STEM CELL INNOVATION CONTEST 2026

11TH EDITION | THEME: STEM CELLS & AI

ID: 415

VALTENE

ORGANIZER

SPONSORING ORGANIZATION

Application of Deep Learning for Early Prediction of Genetic Variants in ESCs

Introduction

Embryonic stem cells (ESCs) during prolonged expansion may develope recurrent genetic mutations (TP53, BCL2L1, trisomies) that drive clonal dominance and tumorigenic risks. Conventional quality control methods fail to detect low-level mosaicism due to morphological indistinguishability under standard microscopy. To address these challenges, our team proposes an innovative diagnostic framework combining big data predictive analytics and 3D machine vision to detect subtle morphological and kinetic cues before downstream clinical translation.

Aim

Develop an integrated deep learning approach that combines static image analysis, kinetic feature extraction, spatiotemporal video analysis to detect, monitor, predict changes in variant-bearing ESC proportions from bright-field and time-lapse images.

Background Knowledge

Clinical Risks

- **Culture Adaptation:** Prolonged ESCs expansion causes recurrent mutations (TP53, BCL2L1, Trisomies 1,12,17,20).
- **YAP-Driven Competition:** Variant cells mechanically compress wild-type cells, triggering apoptosis and achieving clonal dominance.
- **Diagnostic Gap:** Standard QC (G-banding, qPCR, NGS) misses low-level mosaicism (<10–20%); mutants remain morphologically invisible under light microscopy.
- **Clinical Hazard:** High risk of tumorigenesis and failed differentiation during transplantation.

AI Framework

- **Non-Invasive Sensing:** Deep Learning detects subtle kinetic and structural signals (division rate, stiffness, cell ejection) undetectable by eye.
- **Core Toolkit:** Cellpose, ResNet, 3DCNN, VideoMAE and Random Forest.

Objective

Reference dataset: Establish a reference dataset of genetically characterized WT and variant-bearing ESCs at defined proportions, with paired bright-field and fluorescence images.

Model development: Develop and combine static-image, kinetic-feature, spatiotemporal video models using independent samples.

Early detection: Evaluate detection across different variant proportions; determine the detection limit, sensitivity, specificity, false-positive rate, early-warning capacity.

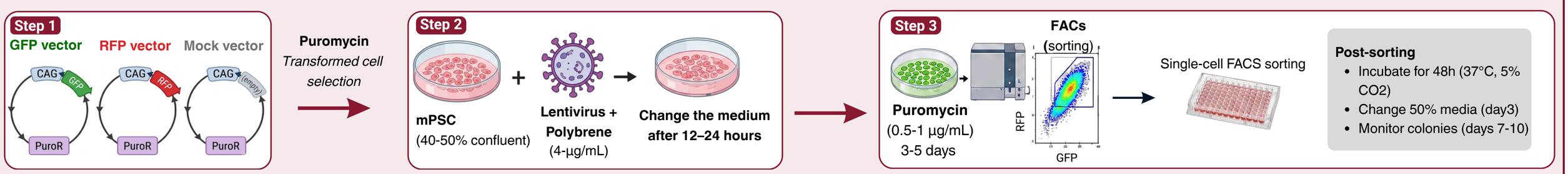
Longitudinal analysis: Track variant proportions over time; assess expansion trends, rates, short-term forecasts, accuracy, warning lead time.

Early-warning workflow: Set alert thresholds, integrate AI with time-lapse imaging and genetic confirmation, assess limitations in accuracy, reliability, culture conditions.

