 12	MCDA – Algorithm	**Current Focus:** Moving to Multi-Criteria Decision Analysis (MCDA) **Thoughts & Progress**: I decided to build a custom weighted formula, assigning weights based on clinical priority: 40% for Specificity (Safety), 30% for Stability, 20% for Equity, and 10% for Protein evidence. I wrote the initial script – for the MCDA Framework to test out the hypothesis with sample datsets. **Challenges & Gaps**: While mapping out the math on paper, I realized a massive scaling issue. The raw biological metrics I plan to use exist on completely different numerical scales. For instance, an entropy score might yield a tiny decimal like 0.05, while stability variance could easily hit numbers like 5.0. If I just multiply 0.4 by 0.05 and 0.3 by 5.0, the larger raw number will completely overpower my intended weights. It would be a mathematical disaster where the variance dictates the score regardless of the 30% cap. Next Iteration Idea: Before I can apply these weights in code, I must normalize every single biological metric to a strict 0 to 1 scale so the percentages actually control the final score. References: 1. Thokala, P., et al. (2016). "Multiple Criteria Decision Analysis for Health Care Decision Making—An Introduction: Report 1 of the ISPOR MCDA Emerging Good Practices Task Force." *Value in Health*, 19(1), 1-13. 2. Zheng, Y., et al. (2020). "Multi-omics data integration using ratio normalization." *Bioinformatics*, 36(12), 3780–3787.
Oct 18 -19	MCDA – Algorithm iteration	**Current Focus:** Normalization and Proteomic Integration **Thoughts & Progress:** Huge breakthrough. I read a paper by Schug et al. (2005) on using Shannon Entropy for tissue specificity. I implemented $1-(H/H_{max})$ so that perfect specificity is exactly 1.0, and "expressed everywhere" is 0.0. I also applied percentile-rank scaling to all my variables. Suddenly, the 40/30/20/10 weights are actually controlling the algorithm the way I intended! **Challenges & Gaps:** While the math works now, I have a biological blind spot. I was reading **Liu et al. (2016) in *Cell*,** and they proved that mRNA levels only predict actual protein levels about 40% of the time. My whole pipeline is built on scRNA-seq (which is just RNA). If I only rely on this, 60% of my targets might be "ghosts" that don't actually exist on the cell surface for the CAR-T to grab. **Next Iteration Idea:** I need to pull in the Human Protein Atlas (HPA) mass spectrometry data. I'll use it as a Bayesian-style reality check. **References:** 1. Schug, J., et al. (2005). "Promoter features related to tissue specificity as measured by Shannon entropy." *Genome Research*, 15(3), 403-410. 2. Liu, Y., Beyer, A., & Aebersold, R. (2016). "On the Dependency of Cellular Protein Levels on mRNA Abundance." *Cell*, 165(3), 535-550.
	MCDA Algorithm - Iteration	Current Focus: Bootstrapping and Official Naming Thoughts & Progress: I finalized the integration of the HPA data today. If a gene has Mass-Spec proof, it gets a 1.0 multiplier; if it's RNA-only, it gets a 0.0 in that category.

		I also spent time formalizing the "Equity" score (R) to penalize Eurocentric datasets, because a therapy is useless if it only works for one demographic. Since this has evolved so far beyond a simple filter, I've officially named the algorithm PRISM (Probabilistic Ranking and Integration for Safe Multi-targeting). Challenges & Gaps: The scRNA-seq data is still inherently noisy due to "dropouts" (false zeros). If I just take the average expression to calculate my stability score, I might be trusting a fluke in the sequencing machine. Next Iteration Idea: Kharchenko et al. (2014) suggests probabilistic modeling for dropouts. I'm going to code a Bootstrap Resampling loop to stress-test the data. References: 1. Kharchenko, P. V., Silberstein, L., & Scadden, D. T. (2014). "Bayesian approach to single-cell differential expression analysis." *Nature Methods*, 11(7), 740-742. 2. Gamazon, E. R., et al. (2018). "The Impact of Ancestry and Geographic Isolation on Gene Expression." *Nature Genetics*, 50(12), 1703-1715. 3. Uhlén, M., et al. (2015). "Proteomics. Tissue-based map of the human proteome." *Science*, 347(6220), 1260419.
