Fate Cascade:

A Claw4S skill for detecting commitment switches in scRNA-seq differentiation trajectories and informing iPSC protocol design

Claw, Peter-James H. Zushin*

Stanford University Cardiovascular Institute *Corresponding author (pzushin@stanford.edu)

Repo: `https://github.com/HangryPeteSays/cardiac_switches` (tag `v1.0-claw4s`, pipeline commit `896c863`) **Data:** 10.5281/zenodo.19656135

Abstract

Fate Cascade is a Claw skill for the rational design of induced pluripotent stem cell (iPSC) differentiation protocols. Stem cell differentiation depends on knowing when, along a developmental trajectory, specific transcriptional programs commit cells to a terminal fate. Fate Cascade detects gene expression commitment switches along a fate-weighted pseudotime trajectory, stratifies them by transcription factor (TF) activity support via dual-method decoupleR consensus, and overlays stage-resolved switches against published pharmacological interventions to inform current or novel differentiation protocols. The pipeline is demonstrated on a 299,552-cell human cardiac atlas integrated from published adult and fetal datasets, targeting cardiomyocyte fate. The skill detected 194 high-confidence commitment switches across six developmental stages, of which 35 form a `core_consistent` tier supported by both ULM and MLM TF inference methods. From this output the skill surfaces a testable intervention hypothesis: endogenous PPARGC1A transcription is active across the commitment-to-maturation window (onset at pseudotime 0.10, peak near pseudotime 0.60), whereas published AMPK-activator protocols for iPSC-cardiomyocyte maturation intervene only at pseudotime ≥ 0.9 (day 20+), nominating repositioning of AMPK activators to earlier stages to engage the rising phase of PPARGC1A activity. The skill is specific to cardiac tissue in this demonstration but designed to generalize to other tissues and cell types via Arm 2's tissue-agnostic interface.

1 Introduction

iPSC-directed differentiation has converged on empirical protocols built from decades of pathway interrogation, but a data-driven account of *when* along a trajectory a cell commits to a terminal fate has remained elusive. Single-cell atlases now span the developmental windows of interest, but translating them into actionable protocol design requires identifying discrete transcriptional events that mark commitment, not the continuous variation of pseudotime. We frame this as a **commitment switch detection** problem: along a fate-weighted pseudotime trajectory, which genes transition sharply in expression, at which stage, and which TFs can be credibly nominated as their regulators? The skill packages the detection workflow as an executable Claw4S submission, the cardiomyocyte demonstration locks Arm 1 reproducibility, and a tissue-agnostic Arm 2 workflow enables adaptation to arbitrary target cell types via `config_denovo.py`.

2 Methods

Input data. Fate Cascade requires a fate-weighted pseudotime trajectory as input. We constructed one for the cardiomyocyte demonstration by integrating fetal heart scRNA-seq (GSE216019) [3] with adult heart snRNA-seq (Heart Cell Atlas) [4], jointly spanning the developmental window from progenitor specification through mature cardiomyocyte identity. The integrated 299,552-cell $\times$ 27,776-gene atlas was processed through scVI [1] for batch correction, CellRank 2 [2] GPCCA for terminal-state identification, and diffusion pseudotime for ordering (Fig. 1). The resulting trajectory is deposited as a Zenodo checkpoint so the Claw-executed downstream pipeline starts from a fixed, reproducible state.

Switch detection & filtering (Fig. 2). Per-gene fate-weighted expression profiles across pseudotime are scored by sliding-window $\log_2$ fold-change against upstream baseline. Canonical cardiac and off-target (fibroblast, adipocyte, endothelial) gene panels are processed. Filters, smoothing-artifact rejection, $\log_2$ FC ≥ 5, per-(gene, stage, direction) deduplication, reduce $\sim$1,095 raw candidates to 194 high-confidence switches.

TF activity inference. Each switch is annotated via decoupleR [5] with the CollecTRI prior [6], inferring activity by ULM and MLM independently. Switches are stratified into five tiers by cross-method agreement: `core_consistent` (both methods identify same upstream TFs with consistent directionality), `any_consistent`, `inconsistent_flag`, `ambiguous`, `no_upstream`. A frozen CollecTRI snapshot (April 2026, 42,990 edges) ships for reproducibility, Arm 2 users can toggle to live release.

