



(12) **BẢN MÔ TẢ GIẢI PHÁP HỮU ÍCH THUỘC BẰNG ĐỘC QUYỀN
GIẢI PHÁP HỮU ÍCH**

(19) **Cộng hòa xã hội chủ nghĩa Việt Nam (VN)
CỤC SỞ HỮU TRÍ TUỆ**

(11)



2-0004412

(51) **C12Q 1/68; G06N 7/08; G06N 3/00**
2022.01

(13) **Y**

(21) 2-2023-00773

(22) 07/12/2023

(45) 27/10/2025 451

(43) 26/02/2024 431

(73) Trường Đại học Y Hà Nội (VN)

Số 1, Tôn Thất Tùng, Quận Đống Đa, Thành Phố Hà Nội

(72) Nguyễn Thị Trang (VN); Đào Thị Huyền Trang (VN); Lê Tú Linh (VN); Nguyễn
Viết Nhung (VN); Nguyễn Hữu Tú (VN); Lê Văn Quảng (VN); Đinh Văn Lượng
(VN); Trịnh Lê Huy (VN); Lê Thị Minh Phương (VN); Nguyễn Ngọc Sơn (VN);
Nguyễn Xuân Hậu (VN); Lưu Hoàng Bách (VN); Vũ Thị Huyền (VN); Tô Thị Thu
Hà (VN); Trần Quỳnh Trang (VN); Nguyễn Thị Minh Ngọc (VN).

(54) **PHƯƠNG PHÁP ĐÁNH GIÁ YẾU TỐ ĐỘT BIẾN DI TRUYỀN ĐẾN ĐÁP ỨNG
THUỐC ĐÍCH ĐIỀU TRỊ UNG THƯ PHỔI KHÔNG TẾ BÀO NHỎ VÀ KIT
DÙNG CHO PHƯƠNG PHÁP NÀY**

(21) 2-2023-00773

(57) Sáng chế đề cập đến phương pháp đánh giá yếu tố di truyền gây phát sinh ung thư phổi không tế bào nhỏ đến đáp ứng thuốc điều trị đích. Phương pháp theo sáng chế dựa trên việc thiết lập cơ sở dữ liệu di truyền chuẩn với 43 gen gây phát sinh ung thư phổi không tế bào nhỏ, thiết lập cơ sở dữ liệu so sánh gồm các dữ liệu liên quan đến mô hình phát sinh bệnh, khởi tạo dữ liệu đánh giá bao gồm các chỉ số lâm sàng, cận lâm sàng và giải trình tự ít nhất hai gen để tạo tập hợp dữ liệu đánh giá. Dựa trên việc so sánh phân tích máy học song song từng chỉ số dựa trên mô hình XGBoost để lọc và đưa ra dự báo mô hình đáp ứng thuốc điều trị đích. Dữ liệu này được cập nhật cho phép máy học liên tục để phát triển mô hình đánh giá dựa trên trí tuệ nhân tạo.

Lĩnh vực kỹ thuật được đề cập

Sáng chế thuộc lĩnh vực công nghệ sinh học và tin sinh ứng dụng trong thiết lập cơ sở dữ liệu nhằm đánh giá các yếu tố đột biến di truyền ảnh hưởng đến việc đáp ứng thuốc đích để sàng lọc và cá nhân hóa dữ liệu di truyền để hỗ trợ cho việc điều trị bệnh ung thư phổi không tế bào nhỏ. Cụ thể là sáng chế đề xuất phương pháp đánh giá yếu tố đột biến di truyền đến đáp ứng thuốc đích điều trị ung thư phổi không tế bào nhỏ kết hợp yếu tố di truyền và phân tích cận lâm sàng ứng dụng trí tuệ nhân tạo.

Tình trạng kỹ thuật của sáng chế

Theo thống kê của Globocan, bệnh ung thư phổi là căn bệnh có tỷ lệ tử vong lớn nhất, khoảng 2,2 triệu ca mắc mới và 1,8 triệu ca tử vong. Tỷ lệ người chẩn đoán bị mắc ung thư phổi chiếm khoảng 11,4% với tỷ lệ tử vong khoảng 18%. Trong đó, tỷ lệ nam giới mắc và tử vong cao hơn nữ giới. Đối với nữ, bệnh ung thư phổi đứng thứ ba về số ca mắc (sau ung thư vú và ung thư trực tràng), và thứ hai về số ca tử vong sau ung thư vú. Tỷ lệ mắc và tử vong vì ung thư phổi ở các nước phát triển gấp 3 đến 4 lần các nước đang phát triển (WHO, 2018).

Ung thư phổi được chia thành 2 nhóm bệnh chính là ung thư phổi tế bào nhỏ chiếm khoảng 15% và ung thư phổi không tế bào nhỏ chiếm 85%. Hai nhóm bệnh này có tiên lượng và phương pháp điều trị tương đối khác nhau. Ung thư phổi không tế bào nhỏ có tiên lượng tốt hơn và có nhiều lựa chọn điều trị hơn ung thư phổi tế bào nhỏ.

Trong thập kỷ qua, nhiều đột biến gen có liên quan đến ung thư phổi tế bào nhỏ đã được phát hiện, như đột biến trên các gen AKT2, ALK, BRAF, DDR2, EGFR, FGFR1, HER2, KRAS, MEK1, MET, NRAS, NTRK1, PIK3CA, PTEN, RET và ROS. Trong số các đột biến gen này một số đột biến đã được nghiên cứu kỹ ở mức độ gen (trật tự các nucleotit) và ở mức độ protein (trình tự các axit amin trong chuỗi polypeptit của protein). Các đột biến này ảnh hưởng nhất định tới đáp ứng thuốc, trong đó một số đột biến làm giảm độ nhạy đối với điều trị đích, nhưng một số đột biến lại làm giảm cũng như chưa rõ ý nghĩa lâm sàng hoặc vẫn đang được thử nghiệm lâm sàng.

Đối với mỗi đột biến sẽ có thuốc đích khác nhau tương ứng với đột biến gen đó, một số đột biến đang được nghiên cứu phát triển thuốc đặc trị đặc hiệu với từng đột biến. Hiện có hai nhóm đột biến gen được áp dụng thuốc điều trị rộng rãi nhất là đột biến gen EGFR và gen ALK.

Trong những năm trở lại đây, cùng với phát triển công nghệ sinh học phân tử và các nghiên cứu phát hiện các loại đột biến gen quan trọng có ảnh hưởng đến ung thư phổi đã được nghiên cứu sâu rộng. Bằng việc nghiên cứu các dấu ấn phân tử này, các nhà khoa học đã tìm đưa ra các phác đồ đáp ứng đối với các dạng đột biến có khả năng ức chế hoạt động của các đột biến gen đó, ngăn không cho tế bào ung thư phát triển, thúc đẩy quá trình chết tự nhiên của tế bào, đồng thời hạn chế tối thiểu tác động lên các tế bào lành của cơ thể, giảm thiểu tác dụng phụ. Các thuốc hướng đích phân tử hay là các điều trị hướng đích (targeted therapy) ảnh hưởng rất nhiều vào các yếu tố, từ các đột biến gen đến các yếu tố lâm sàng, cận lâm sàng và có mối quan hệ rất phức tạp.

Trong quá trình điều trị bệnh, các tế bào ung thư cũng luôn xảy ra các đột biến gọi là các đột biến thứ cấp nên tình trạng đáp ứng thuốc luôn thay đổi. Quá trình này xảy ra đồng thời với việc phát triển các đột biến gen mới hoặc sự xuất hiện đồng thời các đột biến gen mà ảnh hưởng tới hiệu quả của thuốc hướng đích. Điều này tạo ra thách thức trong việc theo dõi, xây dựng phác đồ, đề xuất thuốc điều trị hướng đích hiệu quả để có khả năng đáp ứng tốt và điều trị ung thư phổi hiệu quả theo diễn tiến của bệnh. Một số bệnh nhân có thể xuất hiện các đột biến kháng thuốc hướng đích hoặc kháng lại các liệu pháp hóa trị liệu ngay trong quá trình điều trị thông thường. Đây là một thách thức lớn đối với việc đánh giá chính xác các thuốc điều trị hướng đích liên quan đến các đột biến gây bệnh nhân ung thư phổi không tế bào nhỏ.

Đã có những ứng dụng trí tuệ nhân tạo với mô hình máy học để ứng dụng trong việc điều chỉnh, dựa trên các các chỉ số di truyền, chỉ số lâm sàng, chỉ số cận lâm sàng để đánh giá, dự báo khả năng đáp ứng thuốc điều trị bệnh.

Phương pháp XGBoost (Extreme Gradient Boosting) là phương pháp học máy tối ưu và hiệu quả nhất trong lĩnh vực dự đoán và phân loại nói chung. Phương pháp này được phát triển bởi Chen Tianqi vào năm 2016, XGBoost kết hợp tốt giữa sức mạnh của phương pháp Gradient Boosting và phần cứng để đạt được hiệu suất vượt trội cả về thời gian huấn luyện và bộ nhớ sử dụng. XGBoost được thiết kế để hỗ trợ tính toán song

song trên nhiều lõi CPU, giúp tăng tốc quá trình huấn luyện và dự đoán trên dữ liệu lớn. Điều này đảm bảo XGBoost có thể xử lý hàng tỷ ví dụ trong cài đặt phân tán, cho phép xử lý hiệu quả các tập dữ liệu lớn và phức tạp. XGBoost còn có khả năng xử lý hiệu quả các dữ liệu thừa thớt, như dữ liệu có các giá trị thiếu hoặc các giá trị 0 quá thường xuyên. XGBoost sử dụng kỹ thuật tập hợp, cho phép bổ sung đệ quy các mô hình mới cho đến khi không còn cải thiện đáng kể hiệu suất của các mô hình hiện có. Nhờ vào những cải tiến và tính năng mạnh mẽ này, XGBoost đã trở thành một trong những phương pháp phổ biến và được ưa chuộng trong cả cộng đồng học máy và khoa học dữ liệu.

Trong y học, XGBoost có nhiều ứng dụng quan trọng và hiệu quả có thể kể đến như là dự đoán và phân loại bệnh, phân tích dữ liệu di truyền, dự đoán phản ứng thuốc. Đặc biệt Xgboost thể hiện hiệu quả cao trong dự đoán tình trạng bệnh, ví dụ như dự đoán tổn thương cơ tim, đánh giá nguy cơ mắc bệnh tim mạch, và dự đoán biến chứng sau phẫu thuật và phân tích dữ liệu lâm sàng để xác định các yếu tố quan trọng ảnh hưởng đến kết quả điều trị và lợi ích cho bệnh nhân.

Phương pháp đa chiều giảm thứ nguyên đầu tiên được sử dụng là mô hình của MD Ritchie và cộng sự năm 2001. Nhóm tác giả đề xuất phương pháp để giảm tính đa chiều của thông tin đa điểm nhằm phát hiện các tương tác bậc cao giữa các gen chuyển hóa estrogen trong ung thư vú không có tính chất di truyền. Từ đó đến nay, phương pháp này đã được sử dụng để phân tích tương tác của các yếu tố gen - gen và gen - môi trường trong sàng lọc các bệnh đa gen - đa nhân tố.

Do đó, vẫn cần có phương pháp đánh giá yếu tố đột biến di truyền đến đáp ứng thuốc đích điều trị ung thư phổi không tế bào nhỏ ứng dụng công nghệ sinh học và tin sinh học ứng dụng dựa trên việc ứng dụng công cụ máy học và các dữ liệu trong phân tích di truyền cũng như các yếu tố lâm sàng, cận lâm sàng để đánh giá thuốc nhằm xác định hiệu quả đáp ứng thuốc đích ứng dụng trong phân tích, sàng lọc và điều chỉnh thuốc đích.

Bản chất kỹ thuật của sáng chế

Sáng chế nhằm giải quyết vấn đề trên, cụ thể là dựa trên việc phân tích mô hình máy học kết hợp với dữ liệu và huấn luyện dữ liệu, sáng chế đề xuất phương pháp đánh giá yếu tố đột biến di truyền đến đáp ứng thuốc đích điều trị ung thư phổi không tế bào nhỏ. Trong đó sử dụng dữ liệu lâm sàng và cận lâm sàng kết hợp với xét nghiệm giải

trình tự gen đối với các nhóm đột biến, giúp đánh giá khả năng đáp ứng thuốc đích ứng dụng trong điều trị bệnh ung thư phổi không tế bào nhỏ. Phương pháp theo sáng chế cho phép đánh giá tổng thể các yếu tố tác động đến việc đáp ứng thuốc và dự báo các phát sinh dựa trên mô hình máy học. Ngoài ra sáng chế cũng đề cập đến kit sử dụng trong phương pháp này.

Theo đó, sáng chế đề cập phương pháp đánh giá yếu tố đột biến di truyền đến đáp ứng thuốc đích điều trị ung thư phổi không tế bào nhỏ và kit để giải trình tự xác định đột biến điểm để xác định đồng thời 43 gen gây phát sinh ung thư phổi tế bào không nhỏ dùng trong phương pháp theo sáng chế.

Theo mục đích thứ nhất, sáng chế đề cập đến phương pháp đánh giá yếu tố di truyền gây phát sinh ung thư phổi không tế bào nhỏ đến đáp ứng thuốc điều trị đích, trong đó phương pháp này bao gồm các bước:

a) thiết lập cơ sở dữ liệu di truyền chuẩn đối với các đột biến gen gây phát sinh ung thư phổi không tế bào nhỏ, trong đó tạo tập hợp dữ liệu trình tự gen của 43 gen liên quan bao gồm EGFR, KRAS, ALK, STK11, FGFR1, PIK3CA, ARAF, NRG1, RRM1, HAVCR2, MET, CDK4, BRCA1, CCND1, PDGFRA, AXL, AURKA, TYMS, ETS2, DDR2, ERCC2, ABCB1, TP53, NFE2L2, PDCD4, RET, BRAF, ROS1, NTRK1, XRCC1, CD274, RICTOR, AREG, BCL2L11, BIRC5, ERCC1, GNAS, AKT2, ERBB2, POLD1, RB1, SMARCA4 và WEE1, trong đó dữ liệu của từng gen được so sánh toàn cục để xác định và đánh dấu các vị trí đột biến điểm gây ung thư đã biết;

b) thiết lập cơ sở dữ liệu so sánh với ít nhất mỗi đột biến điểm gây ung thư đã biết được thu thập dữ liệu với các thông số phát sinh bệnh liên quan đến bệnh ung thư phổi không tế bào nhỏ bao gồm các chỉ số lâm sàng, cận lâm sàng, thuốc đích và dữ liệu trình tự gen đột biến trong số các gen EGFR, KRAS, ALK, STK11, FGFR1, PIK3CA, ARAF, NRG1, RRM1, HAVCR2, MET, CDK4, BRCA1, CCND1, PDGFRA, AXL, AURKA, TYMS, ETS2, DDR2, ERCC2, ABCB1, TP53, NFE2L2, PDCD4, RET, BRAF, ROS1, NTRK1, XRCC1, CD274, RICTOR, AREG, BCL2L11, BIRC5, ERCC1, GNAS, AKT2, ERBB2, POLD1, RB1, SMARCA4 và WEE1;

c) khởi tạo dữ liệu đánh giá, trong đó thu thập các chỉ số lâm sàng, cận lâm sàng ứng với mô hình cần đánh giá, đồng thời giải trình tự các đoạn gen chứa đột biến trên gen EGFR, KRAS, ALK, STK11, FGFR1, PIK3CA, ARAF, NRG1, RRM1, HAVCR2,

MET, CDK4, BRCA1, CCND1, PDGFRA, AXL, AURKA, TYMS, ETS2, DDR2, ERCC2, ABCB1, TP53, NFE2L2, PDCD4, RET, BRAF, ROS1, NTRK1, XRCC1, CD274, RICTOR, AREG, BCL2L11, BIRC5, ERCC1, GNAS, AKT2, ERBB2, POLD1, RB1, SMARCA4 và WEE1 của mô hình để tạo tập hợp dữ liệu đánh giá;

d) thiết lập mô hình phân tích máy học dựa trên các chỉ số di truyền và chỉ số lâm sàng, cận lâm sàng bằng cách tiến hành so sánh song song đồng thời từng chỉ số trong tập hợp dữ liệu đánh giá với từng chỉ số trong cơ sở dữ liệu so sánh với các thông số liên quan đến mô hình phát sinh bệnh và so sánh toàn bộ các gen EGFR, KRAS, ALK, STK11, FGFR1, PIK3CA, ARAF, NRG1, RRM1, HAVCR2, MET, CDK4, BRCA1, CCND1, PDGFRA, AXL, AURKA, TYMS, ETS2, DDR2, ERCC2, ABCB1, TP53, NFE2L2, PDCD4, RET, BRAF, ROS1, NTRK1, XRCC1, CD274, RICTOR, AREG, BCL2L11, BIRC5, ERCC1, GNAS, AKT2, ERBB2, POLD1, RB1, SMARCA4 và WEE1 được giải với cơ sở dữ liệu di truyền chuẩn theo mô hình XGBoost tăng cường với truy vấn ngẫu nhiên (random search) và đánh giá chéo (cross-validation), số trường quyết định độc lập (n_estimator) là 100, độ sâu trường (max_depth) là 13, tỷ lệ học (learning_rate) là 0,3 từ đó thu được các thông số đột biến điểm của gen liên quan đến thuốc điều trị đích phù hợp nhất có trong cơ sở dữ liệu mô hình phát sinh bệnh đồng thời nạp dữ liệu này vào dữ liệu so sánh; và

e) hiệu chỉnh dữ liệu so sánh bằng cách khởi tạo lại dữ liệu đánh giá theo một khoảng thời gian nhất định và so sánh với dữ liệu được nạp trước đó theo mô hình máy học đồng thời tiến hành phân tích máy học để hiệu chỉnh các yếu tố di truyền gây phát sinh ung thư phổi không tế bào nhỏ theo mô hình XGBoost theo thư viện SHAP (SHapley Additive exPlanations) và đột biến tương tác gen, trong đó yếu tố gây phát sinh ung thư phổi không tế bào nhỏ được xác định theo thứ tự giảm dần theo gồm đột biến phát sinh trên gen EGFR, ALK, các chỉ số lâm sàng, đột biến phát sinh trên gen KRAS, ROS, FGFR1, AKT1, HER2, MEK, RET, chỉ số cận lâm sàng, BRAF, DDR2, MTRK1, MEK1, MET, NRAS, PTEN, PIK3CA, đồng thời đánh dấu bổ sung vị trí đột biến điểm phát sinh vào cơ sở dữ liệu chuẩn.

Theo mục đích thứ hai, sáng chế đề cập đến kit dùng trong phương pháp đánh giá yếu tố di truyền gây phát sinh ung thư phổi tế bào không nhỏ nhằm giải trình tự xác định đồng thời 43 gen bao gồm EGFR, KRAS, ALK, STK11, FGFR1, PIK3CA, ARAF,

NRG1, RRM1, HAVCR2, MET, CDK4, BRCA1, CCND1, PDGFRA, AXL, AURKA, TYMS, ETS2, DDR2, ERCC2, ABCB1, TP53, NFE2L2, PDCD4, RET, BRAF, ROS1, NTRK1, XRCC1, CD274, RICTOR, AREG, BCL2L11, BIRC5, ERCC1, GNAS, AKT2, ERBB2, POLD1, RB1, SMARCA4 và WEE1, trong đó kit này bao gồm các trình tự nêu trong SEQ ID NO.1 đến SEQ ID NO.942.

Mô tả vắn tắt các hình vẽ

Hình 1 là biểu đồ thống kê giá trị Shapley thể hiện mức độ đóng góp của từng yếu tố gen đến hiệu quả đáp ứng thuốc đích đối với ung thư phổi không tế bào nhỏ.

Hình 2 là hình minh họa mức độ ảnh hưởng của yếu tố đối với từng đầu ra riêng lẻ đến hiệu quả đáp ứng thuốc đích đối với ung thư phổi không tế bào nhỏ.

Hình 3 là ma trận tổng hợp dữ liệu đầu vào và đầu ra của mô hình MDR.

Hình 4 là sơ đồ minh họa mối tương tác giữa đột biến gen và các yếu tố tiền lâm sàng đến hiệu quả đáp ứng thuốc đích đối với ung thư phổi không tế bào nhỏ.

Hình 5 là hình vẽ minh họa khoảng cách tương tác giữa các gen và yếu tố tiền lâm sàng đến hiệu quả đáp ứng thuốc đích đối với ung thư phổi không tế bào nhỏ.

Hình 6 là hình vẽ thể hiện sơ đồ phân tích máy học dựa trên các kết quả phân tích đột biến gen với các yếu tố lâm sàng và cận lâm sàng dự đoán hiệu quả đáp ứng thuốc đích trên mô hình thử nghiệm.

Hình 7 là kết quả đánh giá hình ảnh CT phổi trước và sau khi điều trị dựa trên đánh giá thuốc đích được đề xuất bởi phương pháp đánh giá theo sáng chế. (A) hình ảnh CT phổi trước khi điều trị; (B) hình ảnh CT phổi sau 3 tháng điều trị dựa trên mô hình phân tích, đánh giá thuốc hướng đích bằng phương pháp theo sáng chế.

Hình 8 là hình ảnh chụp MRI sọ não trước và sau khi điều trị dựa trên đánh giá thuốc đích được đề xuất bởi phương pháp đánh giá theo sáng chế. (A) hình ảnh MRI sọ não trước khi điều trị; (B) hình ảnh MRI sọ não sau 3 tháng điều trị dựa trên mô hình phân tích, đánh giá thuốc hướng đích bằng phương pháp theo sáng chế.

Hình 9 là kết quả đánh giá hình ảnh CT phổi trước và sau khi điều trị dựa trên đánh giá thuốc đích được đề xuất bởi phương pháp đánh giá theo sáng chế. (A) hình ảnh CT

phổi trước khi điều trị; (B) hình ảnh CT phổi sau 3 tháng điều trị dựa trên mô hình phân tích, đánh giá thuốc hướng đích bằng phương pháp theo sáng chế.

Mô tả chi tiết sáng chế

Sau đây sáng chế mô tả chi tiết sáng chế với các phương án và các ví dụ thực hiện cụ thể, tuy nhiên các phương án và các ví dụ thực hiện này chỉ nhằm mục đích minh họa làm rõ bản chất của sáng chế chứ không nhằm mục đích giới hạn phạm vi yêu cầu bảo hộ của sáng chế.

Sáng chế đề cập phương pháp đánh giá yếu tố đột biến di truyền đến đáp ứng thuốc đích điều trị ung thư phổi không tế bào nhỏ, trong đó phương pháp này bao gồm các bước: a) thiết lập cơ sở dữ liệu di truyền chuẩn; b) thiết lập cơ sở dữ liệu sao sánh; c) khởi tạo dữ liệu đánh giá; d) thiết lập mô hình phân tích máy học; và e) hiệu chỉnh dữ liệu so sánh.

Theo sáng chế, cơ sở dữ liệu di truyền chuẩn liên quan đến ung thư phổi không tế bào nhỏ được xác định bao gồm 43 gen bao gồm các gen EGFR, KRAS, ALK, STK11, FGFR1, PIK3CA, ARAF, NRG1, RRM1, HAVCR2, MET, CDK4, BRCA1, CCND1, PDGFRA, AXL, AURKA, TYMS, ETS2, DDR2, ERCC2, ABCB1, TP53, NFE2L2, PDCD4, RET, BRAF, ROS1, NTRK1, XRCC1, CD274, RICTOR, AREG, BCL2L11, BIRC5, ERCC1, GNAS, AKT2, ERBB2, POLD1, RB1, SMARCA4 và WEE1. Các đoạn môi đặc hiệu để giải trình tự liên quan đến các gen này cũng đã được phát triển tùy vào từng gen hoặc từng vị trí đột biến đã được công bố. Cơ sở dữ liệu về các gen này có thể được khai thác trên ngân hàng gen (Genbank) của Hoa Kỳ và được công bố, có thể truy cập trực tuyến tại <https://www.ncbi.nlm.nih.gov/genbank/>.

Để thiết lập cơ sở dữ liệu di truyền chuẩn, tiến hành tạo tập hợp dữ liệu trình tự 43 gen liên quan đến phát sinh bệnh ung thư phổi không tế bào nhỏ. Đối với mỗi gen, tiến hành thu thập dữ liệu liên quan đến từng vị trí đột biến điểm gây ung thư theo dữ liệu đã được công bố. Từng tập hợp dữ liệu của 43 gen theo định dạng chuẩn được khai thác, tạo tập dữ liệu. Trong đó tạo tập hợp dữ liệu trình tự gen của 43 gen liên quan bao gồm EGFR, KRAS, ALK, STK11, FGFR1, PIK3CA, ARAF, NRG1, RRM1, HAVCR2, MET, CDK4, BRCA1, CCND1, PDGFRA, AXL, AURKA, TYMS, ETS2, DDR2, ERCC2, ABCB1, TP53, NFE2L2, PDCD4, RET, BRAF, ROS1, NTRK1, XRCC1, CD274, RICTOR, AREG, BCL2L11, BIRC5, ERCC1, GNAS, AKT2, ERBB2,

POLD1, RB1, SMARCA4 và WEE1 được tập hợp và từng gen được so sánh toàn cục để xác định các vị trí đột biến điểm để đưa ra chính xác vị trí đột biến trên gen và các trình tự chuẩn không đột biến. Các dữ liệu này được tập hợp tạo ra các tham số so sánh đối với từng gen để xác định tình trạng gen chuẩn và các vị trí đột biến liên quan đến gen đó đã được công bố. Nguồn dữ liệu di truyền chuẩn liên quan đến các đột biến được thu thập, đánh giá, ví dụ bằng cách so sánh tổng thể (so sánh toàn cục) các trình tự gen đã được công bố. Từ đó thu về trình tự gen chuẩn với các vị trí gen đột biến gây bệnh ung thư được đánh dấu. Trình tự gen chuẩn này được sử dụng như bản đồ đột biến liên quan đến bệnh thu thư phổi tế bào không nhỏ. Danh sách 43 gen liên quan đến bệnh ung thư phổi tế bào không nhỏ được tập hợp mô tả chung trong Bảng 1. Đối với mỗi gen này, sau khi so sánh toàn cục, sẽ thiết lập được các vị trí đột biến điểm gây ung thư theo dữ liệu đã được công bố.

Bảng 1: Danh sách 43 gen ung thư phổi

Gen	Vị trí gen	Tên đầy đủ	Tên khác	Đặc điểm gây bệnh
AKT2	19q13.2	AKT serine/threonine kinase 2	PKBB; PRKBB; HIHGH; PKBBETA; RAC-BETA	Được coi là một gen gây ung thư, mã hóa cho một protein thuộc một phân họ serine/threonine kinase chứa các vùng giống SH2, có liên quan đến các con đường truyền tín hiệu. Protein được mã hóa là một protein kinase có khả năng phosphoryl hóa một số protein khác và liên quan đến việc truyền tín hiệu insulin.
ERBB2	17q12	erb-b2 receptor tyrosine kinase 2	NEU; NGL; HER2; TKR1; CD340; HER-2; VSCN2; MLN 19; MLN-19; c-ERB2; c-ERB-2; HER-2/neu; p185(erbB2)	Mã hóa cho protein là một thành viên thuộc họ thụ thể yếu tố tăng trưởng biểu mô. Sự khuếch đại và/hoặc biểu hiện quá mức của gen này đã được báo cáo ở nhiều bệnh ung thư, bao gồm cả khối u vú và buồng trứng.
RB1	13q14.2	RB transcriptional corepressor	RB; pRb; OSRC; pp110; p105-Rb; PPP1R130; p110-RB1	Mã hóa cho protein là chất điều hòa âm tính chu kỳ tế bào, có tác dụng ức chế khối u. Protein này cũng giúp ổn định chất nhiễm sắc, duy trì cấu trúc nhiễm sắc thể. Đột biến trong gen này

				thường dẫn đến ung thư nguyên bào vồng mạc, ung thư bàng quang và sarcoma xương.
SMARCA4	19p13.2	SWI/SNF related, matrix associated, actin dependent regulator of chromatin, subfamily a, member 4	BRG1; CSS4; SNF2; SWI2; MRD16; RTPS2; BAF190; SNF2L4; SNF2LB; hSNF2b; BAF190A; SNF2-beta	Mã hóa cho một protein là thành viên thuộc họ protein SWI/SNF và tương tự như protein brahma của ruồi giấm; các thành viên của họ này có các hoạt động helicase và ATPase và được cho là có thể điều chỉnh quá trình phiên mã của một số gen nhất định bằng cách thay đổi cấu trúc nhiễm sắc thể xung quanh các gen đó. Protein này còn có thể liên kết với protein BRCA1 và điều hòa sự biểu hiện của protein CD44.
STK11	19p13.3	serine/threonine kinase 11	PJS; LKB1; hLKB1	Protein được mã hóa bởi gen này là serine/threonine kinase có chức năng điều hòa sự phân cực của tế bào và chuyển hóa năng lượng, đồng thời có chức năng như một chất ức chế khối u. Đột biến ở gen này thường liên quan đến hội chứng Peutz-Jeghers, ung thư da, ung thư tuyến tụy và tinh hoàn.
EGFR	7p11.2	epidermal growth factor receptor	ERBB; ERBB2; HER1; mENA; ERBB1; PIG61; NISBD2	Mã hóa cho protein là thụ thể của yếu tố tăng trưởng biểu mô, giúp tăng sinh tế bào. Đột biến xảy ra ở gen này thường liên quan nhiều đến ung thư phổi.
POLD1	19q13.3	DNA polymerase delta 1, catalytic subunit	CDC2; MDPL; POLD; CRCS10	Mã hóa tiểu đơn vị xúc tác 125 kDa của ADN polymerase delta, có vai trò quan trọng trong sao chép và chỉnh sửa ADN.
ERCC1	19q13.32	ERCC excision repair 1, endonuclease non-catalytic subunit	UV20; COFS4; RAD10	Protein mã hóa có vai trò trong quá trình sửa chữa cắt bỏ nucleotid, ví dụ như tổn thương ADN do tia UV. Các đột biến trong gen làm thay đổi mức độ biểu hiện của gen có thể đóng vai trò trong quá trình sinh ung thư.

GNAS	20q13.32	GNAS complex locus	AHO; GSA; GSP; POH; GPSA; NESP; SCG6; SgVI; GNAS1; PITA3; C20orf45	Có rất nhiều biến thể mã hóa các dạng đồng phân khác nhau của gen này. Các đột biến trên gen này có thể dẫn đến suy tuyến cận giáp giả typ 1a, 1b; hội chứng McCune-Albright,... và một số khối u tuyến yên.
BIRC5	17q25.3	baculoviral IAP repeat containing 5	API4; EPR-1	Một gen thuộc họ gen ức chế quá trình chết theo chương trình mã hóa các protein điều hòa âm tính ngăn ngừa sự chết tế bào theo chương trình. Sự biểu hiện của gen thường cao trong quá trình phát triển của thai nhi và trong nhiều khối u.
BCL2L1 1	2q13	BCL2 like 11	BAM; BIM; BOD	Mã hóa protein thuộc họ protein BCL-2, có vai trò điều hòa chống lại hoặc thúc đẩy quá trình chết tế bào theo chương trình, tham gia vào nhiều hoạt động khác của tế bào.
AURKA	20q13.2	aurora kinase A	AIK; ARK1; AURA; BTAK; STK6; STK7; STK15; PPP1R47	Mã hóa protein kinase có vai trò điều hòa chu kỳ tế bào, có thể có liên quan đến quá trình hình thành, ổn định các vi ống trong quá trình phân chia nhiễm sắc thể. Gen này có thể có vai trò trong sự phát triển và tiến triển của khối u.
FGFR1	8p11.23	fibroblast growth factor receptor 1	CEK; FLG; HH2; OGD; ECCL; FLT2; KAL2; BFGFR; CD331; FGFR; FLT-2; HBGFR; N-SAM; FGFR-1; HRTFDS; bFGF-R-1	Mã hóa cho protein thuộc họ thụ thể yếu tố tăng trưởng nguyên bào sợi. Phần ngoại bào của protein tương tác với các yếu tố tăng trưởng nguyên bào sợi, dẫn đến một loạt các tín hiệu nội bào, từ đó ảnh hưởng đến quá trình nguyên phân và biệt hóa tế bào. Các đột biến trong gen này có liên quan đến loạn sản sụn xương, hội chứng Kallmann 2,... Thay đổi nhiễm sắc thể ảnh hưởng đến gen này có thể liên quan đến rối loạn sinh dục.

TYMS	18p11.32	thymidylate synthetase	TS; TMS; DKCD; HST422	Xúc tác cho phản ứng methyl hóa deoxyuridylate thành deoxythymidylate, sử dụng 10-methylenetetrahydrofolate như một cofactor. Các đa hình của gen này có thể liên quan đến sự tạo thành khối u, trong đó có ung thư vú, và liên quan đến cả đáp ứng với điều trị hóa chất.
ETS2	21q22.2	ETS proto-oncogene 2, transcription factor	ETS2IT1	Mã hóa một yếu phiên mã có vai trò điều hòa các gen trong sự phát triển và chết tế bào theo chương trình. Protein mã hóa cũng được coi là một gen tiền ung thư và có ảnh hưởng đến quá trình điều hòa telomerase.
DDR2	1q23.3	discoidin domain receptor tyrosine kinase 2	TKT; WRCN; MIG20a; NTRKR3; TYRO10	Mã hóa protein thuộc họ protein thụ thể tyrosine kinase, đóng vai trò quan trọng trong việc sự thông tin giữa tế bào với vi môi trường. Protein mã hóa kích hoạt các con đường tín hiệu liên quan đến sự kết dính, tăng sinh và tái cấu trúc chất nền ngoại bào.
PIK3CA	3q26.32	phosphatidylinositol-4,5-bisphosphate 3-kinase catalytic subunit alpha	MCM; CCM4; CWS5; MCAP; PI3K; CLAPO; CLOVE; MCMTC; PI3K-alpha; p110-alpha	Gen được phát hiện có liên quan đến sự hình thành ung thư và được cho là có liên quan đến ung thư cổ tử cung.
ABCB1	7q21.12	ATP binding cassette subfamily B member 1	CLCS; MDR1; P-GP; PGY1; ABC20; CD243; GP170; p-170	Protein liên kết với màng được mã hóa bởi gen này là một thành viên của siêu họ vận chuyển cassette liên kết ATP (ABC), làm giảm sự tích lũy thuốc trong các tế bào đa kháng thuốc và thường làm trung gian cho sự hình thành đề kháng với thuốc chống ung thư.
ALK	2p23.2-p23.1	ALK receptor	ALK1; CD246; NBLST3	Mã hóa một thụ thể tyrosine kinase thuộc siêu họ thụ thể insulin. Các thay đổi tái cấu trúc, đột biến, khuếch đại gen này đã

		tyrosine kinase		được tìm thấy trong một loạt khối u bao gồm u lympho, u nguyên bào thần kinh và ung thư phổi không phải tế bào nhỏ. Sự sắp xếp lại nhiễm sắc thể là những thay đổi di truyền phổ biến nhất ở gen này, dẫn đến việc tạo ra nhiều gen fusion gây ra khối u, bao gồm ALK/EML4, ALK/RANBP2, ALK/ATIC, ALK/TFG, ALK/NPM1, ALK/SQSTM1, ALK/KIF5B, ALK/CLTC, ALK/TPM4 và ALK/MSN.
NFE2L2	2q31.2	NFE2 like bZIP transcription factor 2	NRF2; HEBP1; Nrf-2; IMDDHH	Gen này mã hóa một yếu tố phiên mã là thành viên của một họ nhỏ protein cơ bản leucine Zipper (bZIP), có tác dụng điều hòa các gen có chứa các yếu tố phản ứng chống oxy hóa trong vùng khởi động của chúng; nhiều gen trong số này mã hóa các protein liên quan đến phản ứng với chấn thương và viêm, bao gồm cả việc sản xuất các gốc tự do.
PDCD4	10q25.2	programmed cell death 4	H731	Gen này là chất ức chế khối u và mã hóa một protein liên kết với yếu tố khởi đầu dịch mã 4A1 và ức chế chức năng của nó bằng cách ngăn chặn sự liên kết với RNA.
ERCC2	19q13.32	ERCC excision repair 2, TFIIH core complex helicase subunit	EM9; TTD; XPD; TTD1; COFS2; TFIIH	Protein được mã hóa bởi gen này có liên quan đến quá trình sửa chữa cắt bỏ nucleotide kết hợp phiên mã và là thành viên không thể thiếu của phức hợp yếu tố phiên mã cơ bản BTF2/TFIIH. Khiếm khuyết ở gen này có thể dẫn đến hội chứng dễ bị ung thư xeroderma sắc tố bổ sung nhóm D, rối loạn trichothiodystrophy và hội chứng Cockayne.
TP53	17p13.1	tumor protein p53	P53; BCC7; LFS1; BMFS5; TRP53	Mã hóa một protein ức chế khối u có chức năng kích hoạt phiên mã, liên kết DNA và các miền oligome hóa. Protein được mã

				hóa phản ứng với các thay đổi khác nhau của tế bào để điều chỉnh sự biểu hiện của gen mục tiêu, từ đó gây ra sự ngừng chu kỳ tế bào, chết theo chương trình, sửa chữa DNA hoặc thay đổi quá trình trao đổi chất. Đột biến ở gen này có liên quan đến nhiều loại bệnh ung thư ở người, bao gồm cả ung thư di truyền như hội chứng Li-Fraumeni.
CD274	9p24.1	CD274 molecule	B7-H; B7H1; PDL1; PD-L1; hPD-L1; PDCD1L1; PDCD1LG1	Mã hóa một phối tử thụ thể ức chế miễn dịch được biểu hiện bởi các tế bào tạo máu và không tạo máu, chẳng hạn như tế bào T và tế bào B và các loại tế bào khối u khác nhau. Sự biểu hiện của gen này trong các tế bào khối u được coi là tiên lượng cho nhiều loại bệnh ác tính ở người, bao gồm ung thư đại trực tràng và ung thư biểu mô tế bào thận.
RET	10q11.21	ret proto-oncogene	PTC; MTC1; HSCR1; MEN2A; MEN2B; CDHF12; CDHR16; RET-ELE1	Là một gen tiền ung thư, mã hóa một thụ thể xuyên màng và là thành viên của họ protein tyrosine kinase.
BRAF	7q34	B-Raf proto-oncogene, serine/threonine kinase	NS7; B-raf; BRAF1; RAFB1; B-RAF1; BRAF-1	Mã hóa một protein thuộc họ RAF của protein kinase/threonine, đóng vai trò điều chỉnh con đường tín hiệu MAP kinase/ERK, ảnh hưởng đến sự phân chia, biệt hóa và bài tiết của tế bào. Đột biến trong gen này, phổ biến nhất là đột biến V600E, là đột biến gây ung thư được xác định thường xuyên nhất trong khối u ác tính ở nhiều loại ung thư khác, bao gồm u lympho không Hodgkin, ung thư đại trực tràng, ung thư biểu mô tuyến giáp, ung thư biểu mô phổi không tế bào nhỏ, ung thư biểu mô tuyến phổi.

