

HARVARD MEDICAL SCHOOL | DEPARTMENT OF Biomedical Informatics

Carnegie Mellon University

Controllable Sequence Editing for Biological and Clinical Trajectories

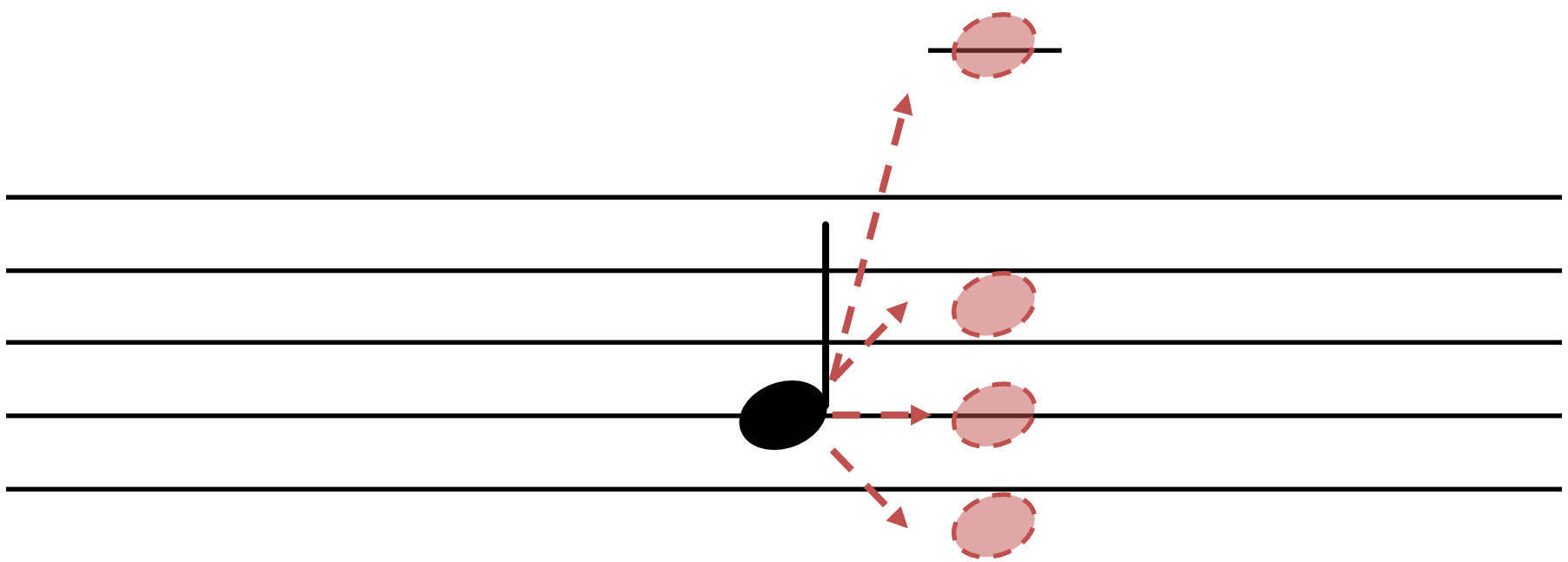
Michelle M. Li, PhD

Berkowitz Postdoctoral Research Fellow | Harvard Medical School
Incoming Assistant Professor | Carnegie Mellon University