Oct 20 – 24	PRISM (MCDA) Algorithm iteration	Current Focus: Uncertainty Propagation in PRISM Thoughts & Progress: The bootstrap is running perfectly. The algorithm now takes the data, injects 20% random dropouts, and recalculates the stability 1,000 times. If the gene survives the simulation, it gets a high stability score. Then I realized something awesome: since I'm running the stability 1,000 times, I can actually calculate the *entire* P.R.I.S.M. score 1,000 times for every antigen. This means I don't just get a single rank; I get a Mean Score $\pm$ Standard Deviation. Improvements: This is the the jump I was looking for. Now, if a target has a PRISM score of 0.82 but an error bar of ± 0.30, I know the data is too noisy to trust. I've added Python assert gates to ensure no value ever breaks the [0,1] bounds during the loops. References: 1. Efron, B. (1979). "Bootstrap Methods: Another Look at the Jackknife." *The Annals of Statistics*, 7(1), 1-26. 2. Hicks, S. C., et al. (2018). "Missing data and technical variability in single-cell RNA-sequencing experiments." *Biostatistics*, 19(4), 562-578.
Oct 25 – 26		Current Focus: Finalizing the Phase 1 Engine Thoughts & Progress: After much deliberatinThe PRISM Algorithm is complete. The Final Formula: $$S_{PRISM} = \left(0.4 \cdot \hat{S}_{spec} + 0.3 \cdot \hat{S}_{stab} + 0.2 \cdot R + 0.1 \cdot S_{prot}\right) \pm \sigma_{agg}$$ With this phase 1 is more or less conceptualized … I may fine tune it later but this should make core of the discovery engine. Looking back at early October, I was basically just using Scanpy as a tool to filter Excel rows. Now, I've engineered a multi-parametric MCDA discovery engine that mathematically de-risks biological noise and RNA-protein discordance. PRISM is ready to feed its top 30 targets into the Phase 2 Logic-Gate pairing script.

		References: 1. Roybal, K. T., et al. (2016). "Precision Tumor Recognition by T Cells With Combinatorial Antigen-Sensing Circuits." *Cell*, 164(4), 770-779. 2. Saltelli, A., et al. (2008). *Global Sensitivity Analysis: The Primer*. John Wiley & Sons.
October 27 - 31	Exploring concepts that can enhance my pipeline	**Now that I have a fairly rigorous discovery engine (MCDA Algorithm) – although fine tune it later. I want to revaluate the pipeline objectively.** **After a lot of reading and research – I am certain I want to stick with these approaches:** - ScRNA single cell data cohort - Logic Gate - MCDA based multiomnic antigen discovery - Transient mRNA inVivio - Human tissues (and not murine related datasets) - Veto against a healthy cell cohort - Agent based simulation for phase 3 I am also exploring other ways I could enhance this pipeline . Research the following concepts Transcription Factors (TFs) - These are master switch proteins inside the cell that bind to DNA to turn specific genes on or off. These are important to prove the targets aren't a fluke. If my PRISM algorithm finds 10 great surface targets, identifying that they all share the same upstream TF proves they are biologically linked by the same disease mechanism (fibroblast activation). AND Gate Logic – Further bushed up on AND Gate Logic. These are essentially biological circuit where a therapeutic cell requires *two* distinct surface targets to trigger a response (Target A + Target B = Attack). These are considered very safe. It ensures the CAR-T cell only destroys diseased heart fibroblasts (which express both targets) and ignores healthy tissue (which might express only one or neither). Glycosylation – Is a biological process where cells attach bulky, complex sugar molecules to the outside of surface proteins. These sugars act like a physical .A target might look perfect in the RNA data, but if it is covered in sugars in real life, the CAR-T cell can't physically dock onto it. It can be used to filter out "hidden" false positives. Pharmacokinetics (PK) – This is essentially what the mathematical study of what the body does to a therapy over time (how it is absorbed, distributed, and eventually destroyed/cleared). Because I am modeling a *transient* therapy (mRNA-LNP), the CAR-T cells don't live forever. PK modeling allows my simulation to accurately reflect the therapy fading away over 48–72 hours, testing if it works fast enough before it degrades. The Hill Function – Is a math equation that creates an S-shaped (sigmoidal) curve to model sudden, non-linear biological responses. It is used where CAR-T cells don't activate in a slow, linear way. They require a specific density threshold of targets to suddenly "switch on" (known as ultrasensitivity). The Hill function mathematically simulates this realistic biological tipping point in my Phase 3 model. These are the concepts I will try to incorporate if it si approrate whin the context of the end to end pipeline

		Research Literature: **Transcription Factors (TFs) & Master Regulators** https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6602523/ **The Cardiac Fibroblast Target Paper: *Spatial multi-omic map of human myocardial infarction* (Kuppe et al., 2022)** https://pubmed.ncbi.nlm.nih.gov/35948637/ **AND Gate Logic (Synthetic Biology)** https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4752902/ **The General Review: *Engineering T Cells to Treat Cancer: The Convergence of Immuno-Oncology and Synthetic Biology* (Roybal & Lim, 2017)** https://pubmed.ncbi.nlm.nih.gov/28226226/ **Glycosylation & Steric Hindrance - The "Sweet CARs" Paper: *Targeting glycans for CAR therapy: The advent of sweet CARs* (He et al., 2022)** https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9664560/ **Influence of Glycosylation Review: *Influence of Target Antigen Glycosylation on CAR T-cell Efficacy* (Review showing why computational PTM filtering is necessary)** https://pubmed.ncbi.nlm.nih.gov/34493645/ **Pharmacokinetics (PK) of mRNA-LNP CAR-T** **The Exact Cardiac Fibrosis mRNA-LNP Paper: *In vivo generation of CAR T cells with targeted mRNA-lipid nanoparticles* (Rurik et al., 2022, Jonathan Epstein Lab)** https://pubmed.ncbi.nlm.nih.gov/35015566/ **The Hill Function & Ultrasensitivity** **The Density/Ultrasensitivity Math Paper: *T cell circuits that sense antigen density with an ultrasensitive threshold* (Hernandez-Lopez et al., 2021) -** https://pubmed.ncbi.nlm.nih.gov/33707234/
October 31	Project Iteration 3.0	**Based on all the research and feed back I have gotten so far, here is my revised iteration of the project** **Phase 1: The PRISM-TF Discovery Engine** **Goal: Find and rank the best individual surface targets.** • Every potential target is fed into the PRISM algorithm. • Targets are scored on Specificity (using Shannon Entropy to ensure tissue restriction), Density, and Robustness. • The Filter: The S_{prot} Kill Switch mathematically punishes and removes any protein that lives inside the cell (like the nucleus) where a CAR-T cell cannot reach it. • A ranked list of the top 30 single targets, verified by upstream Transcription Factors (TFs) to ensure they are driven by real disease mechanisms.
		Phase 2: Combinatorial Logic & Scoring **Goal: Pair the best single targets into ultra-safe antigens** • The top 30 Phase 1 targets are combined into dual-antigen pairs (up to 500 combinations). • Logic Integration: * AND-Gate: The therapy requires both Target A + Target B to activate. ○ NOT-Gate (Shield): A healthy cell marker (e.g., SIRPA) is added. If the CAR-T cell sees this shield, it forcibly aborts the attack. • The pipeline scans the amino acid sequences of the pairs and filters out targets that are likely to be masked by complex sugars in the human body (Glycosylation/Steric hindrance). **Phase 3: Mechanistic Simulation (The Stress Test)** **Goal: Simulate how these engineered cells will behave dynamically in a 3D environment over time.**

		• The champion pairs are imported into an Agent-Based Model (ABM) using the Python framework Mesa. • The Hill Function: Simulates "ultrasensitivity"—ensuring the virtual CAR-T cells only activate when they encounter a high density of targets, ignoring low-level biological noise. • Pharmacokinetics (PK): Applies mathematical decay to simulate the mRNA-LNP therapy fading away over 48–72 hours.