RICTOR	5p13.1	RPTOR independent companion of MTOR complex 2	PIA; AVO3; hAVO3	RICTOR và MTOR là các thành phần của phức hợp protein tích hợp các tín hiệu từ chất dinh dưỡng và yếu tố tăng trưởng để điều chỉnh sự phát triển của tế bào.
ROS1	6q22.1	ROS proto-oncogene 1, receptor tyrosine kinase	ROS; MCF3; c-ros-1	Là gen tiền ung thư, có độ biểu hiện cao trong nhiều dòng tế bào khối u, thuộc phân họ thụ thể insulin tyrosine kinase. Protein được mã hóa bởi gen này là protein màng có hoạt tính tyrosine kinase, có thể hoạt động như một thụ thể của yếu tố tăng trưởng hoặc biệt hóa.
NTRK1	1q23.1	neurotrophic receptor tyrosine kinase 1	MTC; TRK; TRK1; TRKA; Trk-A; p140-TrkA	Mã hóa một thành viên của họ thụ thể tyrosine kinase (NTRK) thần kinh. Kinase này là một thụ thể gắn trên màng, khi liên kết với tế bào thần kinh sẽ phosphoryl hóa chính nó và các thành viên của con đường MAPK. Đột biến ở gen này có thể gây ra khuyết tật về nhận thức và khối u ác tính.
WEE1	11p15.4	WEE1 G2 checkpoint kinase	WEE1A; WEE1hu	Mã hóa một protein là tyrosine kinase thuộc họ protein kinase Ser/Thr.
AREG	4q13.3	amphiregulin	AR; SDGF; AREGB; CRDGF	Protein được mã hóa bởi gen này là thành viên của họ yếu tố tăng trưởng biểu mô, tương tác với thụ thể EGF/TGF- α để thúc đẩy sự phát triển của các tế bào biểu mô bình thường và ức chế sự phát triển của một số dòng tế bào ung thư biểu mô.
XRCC1	19q13.31	X-ray repair cross complementing 1	RCC; SCAR26	Mã hóa protein có liên quan đến quá trình sửa chữa các đứt gãy chuỗi đơn ADN do tiếp xúc với bức xạ ion hóa và các tác nhân kiềm hóa. Protein này tương tác với ADN ligase III, polymerase beta và poly (ADP-ribose)

				polymerase để tham gia vào quá trình sửa chữa cắt bỏ bazơ.
KRAS	12p12.1	KRAS proto-oncogene, GTPase	NS; NS3; OES; CFC2; RALD; K-Ras; KRAS1; KRAS2; RASK2; KI-RAS; C-K-RAS; K-RAS2A; K-RAS2B; K-RAS4A; K-RAS4B; K-Ras 2; 'C-K-RAS; c-Ki-ras; c-Ki-ras2	Mã hóa protein là thành viên của siêu họ GTPase. Đột biến gen này có liên quan đến nhiều loại bệnh ác tính khác nhau, bao gồm ung thư biểu mô tuyến phổi, u tuyến nhầy, ung thư biểu mô ống tuyến tụy và ung thư biểu mô đại trực tràng.
ARAF	Xp11.3	A-Raf proto-oncogene, serine/threonine kinase	PKS2; A-RAF; ARAF1; RAFA1	Là một gen tiền ung thư, kích hoạt protein serine/threonine kinase, điều hòa âm tính quá trình chết theo chương trình, tín hiệu TOR và quá trình trao đổi chất của tế bào.
NRG1	8p12	neuregulin 1	GGF; HGL; nhân sự; NDF; ARIA; GGF2; HRG1; HRGA; SMDF; MST131; MSTP131; NRG1-IT2	Protein được mã hóa bởi gen này là một glycoprotein màng làm trung gian truyền tín hiệu tế bào và đóng vai trò quan trọng trong sự tăng trưởng và phát triển của nhiều hệ cơ quan. Sự thay đổi trong gen này có liên quan đến các bệnh như ung thư, tâm thần phân liệt và rối loạn lưỡng cực.
RRM1	11p15.4	ribonucleotide reductase catalytic subunit M1	R1; RR1; RIR1	Mã hóa tiểu đơn vị lớn và có tính xúc tác của ribonucleotide reductase, một enzyme cần thiết cho việc chuyển đổi ribonucleotide thành deoxyribonucleotide.
HAVCR2	5q33.3	hepatitis A virus cellular receptor 2	TIM3; CD366; KIM-3; SPTCL; TIMD3; Tim-3; TIMD-3; HAVcr-2	Protein được mã hóa bởi gen này thuộc siêu họ immunoglobulin và họ protein TIM, là protein bề mặt tế bào đặc hiệu Th1, có chức năng điều chỉnh hoạt hóa đại thực bào và ức chế các phản ứng tự miễn dịch và miễn dịch qua trung gian Th1, đồng thời thúc đẩy khả năng miễn dịch.

MET	7q31	MET proto-oncogene, receptor tyrosine kinase	DA11; HGFR; AUTS9; RCCP2; c-Met; DFNB97	Mã hóa một thành viên của họ protein thụ thể tyrosine kinase. Đột biến ở gen này có liên quan đến ung thư biểu mô tế bào thận dạng nhú, ung thư biểu mô tế bào gan và các bệnh ung thư đầu cổ khác.
CDK4	12q14.1	cyclin dependent kinase 4	CMM3; PSK-J3	Protein được mã hóa bởi gen này là thành viên của họ protein Serine/Threonine kinase. Các đột biến ở gen này cũng như các protein liên quan của nó bao gồm cyclin loại D, p16 (INK4a) và Rb đều được phát hiện có liên quan đến nguyên nhân khối u của nhiều loại bệnh ung thư.
BRCA1	17q21.31	BRCA1 DNA repair associated	IRIS; PSCP; BRCAI; BRCC1; FANCS; PNCA4; RNF53; BROVCA1; PPP1R53	Gen này mã hóa phosphoprotein 190 kD có vai trò duy trì sự ổn định của bộ gen và nó cũng hoạt động như một chất ức chế khối u. Đột biến ở gen này liên quan đến khoảng 40% bệnh ung thư vú di truyền và hơn 80% bệnh ung thư vú và ung thư buồng trứng di truyền.
CCND1	11q13.3	cyclin D1	BCL1; PRAD1; U21B31; D11S287E	Mã hóa protein thuộc họ cyclin có tính bảo tồn cao, có chức năng điều chỉnh các kinase CDK. Protein này đã được chứng minh là có tương tác với protein ức chế khối u Rb và sự biểu hiện của gen này được điều hòa bởi Rb. Đột biến, khuếch đại và biểu hiện quá mức của gen này, làm thay đổi sự phát triển của chu kỳ tế bào, được tìm thấy ở nhiều loại bệnh ung thư ở người.
PDGFRA	4q12	platelet derived growth factor receptor alpha	CD140A; PDGFR2; PDGFR-2	Gen này mã hóa thụ thể tyrosine kinase bề mặt tế bào cho các thành viên của họ yếu tố tăng trưởng có nguồn gốc từ tiểu cầu. Những yếu tố tăng trưởng này là tác nhân phân bào cho các tế bào có nguồn gốc trung mô. Các nghiên cứu cho thấy gen này đóng vai trò trong việc phát triển

				cơ quan, chữa lành vết thương và phát triển khối u. Đột biến trong gen này có liên quan đến hội chứng tăng bạch cầu ái toan vô căn, khối u mô đệm đường tiêu hóa và nhiều loại ung thư khác.
AXL	19q13.2	AXL receptor tyrosine kinase	ARK; UFO; AXL3; JTK11; Tyro7	Protein được mã hóa bởi gen này là thành viên của phân họ tyrosine kinase thụ thể Tyro3-Axl-Mer (TAM). Gen này có thể tham gia vào một số chức năng của tế bào bao gồm tăng trưởng, di chuyển và chống viêm ở nhiều loại tế bào.

Trong bước thiết lập cơ sở dữ liệu so sánh, dữ liệu này được thiết lập căn cứ theo từng vị trí đột biến điểm đã biết. Ngoài ra, tiến hành thu thập các dữ liệu liên quan đến các thông số phát sinh bệnh liên quan đến ung thư phổi, ưu tiên các thông tin liên quan đến các tình trạng kháng, đáp ứng, không đáp ứng hoặc quá mẫn với các dược chất nghiên cứu. Các thông tin thu thập liên quan đến mô hình phát sinh bệnh liên quan đến bệnh ung thư phổi không tế bào nhỏ bao gồm các chỉ số lâm sàng, cận lâm sàng. Các thuốc đích đã được sử dụng để điều trị và tiến trình điều trị. Các phát sinh trong quá trình điều trị cùng với trình tự gen bất kỳ được giải để xác định đột biến liên quan đến phát sinh bệnh học. Trong số từng đột biến được sử dụng làm dữ liệu so sánh này được thiết lập tạo ra cơ sở dữ liệu tham chiếu tổng số đủ cho ít nhất 43 gen tham chiếu với các exon chứa đột biến điểm. Theo sáng chế, các exon vị trí đột biến xác định bao gồm các đột biến nằm trong các exon 1 đến exon 28 của gen EGFR, các exon 2 đến exon 6 của gen KRAS, các exon 1-10 của gen ALK, các exon 1-10 của gen STK11, các exon 1-10 của gen FGFR1, các exon 2-10 của gen PIK3CA, các exon 2-7 và các exon 9-16 của gen ARAF, các exon 1-3, exon 6 và các exon 8-12 của gen NRG1, các exon 1-4, các exon 7-19 của gen RRM1, exon 1-2, exon 5, exon 7 của gen HAVCR2, các exon 1-21 của gen MET, các exon 1-5 và các exon 7-8 của gen CDK4, các exon 1-2, các exon 6-7, các exon 10-23 của gen BRCA1, các exon 1-5 của gen CCND1, các exon 1-23 của gen PDGFRA, các exon 1-9, các exon 11-20 của gen AXL, các exon 1-3, 5-9 của gen AURKA, các exon 1-7 của gen TYMS, exon 1 của gen ETS2, các exon 1-8 của gen DDR2, các exon 2, exon 5-13 của gen ERCC2, các exon 1, exon 4, exon 6-9, exon 12-15 của gen ABCB1, các exon 1, exon 2a-2c, exon 3, exon 4a, exon 4b, exon 5-8, exon

9a, exon 10-11 của gen TP53, gen NFE2L2, gen PDCD4, các exon 1, exon 3-15, exon 17-18, exon 19b, exon 20 của gen RET, các exon 1-8, exon 10-18 của gen BRAF, các exon 1, exon 5, -7, exon 10-21, exon 23-36, exon 38-43 của gen ROS1, các exon 1-5, exon 7-10, exon 13-19 của gen NTRK1, các exon 1-13, exon 14-15, exon 17 của gen XRCC1, các đoạn chứa đột biến NM_014143.4, NM_001314029.2 của gen CD274, các đoạn chứa đột biến NM_152756.5, NM_001285439.2 của gen RICTOR, đoạn chứa đột biến NM_001657.4 của gen AREG, các exon 1, exon 2a-2c, exon 3, exon 7-12 của gen BCL2L11, các exon 1-4 của gen BIRC5, các exon 1, exon 3-8, exon 10 của gen ERCC1, các exon 1-3, exon 6-7, exon 9, exon 11-12, exon 14 của gen GNAS, các exon 1-6, 8-12 của gen AKT2, các exon 1-2, exon 4, exon 5-9, exon 11-13 của gen ERBB2, các exon 2A-2B, exon 3-10 của gen POLD1, các exon 1-4, exon 7-8, exon 11-13, exon 17 của gen RB1, các exon 1-10 của gen SMARCA4, các exon 1-2, exon 4-11 của gen WEE1. Dữ liệu giải trình tự gen và các thông tin đối với mỗi vị trí đột biến được giải trình tự trên được tập hợp dữ liệu được và tập hợp các thông số được thiết lập thành từng tệp dữ liệu với các thông số liên quan đến vị trí đột biến điểm với mỗi exon và đoạn gen trong số 43 gen nêu trên. Các thông tin cận lâm sàng được khai thác bổ sung bao gồm độ tuổi, cân nặng, nghề nghiệp, tiền sử, thời gian nhập viện, thời gian xuất hiệu các triệu chứng tiền lâm sàng. Các thông tin lâm sàng bao gồm triệu chứng hô hấp (ho, khạc đờm, ho ra máu, đau ngực, khó thở), triệu chứng toàn thân (mệt mỏi, gầy sút cân, hạch ngoại vi, cân nặng chiều cao) đánh giá thể trạng theo tiêu chuẩn của Tổ chức Y tế Thế giới (WHO), trong đó điểm toàn trạng (PS) được đánh giá theo thang điểm từ 0-4. Dữ liệu về chẩn đoán hình ảnh được lượng hóa theo vị trí, kích thước, số lượng và các đặc điểm về hình thái, tính chất xâm lấn của u trên hình ảnh X quang lồng ngực tiêu chuẩn, hình ảnh CLVT lồng ngực có tiêm thuốc cản quang, phim chụp MRI sọ não, xạ hình xương hoặc PET/CT trước, trong và sau quá trình điều trị cũng được tập hợp. Các dữ liệu liên quan đến trạng thái, tiến trình, các phác đồ trị liệu áp dụng và kết quả điều trị cũng được thu thập. Tất cả các dữ liệu được tạo thành các trường tập hợp riêng biệt đối với từng mô hình riêng lẻ.

Trong bước khởi tạo dữ liệu đánh giá, các dữ liệu được thu thập bao gồm các chỉ số lâm sàng, cận lâm sàng và giải trình tự toàn bộ các đoạn gen (exon) chứa đột biến điểm liên quan đến 43 gen gây bệnh ung thư phổi không tế bào nhỏ đối với mô hình cần đánh giá.

Các đoạn gen được giải trình tự bao gồm các đột biến nằm trong các exon 1 đến exon 28 của gen EGFR, các exon 2 đến exon 6 của gen KRAS, các exon 1-10 của gen ALK, các exon 1-10 của gen STK11, các exon 1-10 của gen FGFR1, các exon 2-10 của gen PIK3CA, các exon 2-7 và các exon 9-16 của gen ARAF, các exon 1-3, exon 6 và các exon 8-12 của gen NRG1, các exon 1-4, các exon 7-19 của gen RRM1, exon 1-2, exon 5, exon 7 của gen HAVCR2, các exon 1-21 của gen MET, các exon 1-5 và các exon 7-8 của gen CDK4, các exon 1-2, các exon 6-7, các exon 10-23 của gen BRCA1, các exon 1-5 của gen CCND1, các exon 1-23 của gen PDGFRA, các exon 1-9, các exon 11-20 của gen AXL, các exon 1-3, 5-9 của gen AURKA, các exon 1-7 của gen TYMS, exon 1 của gen ETS2, các exon 1-8 của gen DDR2, các exon 2, exon 5-13 của gen ERCC2, các exon 1, exon 4, exon 6-9, exon 12-15 của gen ABCB1, các exon 1, exon 2a-2c, exon 3, exon 4a, exon 4b, exon 5-8, exon 9a, exon 10-11 của gen TP53, gen NFE2L2, gen PDCD4, các exon 1, exon 3-15, exon 17-18, exon 19b, exon 20 của gen RET, các exon 1-8, exon 10-18 của gen BRAF, các exon 1, exon 5, -7, exon 10-21, exon 23-36, exon 38-43 của gen ROS1, các exon 1-5, exon 7-10, exon 13-19 của gen NTRK1, các exon 1-13, exon 14-15, exon 17 của gen XRCC1, các đoạn chứa đột biến NM_014143.4, NM_001314029.2 của gen CD274, các đoạn chứa đột biến NM_152756.5, NM_001285439.2 của gen RICTOR, đoạn chứa đột biến NM_001657.4 của gen AREG, các exon 1, exon 2a-2c, exon 3, exon 7-12 của gen BCL2L11, các exon 1-4 của gen BIRC5, các exon 1, exon 3-8, exon 10 của gen ERCC1, các exon 1-3, exon 6-7, exon 9, exon 11-12, exon 14 của gen GNAS, các exon 1-6, 8-12 của gen AKT2, các exon 1-2, exon 4, exon 5-9, exon 11-13 của gen ERBB2, các exon 2A-2B, exon 3-10 của gen POLD1, các exon 1-4, exon 7-8, exon 11-13, exon 17 của gen RB1, các exon 1-10 của gen SMARCA4, các exon 1-2, exon 4-11 của gen WEE1 của mô hình để tạo tập hợp dữ liệu đánh giá. Việc giải trình tự các exon và các đoạn gen nêu trên có thể thực hiện bởi kit theo sáng chế kết hợp phương pháp giải trình tự toàn bộ, ví dụ phương pháp giải trình tự (Next Generation Sequencing: NGS). Đây là phương pháp cho phép giải trình tự đồng thời hàng triệu đoạn ADN trong cùng lúc, được ứng dụng để giải trình tự gen người. Kỹ thuật này có thời gian đọc nhanh, dữ liệu đầu ra lớn và tiết kiệm chi phí hơn, khắc phục một số nhược điểm của giải trình tự gen truyền thống và cho phép giải toàn bộ exon của gen. Do exon chỉ chiếm 1% của bộ gen nhưng chứa thông tin trực tiếp mã hoá cho các protein thực hiện chức năng trong cơ thể. Vì vậy, những thay đổi

trình tự của exon có thể liên quan trực tiếp đến bệnh tật hay tình trạng sức khỏe của con người. Quy trình này đã được thương mại hoá nên dễ thực hiện và khá phổ biến. Kỹ thuật này cũng cho phép giải trình tự một hoặc một vài gen đích, đặc biệt hữu ích đối với kỹ thuật đánh giá ung thư di truyền. Vì số lượng ít và mục tiêu rõ ràng nên kết quả và độ chính xác của việc giải trình tự này cao hơn, nhưng chi phí lại thấp hơn.

Vị trí của các gen được giải trình tự và các cặp môi tương ứng được thể hiện trong bảng 2 sau. Các phương pháp giải trình tự gen dựa trên các đoạn môi đặc hiệu đã được biết và được thực hiện bởi người có hiểu biết trung bình trong lĩnh vực kỹ thuật này để có thể thu được thông tin giải trình tự, tốt nhất là sử dụng phương pháp giải trình tự gen NGS nêu trên với các đoạn môi tương ứng nêu trong bảng 2 bên dưới. Các phương pháp giải trình tự khác có thể được thực hiện, ví dụ bằng phương pháp giải, ví dụ phương pháp Stranger hoặc sử dụng các thiết bị giải trình tự gen tự động đã biết trong lĩnh vực kỹ thuật này. Tuy nhiên việc giải trình tự dữ liệu lớn tốt nhất được thực hiện bằng kỹ thuật NGS để giảm chi phí và thời gian giải và phân tích dữ liệu.

Bảng 2. Danh sách các cặp môi được sử dụng để giải trình tự gen tham chiếu gây bệnh ung thư phổi không tế bào nhỏ

SEQ ID NO	Trình tự	Vị trí	Gen
SEQ ID NO.1	CCCCTGACTCCGTCCAGTAT	Exon 1	EGFR
SEQ ID NO.2	CTTTTCCTCCAGAGCCCGAC		
SEQ ID NO.3	ACGAGTAACAAGCTCACGCA	Exon 2	
SEQ ID NO.4	CCAAGGACCACCTCACAGTT		
SEQ ID NO.5	CCAGGAGGTGGCTGGTTATG	Exon 3	
SEQ ID NO.6	CAGCTCCTTCAGTCCGGTTT		
SEQ ID NO.7	AACGTGGAGAGCATCCAGTG	Exon 4	
SEQ ID NO.8	CCCAGGTGGTTCTGGAAGTC		
SEQ ID NO.9	GCCAAAAGTGTGATCCAAGC	Exon 5	
SEQ ID NO.10	CAGTTCTCCTCTCCTGCACC		
SEQ ID NO.11	AATCATCTGTGCCAGCAGT	Exon 6	
SEQ ID NO.12	CCTGCAGCACACTGGTTGT		
SEQ ID NO.13	GCAAATTCCGAGACGAAGCC	Exon 7	
SEQ ID NO.14	CTGTATTTGCCCTCGGGGTT		
SEQ ID NO.15	TATGTGGTGACAGATCACGGC	Exon 8	

SEQ ID NO.16	AAGGCCCTTCGCACTTCTTA		
SEQ ID NO.17	TGTGTAACGGAATAGGTATTGGTG	Exon 9	
SEQ ID NO.18	GGAGATCGCCACTGATGGAG		
SEQ ID NO.19	TCCTTCACACATACTCCTCCTCT	Exon 10	
SEQ ID NO.20	CTGTGATTTCTTTACGGTTTTCA		
SEQ ID NO.21	TTTGCTGATTCAGGCTTGGC	Exon 11	
SEQ ID NO.22	TCCTGCCGCGTATGATTTCT		
SEQ ID NO.23	TTCTCTTGCAGTCGTCAGCC	Exon 12	
SEQ ID NO.24	TTCTGACCGGAGGTCCCAA		
SEQ ID NO.25	AGGTCTGCCATGCCTTGTG	Exon 13	
SEQ ID NO.26	GTTGCACTTGTCCACGCATT		
SEQ ID NO.27	CCAAGGGAGTTTGTGGAGAACT	Exon 14	
SEQ ID NO.28	GTCCTGTGCAGGTGATGTTC		
SEQ ID NO.29	GTGTGCCCACTACATTGACG	Exon 15	
SEQ ID NO.30	CTGCGTACTTCCAGACCAGG		
SEQ ID NO.31	CCTAAGATCCCGTCCATCGC	Exon 17	
SEQ ID NO.32	CTTCGCATGAAGAGGCCGAT		
SEQ ID NO.33	GCTCCCAACCAAGCTCTCTT	Exon 18	
SEQ ID NO.34	TTATACACCGTGCCGAACGC		
SEQ ID NO.35	ACTCTGGATCCCAGAAGGTGA	Exon 19	
SEQ ID NO.36	TGGCTTTCGGAGATGTTGCT		
SEQ ID NO.37	CCTGGACTATGTCCGGGAAC	Exon 20	
SEQ ID NO.38	CTTTGCGATCTGCACACACC		
SEQ ID NO.39	CGTACTGGTGAAAACACCGC	Exon 21	
SEQ ID NO.40	TTCTTTCTCTTCCGCACCCA		
SEQ ID NO.41	ATCAAGTGGATGGCATTGGAATCA	Exon 22	
SEQ ID NO.42	CGTAGCTCCAGACATCACTCTG		
SEQ ID NO.43	GGGTGACCGTTTGGGAGTT	Exon 23	
SEQ ID NO.44	GAGGGAGGCGTTCTCCTTTC		
SEQ ID NO.45	GCCCAAAGTTCCGTGAGTTG	Exon 24	
SEQ ID NO.46	AATGACAAGGTAGCGCTGGG		
SEQ ID NO.47	CGAGTACCTCATCCCACAGC	Exon 25	
SEQ ID NO.48	GAGAGCTCAGGAGGGGAGTC		

SEQ ID NO.49	TGCAAAGCTGTCCCATCAAG	Exon 27	
SEQ ID NO.50	ACTGGGAGGAAGGTGTCGTC		
SEQ ID NO.51	AGTTCATCCAGGCCCAACTG	Exon 28	
SEQ ID NO.52	CAAAGCGTAAGGAGGGGGTT		
SEQ ID NO.53	GCCTGCTGAAAATGACTGAAT	Exon 2	KRAS
SEQ ID NO.54	AGCTGTATCGTCAAGGCACT		
SEQ ID NO.55	TGGAGAAACCTGTCTCTTGGA	Exon 3	
SEQ ID NO.56	TGTACTGGTCCCTCATTGCAC		
SEQ ID NO.57	TGGTGGTGTGCCAAGACATT	Exon 4	
SEQ ID NO.58	AGTCATGGGGCATGTGGAAG		
SEQ ID NO.59	TGGTGAGAGAGATCCGACAA	Exon 5	
SEQ ID NO.60	TTCACACAGCCAGGAGTCTT		
SEQ ID NO.61	TGGTGGTGTGCCAAGACATT	Exon 6	
SEQ ID NO.62	AGTCATGGGGCATGTGGAAG		
SEQ ID NO.63	GGGCTGCTCCAGTTCAATCT	Exon 1	ALK
SEQ ID NO.64	CATCAGGCGGATCCTCAGTC		
SEQ ID NO.65	AGCTCCTTGGAATCACCAACA	Exon 2	
SEQ ID NO.66	TGCGATGAGACAGGAAAGGG		
SEQ ID NO.67	TTCCCCTGTGAGCTGGAGTA	Exon 3	
SEQ ID NO.68	TGGGCATCTCCTTAGAACGC		
SEQ ID NO.69	CTCAGCTGACTCCAAGCACA	Exon 4	
SEQ ID NO.70	AGTGGGCATCAGGAGGATCT		
SEQ ID NO.71	GTGCTCCAGGGAAGAATCGG	Exon 5	
SEQ ID NO.72	ACTGCAGTTCTTCAGGGCAA		
SEQ ID NO.73	GAATGGGACAGTCCTCCAGC	Exon 6	
SEQ ID NO.74	TCTCATCTTCTCCCTGGGCA		
SEQ ID NO.75	AAGATGGCTTCTGTGGCTGG	Exon 7	
SEQ ID NO.76	GGGCATCCTTTAGGGTCCTG		
SEQ ID NO.77	CAGTACCACTGATGTCCCCG	Exon 8	
SEQ ID NO.78	AGAGCTCTTGATCGGTGCAG		
SEQ ID NO.79	ACGTGTCCTTGGTGCTAGTG	Exon 9	
SEQ ID NO.80	AGAGGCAACACCATCCACTG		
SEQ ID NO.81	TTCTGGCTGCAGATGGTCG	Exon 10	

SEQ ID NO.82	GTAGCAGTCCAGGCTGATGG		
SEQ ID NO.83	GAAGGCTCTTACGGCAAGGT	Exon 1	STK 11
SEQ ID NO.84	TGGGGATCCTTCGCAACTTC		
SEQ ID NO.85	AACTACTGAGGAGGTTACGGC	Exon 2	
SEQ ID NO.86	TGCTTCTCTTCGTTGTATAACACAT		
SEQ ID NO.87	ATGGAGTACTGCGTGTGTGG	Exon 3	
SEQ ID NO.88	GCCTGGCACACTGGGAAA		
SEQ ID NO.89	GATTGACGGCCTGGAGTACC	Exon 4	
SEQ ID NO.90	TCGGAGATTTTGAGGGTGCC		
SEQ ID NO.91	GACGACACCTGCCGGAC	Exon 5	
SEQ ID NO.92	AGATGTCCACCTTGAAGCCG		
SEQ ID NO.93	CATCACCACGGGTCTGTACC	Exon 6	
SEQ ID NO.94	TTTCAGCAGGTCAGAGAGCG		
SEQ ID NO.95	CATGACTGTGGTGCCGTACT	Exon 8	
SEQ ID NO.96	CCTCGATGTCTGAAGAGGTCC		
SEQ ID NO.97	AGAAGAGGAGGCCAGTCACA	Exon 9	
SEQ ID NO.98	GCCTCTGTGCCGTTTCATACA		
SEQ ID NO.99	CTCAGTCTTCCTGCCGGTTC	Exon 10	
SEQ ID NO.100	TTGCTCCGGGTTTCTCTGG		
SEQ ID NO.101	CCAAAGGAGACCAGGCTGTAG	Exon 1	FGFR 1
SEQ ID NO.102	GCTCGGACCGGAGAAAAGTC		
SEQ ID NO.103	AGGAGGATCGAGCTCACTGT	Exon 2	
SEQ ID NO.104	CTTCCAGCTCCACATCCCAG		
SEQ ID NO.105	AGTGGAGTCCTTCCTGGTCC	Exon 3	
SEQ ID NO.106	TTACGCAAGCATAGAGGCCG		
SEQ ID NO.107	ATGCTCTCCCCTCCTCGG	Exon 4	
SEQ ID NO.108	GCATACGGTTTGGTTTGGTGT		
SEQ ID NO.109	GGCTGCCAAGACAGTGAAGT	Exon 5	
SEQ ID NO.110	GCCATTTTCAACCAGCGCA		
SEQ ID NO.111	GTCCGTTATGCCACCTGGAG	Exon 6	
SEQ ID NO.112	TGGTTGATGCTGCCGTACTC		
SEQ ID NO.113	GGGTAGCAACGTGGAGTTCA	Exon 7	
SEQ ID NO.114	ATAAGGCAGGTTGTCTGGGC		

SEQ ID NO.115	GACGCAGGGGAGTATACGTG	Exon 8	
SEQ ID NO.116	AGAACGGTCAACCATGCAGA		
SEQ ID NO.117	CTCATCTCCTGCATGGTGGG	Exon 9	
SEQ ID NO.118	CCATCTGGCTGTGGAAGTCA		
SEQ ID NO.119	GTCTGCTGACTCCAGTGCAT	Exon 10	
SEQ ID NO.120	ACTGGAGGAGAGCCGTGAT		
SEQ ID NO.121	AACAATGCCTCCACGACCAT	Exon 2	PIK3 CA
SEQ ID NO.122	TCACGGTTGCCTACTGGTTC		
SEQ ID NO.123	CTATCGGCATGCCAGTGTGT	Exon 3	
SEQ ID NO.124	TGTGCTTTGGCAATTCTGGTG		
SEQ ID NO.125	GGCAAATAATAGTGGTGATCTGGG	Exon 4	
SEQ ID NO.126	TGTTCTGGTACACAGTCATGGT		
SEQ ID NO.127	TAATGCTTGGGAGGATGCCC	Exon 5	
SEQ ID NO.128	GGTGTAGCTGTGGAAATGCG		
SEQ ID NO.129	GTATCTACCATGGAGGAGAACCC	Exon 6	
SEQ ID NO.130	CTGGGATTGGAACAAGGTACT		
SEQ ID NO.131	TCCTGATCTTCCTCGTGCTG	Exon 7	
SEQ ID NO.132	TTTAGCACCCCTTTCGGCCTT		
SEQ ID NO.133	TGTCCATTGGCATGGGGAAA	Exon 8	
SEQ ID NO.134	CCATGAGGTACTGGCCAAAGA		
SEQ ID NO.135	GACTGGTTCAGCAGTGTGGT	Exon 9	
SEQ ID NO.136	CAGTCCTGCGTGGGAATAGC		
SEQ ID NO.137	GGGAAAATGACAAAGAACAGCTCA	Exon 10	
SEQ ID NO.138	TCTCCTGCTCAGTGATTCAGA		
SEQ ID NO. 139	CCACCTCAGCCCATCTTGAC	Exon 2	ARAF
SEQ ID NO. 140	GGTATACTTTGACGGTGCCCA		
SEQ ID NO. 141	CTGTCCGGGATGGCATGAG	Exon 3	
SEQ ID NO. 142	GGTAGACCACACAGCAGTCC		
SEQ ID NO. 143	GAAAGACGGTCACTGCCTGG	Exon 4	
SEQ ID NO. 144	GGTCAGCGGGACATCTTCAA		
SEQ ID NO. 145	CTTCTTCAGCCTGGCGTTCT	Exon 5	
SEQ ID NO. 146	CACAGGTTTGGCAACGGAAG		
SEQ ID NO. 147	GTCCAGGATTTGTCCGGAGG	Exon 6	

SEQ ID NO. 148	TTAGCAACTCATTTCAGGGGGGC		
SEQ ID NO. 149	CTGTGACCCGGAGCACTT	Exon 7	
SEQ ID NO. 150	GACGTTGGGAGTGGACGTG		
SEQ ID NO. 151	GAGGTAGTGATGGAACCCCC	Exon 9	
SEQ ID NO. 152	GCTCTGCTGGTGACTTGGA		
SEQ ID NO. 153	CTCGTTTGGCACCGTGTTTC	Exon 10	
SEQ ID NO. 154	TTCTTGAAAGCCTGGGCCTG		
SEQ ID NO. 155	CGGGATTTGCCATCATCACAC	Exon 11	
SEQ ID NO. 156	TGAGCTGGACCATGTCTGAAG		
SEQ ID NO. 157	CTTCCTACATGAGGGGCTCAC	Exon 13	
SEQ ID NO. 158	CGCTCCATCGAGTCTTCACT		
SEQ ID NO. 159	AGCTGAGGTGATCCGTATGC	Exon 14	
SEQ ID NO. 160	AAGCTCGTAGAGCACAACCC		
SEQ ID NO. 161	TATGGTGGGCCGTGGCTAT	Exon 15	
SEQ ID NO. 162	GCTGGAACCTTGAGGCAGTC		
SEQ ID NO. 163	CACCAGGGAGCCAATCTCAG	Exon 16	
SEQ ID NO. 164	GGGTCCCCCAATTCCAGAAG		
SEQ ID NO. 165	CGTGAGCAGGACGGTGATAA	Exon 1	
SEQ ID NO. 166	CTCGGATCGGGTTTGGGAAA		
SEQ ID NO. 167	GCCTCCCCGATTGAAAGAGA	Exon 2	
SEQ ID NO. 168	ACTGGTTTCACACCGAAGGA		
SEQ ID NO. 169	ACAAAGCATCACTGGCTGATT	Exon 3	
SEQ ID NO. 170	TCCACGATGGTGATATTGGCA		
SEQ ID NO. 171	CCACTGGGACAAGCCATCTT	Exon 6	
SEQ ID NO. 172	TTCACCATGAAGCACTCCCC		
SEQ ID NO. 173	GAGTGCTGACCATAACCGGC	Exon 8	
SEQ ID NO. 174	CAGTAGGCCACCACACACAT		
SEQ ID NO. 175	GCGGAAAAAGCTGCATGACC	Exon 9	
SEQ ID NO. 176	TAGGATGGTGAGGCCCATTG		
SEQ ID NO. 177	AAACGTCATCTCCAGTGAGCAT	Exon 10	
SEQ ID NO. 178	GTGGCTAGGAGTCTGGGTGA		
SEQ ID NO. 179	GGAGCAACGGACACACTGAA	Exon 11	
SEQ ID NO. 180	TCTCGGTAGGAATCAGGGGT		

SEQ ID NO. 181	GATGCCTGTCTGCAGGTTCT	Exon 12	RRM 1
SEQ ID NO. 182	GACCCGGGAAACAAAGTTGC		
SEQ ID NO. 183	AAAGTGCTGTCTGGCTCCAA	Exon 1	
SEQ ID NO. 184	CGGGATCGAGAAGGTGACAG		
SEQ ID NO. 185	CGCCAAGAACGAGTCATGTT	Exon 2	
SEQ ID NO. 186	TCCATATTGAGTCCATAACAAAGC		
SEQ ID NO. 187	TCCAAGGCTTGTACAGTGGG	Exon 3	
SEQ ID NO. 188	GGCTGCCAGGATAGCATAGT		
SEQ ID NO. 189	GTGATGGAAGACCTCTATAACTACA	Exon 4	
SEQ ID NO. 190	CCAATGTTGACTTGGCCACC		
SEQ ID NO. 191	TGGCTGAAAGACCACAACAT	Exon 7	
SEQ ID NO. 192	AAAGTTGTGGGCGGTTGGTA		
SEQ ID NO. 193	AGATGACAGCATTGAAGGCA	Exon 8	
SEQ ID NO. 194	AATGTAGCTGCCAGTAGCCC		
SEQ ID NO. 195	TTCCAATGGCCTTGTACCGA	Exon 9	
SEQ ID NO. 196	TGTTCCACCTTGATCCACA		
SEQ ID NO. 197	CGTCCTGGGGCATTGCTA	Exon 10	
SEQ ID NO. 198	GAAAAGATCTCTGGCACGCT		
SEQ ID NO. 199	GACTGGTCTTTGATGTGTCCA	Exon 11	
SEQ ID NO. 200	AGTTTCTCAAATTCCTCTCCCCA		
SEQ ID NO. 201	TCTCAGACGGAAACAGGCAC	Exon 12	
SEQ ID NO. 202	GCACAGGTTGCTGCATTTGA		
SEQ ID NO. 203	CTTCCCTGGCCCTGAATATGT	Exon 13	
SEQ ID NO. 204	TCAAGTTTCGGACAACGACTT		
SEQ ID NO. 205	CAAATAAACGCCATCGCCCC	Exon 14	
SEQ ID NO. 206	CTGGGCTTCTGCACTCTCAA		
SEQ ID NO. 207	TATTGCCCCGATGCCTACAG	Exon 16	
SEQ ID NO. 208	TCTCCTGACAAGACTCTGCGA		
SEQ ID NO. 209	CAGCTGAGAGAGGTGCTTTC	Exon 18	
SEQ ID NO. 210	CTTCCAGCCGTAGAAGTGCAT		
SEQ ID NO. 211	ACAAGACCAGCGGCTAATCC	Exon 19	
SEQ ID NO. 212	CTTTTGCCTGCAACAGGGTC		
SEQ ID NO. 213	TGTGCCTAACAGAGGTGTCC	Exon 1	

SEQ ID NO. 214	ACACAGTCAAAGGGAAGATGTGA		HAV CR2
SEQ ID NO. 215	GAATACAGAGCGGAGGTCGG	Exon 2	
SEQ ID NO. 216	AGGGACACATCTCCTTTGCG		
SEQ ID NO. 217	GGGACTCTGGAGCAACCATC	Exon 5	
SEQ ID NO. 218	GATAAGAGCCAGAGCCAGCC		
SEQ ID NO. 219	CTGGGACCTGCACTGAACTT	Exon 7	
SEQ ID NO. 220	CTTTCACCTCAGCACCCAGT		
SEQ ID NO. 221	CTTTGTGAGCAGATGCGGAG	Exon 1	MET
SEQ ID NO. 222	GCCTCCTCTCAGCAAGTCAG		
SEQ ID NO. 223	TGCAGCAGCAAAGCCAATTT	Exon 2	
SEQ ID NO. 224	AAAGGACTTTGGCTCCCAGG		
SEQ ID NO. 225	CTTTGCAGCGCGTTGACTTA	Exon 3	
SEQ ID NO. 226	CGACCCTCTGATGTCCCAAG		
SEQ ID NO. 227	TTCTCCTGGACTCCCATCCA	Exon 4	
SEQ ID NO. 228	AACCAGTGTGTAGCCATTTTGG		
SEQ ID NO. 229	CCCCACCCTTTGTTCAGTGT	Exon 5	
SEQ ID NO. 230	TCTGTTGAGTCCATGTCCCG		
SEQ ID NO. 231	CCCCTTGAAGGAGGGACAAG	Exon 6	
SEQ ID NO. 232	AAGGTGCAGCTCTCATTTCCA		
SEQ ID NO. 233	AAATGCACAGTTGGTCCTGC	Exon 7	
SEQ ID NO. 234	TTGTCCCGTGGCCATTTGAA		
SEQ ID NO. 235	TCGCCGAAATACGGTCCTATG	Exon 8	
SEQ ID NO. 236	TGTGTCTAGAATTCCCAGTGT		
SEQ ID NO. 237	TGTTATACCCCAGCCCAAACC	Exon 9	
SEQ ID NO. 238	TGTCTCTCGGTTGGCTAAGTC		
SEQ ID NO. 239	ACTTGGTGGAAAGAACCTCTCA	Exon 10	
SEQ ID NO. 240	ACCTGTTATTGTGCTCCCACC		
SEQ ID NO. 241	AGCTGAATCTGCAACTCCCC	Exon 11	
SEQ ID NO. 242	TCCAGTACATTTTCATTGCCCA		
SEQ ID NO. 243	GACCCTGAAGCAGTTAAAGGTG	Exon 12	
SEQ ID NO. 244	GTCATTGGGGACCGTGCATA		
SEQ ID NO. 245	TCACAGGATTGATTGCTGGTGT	Exon 13	
SEQ ID NO. 246	CAGCCACAGGAAAAACCCAAG		

SEQ ID NO. 247	AGTTCGCTACGATGCAAGAGT	Exon 14	CDK4
SEQ ID NO. 248	AGTTGGGCTTACACTTCGGG		
SEQ ID NO. 249	GTTTCATGCCGACAAGTGCAG	Exon 15	
SEQ ID NO. 250	TGCACAATCAGGCTACTGGG		
SEQ ID NO. 251	TGGTTGTGTATATCATGGGACTT	Exon 16	
SEQ ID NO. 252	TCAAGGATTTCACAGCACAGT		
SEQ ID NO. 253	TCATCCCAATGTCCTCTCGC	Exon 17	
SEQ ID NO. 254	GTATGGTAGGACCACCAGCG		
SEQ ID NO. 255	GGTCTTCAAGTAGCCAAAGGCA	Exon 18	
SEQ ID NO. 256	TTCTTGCAGCCAAGTCTCTGT		
SEQ ID NO. 257	GTTGCTGATTTTGGTCTTGCCA	Exon 19	
SEQ ID NO. 258	CTTCACTGGCAGCTTTGCAC		
SEQ ID NO. 259	AGAGGAGCCCCACCTTATCC	Exon 20	
SEQ ID NO. 260	AAGGGGTCTGGGCAGTATTC		
SEQ ID NO. 261	ATGCGCCCATCCTTTTCTGA	Exon 21	
SEQ ID NO. 262	AATGAGTGGCCTGTTCTGGG		
SEQ ID NO. 263	CTGGTCACATGGTGAGGGTG	Exon 1	
SEQ ID NO. 264	GGTTCCTACGGCCCCATACA		
SEQ ID NO. 265	CAGTGTACAAGGCCCGTGAT	Exon 2	
SEQ ID NO. 266	CAGTCGCCTCAGTAAAGCCA		
SEQ ID NO. 267	CGAACTGACCGGGAGATCAA	Exon 3	
SEQ ID NO. 268	CTTGATCGTTTCGGCTGGCA		
SEQ ID NO. 269	TGCCAATTGCATCGTTCACC	Exon 4	
SEQ ID NO. 270	CCAGCTTGACTGTTCCACCA		
SEQ ID NO. 271	GCAGTCCACATATGCAACACC	Exon 5	
SEQ ID NO. 272	TTCGACGAAACATCTCTGCAA		
SEQ ID NO. 273	GGATGACTGGCCTCGAGATG	Exon 7	
SEQ ID NO. 274	CTCAGGTACCACCGACTGC		
SEQ ID NO. 275	ACACAAGCGAATCTCTGCCT	Exon 8	
SEQ ID NO. 276	AGCCACTCCATTGCTCACTC		
SEQ ID NO. 277	GTAGCCCCTTGGTTTCCGTG	Exon 1	BRC A1
SEQ ID NO. 278	CGCGCAGTCGCAGTTTAAAT		
SEQ ID NO. 279	TGGATTTATCTGCTCTTCGCGT	Exon 2	