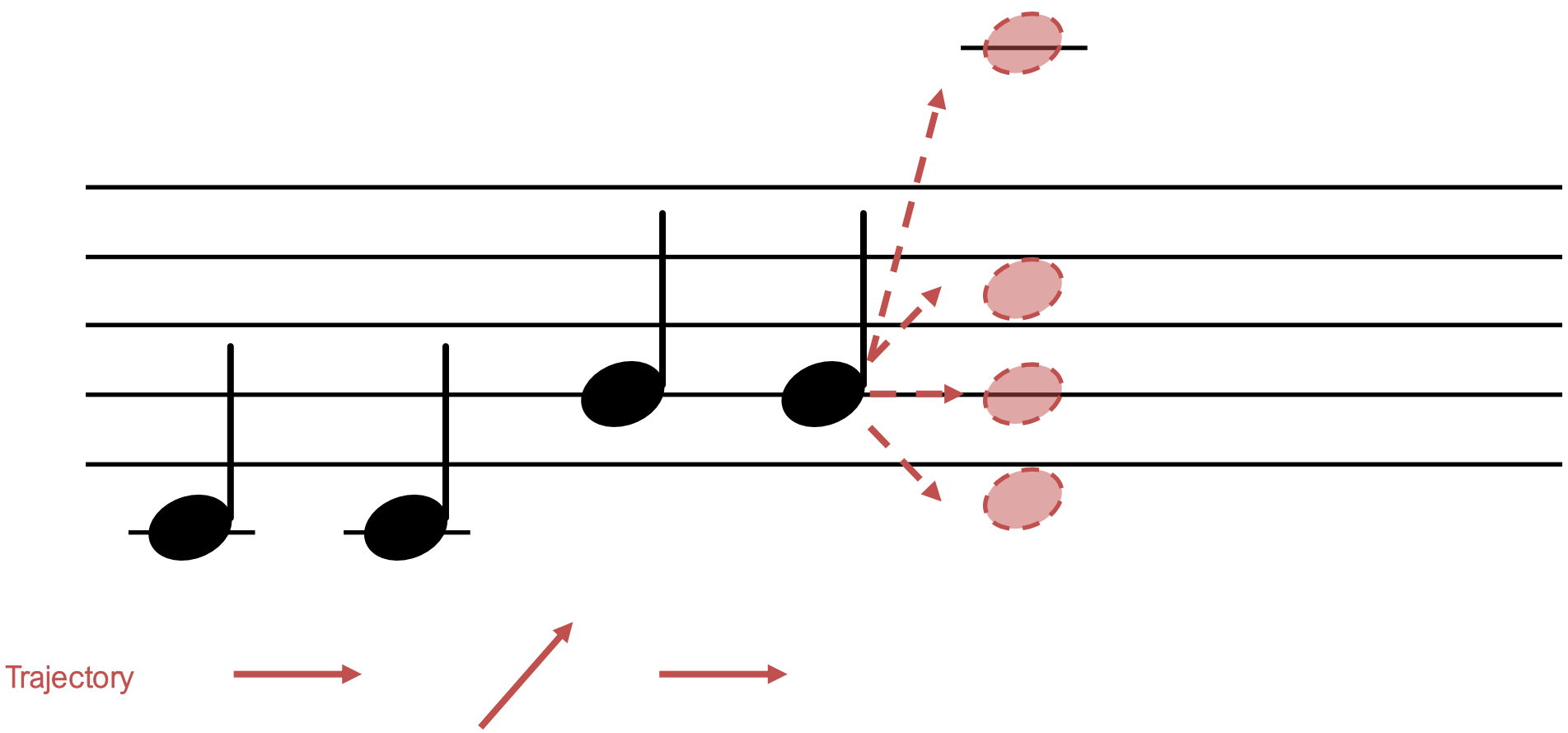
May 26, 2026

michellemli.com

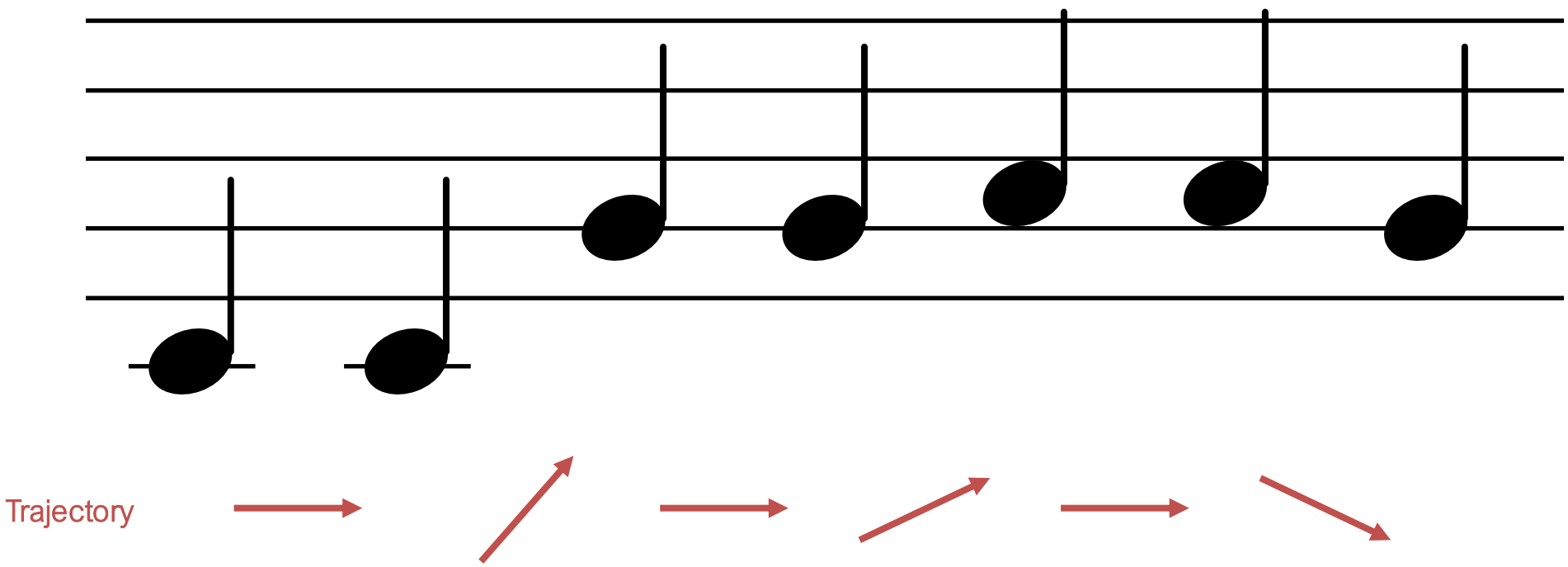
Finish the melody



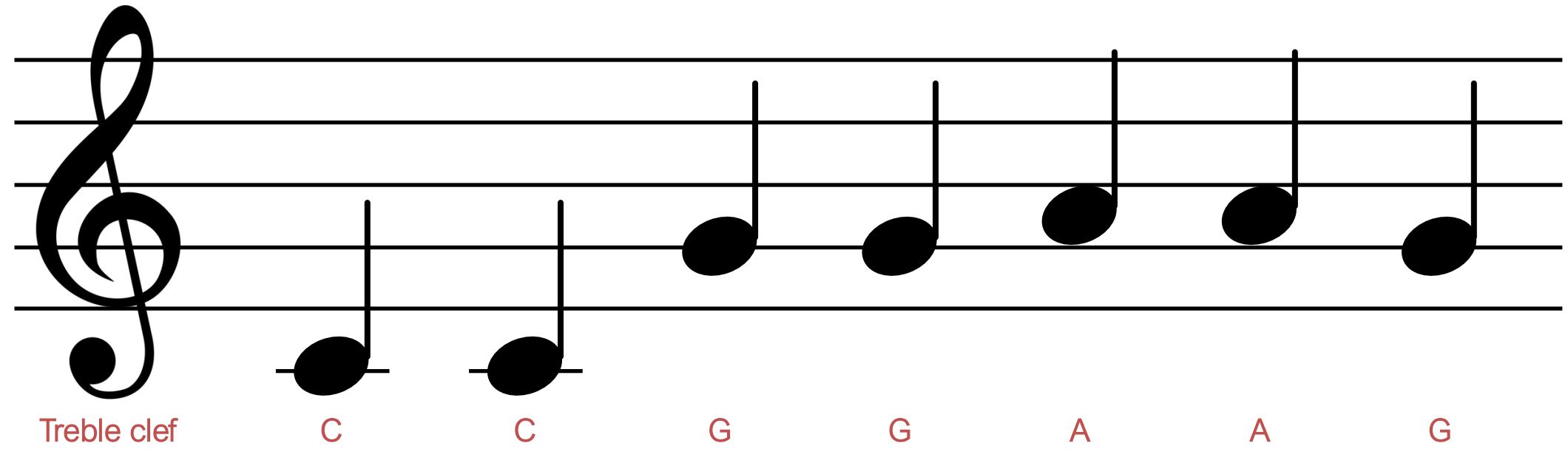
Finish the melody

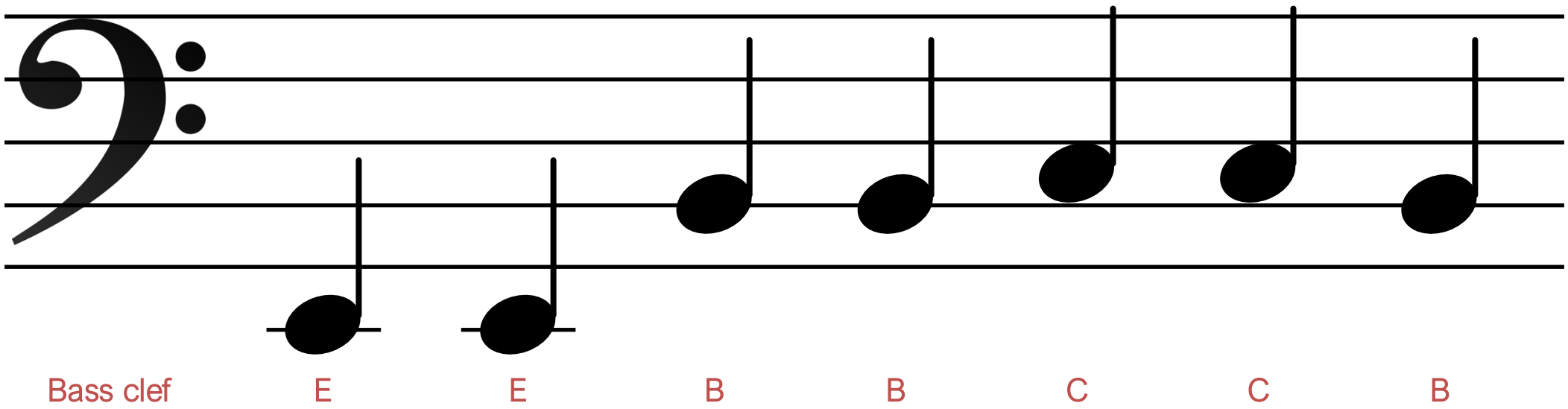


Finish the melody

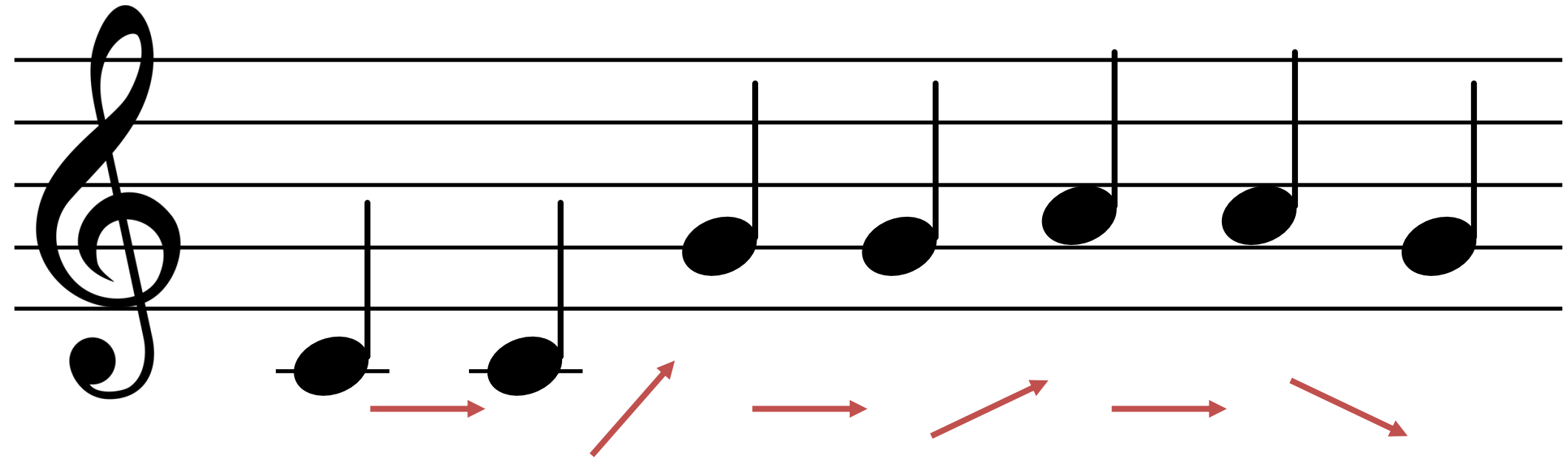