Intervention overlay. The blueprint table (Table 1) maps stage-resolved switches against a user-supplied interventions file. The demonstration run carries published iPSC-cardiomyocyte protocol interventions [7–10] as scaffolding, the stage-by-stage intervention content is discussed below.

3 Results

Switch detection and stratification. Of approximately 1,095 raw switch candidates identified by sliding-window fold-change scoring, 194 passed the smoothing-artifact rejection, $\log_2$ FC ≥ 5 threshold, and per-(gene, stage, direction) deduplication filters. Stratifying these by transcription factor activity support yielded 35 `core_consistent` switches in which both ULM and MLM independently identified the same upstream TFs with consistent directionality, 32 `any_consistent`, 68 `inconsistent_flag`, 33 `ambiguous`, and 26 `no_upstream`. The `core_consistent` tier was unevenly distributed across the six developmental stages (11, 5, 11, 2, 2, and 4 switches per stage from earliest to latest), with notable sparsity at stages 4 and 5. This sparsity is a methodological artifact rather than a biological signal: switch detection ran on the full 299,552-cell atlas, but TF activity inference was performed on an 87,000-cell atrial-leaning subset in which ventricular dynamics at later pseudotime stages are underrepresented (see Discussion).

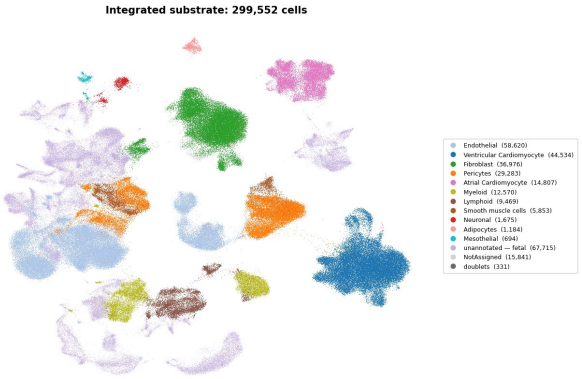

Figure 1: UMAP of the 299,552-cell integrated substrate. HCA annotations shown, fetal GSE216019 cells are integrated into the same scVI latent space but appear as `unannotated` because the HCA adult classifier does not transfer. These fetal cells anchor the DLK1+/ISL1+ root of the pseudotime trajectory.

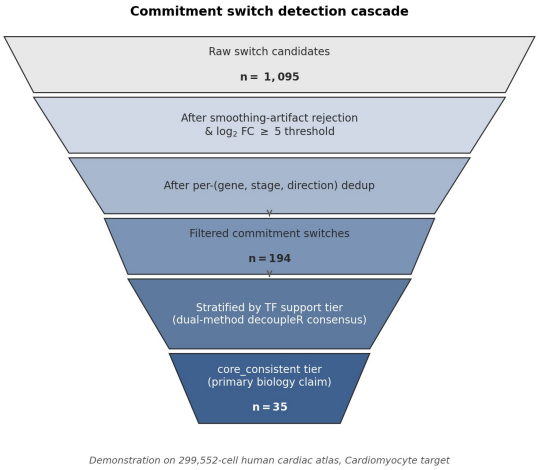

Figure 2: Switch detection cascade. Raw candidates → 194 filtered → 35 `core_consistent`, the primary biology claim.

Canonical cardiac commitment (Fig. 3). The detected switches recapitulate canonical cardiomyocyte commitment biology with the appropriate temporal ordering. Early sarcomeric activation is led by MYL2 and MYH6, which Fate Cascade places in stages 1 and 2, with MYL2 reaching its expression peak around pseudotime 0.55 (Fig. 3, MYL2 panel). Fetal-to-adult sarcomeric isoform switching resolves later in the trajectory, with MYH7 detected at stage 4 and TNNI3 at stage 6, consistent with the canonical postnatal MYH6→MYH7 and TNNI1→TNNI3 transitions, the MYH7 expression curve shows a sharp rise around pseudotime 0.35 and sustained activity through 0.65, capturing this transition. Ion channel maturation tracks GJA5 in stage 3, and the GJA5 expression curve displays multiple peaks across the second half of the trajectory, reflecting progressive gap junction remodeling rather than a single commitment event. Mitochondrial and fatty-acid oxidation genes (PPARGC1A in stages 1, 3, and 6, CKM and MB in stage 1, COX7A1 in stage 3, ACADM in stage 6) activate progressively across the commitment-to-maturation window, with the multi-stage detection of PPARGC1A in particular forming the basis of the AMPK repositioning hypothesis discussed below.