	Revising the Computation Pipeline. Discarding some old features and adopting new ones.	I have been researching and reading a lot over the past few days. I am reconsidering some of the freatures of the pipeline and want to discard them for better ones. **3D Spatiotemporal Environment (Phase 3)** Rendering the Agent-Based Model (ABM) in a full 3D coordinate system. **Rationale:** Calculating spherical volumes and z-axis collision physics over a 48-hour simulation window was paralyzing the CPU. I pivoted to a continuous 2D Mesa environment ($400 \times 400 \mu m$) representing a cross-section of the ventricular wall. The assumption is that z-axis interactions scale linearly with the x-y plane, allowing me to track clearance rates efficiently without losing behavioral accuracy. **Explicit Molecular Dynamics (MD) (Phase 2)** Simulating the full physical forces of the immunological synapse via explicit MD. **Rationale:** MD requires cluster-level HPC resources and is prohibitively slow for screening over 50 candidates. Instead, I adopted AlphaFold-predicted Extracellular Domain (ECD) heights and pLDDT scores as a high-throughput proxy for steric hindrance and structural stability. **Partial Differential Equations (PDEs) for Diffusion (Phase 3)** Explicitly modeling chemokine fluid dynamics using PDEs. **Rationale:** Continuous mathematical solving for diffusion gradients broke the optimized, step-wise logic of the ABM. I replaced this with a vectorized "Dynamic Homing Gradient" (H(x,y,t)), a mathematical proxy that biases CAR-T motility toward clusters without needing fluid dynamics calculations. **Randomized Monte Carlo Escape (Phase 2)** random antigen downregulation models **Rationale:** Assuming tumor or fibrotic escape happens purely randomly lacks biological truth. I upgraded this to a "Bio-Evolutionary Resilience Scan" that explicitly checks UniProt/Gene Ontology (GO) terms for actual pathway redundancy, penalizing target pairs that share functional overlap and are thus susceptible to correlated escape.
		Next Concepts Adopted & Rationale **Spatial Transcriptomics Integration** • Integrating Spatial Transcriptomics (ST) datasets (e.g., Visium or Xenium) to complement the existing single-nucleus RNA-seq atlases. • **Rationale:** Currently, the safety vetoes assume co-expression dictates toxicity (cell-autonomous, direct contact). Incorporating spatial coordinates will allow me to map the "Bystander Effect"—modeling the precise risk of cytokine-mediated toxicity when a diseased fibrotic cell physically neighbors a healthy cardiomyocyte. **2. Systemic CRS Toxicity Modeling** • Expanding the Phase 3 ABM to track total simultaneous synaptic activations over time. • **Rationale:** Cytokine Release Syndrome (CRS) is the deadliest clinical bottleneck for CAR-T therapies. By tracking total activation states

		dynamically, I can output a systemic organ-level cytokine load prediction, rather than just a local cell-clearance metric. **3. Long-Term Darwinian Evolutionary Resistance** • A time-dependent, generational ABM module. • **Rationale:** Instead of using instantaneous density reductions, the next upgrade to Phase 3 will simulate the gradual evolutionary loss of antigens over weeks of continuous CAR-T pressure. This will rigorously stress-test whether the matrifibrocyte population ultimately bounces back or is permanently suppressed. **4. Generative AI Sequence Design (Wet-Lab Translation)** • Implementing deep learning structural models like ProteinMPNN to generate the physical binders. • **Rationale:** The pipeline currently establishes target *discovery*. To translate this *in vitro*, I will use ProteinMPNN to generate the specific Single-chain variable fragments (scFvs) for the champion targets (e.g., GPC6 and CD9). This effectively moves the project from logic-gated theory to tangible biochemical sequences that can be synthesized and tested. Scientific Literature: **Museum of spatial transcriptomics** https://www.nature.com/articles/s41592-022-01409-2 **Mathematical modeling unveils the timeline of CAR-T cell therapy and macrophage-mediated cytokine release syndrome** https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1012908 **Tumor morphology and phenotypic evolution driven by selective pressure from the microenvironment** https://www.cell.com/cell/fulltext/S0092-8674(06)01403-1 **Robust deep learning-based protein sequence design using ProteinMPNN** https://www.science.org/doi/10.1126/science.add2187 **The immunological synapse** https://aacrjournals.org/cancerimmunolres/article/2/11/1023/465715/The-Immunological-Synapse **Strategies to address chimeric antigen receptor tonic signaling** https://aacrjournals.org/mct/article/17/9/1795/115147/Strategies-to-Address-Chimeric-Antigen-Receptor **Intrinsically disordered proteins and their environment** https://link.springer.com/article/10.1007/s10930-013-9485-8 **PD-1–and CTLA-4–based inhibitory chimeric antigen receptors (iCARs) divert off-target immunotherapy responses** https://www.science.org/doi/10.1126/scitranslmed.3006597 **Embracing the dropouts in single-cell RNA-seq analysis** https://www.nature.com/articles/s41467-020-14976-9 **Evolutionary trade-offs, Pareto optimality, and the geometry of phenotype space** https://www.science.org/doi/10.1126/science.1217405