SEQ ID NO. 280	TGGGACACTCTAAGATTTTCTGC		CCN D1
SEQ ID NO. 281	GAGCCTACAAGAAAGTACGAGA	Exon 6	
SEQ ID NO. 282	CTCCAAACCTGTGTCAAGCTG		
SEQ ID NO. 283	GGAAACCAGTCTCAGTGTCCA	Exon 7	
SEQ ID NO. 284	TGTATCCGCTGCTTTGTCCT		
SEQ ID NO. 285	CCGAAGAGGGGGCCAAGAAAT	Exon 10	
SEQ ID NO. 286	TCCAAGCCCGTTCCTCTTTC		
SEQ ID NO. 287	GAAGCAGCATCTGGGTGTGA	Exon 11	
SEQ ID NO. 288	GTCACTCTGAGAGGATAGCCC		
SEQ ID NO. 289	CTCCAGCAGGAAATGGCTGA	Exon 12	
SEQ ID NO. 290	ATTTCGCAGGTCCTCAAGGG		
SEQ ID NO. 291	GGCCTTTCTGCTGACAAGTT	Exon 13	
SEQ ID NO. 292	CTTTCCACTCCTGGTTCTTTATTT		
SEQ ID NO. 293	GGTGGTACATGCACAGTTGC	Exon 14	
SEQ ID NO. 294	AGCTGTTGCTCCTCCACATC		
SEQ ID NO. 295	TCTGAAGACAGAGCCCCAGA	Exon 15	
SEQ ID NO. 296	GCTTCTCCCTGCTCACACTT		
SEQ ID NO. 297	GCTCGTGTACAAGTTTGCCA	Exon 16	
SEQ ID NO. 298	ACAACATGAGTAGTCTCTTCAGTAA		
SEQ ID NO. 299	GCTGAGTTTGTGTGTGAACGG	Exon 17	
SEQ ID NO. 300	AGCTAACTACCCATTTTCCTCCC		
SEQ ID NO. 301	TTGAAGTCAGAGGAGATGTGGT	Exon 19	
SEQ ID NO. 302	TCCTGGGATTCTCTTGCTCG		
SEQ ID NO. 303	CAACTGGAATGGATGGTACAGC	Exon 21	
SEQ ID NO. 304	TGTGCCAAGGGTGAATGATGA		
SEQ ID NO. 305	CTGGGAGCTCCTCTCACTCT	Exon 23	
SEQ ID NO. 306	CCTCCCCAATGTTCCACTCC		
SEQ ID NO. 307	ACACGGACTACAGGGGAGTT	Exon 1	
SEQ ID NO. 308	ACACGGACTACAGGGGAGTT		
SEQ ID NO. 309	TGGAGCCCGTGAAAAAGAGC	Exon 2	
SEQ ID NO. 310	CGGATGGAGTTGTCGGTGT		
SEQ ID NO. 311	CAATGACCCCGCACGATTTC	Exon 3	
SEQ ID NO. 312	CAGAGGGCAACGAAGGTCTG		

SEQ ID NO. 313	TGAGGAGCCCCAACAACCTTC	Exon 4	PDGF RA
SEQ ID NO. 314	CGGGTCACACTTGATCACTCT		
SEQ ID NO. 315	CTGATTGGACAGGCATGGGT	Exon 5	
SEQ ID NO. 316	GTGCCTGGAAGTCAACGGTA		
SEQ ID NO. 317	AAGAGCAAAAAGCGAAGGCG	Exon 1	
SEQ ID NO. 318	TCCGCAATGAATGTCCCACA		
SEQ ID NO. 319	GAGTGAAGTGAGCTGGCAGT	Exon 3	
SEQ ID NO. 320	GTCACAAAAAGGCCGCTGTT		
SEQ ID NO. 321	GCACAACCTGATCCCGAGACT	Exon 4	
SEQ ID NO. 322	TAAAGCCCTGTCTGCTGTCTG		
SEQ ID NO. 323	CGATTGTGGTCACCTGTGCT	Exon 5	
SEQ ID NO. 324	CACTTCTCCAGGGTAAGTCCA		
SEQ ID NO. 325	GCAAAGGCATCACAATGCTG	Exon 6	
SEQ ID NO. 326	TCGGGGACCGTCAAAGTGTA		
SEQ ID NO. 327	TTCAGCCAGTTGGAAGCTGT	Exon 7	
SEQ ID NO. 328	GGATATCCTGGGAGGTGGGTA		
SEQ ID NO. 329	AGCTGATCCGTGCTAAGGAAG	Exon 8	
SEQ ID NO. 330	ATAGCTCTTCACAGCATCTTCA		
SEQ ID NO. 331	TGATCACCATGGCTCAACTGG	Exon 9	
SEQ ID NO. 332	CAATATCAGGAAGCGGCGTG		
SEQ ID NO. 333	ACTCCCGAGACAGGAGTACC	Exon 10	
SEQ ID NO. 334	CTCGGTTCTCAGCTCCAAGG		
SEQ ID NO. 335	TCTGAACTCACGGTGGCTG	Exon 11	
SEQ ID NO. 336	GTTTCCAAATGACAACCAGGACA		
SEQ ID NO. 337	TGAAATTCGCTGGAGGGTCA	Exon 12	
SEQ ID NO. 338	GCTGCATCGGGTCCACATAA		
SEQ ID NO. 339	GTCTTGGGGTCTGGAGCGTT	Exon 13	
SEQ ID NO. 340	CAGGTTGGGACCGGCTTAAT		
SEQ ID NO. 341	CACGGCCAGATCCAGTGAAA	Exon 14	
SEQ ID NO. 342	CACGGCCAGATCCAGTGAAA		
SEQ ID NO. 343	AGGGATAGCTTCCTGAGCCA	Exon 15	
SEQ ID NO. 344	TGTGCTTTCATCAGCAGGGT		
SEQ ID NO. 345	CATGGACATGAAGCAGGCTGA	Exon 16	

SEQ ID NO. 346	AGGCTGGACGATCATAGAGTG			
SEQ ID NO. 347	ACTCAGAAGTCAAAAACCTCCT	Exon 17		
SEQ ID NO. 348	CTCCATTCTCTCGGGCAACTT			
SEQ ID NO. 349	GTGATCTGGCTGCTCGCAA	Exon 18		
SEQ ID NO. 350	GCCAGGCCAAAGTCACAGAT			
SEQ ID NO. 351	CTTTCTGCCCCGTGAAGTGGA	Exon 19		
SEQ ID NO. 352	ATCTCCCAGAGCAGAATGCC			
SEQ ID NO. 353	CCGGCATGATGGTGGATTCTA	Exon 20		
SEQ ID NO. 354	CTTCACTGGTAGCGTGGTCA			
SEQ ID NO. 355	AACAGTGAGCCGGAGAAGAG	Exon 21		
SEQ ID NO. 356	ATTGTCCAGGCAGCAGATTCTC			
SEQ ID NO. 357	GCATGCGTGTGGACTCAGA	Exon 22		
SEQ ID NO. 358	CACTGTCAGCGCTCAGTCTC			
SEQ ID NO. 359	ACTTTGGAACAGGGTTGGCA	Exon 23		
SEQ ID NO. 360	GCCAATGAACCCACACCAAC			
SEQ ID NO. 361	CAGTGCCAAATCCGGGGAG	Exon 1		AXL
SEQ ID NO. 362	TGGGTGCCAAACTTTCCTCA			
SEQ ID NO. 363	GCAGGCTGAAGAAAGTCCCT	Exon 2		
SEQ ID NO. 364	GAGCTGGCTGACCACTATCC			
SEQ ID NO. 365	AATCACCTCCCTGCAGCTTTC	Exon 3		
SEQ ID NO. 366	ACACGAAGGTCTGATGTCCC			
SEQ ID NO. 367	CAACACCCCCTTCAACCTGAG	Exon 4		
SEQ ID NO. 368	CATCCTGGAGCCAGAGTAGGT			
SEQ ID NO. 369	GGCTGAACAAGACATCCTCTTT	Exon 5		
SEQ ID NO. 370	GTGGCTGTGCGGGATGT			
SEQ ID NO. 371	TAACCTCCACCTGGTCTCCC	Exon 6		
SEQ ID NO. 372	AGTGGGTCAGGGGGTAGATG			
SEQ ID NO. 373	CTGTCAGACGATGGGATGGG	Exon 7		
SEQ ID NO. 374	ATGCCACGCGGATGTGATAA			
SEQ ID NO. 375	CCTTCGTGCATTGGCAAGAG	Exon 8		
SEQ ID NO. 376	GTGTCCTGGCCTTGATACGC			
SEQ ID NO. 377	CGGGTCTGTGTCCAATCTGA	Exon 9		
SEQ ID NO. 378	CAGGCCTCCAGGGGTACT			

SEQ ID NO. 379	GGCCCTGGTGGTATGTACTG	Exon 11	
SEQ ID NO. 380	GGTGGACAAGGAAGAGAGCC		
SEQ ID NO. 381	AGTGGAAGAGGTGAACTGGT	Exon 12	
SEQ ID NO. 382	GTAGCTTCAGTGGTCCGACG		
SEQ ID NO. 383	TGGGCATCAGTGAAGAGCTG	Exon 13	
SEQ ID NO. 384	TCTCCCAGAGTCTTCCCCAG		
SEQ ID NO. 385	GAGTTTGGAGCTGTGATGGAAGG	Exon 14	
SEQ ID NO. 386	TCGTCTTCACAGCCACCTTGA		
SEQ ID NO. 387	CCATCTGCACGAGGTCAGAG	Exon 15	
SEQ ID NO. 388	GAGCCTCATGACGTTGGGAT		
SEQ ID NO. 389	TCCAGGGTTCTGAACGAGAGA	Exon 16	
SEQ ID NO. 390	AGAGGAGGAAGCTGTGTAGGT		
SEQ ID NO. 391	ACCTGCCCCACTCAGATGCTA	Exon 17	
SEQ ID NO. 392	CCAGGTCCCGGTGTATGAAT		
SEQ ID NO. 393	AGGGACGTATCGCCAAGATG	Exon 18	
SEQ ID NO. 394	CTGGTGTAGACACGGTCAGC		
SEQ ID NO. 395	TCCTTCGGGGTGACAATGTG	Exon 19	
SEQ ID NO. 396	CGATTTCCTGGCGCAGATA		
SEQ ID NO. 397	CATCCTGCTGGACGCTATGT	Exon 20	
SEQ ID NO. 398	AGTACCAGGTGGAGGGTTGT		
SEQ ID NO. 399	CACGGGTAAGACGCAACTGA	Exon 1	AUR KA
SEQ ID NO. 400	AAACCCAATCAGGCCTACCG		
SEQ ID NO. 401	TCTTCCCAGCGCATTCCTTT	Exon 3	
SEQ ID NO. 402	TGGTTGCCTGCAATTGCTTC		
SEQ ID NO. 403	TGAAATTGGTCGCCCTCTGG	Exon 5	
SEQ ID NO. 404	CTGAGCTGATGCTCCACTCC		
SEQ ID NO. 405	TGCACCACTTGGAACAGTTT	Exon 6	
SEQ ID NO. 406	AGTAGCAGTTCTCTGCTCATCA		
SEQ ID NO. 407	TGGCAAATGCCCTGTCTTACT	Exon 7	
SEQ ID NO. 408	GGAGCATGTACTGACCACCC		
SEQ ID NO. 409	CTCTGTGGCACCCTGGACTA	Exon 8	
SEQ ID NO. 410	TCCAAGGCTCCAGAGATCCA		
SEQ ID NO. 411	ACAGGAGGCAAATCCAGAGC	Exon 9	

SEQ ID NO. 412	AAACCCAATCAGGCCTACCG		
SEQ ID NO. 413	GACGGCCGCGGGAAAAG	Exon 1	TYM S
SEQ ID NO. 414	ATGCCGAATACCGACAGGGT		
SEQ ID NO. 415	ATTCCCTCTGCTGACAACCA	Exon 2	
SEQ ID NO. 416	CTTGATAAACCACAGCAACTCCT		
SEQ ID NO. 417	ATGCCAATGGATCCCGAGAC	Exon 3	
SEQ ID NO. 418	ATAAACTGGGCCCAAGTCCC		
SEQ ID NO. 419	TCAGGACAGGGAGTTGACCA	Exon 4	
SEQ ID NO. 420	TGGATTCCAAGCGCACATGA		
SEQ ID NO. 421	ATCTTCCTCTGATGGCGCTG	Exon 5	
SEQ ID NO. 422	GGACAGCTCACTGTTCACCA		
SEQ ID NO. 423	CAGGTGACTTTATACACACTTTGGG	Exon 6	
SEQ ID NO. 424	TGAATTTTCAGTGGCTCGATGTG		
SEQ ID NO. 425	CAGCGAGAACCCAGACCTTT	Exon 7	
SEQ ID NO. 426	CCAGCCCAACCCCTAAAGAC		
SEQ ID NO. 427	TTCAAGAATGGGGTCGGCTC	Exon 1	ETS2
SEQ ID NO. 428	CTTTGGAATTCCGCAGCGAC		
SEQ ID NO. 429	GATGTATGCCTACCACCGGG	Exon 1	DDR2
SEQ ID NO. 430	AGGGGTGCACAGTTTTTGAA		
SEQ ID NO. 431	CTCCATCAAGGGAGACCTGC	Exon 2	
SEQ ID NO. 432	CAAAAACAGCTGGCCTCTGC		
SEQ ID NO. 433	ACTCCAGTTCCAACACCATCT	Exon 3	
SEQ ID NO. 434	ATAGGCAGCAGCAGGAACAG		
SEQ ID NO. 435	GCATGTCAGGAGGCCAGATT	Exon 4	
SEQ ID NO. 436	ATTTGGCAGCTGTGGACTCT		
SEQ ID NO. 437	AGAAGAAGGGGATGGAGCCT	Exon 5	
SEQ ID NO. 438	GCAAACCTCGATGCCATGACC		
SEQ ID NO. 439	CGGTTCATTCCAGTCACCGA	Exon 6	
SEQ ID NO. 440	AGCCAGACACAGCCGTAAAG		
SEQ ID NO. 441	GGCAGCAGTTTGTACTCCCT	Exon 7	
SEQ ID NO. 442	TCCAACAGCTCCATCATAGACA		
SEQ ID NO. 443	TTGACCGATGGTGTGTCTGG	Exon 8	
SEQ ID NO. 444	GTAGCCATTGGTGGCACTCT		

SEQ ID NO. 445	ACGGGCTCCTGGTCTACTTC	Exon 2	ERCC 2
SEQ ID NO. 446	CGTTTGAGCTCCCGCATGTA		
SEQ ID NO. 447	CTATGAGAAGCAGGAGGGCG	Exon 5	
SEQ ID NO. 448	ACACAAGTTTTTGCGGGAGC		
SEQ ID NO. 449	TTTGGGAAGGACGTCGATGG	Exon 6	
SEQ ID NO. 450	GCTGGTGTCATGCTGGTACT		
SEQ ID NO. 451	GCTGGCATCTACAACCTGGA	Exon 7	
SEQ ID NO. 452	AGCAAGGAAGTATGGGCACC		
SEQ ID NO. 453	GCATGCCAATGTGGTGGTTT	Exon 8	
SEQ ID NO. 454	CCAGTTCCTTGGACACCAGG		
SEQ ID NO. 455	TCGACTCCATGAGCGTCAAC	Exon 9	
SEQ ID NO. 456	TTCTGCAGGGTCTCCAGGTT		
SEQ ID NO. 457	AAAGAGACAGACGAGCAGCG	Exon 10	
SEQ ID NO. 458	ACTTCGTCGGGCAGCAC		
SEQ ID NO. 459	GAGCATTTCTGGGCTTCCT	Exon 11	
SEQ ID NO. 460	CTCTCCTGCACCACATGCT		
SEQ ID NO. 461	TCACCGACCTTGCTGACTTC	Exon 12	
SEQ ID NO. 462	TTTGGCGTAGGTGCTGACAA		
SEQ ID NO. 463	GCTTCACCATCATCATCGAGC	Exon 13	
SEQ ID NO. 464	CTGAAGTGCAGGATGGGGTT		
SEQ ID NO. 465	ATTCGAGTAGCGGCTCTTCC	Exon 1	ABC B1
SEQ ID NO. 466	CTTGGAACGGCCACCAAGA		
SEQ ID NO. 467	GGGAACCTTTGGCTGCCATCA	Exon 4	
SEQ ID NO. 468	TGTCATTTCTCCAAACACCAGC		
SEQ ID NO. 469	CCTGGCAGCTGGAAGACAAA	Exon 6	
SEQ ID NO. 470	GGGTGTTAAGCTCCCCAACA		
SEQ ID NO. 471	GGATTTACACGTGGTTGGAAGC	Exon 7	
SEQ ID NO. 472	CAGCAGCTGACAGTCCAAGA		
SEQ ID NO. 473	AGCGTATGCAAAAGCTGGAG	Exon 8	
SEQ ID NO. 474	TGTCCTCCAAATGCAATCACAG		
SEQ ID NO. 475	TAGGTGCTGCTTTCCTGCTG	Exon 9	
SEQ ID NO. 476	TTCCCCTGAGAGGACCAAGG		
SEQ ID NO. 477	AACCTGAAGGTGCAGAGTGG	Exon 12	

SEQ ID NO. 478	TCCCCTCTGTGGGGTCATAG		Exon 13
SEQ ID NO. 479	AACATTCGCTATGGCCGTGA		
SEQ ID NO. 480	AGGCATTGGCTTCCTTGACA		
SEQ ID NO. 481	TGACACCCTGGTTGGAGAGA	Exon 14	Exon 15
SEQ ID NO. 482	GAGGATCTTGGGGTTGCGAA		
SEQ ID NO. 483	CAGAAAAGGTCGGACCACCA		
SEQ ID NO. 484	CATCGAAACCAGCGATGACG		
SEQ ID NO. 485	TGCTCAAGACTGGCGCTAAA	Exon 1	TP53
SEQ ID NO. 486	CAATCCAGGGAAGCGTGTCA		
SEQ ID NO. 487	CTGCCTTCCGGGTCACTG	Exon 2a	
SEQ ID NO. 488	GTTTCCTGACTCAGAGGGGG		
SEQ ID NO. 489	TGGATCCATTGGAAGGGCAG	Exon 2b	
SEQ ID NO. 490	TTTCGCTTCCCACAGGTCTC		
SEQ ID NO. 491	CTGCCTTCCGGGTCACTG	Exon 2c	
SEQ ID NO. 492	GTTTCCTGACTCAGAGGGGG		
SEQ ID NO. 493	TGAAGCTCCCAGAATGCCAG	Exon 3	
SEQ ID NO. 494	GCTGCCCTGGTAGGTTTTCT		
SEQ ID NO. 495	GCCAACTCTCTCTAGCTCGC	Exon 4a	
SEQ ID NO. 496	CTCCGTCATGTGCTGTGACT		
SEQ ID NO. 497	CCCTGCCCTCAACAAGATGT	Exon 4b	
SEQ ID NO. 498	CTCCGTCATGTGCTGTGACT		
SEQ ID NO. 499	GCCCCTCCTCAGCATCTTATC	Exon 5	
SEQ ID NO. 500	CTCATAGGGCACCACCACAC		
SEQ ID NO. 501	TGGCTCTGACTGTACCACCA	Exon 6	
SEQ ID NO. 502	TGATGATGGTGAGGATGGGC		
SEQ ID NO. 503	TGGGACGGAACAGCTTTGAG	Exon 7	
SEQ ID NO. 504	ATTCTCTTCCTCTGTGCGCC		
SEQ ID NO. 505	CACTGCCCAACAACACCAG	Exon 8	
SEQ ID NO. 506	AGGGTGAAATATTCTCCATCCAGT		
SEQ ID NO. 507	GACCAGACCAGCTTTCAAAAAGA	Exon 9a	
SEQ ID NO. 508	TATCAGGCAAAGTCATAGAACCA		
SEQ ID NO. 509	CGCTTCGAGATGTTCCGAGA	Exon 10	
SEQ ID NO. 510	CTGGCTCCTTCCCAGCCT		

SEQ ID NO. 511	TGCATTTTTCACCCCACCCTT	Exon 11	
SEQ ID NO. 512	GACACACAGGTGGCAGCAAA		
SEQ ID NO. 513	CCCAACACACGGTCCACA		NFE2 L2
SEQ ID NO. 514	CTGCTGGGACGGGAGTC		
SEQ ID NO. 515	AGAGACTCTCCGAGGCGG		PDCD 4
SEQ ID NO. 516	CTGCTCGCCCAGGTTCC		
SEQ ID NO. 517	CTGAGTGCCCCGGAACG	Exon 1	RET
SEQ ID NO. 518	CTGGAGGGACTGCGGCTA		
SEQ ID NO. 519	CTTCCTGTCACCCACATCCC	Exon 3	
SEQ ID NO. 520	AGGCTGGAAAGGAGGTGTTG		
SEQ ID NO. 521	GTGTACGACGAGGACGACTC	Exon 4	
SEQ ID NO. 522	TCCTTCCGCTTGA ACTCCAC		
SEQ ID NO. 523	CAGACGTGGTACCTGCATCA	Exon 5	
SEQ ID NO. 524	GTTCCACCCGGAAGGTCTG		
SEQ ID NO. 525	TCTCAACCGGAACCTCTCCA	Exon 6	
SEQ ID NO. 526	GAGGGAGTAGGTACTGGGCA		
SEQ ID NO. 527	TTCAGTGGCATCAACGTCCA	Exon 7	
SEQ ID NO. 528	GGCCACCACCATGTAGTGAA		
SEQ ID NO. 529	TCCTGTGCAGTCAGCAAGAG	Exon 8	
SEQ ID NO. 530	CCATCTCCTTGCCTCCACTC		
SEQ ID NO. 531	CCAGGAACTTCTCCACCTGC	Exon 9	
SEQ ID NO. 532	GGTCTCCACAACATCGCAGT		
SEQ ID NO. 533	ATTGTTGGGGGACACGAGC	Exon 10	
SEQ ID NO. 534	CCTCCTCAGGGAAGCAGTTG		
SEQ ID NO. 535	GTCTGCCTTCTGCATCCACT	Exon 11	
SEQ ID NO. 536	TTGAAGGCATCCACGGAGAC		
SEQ ID NO. 537	AAAAGTGGTCAAGGCAACGG	Exon 12	
SEQ ID NO. 538	TTTCAGCATCTTCACGGCCA		
SEQ ID NO. 539	CGAGTGAGCTTCGAGACCTG	Exon 13	
SEQ ID NO. 540	TGCAGGCCCCATACAATTTGA		
SEQ ID NO. 541	AGTACGCCAAATACGGCTCC	Exon 14	
SEQ ID NO. 542	GCCAGATACTGCATCCCCTG		
SEQ ID NO. 543	TTGGCAGCCAGAAACATCCT	Exon 15	

SEQ ID NO. 544	TCTCGGGACAAGCCGAAATC		BARF
SEQ ID NO. 545	ATCGTGACCCTAGGGGGAAA	Exon 17	
SEQ ID NO. 546	GCCGGTCTTCAGAAGGTTGA		
SEQ ID NO. 547	TGATGCTGCAATGCTGGAAG	Exon 18	
SEQ ID NO. 548	TCCAGGTCTTTGCTGATGTCC		
SEQ ID NO. 549	TCTCAGAGGAGGAGACACCG	Exon 19b	
SEQ ID NO. 550	ACCATCCTAAGTTGCTGGGC		
SEQ ID NO. 551	GGCCTGGAGAGAGTCCTGTA	Exon 20	
SEQ ID NO. 552	CTTGTGAGCCCTCTTCCCTG		
SEQ ID NO. 553	TATAAGATGGCGGCGCTGAG	Exon 1	
SEQ ID NO. 554	CCATGTCCCCGTTGAACAGA		
SEQ ID NO. 555	AAATGATTAAGTTGACACAGGAACA	Exon 2	
SEQ ID NO. 556	GGTGGATTATGCTCCCCACC		
SEQ ID NO. 557	TCTCTGGGGAACGGAAGTGA	Exon 3	
SEQ ID NO. 558	TCCTCTGTTTGTTGGGCAGG		
SEQ ID NO. 559	GTACCTGCAAGGTGTGGAGT	Exon 4	
SEQ ID NO. 560	CAGCACAGCACTCTGGGATT		
SEQ ID NO. 561	CCAATTGGTTGGGACACTGA	Exon 5	
SEQ ID NO. 562	TGGAACATTCTCCAACACTTCCA		
SEQ ID NO. 563	GCTGCTTTTCCAGGGTTTCC	Exon 6	
SEQ ID NO. 564	AACTTCTGTACTACAACGCTGGT		
SEQ ID NO. 565	AGAGGCGTCCTTAGCAGAGA	Exon 7	
SEQ ID NO. 566	AATAGAGTCCGAGGCGGGT		
SEQ ID NO. 567	TCCAATTCCACAGCCCTTCC	Exon 8	
SEQ ID NO. 568	GCTGATGAGGATCGGTCTCG		
SEQ ID NO. 569	TGCCTCATTACCTGGCTCAC	Exon 10	
SEQ ID NO. 570	CCTTTCTCGCTGAGGTCCTG		
SEQ ID NO. 571	AGACGGGACTCGAGTGATGA	Exon 11	
SEQ ID NO. 572	TGCCACTTTCCCTTGTAAGT		
SEQ ID NO. 573	GTGATGTGGCAGTGAAAATGT	Exon 12	
SEQ ID NO. 574	TCATTTTTGAAGGCTTGTAAGT		
SEQ ID NO. 575	GTTACCCAGTGGTGTGAGGG	Exon 13	
SEQ ID NO. 576	TGTGCAGTCTGTCTGCAAT		

SEQ ID NO. 577	AGACCTCACAGTAAAAATAGGTGA	Exon 15	ROS1
SEQ ID NO. 578	CTGATGGGACCCACTCCATC		
SEQ ID NO. 579	GCACCAGAAGTCATCAGAATGC	Exon 16	
SEQ ID NO. 580	CTGGTCCCTGTTGTTGATGT		
SEQ ID NO. 581	TGGTGGGACGAGGATACCTG	Exon 17	
SEQ ID NO. 582	TGGGGAAAGAGTGGTCTCTCA		
SEQ ID NO. 583	CTCCTTGAATCGGGCTGGTT	Exon 18	
SEQ ID NO. 584	CGACTGCCAACTTCTCACCT		
SEQ ID NO. 585	CCGGCCATCTAAAAATCTGCC	Exon 1	
SEQ ID NO. 586	GGCAGCCAAGAGTTGCAAAA		
SEQ ID NO. 587	ACCAACTGCTCCCTTTGCTT	Exon 5	
SEQ ID NO. 588	CAGCTTCCCAGAAGTTGTGC		
SEQ ID NO. 589	TCCAGACCGTCCTATGTGGT	Exon 6	
SEQ ID NO. 590	TGGGACTTGGAGGGGAGTAG		
SEQ ID NO. 591	GAGAGCTCAAGTCCCGACAC	Exon 7	
SEQ ID NO. 592	CTGGTTCTCTGTGTCCCTGC		
SEQ ID NO. 593	CTGTTGATGTCCACCAGCAAA	Exon 10	
SEQ ID NO. 594	TCAGACATGTTGGCAGCCTT		
SEQ ID NO. 595	AGAACTGCTCAAACATCGAGGA	Exon 11	
SEQ ID NO. 596	CCATTGTATGAATCAGCCACAA		
SEQ ID NO. 597	GCAATCAAGCCACAAGCCAA	Exon 12	
SEQ ID NO. 598	AACAAGAGCCTGGTGAGAGC		
SEQ ID NO. 599	AATTAGGCCCTTCTGCCTGG	Exon 13	
SEQ ID NO. 600	ACTTCAGGAGGGTCTTGGGT		
SEQ ID NO. 601	ACCACCATTTATCATGGCTGTG	Exon 14	
SEQ ID NO. 602	AAGAACTCTCCTGGGCCAAA		
SEQ ID NO. 603	AAGGCGACGTTTTTGTGTGG	Exon 15	
SEQ ID NO. 604	CTGCAATGCTGGGTAGGTGA		
SEQ ID NO. 605	CAGTCTGTGTTGACGGGACA	Exon 16	
SEQ ID NO. 606	TCCCAGAAAACCAAGGCTGT		
SEQ ID NO. 607	TGGGCTCCTGTATTGGTTGG	Exon 17	
SEQ ID NO. 608	CTCTGTCCCCGAAGAACAGC		
SEQ ID NO. 609	ATAGTGGTCGGCTGTTCTGG	Exon 18	

SEQ ID NO. 610	GCAGGGGCTTAAGGGATGTC	
SEQ ID NO. 611	TCCTTTACCCCTAAGGTTATTCCA	Exon 19
SEQ ID NO. 612	GGGGACCATTCCACAGGATT	
SEQ ID NO. 613	ACTGTGGAAGGACTGGAACC	Exon 20
SEQ ID NO. 614	CAGGTGCTCGAAGTGACAGA	
SEQ ID NO. 615	GCACCAGAGAACCCCAGAAT	Exon 21
SEQ ID NO. 616	GCATCTGGGACTCATGCCTT	
SEQ ID NO. 617	TGTGTGGACGTTTTTCAGCAG	Exon 23
SEQ ID NO. 618	AGAGTGAGTCCCCTTCAGCA	
SEQ ID NO. 619	AGGCACCTCTACTTTGCACT	Exon 24
SEQ ID NO. 620	TGTGAATCTTCACCTCTCTGGG	
SEQ ID NO. 621	AGTCACACCTTGCACCGAAT	Exon 25
SEQ ID NO. 622	AGCCTTCCAGATCCATTGCC	
SEQ ID NO. 623	CACAGCAAAGGACAGCACAC	Exon 26
SEQ ID NO. 624	ACTCCTTAGGGCCTTCACCT	
SEQ ID NO. 625	TCAGACACTGTGGAACCAACT	Exon 27
SEQ ID NO. 626	GGTTTGTCTTGGCCAGAGGT	
SEQ ID NO. 627	TCAGGACAGTATAGCTCTTATTGAA	Exon 28
SEQ ID NO. 628	TCCCCAAATCTCTTTTCCTGGT	
SEQ ID NO. 629	GGTCAGACACCAGCCTCATT	Exon 29
SEQ ID NO. 630	CCAGGTGTGAGATTGCCAAC	
SEQ ID NO. 631	CCTGCCACTCTGAGGAAAT	Exon 30
SEQ ID NO. 632	AGTGTTCTCTGGAACCAAGG	
SEQ ID NO. 633	GCAGCCAAGGTCCTGCTTAT	Exon 31
SEQ ID NO. 634	GCTTTCTGGAAGTGAGGTGC	
SEQ ID NO. 635	GGAGTCCCAAATAAACCAGGC	Exon 32
SEQ ID NO. 636	TCTTCAGCTTTCTCCCACTGT	
SEQ ID NO. 637	TCCTGCAGTAGTGTTTGCAC	Exon 33
SEQ ID NO. 638	TGCAGCTACTACTCTGAACTGA	
SEQ ID NO. 639	CCAAGGAAGGGGTGACAGTG	Exon 35
SEQ ID NO. 640	CATTAGCCAGGCCTACTCCG	
SEQ ID NO. 641	CTGCCTTCCCTCGGGAAAAA	Exon 36
SEQ ID NO. 642	TGTCCACTGCTGTTTCCTTCA	

SEQ ID NO. 643	GCAGCTTGGAGTTTGTCTGC	Exon 38	
SEQ ID NO. 644	AAGGTCTCCTCCCTCCATCA		
SEQ ID NO. 645	ACTCACCTTGGTTGACCTTGT	Exon 39	
SEQ ID NO. 646	TGCATCCGTTCCAAGTAGACA		
SEQ ID NO. 647	TTGGAGACTTTGGACTCGCC	Exon 40	
SEQ ID NO. 648	ACTTTCTGGAGCCATCCACC		
SEQ ID NO. 649	TGGTCATCAGCCTTATCCAGC	Exon 41	
SEQ ID NO. 650	CCAGTCTCCCTCCTGTTTGC		
SEQ ID NO. 651	GGCTCAAGAACCCGACCAAA	Exon 42	
SEQ ID NO. 652	GACTCCACTGTTGTTTGCTTCA		
SEQ ID NO. 653	GTCCTCTAGGCTCCCAGGAA	Exon 43	
SEQ ID NO. 654	CCACTGTGAGTGAGACAGGC		
SEQ ID NO. 655	CCTGGTCATCTGCGGACTC	Exon 1	NTR K1
SEQ ID NO. 656	GAGTTGCCCACTCACTCCC		
SEQ ID NO. 657	CTGTACCCCCGATCTTGACG	Exon 2	
SEQ ID NO. 658	TGCTGCCAACAAGTTTGTGC		
SEQ ID NO. 659	CTGCTGGCTTGGCTGATACT	Exon 3	
SEQ ID NO. 660	AGCTCAGTCAGGTTCTCTGC		
SEQ ID NO. 661	ACATCGAGAACCAGCAGCAT	Exon 4	
SEQ ID NO. 662	AGGTTTCTCAGCTCCCCCA		
SEQ ID NO. 663	CACCATCGTGAAGAGTGGTCT	Exon 5	
SEQ ID NO. 664	GCGACTGAGCCGAGGAGT		
SEQ ID NO. 665	GGAACCTCTGCACTGTTCT	Exon 7	
SEQ ID NO. 666	AGCTTCTGTTCAGGCACTCC		
SEQ ID NO. 667	GAAGGTCCAGGTGCCCAATG	Exon 8	
SEQ ID NO. 668	CAGCTCTGTGAGGATCCAGC		
SEQ ID NO. 669	AAATCTGGGGGTCTGCCATC	Exon 9	
SEQ ID NO. 670	GCACGTCACGTTCTTCCTGT		
SEQ ID NO. 671	TCAATGGCTCCGTGCTCAAT	Exon 10	
SEQ ID NO. 672	TCGGGGTTGAACTCGAAAGG		
SEQ ID NO. 673	GTCTTTGCCTGCCTCTTCCT	Exon 13	
SEQ ID NO. 674	CCCAAACCTTGTTTCTCCGTCC		
SEQ ID NO. 675	ATGACATTGGGTGGCAGCTC	Exon 14	

SEQ ID NO. 676	TGGGTTCTCGATGATGTGGC		XRC C1
SEQ ID NO. 677	ACATCGTGCTCAAGTGGGAG	Exon 15	
SEQ ID NO. 678	ACCAGCATCTTGTCTGCTC		
SEQ ID NO. 679	CACATCGTGCGCTTCTTCG	Exon 16	
SEQ ID NO. 680	CCCGTGCCGCATATACTCAA		
SEQ ID NO. 681	ATGGACCTGATGCCAAGCTG	Exon 17	
SEQ ID NO. 682	CCCACTAGACAGTTGCGTGT		
SEQ ID NO. 683	ACCGTAAGTTCACCACCGAG	Exon 18	
SEQ ID NO. 684	CTGCTTGCCGTAGGTGAAGA		
SEQ ID NO. 685	CACCAGAGGTCTACGCCATC	Exon 19	
SEQ ID NO. 686	CCCACATGCTGAGGGTGAAT		
SEQ ID NO. 687	CGCTTGCGCACTTTAGCCAG	Exon 1	
SEQ ID NO. 688	TGAGGTAGAGTATGGGGTCCG		
SEQ ID NO. 689	CTCAAGGCAGACACTTACCGA	Exon 2	
SEQ ID NO. 690	TGTAGGACCACAGAGATGGTCT		
SEQ ID NO. 691	TGGGAATGATGGCTCAGCTT	Exon 3	
SEQ ID NO. 692	CATAGTCTTGCTCCCCAGCG		
SEQ ID NO. 693	GTTCGCATGTTTGGGCCTG	Exon 4	
SEQ ID NO. 694	GGGCTGGCTGCAAACAATTT		
SEQ ID NO. 695	CTCCCCCTTTGGCTTGAGTT	Exon 5	
SEQ ID NO. 696	CTGGGACGGGGCCTCT		
SEQ ID NO. 697	CCAGTTCCGTGTGAAGGAGG	Exon 6	
SEQ ID NO. 698	ATCCGGCTGAAGAAGAGAGC		
SEQ ID NO. 699	CTATGCAGCTGCTACCCTCC	Exon 7	
SEQ ID NO. 700	CTTGGAGGTGCTGCCTATGG		
SEQ ID NO. 701	AGTCTCCCAAAGGGAAGAGGA	Exon 8	
SEQ ID NO. 702	TCTTGGGAACAGATGGCGAC		
SEQ ID NO. 703	CCTTCAGGGTGTGGTAGTGG	Exon 9	
SEQ ID NO. 704	CTAGGGCCTTATCTCGCAGC		
SEQ ID NO. 705	TTTGCCAACACCCCCAAGTA	Exon 10	
SEQ ID NO. 706	GTCCAGCACCCACTCCTTAC		
SEQ ID NO. 707	CAGGTTCCAGCAGTGAGGAG	Exon 11	
SEQ ID NO. 708	CTTCTGAGGAAGCTTGGGGG		

SEQ ID NO. 709	AGACCAAAACCAAGCCCACT	Exon 12	
SEQ ID NO. 710	CTTCCTGGAGCACTGGTGAG		
SEQ ID NO. 711	GGCAGAGCAGAAGGAACACA	Exon 14	
SEQ ID NO. 712	TGTCCGTGTTCTCATCCGTG		
SEQ ID NO. 713	TCTTCCAGGGCAAGCACTTC	Exon 15	
SEQ ID NO. 714	TATCGGATGAGTTTCCGCCG		
SEQ ID NO. 715	TCGTCCCCGATGGATCTACA	Exon 17	
SEQ ID NO. 716	GCACATACTTCAGGCTTGCG		
SEQ ID NO. 717	GTTGAAGGACCAGCTCTCCC	NM_014	CD27 4
SEQ ID NO. 718	CTTGTAAGTCGGCACCACCAT	143.4	
SEQ ID NO. 719	CATGTCAGGCTGAGGGCTAC	NM_014	
SEQ ID NO. 720	GGAATTGGTGGTGGTGGTCT	143.4	
SEQ ID NO. 721	TACCTCTGGCACATCCTCCA	NM_014	
SEQ ID NO. 722	GTCAGTGCTACACCAAGGCA	143.4	
SEQ ID NO. 723	GGCCGAAGTCATCTGGACAA	NM_001	
SEQ ID NO. 724	TCATGCTAGGTGCTAGGGGA	314029.2	
SEQ ID NO. 725	AGACCACCACCACCAATTCC	NM_001	
SEQ ID NO. 726	TCATGCTAGGTGCTAGGGGA	314029.2	
SEQ ID NO. 727	AAGAGAGATTCTCCAAAATGTGGC	NM_152	RICT OR
SEQ ID NO. 728	ATTCAGATGGCCTAGCTTTCTC	756.5	
SEQ ID NO. 729	CAAAAGAAGTGCGAGCAGCA	NM_152	
SEQ ID NO. 730	GCACCTTCTGGAGAATACTGGA	756.5	
SEQ ID NO. 731	ACTGTGAATGCTTCCTTGTTTCC	NM_152	
SEQ ID NO. 732	TGCTCGGACCATTCTGTCTC	756.5	
SEQ ID NO. 733	TCCAGGGAGGTTCCAAGACA	NM_001	
SEQ ID NO. 734	CTGGATCTGGCACGATGAGG	285439.2	
SEQ ID NO. 735	GCCCTTCGAGGTGGACTAAA	NM_001	
SEQ ID NO. 736	CTCGCACATACTGTGAGTCT	285439.2	
SEQ ID NO. 737	TCTTCCTCATTACATAGCCATCA	NM_001	
SEQ ID NO. 738	CCTTGGGGATATCAAAGGATGC	285439.2	
SEQ ID NO. 739	CAGTGCAGCCTTGAAGTGT	NM_001	
SEQ ID NO. 740	TGATCCCGTTTCTGGTGTGT	285439.2	
SEQ ID NO. 741	GCCCAGCAGTAAATTATATGCCA		

SEQ ID NO. 742	AACTGGCAACCTACAACCGT	NM_001 285439.2	ARE G
SEQ ID NO. 743	TGGCTCAATGCTTCATCTGGA	NM_001 285439.2	
SEQ ID NO. 744	AACTCCATGAGGGTGGCAAG		
SEQ ID NO. 745	GCCTGCAGACTCTATGCAAC	NM_001 285439.2	
SEQ ID NO. 746	AGCTGGGTCACTAACAACCTCA		
SEQ ID NO. 747	GTGTCCCAGAGACCGAGTTG	NM_001 657.4	
SEQ ID NO. 748	GCCGAGTATCAAGAGCGACA		
SEQ ID NO. 749	ATGCTGCTGGATTGGACCTC	NM_001 657.4	
SEQ ID NO. 750	AGGCATTTCACTCACAGGGG		
SEQ ID NO. 751	ATTTCGGTGAACGGTGTGGG	NM_001 657.4	
SEQ ID NO. 752	TAACAGCAACAGCTGTGAGGA		
SEQ ID NO. 753	ATTTCGGTGAACGGTGTGGG	NM_001 657.4	
SEQ ID NO. 754	TAACAGCAACAGCTGTGAGGA		
SEQ ID NO. 755	AGTCGGTCCTCTTTCCAGTG	NM_001 657.4	
SEQ ID NO. 756	ACTTTTTACCTTCGTGCACCT		
SEQ ID NO. 757	TTGATTCTTGCAGCCACCCT	Exon 1	BCL2 L11
SEQ ID NO. 758	GAACGCAGCGAACCGAATAC		
SEQ ID NO. 759	AGGCAATCACGGAGGTGAAG	Exon 2b	
SEQ ID NO. 760	TCTGGTAGCAAAAGGGCCAG		
SEQ ID NO. 761	CTGAGTGTGACCGAGAAGGT	Exon 2a	
SEQ ID NO. 762	TGTGGCTCTGTCTGTAGGGA		
SEQ ID NO. 763	CAGCACCCATGAGTTGTGAC	Exon 2c	
SEQ ID NO. 764	AGATAGTGGTTGAAGGCCTGG		
SEQ ID NO. 765	TCATCCTAGAGGATATAGGTGATCT	Exon 3	
SEQ ID NO. 766	ACCAATGACCATATGACAAAATGCT		
SEQ ID NO. 767	TGCCTGAGTCTGCTACTTGC	Exon 7	
SEQ ID NO. 768	GGGGAAATTCTGATGAGCTGC		
SEQ ID NO. 769	CTTCCATGAGGCAGGCTGAA	Exon 8	
SEQ ID NO. 770	TCCAATACGCCGCAACTCTT		
SEQ ID NO. 771	CTGGCAAAACTCCTGGCATC	Exon 9	
SEQ ID NO. 772	TTTCAGATCTAGGCTGGAATGTTCT		
SEQ ID NO. 773	CACTTAAGGGCAGTGGGGAA	Exon 10	