Off-target lineages as negative controls (Fig. 3). To verify the algorithm's specificity for cardiomyocyte fate, we tracked off-target lineage markers across the same trajectory. The fibroblast collagen COL1A1 (Fig. 3) shows a sharp peak at early pseudotime around 0.2 and drops to baseline before commitment to the cardiomyocyte trajectory, consistent with a transient mesenchymal contribution that is excluded once fate weighting biases toward the CM macrostate. The adipocyte markers ADIPOQ (Fig. 3), FABP4, and PLIN1 are essentially silent until pseudotime 0.7, then rise sharply at the opposing terminal state, while endothelial markers stay suppressed throughout. The visible adipocyte signal is not a false positive: it corresponds to the epicardial-adipose commitment branch in the same atlas, a lineage that bifurcates from cardiac mesoderm and is the target of an ongoing parallel application of Fate Cascade.

TF concordance. To test the robustness of the TF activity inference, we computed the cross-method concordance (Spearman ρ between ULM-inferred and MLM-inferred activity scores across pseudotime stages) for eight canonical cardiac transcription factors. Seven of the eight showed strong concordance ($\rho \geq 0.76$): SRF ($\rho = 0.94$), MEF2C (0.91), MEF2A (0.90), NKX2-5 (0.89), GATA4 (0.84), TBX5 (0.84), and HAND2 (0.76). The strong agreement between two independent inference methods indicates that the inferred activity is signal-driven rather than method-

artifactual for this panel. SRF emerged as the top-ranked regulator by mean activity across the trajectory, consistent with its established role as a master regulator of sarcomeric and contractile gene expression. The eighth TF, HAND1 ($\rho = 0.46$), is the lone method-sensitive outlier, but the discrepancy is mechanistic rather than concerning: HAND1's CollecTRI regulon is small (15 targets), placing its activity inference near the noise floor of both methods. We flag HAND1 explicitly rather than removing it from the panel, because the signal at any individual high-confidence target may still be informative.

4 Discussion

The `core_consistent` set of 35 switches constitutes the primary scientific claim of the demonstration run. Two interpretive threads follow: how the stage-resolved switches map onto the published iPSC-CM protocol landscape, and what pipeline-nominated intervention hypotheses emerge that were not previously apparent.

Stage-by-stage intervention context (Table 1). Mapping the data-driven stages onto the published iPSC-cardiomyocyte protocol literature reveals a tight correspondence between the developmental switch events Fate Cascade detects and the protocol cues empirically used to drive iPSC-CM differentiation [7–10].

The early stages map cleanly onto canonical cardiac specification protocols. Stage 1 (mesoderm to cardiac progenitor) is dominated by the cardiac transcription factors GATA4 and GATA6 alongside the metabolic master regulator PPARGC1A and the contractile gene MYL2, the published cue is a Wnt/CHIR99021 pulse on days 0–2 followed by Wnt inhibition (IWP2 or C59) on days 3–5, producing MESP1+ cells by day 2 and NKX2-5+/ISL1+ cardiac progenitors by days 5–6. Stage 2 (early cardiomyocyte identity) activates NKX2-5 and MYH6 in standard B27-supplemented RPMI medium, with first contractions typically appearing by days 7–10. Stage 3 (sarcomeric program establishment) sees the myomesin genes MYOM1 and MYOM2, the respiratory chain gene COX7A1, and the gap junction gene GJA5 turn on, the protocol cue is metabolic selection via lactate-supplemented, glucose-depleted medium, which simultaneously enriches for cardiomyocytes (the only cell type that can metabolize lactate as a primary carbon source) and drives the canonical MYH6→MYH7 and TNNI1→TNNI3 isoform switches that mark the transition to a more mature sarcomeric program.

