CUỘC THI SÁNG TẠO TẾ BÀO GỐC 2026
LẦN THỨ 11 | CHỦ ĐỀ: TẾ BÀO GỐC & AI
MĐT: 506
NEOSTEM
ĐƠN VỊ TỔ CHỨC
ĐƠN VỊ TÀI TRỢ

HỆ THỐNG PHÂN TÍCH, ĐỐI CHIẾU TẾ BÀO GỐC TẠO MÁU DỰA TRÊN MÔ HÌNH MACHINE LEARNING SUY LUẬN ĐA ĐỘ PHÂN GIẢI HLA VÀ THUẬT TOÁN GHÉP CẶP

I. Bối cảnh

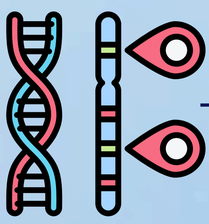
Việc tìm kiếm tế bào gốc phù hợp gặp rào cản lớn do tỷ lệ hòa hợp HLA ngoài huyết thống cực thấp, đặc biệt tại Việt Nam với độ đa dạng di truyền cao. Việc lưu trữ dữ liệu y tế phân mảnh khiến quá trình đối chiếu kéo dài nhiều tuần, làm mất đi "thời gian vàng" cứu sống bệnh nhân. Do đó, xây dựng hệ thống phân tích, đối chiếu dựa trên mô hình machine learning và thuật toán ghép cặp là giải pháp thiết yếu giúp truy xuất tức thời, rút ngắn thời gian chờ và tăng cơ hội sống sót.

II. Mục tiêu nghiên cứu

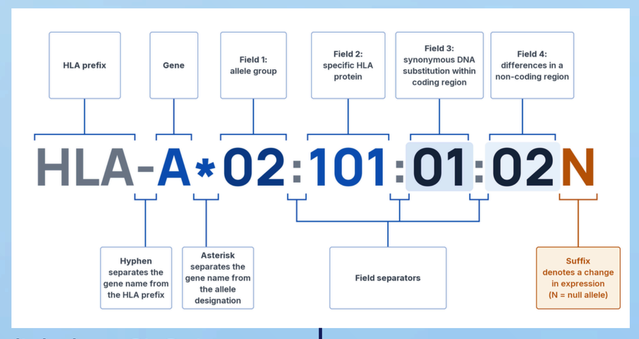
Thiết kế và xây dựng hệ thống phân tích, đối chiếu hàng triệu tế bào gốc được hiến tặng dựa trên mô hình machine learning suy luận đa độ phân giải HLA (HLA-GRAPHNET) và thuật toán ghép cặp (EPLET MATCHING) từ đó tìm ra tế bào gốc phù hợp để hiến trong thời gian ngắn góp phần tăng tỉ lệ sống sót cho bệnh nhân.

III. Phương pháp nghiên cứu

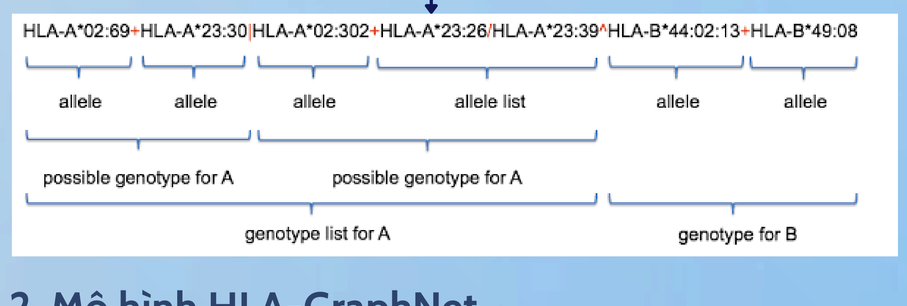
1. Thu thập, chuẩn hóa dữ liệu



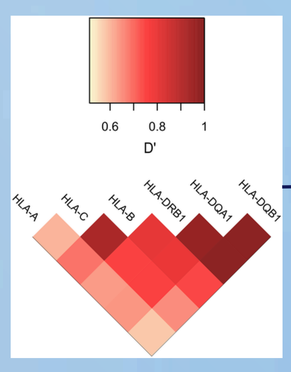
Ánh xạ alen về danh pháp chuẩn IPD-IMGT/HLA