SEQ ID NO. 774	GCCTCTCTCACCTCCTCCTT			
SEQ ID NO. 775	AGAACCACTGTAGCAGACGC	Exon 11		
SEQ ID NO. 776	ATGCCTTGACCGTCTTCCTG			
SEQ ID NO. 777	GCTTTGCCGCCCTTATGATG	Exon 12		
SEQ ID NO. 778	CACGTGTACGTGAGTGGACA			
SEQ ID NO. 779	AACCGCCAGATTTGAATCGC	Exon 1		
SEQ ID NO. 780	GATGCGGTGGTCCTTGAGAA			
SEQ ID NO. 781	CCACTGAGAACGAGCCAGAC	Exon 2	BIRC 5	
SEQ ID NO. 782	CATCTGGCTCCCAGCCTTC			
SEQ ID NO. 783	CGGTTGCGCTTTCCTTTCTG	Exon 3		
SEQ ID NO. 784	TGGCTCTTTCTCTGTCCAGT			
SEQ ID NO. 785	AGTCCCTGGCTCCTCTACTG	Exon 4		
SEQ ID NO. 786	CGTGTGGAGAACGTGACAGA			
SEQ ID NO. 787	TCCCTCTGCAGTCTTTCCT	Exon 1		ERCC 1
SEQ ID NO. 788	CGGGCCTCACGGTTTCAG			
SEQ ID NO. 789	GACCTACGCCGAATATGCCA	Exon 3		
SEQ ID NO. 790	GATTTTGCCCCGGGTTTCAG			
SEQ ID NO. 791	GGGGCAATCCCGTACTGAAG	Exon 4		
SEQ ID NO. 792	CTCTGGCCCAGCACATAGTC			
SEQ ID NO. 793	CCCAGACTACATCCATGGGC	Exon 5		
SEQ ID NO. 794	ACATCCACCTGGACAAGCAG			
SEQ ID NO. 795	AAAGATCCCCAGCAGGCCC	Exon 6		
SEQ ID NO. 796	TCCAGGCGAGGATCAATGTG			
SEQ ID NO. 797	GGCGGTACCTGGAGACCTA	Exon 7		
SEQ ID NO. 798	ACGAAGTCCTGCTCTAGCTTC			
SEQ ID NO. 799	TGACTGAATGTCTGACCACCG	Exon 8		
SEQ ID NO. 800	CCAAATGTGGTCAGGAGGGT			
SEQ ID NO. 801	CGGAGGCTGTTTGATGTCCT	Exon 10		
SEQ ID NO. 802	CCAGAATCCCTCCCCAGAGA			
SEQ ID NO. 803	AGGTATTCCCTGAGTCCCCC	Exon 1	GNA S	
SEQ ID NO. 804	TTGGATCTTTGTCCTCGGGC			
SEQ ID NO. 805	CAGTAAGACCGAGGACCAGC	Exon 2		
SEQ ID NO. 806	GCCCGGTAGACCTGCTTG			

SEQ ID NO. 807	GTGCTGGAGAATCTGGTAAAAGC	Exon 3	
SEQ ID NO. 808	CCATTAAACCCATTAACATGCAGGA		
SEQ ID NO. 809	CCCCGAGAACCAGTTCAGAG	Exon 6	
SEQ ID NO. 810	GGGAGGGAAGTCAAAGTCAGG		
SEQ ID NO. 811	CAAGGCTCTGTGGGAGGATG	Exon 7	
SEQ ID NO. 812	ACTGGGCACAGTCAATCAGC		
SEQ ID NO. 813	CTGCTTCGCTGCCGTGT	Exon 9	
SEQ ID NO. 814	GGAAGTTGACTTTGTCCACCTG		
SEQ ID NO. 815	ACTGCCATCATCTTCGTGGT	Exon 11	
SEQ ID NO. 816	GGTTGGTCTGGTTGTCCTCC		
SEQ ID NO. 817	CGCACCATCTCTGTGATCCT	Exon 12	
SEQ ID NO. 818	TCGATTTCAGCAAGGACT		
SEQ ID NO. 819	GCGTCACTACTGCTACCCTC	Exon 14	
SEQ ID NO. 820	ATGATGTCACGGCAGTCGTT		
SEQ ID NO. 821	CCGTGCTAGCCGTTGGG	Exon 1	AKT2
SEQ ID NO. 822	CGCTGGTTCCCTTTCCTTGT		
SEQ ID NO. 823	CTAGGTGACAGCGTACCACG	Exon 2	
SEQ ID NO. 824	CTTGTGGAGCCAGCCTTCTT		
SEQ ID NO. 825	ACTTCCTGCTGAAGAGCGAC	Exon 3	
SEQ ID NO. 826	GGGTAGAGTCTGATCAGGGG		
SEQ ID NO. 827	ATGCCAGCTGATGAAGACCG	Exon 4	
SEQ ID NO. 828	ACGTGGAAGGTCCTCTCGAT		
SEQ ID NO. 829	CGAGGACCCCATGGACTACA	Exon 5	
SEQ ID NO. 830	CGCCACTTCCATCTCCTCAG		
SEQ ID NO. 831	TGGCAAAGTCATCCTGGTGC	Exon 6	
SEQ ID NO. 832	GCAATGATGACTTCCTTCCGC		
SEQ ID NO. 833	CTGTTCTTCCACCTGTCCCG	Exon 8	
SEQ ID NO. 834	AGCCGAGACAATCTCTGCAC		
SEQ ID NO. 835	CTGTTCTTCCACCTGTCCCG	Exon 9	
SEQ ID NO. 836	AGCCGAGACAATCTCTGCAC		
SEQ ID NO. 837	TCACTGACTTTGGCCTCTGC	Exon 10	
SEQ ID NO. 838	GGGGTCCCACAGAAGGTTTT		
SEQ ID NO. 839	GCCTCTTCGAGCTCATCCTC	Exon 11	

SEQ ID NO. 840	TCCTTCTTAAGCAGCCCAGC		Exon 12	ERBB 2
SEQ ID NO. 841	CAGCGATGCCAAGGAGGTC			
SEQ ID NO. 842	CTTCTGGACCACGTCCTGC			
SEQ ID NO. 843	GTTCCCGGATTTTTGTGGGC	Exon 1	Exon 2	
SEQ ID NO. 844	TCCGCCTTCCCTTAACCCTA			
SEQ ID NO. 845	CTCCCGAGGGTGATTGCTTC	Exon 2	Exon 4	
SEQ ID NO. 846	TGCTGGAGGTAGAGTGGTGA			
SEQ ID NO. 847	TGCAAAGCTACTCCCTGAGC	Exon 4	Exon 6	
SEQ ID NO. 848	AGGGTCTGAGTCTCTGTGCT			
SEQ ID NO. 849	GCTGCTTGAGGAAGTATAAGAATGA	Exon 6	Exon 7	
SEQ ID NO. 850	CTGGTTTCTCCGGTCCCAAT			
SEQ ID NO. 851	GCACAGACATGAAGCTGCG	Exon 7	Exon 8	
SEQ ID NO. 852	GTGGGCAGGTAGGTGAGTTC			
SEQ ID NO. 853	ACAACCAAGTGAGGCAGGTC	Exon 8	Exon 9	
SEQ ID NO. 854	AGGGGTGGTATTGTTCAGCG			
SEQ ID NO. 855	CAGCTCTGCTACCAGGACAC	Exon 9	Exon 11	
SEQ ID NO. 856	CGAGAGCGGTTGGTGTCTAT			
SEQ ID NO. 857	CACTGTCTGTGCCGGTGG	Exon 11	Exon 12	
SEQ ID NO. 858	AGGCAGTCAGAGTGCTTGG			
SEQ ID NO. 859	GCCTCCACTTCAACCACAGT	Exon 12	Exon 13	
SEQ ID NO. 860	GGGATTGGGCATGGACTCAA			
SEQ ID NO. 861	TCTACGGACGTGGGATCCTG	Exon 13	Exon 2A	POLD 1
SEQ ID NO. 862	GCTGTGTTCCATCCTCTGCT			
SEQ ID NO. 863	GGGCCTCTGGGATGATGATG	Exon 2A	Exon 2B	
SEQ ID NO. 864	CAACCCCCTCCAGGACTGA			
SEQ ID NO. 865	GGGCCTCTGGGATGATGATG	Exon 2B	Exon 3	
SEQ ID NO. 866	CAACCCCCTCCAGGACTGA			
SEQ ID NO. 867	GGTCCCACCATCAGCCATAG	Exon 3	Exon 4	
SEQ ID NO. 868	TGGAAGATGAGGGGCTCTGT			
SEQ ID NO. 869	GTCACCGATGAGGGGTTCTC	Exon 4	Exon 5	
SEQ ID NO. 870	GCGCTGGGGTGTAGAAGTAG			
SEQ ID NO. 871	GGAGCTGAACTTGGCCATCA	Exon 5		
SEQ ID NO. 872	CTCTCGGGAGCACAGTTCC			

SEQ ID NO. 873	GTCTCCTGGAACAGGGGCATC	Exon 6	
SEQ ID NO. 874	AAGTCGACGTTGGCCTCGTA		
SEQ ID NO. 875	TCATGGTGGACACGGACATC	Exon 7	
SEQ ID NO. 876	TCAGCCTCAGGGCGTATTTC		
SEQ ID NO. 877	CTGACGTGGTCAGTCACCC	Exon 8	
SEQ ID NO. 878	GATATCGAAGCTGAGCACGC		
SEQ ID NO. 879	TGTCATCCAGATCTGCTCGC	Exon 9	
SEQ ID NO. 880	CCTTCTCGTAGCTCTGCACC		
SEQ ID NO. 881	CATCCGTATCATGGACCCCG	Exon 10	
SEQ ID NO. 882	CCCGAGAGATGAGGTACGGA		
SEQ ID NO. 883	TTGTAACGGGAGTCGGGAGA	Exon 1	RB1
SEQ ID NO. 884	CAGCGAGCTGTGGAGGAG		
SEQ ID NO. 885	GAGTTTGAAGAAACAGAAGAACCTG	Exon 2	
SEQ ID NO. 886	TCTCCCAAGTTAACCAAGCTCTC		
SEQ ID NO. 887	AGGAACTGTGGGGAATCTGT	Exon 3	
SEQ ID NO. 888	AGCTCAGTAAAAGTGAACGACA		
SEQ ID NO. 889	CCAGTACCAAAGTTGATAATGCTAT	Exon 4	
SEQ ID NO. 890	TCCAATTTGCTGAAGAGTGCAA		
SEQ ID NO. 891	ACAAATGGAAGATGATCTGGTGA	Exon 7	
SEQ ID NO. 892	TGAGCAACATGGGAGGTGAG		
SEQ ID NO. 893	CATTAATGGTTCACCTCGAACACC	Exon 8	
SEQ ID NO. 894	TTCTAGTTGTTTTGCTATCCGTGC		
SEQ ID NO. 895	TGAAACACAGAGAACACCACGA	Exon 11	
SEQ ID NO. 896	AACTGGAGTGTGTGGAGGAA		
SEQ ID NO. 897	GACTGTTATGAACACTATCCAACAA	Exon 12	
SEQ ID NO. 898	ATTTTCTGAAGGTTGATCACTTGCT		
SEQ ID NO. 899	ACTGCACAGTGAATCCAAAAGA	Exon 13	
SEQ ID NO. 900	TTTCGACACAACCCTGTCCC		
SEQ ID NO. 901	GCAGAAGGCAACTTGACAAGAG	Exon 17	
SEQ ID NO. 902	GAGCCATGCAAGGGATTCCA		
SEQ ID NO. 903	GAGGCTTCCCCTCGTTTGG	Exon 1	SMA RCA4
SEQ ID NO. 904	CAGAATGGGCGTCTCGCTTC		
SEQ ID NO. 905	CCCGTGAAGATGTCCACTCC	Exon 2	

SEQ ID NO. 906	GCATCTGGTGCATGTTGTCC			
SEQ ID NO. 907	AGTCCATGCATGAGAAGGGC	Exon 3		
SEQ ID NO. 908	GACCGCATCCCCATTCCTTT			
SEQ ID NO. 909	GCCTCTAGTCCAGTTCAGC	Exon 4		
SEQ ID NO. 910	CCTGGCCAGCATCTTGTAGG			
SEQ ID NO. 911	AACATGCCTCCCCCAGGA	Exon 5		
SEQ ID NO. 912	CTTCAGGCCAGGGCTTGGG			
SEQ ID NO. 913	GCACCCCTCAGAAGCTGATT	Exon 6		
SEQ ID NO. 914	CTTCTGGATGGGGGTGATGC			
SEQ ID NO. 915	CGCACACCGAATTCAGGAAC	Exon 7		
SEQ ID NO. 916	TGGAAGTTCAGCAGCCTGAG			
SEQ ID NO. 917	ACAGCCCTCAATGCTAAGGC	Exon 8		
SEQ ID NO. 918	CTGCTGCTTCTCCAGCTTCT			
SEQ ID NO. 919	GCATTCTCCAGCATGCCAAG	Exon 9		
SEQ ID NO. 920	TCTTTCTTCTGCTCCCGCTC			
SEQ ID NO. 921	CGCAAGCTCATCGACCAGAA	Exon 10		
SEQ ID NO. 922	CACCAGCTCCGTGAGGTTAG			
SEQ ID NO. 923	TGAGACTGGACCTGAGGAGAC	Exon 1	WEE1	
SEQ ID NO. 924	GGGTGTCTGAAGAGTCGCAG			
SEQ ID NO. 925	GTTTGCTCTCCAAAGCTCGG	Exon 2		
SEQ ID NO. 926	CACTTGAGGAGTCTGTCTGCA			
SEQ ID NO. 927	AAGTGTGTGAAGAGGCTGGA	Exon 4		
SEQ ID NO. 928	CATCAACAGAGCCCGCCAA			
SEQ ID NO. 929	GCAGAACGCTTTGAGAGAAG	Exon 5		
SEQ ID NO. 930	GCCCACGCAGAGAAATATCG			
SEQ ID NO. 931	GGAAGTTTAGCTGATGCTATAAGTG	Exon 6		
SEQ ID NO. 932	CGGCCAACTTGCAAAAGGAG			
SEQ ID NO. 933	TCGAACCTCAATCCCAAATGC	Exon 7		
SEQ ID NO. 934	ACATAACTTTGTTGGATGCCCA			
SEQ ID NO. 935	GTGATCTTGGGCATGTAACAAGG	Exon 8		
SEQ ID NO. 936	TTTGCAAGAAAACGACTATCGCC			
SEQ ID NO. 937	GCTTGCCCTCACAGTGGTAT	Exon 9		
SEQ ID NO. 938	GCACTTGTGGTATCCGAGGT			

SEQ ID NO. 939	TCCAGAGAGAAGACCTTCAGC	Exon 10	
SEQ ID NO. 940	ACTCTTTCTAGAAGCGGACAGC		
SEQ ID NO. 941	CTCCCCCTGAACACTGTGAC	Exon 11	
SEQ ID NO. 942	TGCCTACAAAGTGCTCCCAG		

Với dữ liệu giải trình tự, được chuẩn hóa dữ liệu và được tập hợp thành các thông số bao gồm trình tự, gen các thông tin sánh, các đột biến điểm trên gen, nếu có. Cần lưu ý rằng, trong quá trình giải trình tự này có thể xuất hiện các đột biến điểm mới và việc đánh giá các đột biến điểm này liên quan đến phát sinh bệnh ung thư phổi tế bào không nhỏ sẽ được phân tích và cập nhật trong dữ liệu bằng kỹ thuật máy học để tối ưu hóa và thiết lập cơ sở dữ liệu đột biến điểm bổ sung.

Dữ liệu có được sau khi giải trình tự để chuyển thành dữ liệu phân tích sẽ trải qua các bước phân tích kết quả bao gồm:

Tiền xử lý dữ liệu: Nhằm giải quyết các vấn đề liên quan tới dữ liệu như sự hiện diện của nhiễu, các phần tử ngoại lai, dữ liệu bị thiếu hoặc trùng lặp và dữ liệu bị sai lệch - không thể hiện được và để chuẩn hóa dữ liệu đầu vào cho mô hình học máy. Tiền xử lý dữ liệu với các bước cụ thể như sau: Cắt bỏ (trimming) một phần đoạn trình tự dựa vào chỉ tiêu chất lượng (dưới 20) hay độ dài cần thiết (thông thường là phần đầu hoặc phần đuôi trình tự đọc) bằng cách sử dụng công cụ mã nguồn trimmomatic (Bolger et al., 2014; chuyển đổi định dạng từ FASTQ sang FASTA sử dụng các công cụ FastXtoolkit; ghép nối đối với cặp trình tự paired-end (PE) có các lần đọc 1 và 2 trùng lặp nhau thì có thể lựa chọn ghép nối để thành đoạn trình tự dài hơn.

Theo quy trình của sáng chế, dữ liệu xuất ra sau bước tiền xử lý dữ liệu phải đạt chất lượng thỏa mãn: độ sâu của giải trình tự (read depth) trên 20, độ bao phủ (sequencing coverage) đạt 200x mới tiếp tục thực hiện các giai đoạn tiếp theo.

Định vị trình tự ADN ngắn so với hệ tham chiếu bằng cách so sánh trình tự (Read Alignment) với hệ gen tham chiếu GRCh38 và loại bỏ vị trí phân tử trùng lặp bằng cách sử dụng công cụ BWA-MEM.

Xác định biến thể bằng mô hình học máy bằng cách gọi tên biến thể (Variant Calling) áp dụng ba công cụ quan trọng: mô hình Convolutional Neural Network (CNN), Genomic Analysis Toolkit (GATK) và thuật toán XGBoost để chú thích biến thể.

Chú thích biến thể (Variant Annotation): Thêm thông tin về mỗi biến thể được phát hiện. Thông thường, chú thích cả thông tin để tạo điều kiện thuận lợi cho việc giải thích ý nghĩa lâm sàng của biến thể. Theo sáng chế, chú thích biến thể sử dụng công cụ ANNOVAR với bộ lọc gồm:

- + Lọc các biến thể có $MQ > 40$.

- + Lọc các biến thể dựa trên tần số alen ($MAF < 0.01$)

- + Lọc các biến thể có tiềm năng gây bệnh: các biến thể gây thêm/mất, mất bộ ba kết thúc, stop-gain, dịch khung, và thay đổi vị trí trượt gen thường được coi là có tiềm năng bệnh, các biến thể sai nghĩa được chú trọng bởi khả năng gây hại cho chức năng protein.

- + Đánh giá các đa hình (SNP và In/del) trên vùng mã hóa của các bệnh nhân mắc bệnh ung thư vú qua chỉ số SIFT_Pred và Polyphen 2.

Từ kết quả giải trình tự, các dữ liệu liên quan giữa các biến thể gen có đột biến gây bệnh được xác định theo các dược chất hướng đích hướng dẫn của FDA thông qua việc truy cập và tìm kiếm thông tin trên cơ sở dữ liệu như PharmVar và PharmVarGKB để đề xuất hoạt chất hướng đích tiềm năng.

Theo sáng chế, một số mô hình liên quan đến khả năng đáp ứng thuốc từ kết quả đột biến gen di truyền dựa trên bệnh phẩm mô u hoặc máu của bệnh nhân, kết quả phân tích trả lời bệnh nhân có hay không có xuất hiện đột biến. Nếu đột biến xuất hiện trên các gen EGFR, ALK, KRAS hoặc các đột biến gen hiếm gặp khác như PIK3CA, BRAF, NRAS, ROS1, HER2 thì các thuốc hướng đích hiệu quả được căn cứ theo các thuốc điều trị hướng đích được FDA phê duyệt. Ví dụ thuốc điều trị hướng đích cho đột biến gen EGFR là Erlotinib, Gefitinib, Afatinib, Osimertinib trong khi thuốc điều trị hướng đích cho đột biến gen ALK lại là Crizotinib, Ceritinib, Alectinib, Loratinib.

Căn cứ vào mức độ đột biến và thứ tự đột biến từng gen dương tính được thống kê, kết quả phân tích sẽ gán thuốc điều trị hướng đích phù hợp đối với từng đột biến theo thứ tự mức độ ưu tiên của thuốc điều trị, mức độ đáp ứng hay kháng lại thuốc điều trị hướng đích được khuyến cáo. Quá trình này được thực hiện thông qua quy trình phân tích gen liên quan đến ung thư phổi không tế bào nhỏ và phương pháp đánh giá khả năng đáp ứng với thuốc điều trị hướng đích.

Theo sáng chế, nếu mô hình phân tích xuất hiện gen nhạy cảm với thuốc điều trị hướng đích được khuyến cáo thì kết quả sẽ trả về ưu tiên thuốc điều trị hướng đích tương ứng với đột biến gen đó. Khi mô hình phân tích có đột biến gen kháng với thuốc điều trị hướng đích hoặc thuốc điều trị hướng đích được khuyến cáo sử dụng cho giai đoạn sau thì kết quả trả về gán cho phác đồ điều trị hoá chất hoặc miễn dịch trước. Sau đó, nếu xuất hiện đột biến kháng lại thuốc hướng đích thì sẽ lựa chọn đề xuất thuốc hướng đích ở bước tiếp theo. Việc thiết lập, đánh giá, khảo sát các đột biến điểm mới phát sinh liên quan đến bệnh cũng được cập nhật theo thời gian cho phép mô hình máy học cập nhật, đánh giá bổ sung có áp dụng mô hình phân tích tối ưu, cho phép lặp lại kết quả mong muốn hoặc loại trừ kết quả không mong muốn.

Theo đó, nếu mô hình phân tích cho thấy xuất hiện đột biến gen kháng lại thuốc điều trị hướng đích hoặc đột biến gen phù hợp với thuốc điều trị hướng đích ở bước 2. Khi đó kết quả sẽ gán phác đồ điều trị ưu tiên, sau đó là hoá chất hoặc điều trị miễn dịch.

Trường hợp xuất hiện nhiều đột biến gen, chẳng hạn đột biến gen EGFR kết hợp với đột biến gen khác hoặc xuất hiện đột biến mới, kết quả phân tích sẽ lựa chọn ưu tiên thuốc điều trị hướng đích nào phù hợp với đột biến gen đó và sắp xếp theo mức độ ưu tiên. Ngoài ra kết quả sẽ cho thấy rằng thuốc đó khi điều trị sẽ tương tác với loại đột biến gen đó hay không. Ví dụ: khi bệnh nhân có đồng thời đột biến gen EGFR và KRAS, kết quả phân tích đột biến gen sẽ trả lời các gen liên quan đến điều trị đích EGFR là Gefitinib, Erlotinib, Afatinib, Osimertinib, nhưng hiệu quả điều trị sẽ ảnh hưởng bởi đột biến gen KRAS đi kèm, đây là đột biến được công bố làm kháng lại thuốc điều trị hướng đích EGFR. Các dữ liệu đột biến điểm mới phát sinh được duy trì cho đến bước phân tích tiếp theo.

Các chỉ số số cận lâm sàng được khai thác và tập hợp bao gồm các thông tin về tuổi, giới tính, lý do nhập viện, thời gian từ khi xuất hiện triệu chứng đến khi nhập viện, tiền sử bệnh trước đó (bệnh lý liên quan đến phổi như lao, viêm phổi, COPD, hen...). Tiếp đến là các triệu chứng hô hấp: ho, khạc đờm, ho ra máu, đau ngực, khó thở, các triệu chứng toàn thân: mệt mỏi, gầy sút cân, hạch ngoại vi, cân nặng chiều cao.

Các chỉ số lâm sàng được thu thập đầy đủ các triệu chứng, bao gồm triệu chứng toàn thân, cơ năng, và triệu chứng cơ quan di căn, đánh giá điểm thể trạng theo tiêu chuẩn của Tổ chức Y tế Thế giới (WHO), điểm toàn trạng PS từ 0-4 đánh giá bệnh nhân

vận động hoàn toàn bình thường cho đến bệnh nhân nằm liệt giường không thể đi lại vận động được.

Các thông tin chẩn đoán hình ảnh, bệnh nhân được mô tả vị trí, kích thước, số lượng và các đặc điểm hình thái, tính chất xâm lấn của u trên hình ảnh XQ lồng ngực tiêu chuẩn và hình ảnh CLVT lồng ngực có tiêm thuốc cản quang, phim chụp MRI sọ não, xạ hình xương hoặc PET/CT nếu có. Kết quả chẩn đoán mô bệnh học: áp dụng theo hệ thống phân loại mô bệnh học của TCYTTG đưa ra năm 2021 cho các khối u phổi, màng phổi.

Như vậy, các chỉ số về gen, đột biến gen, các chỉ số cận lâm sàng, lâm sàng liên quan đến mô hình đánh giá cùng với các thông tin bổ sung như tình trạng hút thuốc, vị trí, kích thước, số lượng và các đặc điểm hình thái, tính chất xâm lấn của khối u, kết quả chẩn đoán mô bệnh học, tế bào học ung thư, kết quả hoá mô miễn dịch, giai đoạn bệnh theo phân loại TNM sẽ được chuẩn hóa thành các dữ liệu số để có thể đưa vào để đào tạo máy học.

Trong bước thiết lập mô hình phân tích máy học dựa trên các chỉ số di truyền và chỉ số lâm sàng, cận lâm sàng, tiến hành so sánh song song đồng thời từng chỉ số trong tập hợp dữ liệu đánh giá với cơ sở dữ liệu các thông số liên quan đến mô hình phát sinh bệnh. Tiến hành so sánh ít nhất hai trình tự được giải với cơ sở dữ liệu di truyền chuẩn theo mô hình XGBoost tăng cường với truy vấn ngẫu nhiên (random search) và đánh giá chéo (cross-validation). Trong đó, số trường quyết định độc lập ($n_estimator$) là 100, độ sâu trường (max_depth) là 13, tỷ lệ học ($learning_rate$) là 0,3. Từ các dữ liệu đầu vào và dữ liệu so sánh, dữ liệu gốc để thu được các thông số đột biến gen liên quan đến thuốc điều trị đích phù hợp nhất có trong cơ sở dữ liệu mô hình phát sinh bệnh đồng thời nạp dữ liệu này vào dữ liệu so sánh.

Mô hình XGBoost (Extreme Gradient Boosting) là một giải thuật dựa trên gradient boosting, nhưng với những cải tiến đáng kể về tối ưu thuật toán và hiệu suất. XGBoost kết hợp sự mạnh mẽ của phần mềm và phần cứng để đạt được kết quả vượt trội cả về thời gian huấn luyện và sử dụng bộ nhớ.

Tối ưu hóa thuật toán: XGBoost sử dụng một số kỹ thuật tối ưu hóa để tăng tốc độ và hiệu suất của thuật toán gradient boosting. Cụ thể, XGBoost sử dụng các phương

pháp dự đoán khởi tạo (predictive initialization) để ước lượng giá trị dự đoán ban đầu, giúp tiết kiệm thời gian trong việc huấn luyện các mô hình con sau này.

Cải tiến phần cứng: XGBoost tối ưu hóa việc tính toán trên phần cứng, chẳng hạn như sử dụng các thư viện tối ưu hóa BLAS (Basic Linear Algebra Subprograms) và hỗ trợ tính toán song song để tăng tốc độ huấn luyện mô hình trên các bộ dữ liệu lớn.

Regularization: XGBoost sử dụng regularization để kiểm soát overfitting và cải thiện độ chính xác của mô hình. Điều này bao gồm việc sử dụng hàm mất mát có thêm các thành phần penalty để giảm sự phức tạp của mô hình.

Hỗ trợ các loại bài toán: XGBoost có khả năng xử lý cả bài toán hồi quy và phân loại. Nó hỗ trợ nhiều hàm mất mát (loss functions) phù hợp với từng loại bài toán và cho phép tùy chỉnh các tham số để tối ưu mô hình cho mục tiêu cụ thể.

Xgboost được xây dựng dựa trên nguyên lý của Học Kết Hợp (Ensemble Method). Đây là một phương pháp trong lĩnh vực học máy với tư tưởng là thay vì xây dựng một mô hình tốt duy nhất, người ta sẽ xây dựng một tập hợp các mô hình yếu hơn một chút, nhưng khi kết hợp các mô hình lại với nhau một cách chính xác có thể thu được một mô hình cuối cùng vượt trội hơn cả.

Với quần thể gồm M cây. Tree1 được đào tạo bằng cách sử dụng ma trận đặc trưng X và nhãn y . Các dự đoán với nhãn $\hat{y}_1$ từ Tree1 được sử dụng để xác định các lỗi còn lại của tập huấn luyện r_1 . Sau đó, Tree2 được đào tạo bằng cách sử dụng ma trận đặc trưng X và các lỗi còn lại r_1 của Tree1 làm nhãn. Các kết quả dự đoán $\hat{r}_1$ sau đó được sử dụng để xác định phần dư r_2 . Quá trình này tiếp tục lặp lại cho đến khi tất cả M cây trong quần thể đều được huấn luyện.

Trong kỹ thuật này, có một tham số quan trọng được gọi là "Độ co rút" (shrinkage rate). "Độ co rút" thể hiện thực tế là dự đoán của từng cây trong quần thể bị thu hẹp lại sau khi nó được nhân với tốc độ học (η) nằm trong khoảng từ 0 đến 1. Có sự đánh đổi giữa η và số lượng ước tính (M), điều này làm giảm tốc độ học cần được bù bằng các công cụ ước tính tăng dần để đạt được hiệu suất mô hình nhất định. Vì tất cả các cây đều được huấn luyện ngay bây giờ, nên có thể đưa ra các dự đoán.

Thuật toán phân tích và máy học theo sáng chế như sau:

- Khởi tạo giá trị pseudo-residuals là bằng nhau cho từng điểm dữ liệu.

- Tại vòng lặp thứ i :

+ Một tập hợp dữ liệu đào tạo (Train model) mới được thêm vào để gán vào giá trị của pseudo-residuals đã có

+ Tính giá trị confidence score c_i của mô hình vừa huấn luyện.

+ Cập nhật mẫu chính theo công thức $W = W + c_i * w_i$

+ Cuối cùng, tính toán giá trị pseudo-residuals $-\eta \partial w \partial L(W_{n-1})$ để làm đánh dấu (label) cho mô hình (model) tiếp theo

- Sau đó lặp lại với vòng lặp $i + 1$.

Dự đoán cuối cùng được đưa ra bởi công thức

$$y(\text{pred}) = y_1 + (\eta * r_1) + (\eta * r_2) + \dots + (\eta * r_N)$$

Để tìm kiếm tham số tốt nhất cho mô hình XGBoost, sáng chế đề xuất sử dụng kỹ thuật Random Search kết hợp với cross-validation. Phương pháp này giúp tìm ra các giá trị tham số tối ưu cho mô hình XGBoost dựa trên các giá trị thử nghiệm chỉ định. Random Search sẽ thử một số lượng tham số ngẫu nhiên được chỉ định và đánh giá mô hình với từng tổ hợp này để tìm ra tổ hợp tốt nhất.

Theo sáng chế, các tham số đầu vào bao gồm kết quả giải trình tự của 471 đoạn gen của 43 gen liên quan đến ung thư phổi tế bào không nhỏ bao gồm AKT2, ERBB2, RB1, SMARCA4, STK11, EGFR, POLD1, ERCC1, GNAS, BIRC5, BCL2L11, AURKA, FGFR1, TYMS, ETS2, DDR2, PIK3CA, ABCB1, ALK, NFE2L2, PDCD4, ERCC2, TP53, CD274, RET, BRAF, RICTOR, ROS1, NTRK1, WEE1, AREG, XRCC1, KRAS, ARAF, NRG1, RRM1, HAVCR2, MET, CDK4, BRCA1, CCND1, PDGFRA, AXL, các chỉ số lâm sàng, cận lâm sàng như đặc điểm tuổi, kết quả mô bệnh học của mô hình cần đánh giá.

Quá trình xây dựng mô hình dự báo đáp ứng điều trị đích ở bệnh nhân ung thư phổi không tế bào nhỏ bằng XGBoost có thể được thực hiện như sau:

Bước 1: Thu thập dữ liệu và tiền xử lý dữ liệu:

Thu thập dữ liệu về kết quả giải trình tự của 417 đoạn gen có thông tin đột biến gen của 43 gen bao gồm EGFR, KRAS, ALK, STK11, FGFR1, PIK3CA, ARAF, NRG1, RRM1, HAVCR2, MET, CDK4, BRCA1, CCND1, PDGFRA, AXL, AURKA,

TYMS, ETS2, DDR2, ERCC2, ABCB1, TP53, NFE2L2, PDCD4, RET, BRAF, ROS1, NTRK1, XRCC1, CD274, RICTOR, AREG, BCL2L11, BIRC5, ERCC1, GNAS, AKT2, ERBB2, POLD1, RB1, SMARCA4 và WEE1 nêu trên, các chỉ số lâm sàng, cận lâm sàng như đặc điểm tuổi, kết quả mô bệnh học của mô hình cần đánh giá.

Tiền xử lý dữ liệu bằng cách chuẩn hóa, xử lý các giá trị thiếu, và chuyển đổi các biến hạng mục thành biến số nếu cần thiết.

Bước 2: Chia dữ liệu thành tập huấn luyện và tập kiểm tra:

Chia dữ liệu thành hai phần: tập huấn luyện và tập kiểm tra. Tập huấn luyện được sử dụng để huấn luyện mô hình, trong khi tập kiểm tra được sử dụng để đánh giá hiệu suất của mô hình sau khi huấn luyện.

Bước 3: Xây dựng mô hình XGBoost:

Sử dụng XGBoost để xây dựng mô hình dự báo đáp ứng điều trị đích ở bệnh nhân ung thư phổi không tế bào nhỏ.

Đầu vào cho mô hình bao gồm các biến: kết quả giải trình tự của các gen, đặc điểm tuổi, kết quả mô bệnh học và tình trạng hút thuốc.

Đầu ra của mô hình là khả năng đáp ứng hay không của bệnh nhân ung thư phổi không tế bào nhỏ đối với điều trị đích.

Bước 4: Huấn luyện và đánh giá mô hình:

Huấn luyện mô hình XGBoost trên tập huấn luyện.

Đánh giá hiệu suất của mô hình trên tập kiểm tra bằng các độ đo phù hợp như độ chính xác, độ nhạy (sensitivity), độ đặc hiệu (specificity), F1-score, AUC-ROC,...

Để tính độ nhạy (sensitivity) và độ đặc hiệu (specificity) trong mô hình MLP (Multilayer Perceptron) đã xây dựng, bạn cần sử dụng ma trận hỗn hợp (confusion matrix) của mô hình và tính các chỉ số dựa trên các giá trị trong ma trận hỗn hợp. Ma trận hỗn hợp là một bảng chứa các giá trị thực tế của nhãn và các giá trị dự đoán của mô hình cho dữ liệu kiểm tra.

Ma trận hỗn hợp thường có dạng như sau:

	Actual Positive	Actual Negative
Predicted Positive	True Positive (TP)	False Positive (FP)
Predicted Negative	False Negative (FN)	True Negative (TN)

Trong đó:

True Positive (TP): Số lượng điểm dữ liệu được dự đoán đúng là Positive (dương tính).

False Positive (FP): Số lượng điểm dữ liệu được dự đoán sai là Positive (dương tính) nhưng thực tế là Negative (âm tính).

False Negative (FN): Số lượng điểm dữ liệu được dự đoán sai là Negative (âm tính) nhưng thực tế là Positive (dương tính).

True Negative (TN): Số lượng điểm dữ liệu được dự đoán đúng là Negative (âm tính).

Dựa trên ma trận hỗn hợp, bạn có thể tính các chỉ số sau:

Độ Nhảy (Sensitivity hay Recall hay True Positive Rate - TPR): Được tính bằng tỷ lệ số lượng True Positive (TP) trên tổng số Positive thực tế (TP + FN). Chỉ số này cho biết khả năng của mô hình dự đoán chính xác các điểm dữ liệu Positive.

$$\text{Sensitivity (TPR)} = \text{TP} / (\text{TP} + \text{FN})$$

Độ đặc hiệu (Specificity hay True Negative Rate - TNR): Được tính bằng tỷ lệ số lượng True Negative (TN) trên tổng số Negative thực tế (TN + FP). Chỉ số này cho biết khả năng của mô hình dự đoán chính xác các điểm dữ liệu Negative.

$$\text{Specificity (TNR)} = \text{TN} / (\text{TN} + \text{FP})$$

Sáng chế sử dụng các phương pháp tối ưu hóa tham số Random Search để tìm các tham số tốt nhất cho mô hình.

Tham số tốt nhất dành cho mô hình theo sáng chế bao gồm Số lượng cây quyết định độc lập tạo ra (n_estimators): 100, Độ sâu tối đa của cây (max_depth): 13, tỷ lệ học (learning_rate): 0.3.

Bước 6: Đưa ra dự đoán:

Sau khi huấn luyện mô hình, dựa trên dữ liệu từ cơ sở dữ liệu di truyền chuẩn, cơ sở dữ liệu so sánh sẽ thiết lập được các dự đoán về khả năng đáp ứng của thuốc đích. Việc nạp dữ liệu đánh giá với các thông số xác định có thể đưa ra dự đoán về khả năng đáp ứng thuốc đích dùng trong điều trị ung thư phổi không tế bào nhỏ.

Trong bước hiệu chỉnh dữ liệu so sánh, tiến hành khởi tạo lại dữ liệu đánh giá theo một khoảng thời gian nhất định và so sánh với dữ liệu được nạp trước đó theo mô hình máy học. Đồng thời tiến hành phân tích máy học để hiệu chỉnh các yếu tố di truyền gây phát sinh ung thư phổi không tế bào nhỏ theo mô hình XGBoost theo thư viện SHAP (SHapley Additive exPlanations) và đột biến tương tác gen. Trong đó yếu tố gây phát sinh ung thư phổi không tế bào nhỏ được xác định theo thứ tự giảm dần theo gồm đột biến phát sinh trên gen EGFR, ALK, các chỉ số lâm sàng, đột biến phát sinh trên gen KRAS, ROS, FGFR1, AKT1, HER2, MEK, RET, chỉ số cận lâm sàng, BRAF, DDR2, MTRK1, MEK1, MET, NRAS, PTEN, PIK3CA và các gen còn lại.

Để tính giá trị Shapley cho từng đặc trưng thể hiện đóng góp của từng đặc trưng (features) đến kết quả dự báo trong mô hình XGBoost. Thư viện SHAP (SHapley Additive exPlanations) được sử dụng để cung cấp công cụ nhằm đánh giá các dự đoán của mô hình và tính toán giá trị Shapley cho từng đặc trưng. Sáng chế đã sử dụng thư viện SHAP để tính giá trị Shapley cho từng đặc trưng trong mô hình XGBoost. Hàm `shap.summary_plot` sẽ vẽ biểu đồ dạng cột (bar chart) thể hiện giá trị Shapley của các đặc trưng. Các cột càng cao thể hiện đóng góp tích cực của đặc trưng đó đến dự đoán kết quả, còn các cột càng thấp thể hiện đóng góp tiêu cực của đặc trưng đó đến dự đoán kết quả.

Dựa trên kết quả đánh giá và hiệu chỉnh mô hình phân tích theo thời gian và dữ liệu bổ sung, mô hình cho phép đánh giá khả năng đáp ứng thuốc và có cơ chế phản hồi tích cực, cho phép đánh giá được chính xác các mối liên quan cũng như các mối tương tác quan hệ giữa đột biến gen, các chỉ số lâm sàng và các chỉ số cận lâm sàng tương hỗ hoặc kìm hãm khả năng hướng đích của thuốc.

Để đánh giá các yếu tố tác động tương tác giữa các yếu tố di truyền và môi trường đến khả năng đáp ứng thuốc đích, thực hiện phương pháp phân tích bằng mô hình MDR.

Mô hình MDR (Multifactor Dimensionality Reduction) là một phương pháp phân tích dữ liệu được sử dụng để xác định và đánh giá các tương tác giữa nhiều yếu tố (biến) trong dữ liệu. Phương pháp này được sử dụng phổ biến trong nghiên cứu di truyền học và y sinh nhằm đánh giá các yếu tố tăng cường và/hoặc kìm hãm giữa các yếu tố đầu vào.

Mục tiêu chính của MDR là tìm ra các tập con (subsets) của các yếu tố đa biến (multifactor) mà có tương tác đáng kể đối với đầu ra (như khả năng đáp ứng điều trị hay bệnh lý). Kỹ thuật này giúp giảm chiều dữ liệu và tìm ra các mẫu tương tác giữa các yếu tố phức tạp.

Đáp ứng điều trị bệnh ung thư phổi được cho là do sự tương tác của nhiều yếu tố, yếu tố di truyền và yếu tố lâm sàng, cận lâm sàng (yếu tố môi trường). Để xác định mức độ ảnh hưởng của các yếu tố cần đánh giá và từ đó chọn ra 10 yếu tố quan trọng nhất đưa vào mô hình MDR.

Các bước chính trong MDR bao gồm:

Xây dựng tập con: dữ liệu được chia thành hai phần: tập huấn luyện (training set) và tập kiểm tra độc lập (independent testing set) để thực hiện kiểm tra chéo (cross-validation). Sau đó, một tập hợp n các yếu tố (factors) được chọn từ tất cả các yếu tố có sẵn. Các n yếu tố và các tập hợp tương tác (multifactor cells) có thể có của chúng được biểu thị trong không gian n chiều.