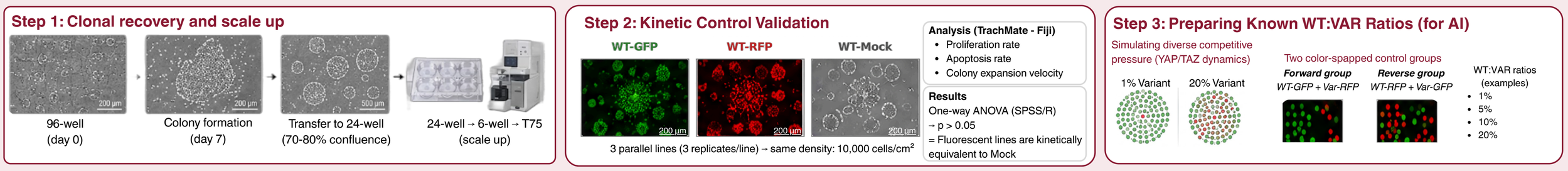
Experimental Design

Protocol for Generation and Preparation of Stable Fluorescently Labeled PSC lines

Stage 1: Lentiviral-based Fluorescent Gene Transfer

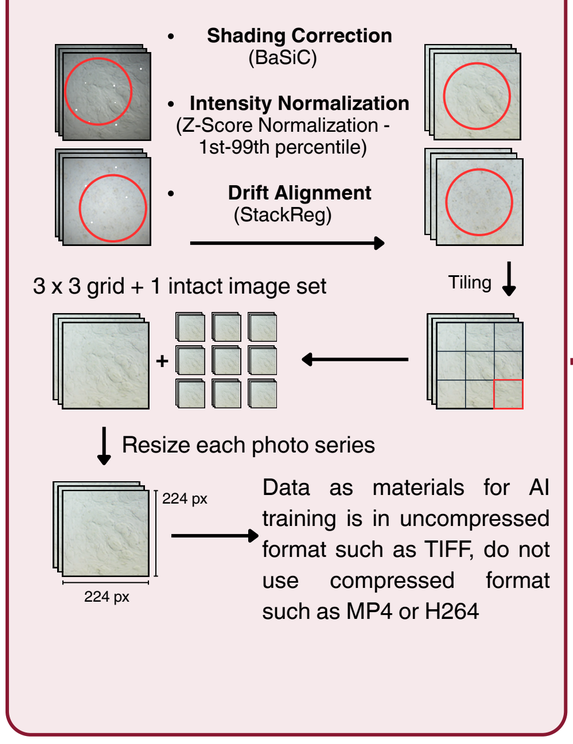


Stage 2: Selection and Kinetic Validation



Method

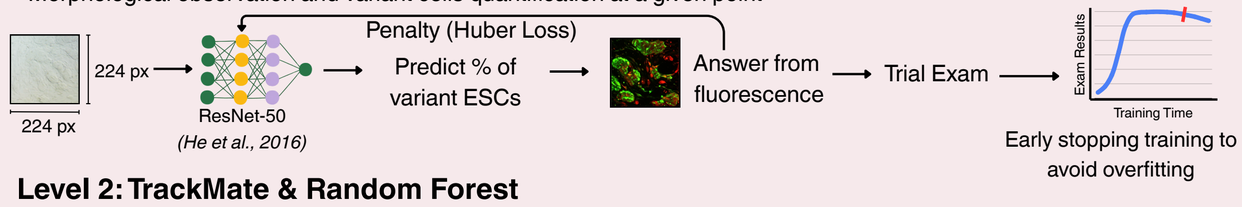
Pre-processing



AI Training

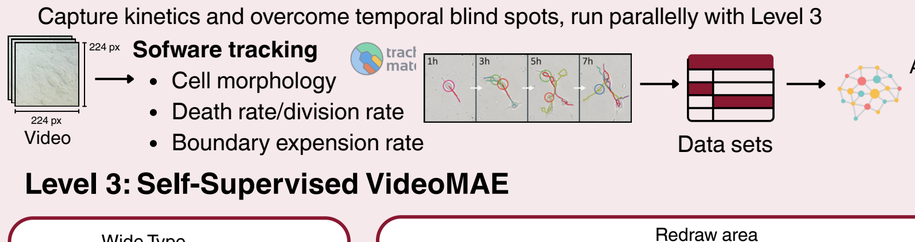
Level 1: Single-Frame ResNet

Morphological observation and variant cells quantification at a given point

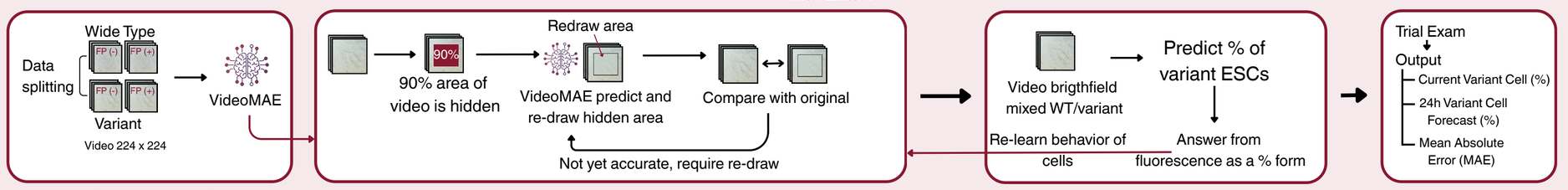


Level 2: TrackMate & Random Forest

Capture kinetics and overcome temporal blind spots, run parallelly with Level 3



Level 3: Self-Supervised VideoMAE



VideoMAE *(Tong et al., 2022)* Delivers top accuracy across all densities by learning spatiotemporal dynamics without cell segmentation

Dual-Routing Ensemble

- Input data is single-frame: Run only Level 1 (ResNet-50) (warning reducing reliability if confluence >= 60%) → **Result is 100% of Level 1**
- Input data is Time-lapse: Turn off Level 1, Run Ensemble Level 2 (TrackMate + RF) + Level 3 (VideoMAE) → **Result is x% of Level 2 + y% of Level 3 (see below)**

Confluence (C)	Density	Level 2	Level 3
<= 60%	< 1.0E5 cells/cm2	40%	60%
70%	1.2E5 cells/cm2	30%	70%
80%	1.4E5 cells/cm2	15%	85%
>= 85%	>= 1.5E5 cells/cm2	10%	90%

Expected result

Reference dataset: Establish a validated dataset of bright-field and time-lapse images with known WT and variant-cell proportions.