Melodies vary depending on context

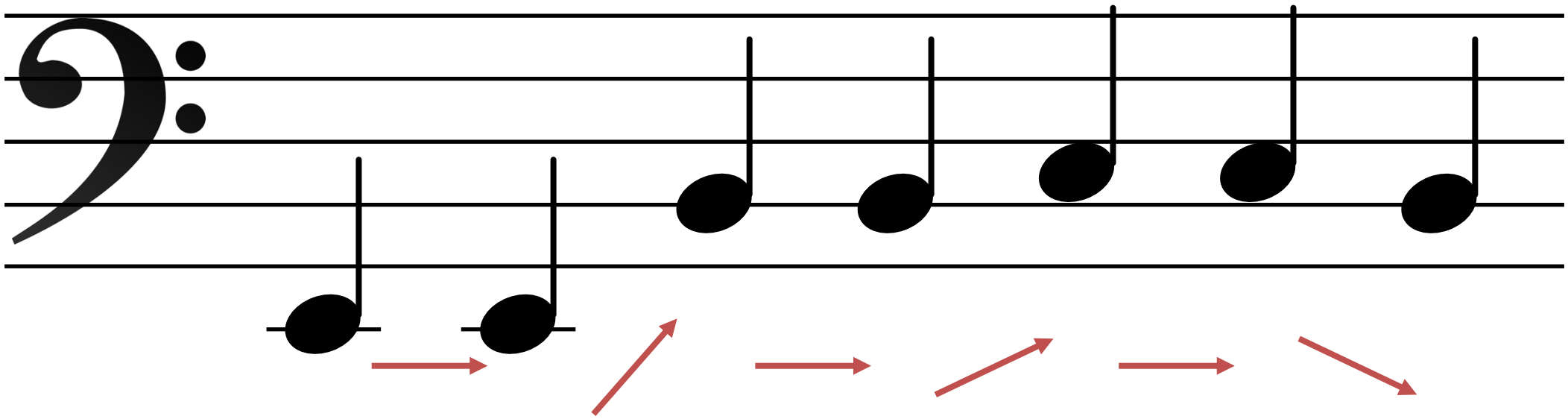




Melodies vary depending on context

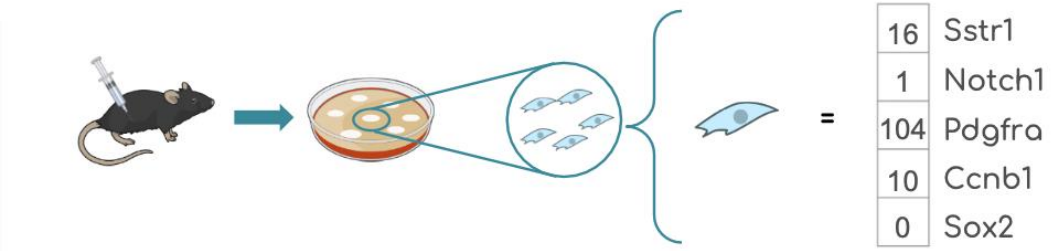


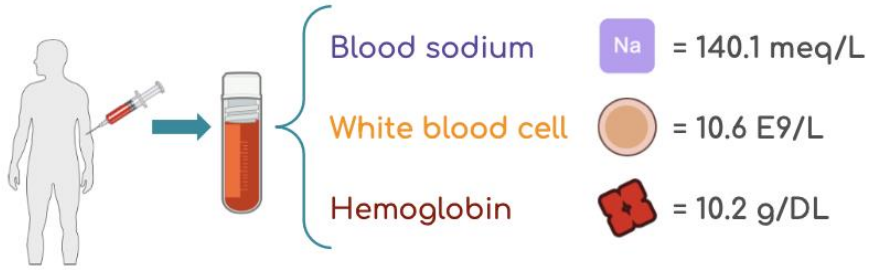
"Happy"
Twinkle Twinkle
Little Star

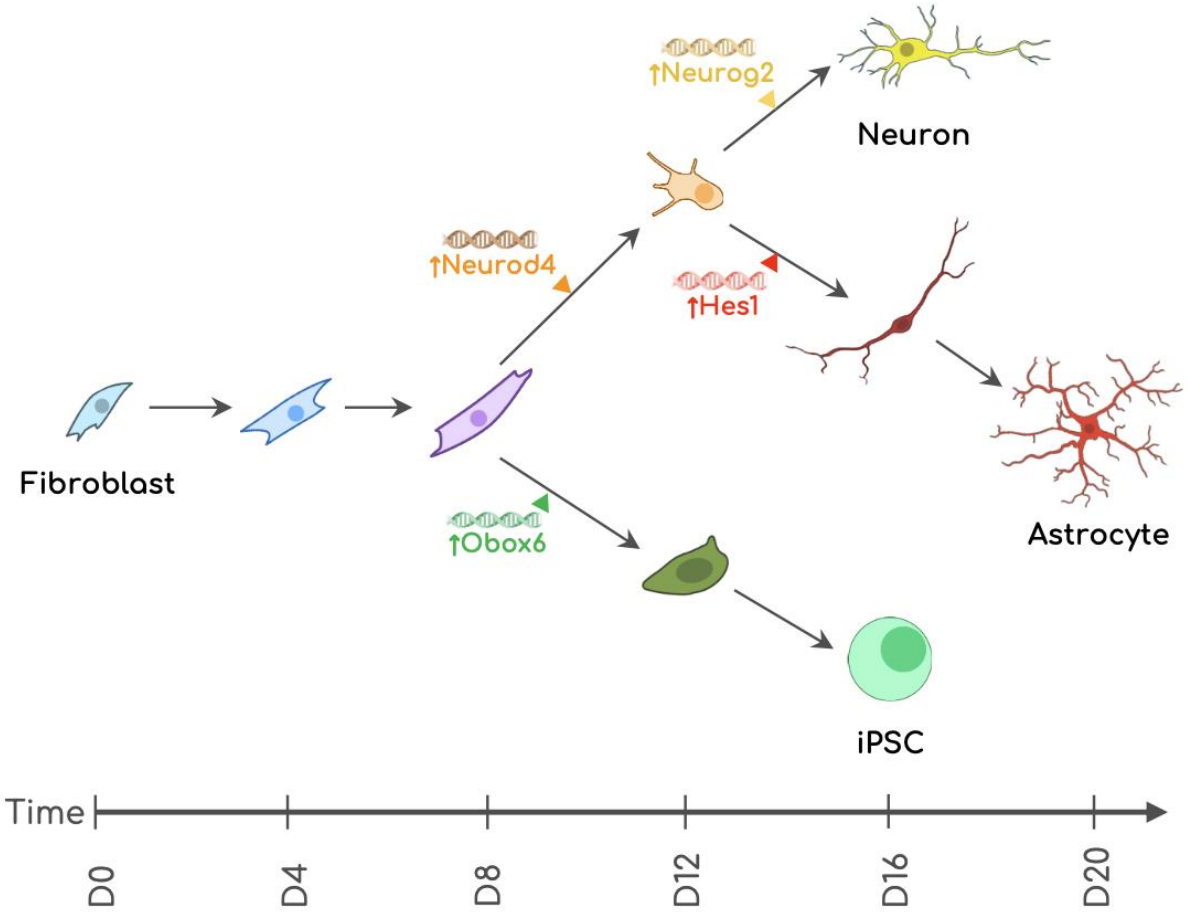


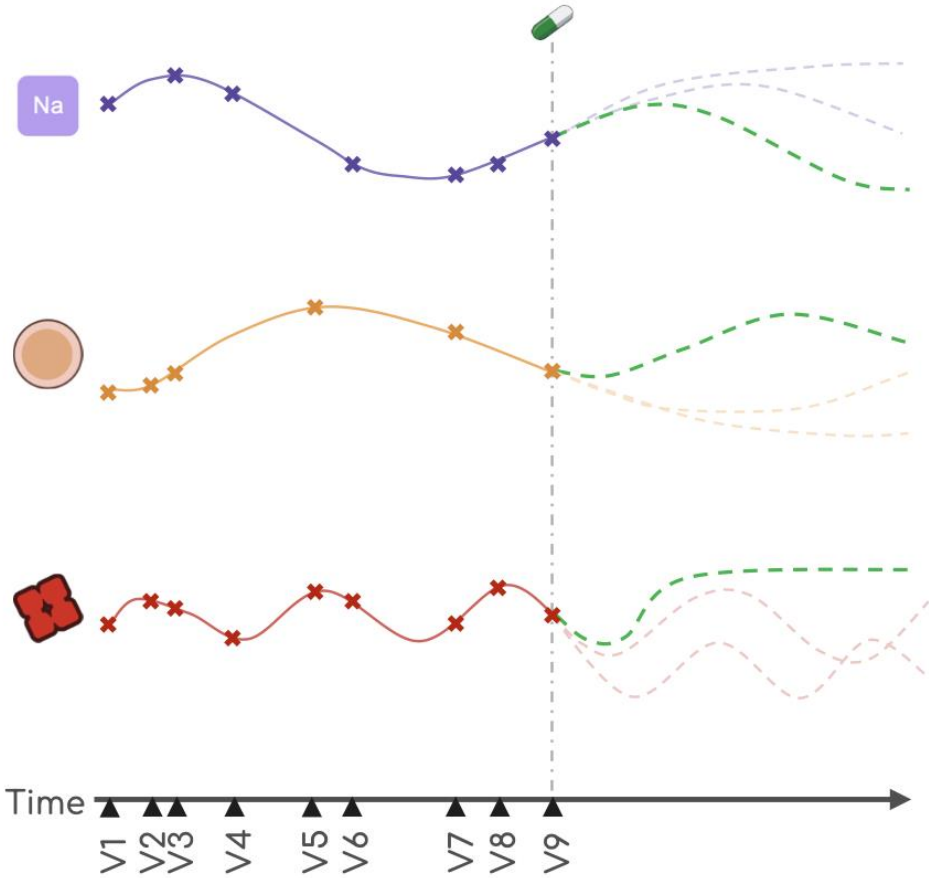