November 1	Research some concepts related to the updated Pipeline	Did some more research on these concepts: **The Immunological Synapse** The microscopic, physical interface between a T-cell and its target cell where receptor-ligand binding and signal transduction occur. In CAR-T engineering, the structural dimensions (such as spacer length and antigen height) must perfectly align to allow proper synapse formation without steric hindrance. **Tonic Signaling & Ultrasensitivity** Tonic signaling is the constant, low-level background activation of a receptor in the absence of its target, which can exhaust CAR-T cells prematurely. Ultrasensitivity mitigates this by functioning as a biological

		threshold, ensuring the cell only triggers a massive response when a high density of antigen is present, effectively ignoring the background noise. **Intrinsic Protein Disorder** This refers to segments of a protein that lack a fixed, rigid 3D structure, making them highly flexible or "floppy." In target discovery, highly disordered extracellular domains are heavily penalized because their physical instability prevents a CAR from establishing a firm, sustained binding event. **Dominant Negative Receptor Logic** A synthetic biology mechanism where an inhibitory receptor (like your "Heart Shield" NOT-gate) actively recruits intracellular phosphatases to dismantle an activation signal. Rather than passively competing for physical space, it acts as a biological multiplier that rapidly crushes the intracellular kill signal to zero, forcing the T-cell to abort the atta **Monte Carlo Bootstrapping (Zero-Injection)** A statistical stress-test that randomly resamples single-cell data thousands of times while artificially injecting "zero" values (dropouts) to simulate technical sequencing noise. This ensures that a gene's expression score is mathematically robust across a population, rather than being a false positive driven by a few highly expressing outlier cells. **Pareto Composite Ranking** A multi-criteria optimization method used to find the best compromise between competing variables, such as safety, efficacy, and evolutionary resilience. It identifies the "Pareto frontier" where no single metric can be improved without sacrificing another, ensuring your final CAR-T targets possess the most balanced and viable clinical profile. **Vectorized Homing Gradients** A computational abstraction used in agent-based modeling to simulate chemotaxis without calculating computationally heavy fluid dynamics. It assigns directional mathematical vectors across a 2D grid, continuously biasing the physical movement of virtual CAR-T cells toward the highest concentration of fibrotic clusters. Dustin, M. L. (2014). The immunological synapse. *Cancer Immunology Research*, 2(11), 1023-1033. Link Ajina, A., & Maher, J. (2018). Strategies to address chimeric antigen receptor tonic signaling. *Molecular Cancer Therapeutics*, 17(9), 1795-1815. Link Macal, C. M., & North, M. J. (2010). Tutorial on agent-based modelling and simulation. *Journal of Simulation*, 4(3), 151-162. Link Shoval, O., et al. (2012). Evolutionary trade-offs, Pareto optimality, and the geometry of phenotype space. *Science*, 336(6085), 1157-1160. Link Qiu, P. (2020). Embracing the dropouts in single-cell RNA-seq analysis. *Nature Communications*, 11(1), 1-9. Link ck. Fedorov, V. D., et al. (2013). PD-1–and CTLA-4–based inhibitory chimeric antigen receptors (iCARs) divert off-target immunotherapy responses. *Science Translational Medicine*, 5(215). Link Uversky, V. N. (2013). Intrinsically disordered proteins and their environment. *The Protein Journal*, 32(3), 156-175. Link
Nov 2	Final Model – **SCARGATE: A Novel Multi-Omic MCDA Algorithm and Logic-Gated Pipeline for**	**So here's my draft of the model – I don't expect to change much unless when I run the data and the hypothesis or assumptions are not working. However I might make changes regarding the tresholds and variables of the pipeline itself – other than that this framework should be solid.** **Project Model: Engineering the Heart Shield**

	the In Silico Discovery of Safe, Transient CAR-T Targets in Cardiac Fibrosis Reversal	**Objective:** To design an *in silico* framework that discovers and engineers logic-gated (AND/NOT) CAR-T cells capable of targeting cardiac fibrosis without damaging healthy tissue. The pipeline integrates three landmark single-nucleus RNA-sequencing (snRNA-seq) atlases totaling nearly 2.5 million nuclei: Litviňuková (Healthy reference), Kuppe (Ischemic transition), and Koenig (Chronic failure). The data is harmonized and strictly filtered to isolate pathological fibroblasts expressing core disease markers (POSTN and CTHRC1).
