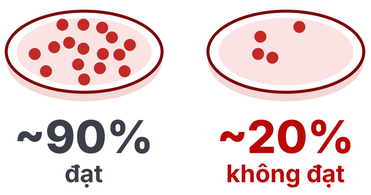
CUỘC THI SÁNG TẠO TẾ BÀO GỐC 2026
LẦN THỨ 11 | CHỦ ĐỀ: TẾ BÀO GỐC & AI
MĐT: 409
FORCELLS
ĐƠN VỊ TỔ CHỨC
ĐƠN VỊ TÀI TRỢ

FATEPOINT: MODEL AI PHÁT HIỆN SỚM SỰ PHÂN KỲ PHÁT TRIỂN TRONG BIỆT HÓA TẾ BÀO GỐC NGƯỜI THÀNH TẾ BÀO CƠ TIM

1 VẤN ĐỀ

Cùng một quy trình biệt hóa tế bào gốc đa năng người (hPSC) thành tế bào cơ tim, các lô có thể cho kết quả rất khác nhau: tỷ lệ tế bào **cTnT⁺** dao động từ khoảng **90%** đến khoảng **20%**. Chất lượng hiện chỉ được đánh giá ở cuối quy trình (**D16**), khi thời gian, hóa chất và nhân lực đã được sử dụng.



~90% đạt

~20% không đạt

Các giai đoạn biệt hóa hPSC → tế bào cơ tim

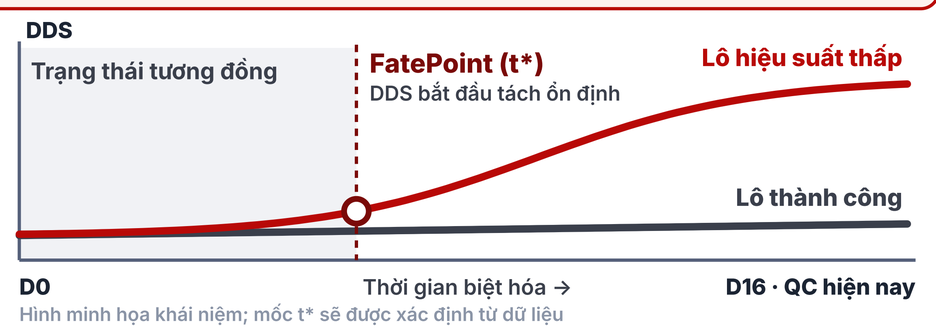
D0 hPSC › **D2/3** Trung bì › **D4/5** Trung bì tim › **D6/7** Tiền thân tim › **D16** Tế bào cơ tim

Sai lệch có thể khởi phát sớm, ở giai đoạn trung bì hoặc tiền thân tim, rồi được khuếch đại ở các giai đoạn sau; nhưng chỉ lộ ra ở D16.

2 Ý TƯỞNG

CÂU HỎI NGHIÊN CỨU

Một lô biệt hóa bắt đầu lệch khỏi quỹ đạo phát triển bình thường từ thời điểm nào?



Hình minh họa khái niệm; mốc t* sẽ được xác định từ dữ liệu

Giả thuyết: lô thành công và lô hiệu suất thấp ban đầu có trạng thái phát triển tương tự; sau đó lô hiệu suất thấp phân kỳ phân tử theo cách tái lập được, trước khi khác biệt lộ ra ở D16.

BỐN ĐẠI LƯỢNG CHÍNH CỦA FATEPOINT (sẽ xuất hiện lần lượt trong quy trình bên dưới)

τ Head 1 · vị trí phát triển
Mẫu đang ở đâu trên con đường từ hPSC đến tế bào cơ tim.

P(success) Head 2 · xác suất thành công
Khả năng lô thuộc nhóm hiệu suất cao, tính từ trạng thái phát triển.

DDS Developmental Divergence Score
Khoảng cách từ mẫu đến quỹ đạo của các lô thành công; càng lớn càng lệch.

FatePoint mốc t*
Mốc sớm nhất mà DDS của lô hiệu suất thấp tách khỏi lô thành công một cách ổn định.

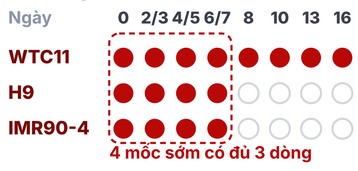
3 QUY TRÌNH FATEPOINT

1 Dữ liệu › 2 Gene program › 3 Học trạng thái › 4 Đo độ lệch › 5 Xác định FatePoint

Kết hợp hai bộ dữ liệu công khai bổ trợ nhau

Grancharova 2021 tham chiếu
- scRNA-seq, >20.000 tế bào, 55 mẫu
- Nhiều dòng hiPSC, D0–D90
- Mô tả phát triển tim **bình thường**

GSE314132 cao / thấp
- Bulk RNA-seq, 80 mẫu
- Nhóm cao/thấp theo cTnT⁺ ở D16
- So sánh lô cao và lô thấp **theo thời gian**

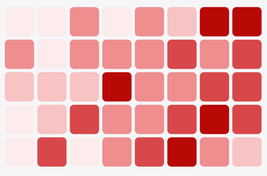


ĐẦU RA
Ma trận biểu hiện và metadata

Đưa hai loại dữ liệu về cùng một ngôn ngữ

scRNA-seq và bulk RNA-seq có phân bố đo khác nhau nên **không gộp trực tiếp**. Mỗi bộ được chuẩn hóa riêng, rồi quy về hoạt động của các **gene program** (nhóm gene cùng tham gia một quá trình sinh học).