"Creepy"
Twinkle Twinkle
Little Star

Finish the cellular and patient sequences

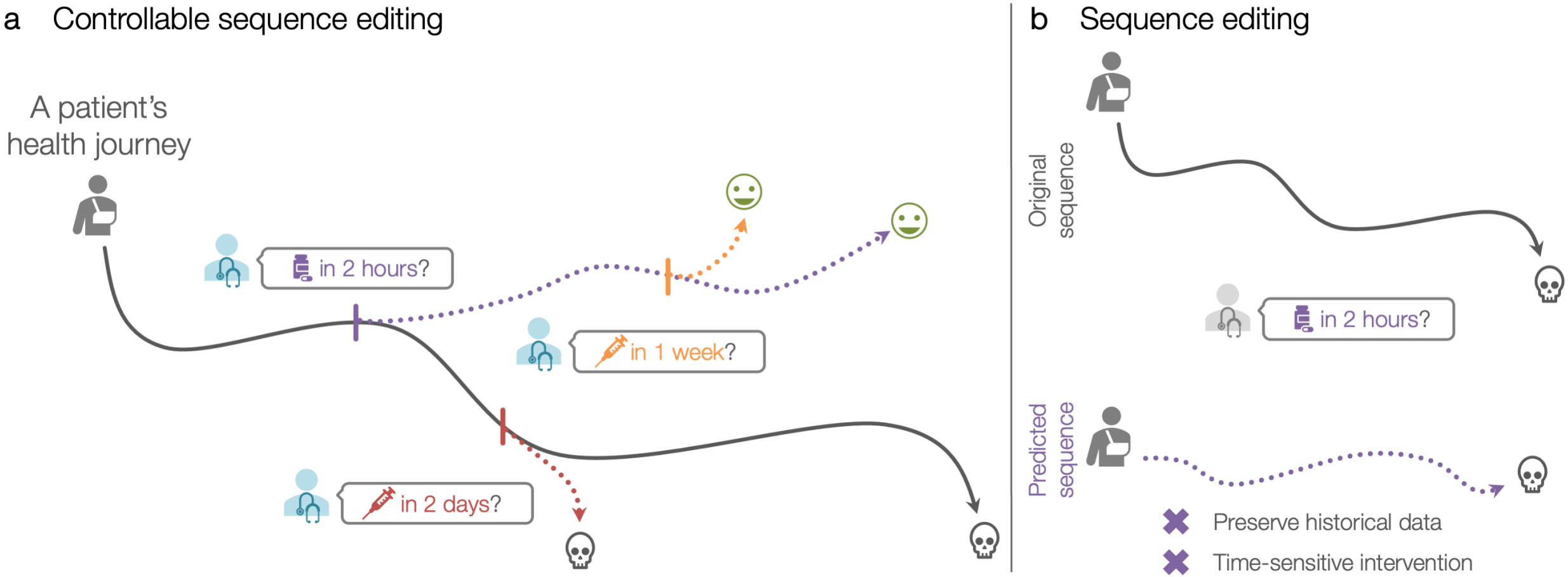




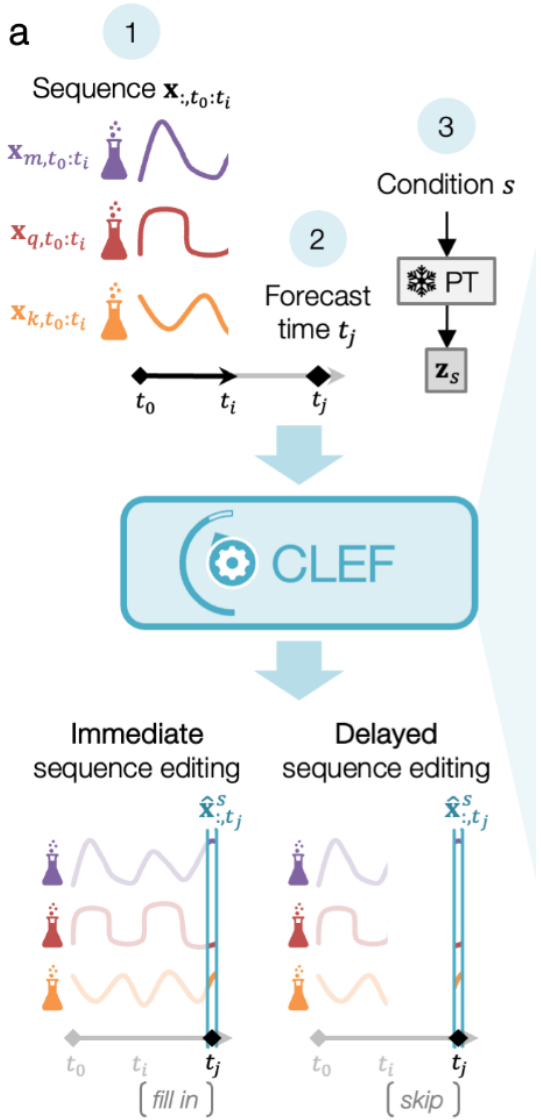




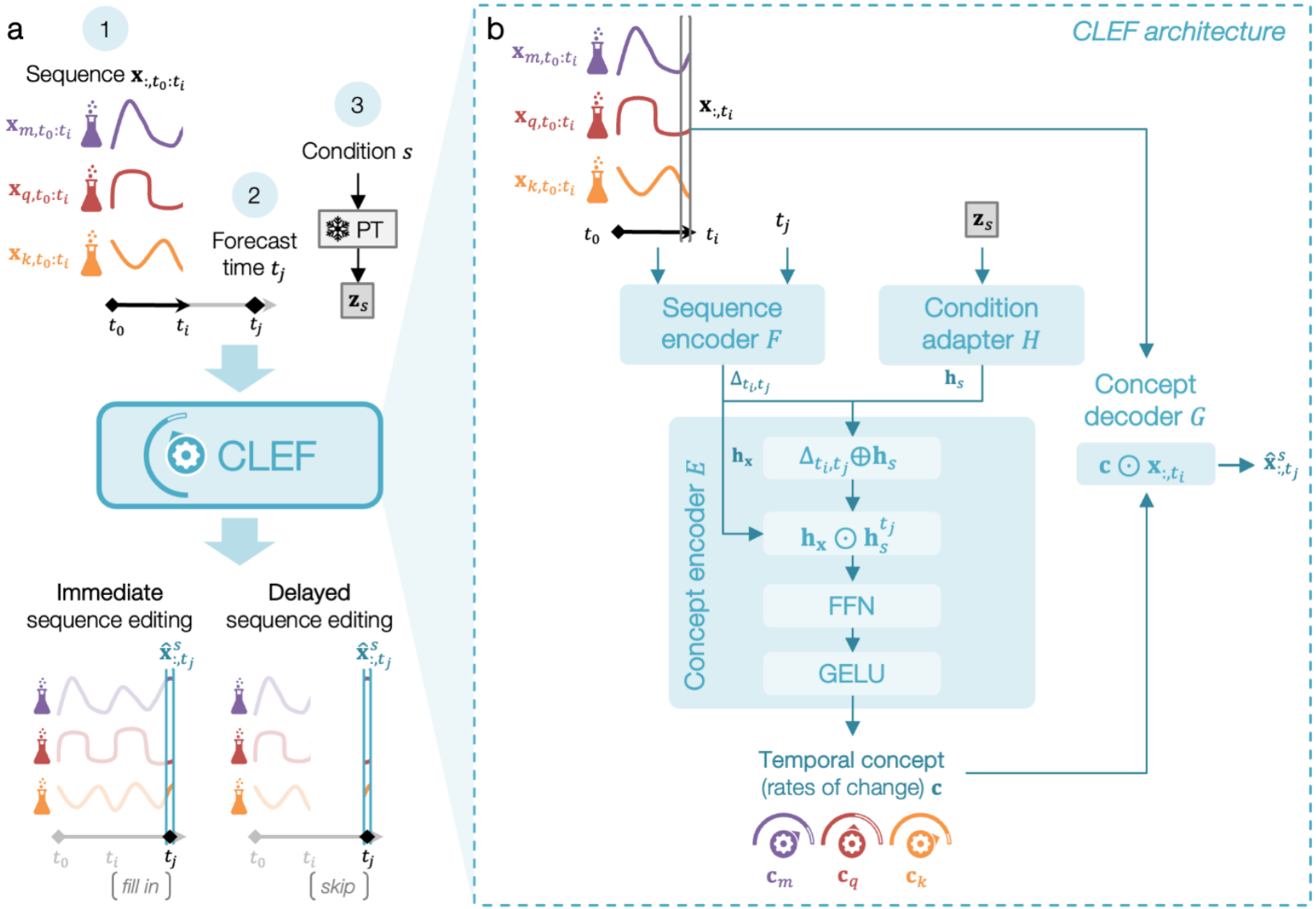
CLEF: Controllable sequence editing for conditional generation of cellular and patient trajectories