		Phase 1: Target Discovery (The FSI v2.1 Engine) **Goal:** Identify the safest and most prominent single surface antigens on fibrotic cells. The pipeline processes targets through the custom **Fibro-Safe Index (FSI v2.1)**, an algorithm that scores proteins based on four mathematical pillars: • **Specificity:** Evaluates tissue restriction using Shannon Entropy to prevent systemic off-target effects. • **Density:** Ensures expression levels are high enough to trigger CAR-T activation. • **Robustness:** De-risks single-cell sequencing noise using 1,000 bootstrap simulations. • **Surface-Safe (S_{prot}):** Applies a mathematical "Kill Switch" (a negative penalty) to intracellular proteins, guaranteeing the algorithm only prioritizes targets physically accessible on the cell membrane. Before passing candidates forward, the pipeline applies a strict **Safety Veto**, instantly discarding any target expressed in vital organs (Brain, Lung, Liver, healthy Heart) to generate a pool of highly specific dual-antigen pairs.
		Phase 2: Mechanistic Engineering (The Synaptic Filter) **Goal:** Stress-test the genetic targets against the physical and biological constraints of the immunological synapse. • **Tunable Logic Engine:** Utilizes a Hill-function probability model to set activation thresholds, mathematically penalizing pairs that exhibit dangerous baseline "leakiness" (tonic signaling). • **Structural Gating:** Imports AlphaFold 3D predictions to assess Extracellular Domain (ECD) heights. It applies penalties if a target is too short or masked by complex sugars (glycosylation), which would physically block the CAR-T cell from binding. • **Bio-Evolutionary Resilience:** Runs a Monte Carlo escape simulation and queries biological pathways (UniProt/GO terms) to prioritize antigen pairs that are mechanistically independent, making it harder for the disease to mutate and evade the therapy. • **Single-Cell Veto:** Brute-force scans ~480,000 individual healthy cells, vetoing any pair that co-expresses on critical subtypes like pacemaker cells.
		Phase 3: Spatiotemporal Simulation (The Agent-Based Model) **Goal:** Simulate the dynamic, time-dependent behavior of the engineered CAR-T cells in a spatial environment. The final pairs are imported into a continuous 2D Agent-Based Model (ABM) built in Python's Mesa framework, representing a cross-section of fibrotic heart tissue over a 48-hour window. • **Motility:** Virtual CAR-T agents navigate the tissue guided by a vectorized "Dynamic Homing Gradient" that biases movement toward fibrotic clusters.

		• **Synaptic State Machine:** Upon physical contact with a cell, the CAR-T agent enters a "Synapse" state, utilizing a dominant-negative accumulator to calculate the activation signal over 5 virtual minutes. • **The NOT-Gate (Heart Shield):** If the CAR-T cell detects the designated healthy tissue marker (the Shield), the accumulator multiplies the activation signal by a suppressive fraction. This mathematically crushes the kill signal to zero, forcing the agent to abort the attack and detach.
Nov 3 - 7	Code Architecture	1. |
Nov 8 -9	Computational Pipeline – Coding	**Data Ingestion & Standardization (Steps 1–2)** • Action: Ingested the "Core-4" data pillars. Mapped Kuppe, Wang, and Koenig datasets to a unified schema (cell_type_unified and condition_unified). • Design Choice: Implemented "fail-fast validation." Applied the Acid Test to require POSTN and CTHRC1 expression in ≥1% of fibroblasts. If a dataset failed, it was immediately halted. Handled massive h5ad files in backed mode (backed="r") to prevent memory overload. Results: The pipeline successfully ingested and standardized roughly 2.5 million nuclei across the three discovery pillars (Kuppe, Koenig, Wang). Crucially, all three datasets passed the initial Acid Test, confirming that POSTN and CTHRC1 were expressed in at least 1% of the fibroblasts.