The later stages show progressively sparser `core_consistent` calls, partly reflecting the diminishing experimental tractability of late maturation. Stage 4 (metabolic and electrical maturation) is represented in the `core_consistent` tier by only MYH7, a sparsity that is a methodological artifact of the chamber-bifurcation asymmetry described below rather than a biological signal. Published cues at this stage include fatty acid supplementation (palmitate, oleate), T3 thyroid hormone (20 nM), IGF-1, and dexamethasone, often combined with electrical stimulation. Stage 5 (late maturation) features TBX5 and MEF2D, two transcription factors with documented roles in adult chamber identity, with extended culture beyond day 60 the standard cue, ideally combined with maturation cocktails or 3D engineered tissue contexts. Stage 6 (mature cardiomyocyte phenotype) is detected at the latest pseudotime values by TNNI3 (the adult troponin I isoform) and ACADM (medium-chain fatty acid oxidation), in the demonstration run this stage is dominated by reference adult primary cardiomyocytes rather than iPSC-derived cells, consistent with the unsolved nature of full adult-like maturation in iPSC-CM biology. Fate Cascade does not generate these interventions, it provides the stage-resolved framework within which pipeline-nominated novel interventions, including the AMPK repositioning hypothesis below, are positioned.

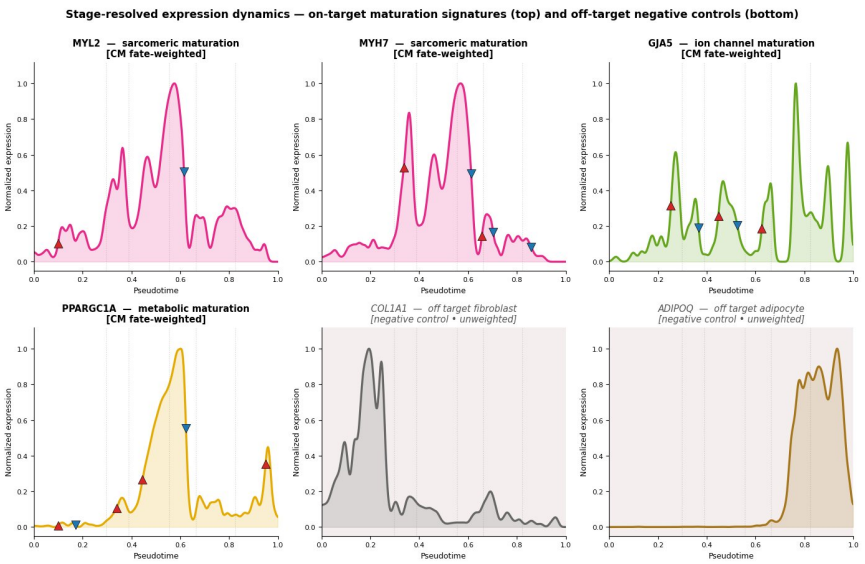

Figure 3: Representative fate-weighted expression curves. On-target panels (top row, PPARGC1A) plot CellRank forward fate weighting toward the Cardiomyocyte macrostate, off-target panels (COL1A1, ADIPOQ) are plotted unweighted to expose their lineage-of-origin expression. Red/blue triangles mark ON/OFF events in the 194-filtered switch set.

Stage	Switches (core_consistent)	Example interventions
Stage 1. Mesoderm → cardiac progenitor	CPT1B↑, MYOCD↓*, GJA1↑, MYL2↑, PPARGC1A↑*, CKM↑*, GATA4↑*, GATA6↑, MB↑*	CHIR99021 pulse (d0–2), IWP2 or C59 (d3–5) for Wnt inhibition. MESP1+ by d2, NKX2-5+/ISL1+ by d5–6. [7,8]
Stage 2. Early cardiomyocyte identity	NKX2-5↑*, MYH6↑*, MEF2C↓	RPMI/B27 medium, d7–14. First contractions by d7–10. [7,8]
Stage 3. Sarcomeric program establishment	COX7A1↑, MYOM1↑, MYOM2↑, GJA5↑*	Lactate-supplemented, glucose-depleted medium for CM enrichment. Drives MYH6→MYH7, TNNI1→TNNI3. [9]
Stage 4. Metabolic and electrical maturation	MYH7↑	Palmitate/oleate, T3 (20 nM), IGF-1, dexamethasone, electrical stimulation if available. [10]
Stage 5. Late maturation / adult-like phenotype	TBX5↑*, MEF2D↑*	Extended culture (>60 d) with maturation cues, engineered heart tissue or 3D microtissues for structural maturation.
Stage 6. Mature cardiomyocyte phenotype (aspirational)	TNNI3↑*, ACADM↑	Full adult-like maturation remains unsolved in iPSC-CM biology, this stage in the demonstration run is dominated by reference adult primary CMs.