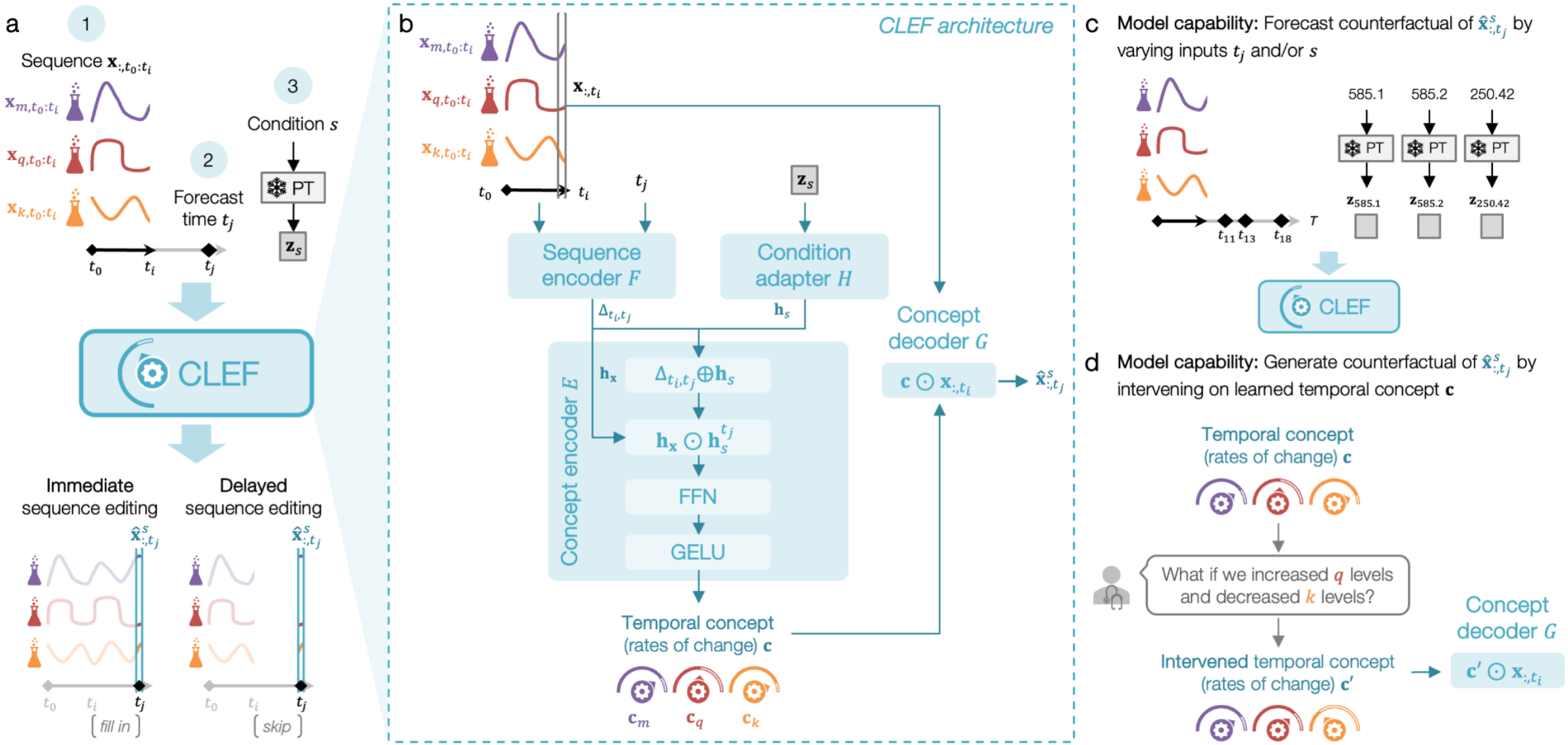
CLEF model architecture



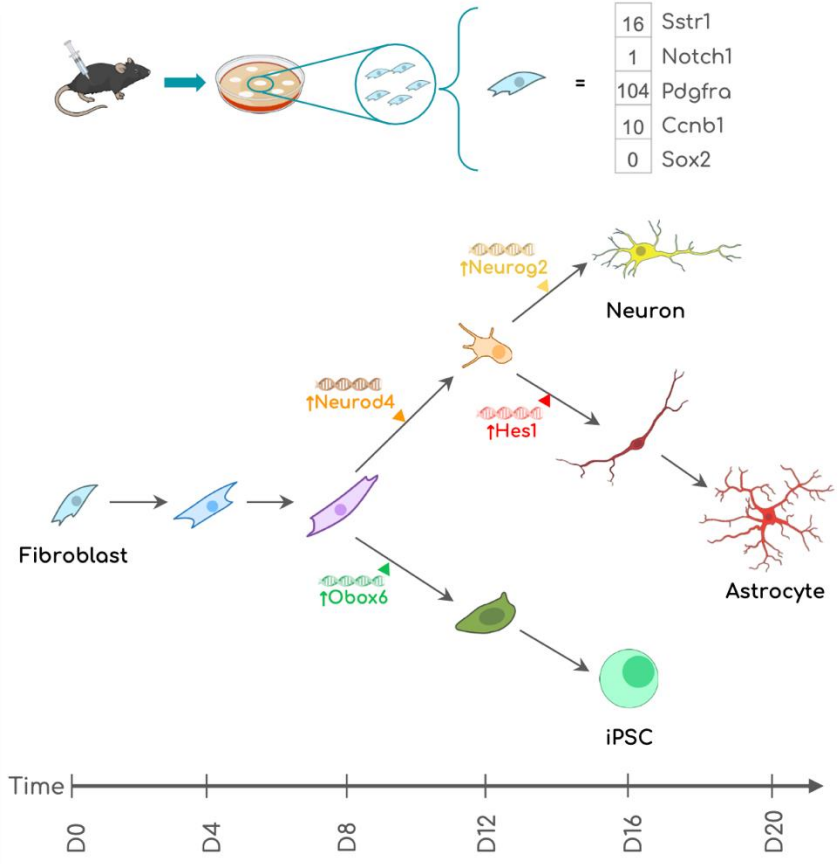
CLEF model architecture

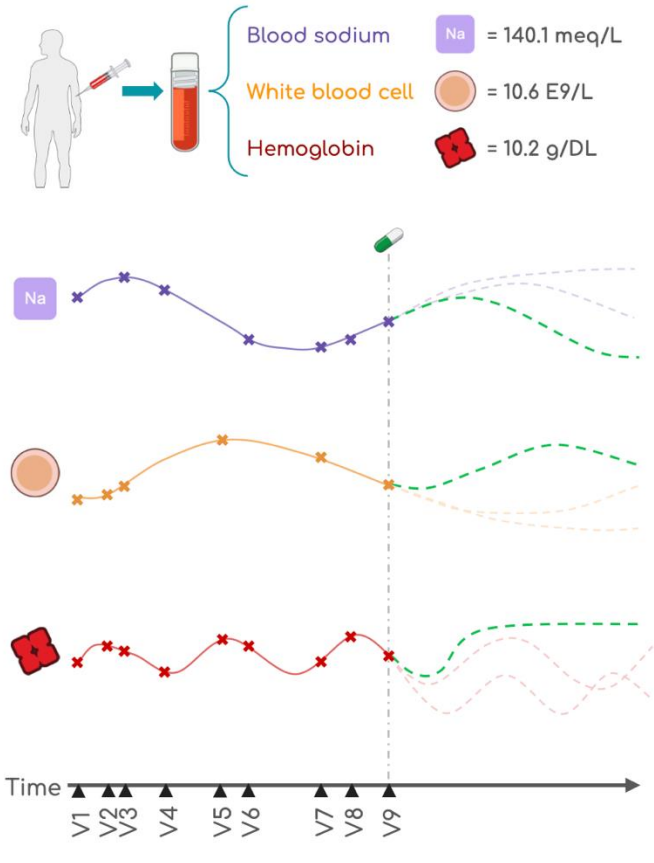


CLEF model architecture



Data: Constructing trajectories of cellular development and patient immune dynamics

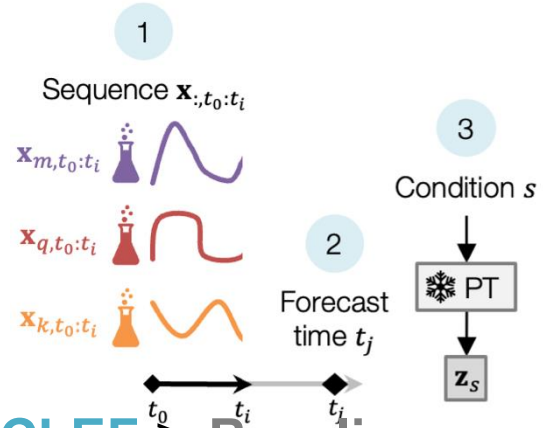




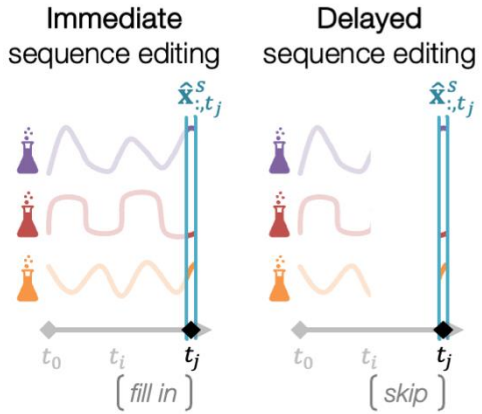
Dataset	N	V	Mean L	Max L
WOT	$3,000$	$1,480$	27.03 ± 6.04	37
WOT-CF	$2,560$	$1,480$	27.31 ± 6.11	37
eICU	$108,346$	17	20.27 ± 25.23	858
MIMIC-IV	$156,310$	16	15.56 ± 24.43	949

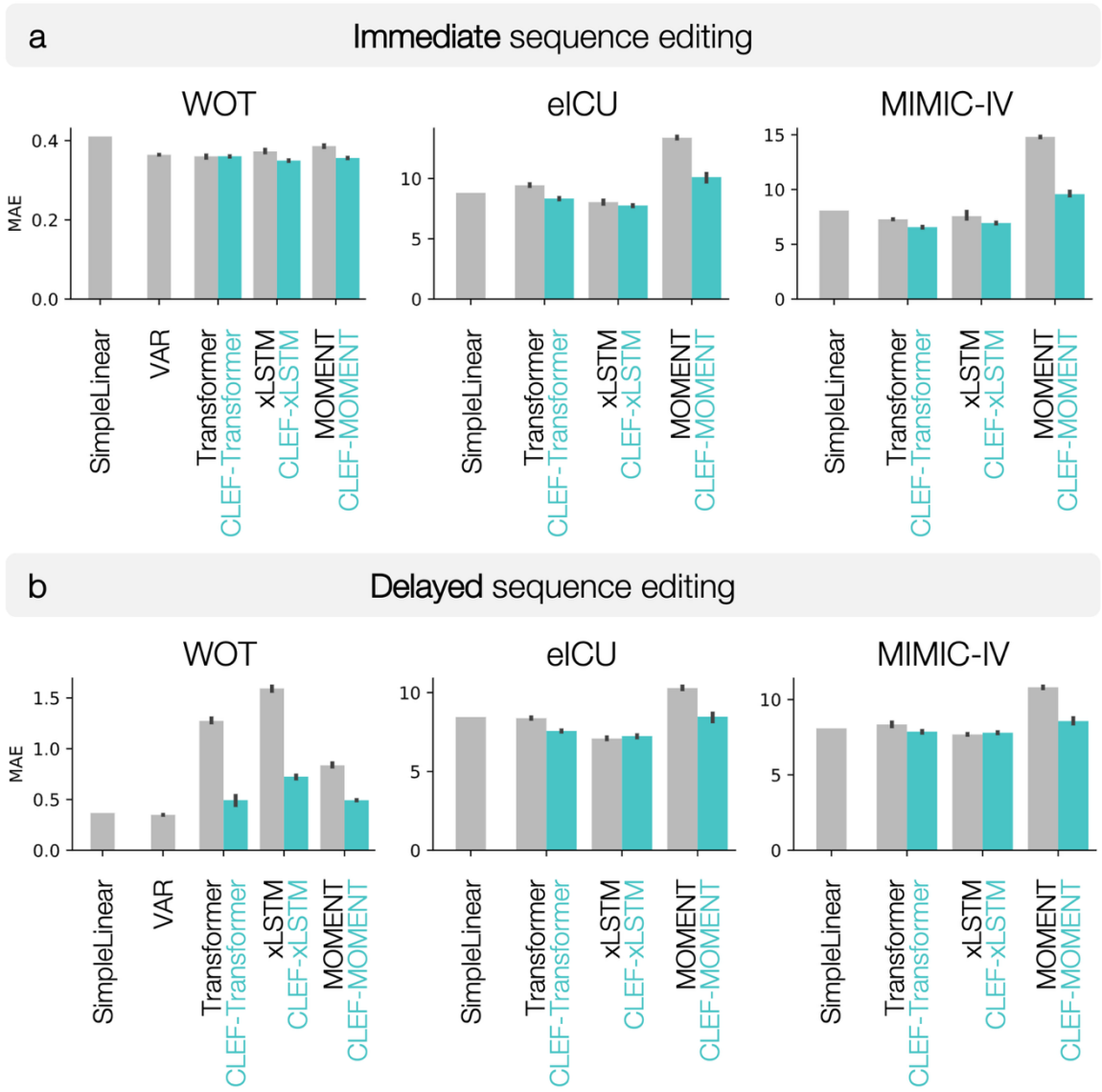
N = Number of trajectories
V = Number of covariates
L = Length of trajectories