~15.000–20.000 gene
▼
50–150 gene program



Wnt · MAPK · TGF-β/BMP · Trung bì · Biệt hóa tim · Glycolysis · Oxy hóa · Stress · Đa năng · Chu kỳ tế bào

Mỗi mẫu trở thành một vector 50–150 giá trị; mỗi giá trị là mức hoạt động của một chương trình sinh học.

ĐẦU RA
Hoạt động gene program của mỗi mẫu

Học xem mẫu đang ở trạng thái phát triển nào

Gene program (50–150)
▼
Encoder 256 → 128
▼
Trạng thái latent z (32–64)
▼
Head 1 · τ vị trí phát triển | **Head 2 · P(success)** xác suất thành công

τ được huấn luyện để tăng đúng thứ tự thời gian: τ(D0) < τ(D2) < τ(D4) < τ(D6).

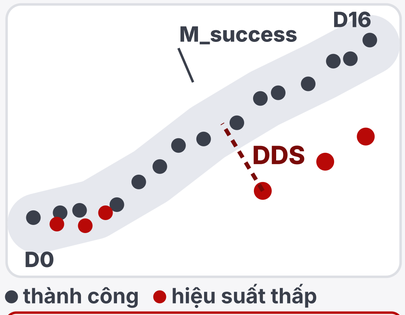
P(success | z) dự đoán lô thuộc nhóm cao hay thấp.

z là bản tóm tắt trạng thái phát triển của mẫu, dùng chung cho cả hai head. Mô hình gọn, ~10^5–10^6 tham số.

ĐẦU RA
z, τ và P(success) của mỗi mẫu

Đo mỗi mẫu lệch bao xa khỏi quỹ đạo của lô thành công

Từ các mẫu thành công/tham chiếu, dựng quỹ đạo phát triển bình thường **M_success** trong không gian z.



DDS = d(z, M_success)

DDS thấp: đi đúng quỹ đạo. DDS cao: đã lệch.

ĐẦU RA
DDS của mỗi mẫu tại mỗi mốc

Tìm mốc sớm nhất mà sự lệch trở nên rõ ràng và đáng tin

FatePoint t* là mốc sớm nhất thỏa đồng thời 4 tiêu chí:
1. DDS giữa lô thấp và lô cao khác biệt có ý nghĩa thống kê
2. Quan sát được ở ≥ 2 dòng tế bào
3. Duy trì ở các mốc sau
4. Đóng góp vào dự đoán kết quả D16

Kiểm chứng

Time truncation: chỉ dùng dữ liệu đến từng mốc để tìm mốc sớm nhất đủ thông tin.

D0 › D0–2/3 › D0–4/5 › D0–6/7

So sánh với: marker score phát triển; hồi quy logistic/cây quyết định trên gene program; mô hình bỏ hàm mất mát thời gian.

Chia tập theo lô biệt hóa: các giếng lặp trong cùng một lô không phải mẫu độc lập.

ĐẦU RA
Mốc FatePoint t* và cửa sổ cảnh báo sớm

4 KẾT QUẢ KỲ VỌNG

Học được tiến trình	τ tăng đúng thứ tự các giai đoạn phát triển
Phân kỳ sớm	DDS của lô hiệu suất thấp tăng sớm hơn lô thành công, cho phép xác định t*
Phân kỳ bền, nhiều dòng	Độ lệch duy trì ở các mốc sau và có mặt ở ≥ 2 dòng tế bào
Dự đoán kết quả	DDS và z sớm dự đoán cTnT⁺ ở D16 tốt hơn marker score

5 GIẢI THÍCH SINH HỌC

- Tại t*: phân tích biểu hiện khác biệt, module/pathway và SHAP hoặc Integrated Gradients.
- Đối chứng dương: **Wnt, MAPK, chuyển hóa glucose** (đã báo cáo trong GSE314132).
- Chương trình lệch sớm ngoài nhóm trên là hướng phát hiện mới.

Kịch bản kỳ vọng: D0 tương đồng → lệch nhẹ ở D2/3 → phân kỳ bền ở chuyển tiếp trung bì – tiền thân tim → khác biệt cTnT⁺ ở D16. Mốc và con đường cụ thể do dữ liệu quyết định.

6 ỨNG DỤNG

Cảnh báo sớm trong kiểm soát chất lượng hPSC-CM

Lấy mẫu sớm › Tính gene program › Ước lượng τ, DDS › Cảnh báo QC

DDS thấp
Tiếp tục quy trình chuẩn

DDS cao
Gắn cờ lô, rà soát quy trình, tăng giám sát

7 ĐIỂM MỚI

Nghiên cứu gốc GSE314132 đã có phân loại cao/thấp; FatePoint khác ở chỗ:

1. **Tham chiếu phát triển chuẩn:** dùng scRNA-seq của các lô thành công để xác định quỹ đạo phát triển tim bình thường.
2. **Mô hình hóa độ lệch quỹ đạo:** ước lượng τ và DDS thay vì chỉ phân loại thành công/thất bại.
3. **Định nghĩa FatePoint có tiêu chí:** ý nghĩa thống kê, bền vững, nhiều dòng tế bào, liên quan kết quả cuối.

Nguồn dữ liệu: Feeney AK et al., bioRxiv 2026 (GSE314132); Grancharova T et al., Scientific Reports 2021;11:15845.