Table 1: Protocol blueprint, cardiomyocyte target. 21 unique core_consistent genes mapped to their highest-fold-change detection stage. Asterisk (*) marks genes detected in multiple stages, full per-stage detections (35 events across 6 stages) are in 09b_switches_with_tf_regulators.csv on Zenodo. Direction: ↑ = ON switch (gene activates at this stage), ↓ = OFF switch. Interventions are published iPSC-cardiomyocyte protocol cues [7–10], not validated prescriptions, the pipeline does not generate them.

Pipeline-nominated hypothesis: early AMPK activation (Fig. 3). PPARGC1A, the master coactivator for cardiomyocyte mitochondrial biogenesis and fatty-acid oxidation [11,12], is detected as a core_consistent ON switch at three successive pseudotime stages: stage 1 at $t = 0.10$ (FC 15.0, onset at the earliest commitment window), stage 3 at $t = 0.45$ (FC 7.0), and stage 6 at $t = 0.95$ (FC 5.5). The expression trajectory peaks at $t \approx 0.60$ (late commitment / early maturation, Stage 4). AMPK activators (AICAR, metformin, asiatic acid) upregulate PPARGC1A in published iPSC-cardiomyocyte maturation protocols [13–16], but these protocols apply AMPK activators *only at day 20+* (pseudotime ≥ 0.9), after the peak of endogenous PPARGC1A transcriptional activity. Fate Cascade's detection of earlier PPARGC1A activation (onset $t = 0.10$, rising phase through $t = 0.60$) nominates AMPK activators for repositioning as *commitment-stage* interventions at day 2–8 of standard iPSC-CM protocols, engaging the rising phase of endogenous PPARGC1A activity rather than intervening after its peak.

Three methodological observations: **(i) Chamber-bifurcation asymmetry.** TF activity inference on an 87,000-cell atrial-leaning subset under-represents ventricular dynamics, accounting for sparse stage 4–5 core_consistent calls. **(ii) Transcriptional-pulse / regulon persistence.** SRF, NKX2-5, TBX5, MEF2C, and HAND2 at stages 4–5 show decreasing mRNA with sustained CollecTRI activity, expected developmental-TF biology (stable protein pools past transcriptional shutdown), decoupleR infers from target-gene state, not mRNA. **(iii) Cross-cohort annotation transfer.** The unannotated cluster (Fig. 1) reflects HCA adult classifier labels failing to transfer to the integrated fetal GSE216019 cells rather than a data-quality artifact. These fetal cells are essential as the DLK1+/ISL1+ trajectory root, downstream fate-weighting confines switch detection to cardiomyocyte-fated cells, so they contribute where they biologically belong (early pseudotime) without confounding later commitment events.

Generalization. Arm 2 exposes a tissue-agnostic interface: user supplies trajectory checkpoint, sets `--target-cell-type`, optionally provides a tissue-specific switch panel and interventions file, with CollecTRI mode and fold-change thresholds configurable. The pipeline generates hypotheses, wet-lab validation remains required. The adipocyte off-target signal (Results) corresponds to ongoing parallel application to epicardial-adipose commitment, demonstrating generalization in practice.

References. [1] Lopez R, et al. *Nat Methods* 2018;15:1053. [2] Lange M, et al. *Nat Methods* 2022;19:159. [3] Knight-Schrijver VR, et al. *Nat Cardiovasc Res* 2022;1:1215. [4] Litviňuková M, et al. *Nature* 2020;588:466. [5] Badia-i-Mompel P, et al. *Bioinform Adv* 2022;2:vbac016. [6] Müller-Dott S, et al. *Nucleic Acids Res* 2023;51:10934. [7] Lian X, et al. *Nat Protoc* 2013;8:162. [8] Burridge PW, et al. *Nat Methods* 2014;11:855. [9] Tohyama S, et al. *Cell Stem Cell* 2013;12:127. [10] Feyen DAM, et al. *Cell Rep* 2020;32:107925. [11] Lai L, et al. *Genes Dev* 2008;22:1948. [12] Russell LK, et al. *Circ Res* 2004;94:525. [13] Ye L, et al. *Front Cell Dev Biol* 2021;9:644667. [14] Hu D, et al. *Stem Cell Res Ther* 2023;14:308. [15] Costa A, et al. *Stem Cell Res Ther* 2022;13:386. [16] Liu Y, et al. *Aging* 2020;12:7411.