Results: CLEF outperforms SOTA in conditional generation



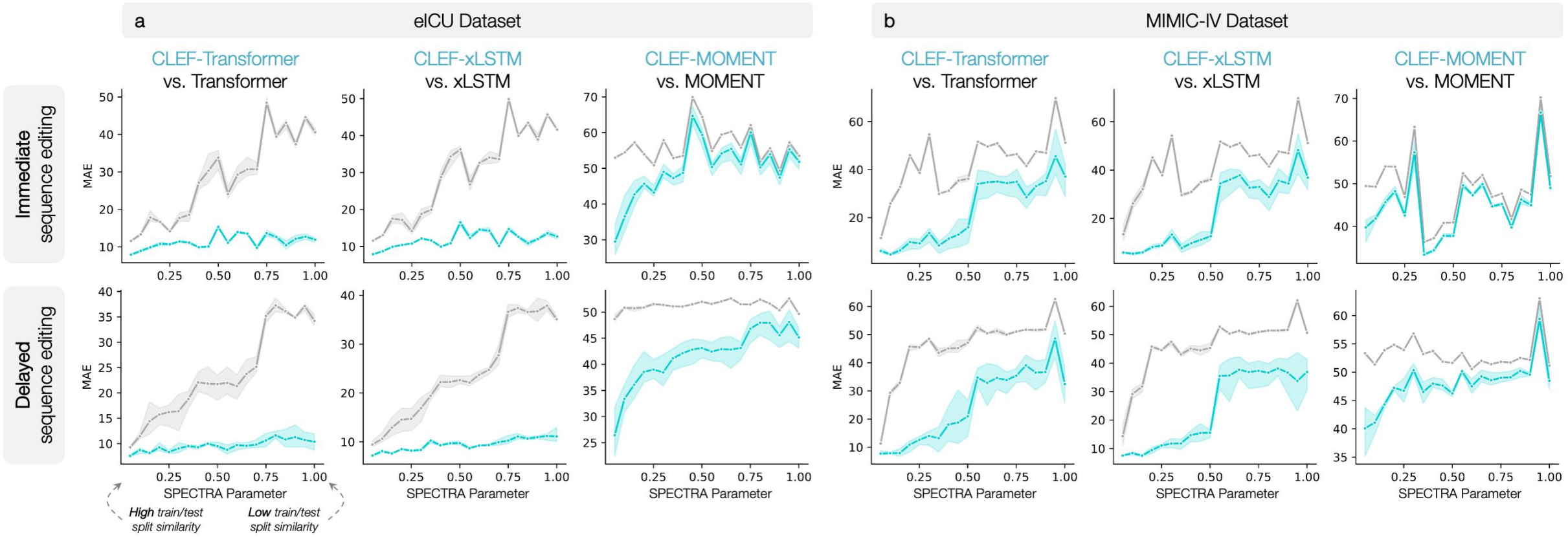
CLEF ≥ **Baseline** on accurately generating cellular and patient trajectories, suggesting that temporal concepts can improve conditional generation.





Results: CLEF outperforms SOTA in conditional generation

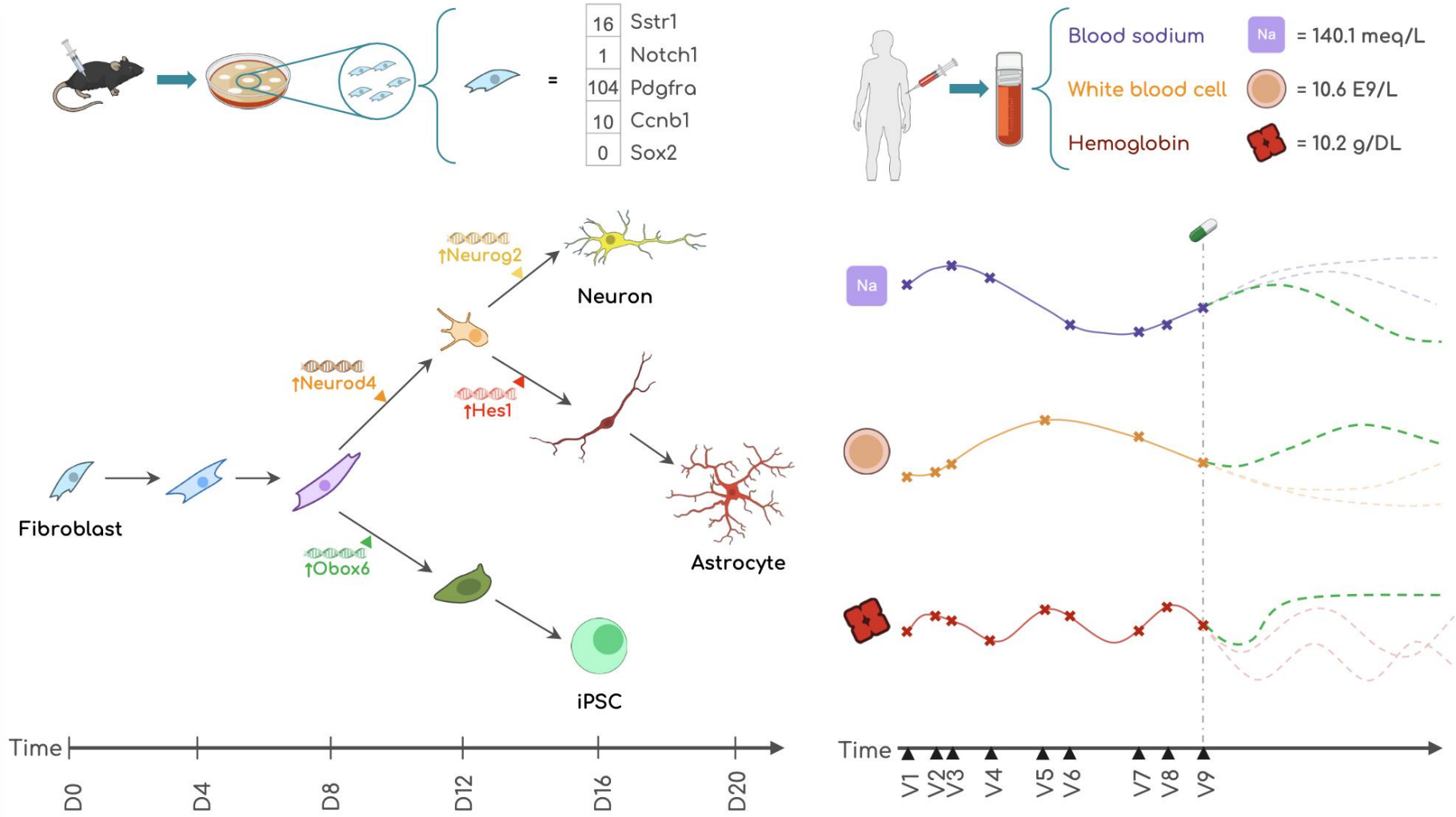
CLEF > **Baseline** on generalizing to unseen patients, even when they are very different from the patients seen during training.



Contextual learning for *simulating* cellular developmental trajectories and patient immune dynamics

Ask **“what if”** questions about the effects of a candidate intervention at a specific time point.

What medication should I prescribe? Should I prescribe it now? In 1 year? In 10 years?



CLEF's connection to counterfactual prediction

The *only* treatment between any two time points is s.

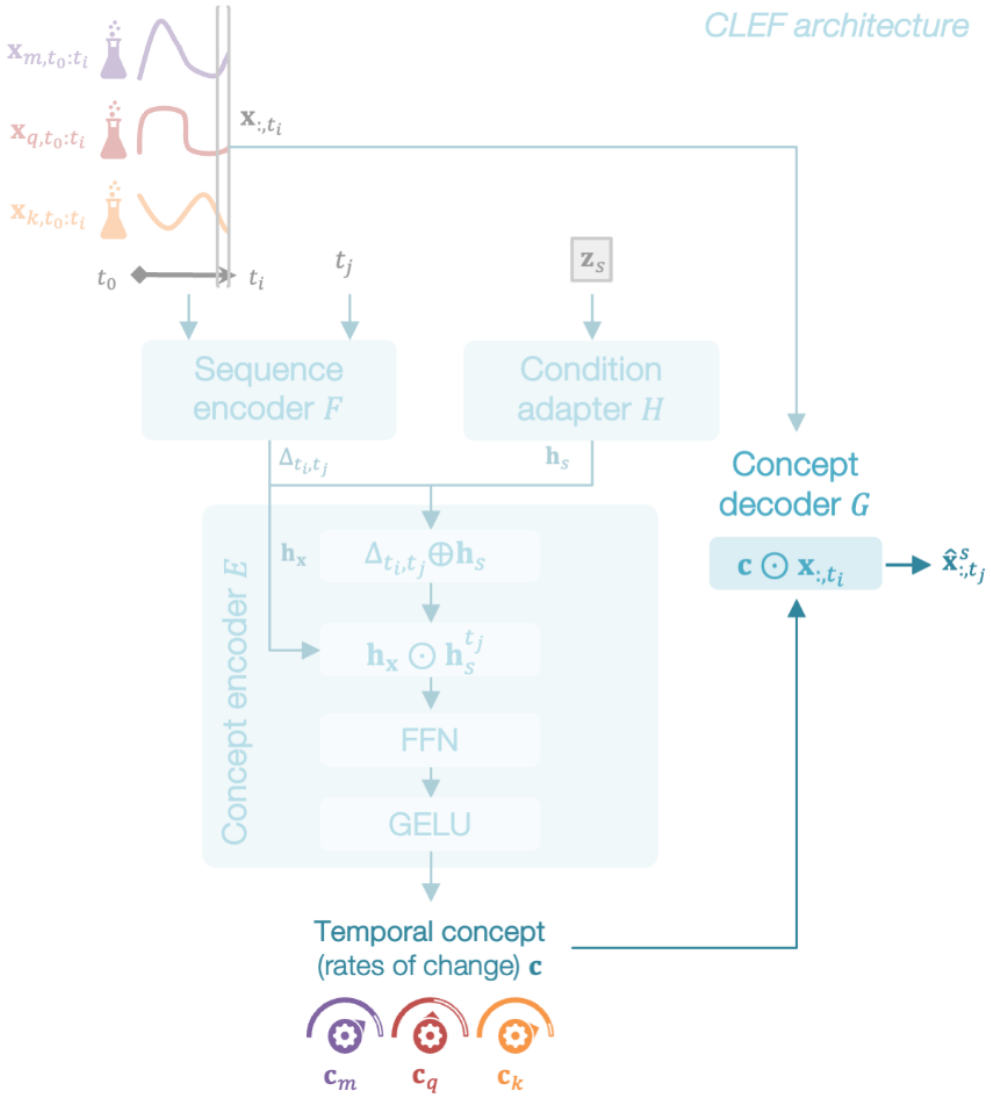
Formally, under the **potential outcomes framework** (Neyman, 1923; Rubin, 1978) and **its extension to time-varying treatments and outcomes** (Robins & Hernan, 2008), the potential counterfactual outcomes over time are identifiable from the observational data under three standard assumptions: **consistency, positivity, sequential ignorability**.