Nov 10 – 14	Computational Pipeline – Coding	**Pre-processing & Merging (Steps 3–4)** • **Action:** Merged the three discovery pillars using an inner join on genes to ensure fair cross-study differential expression. Applied Harmony batch correction. • **Design Choice:** Masked the matrix with the Bausch-Fluck/CSPA surfaceome list (~650 genes) *before* running the Monte Carlo zero-injection bootstrap. Masking early drastically reduced compute time and RAM. Used column-wise extraction for sparse matrices to prevent the 33 GB memory allocation crashes experienced during development. After merging the data and applying the Harmony batch correction, the pipeline identified 120 genes heavily upregulated in fibrosis (Log2FC ≥ 1.5, adj p < 0.05). I then applied the Bausch-Fluck/CSPA surfaceome mask to strip away useless intracellular targets, leaving ~620 surface proteins.
Nov 15 – 16		**PRISM Target Scoring** Action: Fed the standardized matrix into the PRISM algorithm. Design Choice: Calculated scores for Specificity, Density, and Robustness. Applied the S_{prot} filter to mathematically penalize and remove any proteins located inside the cell, guaranteeing the algorithm only prioritizes targets physically accessible to a CAR-T cell. *The* The TF quantification steps (5b/7b) are computationally heavy and optional. When those files were missing, the script would fail to generate the Top 30 Golden Pairs.

		T*he Solution:* I engineered a graceful fallback. If the script cannot find the S_TF files, it defaults the TF score to 0 and successfully ranks the top 30 pairs using just the Mean FSI score, preventing a hard crash.
Nov 17 - 21	Computational Pipeline – Coding	**Logic Gating & Safety Veto (Steps 6–7)** **Action:** Set the Logic Gate threshold to >25% expression within the fibrotic niche. Executed the Pillar 4 (Tabula Sapiens/Litviňuková) safety veto. **Design Choice:** Adjusted the niche threshold down from 50% to 25% to balance sensitivity and specificity, retaining biologically relevant but technically sparse genes. For the veto: any gene with >1% expression in the brain or >20% in cardiomyocytes was strictly discarded. I initially set a strict rule that a gene must be expressed in >50% of the entire fibroblast population to pass. This was a massive mistake; it left only 8 genes and completely deleted highly ranked, biologically relevant targets (like PIEZO2 and ITGAX) simply because of technical dropout in the sequencing data. *The Solution:* I lowered the threshold to >25% and redefined the "niche." Instead of looking at all fibroblasts, the pipeline now strictly calculates prevalence within the specific POSTN/CTHRC1+ disease cells. Loading the massive Pillar 4 safety datasets (especially the Tabula Sapiens Brain file) completely maxed out my laptop's RAM. The system would hang, the cursor became unresponsive, and the IDE terminal would frequently crash. *he Solution:* I completely overhauled the memory management. I implemented a 2 GB "backed mode" threshold so the matrix stays on the hard drive instead of RAM. I also forced batched column reads (pulling data in 25k chunks), combined the Heart datasets into a single-pass read, and moved execution to an external terminal.
Nov 22 - 23	Computational Pipeline – Coding	**Combinatorial Pairing (Step 8)** • **Action:** Combined surviving Phase 1 targets into dual-antigen pairs. • **Design Choice:** Applied the specific AND-gate safety logic. If a gene showed ≥20% expression in lung, liver, or healthy heart fibroblasts, it was flagged. However, the pair was considered safe and passed if its partner gene showed 0% expression in that same tissue. Output the top 30 candidate pairs.
Nov 24 – Nov 28	Computational Pipeline – Coding	**Tunable Logic & Hill Function (Step 9)** • **Action:** Transitioned to Phase 2 (Mechanistic Engineering). Simulated the probability of activation (P_{act}) using a Hill-function logic engine. • **Design Choice:** Conducted an affinity (K_d) stress test to check tunability. Penalized target pairs that exhibited high tonic signaling risk to prevent theoretical CAR-T cell exhaustion.