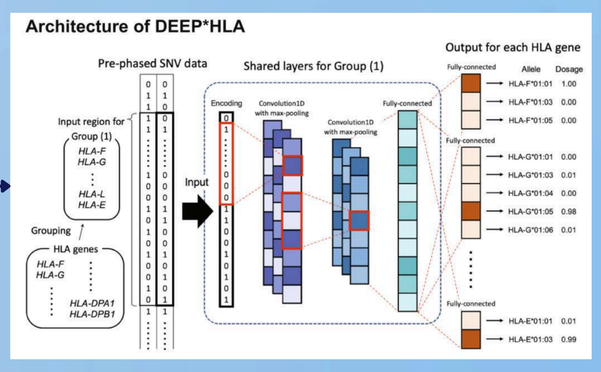


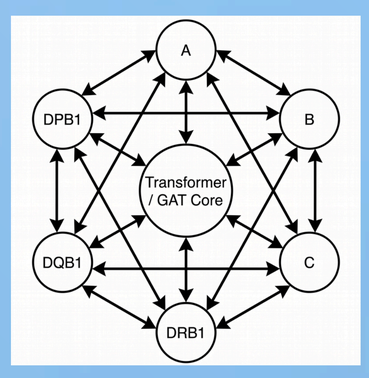
Chuyển sang định dạng GL String



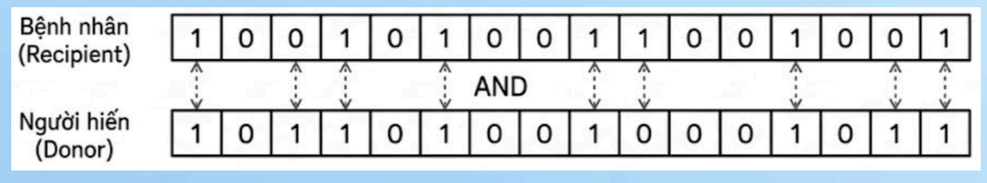
2. Mô hình HLA-GraphNet





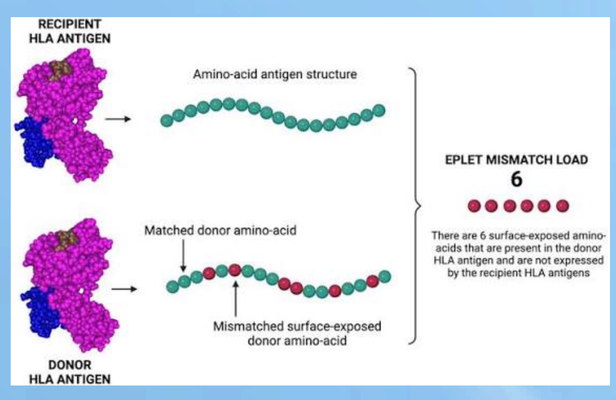


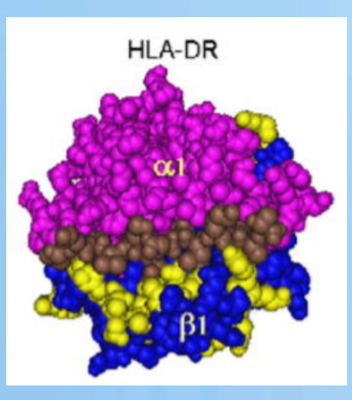
3. Thuật toán Chỉ mục Bitmask Nhị phân



Mã hóa 242+ alen phổ biến tại Việt Nam thành Candidate Bitmask từ tập alen có xác suất cao.
Bitwise AND đối chiếu hàng trăm vị trí kháng nguyên và trả kết quả dương tính khi cả tế bào người hiến lẫn người nhận đều biểu hiện chung một alen.
→ Top 1000 ứng viên có số locus khớp cao nhất từ hàng triệu dữ liệu.

4. Thuật toán Eplet Matching và MCDA Ranking





Thống kê di truyền

Ma trận Linkage Disequilibrium định lượng các bộ gene đặc trưng và cung cấp dữ liệu đầu vào cho DEEP*HLA và mô hình.

Học sâu Đồ thị & Transformer

DEEP*HLA mô phỏng các tương tác di truyền và khoảng cách giữa các điểm đa hình đơn nucleotide và các locus (A, B, C, DRB1, DQB1, DPB1) trên cùng một nhiễm sắc thể. Nhờ đó, giúp điền khuyết chính xác các alen độ phân giải cao (4-digit) hoặc các locus bị thiếu (như DQB1, DPB1).

Eplet Matching: Đối chiếu cấu trúc 3D kháng nguyên HLA, đếm số lượng Eplet bất tương đồng trên rãnh gắn peptide và giữ lại các mẫu Permissive mismatch và loại bỏ Non-permissive mismatch.
MCDA Ranking: Tính điểm an toàn miễn dịch (Eplet) với các tiêu chí (như ngưỡng tế bào gốc CD34+, CMV, ABO,...) vào Hàm điểm lâm sàng tổng hợp (CTS) để chọn Top 3-5 mẫu tốt nhất.

IV. Kết quả dự kiến

- Dự kiến đạt độ chính xác cao khi suy luận từ độ phân giải 2-digit lên 4-digit trên tập kiểm thử dữ liệu người Việt.
- Nhận diện các cặp lệch alen nhưng tương đồng về mặt cấu trúc Eplet và tối ưu hóa ghép máu dây rốn.
- Chỉ định Top 3–5 mẫu có xác suất khớp thực tế cao nhất để xét nghiệm NGS kiểm chứng, loại bỏ việc xét nghiệm tràn lan.

V. Ứng dụng

- Nhanh chóng tìm ra tế bào gốc phù hợp nhất cho từng bệnh nhân, góp phần tăng tỉ lệ sống sót.
- Chủ động lưu trữ các mẫu tế bào gốc còn thiếu và hiếm gặp.
- Tiết kiệm chi phí xét nghiệm NGS không cần thiết và cung cấp bản đồ Haplotype giúp các ngân hàng tế bào gốc tối ưu hóa việc lưu trữ.

VI. Tài liệu tham thảo

[1] Tran N. Q., et al., "Allele and Haplotype Frequencies... in a Kinh Vietnamese Population," Front. Immunol., 13:875283, 2022.
[2] Naito T., et al., "A deep learning method for HLA imputation...," Nat. Commun., 12:1639, 2021.
[3] Choudhery M. S., et al., "Applications of artificial intelligence in stem cell therapy," World J. Stem Cells, 17(8):106086, 2025.
[4] Steiner D., "Computer Algorithms in the Search for Unrelated Stem Cell Donors," Bone Marrow Res., 2012:175419, 2012.

Phone: 028 3636 1206 (nhánh 3)
Website: 2026.stemcellinnovation.com.vn
Email: sci@sci.edu.vn