Quantitative estimation: Achieve reliable AI-based estimates of variant cell proportions, assessed by MAE and RMSE.

Early detection: Determine the lowest reliably detectable variant proportion via sensitivity, specificity, false-positive rate.

Temporal analysis: Assess variant expansion trends, rates, short-term forecasts, early-warning lead time.

AI-assisted workflow: Establish an alert and genetic confirmation workflow, assess its feasibility and limitations across culture conditions.

[1] Marzec-Schmidt, K., et al. (2023). Artificial Intelligence Supports Automated Characterization of Differentiated Human Pluripotent Stem Cells. Stem Cells, 41(9), 850–861

[2] He, K., Zhang, X., Ren, S. and Sun, J. (2016) Deep Residual Learning for Image Recognition. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, 27-30 June 2016, 770-778.

[3] Breiman, L. (2001). Random forests. Machine Learning, 45(1), 5–32.

[4] Tong, Z., Song, Y., Wang, J., & Wang, L. (2022). VIDEOMAE: Masked Autoencoders are Data-Efficient Learners for Self-Supervised Video Pre-Training. Neural Information Processing Systems, 10078–10093.

[5] Davis, H. E., Morgan, J. R., & Yarmush, M. L. (2004). Polybrene increases retrovirus gene transfer efficiency by enhancing receptor-independent virus adsorption on target cell membranes. Biophysical Chemistry, 97(2-3), 159-172.

Phone: 028 3636 1206 (nhánh 3)

Website: 2026.stemcellinnovation.com.vn

Email: sci@sci.edu.vn