Nov 29 – 30	Computational Pipeline – Coding	**Structural & Synaptic Filtering (Step 10)** • **Action:** Filtered pairs based on the physical constraints of the immunological synapse. • **Design Choice:** Utilized AlphaFold-predicted Extracellular Domain (ECD) heights to algorithmically assign CAR spacers (short vs. long) to prevent physical buckling. Applied the Glyco-Shield penalty to targets masked by glycosylation or intrinsic disorder. **Evolutionary Resilience (Step 11)** • **Action:** Ran a Monte Carlo escape simulation to model tumor/fibrotic escape.

		• **Design Choice:** Cross-referenced UniProt and Gene Ontology (GO) terms to check for pathway redundancy. Penalized pairs that operate in the same biological pathway, prioritizing independent targets to mathematically reduce the risk of correlated evolutionary escape. Results Running the 1,000-iteration Monte Carlo zero-injection bootstrap proved highly effective at rooting out technical noise. When run through the full P.R.I.S.M. scoring matrix (S_{spec}, S_{dens}^{*}, S_{rob}, S_{prot}), targets like PIEZO2, ITGAX, CDH11, and GPC6 emerged at the very top of the functional ranking.
Dec 1 – 5	Computational Pipeline – Coding	**Stochastic Single-Cell Safety Scan (Step 12)** • **Action:** Brute-force verified the pairs against ~480,000 cells in the Litviňuková healthy heart atlas. • **Design Choice:** Any pair showing >0.1% co-expression in vital cells (like pacemakers) was immediately vetoed. Verified the presence of the surviving targets at the protein level using a manual ProteomeXchange fibrosis atlas. Scanning the Litviňuková healthy heart dataset cell-by-cell (~480,000 cells) was incredibly slow and memory-intensive. Furthermore, ambient RNA noise was triggering false-positive vetoes. *The Solution- I* implemented chunked reading to control RAM spikes and added an expression threshold of 0.5. By ignoring trace-level ambient noise, the veto became both faster and highly accurate.
Dec 6 – 7	Computational Pipeline – Coding	**Pareto Composite Ranking (Step 13)** • **Action:** Ranked the final pairs. • **Design Choice:** Used Pareto ranking across Safety, Efficacy, and Resilience metrics. This ensured no single metric was optimized at the expense of another, resulting in the top 3 most balanced pairs: GPC6-CD9, CDH11-CD9, and CDH11-CD44.
Jan 2 – Feb 7	Computational Pipeline – Coding	**Spatiotemporal Simulation (Step 14)** • **Action:** Executed Phase 3. Imported the top pairs into a continuous 2D Mesa Agent-Based Model. • I simulated a 48-hour physical tissue environment containing roughly 400 fibrotic cells, 1,000 healthy cells, and 150 decoy cells. I then executed the virtual CAR-T cells programmed with my Top 3 champion pairs: GPC6-CD9, CDH11-CD9, and CDH11-CD44. T T**he Results** • **Dual Mode (The AND-Gate):** When the CAR-T cells required *both* antigens to activate, they successfully navigated the tissue and achieved **7 fibrotic cell kills** per pair. • **The Safety Proof:** Across all runs, the logic-gated CAR-T cells registered exactly **0 healthy cell kills** and **0 decoy cell kills**. They completely ignored the healthy tissue. • **The Controls (Iso-A and Iso-B):** To prove the AND-gate was actually responsible for this safety, I ran control simulations where the CAR-T cells only looked for a single antigen (Antigen A only, or Antigen B only). These single-target runs resulted in **0 kills across the board** because they lacked the accumulated signal threshold to fire.

		The Conclusion Step 14 mathematically proved that the engineered logic-gate circuitry works. The simulated CAR-T cells remained entirely peaceful and inactive until they encountered the precise dual-antigen signature of a fibrotic cell. This demonstrates a massive therapeutic window and confirms that the combinatorial pairs discovered in Phase 1 and 2 can safely clear cardiac fibrosis without triggering lethal off-target toxicity.
Feb 8 – 14	Final stage!!!	Focussing on the write-up now
Feb 15		Working on final presentation
Feb 22		Finalizing write up.
March 2		Upload final presentation to CYSF platform
March 1 - 14		Continue to work on tri-fold poster