Tính toán ma trận suy biến: Đối với mỗi tập con, MDR tính toán ma trận suy biến để xác định mức độ tương tác giữa các yếu tố trong tập con đó. Ma trận suy biến đo lường tương tác giữa các yếu tố bằng cách đánh giá hiệu suất dự báo khi kết hợp các yếu tố lại với nhau.

Xác định tập con tốt nhất: MDR đánh giá mức độ tương tác cho tất cả các tập con và chọn tập con có mức độ tương tác tốt nhất (chẳng hạn như có hiệu suất dự báo tốt nhất) để xem xét tương tác giữa các yếu tố trong đó. Chọn mô hình có tỷ lệ sai lầm phân loại thấp nhất và ước tính lỗi dự đoán của mô hình bằng cách sử dụng dữ liệu kiểm tra độc lập. Các bước trên được lặp lại cho mỗi khoảng kiểm tra chéo có thể có.

MDR là một phương pháp phi tham số dựa trên xây dựng mô hình tuyến tính (linear programming) và kỹ thuật nhóm dữ liệu (data pooling), điều này có nghĩa là nó không dựa vào các giả định cụ thể về phân phối dữ liệu và không yêu cầu các tham số cụ thể. Nó có thể xử lý các biến hạng mục và biến liên tục và phù hợp cho các tập dữ liệu lớn.

Dựa trên mô hình MDR (Multifactor Dimensionality Reduction), các yếu tố tương tác để đánh giá bao gồm giữa các yếu tố di truyền và môi trường đối với khả năng đáp ứng thuốc đích. Điều này cho phép giúp xác định các mối quan hệ giữa các yếu tố di

truyền, môi trường và từ đó đưa ra đánh giá ảnh hưởng của từng yếu tố này, so sánh thứ tự ưu tiên chúng đến khả năng đáp ứng thuốc.

Mô hình MDR nhằm dự báo đáp ứng thuốc đích dựa trên sáng chế bộ dữ liệu chứa thông tin về các yếu tố di truyền, môi trường và đáp ứng thuốc điều trị bệnh ung thư phổi không tế bào nhỏ đã được xây dựng ở trên. Mô hình này cho phép dự báo dựa trên các chỉ số có giá trị nhất ảnh hưởng đến dự báo kết quả điều trị bao gồm đa hình trên 16 gen AKT2, ALK, BRAF, DDR2, EGFR, FGFR1, HER2, KRAS, MEK1, MET, NRAS, NTRK1, PIK3CA, PTEN, RET và ROS1, cũng như các yếu tố ngoài gen khác có liên quan đến bệnh ung thư phổi không tế bào nhỏ như tình trạng hút thuốc, phân loại mô bệnh học, giai đoạn bệnh.

Khi hiệu chỉnh dữ liệu so sánh, cần bổ sung ít nhất một thông số dựa trên 16 gen ảnh hưởng bao gồm AKT2, ALK, BRAF, DDR2, EGFR, FGFR1, HER2, KRAS, MEK1, MET, NRAS, NTRK1, PIK3CA, PTEN, RET và ROS1. Do đó, dữ liệu đầu vào của mô hình MDR bao gồm các biến di truyền liên quan đến ung thư phổi không tế bào nhỏ như các đa hình trên gen bất kỳ trong số AKT2, ALK, BRAF, DDR2, EGFR, FGFR1, HER2, KRAS, MEK1, MET, NRAS, NTRK1, PIK3CA, PTEN, RET và ROS1, các yếu tố ngoài gen khác có liên quan đến bệnh ung thư phổi không tế bào nhỏ như tình trạng hút thuốc, phân loại mô bệnh học, giai đoạn bệnh.

Sau khi chuẩn hóa bộ dữ liệu, sáng chế thực hiện các bước sau sử dụng MDR:

Chuẩn bị dữ liệu: Xác định và chọn các biến di truyền và đáp ứng thuốc từ bộ dữ liệu. Tiền xử lý dữ liệu bằng cách xử lý dữ liệu bị thiếu, chuẩn hóa dữ liệu và chuyển đổi dữ liệu thành định dạng phù hợp cho MDR.

Sau khi sử dụng mô hình phân tích MDR bằng phần mềm MDR trên bộ dữ liệu đã chuẩn hóa. MDR sẽ xây dựng các tập con từ các biến di truyền và đánh giá tương tác giữa chúng để xác định tập con quan trọng có tương tác với đáp ứng thuốc. Từ kết quả đánh giá theo mô hình máy cho ra các thông số liên quan đến khả năng đáp ứng tích cực, đáp ứng một phần hoặc không đáp ứng để từ đó xác định được mối tương tác giữa các biến di truyền và khả năng đáp ứng thuốc đích. Kết quả này sẽ được sử dụng trở lại cho cơ sở dữ liệu so sánh nhằm tăng cường khả năng tương tác dự đoán cho các mô hình tiếp theo.

Theo một phương án, quy trình theo sáng chế còn được đánh giá nhằm hoàn thiện quá trình và tăng cường hiệu quả của mô hình học máy. Quá trình đánh giá dựa trên các phản hồi lâm sàng thu được từ quá trình điều trị bệnh từ bệnh nhân áp dụng mô hình. Tuy nhiên, việc áp dụng mô hình không thuộc phạm vi của sáng chế, các dữ liệu thu được có thể thông qua các cơ sở tập hợp dữ liệu dựa trên các thông tin hoặc các điều chỉnh từ cơ sở dữ liệu thuốc đích.

Quy trình đánh giá kết quả điều trị được thực hiện liên tục và đều đặn kể từ khi bệnh nhân áp dụng kết quả mô hình phân tích. Sau khi bệnh nhân điều trị thuốc đích, các thông tin thu thập được cập nhật nhằm theo dõi đáp ứng của thuốc và tác dụng phụ, nếu có. Định kỳ các thông số phân tích lâm sàng được thu thập kèm theo các kết quả xét nghiệm để đánh giá lại kết quả sơ bộ, kèm theo xét nghiệm công thức máu, sinh hoá máu kiểm tra bất thường nếu có. Các dữ liệu này được cung cấp để cập nhật vào cơ sở dữ liệu, điều chỉnh nếu các kết quả đánh giá trả về xuất hiện kháng hoặc không hiệu quả. Các dữ liệu này được cập nhật để định hướng phân tích cho mô hình tiếp theo.

Theo mục đích thứ hai, sáng chế đề cập đến kit dùng trong phương pháp đánh giá yếu tố di truyền gây phát sinh ung thư phổi tế bào không nhỏ nhằm giải trình tự xác định đồng thời 43 gen gây phát sinh ung thư phổi tế bào không nhỏ bao gồm EGFR, KRAS, ALK, STK11, FGFR1, PIK3CA, ARAF, NRG1, RRM1, HAVCR2, MET, CDK4, BRCA1, CCND1, PDGFRA, AXL, AURKA, TYMS, ETS2, DDR2, ERCC2, ABCB1, TP53, NFE2L2, PDCD4, RET, BRAF, ROS1, NTRK1, XRCC1, CD274, RICTOR, AREG, BCL2L11, BIRC5, ERCC1, GNAS, AKT2, ERBB2, POLD1, RB1, SMARCA4 và WEE1, trong đó kit này bao gồm các trình tự nêu trong SEQ ID NO.1 đến SEQ ID NO.942.

Ví dụ thực hiện sáng chế

Ví dụ 1. Thiết lập cơ sở dữ liệu và mô hình phân tích máy học

+ Thiết lập cơ sở dữ liệu di truyền chuẩn

Dữ cơ sở dữ liệu di truyền chuẩn được thiết lập với các việc tiếp cận và tập hợp dữ liệu 43 gen trên ngân hàng gen tại địa chỉ <https://www.ncbi.nlm.nih.gov/>, các dữ liệu được tập hợp bao gồm EGFR, KRAS, ALK, STK11, FGFR1, PIK3CA, ARAF, NRG1, RRM1, HAVCR2, MET, CDK4, BRCA1, CCND1, PDGFRA, AXL, AURKA, TYMS,

ETS2, DDR2, ERCC2, ABCB1, TP53, NFE2L2, PDCD4, RET, BRAF, ROS1, NTRK1, XRCC1, CD274, RICTOR, AREG, BCL2L11, BIRC5, ERCC1, GNAS, AKT2, ERBB2, POLD1, RB1, SMARCA4 và WEE1.

Từng gen được phân tích toàn cục để xác định các vị trí đột biến gen, mỗi thông số của gen được tập hợp tạo thành bộ dữ liệu di truyền chuẩn và được thiết lập theo các trường dữ liệu bao gồm các thông tin liên quan đến gen, trình tự, kích thước, vị trí các đột biến điểm.

+ Thiết lập cơ sở dữ liệu so sánh

Dựa trên cơ sở dữ liệu được tập hợp từ các bệnh viên chuyên khoa, dữ liệu được tập hợp bao gồm 1450 dữ liệu về bệnh nhân được thăm khám xác định liên quan đến ung thư phổi không tế bào nhỏ. Trong đó các gen được phân tích bao gồm EGFR, KRAS, ALK, STK11, FGFR1, PIK3CA, ARAF, NRG1, RRM1, HAVCR2, MET, CDK4, BRCA1, CCND1, PDGFRA, AXL, AURKA, TYMS, ETS2, DDR2, ERCC2, ABCB1, TP53, NFE2L2, PDCD4, RET, BRAF, ROS1, NTRK1, XRCC1, CD274, RICTOR, AREG, BCL2L11, BIRC5, ERCC1, GNAS, AKT2, ERBB2, POLD1, RB1, SMARCA4 và WEE1, trong đó mỗi vị trí đột biến điểm được ưu tiên sao cho dữ liệu chứa ít nhất một hồ sơ phân tích liên quan đến vị trí đột biến điểm liên quan đến phát sinh bệnh của mỗi gen. Các dữ liệu được tập hợp bao gồm các thông số liên quan đến mô hình phát sinh bệnh ung thư phổi không tế bào nhỏ bao gồm vị trí các đột biến trên các gen, các chỉ số lâm sàng, cận lâm sàng, thuốc đích và trình tự các gen được giải.

Mỗi dữ liệu được tập hợp dưới dạng các thông số kỹ thuật về các chỉ số xét nghiệm, các thông tin về gen, đột biến. Các thông tin về thăm khám lâm sàng, cận lâm sàng cũng như tiền sử phát sinh bệnh cũng được tập hợp với các trường mở. Các thông số độc lập được thiết lập theo các trường phân tích nhằm đưa tối đa các tham số phân tích cho phép bao phủ toàn bộ các yếu tố tác động.

+ Khởi tạo dữ liệu đánh giá

Mô hình cần đánh giá được thiết lập tạo dữ liệu, trong đó tiến hành thu thập các thông tin tối đa theo thông tin từ các trường khởi tạo từ dữ liệu so sánh. Các trình tự được giải từ mẫu mô ung thư đúc parafin thu được từ đối tượng cần áp dụng mô hình đánh giá.

Tiến hành chiết ADN từ mẫu mô ung thư đúc paraffin (FFPE) bằng cách sử dụng kit HiPure FFPE DNA Mini Kit theo hướng dẫn của nhà cung cấp. Kiểm tra chất lượng và nồng độ ADN sau tách chiết bằng máy quang phổ Qubit 4 Fluorometer với bộ kit Qubit DNA HS Assay Kit, yêu cầu nồng độ ADN ≥ 2 ng/ μ L. Bảo quản mẫu ADN ở -20°C cho các bước thực hiện tiếp theo.

Sau đó các mẫu ADN được dùng để tạo thư viện sử dụng bộ hóa chất Ion Plus Fragment Library. Dựa vào kết quả định lượng mẫu thư viện, pha loãng và trộn các mẫu thư viện theo tỉ lệ vào 1 ống 1,5 ml sao cho hỗn hợp thư viện có nồng độ 30 pM. Hỗn hợp thư viện được khuếch đại một lần nữa bằng hệ thống máy Ion OneTouch2 bằng bộ kit hóa chất Ion PI Hi-Q OT2 200.

Mẫu thư viện được làm giàu bằng hệ thống máy ES, sử dụng bộ kit hóa chất Ion PI Hi-Q OT2 200 rồi giải trình tự tự động trên hệ thống máy Ion Torrent, sử dụng bộ kit Ion PI Hi-Q 200 Sequence: Hỗn hợp mẫu ADN thư viện được nạp vào Ion PI Chip V2 và giải tự động trên máy Ion Torrent bằng kit theo sáng chế cho phép giải đồng thời 471 đoạn gen tương ứng với 43 gen cần đánh giá. Dữ liệu giải trình tự được chuyển trực tiếp vào thư mục ADN trong máy chủ của hệ thống Ion Torrent (Ion Torrent Server). Dữ liệu giải trình tự được đối chiếu với hệ gen tham chiếu GRCh37 (NCBI) để xác định các biến thể và thu kết quả.

+ Thiết lập mô hình phân tích máy học

Dữ liệu sau khi được thu thập được đưa vào thiết lập phân tích máy học theo mô hình XGBoost tăng cường. Các thông số khởi tạo được nạp vào và máy tiến hành với truy vấn ngẫu nhiên (random search) và đánh giá chéo (cross-validation), số trường quyết định độc lập ($n_estimator$) là 100, độ sâu trường (max_depth) là 13, tỷ lệ học ($learning_rate$) là 0,3.

Các chỉ số được thiết lập trong tập hợp dữ liệu đánh giá với cơ sở dữ liệu các thông số liên quan đến mô hình phát sinh bệnh và so sánh ít nhất hai trình tự được giải với cơ sở dữ liệu di truyền chuẩn theo mô hình XGBoost tăng cường nêu trên. Kết quả thu được các thông số đột biến gen liên quan đến thuốc điều trị đích trên cơ sở so sánh và phân tích chéo các dữ liệu liên quan đến mô hình phát sinh bệnh.

+ Hiệu chỉnh dữ liệu so sánh

Dữ liệu đánh giá đối với một tập hợp được hiệu chỉnh bằng cách khởi tạo lại dữ liệu đánh giá theo một khoảng thời gian nhất định, thường từ 2 đến 3 tháng. Thường là sau khi tiến hành đánh giá và áp dụng đối với mô hình phát sinh bệnh với mỗi khoảng thời gian sẽ tiến hành thu thập lại dữ liệu phân tích. Trên cơ sở hiệu so sánh lần hai với kết quả dữ liệu được nạp trước đó theo mô hình máy học. Tiến hành phân tích trên máy để hiệu chỉnh các yếu tố di truyền lần tiếp theo bằng cách lặp lại phương pháp đánh giá theo mô hình XGBoost trên cơ sở thư viện SHAP (SHapley Additive exPlanations) và đột biến tương tác gen. Các yếu tố đánh giá được thiết lập lại dựa trên các yếu tố so sánh. Trường hợp các đột biến được xác định không hiệu quả hoặc bị kháng bởi các đột biến, các yếu tố phân tích tiếp theo sẽ được đề xuất dựa trên cơ sở ưu tiên các chỉ số ảnh hưởng.

Giá trị Shapley cho từng đặc trưng đối với các gen được đề xuất. Kết quả đánh giá các yếu tố gây phát sinh ung thư phổi không tế bào nhỏ qua mô hình thư viện SHAP được thể hiện trên Hình 1. Giá trị Shapley thể hiện mức độ đóng góp của từng yếu tố đến hiệu quả đáp ứng thuốc đích được đề xuất theo mô hình phân tích. Trong đó các yếu tố gây phát sinh ung thư phổi không tế bào nhỏ giảm dần theo thứ tự AGFR, ALK, hút thuốc (Smok), Hist, KRAS, ROS, FGFR1, AKT1, HER2, MEK, RET, độ tuổi, BRAF, DDR2, MTRK1, MEK1, MET, NRAS, PTEN, PIK3CA và các gen còn lại. Trong đó các giá trị ảnh hưởng được ưu tiên xác định theo mô hình từ trên xuống dưới.

Biểu đồ Shapley là một công cụ hữu ích để hiểu mức độ đóng góp của từng đặc trưng (feature) đến kết quả (outcome) trong một mô hình dự đoán được thiết lập. Dựa trên cơ sở thiết lập mô hình nhiệm vụ dưới dạng bài toán phân loại nhị phân, trong đó một gen được dán nhãn có “1” hoặc không “0” (Hình 3). Các yếu tố môi trường như hút thuốc, tiền sử gia đình có người mắc ung thư hay không cũng được quy ước theo cách này. Kết quả đánh giá dựa trên tổ hợp theo quy ước: Class 0 tương ứng với đáp ứng hoàn toàn, class 1 là đáp ứng một phần và class 2 tương ứng với bệnh ổn định (không tiến triển cũng không nặng thêm).

Phân tích được tiến hành bằng cách chia bộ dữ liệu thành 2 tập con: tập huấn luyện (training) và tập kiểm tra (test) theo tỷ lệ 8:2.

Kết quả, biểu đồ Shapley cho mô hình dự đoán về mức độ đáp ứng của thuốc điều trị đích đối với bệnh nhân ung thư phổi không tế bào nhỏ, gồm có một biểu đồ chứa các

thanh cột đại diện cho đặc trưng và chiều dài của mỗi cột thể hiện mức độ đóng góp của đặc trưng đó đối với kết quả dự đoán. Với mô hình theo sáng chế, các biến đổi trên gen EGFR và ALK trong khối u của người bệnh có mức ảnh hưởng lớn nhất đến khả năng đáp ứng thuốc của bệnh nhân. Tiếp theo đó là tiền sử hút thuốc lá. Ngoài ra, một số yếu tố tiềm năng có thể ảnh hưởng đến khả năng đáp ứng thuốc điều trị đích của bệnh nhân có thể tiếp tục nghiên cứu và cảnh báo bao gồm KRAS, ROS, FGFR1, AKT1.

Trường hợp xác định mô hình xuất hiện đột biến kháng thuốc, ví dụ xuất hiện đột biến trên EGFR, khi đó mô hình sẽ đề xuất thuốc hướng đích được loại bỏ đối với đột biến EGFR, ưu tiên đến thuốc hướng đích đặc hiệu ALK và tương tự. Trường hợp xuất hiện đột biến mới, mô hình sẽ đánh dấu vị trí nhằm xác định mức độ ảnh hưởng của đột biến đó liên quan đến phát sinh bệnh.

Trong đó yếu tố gây phát sinh ung thư phổi không tế bào nhỏ được xác định theo thứ tự giảm dần theo gồm đột biến phát sinh trên gen EGFR, ALK, các chỉ số lâm sàng, đột biến phát sinh trên gen KRAS, ROS, FGFR1, AKT1, HER2, MEK, RET, chỉ số cận lâm sàng, BRAF, DDR2, MTRK1, MEK1, MET, NRAS, PTEN, PIK3CA.

Trên mô hình phân tích máy học này thử nghiệm với thông số kiểm tra nhằm minh họa mức độ ảnh hưởng của từng đặc trưng đối với từng đầu ra riêng lẻ đến khả năng đáp ứng thuốc đích. Như sơ đồ kết quả đánh giá trên hình 2, các thang bên trái màu đỏ thể hiện đóng góp tích cực của đặc trưng đó đến dự đoán kết quả, còn các thang bên phải thể hiện đóng góp tiêu cực của đặc trưng đó đến dự đoán kết quả. Trong trường hợp này, việc kiểm tra các thông số như tiền sử hút thuốc, không có biến đổi trên các gen AKT1, NRAS, KRAS, ALK, MET thể hiện khả năng đóng góp tích cực đến việc bệnh nhân không kháng thuốc điều trị đích đề xuất. Tuy nhiên, nếu xuất hiện đột biến trên gen EGFR và mô bệnh học cho thấy tăng nguy cơ kháng thuốc hướng đích dự kiến. Tổng hợp các yếu tố, bệnh nhân có nguy cơ kháng thuốc tăng 3,06 lần.

Hình 4 là kết quả chạy mô hình thử nghiệm kiểm tra tương tác của các gen và tiền sử hút thuốc đến khả năng đáp ứng thuốc hướng đích. Dữ liệu được quy ước đối với từng biến đổi trên các gen là 0 tương ứng với mang đột biến dạng đồng hợp tử, 1 là đột biến dị hợp tử và 2 là không mang đột biến. Trong trường hợp này, nếu bệnh nhân có hút thuốc, đồng thời có mang cả đột biến trên EGFR và nhưng không mang đột biến trên gen KRAS làm tăng mức độ đáp ứng thuốc lên 1,283 lần. Ngược lại, mang đồng thời

đột biến trên gen EGFR và KRAS sẽ làm giảm đáp ứng với thuốc điều trị đích 0,391 lần.

Mỗi tương tác gen được đánh giá dựa trên các thông tin và được lượng hóa liên quan đến khoảng cách tương tác giữa các gen và các yếu tố lâm sàng dựa trên kết quả tính toán mô hình (Hình 5) thể hiện mỗi tương tác giữa các yếu tố khác nhau đến khả năng đáp ứng thuốc điều trị đích. Trường hợp đánh giá lệch chuẩn, các đột biến mới phát sinh được đưa vào đánh giá và thực hiện tiến trình lặp tiếp theo với mô hình giảm thứ tự ưu tiên. Với vòng lặp thứ ba, sau khi đánh giá sẽ cho phép xác định được mức độ ảnh hưởng của đột biến mới phát sinh đến mô hình phát sinh bệnh. Từ đó cho phép loại bỏ hoặc cập nhật vị trí đột biến mới này vào cơ sở dữ liệu chuẩn và cơ sở dữ liệu so sánh cho phép dữ liệu được hoàn thiện tốt hơn theo thời gian thực.

Ví dụ 2. Áp dụng mô hình phân tích đánh giá yếu tố di truyền đến đáp ứng thuốc điều trị đích trên bệnh nhân

Dữ liệu bệnh học từ đối tượng được bệnh nhân nam, 58 tuổi được thu thập. Các thông tin lâm sàng, cận lâm sàng được tiền sử bệnh được thiết lập. Dữ liệu cho thấy bệnh nhân bị ho, đau ngực kèm theo đau đầu, tiền sử khoẻ mạnh, hút thuốc lá 20 bao/năm, tiến triển bệnh trong 1 tháng xuất hiện ho khan, đau tức ngực, đôi khi có đau đầu buồn nôn, đi khám kiểm tra tuyến YTCS nhập viện điều trị, nhập viện ngày 01/09/2022. Khám lúc vào viện bệnh nhân tỉnh, tiếp xúc được, còn ho và đau đầu, buồn nôn nhiều, mệt mỏi, ăn uống kém gây sút cân, phổi thông khí giảm 2 bên.

Kết quả CLVT: hình ảnh khối mờ thùy dưới phổi T kích thước 3x4cm, nốt mờ lan toả 2 phổi, hạch trung thất to

Kết quả mô bệnh qua sinh thiết phổi/CT: ung thư biểu mô tuyến phổi

MRI não: hình ảnh di căn não đa ổ

Bệnh nhân được chỉ định làm xét nghiệm ĐBG bằng phương pháp NGS

Kết quả phân tích gen cho thấy xuất hiện đột biến EGFR xoá đoạn trên exon 19. Đây là đột biến nhạy thuốc điều trị nhắm trúng đích. Ngoài ra không phát hiện đột biến gen khác. Phần mềm chỉ ra các thuốc phù hợp cho điều trị đối với đột biến gen này là: Erlotinib, Gefitinib, Afatinib, Osimertinib. Sau khi sử dụng thuốc kháng TKIs Iressa (Gefitinib) 250mg/ngày sau 30 ngày, kết quả được thể hiện trên Hình 7 và Hình 8.

Bệnh nhân nam 54 tuổi vào viện vì ho, đau ngực. Tiền sử THA, ĐTĐ 3 năm nay điều trị thường xuyên. Bệnh diễn biến 2 tháng trước vào viện, bệnh nhân xuất hiện ho khan, đau tức ngực, mệt mỏi, ăn kém kèm theo đau đầu nhiều. Nhập viện ngày 14/07/2022. Khám lúc vào viện: bệnh nhân tỉnh, tiếp xúc chậm, mệt mỏi, ăn kém, hạn chế vận động, liệt $\frac{1}{2}$ người phải.

Phim CLVT: Hình ảnh khối mờ thùy trên phổi phải kích thước 5x6cm, hạch lớn trung thất

Kết quả mô bệnh qua sinh thiết/CT: ung thư biểu mô tuyến

MRI não: có hình ảnh di căn não ổ lớn kèm theo phù não diện rộng xung quanh

Bệnh nhân được chỉ định làm xét nghiệm đột biến gen bằng công nghệ NGS. Kết quả : phát hiện đột biến EGFR, PIK3CA, KRAS, KIT, PTEN. Kết quả chỉ ra các thuốc điều trị nhắm trúng đích EGFR có thể tương tác với các đột biến gen cùng xuất hiện trên phân tích sinh học phân tử ở bệnh nhân này. Thuốc này có thể gây tình trạng kháng lại với thuốc điều trị. Sau khi dùng thuốc kháng TKIs thế hệ 2: Giotrif (Afatinib) 30mg/ngày. Kết quả điều trị 30 ngày được thể hiện trên Hình 9.

Hiệu quả đạt được của sáng chế

Phương pháp đánh giá yếu tố di truyền gây phát sinh ung thư phổi không tế bào nhỏ đến đáp ứng thuốc điều trị đích theo sáng chế cho phép xây dựng, áp dụng mô hình phân tích máy học (AI) vào phân tích các yếu tố di truyền tác động đến các đáp ứng thuốc điều trị đích. Bằng việc xây dựng cơ sở dữ liệu di truyền chuẩn, cơ sở dữ liệu so sánh và dữ liệu đánh giá cho phép phân tích các yếu tố tác động trên cơ sở máy học theo mô hình XGBoost tăng cường với truy vấn ngẫu nhiên (random search) và đánh giá chéo (cross-validation), số trường quyết định độc lập (n_estimator) là 100, độ sâu trường (max_depth) là 13, tỷ lệ học (learning_rate) là 0,3 từ đó thu được các thông số đột biến gen liên quan đến thuốc điều trị đích phù hợp nhất. Phương pháp này cho phép đánh giá được đa yếu tố tác động cho phép đánh giá khả năng đáp ứng thuốc điều trị đích của đối với các yếu tố di truyền gây phát sinh ung thư phổi không tế bào nhỏ.

Phương pháp đánh giá dựa trên mô hình máy học có phản hồi kết quả phân tích cho phép rút ngắn quá trình đánh giá, phân tích, thử nghiệm, phù hợp ứng dụng trong các phương pháp phân tích với mô hình tác động đa biến, ma trận chéo phức tạp với

nhiều yếu tố tác động. Phương pháp này thích hợp ứng dụng trong phân tích thuốc hướng đích, mô hình thử nghiệm, nghiên cứu dược phẩm, đào tạo máy học cho phép ứng dụng công nghệ sinh học và công nghệ máy tính vào cuộc sống, giảm thời gian phân tích.

Phương pháp theo sáng chế cho phép lọc và cập nhật những đột biến gen ảnh hưởng đến mô hình phát sinh bệnh cho phép loại bỏ hoặc kế thừa dữ liệu phân tích để tăng cường hiệu quả đánh giá mô hình bệnh học với các đột biến điểm và các yếu tố di truyền, lâm sàng, cận lâm sàng phức tạp.

YÊU CẦU BẢO HỘ

1. Phương pháp đánh giá yếu tố di truyền gây phát sinh ung thư phổi không tế bào nhỏ đến đáp ứng thuốc điều trị đích, trong đó phương pháp này bao gồm các bước:

a) thiết lập cơ sở dữ liệu di truyền chuẩn đối với các đột biến gen gây phát sinh ung thư phổi không tế bào nhỏ, trong đó tạo tập hợp dữ liệu trình tự gen của 43 gen liên quan bao gồm EGFR, KRAS, ALK, STK11, FGFR1, PIK3CA, ARAF, NRG1, RRM1, HAVCR2, MET, CDK4, BRCA1, CCND1, PDGFRA, AXL, AURKA, TYMS, ETS2, DDR2, ERCC2, ABCB1, TP53, NFE2L2, PDCD4, RET, BRAF, ROS1, NTRK1, XRCC1, CD274, RICTOR, AREG, BCL2L11, BIRC5, ERCC1, GNAS, AKT2, ERBB2, POLD1, RB1, SMARCA4 và WEE1, trong đó dữ liệu của từng gen được so sánh toàn cục để xác định và đánh dấu các vị trí đột biến điểm gây ung thư đã biết;

b) thiết lập cơ sở dữ liệu so sánh với ít nhất mỗi đột biến điểm gây ung thư đã biết được thu thập dữ liệu với các thông số phát sinh bệnh liên quan đến bệnh ung thư phổi không tế bào nhỏ bao gồm các chỉ số lâm sàng, cận lâm sàng, thuốc đích và dữ liệu trình tự gen đột biến trong số các gen EGFR, KRAS, ALK, STK11, FGFR1, PIK3CA, ARAF, NRG1, RRM1, HAVCR2, MET, CDK4, BRCA1, CCND1, PDGFRA, AXL, AURKA, TYMS, ETS2, DDR2, ERCC2, ABCB1, TP53, NFE2L2, PDCD4, RET, BRAF, ROS1, NTRK1, XRCC1, CD274, RICTOR, AREG, BCL2L11, BIRC5, ERCC1, GNAS, AKT2, ERBB2, POLD1, RB1, SMARCA4 và WEE1;

c) khởi tạo dữ liệu đánh giá, trong đó thu thập các chỉ số lâm sàng, cận lâm sàng ứng với mô hình cần đánh giá, đồng thời giải trình tự các đoạn gen chứa đột biến trên gen EGFR, KRAS, ALK, STK11, FGFR1, PIK3CA, ARAF, NRG1, RRM1, HAVCR2, MET, CDK4, BRCA1, CCND1, PDGFRA, AXL, AURKA, TYMS, ETS2, DDR2, ERCC2, ABCB1, TP53, NFE2L2, PDCD4, RET, BRAF, ROS1, NTRK1, XRCC1, CD274, RICTOR, AREG, BCL2L11, BIRC5, ERCC1, GNAS, AKT2, ERBB2, POLD1, RB1, SMARCA4 và WEE1 của mô hình để tạo tập hợp dữ liệu đánh giá;

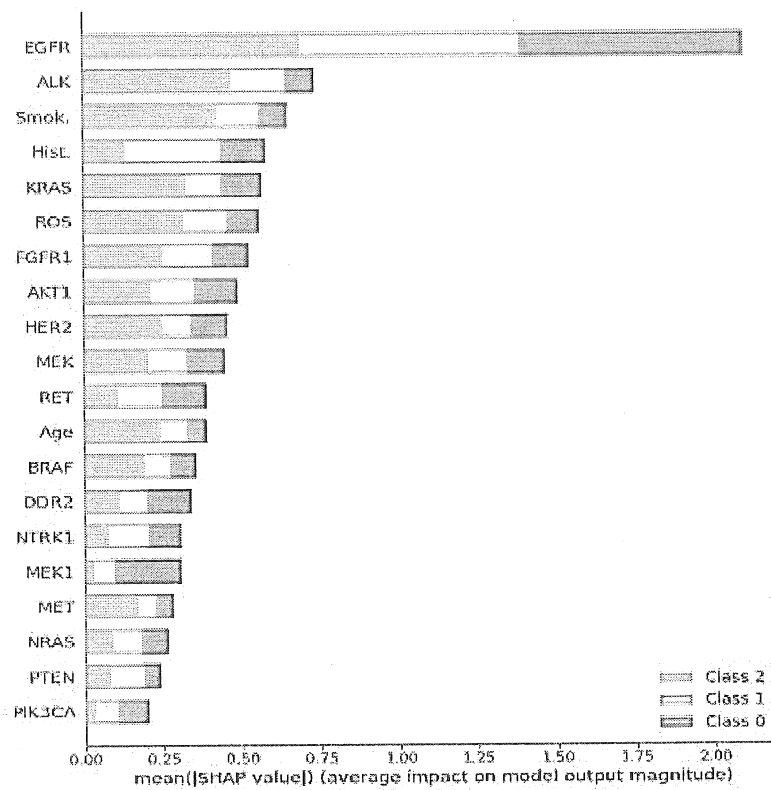
d) thiết lập mô hình phân tích máy học dựa trên các chỉ số di truyền và chỉ số lâm sàng, cận lâm sàng bằng cách tiến hành so sánh song song đồng thời từng chỉ số trong tập hợp dữ liệu đánh giá với từng chỉ số trong cơ sở dữ liệu so sánh với các thông số liên quan đến mô hình phát sinh bệnh và so sánh các đoạn gen chứa đột biến trong số 43 gen liên quan đến bệnh ung thư phổi tế bào không nhỏ bao gồm EGFR, KRAS, ALK, STK11,

FGFR1, PIK3CA, ARAF, NRG1, RRM1, HAVCR2, MET, CDK4, BRCA1, CCND1, PDGFRA, AXL, AURKA, TYMS, ETS2, DDR2, ERCC2, ABCB1, TP53, NFE2L2, PDCD4, RET, BRAF, ROS1, NTRK1, XRCC1, CD274, RICTOR, AREG, BCL2L11, BIRC5, ERCC1, GNAS, AKT2, ERBB2, POLD1, RB1, SMARCA4 và WEE1 được giải với cơ sở dữ liệu di truyền chuẩn theo mô hình XGBoost tăng cường với truy vấn ngẫu nhiên (random search) và đánh giá chéo (cross-validation), số trường quyết định độc lập (n_estimator) là 100, độ sâu trường (max_depth) là 13, tỷ lệ học (learning_rate) là 0,3 từ đó thu được các thông số đột biến điểm của gen liên quan đến thuốc điều trị đích phù hợp nhất có trong cơ sở dữ liệu mô hình phát sinh bệnh đồng thời nạp dữ liệu này vào dữ liệu so sánh; và

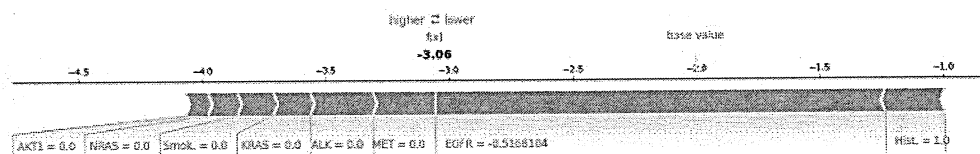
e) hiệu chỉnh dữ liệu so sánh bằng cách khởi tạo lại dữ liệu đánh giá theo một khoảng thời gian nhất định và so sánh với dữ liệu được nạp trước đó theo mô hình máy học đồng thời tiến hành phân tích máy học để hiệu chỉnh các yếu tố di truyền gây phát sinh ung thư phổi không tế bào nhỏ theo mô hình XGBoost theo thư viện SHAP (SHapley Additive exPlanations) và đột biến tương tác gen, trong đó yếu tố gây phát sinh ung thư phổi không tế bào nhỏ được xác định theo thứ tự giảm dần theo gồm đột biến phát sinh trên gen EGFR, ALK, các chỉ số lâm sàng, đột biến phát sinh trên gen KRAS, ROS, FGFR1, AKT1, HER2, MEK, RET, chỉ số cận lâm sàng, BRAF, DDR2, MTRK1, MEK1, MET, NRAS, PTEN, PIK3CA, đồng thời đánh dấu bổ sung vị trí đột biến điểm phát sinh vào cơ sở dữ liệu chuẩn.

2. Kit dùng trong phương pháp đánh giá yếu tố di truyền gây phát sinh ung thư phổi tế bào không nhỏ nhằm giải trình tự xác định đồng thời 43 gen bao gồm EGFR, KRAS, ALK, STK11, FGFR1, PIK3CA, ARAF, NRG1, RRM1, HAVCR2, MET, CDK4, BRCA1, CCND1, PDGFRA, AXL, AURKA, TYMS, ETS2, DDR2, ERCC2, ABCB1, TP53, NFE2L2, PDCD4, RET, BRAF, ROS1, NTRK1, XRCC1, CD274, RICTOR, AREG, BCL2L11, BIRC5, ERCC1, GNAS, AKT2, ERBB2, POLD1, RB1, SMARCA4 và WEE1, trong đó kit này bao gồm các trình tự nêu trong SEQ ID NO.1 đến SEQ ID NO.942.