Temporal concepts $\mathbf{c}$ learned based on the next treatment $\mathbf{s}_{t_j}$, historical treatments $\mathbf{s}_{t_0:t_i}$, and historical covariates $\mathbf{x}_{:,t_0:t_i}$ capture **(balanced) representations** such that the **concept decoder**

$$\mathbf{c}\left(\mathbf{s}_{t_j}, \mathbf{s}_{t_0:t_i}, \mathbf{x}_{:,t_0:t_i}\right) \odot \mathbf{x}_{:,t_i}$$

approximates the **conditional mean function**

$$\mathbb{E}\left[\mathbf{x}_{:,t_j+\epsilon}\left(\mathbf{s}_{t_j}, \mathbf{s}_{t_0:t_i}\right) \middle| \mathbf{s}_{t_0:t_i}, \mathbf{x}_{:,t_0:t_i}\right]$$

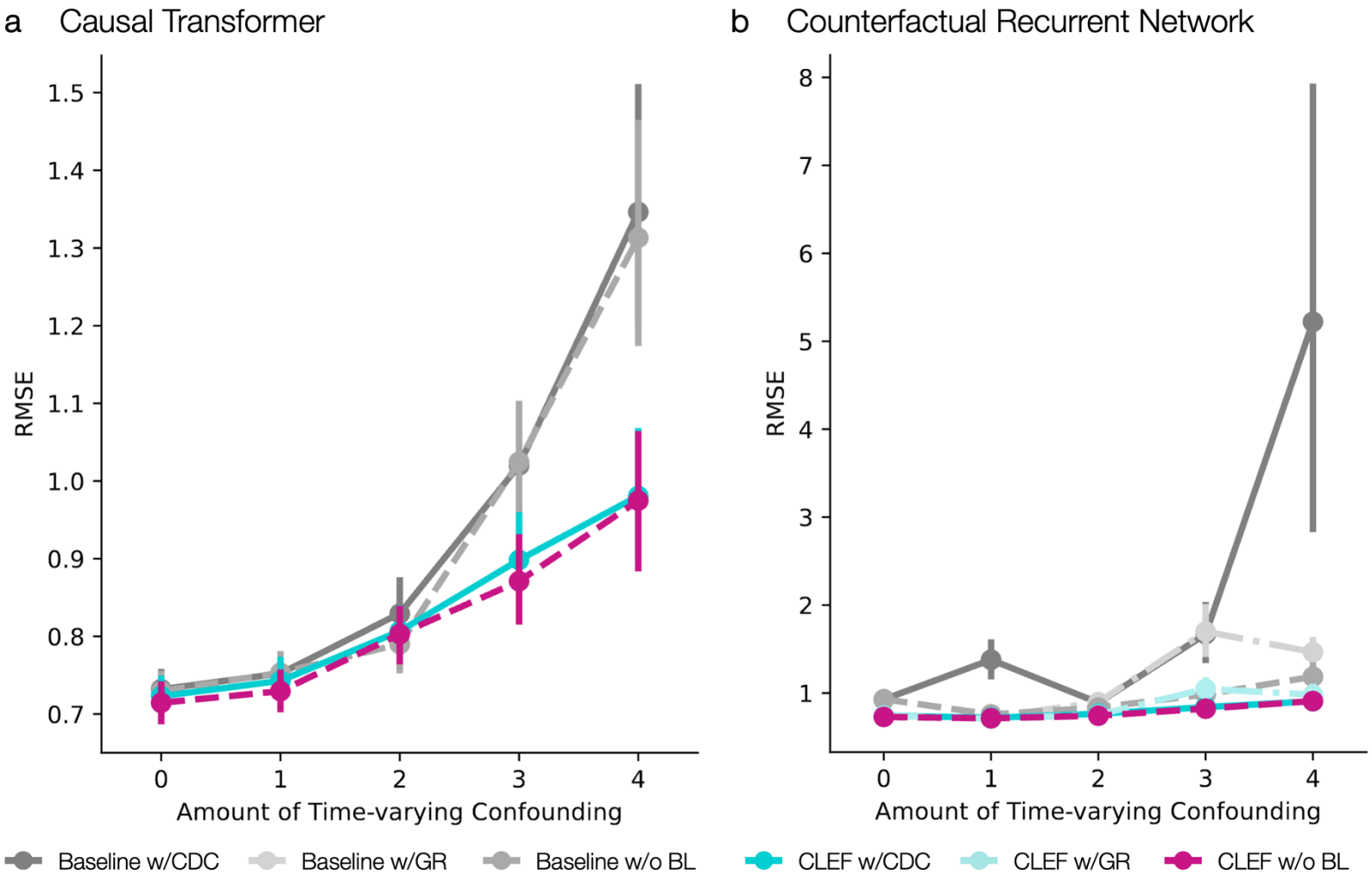


Results: CLEF outperforms SOTA counterfactual prediction

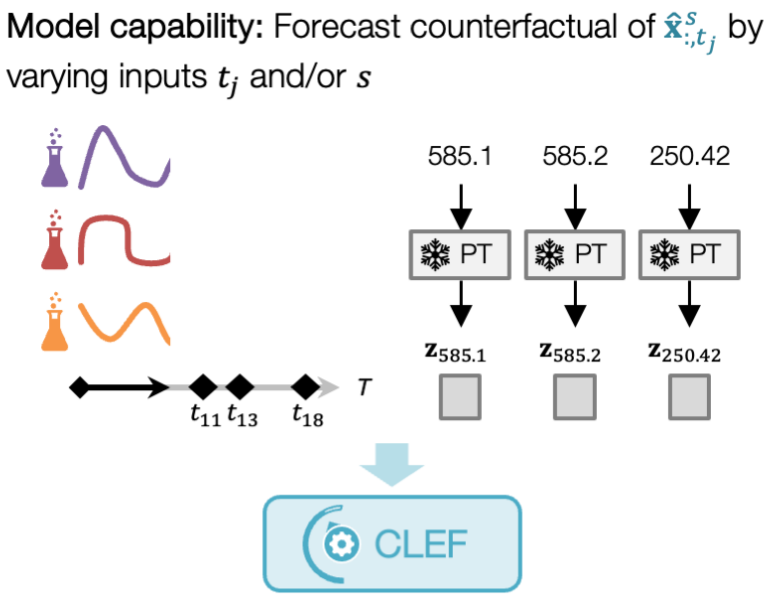
CLEF ≥ **Baseline** with **low** time-varying confounding.

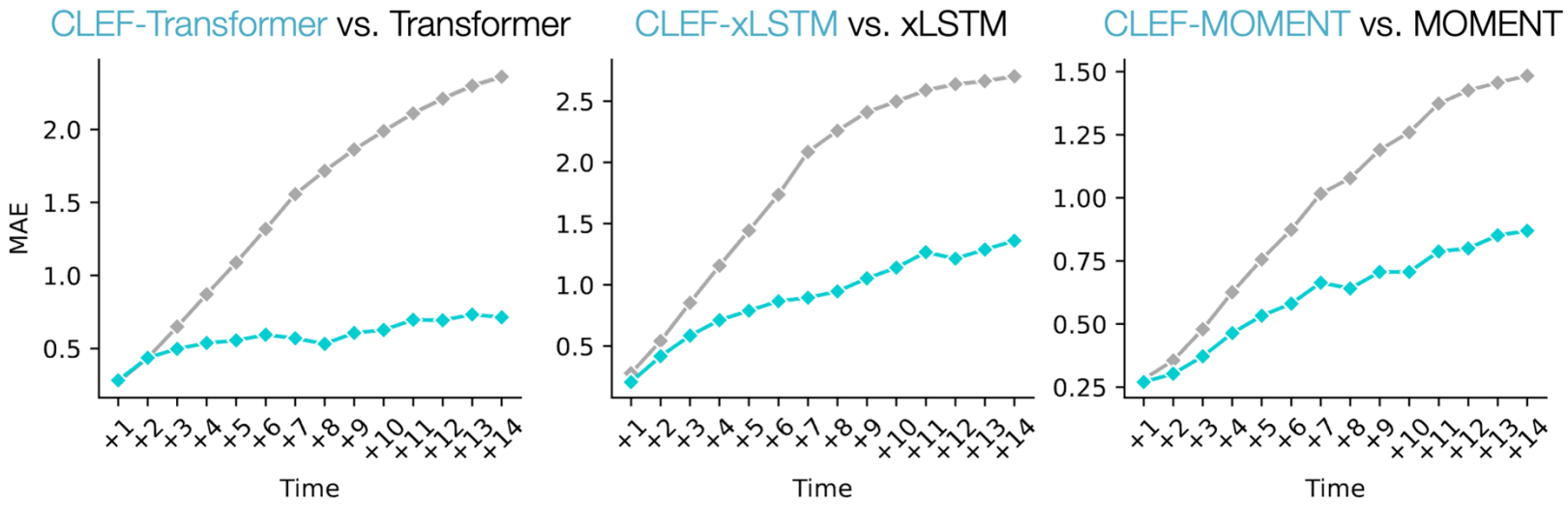
CLEF > **Baseline** with **high** time-varying confounding.