HÌNH 1



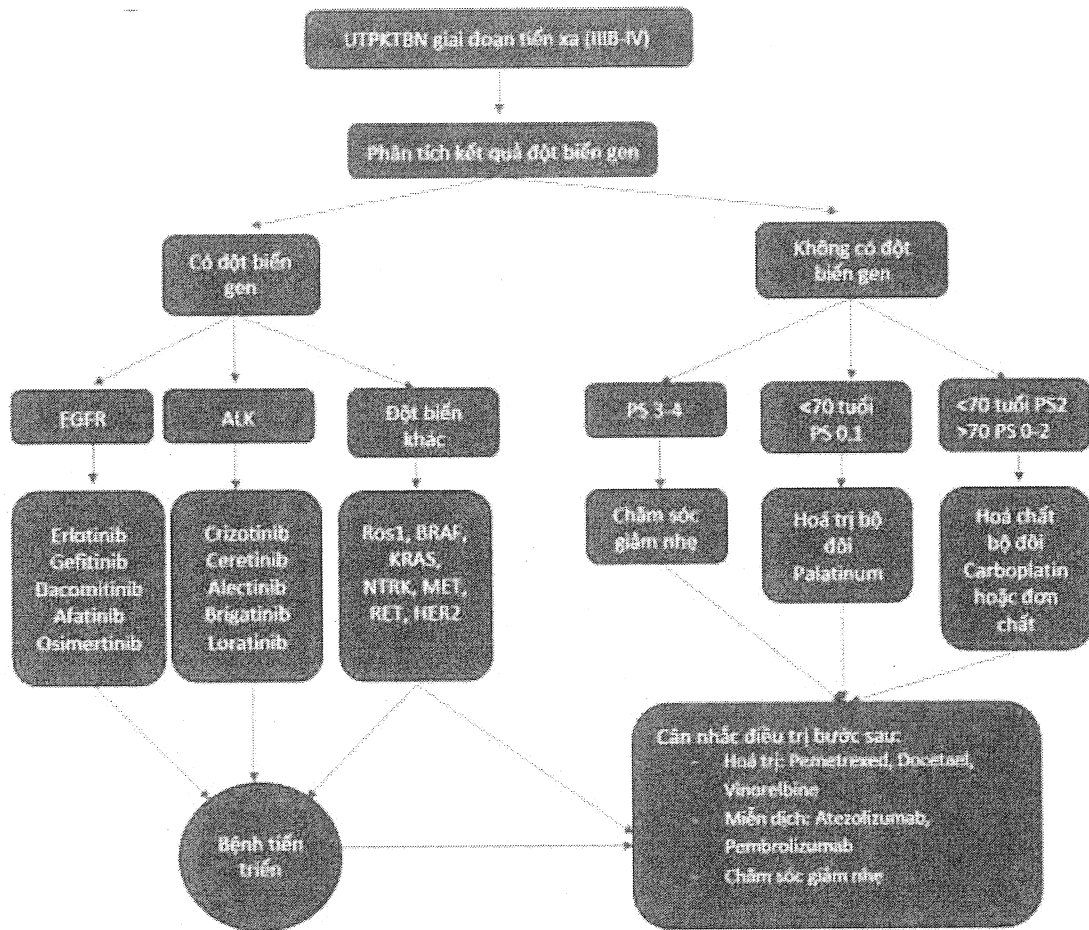
HÌNH 2



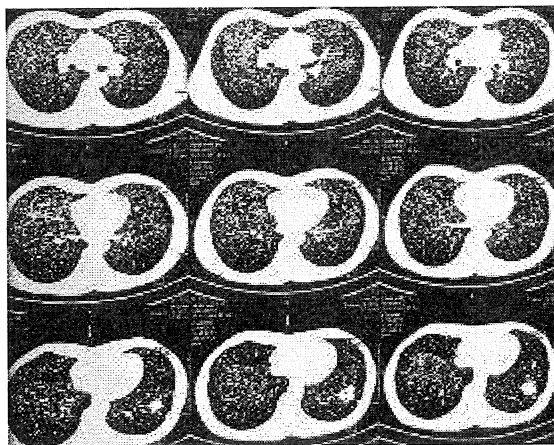
HÌNH 3

SHAP	KRAS	ALK	PIK3CA	NRAS	Age	HisL	Pathology	HER2	ROS	EGFR	DDR2	PTEN	RET	ROS	AGE	ROS	BRAF	AKT1	MET	Smoking Du
1	0	0	0	0	0	0	1	0	1	0	0	0	0	1	0	1	0	0	0	0.403089
0	0	0	1	1	1	0	0	0	1	0	0	0	1	1	0	0	0	0	0	0.424545
0	0	0	1	1	0	0	0	0	1	0	0	0	1	0	0	0	1	0	0	-1.521858
0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	-0.362833
1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2.576499
1	1	0	0	1	1	0	0	1	1	0	0	0	1	0	1	0	1	0	0	-0.814371
1	1	1	0	1	1	1	0	0	0	1	1	1	1	1	1	0	0	1	0	2.373666
0	0	0	0	0	0	1	0	1	0	1	0	0	0	0	0	0	0	0	0	1.424618
0	0	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.954277
1	0	1	0	1	0	0	0	0	1	1	0	1	1	1	0	0	0	1	0	-0.884320
0	0	1	1	0	1	0	0	1	1	0	0	1	0	0	0	0	0	0	1	0.173568
0	0	1	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0.651645
0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	1	0	0	-1.518255
0	0	0	0	0	0	0	1	0	1	0	1	0	0	0	0	0	0	0	0	-0.255565
0	1	0	0	1	0	0	1	0	1	0	1	0	0	1	1	0	1	0	0	-0.142945
0	0	0	0	0	0	0	1	0	0	1	0	1	0	0	0	0	0	0	0	-0.012236
0	1	0	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0.277639
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.885176
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	-0.582355
0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.144319
1	0	1	0	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0.923636
1	0	1	0	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0.844316
0	1	0	0	1	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	-0.523382
1	0	0	0	0	0	1	0	0	0	0	0	1	0	0	0	0	0	0	0	1.352527
0	1	0	0	0	0	1	0	0	0	0	0	1	0	0	0	0	0	0	0	1.475371
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1.007229
1	0	0	0	0	1	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0.823399
1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.204294
1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.778342
1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	-0.383398
0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	-0.284284
0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.881848
0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.878827

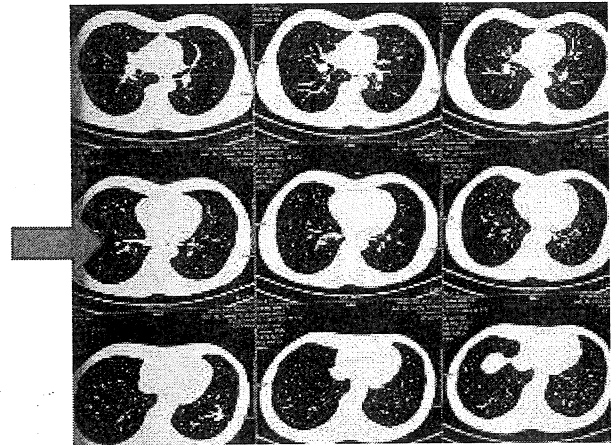
HÌNH 4



HÌNH 7

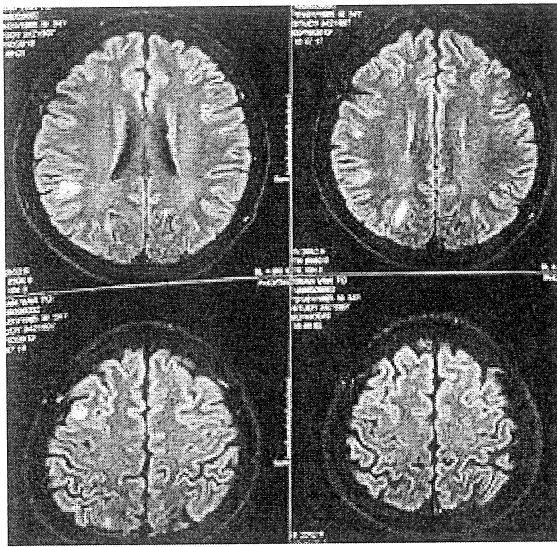


Trước điều trị

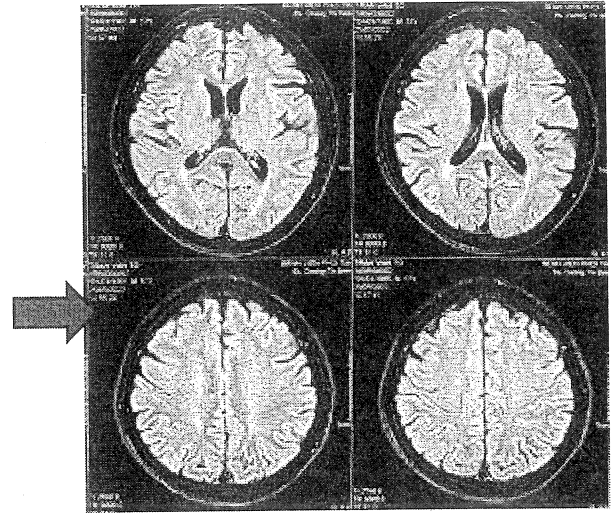


Sau điều trị 3 tháng

HÌNH 8

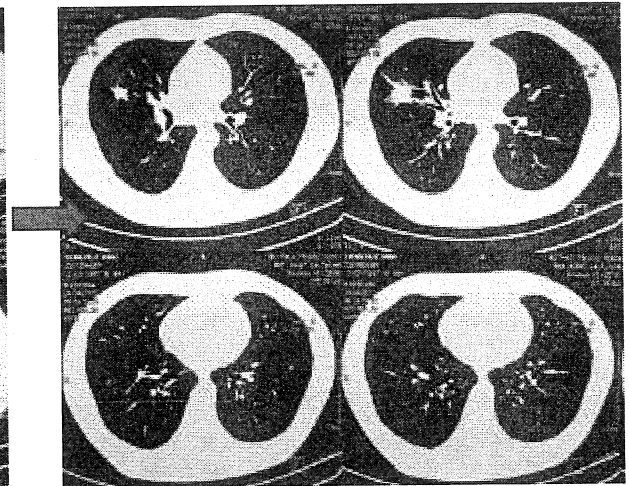
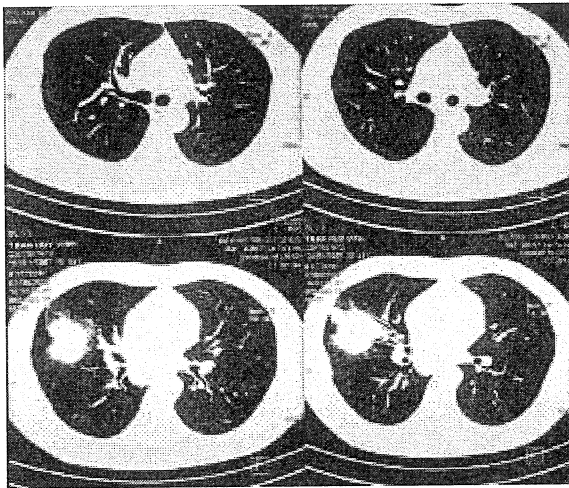


Trước điều trị



Sau điều trị 3 tháng

HÌNH 9



DANH MỤC TRÌNH TỰ

<110> Trường Đại học Y Hà Nội	
<120> Phương pháp đánh giá yếu tố đột biến di truyền đến đáp ứng thuốc đích điều trị ung thư phổi không tế bào nhỏ và kit dùng cho phương pháp này	
<130>	
<160> 942	
<170>	
<210> 1	
<211> 20	
<212> ADN	
<213> Đoạn mỗi	
<220> Nhân tạo	
<223>	
<220>	
<400> 1	
CCCCTGACTC CGTCCAGTAT	20
<210> 2	
<211> 20	
<212> ADN	
<213> Đoạn mỗi	
<220> Nhân tạo	
<223>	
<220>	
<400> 2	
CTTTTCCTCC AGAGCCCGAC	20
<210> 3	
<211> 20	
<212> ADN	
<213> Đoạn mỗi	
<220> Nhân tạo	
<223>	
<220>	
<400> 3	
ACGAGTAACA AGCTCACGCA	20
<210> 4	
<211> 20	
<212> ADN	
<213> Đoạn mỗi	
<220> Nhân tạo	
<223>	
<220>	
<400> 4	
CCAAGGACCA CCTCACAGTT	20
<210> 5	
<211> 20	
<212> ADN	
<213> Đoạn mỗi	
<220> Nhân tạo	
<223>	
<220>	

<400> 5
CCAGGAGGTG GCTGGTTATG

20

<210> 6
<211> 20
<212> ADN
<213> Đoạn mỗi
<220> Nhân tạo
<223>

<220>
<400> 6
CAGCTCCTTC AGTCCGGTTT

20

<210> 7
<211> 20
<212> ADN
<213> Đoạn mỗi
<220> Nhân tạo
<223>

<220>
<400> 7
AACGTGGAGA GCATCCAGTG

20

<210> 8
<211> 20
<212> ADN
<213> Đoạn mỗi
<220> Nhân tạo
<223>

<220>
<400> 8
CCCAGGTGGT TCTGGAAGTC

20

<210> 9
<211> 20
<212> ADN
<213> Đoạn mỗi
<220> Nhân tạo
<223>

<220>
<400> 9
GCCAAAAGTG TGATCCAAGC

20

<210> 10
<211> 20
<212> ADN
<213> Đoạn mỗi
<220> Nhân tạo
<223>

<220>
<400> 10
CAGTTCTCCT CTCCTGCACC

20

<210> 11
<211> 20
<212> ADN
<213> Đoạn mỗi
<220> Nhân tạo

<223>

<220>

<400> 11

AATCATCTGT GCCCAGCAGT

20

<210> 12

<211> 20

<212> ADN

<213> Đoạn mỗi

<220> Nhân tạo

<223>

<220>

<400> 12

CCTGCAGCAC ACTGGTTGT

19

<210> 13

<211> 20

<212> ADN

<213> Đoạn mỗi

<220> Nhân tạo

<223>

<220>

<400> 13

GCAAATTCCG AGACGAAGCC

20

<210> 14

<211> 20

<212> ADN

<213> Đoạn mỗi

<220> Nhân tạo

<223>

<220>

<400> 14

CTGTATTTGC CCTCGGGGTT

20

<210> 15

<211> 21

<212> ADN

<213> Đoạn mỗi

<220> Nhân tạo

<223>

<220>

<400> 15

TATGTGGTGA CAGATCACGG C

21

<210> 16

<211> 20

<212> ADN

<213> Đoạn mỗi

<220> Nhân tạo

<223>

<220>

<400> 16

AAGGCCCTTC GCACTTCTTA

20

<210> 17

<211> 24

<212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 17
 TGTGTAACGG AATAGGTATT GGTG

24

<210> 18
 <211> 20
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 18
 GGAGATCGCC ACTGATGGAG

20

<210> 19
 <211> 23
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 19
 TCCTTCACAC ATACTCCTCC TCT

23

<210> 20
 <211> 24
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 20
 CTGTGATTTC CTTTACGGTT TTCA

24

<210> 21
 <211> 20
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 21
 TTTGCTGATT CAGGCTTGGC

20

<210> 22
 <211> 20
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 22
 TCCTGCCGCG TATGATTCT

20

<210> 23	
<211> 20	
<212> ADN	
<213> Đoạn mỗi	
<220> Nhân tạo	
<223>	
<220>	
<400> 23	
TTCTCTTGCA GTCGTCAGCC	20
<210> 24	
<211> 20	
<212> ADN	
<213> Đoạn mỗi	
<220> Nhân tạo	
<223>	
<220>	
<400> 24	
TTCTGACCGG AGGTCCCAA	20
<210> 25	
<211> 19	
<212> ADN	
<213> Đoạn mỗi	
<220> Nhân tạo	
<223>	
<220>	
<400> 25	
AGGTCTGCCA TGCCTTG	19
<210> 26	
<211> 20	
<212> ADN	
<213> Đoạn mỗi	
<220> Nhân tạo	
<223>	
<220>	
<400> 26	
GTTGCACTTG TCCACGCATT	20
<210> 27	
<211> 22	
<212> ADN	
<213> Đoạn mỗi	
<220> Nhân tạo	
<223>	
<220>	
<400> 27	
CCAAGGGAGT TTGTGGAGAA CT	22
<210> 28	
<211> 20	
<212> ADN	
<213> Đoạn mỗi	
<220> Nhân tạo	
<223>	

<220>
 <400> 28
 GTCCTGTGCA GGTGATGTTC

20

<210> 29
 <211> 20
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 29
 GTGTGCCCAC TACATTGACG

20

<210> 30
 <211> 20
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 30
 CTGCGTACTT CCAGACCAGG

20

<210> 31
 <211> 20
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 31
 CCTAAGATCC CGTCCATCGC

20

<210> 32
 <211> 20
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 32
 CTTGCGCATGA AGAGGCCGAT

20

<210> 33
 <211> 20
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 33
 GCTCCCAACC AAGCTCTCTT

20

<210> 34
 <211> 20
 <212> ADN
 <213> Đoạn mỗi

<220> Nhân tạo
<223>

<220>
<400> 34
TTATACACCG TGCCGAACGC

20

<210> 35
<211> 21
<212> ADN
<213> Đoạn mỗi
<220> Nhân tạo
<223>

<220>
<400> 35
ACTCTGGATC CCAGAAGGTG A

21

<210> 36
<211> 20
<212> ADN
<213> Đoạn mỗi
<220> Nhân tạo
<223>

<220>
<400> 36
TGGCTTTCGG AGATGTTGCT

20

<210> 37
<211> 20
<212> ADN
<213> Đoạn mỗi
<220> Nhân tạo
<223>

<220>
<400> 37
CCTGGACTAT GTCCGGGAAC

20

<210> 38
<211> 20
<212> ADN
<213> Đoạn mỗi
<220> Nhân tạo
<223>

<220>
<400> 38
CTTTGCGATC TGCACACACC

20

<210> 39
<211> 20
<212> ADN
<213> Đoạn mỗi
<220> Nhân tạo
<223>

<220>
<400> 39
CGTACTGGTG AAAACACCGC

20

<210> 40

<211> 20
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 40
 TTCTTTCTCT TCCGCACCCA

20

<210> 41
 <211> 24
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 41
 ATCAAGTGGA TGGCATTGGA ATCA

24

<210> 42
 <211> 22
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 42
 CGTAGCTCCA GACATCACTC TG

22

<210> 43
 <211> 19
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 43
 GGGTGACCGT TTGGGAGTT

19

<210> 44
 <211> 20
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 44
 GAGGGAGGCG TTCTCCTTTC

20

<210> 46
 <211> 20
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 45

GCCCAAAGTT CCGTGAGTTG

20

<210> 46

<211> 20

<212> ADN

<213> Đoạn mỗi

<220> Nhân tạo

<223>

<220>

<400> 46

AATGACAAGG TAGCGCTGGG

20

<210> 47

<211> 20

<212> ADN

<213> Đoạn mỗi

<220> Nhân tạo

<223>

<220>

<400> 47

CGAGTACCTC ATCCACAGC

20

<210> 48

<211> 20

<212> ADN

<213> Đoạn mỗi

<220> Nhân tạo

<223>

<220>

<400> 48

GAGAGCTCAG GAGGGGAGTC

20

<210> 49

<211> 20

<212> ADN

<213> Đoạn mỗi

<220> Nhân tạo

<223>

<220>

<400> 49

TGCAAAGCTG TCCCATCAAG

20

<210> 50

<211> 20

<212> ADN

<213> Đoạn mỗi

<220> Nhân tạo

<223>

<220>

<400> 50

ACTGGGAGGA AGGTGTCGTC

20

<210> 51

<211> 20

<212> ADN

<213> Đoạn mỗi

<220> Nhân tạo

<223>

<220>
 <400> 51
 AGTTCATCCA GGCCCAACTG

20

<210> 52
 <211> 20
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 52
 CAAAGCGTAA GGAGGGGGTT

20

<210> 53
 <211> 21
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 53
 GCCTGCTGAA AATGACTGAA T

21

<210> 54
 <211> 20
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 54
 AGCTGTATCG TCAAGGCACT

20

<210> 55
 <211> 20
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 55
 TGGAGAAACCTGTCTCTTGGA

20

<210> 56
 <211> 21
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 56
 TGTACTGGTC CCTCATTGCA C

21

<210> 57
 <211> 20
 <212> ADN

<213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 57
 TGGTGGTGTG CCAAGACATT

20

<210> 58
 <211> 20
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 58
 AGTCATGGGG CATGTGGAAG

20

<210> 59
 <211> 20
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 59
 TGGTGAGAGA GATCCGACAA

20

<210> 60
 <211> 20
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 60
 TTCACACAGC
 CAGGAGTCTT

20

<210> 61
 <211> 20
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 61
 TGGTGGTGTG CCAAGACATT

20

<210> 62
 <211> 20
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 62

AGTCATGGGG CATGTGGAAG

20

<210> 63

<211> 20

<212> ADN

<213> Đoạn mỗi

<220> Nhân tạo

<223>

<220>

<400> 63

GGGCTGCTCC AGTTCAATCT

20

<210> 64

<211> 20

<212> ADN

<213> Đoạn mỗi

<220> Nhân tạo

<223>

<220>

<400> 64

CATCAGGCGG ATCCTCAGTC

20

<210> 65

<211> 21

<212> ADN

<213> Đoạn mỗi

<220> Nhân tạo

<223>

<220>

<400> 65

AGCTCCTTGG AATCACCAAC A

21

<210> 66

<211> 20

<212> ADN

<213> Đoạn mỗi

<220> Nhân tạo

<223>

<220>

<400> 66

TGCGATGAGA CAGGAAAGGG

20

<210> 67

<211> 20

<212> ADN

<213> Đoạn mỗi

<220> Nhân tạo

<223>

<220>

<400> 67

TTCCCCTGTG AGCTGGAGTA

20

<210> 68

<211> 20

<212> ADN

<213> Đoạn mỗi

<220> Nhân tạo

<223>

<220> <400> 68 TGGGCATCTC CTTAGAACGC	20
<210> 69 <211> 20 <212> ADN <213> Đoạn mỗi <220> Nhân tạo <223>	
<220> <400> 69 CTCAGCTGAC TCCAAGCACA	20
<210> 70 <211> 20 <212> ADN <213> Đoạn mỗi <220> Nhân tạo <223>	
<220> <400> 70 AGTGGGCATC AGGAGGATCT	20
<210> 71 <211> 20 <212> ADN <213> Đoạn mỗi <220> Nhân tạo <223>	
<220> <400> 71 GTGCTCCAGG GAAGAATCGG	20
<210> 72 <211> 20 <212> ADN <213> Đoạn mỗi <220> Nhân tạo <223>	
<220> <400> 72 ACTGCAGTTC TTCAGGGCAA	20
<210> 73 <211> 20 <212> ADN <213> Đoạn mỗi <220> Nhân tạo <223>	
<220> <400> 73 GAATGGGACA GTCCTCCAGC	20
<210> 74	

<211> 20
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 74
 TCTCATCTTC TCCCTGGGCA

20

<210> 75
 <211> 20
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 75
 AAGATGGCTT CTGTGGCTGG

20

<210> 76
 <211> 20
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 76
 GGGCATCCTT TAGGGTCCTG

20

<210> 77
 <211> 20
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 77
 CAGTACCACT GATGTCCCCG

20

<210> 78
 <211> 20
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 78
 AGAGCTCTTG ATCGGTGCAG

20

<210> 79
 <211> 20
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 79

ACGTGTCCTT GGTGCTAGTG

20

<210> 80

<211> 20

<212> ADN

<213> Đoạn mỗi

<220> Nhân tạo

<223>

<220>

<400> 80

AGAGGCAACA CCATCCACTG

20

<210> 81

<211> 20

<212> ADN

<213> Đoạn mỗi

<220> Nhân tạo

<223>

<220>

<400> 81

TTCTGGCTGC AGATGGTCG

19

<210> 82

<211> 20

<212> ADN

<213> Đoạn mỗi

<220> Nhân tạo

<223>

<220>

<400> 82

GTAGCAGTCC AGGCTGATGG

20

<210> 83

<211> 20

<212> ADN

<213> Đoạn mỗi

<220> Nhân tạo

<223>

<220>

<400> 83

GAAGGCTCTT ACGGCAAGGT

20

<210> 84

<211> 20

<212> ADN

<213> Đoạn mỗi

<220> Nhân tạo

<223>

<220>

<400> 84

TGGGGATCCT TCGCAACTTC

20

<210> 84

<211> 21

<212> ADN

<213> Đoạn mỗi

<220> Nhân tạo

<223>

<220>		
<400> 85		
AACTACTGAG GAGGTTACGG C		21
<210> 86		
<211> 26		
<212> ADN		
<213> Đoạn mỗi		
<220> Nhân tạo		
<223>		
<220>		
<400> 86		
TGCTTCTCT	TCGTTGTATA	
ACACAT		26
<210> 87		
<211> 20		
<212> ADN		
<213> Đoạn mỗi		
<220> Nhân tạo		
<223>		
<220>		
<400> 87		
ATGGAGTACT GCGTGTGTGG		20
<210> 88		
<211> 18		
<212> ADN		
<213> Đoạn mỗi		
<220> Nhân tạo		
<223>		
<220>		
<400> 88		
GCCTGGCACA CTGGGAAA		18
<210> 89		
<211> 20		
<212> ADN		
<213> Đoạn mỗi		
<220> Nhân tạo		
<223>		
<220>		
<400> 89		
GATTGACGGC CTGGAGTACC		20
<210> 90		
<211> 20		
<212> ADN		
<213> Đoạn mỗi		
<220> Nhân tạo		
<223>		
<220>		
<400> 90		
TCGGAGATTT TGAGGGTGCC		20
<210> 91		

<211> 17
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 91
 GACGACACCT GCCGGAC

17

<210> 92
 <211> 20
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 92
 AGATGTCCAC CTTGAAGCCG

20

<210> 93
 <211> 20
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 93
 CATCACCACG GGTCTGTACC

20

<210> 94
 <211> 20
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 94
 TTTCAGCAGG TCAGAGAGCG

20

<210> 95
 <211> 20
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 95
 CATGACTGTG GTGCCGTACT

20

<210> 96
 <211> 20
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 96

CCTCGATGTC GAAGAGGTCC

20

<210> 97

<211> 20

<212> ADN

<213> Đoạn mỗi

<220> Nhân tạo

<223>

<220>

<400> 97

AGAAGAGGAG GCCAGTCACA

20

<210> 98

<211> 20

<212> ADN

<213> Đoạn mỗi

<220> Nhân tạo

<223>

<220>

<400> 98

GCCTCTGTGC CGTTCATACA

20

<210> 99

<211> 20

<212> ADN

<213> Đoạn mỗi

<220> Nhân tạo

<223>

<220>

<400> 99

CTCAGTCTTC CTGCCGGTTC

20

<210> 100

<211> 20

<212> ADN

<213> Đoạn mỗi

<220> Nhân tạo

<223>

<220>

<400> 100

TTGCTCCGGG TTTTCTCTGG

20

<210> 101

<211> 21

<212> ADN

<213> Đoạn mỗi

<220> Nhân tạo

<223>

<220>

<400> 101

CCAAAGGAGA CCAGGCTGTA G

21

<210> 102

<211> 20

<212> ADN

<213> Đoạn mỗi

<220> Nhân tạo

<223>

<220> <400> 102 GCTCGGACCG GAGAAAAGTC	20
<210> 103 <211> 20 <212> ADN <213> Đoạn mỗi <220> Nhân tạo <223>	
<220> <400> 103 AGGAGGATCG AGCTCACTGT	20
<210> 104 <211> 20 <212> ADN <213> Đoạn mỗi <220> Nhân tạo <223>	
<220> <400> 104 CTTCCAGCTC CACATCCCAG	20
<210> 105 <211> 20 <212> ADN <213> Đoạn mỗi <220> Nhân tạo <223>	
<220> <400> 105 AGTGGAGTCC TTCCTGGTCC	20
<210> 106 <211> 20 <212> ADN <213> Đoạn mỗi <220> Nhân tạo <223>	
<220> <400> 106 TTACGCAAGC ATAGAGGCCG	20
<210> 107 <211> 18 <212> ADN <213> Đoạn mỗi <220> Nhân tạo <223>	
<220> <400> 107 ATGCTCTCCC CTCCTCGG	18
<210> 108 <211> 21 <212> ADN	

<213> Đoạn mỗi
<220> Nhân tạo
<223>

<220>
<400> 108
GCATACGGTT TGGTTTGGTG T

21

<210> 109
<211> 20
<212> ADN
<213> Đoạn mỗi
<220> Nhân tạo
<223>

<220>
<400> 109
GGCTGCCAAG ACAGTGAAGT

20

<210> 110
<211> 20
<212> ADN
<213> Đoạn mỗi
<220> Nhân tạo
<223>

<220>
<400> 110
GCCATTTTTC AACCAGCGCA

20

<210> 111
<211> 20
<212> ADN
<213> Đoạn mỗi
<220> Nhân tạo
<223>

<220>
<400> 111
GTCCGTTATG CCACCTGGAG

20

<210> 112
<211> 20
<212> ADN
<213> Đoạn mỗi
<220> Nhân tạo
<223>

<220>
<400> 112
TGGTTGATGC TGCCGTACTC

20

<210> 113
<211> 20
<212> ADN
<213> Đoạn mỗi
<220> Nhân tạo
<223>

<220>
<400> 113
GGGTAGCAAC GTGGAGTTCA

20

<210> 114
 <211> 20
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 114
 ATAAGGCAGG TTGTCTGGGC

20

<210> 115
 <211> 20
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 115
 GACGCAGGGG AGTATACGTG

20

<210> 116
 <211> 20
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 116
 AGAACGGTCA ACCATGCAGA

20

<210> 117
 <211> 20
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 117
 CTCATCTCCT GCATGGTGGG

20

<210> 118
 <211> 20
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 118
 CCATCTGGCT GTGGAAGTCA

20

<210> 119
 <211> 20
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>

<400> 119
GTCTGCTGAC TCCAGTGCAT

20

<210> 120
<211> 19
<212> ADN
<213> Đoạn mỗi
<220> Nhân tạo
<223>

<220>
<400> 120
ACTGGAGGAG AGCCGTGAT

19

<210> 121
<211> 20
<212> ADN
<213> Đoạn mỗi
<220> Nhân tạo
<223>

<220>
<400> 121
ACAATGCCT CCACGACCAT

20

<210> 122
<211> 20
<212> ADN
<213> Đoạn mỗi
<220> Nhân tạo
<223>

<220>
<400> 12
TCACGGTTGC CTACTGGTTC

20

<210> 123
<211> 20
<212> ADN
<213> Đoạn mỗi
<220> Nhân tạo
<223>

<220>
<400> 123
CTATCGGCAT GCCAGTGTGT

20

<210> 124
<211> 21
<212> ADN
<213> Đoạn mỗi
<220> Nhân tạo
<223>

<220>
<400> 124
TGTGCTTTGG CAATTCTGGT G

21

<210> 125
<211> 24
<212> ADN
<213> Đoạn mỗi
<220> Nhân tạo

<223>

<220>

<400> 125

GGCAAATAAT AGTGGTGATC TGGG

24

<210> 126

<211> 22

<212> ADN

<213> Đoạn mỗi

<220> Nhân tạo

<223>

<220>

<400> 126

TGTTCTGGTA CACAGTCATG GT

22

<210> 127

<211> 20

<212> ADN

<213> Đoạn mỗi

<220> Nhân tạo

<223>

<220>

<400> 127

TAATGCTTGG GAGGATGCCC

20

<210> 128

<211> 20

<212> ADN

<213> Đoạn mỗi

<220> Nhân tạo

<223>

<220>

<400> 128

GGTGTAAGCTG TGGAAATGCG

20

<210> 129

<211> 23

<212> ADN

<213> Đoạn mỗi

<220> Nhân tạo

<223>

<220>

<400> 129

GTATCTACCA TGGAGGAGAA CCG

23

<210> 130

<211> 21

<212> ADN

<213> Đoạn mỗi

<220> Nhân tạo

<223>

<220>

<400> 130

CTGGGATTGG AACAAAGGTAC T

21

<210> 131

<211> 20

<212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 131
 TCCTGATCTT CCTCGTGCTG

20

<210> 132
 <211> 20
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 132
 TTTAGCACCC TTTCGGCCTT

20

<210> 133
 <211> 20
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 133
 TGTCCATTGG CATGGGGAAA

20

<210> 134
 <211> 21
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 134
 CCATGAGGTA CTGGCCAAAG A

21

<210> 135
 <211> 20
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 135
 GACTGGTTCA GCAGTGTGGT

20

<210> 136
 <211> 20
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 136
 CAGTCCTGCG TGGGAATAGC

20

<210> 137	
<211> 24	
<212> ADN	
<213> Đoạn mỗi	
<220> Nhân tạo	
<223>	
<220>	
<400> 137	
GGGAAAATGA CAAAGAACAG CTCA	24
<210> 138	
<211> 22	
<212> ADN	
<213> Đoạn mỗi	
<220> Nhân tạo	
<223>	
<220>	
<400> 138	
TCTCCTGCTC AGTGATTTC GA	22
<210> 139	
<211> 20	
<212> ADN	
<213> Đoạn mỗi	
<220> Nhân tạo	
<223>	
<220>	
<400> 139	
CCACCTCAGC CCATCTTGAC	20
<210> 140	
<211> 21	
<212> ADN	
<213> Đoạn mỗi	
<220> Nhân tạo	
<223>	
<220>	
<400> 140	
GGTATACTTT GACGGTGCCC A	21
<210> 141	
<211> 19	
<212> ADN	
<213> Đoạn mỗi	
<220> Nhân tạo	
<223>	
<220>	
<400> 141	
CTGTCCGGGA TGGCATGAG	19
<210> 142	
<211> 20	
<212> ADN	
<213> Đoạn mỗi	
<220> Nhân tạo	
<223>	

<220>
 <400> 142
 GGTAGACCAC ACAGCAGTCC

20

<210> 143
 <211> 20
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 143
 GAAAGACGGT CACTGCCTGG

20

<210> 144
 <211> 20
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 144
 GGTCAGCGGG ACATCTTCAA

20

<210> 145
 <211> 20
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 145
 CTTCTTCAGC CTGGCGTTCT

20

<210> 146
 <211> 20
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 146
 CACAGGTTTG GCAACGGAAG

20

<210> 147
 <211> 20
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 147
 GTCCAGGATT TGTCCGGAGG

20

<210> 148
 <211> 21
 <212> ADN
 <213> Đoạn mỗi

<220> Nhân tạo
<223>

<220>
<400> 148
TTAGCAACTC ATTCAGGGGG C

21

<210> 149
<211> 18
<212> ADN
<213> Đoạn mỗi
<220> Nhân tạo
<223>

<220>
<400> 149
CTGTGACCCG GAGCACTT

18

<210> 150
<211> 19
<212> ADN
<213> Đoạn mỗi
<220> Nhân tạo
<223>

<220>
<400> 150
GACGTTGGGA GTGGACGTG

19

<210> 151
<211> 20
<212> ADN
<213> Đoạn mỗi
<220> Nhân tạo
<223>

<220>
<400> 151
GAGGTAGTGA TGGAACCCCC

20

<210> 152
<211> 20
<212> ADN
<213> Đoạn mỗi
<220> Nhân tạo
<223>

<220>
<400> 152
GCTCTGCTGG TGA CTGGA

20

<210> 153
<211> 20
<212> ADN
<213> Đoạn mỗi
<220> Nhân tạo
<223>

<220>
<400> 153
CTCGTTTGGC ACCGTGTTTC

20

<210> 154

<211> 20
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 154
 TTCTTGAAAG CCTGGGCCTG

20

<210> 155
 <211> 20
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 155
 CGGGATTTCG CATCATCACA C

21

<210> 156
 <211> 20
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 156
 TGAGCTGGAC CATGTCGAAG

20

<210> 157
 <211> 20
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 157
 CTCCTACAT GAGGGGCTCA C

21

<210> 158
 <211> 20
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 158
 CGCTCCATCG AGTCTTCACT

20

<210> 159
 <211> 20
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 159

AGCTGAGGTG ATCCGTATGC

20

<210> 160

<211> 20

<212> ADN

<213> Đoạn mỗi

<220> Nhân tạo

<223>

<220>

<400> 160

AAGCTCGTAG AGCACAACCC

20

<210> 161

<211> 19

<212> ADN

<213> Đoạn mỗi

<220> Nhân tạo

<223>

<220>

<400> 161

TATGGTGGGC CGTGGCTAT

19

<210> 162

<211> 19

<212> ADN

<213> Đoạn mỗi

<220> Nhân tạo

<223>

<220>

<400> 162

GCTGGAACCTT GAGGCAGTC

19

<210> 163

<211> 20

<212> ADN

<213> Đoạn mỗi

<220> Nhân tạo

<223>

<220>

<400> 163

CACCAGGGAG CCAATCTCAG

20

<210> 164

<211> 20

<212> ADN

<213> Đoạn mỗi

<220> Nhân tạo

<223>

<220>

<400> 164

GGGTCCCCCA ATTCCAGAAG

20

<210> 165

<211> 20

<212> ADN

<213> Đoạn mỗi

<220> Nhân tạo

<223>

<220>		
<400> 165		
CGTGAGCAGG ACGGTGATAA		20
<210> 166		
<211> 20		
<212> ADN		
<213> Đoạn mỗi		
<220> Nhân tạo		
<223>		
<220>		
<400> 166		
CTCGGATCGG GTTTGGGAAA		20
<210> 167		
<211> 20		
<212> ADN		
<213> Đoạn mỗi		
<220> Nhân tạo		
<223>		
<220>		
<400> 167		
GCCTCCCCGA TTGAAAGAGA		20
<210> 168		
<211> 20		
<212> ADN		
<213> Đoạn mỗi		
<220> Nhân tạo		
<223>		
<220>		
<400> 168		
ACTGGTTTCA CACCGAAGGA		20
<210> 169		
<211> 21		
<212> ADN		
<213> Đoạn mỗi		
<220> Nhân tạo		
<223>		
<220>		
<400> 169		
ACAAAGCATC ACTGGCTGAT T		21
<210> 170		
<211> 21		
<212> ADN		
<213> Đoạn mỗi		
<220> Nhân tạo		
<223>		
<220>		
<400> 170		
TCCACGATGG TGATATTGGC A		21
<210> 171		
<211> 20		
<212> ADN		

<213> Đoạn mỗi
<220> Nhân tạo
<223>

<220>
<400> 171
CCACTGGGAC AAGCCATCTT

20

<210> 172
<211> 20
<212> ADN
<213> Đoạn mỗi
<220> Nhân tạo
<223>

<220>
<400> 172
TTCACCATGA AGCACTCCCC

20

<210> 173
<211> 20
<212> ADN
<213> Đoạn mỗi
<220> Nhân tạo
<223>

<220>
<400> 173
GAGTGCTGAC CATAACCGGC

20

<210> 174
<211> 20
<212> ADN
<213> Đoạn mỗi
<220> Nhân tạo
<223>

<220>
<400> 174
CAGTAGGCCA CCACACACAT

20

<210> 175
<211> 20
<212> ADN
<213> Đoạn mỗi
<220> Nhân tạo
<223>

<220>
<400> 175
GCGGAAAAAG CTGCATGACC

20

<210> 176
<211> 20
<212> ADN
<213> Đoạn mỗi
<220> Nhân tạo
<223>

<220>
<400> 176
TAGGATGGTG AGGCCCATTG

20

<210> 177

<211> 22

<212> ADN

<213> Đoạn mỗi

<220> Nhân tạo

<223>

<220>

<400> 177

AAACGTCATC TCCAGTGAGC AT

22

<210> 178

<211> 20

<212> ADN

<213> Đoạn mỗi

<220> Nhân tạo

<223>

<220>

<400> 178

GTGGCTAGGA GTCTGGGTGA

20

<210> 179

<211> 20

<212> ADN

<213> Đoạn mỗi

<220> Nhân tạo

<223>

<220>

<400> 179

GGAGCAACGG ACACACTGAA

20

<210> 180

<211> 20

<212> ADN

<213> Đoạn mỗi

<220> Nhân tạo

<223>

<220>

<400> 180

TCTCGGTAGG AATCAGGGGT

20

<210> 181

<211> 20

<212> ADN

<213> Đoạn mỗi

<220> Nhân tạo

<223>

<220>

<400> 181

GATGCCTGTC TGCAGGTTCT

20

<210> 182

<211> 20

<212> ADN

<213> Đoạn mỗi

<220> Nhân tạo

<223>

<220>

<400> 182
GACCCGGGAA ACAAAGTTGC

20

<210> 183
<211> 20
<212> ADN
<213> Đoạn mỗi
<220> Nhân tạo
<223>

<220>
<400> 183
AAAGTGCTGT CTGGCTCCAA

20

<210> 184
<211> 20
<212> ADN
<213> Đoạn mỗi
<220> Nhân tạo
<223>

<220>
<400> 184
CGGGATCGAG AAGGTGACAG

20

<210> 185
<211> 20
<212> ADN
<213> Đoạn mỗi
<220> Nhân tạo
<223>

<220>
<400> 185
CGCCAAGAAC GAGTCATGTT

20

<210> 186
<211> 24
<212> ADN
<213> Đoạn mỗi
<220> Nhân tạo
<223>

<220>
<400> 186
TCCATATTGA GTCCATAACA AAGC

24

<210> 187
<211> 20
<212> ADN
<213> Đoạn mỗi
<220> Nhân tạo
<223>

<220>
<400> 187
TCCAAGGCTT GTACAGTGGG

20

<210> 188
<211> 20
<212> ADN
<213> Đoạn mỗi
<220> Nhân tạo

<223>

<220>

<400> 188

GGCTGCCAGG ATAGCATAGT

20

<210> 189

<211> 25

<212> ADN

<213> Đoạn mỗi

<220> Nhân tạo

<223>

<220>

<400> 189

GTGATGGAAG ACCTCTATAA CTACA

25

<210> 190

<211> 20

<212> ADN

<213> Đoạn mỗi

<220> Nhân tạo

<223>

<220>

<400> 190

CCAATGTTGA CTTGGCCACC

20

<210> 191

<211> 20

<212> ADN

<213> Đoạn mỗi

<220> Nhân tạo

<223>

<220>

<400> 191

TGGCTGAAAG ACCACAACAT

20

<210> 192

<211> 20

<212> ADN

<213> Đoạn mỗi

<220> Nhân tạo

<223>

<220>

<400> 192

AAAGTTGTGG GCGGTTGGTA

20

<210> 193

<211> 20

<212> ADN

<213> Đoạn mỗi

<220> Nhân tạo

<223>

<220>

<400> 193

AGATGACAGC ATTGAAGGCA

20

<210> 194

<211> 20

<212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

 <220>
 <400> 194
 AATGTAGCTG CCAGTAGCCC 20

 <210> 195
 <211> 20
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

 <220>
 <400> 195
 TTCCAATGGC CTTGTACCGA 20

 <210> 196
 <211> 20
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

 <220>
 <400> 196
 TGTTCCCACC TTGATCCACA 20

 <210> 197
 <211> 19
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

 <220>
 <400> 197
 CGTCCTGGGG CATTTGCTA 19

 <210> 198
 <211> 20
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

 <220>
 <400> 198
 GAAAAGATCT CTGGCACGCT 20

 <210> 199
 <211> 21
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

 <220>
 <400> 199
 GACTGGTCTT TGATGTGTCC A 21

<210> 200
 <211> 23
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 200
 AGTTTCTCAA ATTCCTCTCC CCA

23

<210> 201
 <211> 20
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 201
 TCTCAGACGG AACAGGCAC

20

<210> 202
 <211> 20
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 202
 GCACAGGTTG CTGCATTGA

20

<210> 203
 <211> 21
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 203
 CTTCCCTGGC CCTGAATATG T

21

<210> 204
 <211> 21
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 204
 TCAAGTTTCG GACAACGACT T

21

<210> 205
 <211> 20
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 205
 CAAATAAACG CCATCGCCCC

20

<210> 206
 <211> 20
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 206
 CTGGGCTTCT GCACTCTCAA

20

<210> 207
 <211> 20
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 207
 TATTGCCCCG ATGCCTACAG

20

<210> 208
 <211> 21
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 208
 TCTCCTGACA AGACTCTGCG A

21

<210> 209
 <211> 20
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 209
 CAGCTGAGAG AGGTGCTTTC

20

<210> 210
 <211> 21
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 210
 CTTCCAGCCG TAGAAGTGCA T

21

<210> 211
 <211> 20
 <212> ADN
 <213> Đoạn mỗi

<220> Nhân tạo
<223>

<220>
<400> 211
ACAAGACCAG CGGCTAATCC

20

<210> 212
<211> 20
<212> ADN
<213> Đoạn mỗi
<220> Nhân tạo
<223>

<220>
<400> 212
CTTTTGCCTG CAACAGGGTC

20

<210> 213
<211> 20
<212> ADN
<213> Đoạn mỗi
<220> Nhân tạo
<223>

<220>
<400> 213
TGTGCCTAAC AGAGGTGTCC

20

<210> 214
<211> 22
<212> ADN
<213> Đoạn mỗi
<220> Nhân tạo
<223>

<220>
<400> 214
ACACAGTCAA AGGGAAGATGT GA

22

<210> 215
<211> 20
<212> ADN
<213> Đoạn mỗi
<220> Nhân tạo
<223>

<220>
<400> 215
GAATACAGAG CGGAGGTCGG

20

<210> 216
<211> 20
<212> ADN
<213> Đoạn mỗi
<220> Nhân tạo
<223>

<220>
<400> 216
AGGGACACAT CTCCTTTGCG

20

<210> 217

<211> 20
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

 <220>
 <400> 217
 GGGACTCTGG AGCAACCATC 20

 <210> 218
 <211> 20
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

 <220>
 <400> 218
 GATAAGAGCC AGAGCCAGCC 20

 <210> 219
 <211> 20
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

 <220>
 <400> 219
 CTGGGACCTG CACTGAACTT 20

 <210> 220
 <211> 20
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

 <220>
 <400> 220
 CTTTCACCTC AGCACCCAGT 20

 <210> 221
 <211> 20
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

 <220>
 <400> 221
 CTTTGTGAGC AGATGCGGAG 20

 <210> 222
 <211> 20
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

 <220>
 <400> 222

GCCTCCTCTC AGCAAGTCAG

20

<210> 223

<211> 20

<212> ADN

<213> Đoạn mỗi

<220> Nhân tạo

<223>

<220>

<400> 223

TGCAGCAGCA AAGCCAATT

20

<210> 224

<211> 20

<212> ADN

<213> Đoạn mỗi

<220> Nhân tạo

<223>

<220>

<400> 224

AAAGGACTTT GGCTCCCAGG

20

<210> 225

<211> 20

<212> ADN

<213> Đoạn mỗi

<220> Nhân tạo

<223>

<220>

<400> 225

CTTTGCAGCG CGTTGACTTA

20

<210> 226

<211> 20

<212> ADN

<213> Đoạn mỗi

<220> Nhân tạo

<223>

<220>

<400> 226

CGACCCTCTG ATGTCCAAG

20

<210> 227

<211> 20

<212> ADN

<213> Đoạn mỗi

<220> Nhân tạo

<223>

<220>

<400> 227

TTCTCCTGGA CTCCCATCCA

20

<210> 228

<211> 22

<212> ADN

<213> Đoạn mỗi

<220> Nhân tạo

<223>

<220>	
<400> 228	
AACCAGTGTG TAGCCATTTT GG	22
<210> 229	
<211> 20	
<212> ADN	
<213> Đoạn mỗi	
<220> Nhân tạo	
<223>	
<220>	
<400> 229	
CCCCACCCTT TGTTCAAGTGT	20
<210> 230	
<211> 20	
<212> ADN	
<213> Đoạn mỗi	
<220> Nhân tạo	
<223>	
<220>	
<400> 230	
TCTGTTGAGT CCATGTCCCG	20
<210> 231	
<211> 20	
<212> ADN	
<213> Đoạn mỗi	
<220> Nhân tạo	
<223>	
<220>	
<400> 231	
CCCCTTGAAG GAGGGACAAG	20
<210> 232	
<211> 21	
<212> ADN	
<213> Đoạn mỗi	
<220> Nhân tạo	
<223>	
<220>	
<400> 232	
AAGGTGCAGC TCTCATTTCC A	21
<210> 233	
<211> 20	
<212> ADN	
<213> Đoạn mỗi	
<220> Nhân tạo	
<223>	
<220>	
<400> 233	
AAATGCACAG TTGGTCCTGC	20
<210> 234	
<211> 20	
<212> ADN	

<213> Đoạn mỗi
<220> Nhân tạo
<223>

<220>
<400> 234
TTGTCCCGTG GCCATTTGAA

20

<210> 235
<211> 21
<212> ADN
<213> Đoạn mỗi
<220> Nhân tạo
<223>

<220>
<400> 235
TCGCCGAAAT ACGGTCCTAT G

21

<210> 236
<211> 20
<212> ADN
<213> Đoạn mỗi
<220> Nhân tạo
<223>

<220>
<400> 236
TGTGTCTAGA ATTCCCACTG TT

22

<210> 237
<211> 20
<212> ADN
<213> Đoạn mỗi
<220> Nhân tạo
<223>

<220>
<400> 237
TGTTATACCC CAGCCCAAAC C

21

<210> 238
<211> 21
<212> ADN
<213> Đoạn mỗi
<220> Nhân tạo
<223>

<220>
<400> 238
TGTCTCTCGG TTGGCTAAGT C

21

<210> 239
<211> 22
<212> ADN
<213> Đoạn mỗi
<220> Nhân tạo
<223>

<220>
<400> 239
ACTTGGTGGA AAGAACCTCT CA

22

<210> 240
 <211> 21
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 240
 ACCTGTTATT GTGCTCCCAC C

21

<210> 241
 <211> 20
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 241
 AGCTGAATCT GCAACTCCCC

20

<210> 242
 <211> 22
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 242
 TCCAGTACAT TTTCATTGCC CA

22

<210> 243
 <211> 22
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 243
 GACCCTGAAG CAGTTAAAGG TG

22

<210> 244
 <211> 20
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 244
 GTCATTGGGG ACCGTGCATA

20

<210> 245
 <211> 22
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>

<400> 245 TCACAGGATT GATTGCTGGT GT	22
<210> 246 <211> 21 <212> ADN <213> Đoạn mỗi <220> Nhân tạo <223>	
<220> <400> 246 CAGCCACAGG AAAAACCCTAA G	21
<210> 247 <211> 21 <212> ADN <213> Đoạn mỗi <220> Nhân tạo <223>	
<220> <400> 247 AGTTCGCTAC GATGCAAGAG T	21
<210> 248 <211> 20 <212> ADN <213> Đoạn mỗi <220> Nhân tạo <223>	
<220> <400> 248 AGTTGGGCTT ACACTTCGGG	20
<210> 249 <211> 20 <212> ADN <213> Đoạn mỗi <220> Nhân tạo <223>	
<220> <400> 249 GTTTCATGCCG ACAAGTGCAG	20
<210> 250 <211> 20 <212> ADN <213> Đoạn mỗi <220> Nhân tạo <223>	
<220> <400> 250 TGCACAATCA GGCTACTGGG	20
<210> 251 <211> 23 <212> ADN <213> Đoạn mỗi <220> Nhân tạo	

<223>

<220>

<400> 251

TGGTTGTGTA TATCATGGGA CTT

23

<210> 252

<211> 21

<212> ADN

<213> Đoạn mỗi

<220> Nhân tạo

<223>

<220>

<400> 252

TCAAGGATTT CACAGCACAG T

21

<210> 253

<211> 20

<212> ADN

<213> Đoạn mỗi

<220> Nhân tạo

<223>

<220>

<400> 253

TCATCCCAAT GTCCTCTCGC

20

<210> 254

<211> 20

<212> ADN

<213> Đoạn mỗi

<220> Nhân tạo

<223>

<220>

<400> 254

GTATGGTAGG ACCACCAGCG

20

<210> 255

<211> 22

<212> ADN

<213> Đoạn mỗi

<220> Nhân tạo

<223>

<220>

<400> 255

GGTCTTCAAG TAGCCAAAGG CA

22

<210> 256

<211> 21

<212> ADN

<213> Đoạn mỗi

<220> Nhân tạo

<223>

<220>

<400> 256

TTCTTGCAGC CAAGTCTCTG T

21

<210> 257

<211> 22

<212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 257
 GTTGCTGATT TTGGTCTTGC CA

22

<210> 258
 <211> 20
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 258
 CTTCACTGGC AGCTTTGCAC

20

<210> 259
 <211> 20
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 259
 AGAGGAGCCC CACCTTATCC

20

<210> 260
 <211> 20
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 260
 AAGGGGTCTG GGCAGTATTC

20

<210> 261
 <211> 20
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 261
 ATGCGCCCAT CCTTTTCTGA

20

<210> 262
 <211> 20
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 262
 AATGAGTGGC CTGTTCTGGG

20

<210> 263
 <211> 20
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 263
 CTGGTCACAT GGTGAGGGTG

20

<210> 264
 <211> 20
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 264
 GGTTCTACG GCCCCATACA

20

<210> 265
 <211> 20
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 265
 CAGTGTACAA GGCCCGTGAT

20

<210> 266
 <211> 20
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 266
 CAGTCGCCTC AGTAAAGCCA

20

<210> 267
 <211> 20
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 267
 CGAACTGACC GGGAGATCAA

20

<210> 268
 <211> 20
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220> <400> 268 CTTGATCGTT TCGGCTGGCA	20
<210> 269 <211> 20 <212> ADN <213> Đoạn mỗi <220> Nhân tạo <223>	
<220> <400> 269 TGCCAATTGC ATCGTTCACC	20
<210> 270 <211> 20 <212> ADN <213> Đoạn mỗi <220> Nhân tạo <223>	
<220> <400> 270 CCAGCTTGAC TGTTCACCA	20
<210> 271 <211> 21 <212> ADN <213> Đoạn mỗi <220> Nhân tạo <223>	
<220> <400> 271 GCAGTCCACA TATGCAACAC C	21
<210> 272 <211> 21 <212> ADN <213> Đoạn mỗi <220> Nhân tạo <223>	
<220> <400> 272 TTCGACGAAA CATCTCTGCA A	21
<210> 273 <211> 20 <212> ADN <213> Đoạn mỗi <220> Nhân tạo <223>	
<220> <400> 273 GGATGACTGG CCTCGAGATG	20
<210> 274 <211> 19 <212> ADN <213> Đoạn mỗi	

<220> Nhân tạo
<223>

<220>
<400> 274
CTCAGGTACC ACCGACTGC

19

<210> 275
<211> 20
<212> ADN
<213> Đoạn mỗi
<220> Nhân tạo
<223>

<220>
<400> 275
ACACAAGCGA ATCTCTGCCT

20

<210> 276
<211> 20
<212> ADN
<213> Đoạn mỗi
<220> Nhân tạo
<223>

<220>
<400> 276
AGCCACTCCA TTGCTCACTC

20

<210> 277
<211> 20
<212> ADN
<213> Đoạn mỗi
<220> Nhân tạo
<223>

<220>
<400> 277
GTAGCCCCTT GGTTCCTG

20

<210> 278
<211> 20
<212> ADN
<213> Đoạn mỗi
<220> Nhân tạo
<223>

<220>
<400> 278
CGCGCAGTCG CAGTTTAAAT

20

<210> 279
<211> 22
<212> ADN
<213> Đoạn mỗi
<220> Nhân tạo
<223>

<220>
<400> 279
TGGATTATC TGCTCTTCGC GT

22

<210> 280

<211> 23
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 280
 TGGGACACTC TAAGATTTTC TGC

23

<210> 281
 <211> 22
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 281
 GAGCCTACAA GAAAGTACGA GA

22

<210> 282
 <211> 21
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 282
 CTCCAAACCT GTGTCAAGCT G

21

<210> 283
 <211> 21
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 283
 GGAAACCAGT CTCAGTGTCC A

21

<210> 284
 <211> 20
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 284
 TGTATCCGCT GCTTTGTCCT

20

<210> 285
 <211> 20
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 285

CCGAAGAGGG GCCAAGAAAT

20

<210> 286

<211> 20

<212> ADN

<213> Đoạn mỗi

<220> Nhân tạo

<223>

<220>

<400> 286

TCCAAGCCCG TTCCTCTTTC

20

<210> 287

<211> 20

<212> ADN

<213> Đoạn mỗi

<220> Nhân tạo

<223>

<220>

<400> 287

GAAGCAGCAT CTGGGTGTGA

20

<210> 288

<211> 21

<212> ADN

<213> Đoạn mỗi

<220> Nhân tạo

<223>

<220>

<400> 288

GTCACTCTGA GAGGATAGCC C

21

<210> 289

<211> 20

<212> ADN

<213> Đoạn mỗi

<220> Nhân tạo

<223>

<220>

<400> 289

CTCCAGCAGG AAATGGCTGA

20

<210> 290

<211> 20

<212> ADN

<213> Đoạn mỗi

<220> Nhân tạo

<223>

<220>

<400> 290

ATTTTCGCAGG TCCTCAAGGG

20

<210> 291

<211> 20

<212> ADN

<213> Đoạn mỗi

<220> Nhân tạo

<223>

<220>
 <400> 291
 GGCCTTTCTG CTGACAAGTT

20

<210> 292
 <211> 24
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 292
 CTTTCCACTC CTGGTTCTTT ATTT

24

<210> 293
 <211> 20
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 293
 GGTGGTACAT GCACAGTTGC

20

<210> 294
 <211> 20
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 294
 AGCTGTTGCT CCTCCACATC

20

<210> 295
 <211> 20
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 295
 TCTGAAGACA GAGCCCCAGA

20

<210> 296
 <211> 20
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 296
 GCTTCTCCCT GCTCACACTT

20

<210> 297
 <211> 20
 <212> ADN

<213> Đoạn mỗi
<220> Nhân tạo
<223>

<220>
<400> 297
GCTCGTGTAC AAGTTTGCCA

20

<210> 298
<211> 25
<212> ADN
<213> Đoạn mỗi
<220> Nhân tạo
<223>

<220>
<400> 298
ACAACATGAG TAGTCTCTTC AGTAA

25

<210> 299
<211> 21
<212> ADN
<213> Đoạn mỗi
<220> Nhân tạo
<223>

<220>
<400> 299
GCTGAGTTTG TGTGTGAACG G

21

<210> 300
<211> 23
<212> ADN
<213> Đoạn mỗi
<220> Nhân tạo
<223>

<220>
<400> 300
AGCTAACTAC CCATTTTCCT CCC

23

<210> 301
<211> 22
<212> ADN
<213> Đoạn mỗi
<220> Nhân tạo
<223>

<220>
<400> 301
TTGAAGTCAG AGGAGATGTG GT

22

<210> 302
<211> 20
<212> ADN
<213> Đoạn mỗi
<220> Nhân tạo
<223>

<220>
<400> 302
TCCTGGGATT CTCTTGCTCG

20

<210> 303
 <211> 22
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 303
 CAACTGGAAT GGATGGTACA GC

22

<210> 304
 <211> 21
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 304
 TGTGCCAAGG GTGAATGATG A

21

<210> 305
 <211> 20
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 305
 CTGGGAGCTC CTCTCACTCT

20

<210> 306
 <211> 20
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 306
 CCTCCCCAAT GTTCCACTCC

20

<210> 307
 <211> 20
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 307
 ACACGGACTA CAGGGGAGTT

20

<210> 308
 <211> 20
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>

<400> 308
ACACGGACTA CAGGGGAGTT

20

<210> 309
<211> 20
<212> ADN
<213> Đoạn mỗi
<220> Nhân tạo
<223>

<220>
<400> 309
TGGAGCCCGT GAAAAAGAGC

20

<210> 310
<211> 19
<212> ADN
<213> Đoạn mỗi
<220> Nhân tạo
<223>

<220>
<400> 310
CGGATGGAGT TGTCGGTGT

19

<210> 311
<211> 20
<212> ADN
<213> Đoạn mỗi
<220> Nhân tạo
<223>

<220>
<400> 311
CAATGACCCC GCACGATTTC

20

<210> 312
<211> 20
<212> ADN
<213> Đoạn mỗi
<220> Nhân tạo
<223>

<220>
<400> 312
CAGAGGGCAA CGAAGGTCTG

20

<210> 313
<211> 20
<212> ADN
<213> Đoạn mỗi
<220> Nhân tạo
<223>

<220>
<400> 313
TGAGGAGCCC CAACAACTTC

20

<210> 314
<211> 21
<212> ADN
<213> Đoạn mỗi
<220> Nhân tạo

<223>

<220>

<400> 314

CGGGTCACAC TTGATCACTC T

21

<210> 315

<211> 20

<212> ADN

<213> Đoạn mỗi

<220> Nhân tạo

<223>

<220>

<400> 315

CTGATTGGAC AGGCATGGGT

20

<210> 316

<211> 20

<212> ADN

<213> Đoạn mỗi

<220> Nhân tạo

<223>

<220>

<400> 316

GTGCCTGGAA GTCAACGGTA

20

<210> 317

<211> 20

<212> ADN

<213> Đoạn mỗi

<220> Nhân tạo

<223>

<220>

<400> 317

AAGAGCAAAA AGCGAAGGCG

20

<210> 318

<211> 20

<212> ADN

<213> Đoạn mỗi

<220> Nhân tạo

<223>

<220>

<400> 318

TCCGCAATGA ATGTCCCACA

20

<210> 319

<211> 20

<212> ADN

<213> Đoạn mỗi

<220> Nhân tạo

<223>

<220>

<400> 319

GAGTGAAGTG AGCTGGCAGT

20

<210> 320

<211> 20

<212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 320
 GTCACAAAAA GGCCGCTGTT

20

<210> 321
 <211> 20
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 321
 GCACAACCTGA TCCCGAGACT

20

<210> 322
 <211> 20
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 322
 TAAAGCCCTG TCTGCTGTCTG

20

<210> 323
 <211> 20
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 323
 CGATTGTGGT CACCTGTGCT

20

<210> 324
 <211> 21
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 324
 CACTTCTCCA GGGTAAGTCC A

21

<210> 325
 <211> 20
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 325
 GCAAAGGCAT CACAATGCTG

20

<210> 326
 <211> 20
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 326
 TCGGGGACCG TCAAAGTGTA

20

<210> 327
 <211> 20
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 327
 TTCAGCCAGT TGGAAGCTGT

20

<210> 328
 <211> 21
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 328
 GGATATCCTG GGAGGTGGGT A

21

<210> 329
 <211> 20
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 329
 AGCTGATCCG TGCTAAGGAA G

21

<210> 330
 <211> 22
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 330
 ATAGCTCTTC ACAGCATCTT CA

22

<210> 331
 <211> 21
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220> <400> 331 TGATCACCAT GGCTCAACTG G	21
<210> 332 <211> 20 <212> ADN <213> Đoạn mỗi <220> Nhân tạo <223>	
<220> <400> 332 CAATATCAGG AAGCGGCGTG	20
<210> 333 <211> 20 <212> ADN <213> Đoạn mỗi <220> Nhân tạo <223>	
<220> <400> 333 ACTCCCGAGA CAGGAGTACC	20
<210> 334 <211> 20 <212> ADN <213> Đoạn mỗi <220> Nhân tạo <223>	
<220> <400> 334 CTCGGTTCTC AGCTCCAAGG	20
<210> 335 <211> 19 <212> ADN <213> Đoạn mỗi <220> Nhân tạo <223>	
<220> <400> 335 TCTGAACTCA CGGTGGCTG	19
<210> 336 <211> 23 <212> ADN <213> Đoạn mỗi <220> Nhân tạo <223>	
<220> <400> 336 GTTTCCAAAT GACAACCAGG ACA	23
<210> 337 <211> 20 <212> ADN <213> Đoạn mỗi	

<220> Nhân tạo
<223>

<220>
<400> 337
TGAAATTCGC TGGAGGGTCA

20

<210> 338
<211> 20
<212> ADN
<213> Đoạn mỗi
<220> Nhân tạo
<223>

<220>
<400> 338
GCTGCATCGG GTCCACATAA

20

<210> 339
<211> 20
<212> ADN
<213> Đoạn mỗi
<220> Nhân tạo
<223>

<220>
<400> 339
GTCTTGCGGT CTGGAGCGTT

20

<210> 340
<211> 20
<212> ADN
<213> Đoạn mỗi
<220> Nhân tạo
<223>

<220>
<400> 340
CAGGTTGGGA CCGGCTTAAT

20

<210> 341
<211> 20
<212> ADN
<213> Đoạn mỗi
<220> Nhân tạo
<223>

<220>
<400> 341
CACGGCCAGA TCCAGTGAAA

20

<210> 342
<211> 20
<212> ADN
<213> Đoạn mỗi
<220> Nhân tạo
<223>

<220>
<400> 342
CACGGCCAGA TCCAGTGAAA

20

<210> 343

<211> 20
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 343
 AGGGATAGCT TCCTGAGCCA

20

<210> 344
 <211> 20
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 344
 TGTGCTTTCA TCAGCAGGGT

20

<210> 345
 <211> 21
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 345
 CATGGACATG AAGCAGGCTG A

21

<210> 346
 <211> 21
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 346
 AGGCTGGACG ATCATAGAGT G

21

<210> 347
 <211> 22
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 347
 ACTCAGAAGT CAAAAACCTC CT

22

<210> 348
 <211> 20
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 348

CTCCATTCCT CGGGCAACTT

20

<210> 349

<211> 19

<212> ADN

<213> Đoạn mỗi

<220> Nhân tạo

<223>

<220>

<400> 349

GTGATCTGGC TGCTCGCAA

19

<210> 350

<211> 20

<212> ADN

<213> Đoạn mỗi

<220> Nhân tạo

<223>

<220>

<400> 350

GCCAGGCCAA AGTCACAGAT

20

<210> 351

<211> 20

<212> ADN

<213> Đoạn mỗi

<220> Nhân tạo

<223>

<220>

<400> 351

CTTTCTGCCC GTGAAGTGGA

20

<210> 352

<211> 20

<212> ADN

<213> Đoạn mỗi

<220> Nhân tạo

<223>

<220>

<400> 352

ATCTCCCAGA GCAGAATGCC

20

<210> 353

<211> 21

<212> ADN

<213> Đoạn mỗi

<220> Nhân tạo

<223>

<220>

<400> 353

CCGGCATGAT GGTGGATTCT A

21

<210> 354

<211> 20

<212> ADN

<213> Đoạn mỗi

<220> Nhân tạo

<223>

<220>		
<400> 354		
CTTCACTGGT AGCGTGGTCA		20
<210> 355		
<211> 20		
<212> ADN		
<213> Đoạn mỗi		
<220> Nhân tạo		
<223>		
<220>		
<400> 355		
AACAGTGAGC CGGAGAAGAG		20
<210> 356		
<211> 22		
<212> ADN		
<213> Đoạn mỗi		
<220> Nhân tạo		
<223>		
<220>		
<400> 356		
ATTGTCCAGG CAGCAGATTC TC		22
<210> 357		
<211> 19		
<212> ADN		
<213> Đoạn mỗi		
<220> Nhân tạo		
<223>		
<220>		
<400> 357		
GCATGCGTGT GGA CTCAGA		19
<210> 358		
<211> 20		
<212> ADN		
<213> Đoạn mỗi		
<220> Nhân tạo		
<223>		
<220>		
<400> 358		
CACTGTCAGC GCTCAGTCTC		20
<210> 359		
<211> 20		
<212> ADN		
<213> Đoạn mỗi		
<220> Nhân tạo		
<223>		
<220>		
<400> 359		
ACTTTGGAAC AGGGTTGGCA		20
<210> 360		
<211> 20		
<212> ADN		

<213> Đoạn mỗi
<220> Nhân tạo
<223>

<220>
<400> 360
GCCAATGAAC CCACACCAAC

20

<210> 361
<211> 19
<212> ADN
<213> Đoạn mỗi
<220> Nhân tạo
<223>

<220>
<400> 361
CAGTGCCAAA TCCGGGGAG

19

<210> 362
<211> 20
<212> ADN
<213> Đoạn mỗi
<220> Nhân tạo
<223>

<220>
<400> 362
TGGGTGCCAA ACTTTCCTCA

20

<210> 363
<211> 20
<212> ADN
<213> Đoạn mỗi
<220> Nhân tạo
<223>

<220>
<400> 363
GCAGGCTGAA GAAAGTCCCT

20

<210> 364
<211> 20
<212> ADN
<213> Đoạn mỗi
<220> Nhân tạo
<223>

<220>
<400> 364
GAGCTGGCTG ACCACTATCC

20

<210> 365
<211> 21
<212> ADN
<213> Đoạn mỗi
<220> Nhân tạo
<223>

<220>
<400> 365
AATCACCTCC CTGCAGCTTT C

21

<210> 366
 <211> 20
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 366
 ACACGAAGGT CTGATGTCCC

20

<210> 367
 <211> 21
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 367
 CAACACCCCC TTCAACCTGA G

21

<210> 368
 <211> 21
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 368
 CATCCTGGAG CCAGAGTAGG T

21

<210> 369
 <211> 22
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 369
 GGCTGAACAA GACATCCTCT TT

22

<210> 370
 <211> 19
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 370
 GTGGCTGTGC GGGATGT

19

<210> 371
 <211> 20
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>

<400> 371
TAACCTCCAC CTGGTCTCCC

20

<210> 372
<211> 20
<212> ADN
<213> Đoạn mỗi
<220> Nhân tạo
<223>

<220>
<400> 372
AGTGGGTCAG GGGGTAGATG

20

<210> 373
<211> 20
<212> ADN
<213> Đoạn mỗi
<220> Nhân tạo
<223>

<220>
<400> 373
CTGTCAGACG ATGGGATGGG

20

<210> 374
<211> 20
<212> ADN
<213> Đoạn mỗi
<220> Nhân tạo
<223>

<220>
<400> 374
ATGCCACGCG GATGTGATAA

20

<210> 375
<211> 20
<212> ADN
<213> Đoạn mỗi
<220> Nhân tạo
<223>

<220>
<400> 375
CCTTCGTGCA TTGGCAAGAG

20

<210> 376
<211> 20
<212> ADN
<213> Đoạn mỗi
<220> Nhân tạo
<223>

<220>
<400> 376
GTGTCCTGGC CTTGATACGC

20

<210> 377
<211> 20
<212> ADN
<213> Đoạn mỗi
<220> Nhân tạo

<223>

<220>

<400> 377

CGGGTCTGTG TCCAATCTGA

20

<210> 378

<211> 18

<212> ADN

<213> Đoạn mỗi

<220> Nhân tạo

<223>

<220>

<400> 378

CAGGCCTCCA GGGGTACT

18

<210> 379

<211> 20

<212> ADN

<213> Đoạn mỗi

<220> Nhân tạo

<223>

<220>

<400> 379

GGCCCTGGTG GTATGTACTG

20

<210> 380

<211> 20

<212> ADN

<213> Đoạn mỗi

<220> Nhân tạo

<223>

<220>

<400> 380

GGTGGACAAG GAAGAGAGCC

20

<210> 381

<211> 21

<212> ADN

<213> Đoạn mỗi

<220> Nhân tạo

<223>

<220>

<400> 381

AGTGGAAGA GGTGAACTGG T

21

<210> 382

<211> 20

<212> ADN

<213> Đoạn mỗi

<220> Nhân tạo

<223>

<220>

<400> 382

GTAGCTTCAG TGGTCCGACG

20

<210> 383

<211> 20

<212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 383
 TGGGCATCAG TGAAGAGCTG

20

<210> 384
 <211> 20
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 384
 TCTCCCAGAG TCTTCCCCAG

20

<210> 385
 <211> 23
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 385
 GAGTTTGGAG CTGTGATGGA AGG

23

<210> 386
 <211> 21
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 386
 TCGTCTTCAC AGCCACCTTG A

21

<210> 387
 <211> 20
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 387
 CCATCTGCAC GAGGTCAGAG

20

<210> 388
 <211> 20
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 388
 GAGCCTCATG ACGTTGGGAT

20

<210> 389
 <211> 21
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 389
 TCCAGGGTTC TGAACGAGAG A

21

<210> 390
 <211> 21
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 390
 AGAGGAGGAA GCTGTGTAGG T

21

<210> 391
 <211> 20
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 391
 ACCTGCCCCAC TCAGATGCTA

20

<210> 392
 <211> 20
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 392
 CCAGGTCCCG GTGTATGAAT

20

<210> 393
 <211> 20
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 393
 AGGGACGTAT CGCCAAGATG

20

<210> 394
 <211> 20
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>	
<400> 394	
CTGGTGTAGA CACGGTCAGC	20
<210> 395	
<211> 20	
<212> ADN	
<213> Đoạn mỗi	
<220> Nhân tạo	
<223>	
<220>	
<400> 395	
TCCTTCGGGG TGACAATGTG	20
<210> 396	
<211> 20	
<212> ADN	
<213> Đoạn mỗi	
<220> Nhân tạo	
<223>	
<220>	
<400> 396	
CGATTTCCCT GGC GCAGATA	20
<210> 397	
<211> 20	
<212> ADN	
<213> Đoạn mỗi	
<220> Nhân tạo	
<223>	
<220>	
<400> 397	
CATCCTGCTG GACGCTATGT	20
<210> 398	
<211> 20	
<212> ADN	
<213> Đoạn mỗi	
<220> Nhân tạo	
<223>	
<220>	
<400> 398	
AGTACCAGGT GGAGGGTTGT	20
<210> 399	
<211> 20	
<212> ADN	
<213> Đoạn mỗi	
<220> Nhân tạo	
<223>	
<220>	
<400> 399	
CACGGGTAAG ACGCAACTGA	20
<210> 400	
<211> 20	
<212> ADN	
<213> Đoạn mỗi	

<220> Nhân tạo
<223>

<220>
<400> 400
AAACCCAATC AGGCCTACCG

20

<210> 401
<211> 20
<212> ADN
<213> Đoạn mỗi
<220> Nhân tạo
<223>

<220>
<400> 401
TCTTCCCAGC GCATTCCTTT

20

<210> 402
<211> 21
<212> ADN
<213> Đoạn mỗi
<220> Nhân tạo
<223>

<220>
<400> 402
TGGTTGCCT GCAATTGCTT C

21

<210> 403
<211> 20
<212> ADN
<213> Đoạn mỗi
<220> Nhân tạo
<223>

<220>
<400> 403
TGAAATTGGT CGCCCTCTGG

20

<210> 404
<211> 20
<212> ADN
<213> Đoạn mỗi
<220> Nhân tạo
<223>

<220>
<400> 404
CTGAGCTGAT GCTCCACTCC

20

<210> 405
<211> 20
<212> ADN
<213> Đoạn mỗi
<220> Nhân tạo
<223>

<220>
<400> 405
TGCACCACTT GGAACAGTTT

20

<210> 406

<211> 22
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 406
 AGTAGCAGTT CTCTGCTCAT CA

22

<210> 407
 <211> 21
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 407
 TGGCAAATGC CCTGTCTTAC T

21

<210> 408
 <211> 20
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 408
 GGAGCATGTA CTGACCACCC

20

<210> 409
 <211> 20
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 409
 CTCTGTGGCA CCCTGGACTA

20

<210> 410
 <211> 20
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 410
 TCCAAGGCTC CAGAGATCCA

20

<210> 411
 <211> 20
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 411

ACAGGAGGCA AATCCAGAGC

20

<210> 421

<211> 20

<212> ADN

<213> Đoạn mỗi

<220> Nhân tạo

<223>

<220>

<400> 412

AAACCCAATC AGGCCTACCG

20

<210> 413

<211> 17

<212> ADN

<213> Đoạn mỗi

<220> Nhân tạo

<223>

<220>

<400> 413

GACGGCCGCG GGAAAAG

17

<210> 414

<211> 20

<212> ADN

<213> Đoạn mỗi

<220> Nhân tạo

<223>

<220>

<400> 414

ATGCCGAATA CCGACAGGGT

20

<210> 415

<211> 20

<212> ADN

<213> Đoạn mỗi

<220> Nhân tạo

<223>

<220>

<400> 415

ATTCCCTCTG CTGACAACCA

20

<210> 416

<211> 23

<212> ADN

<213> Đoạn mỗi

<220> Nhân tạo

<223>

<220>

<400> 416

CTTGATAAAC CACAGCAACT CCT

23

<210> 417

<211> 20

<212> ADN

<213> Đoạn mỗi

<220> Nhân tạo

<223>

<220>	
<400> 417	
ATGCCAATGG ATCCCGAGAC	20
<210> 418	
<211> 20	
<212> ADN	
<213> Đoạn mỗi	
<220> Nhân tạo	
<223>	
<220>	
<400> 418	
ATAAACTGGG CCCAAGTCCC	20
<210> 419	
<211> 20	
<212> ADN	
<213> Đoạn mỗi	
<220> Nhân tạo	
<223>	
<220>	
<400> 419	
TCAGGACAGG GAGTTGACCA	20
<210> 420	
<211> 20	
<212> ADN	
<213> Đoạn mỗi	
<220> Nhân tạo	
<223>	
<220>	
<400> 420	
TGGATTCCAA GCGCACATGA	20
<210> 421	
<211> 20	
<212> ADN	
<213> Đoạn mỗi	
<220> Nhân tạo	
<223>	
<220>	
<400> 421	
ATCTTCCTCT GATGGCGCTG	20
<210> 422	
<211> 20	
<212> ADN	
<213> Đoạn mỗi	
<220> Nhân tạo	
<223>	
<220>	
<400> 422	
GGACAGCTCA CTGTTACCA	20
<210> 423	
<211> 25	
<212> ADN	

<213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 423
 CAGGTGACTT TATACACACT TTGGG

25

<210> 424
 <211> 23
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 434
 TGAATTTTCA GTGGCTCGAT GTG

23

<210> 425
 <211> 20
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 425
 CAGCGAGAAC CCAGACCTTT

20

<210> 426
 <211> 20
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 426
 CCAGCCCAAC CCCTAAAGAC

20

<210> 427
 <211> 20
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 427
 TTCAAGAATG GGGTCGGCTC

20

<210> 428
 <211> 20
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 428
 CTTTGGAATT CCGCAGCGAC

20

<210> 429
 <211> 20
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 429
 GATGTATGCC TACCACCGGG

20

<210> 430
 <211> 20
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 430
 AGGGGTGCAC AGTTTTGGAA

20

<210> 431
 <211> 20
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 431
 CTCCATCAAG GGAGACCTGC

20

<210> 432
 <211> 20
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 432
 CAAAAACAGC TGGCCTCTGC

20

<210> 433
 <211> 21
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 433
 ACTCCAGTTC CAACACCATC T

21

<210> 434
 <211> 20
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>

<400> 434
ATAGGCAGCA GCAGGAACAG

20

<210> 435
<211> 20
<212> ADN
<213> Đoạn mỗi
<220> Nhân tạo
<223>

<220>
<400> 435
GCATGTCAGG AGGCCAGATT

20

<210> 436
<211> 20
<212> ADN
<213> Đoạn mỗi
<220> Nhân tạo
<223>

<220>
<400> 436
ATTTGGCAGC TGTGGACTCT

20

<210> 437
<211> 20
<212> ADN
<213> Đoạn mỗi
<220> Nhân tạo
<223>

<220>
<400> 437
AGAAGAAGGG GATGGAGCCT

20

<210> 438
<211> 20
<212> ADN
<213> Đoạn mỗi
<220> Nhân tạo
<223>

<220>
<400> 438
GCAAACTCGA TGCCATGACC

20

<210> 439
<211> 20
<212> ADN
<213> Đoạn mỗi
<220> Nhân tạo
<223>

<220>
<400> 439
CGGTTCATTC CAGTCACCGA

20

<210> 440
<211> 20
<212> ADN
<213> Đoạn mỗi
<220> Nhân tạo

<223>

<220>

<400> 440

AGCCAGACAC AGCCGTAAAG

20

<210> 441

<211> 20

<212> ADN

<213> Đoạn mỗi

<220> Nhân tạo

<223>

<220>

<400> 441

GGCAGCAGTT TGTACTCCCT

20

<210> 442

<211> 22

<212> ADN

<213> Đoạn mỗi

<220> Nhân tạo

<223>

<220>

<400> 442

TCCAACAGCT CCATCATAGA CA

22

<210> 443

<211> 20

<212> ADN

<213> Đoạn mỗi

<220> Nhân tạo

<223>

<220>

<400> 443

TTGACCGATG GTGTGTCTGG

20

<210> 444

<211> 20

<212> ADN

<213> Đoạn mỗi

<220> Nhân tạo

<223>

<220>

<400> 444

GTAGCCATTG GTGGCACTCT

20

<210> 445

<211> 20

<212> ADN

<213> Đoạn mỗi

<220> Nhân tạo

<223>

<220>

<400> 445

ACGGGCTCCT GGTCTACTTC

20

<210> 446

<211> 20

<212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 446
 CGTTTGAGCT CCCGCATGTA

20

<210> 447
 <211> 20
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 447
 CTATGAGAAG CAGGAGGGCG

20

<210> 448
 <211> 20
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 448
 ACACAAGTTT TTGCGGGAGC

20

<210> 449
 <211> 20
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 449
 TTTGGGAAGG ACGTCGATGG

20

<210> 450
 <211> 20
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 450
 GCTGGTGTCA TGCTGGTACT

20

<210> 451
 <211> 20
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 451
 GCTGGCATCT ACAACCTGGA

20

<210> 452
 <211> 20
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 452
 AGCAAGGAAG TATGGGCACC

20

<210> 453
 <211> 20
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 453
 GCATGCCAAT GTGGTGGTTT

20

<210> 454
 <211> 20
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 454
 CCAGTTCCTT GGACACCAGG

20

<210> 455
 <211> 20
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 455
 TCGACTCCAT GAGCGTCAAC

20

<210> 456
 <211> 20
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 456
 TTCTGCAGGG TCTCCAGGTT

20

<210> 457
 <211> 20
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220> <400> 457 AAAGAGACAG ACGAGCAGCG	20
<210> 458 <211> 17 <212> ADN <213> Đoạn mỗi <220> Nhân tạo <223>	
<220> <400> 458 ACTTCGTCGG GCAGCAC	17
<210> 459 <211> 20 <212> ADN <213> Đoạn mỗi <220> Nhân tạo <223>	
<220> <400> 459 GAGCATTTCC TGGGCTTCCT	20
<210> 460 <211> 19 <212> ADN <213> Đoạn mỗi <220> Nhân tạo <223>	
<220> <400> 460 CTCTCCTGCA CCACATGCT	19
<210> 461 <211> 20 <212> ADN <213> Đoạn mỗi <220> Nhân tạo <223>	
<220> <400> 461 TCACCGACCT TGCTGACTTC	20
<210> 462 <211> 20 <212> ADN <213> Đoạn mỗi <220> Nhân tạo <223>	
<220> <400> 462 TTTGGCGTAG GTGCTGACAA	20
<210> 463 <211> 21 <212> ADN <213> Đoạn mỗi	

<220> Nhân tạo
<223>

<220>
<400> 463
GCTTCACCAT CATCATCGAG C

21

<210> 464
<211> 20
<212> ADN
<213> Đoạn mỗi
<220> Nhân tạo
<223>

<220>
<400> 464
CTGAAGTGCA GGATGGGGTT

20

<210> 465
<211> 20
<212> ADN
<213> Đoạn mỗi
<220> Nhân tạo
<223>

<220>
<400> 465
ATTCGAGTAG CGGCTCTTCC

20

<210> 466
<211> 19
<212> ADN
<213> Đoạn mỗi
<220> Nhân tạo
<223>

<220>
<400> 466
CTTGGAACGG CCACCAAGA

19

<210> 467
<211> 20
<212> ADN
<213> Đoạn mỗi
<220> Nhân tạo
<223>

<220>
<400> 467
GGGAACTTTG GCTGCCATCA

20

<210> 468
<211> 22
<212> ADN
<213> Đoạn mỗi
<220> Nhân tạo
<223>

<220>
<400> 468
TGTCATTCT CCAAACACCA GC

22

<210> 469

<211> 20
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 469
 CCTGGCAGCT GGAAGACAAA

20

<210> 470
 <211> 20
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 470
 GGGTGTTAAG CTCCCCAACA

20

<210> 471
 <211> 22
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 471
 GGATTTACAC GTGGTTGGAA GC

22

<210> 472
 <211> 20
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 472
 CAGCAGCTGA CAGTCCAAGA

20

<210> 473
 <211> 20
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 473
 AGCGTATGCA AAAGCTGGAG

20

<210> 474
 <211> 22
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 474

TGCCTCCAA ATGCAATCAC AG

22

<210> 475

<211> 20

<212> ADN

<213> Đoạn mỗi

<220> Nhân tạo

<223>

<220>

<400> 475

TAGGTGCTGC TTTCCTGCTG

20

<210> 476

<211> 20

<212> ADN

<213> Đoạn mỗi

<220> Nhân tạo

<223>

<220>

<400> 476

TTCCCCTGAG AGGACCAAGG

20

<210> 477

<211> 20

<212> ADN

<213> Đoạn mỗi

<220> Nhân tạo

<223>

<220>

<400> 477

AACCTGAAGG TGCAGAGTGG

20

<210> 478

<211> 20

<212> ADN

<213> Đoạn mỗi

<220> Nhân tạo

<223>

<220>

<400> 478

TCCCCTCTGT GGGGTCATAG

20

<210> 479

<211> 20

<212> ADN

<213> Đoạn mỗi

<220> Nhân tạo

<223>

<220>

<400> 479

AACATTTCGCT ATGGCCGTGA

20

<210> 480

<211> 20

<212> ADN

<213> Đoạn mỗi

<220> Nhân tạo

<223>

<220> <400> 480 AGGCATTGGC TTCCTTGACA	20
<210> 481 <211> 20 <212> ADN <213> Đoạn mỗi <220> Nhân tạo <223>	
<220> <400> 481 TGACACCCTG GTTGGAGAGA	20
<210> 482 <211> 20 <212> ADN <213> Đoạn mỗi <220> Nhân tạo <223>	
<220> <400> 482 GAGGATCTTG GGGTTGCGAA	20
<210> 483 <211> 20 <212> ADN <213> Đoạn mỗi <220> Nhân tạo <223>	
<220> <400> 483 CAGAAAAGGT CGGACCACCA	20
<210> 484 <211> 20 <212> ADN <213> Đoạn mỗi <220> Nhân tạo <223>	
<220> <400> 484 CATCGAAACC AGCGATGACG	20
<210> 485 <211> 20 <212> ADN <213> Đoạn mỗi <220> Nhân tạo <223>	
<220> <400> 485 TGCTCAAGAC TGGCGCTAAA	20
<210> 486 <211> 20 <212> ADN	

<213> Đoạn mỗi
<220> Nhân tạo
<223>

<220>
<400> 486
CAATCCAGGG AAGCGTGTCA

20

<210> 487
<211> 18
<212> ADN
<213> Đoạn mỗi
<220> Nhân tạo
<223>

<220>
<400> 487
CTGCCTTCCG GGTCCTG

18

<210> 488
<211> 20
<212> ADN
<213> Đoạn mỗi
<220> Nhân tạo
<223>

<220>
<400> 488
GTTTCCTGAC TCAGAGGGGG

20

<210> 489
<211> 20
<212> ADN
<213> Đoạn mỗi
<220> Nhân tạo
<223>

<220>
<400> 489
TGGATCCATT GGAAGGGCAG

20

<210> 490
<211> 20
<212> ADN
<213> Đoạn mỗi
<220> Nhân tạo
<223>

<220>
<400> 490
TTTCGCTTCC CACAGGTCTC

20

<210> 491
<211> 18
<212> ADN
<213> Đoạn mỗi
<220> Nhân tạo
<223>

<220>
<400> 491
CTGCCTTCCG GGTCCTG

18

<210> 492
 <211> 20
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 492
 GTTTCCTGAC TCAGAGGGGG