CLEF w/o BL yields the best performance, suggesting that CLEF's temporal concepts learn balanced representations for counterfactual outcomes estimation



Results: CLEF excels in counterfactual generation of cellular reprogramming experiments

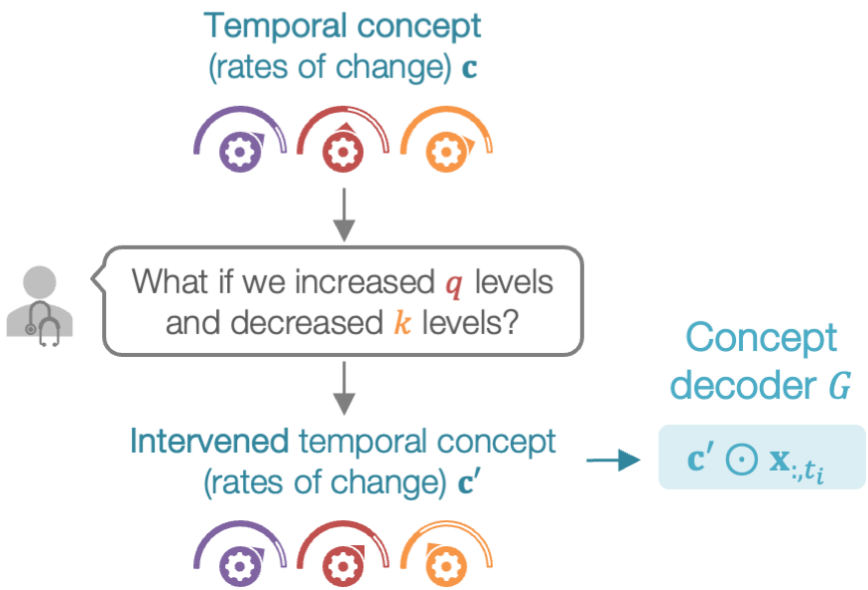




Number of time steps **after** performing the counterfactual action

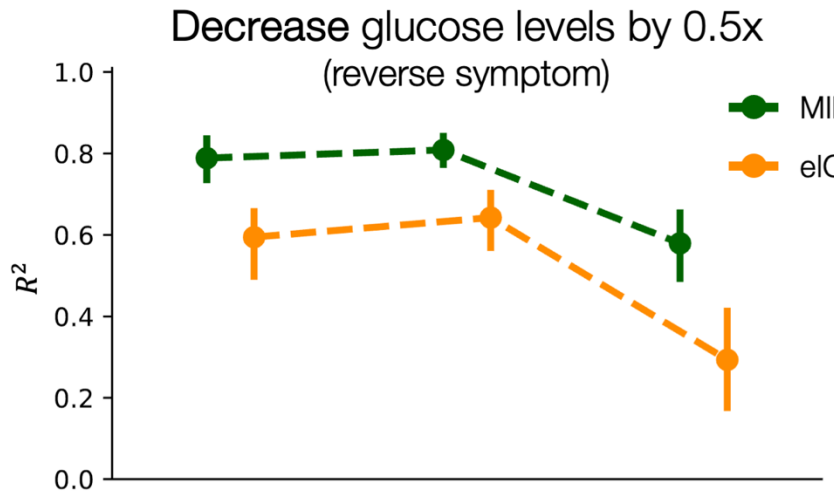
Results: Intervening on T1D patient trajectories to identify clinical interventions that improve patient health outcomes

Model capability: Generate counterfactual of $\hat{\mathbf{x}}^{s}_{:,t_j}$ by intervening on learned temporal concept $\mathbf{c}$

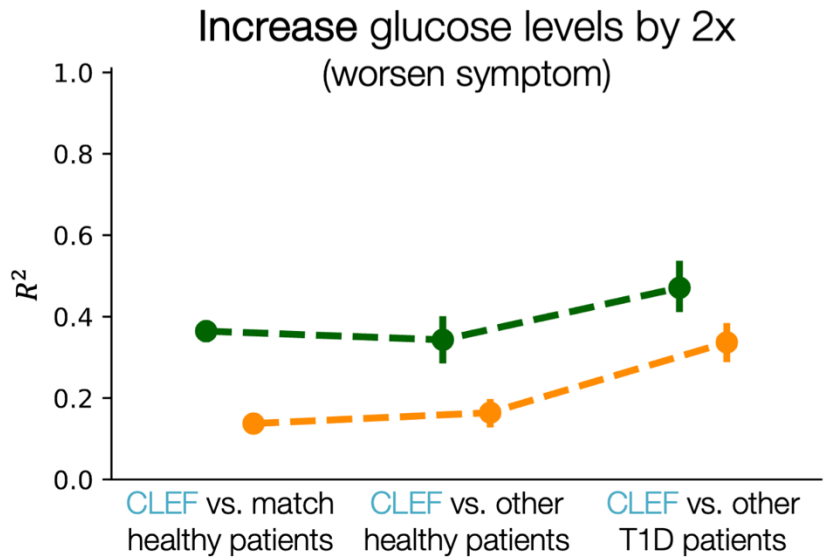


Concept decoder G

$\mathbf{c}' \odot \mathbf{x}_{:,t_i}$

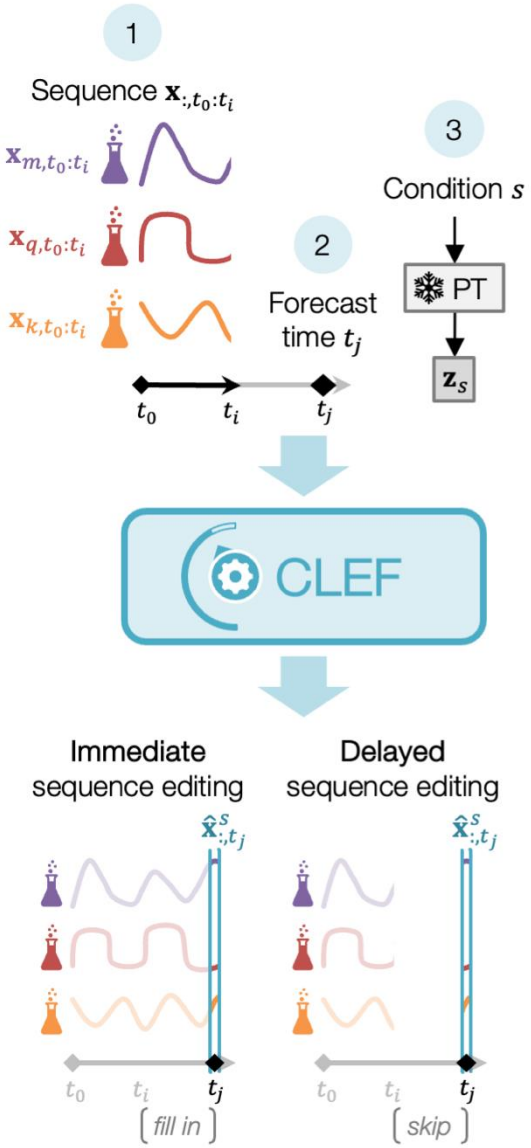


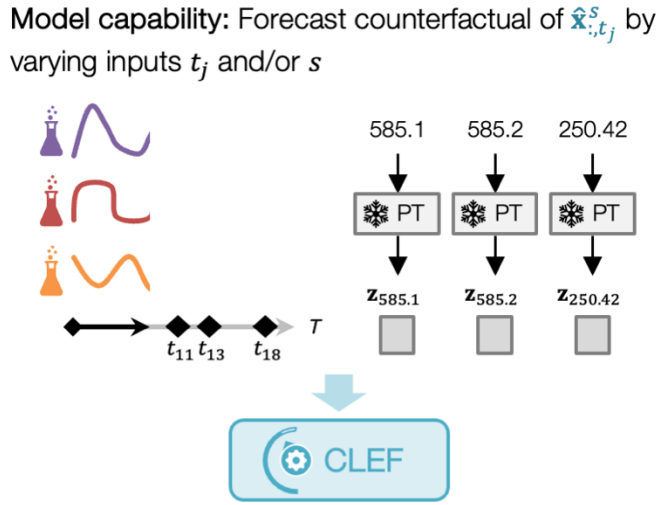
CLEF patients are **MORE SIMILAR** to **HEALTHY** patients than **OTHER T1D** patients

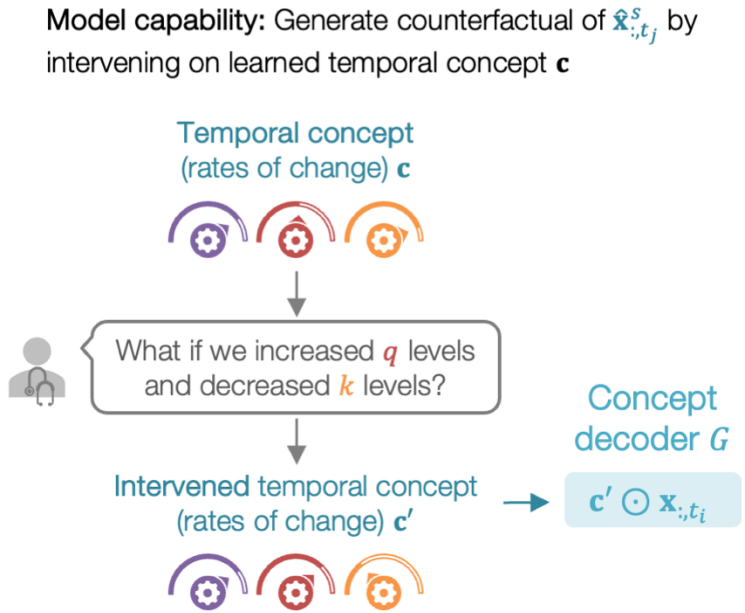


CLEF patients are **MORE SIMILAR** to **OTHER T1D** patients than **HEALTHY** patients

CLEF: Controllable sequence editing for conditional generation of cellular and patient trajectories







Paper

Contact Information
michellemli.com
michelleli@g.harvard.edu