20

<210> 492
 <211> 20
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 493
 TGAAGCTCCC AGAATGCCAG

20

<210> 494
 <211> 20
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 494
 GCTGCCCTGG TAGGTTTTCT

20

<210> 495
 <211> 20
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 495
 GCCAACTCTC TCTAGCTCGC

20

<210> 496
 <211> 20
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 496
 CTCCGTCATG TGCTGTGACT

20

<210> 497
 <211> 20
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>

<400> 497 CCCTGCCCTC AACAAAGATGT	20
<210> 498 <211> 20 <212> ADN <213> Đoạn mỗi <220> Nhân tạo <223>	
<220> <400> 498 CTCCGTCATG TGCTGTGACT	20
<210> 499 <211> 21 <212> ADN <213> Đoạn mỗi <220> Nhân tạo <223>	
<220> <400> 499 GCCCCTCCTC AGCATCTTAT C	21
<210> 500 <211> 20 <212> ADN <213> Đoạn mỗi <220> Nhân tạo <223>	
<220> <400> 500 CTCATAGGGC ACCACCACAC	20
<210> 501 <211> 20 <212> ADN <213> Đoạn mỗi <220> Nhân tạo <223>	
<220> <400> 501 TGGCTCTGAC TGTACCACCA	20
<210> 502 <211> 20 <212> ADN <213> Đoạn mỗi <220> Nhân tạo <223>	
<220> <400> 502 TGATGATGGT GAGGATGGGC	20
<210> 503 <211> 20 <212> ADN <213> Đoạn mỗi <220> Nhân tạo	

<223>

<220>

<400> 503

TGGGACGGAA CAGCTTTGAG

20

<210> 504

<211> 20

<212> ADN

<213> Đoạn mỗi

<220> Nhân tạo

<223>

<220>

<400> 504

ATTCTCTTCC TCTGTGCGCC

20

<210> 505

<211> 19

<212> ADN

<213> Đoạn mỗi

<220> Nhân tạo

<223>

<220>

<400> 505

CACTGCCCAA CAACACCAG

19

<210> 506

<211> 24

<212> ADN

<213> Đoạn mỗi

<220> Nhân tạo

<223>

<220>

<400> 506

AGGGTGAAAT ATTCTCCATC CAGT

24

<210> 507

<211> 23

<212> ADN

<213> Đoạn mỗi

<220> Nhân tạo

<223>

<220>

<400> 507

GACCAGACCA GCTTTCAAAA AGA

23

<210> 508

<211> 23

<212> ADN

<213> Đoạn mỗi

<220> Nhân tạo

<223>

<220>

<400> 508

TATCAGGCAA AGTCATAGAA CCA

23

<210> 509

<211> 20

<212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 509
 CGCTTCGAGA TGTTCCGAGA

20

<210> 510
 <211> 18
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 510
 CTGGCTCCTT CCCAGCCT

18

<210> 511
 <211> 20
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 511
 TGCATTTTCA CCCCACCCTT

20

<210> 512
 <211> 20
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 512
 GACACACAGG TGGCAGCAAA

20

<210> 513
 <211> 18
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 513
 CCCAACACAC GGTCCACA

18

<210> 514
 <211> 17
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 514
 CTGCTGGGAC GGGAGTC

17

<210> 515
 <211> 18
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 515
 AGAGACTCTC CGAGGCGG

18

<210> 516
 <211> 17
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 516
 CTGCTCGCCC AGGTTCC

17

<210> 517
 <211> 17
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 517
 CTGAGTGCCC CGGAACG

17

<210> 518
 <211> 18
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 518
 CTGGAGGGAC TCGGGCTA

18

<210> 519
 <211> 20
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 519
 CTTCTGTCA CCCACATCCC

20

<210> 520
 <211> 20
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 520
 AGGCTGGAAA GGAGGTGTTG

20

<210> 521
 <211> 20
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 521
 GTGTACGACG AGGACGACTC

20

<210> 522
 <211> 20
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 522
 TCCTTCCGCT TGAACTCCAC

20

<210> 523
 <211> 20
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 523
 CAGACGTGGT ACCTGCATCA

20

<210> 524
 <211> 19
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 524
 GTTCCACCCG GAAGGTCTG

19

<210> 525
 <211> 20
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 525
 TCTCAACCGG AACCTCTCCA

20

<210> 526
 <211> 20
 <212> ADN
 <213> Đoạn mỗi

<220> Nhân tạo
<223>

<220>
<400> 526
GAGGGAGTAG GTACTGGGCA

20

<210> 527
<211> 20
<212> ADN
<213> Đoạn mỗi
<220> Nhân tạo
<223>

<220>
<400> 527
TTCAGTGGCA TCAACGTCCA

20

<210> 528
<211> 20
<212> ADN
<213> Đoạn mỗi
<220> Nhân tạo
<223>

<220>
<400> 528
GGCCACCACC ATGTAGTGAA

20

<210> 529
<211> 20
<212> ADN
<213> Đoạn mỗi
<220> Nhân tạo
<223>

<220>
<400> 529
TCCTGTGCAG TCAGCAAGAG

20

<210> 530
<211> 20
<212> ADN
<213> Đoạn mỗi
<220> Nhân tạo
<223>

<220>
<400> 530
CCATCTCCTT GCCTCCACTC

20

<210> 531
<211> 20
<212> ADN
<213> Đoạn mỗi
<220> Nhân tạo
<223>

<220>
<400> 531
CCAGGAACTT CTCCACCTGC

20

<210> 532

<211> 20
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 532
 GGTCTCCACA ACATCGCAGT

20

<210> 533
 <211> 19
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 533
 ATTGTTGGGG GACACGAGC

19

<210> 534
 <211> 20
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 534
 CCTCCTCAGG GAAGCAGTTG

20

<210> 535
 <211> 20
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 535
 GTCTGCCTTC TGCATCCACT

20

<210> 536
 <211> 20
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 536
 TTGAAGGCAT CCACGGAGAC

20

<210> 537
 <211> 20
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 537

AAAAGTGGTC AAGGCAACGG

20

<210> 538
<211> 20
<212> ADN
<213> Đoạn mỗi
<220> Nhân tạo
<223>

<220>
<400> 538
TTTCAGCATC TTCACGGCCA

20

<210> 539
<211> 20
<212> ADN
<213> Đoạn mỗi
<220> Nhân tạo
<223>

<220>
<400> 539
CGAGTGAGCT TCGAGACCTG

20

<210> 540
<211> 21
<212> ADN
<213> Đoạn mỗi
<220> Nhân tạo
<223>

<220>
<400> 540
TGCAGGCCCC ATACAATTTG A

21

<210> 541
<211> 20
<212> ADN
<213> Đoạn mỗi
<220> Nhân tạo
<223>

<220>
<400> 541
AGTACGCCAA ATACGGCTCC

20

<210> 542
<211> 20
<212> ADN
<213> Đoạn mỗi
<220> Nhân tạo
<223>

<220>
<400> 542
GCCAGATACT GCATCCCCTG

20

<210> 543
<211> 20
<212> ADN
<213> Đoạn mỗi
<220> Nhân tạo
<223>

<220> <400> 543 TTGGCAGCCA GAAACATCCT	20
<210> 544 <211> 20 <212> ADN <213> Đoạn mỗi <220> Nhân tạo <223>	
<220> <400> 544 TCTCGGGACA AGCCGAAATC	20
<210> 545 <211> 20 <212> ADN <213> Đoạn mỗi <220> Nhân tạo <223>	
<220> <400> 545 ATCGTGACCC TAGGGGGAAA	20
<210> 546 <211> 20 <212> ADN <213> Đoạn mỗi <220> Nhân tạo <223>	
<220> <400> 546 GCCGGTCTTC AGAAGGTTGA	20
<210> 547 <211> 20 <212> ADN <213> Đoạn mỗi <220> Nhân tạo <223>	
<220> <400> 547 TGATGCTGCA ATGCTGGAAG	20
<210> 548 <211> 21 <212> ADN <213> Đoạn mỗi <220> Nhân tạo <223>	
<220> <400> 548 TCCAGGTCTT TGCTGATGTC C	21
<210> 549 <211> 20 <212> ADN	

<213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 549
 TCTCAGAGGA GGAGACACCG

20

<210> 550
 <211> 20
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 550
 ACCATCCTAA GTTGCTGGGC

20

<210> 551
 <211> 20
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 551
 GGCCTGGAGA GAGTCCTGTA

20

<210> 552
 <211> 20
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 552
 CTTGTGAGCC CTCTTCCCTG

20

<210> 553
 <211> 20
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 553
 TATAAGATGG CGGCGCTGAG

20

<210> 554
 <211> 20
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 554
 CCATGTCCCC GTTGAACAGA

20

<210> 555
 <211> 25
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 555
 AAATGATTAA GTTGACACAG GAACA

25

<210> 556
 <211> 20
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 556
 GGTGGATTAT GCTCCCCACC

20

<210> 557
 <211> 20
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 557
 TCTCTGGGGA ACGGAACTGA

20

<210> 558
 <211> 20
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 558
 TCCTCTGTTT GTTGGGCAGG

20

<210> 559
 <211> 20
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 559
 GTACCTGCAA GGTGTGGAGT

20

<210> 560
 <211> 20
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>

<400> 560
CAGCACAGCA CTCTGGGATT

20

<210> 561
<211> 20
<212> ADN
<213> Đoạn mỗi
<220> Nhân tạo
<223>

<220>
<400> 561
CCAATTGGTT GGGACACTGA

20

<210> 562
<211> 23
<212> ADN
<213> Đoạn mỗi
<220> Nhân tạo
<223>

<220>
<400> 562
TGGAACATTC TCCAACACTT CCA

23

<210> 563
<211> 20
<212> ADN
<213> Đoạn mỗi
<220> Nhân tạo
<223>

<220>
<400> 563
GCTGCTTTTC CAGGGTTTCC

20

<210> 564
<211> 23
<212> ADN
<213> Đoạn mỗi
<220> Nhân tạo
<223>

<220>
<400> 564
AACTTCTGTA CTACAACGCT GGT

23

<210> 565
<211> 20
<212> ADN
<213> Đoạn mỗi
<220> Nhân tạo
<223>

<220>
<400> 565
AGAGGCGTCC TTAGCAGAGA

20

<210> 566
<211> 19
<212> ADN
<213> Đoạn mỗi
<220> Nhân tạo

<223>

<220>

<400> 566

AATAGAGTCC GAGGCGGGT

19

<210> 567

<211> 20

<212> ADN

<213> Đoạn mỗi

<220> Nhân tạo

<223>

<220>

<400> 567

TCCAATTCCA CAGCCCTTCC

20

<210> 568

<211> 20

<212> ADN

<213> Đoạn mỗi

<220> Nhân tạo

<223>

<220>

<400> 568

GCTGATGAGG ATCGGTCTCG

20

<210> 569

<211> 20

<212> ADN

<213> Đoạn mỗi

<220> Nhân tạo

<223>

<220>

<400> 569

TGCCTCATTA CCTGGCTCAC

20

<210> 570

<211> 20

<212> ADN

<213> Đoạn mỗi

<220> Nhân tạo

<223>

<220>

<400> 570

CCTTTCTCGC TGAGGTCCTG

20

<210> 571

<211> 20

<212> ADN

<213> Đoạn mỗi

<220> Nhân tạo

<223>

<220>

<400> 571

AGACGGGACT CGAGTGATGA

20

<210> 572

<211> 21

<212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 572
 TGCCACTTTC CCTTGTAGAC T

21

<210> 573
 <211> 21
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 573
 GTGATGTGGC AGTGAAAATG T

21

<210> 574
 <211> 25
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 574
 TCATTTTTGA AGGCTTGTA CTGCT

25

<210> 575
 <211> 20
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 575
 GTTACCCAGT GGTGTGAGGG

20

<210> 576
 <211> 20
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 576
 TGTGCAGTCT GTCGTGCAAT

20

<210> 577
 <211> 24
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 577
 AGACCTCACA GTAAAAATAG GTGA

24

<210> 578
 <211> 20
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 578
 CTGATGGGAC CCACTCCATC

20

<210> 579
 <211> 22
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 579
 GCACCAGAAG TCATCAGAAT GC

22

<210> 580
 <211> 20
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 580
 CTGGTCCCTG TTGTTGATGT

20

<210> 581
 <211> 20
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 581
 TGGTGGGACG AGGATACCTG

20

<210> 582
 <211> 21
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 582
 TGGGGAAAGA GTGGTCTCTC A

21

<210> 583
 <211> 20
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 583
 CTCCTTGAAT CGGGCTGGTT

20

<210> 584
 <211> 20
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 584
 CGACTGCCAA CTTCTCACCT

20

<210> 585
 <211> 21
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 585
 CCGGCCATCT AAAAATCTGC C

21

<210> 586
 <211> 20
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 586
 GGCAGCCAAG AGTTGCAAAA

20

<210> 587
 <211> 20
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 587
 ACCAACTGCT CCCTTTGCTT

20

<210> 588
 <211> 20
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 588
 CAGCTTCCCA GAAGTTGTGC

20

<210> 589
 <211> 20
 <212> ADN
 <213> Đoạn mỗi

<220> Nhân tạo
<223>

<220>
<400> 589
TCCAGACCGT CCTATGTGGT

20

<210> 590
<211> 20
<212> ADN
<213> Đoạn mỗi
<220> Nhân tạo
<223>

<220>
<400> 590
TGGGACTTGG AGGGGAGTAG

20

<210> 591
<211> 20
<212> ADN
<213> Đoạn mỗi
<220> Nhân tạo
<223>

<220>
<400> 591
GAGAGCTCAA GTCCCGACAC

20

<210> 592
<211> 20
<212> ADN
<213> Đoạn mỗi
<220> Nhân tạo
<223>

<220>
<400> 592
CTGGTTCTCT GTGTCCCTGC

20

<210> 593
<211> 21
<212> ADN
<213> Đoạn mỗi
<220> Nhân tạo
<223>

<220>
<400> 593
CTGTTGATGT CCACCAGCAA A

21

<210> 594
<211> 20
<212> ADN
<213> Đoạn mỗi
<220> Nhân tạo
<223>

<220>
<400> 594
TCAGACATGT TGGCAGCCTT

20

<210> 595

<211> 22
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 595
 AGAACTGCTC AAACATCGAG GA

22

<210> 596
 <211> 22
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 596
 CCATTGTATG AATCAGCCAC AA

22

<210> 597
 <211> 20
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 597
 GCAATCAAGC CACAAGCCAA

20

<210> 598
 <211> 20
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 598
 AACAAGAGCC TGGTGAGAGC

20

<210> 599
 <211> 20
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 599
 AATTAGGCCC TTCTGCCTGG

20

<210> 600
 <211> 20
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 600

ACTTCAGGAG GGTCTTGGGT

20

<210> 601

<211> 22

<212> ADN

<213> Đoạn mỗi

<220> Nhân tạo

<223>

<220>

<400> 601

ACCACCATTT ATCATGGCTG TG

22

<210> 602

<211> 20

<212> ADN

<213> Đoạn mỗi

<220> Nhân tạo

<223>

<220>

<400> 602

AAGAACTCTC CTGGGCCAAA

20

<210> 603

<211> 20

<212> ADN

<213> Đoạn mỗi

<220> Nhân tạo

<223>

<220>

<400> 603

AAGGCGACGT TTTTGTGTGG

20

<210> 604

<211> 20

<212> ADN

<213> Đoạn mỗi

<220> Nhân tạo

<223>

<220>

<400> 604

CTGCAATGCT GGGTAGGTGA

20

<210> 605

<211> 20

<212> ADN

<213> Đoạn mỗi

<220> Nhân tạo

<223>

<220>

<400> 605

CAGTCTGTGT TGACGGGACA

20

<210> 606

<211> 20

<212> ADN

<213> Đoạn mỗi

<220> Nhân tạo

<223>

<220> <400> 606 TCCCAGAAAA CCAAGGCTGT	20
<210> 607 <211> 20 <212> ADN <213> Đoạn mỗi <220> Nhân tạo <223>	
<220> <400> 607 TGGGCTCCTG TATTGGTTGG	20
<210> 608 <211> 20 <212> ADN <213> Đoạn mỗi <220> Nhân tạo <223>	
<220> <400> 608 CTCTGTCCCC GAAGAACAGC	20
<210> 609 <211> 20 <212> ADN <213> Đoạn mỗi <220> Nhân tạo <223>	
<220> <400> 609 ATAGTGGTCG GCTGTTCTGG	20
<210> 610 <211> 20 <212> ADN <213> Đoạn mỗi <220> Nhân tạo <223>	
<220> <400> 610 GCAGGGGCTT AAGGGATGTC	20
<210> 611 <211> 24 <212> ADN <213> Đoạn mỗi <220> Nhân tạo <223>	
<220> <400> 611 TCCTTTACCC CTAAGGTTAT TCCA	24
<210> 612 <211> 20 <212> ADN	

<213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 612
 GGGGACCATT CCACAGGATT

20

<210> 613
 <211> 20
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 613
 ACTGTGGAAG GACTGGAACC

20

<210> 614
 <211> 20
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 614
 CAGGTGCTCG AAGTGACAGA

20

<210> 615
 <211> 20
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 615
 GCACCAGAGA ACCCCAGAAT

20

<210> 616
 <211> 20
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 616
 GCATCTGGGA CTCATGCCTT

20

<210> 617
 <211> 20
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 617
 TGTGTGGACG TTTTCAGCAG

20

<210> 618
 <211> 20
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 618
 AGAGTGAGTC CCCTTCAGCA

20

<210> 619
 <211> 20
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 619
 AGGCACCTCT ACTTTGCACT

20

<210> 620
 <211> 22
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 620
 TGTGAATCTT CACCTCTCTG GG

22

<210> 621
 <211> 20
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 621
 AGTCACACCT TGCACCGAAT

20

<210> 622
 <211> 20
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 622
 AGCCTTCCAG ATCCATTGCC

20

<210> 623
 <211> 20
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>

<400> 623
CACAGCAAAG GACAGCACAC

20

<210> 624
<211> 20
<212> ADN
<213> Đoạn mỗi
<220> Nhân tạo
<223>

<220>
<400> 624
ACTCCTTAGG GCCTTCACCT

20

<210> 625
<211> 21
<212> ADN
<213> Đoạn mỗi
<220> Nhân tạo
<223>

<220>
<400> 625
TCAGACACTG TGGAACCAAC T

21

<210> 626
<211> 20
<212> ADN
<213> Đoạn mỗi
<220> Nhân tạo
<223>

<220>
<400> 626
GGTTTGTCTT GGCCAGAGGT

20

<210> 627
<211> 25
<212> ADN
<213> Đoạn mỗi
<220> Nhân tạo
<223>

<220>
<400> 627
TCAGGACAGT ATAGCTCTTA TTGAA

25

<210> 628
<211> 22
<212> ADN
<213> Đoạn mỗi
<220> Nhân tạo
<223>

<220>
<400> 628
TCCCCAAATC TCTTTTCCTG GT

22

<210> 629
<211> 20
<212> ADN
<213> Đoạn mỗi
<220> Nhân tạo

<223>

<220>

<400> 629

GGTCAGACAC CAGCCTCATT

20

<210> 630

<211> 20

<212> ADN

<213> Đoạn mỗi

<220> Nhân tạo

<223>

<220>

<400> 630

CCAGGTGTGA GATTGCCAAC

20

<210> 631

<211> 19

<212> ADN

<213> Đoạn mỗi

<220> Nhân tạo

<223>

<220>

<400> 631

CCTGCCACTC TGAGGAAAT

19

<210> 632

<211> 20

<212> ADN

<213> Đoạn mỗi

<220> Nhân tạo

<223>

<220>

<400> 632

AGTGTTCTCT GGAACCAAGG

20

<210> 633

<211> 20

<212> ADN

<213> Đoạn mỗi

<220> Nhân tạo

<223>

<220>

<400> 633

GCAGCCAAGG TCCTGCTTAT

20

<210> 634

<211> 20

<212> ADN

<213> Đoạn mỗi

<220> Nhân tạo

<223>

<220>

<400> 634

GCTTTCTGGA AGTGAGGTGC

20

<210> 635

<211> 21

<212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 635
 GGAGTCCCAA ATAAACCAGG C

21

<210> 636
 <211> 21
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 636
 TCTTCAGCTT TCTCCCACTG T

21

<210> 637
 <211> 20
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 637
 TCCTGCAGTA GTGTTTGCAC

20

<210> 638
 <211> 22
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 638
 TGCAGCTACT ACTCTGAACT GA

22

<210> 639
 <211> 20
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 639
 CCAAGGAAGG GGTGACAGTG

20

<210> 640
 <211> 20
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 640
 CATTAGCCAG GCCTACTCCG

20

<210> 641
 <211> 20
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 641
 CTGCCTTCCC TCGGGAAAAA

20

<210> 642
 <211> 20
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 642
 TGTCCACTGC TGTTCCTTCA

20

<210> 643
 <211> 20
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 643
 GCAGCTTGGA GTTTGTCTGC

20

<210> 644
 <211> 20
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 644
 AAGGTCTCCT CCCTCCATCA

20

<210> 645
 <211> 21
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 645
 ACTCACCTTG GTTGACCTTG T

21

<210> 646
 <211> 21
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>	
<400> 646	
TGCATCCGTT CCAAGTAGAC A	21
<210> 647	
<211> 20	
<212> ADN	
<213> Đoạn mỗi	
<220> Nhân tạo	
<223>	
<220>	
<400> 647	
TTGGAGACTT TGGACTCGCC	20
<210> 648	
<211> 20	
<212> ADN	
<213> Đoạn mỗi	
<220> Nhân tạo	
<223>	
<220>	
<400> 648	
ACTTTCTGGA GCCATCCACC	20
<210> 649	
<211> 21	
<212> ADN	
<213> Đoạn mỗi	
<220> Nhân tạo	
<223>	
<220>	
<400> 649	
TGGTCATCAG CCTTATCCAG C	21
<210> 650	
<211> 20	
<212> ADN	
<213> Đoạn mỗi	
<220> Nhân tạo	
<223>	
<220>	
<400> 650	
CCAGTCTCCC TCCTGTTTGC	20
<210> 651	
<211> 20	
<212> ADN	
<213> Đoạn mỗi	
<220> Nhân tạo	
<223>	
<220>	
<400> 651	
GGCTCAAGAA CCCGACCAAA	20
<210> 652	
<211> 22	
<212> ADN	
<213> Đoạn mỗi	

<220> Nhân tạo
<223>

<220>
<400> 652
GACTCCACTG TTGTTTGCTT CA

22

<210> 653
<211> 20
<212> ADN
<213> Đoạn mỗi
<220> Nhân tạo
<223>

<220>
<400> 653
GTCCTCTAGG CTCCCAGGAA

20

<210> 654
<211> 20
<212> ADN
<213> Đoạn mỗi
<220> Nhân tạo
<223>

<220>
<400> 654
CCACTGTGAG TGAGACAGGC

20

<210> 655
<211> 19
<212> ADN
<213> Đoạn mỗi
<220> Nhân tạo
<223>

<220>
<400> 655
CCTGGTCATC TGCGGACTC

19

<210> 656
<211> 19
<212> ADN
<213> Đoạn mỗi
<220> Nhân tạo
<223>

<220>
<400> 656
GAGTTGCCCC CTCACTCCC

19

<210> 657
<211> 20
<212> ADN
<213> Đoạn mỗi
<220> Nhân tạo
<223>

<220>
<400> 657
CTGTACCCCC GATCTTGACG

20

<210> 658

<211> 20
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 658
 TGCTGCCAAC AAGTTTGTGC

20

<210> 659
 <211> 20
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 659
 CTGCTGGCTT GGCTGATACT

20

<210> 660
 <211> 20
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 660
 AGCTCAGTCA GGTCTCTCTGC

20

<210> 661
 <211> 20
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 661
 ACATCGAGAA CCAGCAGCAT

20

<210> 662
 <211> 19
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 662
 AGGTTTCTCA GCTCCCCCA

19

<210> 663
 <211> 21
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 663

CACCATCGTG AAGAGTGGTC T

21

<210> 664

<211> 18

<212> ADN

<213> Đoạn mỗi

<220> Nhân tạo

<223>

<220>

<400> 664

GCGACTGAGC CGAGGAGT

18

<210> 665

<211> 20

<212> ADN

<213> Đoạn mỗi

<220> Nhân tạo

<223>

<220>

<400> 665

GGAACCCTCT GCACTGTTCT

20

<210> 666

<211> 20

<212> ADN

<213> Đoạn mỗi

<220> Nhân tạo

<223>

<220>

<400> 666

AGCTTCTGTT CAGGCACTCC

20

<210> 667

<211> 20

<212> ADN

<213> Đoạn mỗi

<220> Nhân tạo

<223>

<220>

<400> 667

GAAGGTCCAG GTGCCCAATG

20

<210> 668

<211> 20

<212> ADN

<213> Đoạn mỗi

<220> Nhân tạo

<223>

<220>

<400> 668

CAGCTCTGTG AGGATCCAGC

20

<210> 669

<211> 20

<212> ADN

<213> Đoạn mỗi

<220> Nhân tạo

<223>

<220>	
<400> 669	
AAATCTGGGG GTCTGCCATC	20
<210> 670	
<211> 20	
<212> ADN	
<213> Đoạn mỗi	
<220> Nhân tạo	
<223>	
<220>	
<400> 670	
GCACGTCACG TTCTTCCTGT	20
<210> 671	
<211> 20	
<212> ADN	
<213> Đoạn mỗi	
<220> Nhân tạo	
<223>	
<220>	
<400> 671	
TCAATGGCTC CGTGCTCAAT	20
<210> 672	
<211> 20	
<212> ADN	
<213> Đoạn mỗi	
<220> Nhân tạo	
<223>	
<220>	
<400> 672	
TCGGGGTTGA ACTCGAAAGG	20
<210> 673	
<211> 20	
<212> ADN	
<213> Đoạn mỗi	
<220> Nhân tạo	
<223>	
<220>	
<400> 673	
GTCTTTGCCT GCCTCTTCCT	20
<210> 674	
<211> 21	
<212> ADN	
<213> Đoạn mỗi	
<220> Nhân tạo	
<223>	
<220>	
<400> 674	
CCCAAACCTG TTTCTCCGTC C	21
<210> 675	
<211> 20	
<212> ADN	

<213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 675
 ATGACATTGG GTGGCAGCTC

20

<210> 676
 <211> 20
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 676
 TGGGTTCTCG ATGATGTGGC

20

<210> 677
 <211> 20
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 677
 ACATCGTGCT CAAGTGGGAG

20

<210> 678
 <211> 20
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 678
 ACCAGCATCT TGTCTGCTC

20

<210> 679
 <211> 19
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 679
 CACATCGTGC GCTTCTTCG

19

<210> 680
 <211> 20
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 680
 CCCGTGCCGC ATATACTCAA

20

<210> 681
 <211> 20
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 681
 ATGGACCTGA TGCCAAGCTG

20

<210> 682
 <211> 20
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 682
 CCCACTAGAC AGTTGCGTGT

20

<210> 683
 <211> 20
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 683
 ACCGTAAGTT CACCACCGAG

20

<210> 684
 <211> 20
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 684
 CTGCTTGCCG TAGGTGAAGA

20

<210> 685
 <211> 20
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 685
 CACCAGAGGT CTACGCCATC

20

<210> 686
 <211> 20
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>

<400> 686
CCCACATGCT GAGGGTGAAT

20

<210> 687
<211> 20
<212> ADN
<213> Đoạn mỗi
<220> Nhân tạo
<223>

<220>
<400> 687
CGCTTGCGCA CTTTAGCCAG

20

<210> 688
<211> 21
<212> ADN
<213> Đoạn mỗi
<220> Nhân tạo
<223>

<220>
<400> 688
TGAGGTAGAG TATGGGGTCC G

21

<210> 689
<211> 21
<212> ADN
<213> Đoạn mỗi
<220> Nhân tạo
<223>

<220>
<400> 689
CTCAAGGCAG AACTTACCG A

21

<210> 690
<211> 22
<212> ADN
<213> Đoạn mỗi
<220> Nhân tạo
<223>

<220>
<400> 690
TGTAGGACCA CAGAGATGGT CT

22

<210> 691
<211> 20
<212> ADN
<213> Đoạn mỗi
<220> Nhân tạo
<223>

<220>
<400> 691
TGGGAATGAT GGCTCAGCTT

20

<210> 692
<211> 20
<212> ADN
<213> Đoạn mỗi
<220> Nhân tạo

<223>

<220>

<400> 692

CATAGTCTTG CTCCCCAGCG

20

<210> 693

<211> 19

<212> ADN

<213> Đoạn mỗi

<220> Nhân tạo

<223>

<220>

<400> 693

GTTCGCATGT TTGGGCCTG

19

<210> 694

<211> 20

<212> ADN

<213> Đoạn mỗi

<220> Nhân tạo

<223>

<220>

<400> 694

GGGCTGGCTG CAAACAATTT

20

<210> 695

<211> 20

<212> ADN

<213> Đoạn mỗi

<220> Nhân tạo

<223>

<220>

<400> 695

CTCCCCCTTT GGCTTGAGTT

20

<210> 696

<211> 16

<212> ADN

<213> Đoạn mỗi

<220> Nhân tạo

<223>

<220>

<400> 696

CTGGGACGGG GCCTCT

16

<210> 697

<211> 20

<212> ADN

<213> Đoạn mỗi

<220> Nhân tạo

<223>

<220>

<400> 697

CCAGTTCCGT GTGAAGGAGG

20

<210> 698

<211> 20

<212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 698
 ATCCGGCTGA AGAAGAGAGC

20

<210> 699
 <211> 20
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 699
 CTATGCAGCT GCTACCCTCC

20

<210> 700
 <211> 20
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 700
 CTTGGAGGTG CTGCCTATGG

20

<210> 701
 <211> 21
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 701
 AGTCTCCCAA AGGGAAGAGG A

21

<210> 702
 <211> 20
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 702
 TCTTGGGAAC AGATGGCGAC

20

<210> 703
 <211> 20
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 703
 CCTTCAGGGT GTGGTAGTGG

20

<210> 704
 <211> 20
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 704
 CTAGGGCCTT ATCTCGCAGC

20

<210> 705
 <211> 20
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 705
 TTTGCCAACA CCCCCAAGTA

20

<210> 706
 <211> 20
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 706
 GTCCAGCACC CACTCCTTAC

20

<210> 707
 <211> 20
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 707
 CAGGTTCCAG CAGTGAGGAG

20

<210> 708
 <211> 20
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 708
 CTTCTGAGGA AGCTTGGGGG

20

<210> 709
 <211> 20
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>	
<400> 709	
AGACCAAAAC CAAGCCCCT	20
<210> 710	
<211> 20	
<212> ADN	
<213> Đoạn mỗi	
<220> Nhân tạo	
<223>	
<220>	
<400> 710	
CTTCCTGGAG CACTGGTGAG	20
<210> 711	
<211> 20	
<212> ADN	
<213> Đoạn mỗi	
<220> Nhân tạo	
<223>	
<220>	
<400> 711	
GGCAGAGCAG AAGGAACACA	20
<210> 712	
<211> 20	
<212> ADN	
<213> Đoạn mỗi	
<220> Nhân tạo	
<223>	
<220>	
<400> 712	
TGTCCGTGTT CTCATCCGTG	20
<210> 713	
<211> 20	
<212> ADN	
<213> Đoạn mỗi	
<220> Nhân tạo	
<223>	
<220>	
<400> 713	
TCTTCCAGGG CAAGCACTTC	20
<210> 714	
<211> 20	
<212> ADN	
<213> Đoạn mỗi	
<220> Nhân tạo	
<223>	
<220>	
<400> 714	
TATCGGATGA GTTTCGCCG	20
<210> 715	
<211> 20	
<212> ADN	
<213> Đoạn mỗi	

<220> Nhân tạo
<223>

<220>
<400> 715
TCGTCCCCGA TGGATCTACA

20

<210> 716
<211> 20
<212> ADN
<213> Đoạn mỗi
<220> Nhân tạo
<223>

<220>
<400> 716
GCACATACTT CAGGCTTGCG

20

<210> 717
<211> 20
<212> ADN
<213> Đoạn mỗi
<220> Nhân tạo
<223>

<220>
<400> 717
GTTGAAGGAC CAGCTCTCCC

20

<210> 718
<211> 20
<212> ADN
<213> Đoạn mỗi
<220> Nhân tạo
<223>

<220>
<400> 718
CTTGTAAGTCG GCACCACCAT

20

<210> 719
<211> 20
<212> ADN
<213> Đoạn mỗi
<220> Nhân tạo
<223>

<220>
<400> 719
CATGTCAGGC TGAGGGCTAC

20

<210> 720
<211> 20
<212> ADN
<213> Đoạn mỗi
<220> Nhân tạo
<223>

<220>
<400> 720
GGAATTGGTG GTGGTGGTCT

20

<210> 721

<211> 20
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 721
 TACCTCTGGC ACATCCTCCA

20

<210> 722
 <211> 20
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 722
 GTCAGTGCTA CACCAAGGCA

20

<210> 723
 <211> 20
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 723
 GGCCGAAGTC ATCTGGACAA

20

<210> 724
 <211> 20
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 724
 TCATGCTAGG TGCTAGGGGA

20

<210> 725
 <211> 20
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 725
 AGACCACCAC CACCAATTCC

20

<210> 726
 <211> 20
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 726

TCATGCTAGG TGCTAGGGGA

20

<210> 727

<211> 24

<212> ADN

<213> Đoạn mỗi

<220> Nhân tạo

<223>

<220>

<400> 727

AAGAGAGATT CTCCAAAATG TGGC

24

<210> 728

<211> 22

<212> ADN

<213> Đoạn mỗi

<220> Nhân tạo

<223>

<220>

<400> 728

ATTCAGATGG CCTAGCTTTC TC

22

<210> 729

<211> 20

<212> ADN

<213> Đoạn mỗi

<220> Nhân tạo

<223>

<220>

<400> 729

CAAAAGAAGT GCGAGCAGCA

20

<210> 730

<211> 22

<212> ADN

<213> Đoạn mỗi

<220> Nhân tạo

<223>

<220>

<400> 730

GCACCTTCTG GAGAATACTG GA

22

<210> 731

<211> 23

<212> ADN

<213> Đoạn mỗi

<220> Nhân tạo

<223>

<220>

<400> 731

ACTGTGAATG CTCCTTGTT TCC

23

<210> 732

<211> 20

<212> ADN

<213> Đoạn mỗi

<220> Nhân tạo

<223>

<220> <400> 732 TGCTCGGACC ATTCTGTCTC	20
<210> 733 <211> 20 <212> ADN <213> Đoạn mỗi <220> Nhân tạo <223>	
<220> <400> 733 TCCAGGGAGG TTCCAAGACA	20
<210> 734 <211> 20 <212> ADN <213> Đoạn mỗi <220> Nhân tạo <223>	
<220> <400> 734 CTGGATCTGG CACGATGAGG	20
<210> 735 <211> 20 <212> ADN <213> Đoạn mỗi <220> Nhân tạo <223>	
<220> <400> 735 GCCCTTCGAG GTGGACTAAA	20
<210> 736 <211> 21 <212> ADN <213> Đoạn mỗi <220> Nhân tạo <223>	
<220> <400> 736 CTCGCACATA CTGTCGAGTC T	21
<210> 737 <211> 24 <212> ADN <213> Đoạn mỗi <220> Nhân tạo <223>	
<220> <400> 737 TCTTCCTCAT TCACATAGCC ATCA	24
<210> 738 <211> 22 <212> ADN	

<213> Đoạn mỗi
<220> Nhân tạo
<223>

<220>
<400> 738
CCTTGGGGAT ATCAAAGGAT GC

22

<210> 739
<211> 21
<212> ADN
<213> Đoạn mỗi
<220> Nhân tạo
<223>

<220>
<400> 739
CAGTGCAGCC TTGAACTGTT T

21

<210> 740
<211> 20
<212> ADN
<213> Đoạn mỗi
<220> Nhân tạo
<223>

<220>
<400> 740
TGATCCCGTT TCTGGTGTGT

20

<210> 741
<211> 23
<212> ADN
<213> Đoạn mỗi
<220> Nhân tạo
<223>

<220>
<400> 741
GCCCAGCAGT AAATTATATG CCA

23

<210> 742
<211> 20
<212> ADN
<213> Đoạn mỗi
<220> Nhân tạo
<223>

<220>
<400> 742
AACTGGCAAC CTACAACCGT

20

<210> 743
<211> 21
<212> ADN
<213> Đoạn mỗi
<220> Nhân tạo
<223>

<220>
<400> 743
TGGCTCAATG CTTTCATCTGG A

21

<210> 744
 <211> 20
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 744
 AACTCCATGA GGGTGGCAAG

20

<210> 745
 <211> 20
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 745
 GCCTGCAGAC TCTATGCAAC

20

<210> 746
 <211> 21
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 746
 AGCTGGGTCA CTAACAACTC A

21

<210> 747
 <211> 20
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 747
 GTGTCCCAGA GACCGAGTTG

20

<210> 748
 <211> 20
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 748
 GCCGAGTATC AAGAGCGACA

20

<210> 749
 <211> 20
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>

<400> 749 ATGCTGCTGG ATTGGACCTC	20
<210> 750 <211> 20 <212> ADN <213> Đoạn mỗi <220> Nhân tạo <223>	
<220> <400> 750 AGGCATTTC CTCACAGGGG	20
<210> 751 <211> 20 <212> ADN <213> Đoạn mỗi <220> Nhân tạo <223>	
<220> <400> 751 ATTTTCGGTGA ACGGTGTGGG	20
<210> 752 <211> 21 <212> ADN <213> Đoạn mỗi <220> Nhân tạo <223>	
<220> <400> 752 TAACAGCAAC AGCTGTGAGG A	21
<210> 753 <211> 20 <212> ADN <213> Đoạn mỗi <220> Nhân tạo <223>	
<220> <400> 753 ATTTTCGGTGA ACGGTGTGGG	20
<210> 754 <211> 21 <212> ADN <213> Đoạn mỗi <220> Nhân tạo <223>	
<220> <400> 754 TAACAGCAAC AGCTGTGAGG A	21
<210> 755 <211> 20 <212> ADN <213> Đoạn mỗi <220> Nhân tạo	

<223>

<220>

<400> 755

AGTCGGTCCT CTTTCCAGTG

21

<210> 756

<211> 21

<212> ADN

<213> Đoạn mỗi

<220> Nhân tạo

<223>

<220>

<400> 756

ACTTTTTACC TTCGTGCACC T

21

<210> 757

<211> 20

<212> ADN

<213> Đoạn mỗi

<220> Nhân tạo

<223>

<220>

<400> 757

TTGATTCTTG CAGCCACCCT

20

<210> 758

<211> 20

<212> ADN

<213> Đoạn mỗi

<220> Nhân tạo

<223>

<220>

<400> 758

GAACGCAGCG AACCGAATAC

20

<210> 759

<211> 20

<212> ADN

<213> Đoạn mỗi

<220> Nhân tạo

<223>

<220>

<400> 759

AGGCAATCAC GGAGGTGAAG

20

<210> 760

<211> 20

<212> ADN

<213> Đoạn mỗi

<220> Nhân tạo

<223>

<220>

<400> 760

TCTGGTAGCA AAAGGGCCAG

20

<210> 761

<211> 20

<212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 761
 CTGAGTGTGA CCGAGAAGGT

20

<210> 762
 <211> 20
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 762
 TGTGGCTCTG TCTGTAGGGA

20

<210> 763
 <211> 20
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 763
 CAGCACCCAT GAGTTGTGAC

20

<210> 764
 <211> 21
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 764
 AGATAGTGGT TGAAGGCCTG G

21

<210> 765
 <211> 25
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 765
 TCATCCTAGA GGATATAGGT GATCT

25

<210> 766
 <211> 25
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 766
 ACCAATGACC ATATGACAAA ATGCT

25

<210> 767
 <211> 20
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 767
 TGCCTGAGTC TGCTACTTGC

20

<210> 768
 <211> 21
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 768
 GGGGAAATTC TGATGAGCTG C

21

<210> 769
 <211> 20
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 769
 CTTCCATGAG GCAGGCTGAA

20

<210> 770
 <211> 20
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 770
 TCCAATACGC CGCAACTCTT

20

<210> 771
 <211> 20
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 771
 CTGGCAAAAC TCCTGGCATC

20

<210> 772
 <211> 25
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>	
<400> 772	
TTTCAGATCT AGGCTGGAAT GTTCT	25
<210> 773	
<211> 20	
<212> ADN	
<213> Đoạn mỗi	
<220> Nhân tạo	
<223>	
<220>	
<400> 773	
CACTTAAGGG CAGTGGGGAA	20
<210> 774	
<211> 20	
<212> ADN	
<213> Đoạn mỗi	
<220> Nhân tạo	
<223>	
<220>	
<400> 774	
GCCTCTCTCA CCTCCTCCTT	20
<210> 775	
<211> 20	
<212> ADN	
<213> Đoạn mỗi	
<220> Nhân tạo	
<223>	
<220>	
<400> 775	
AGAACCACTG TAGCAGACGC	20
<210> 776	
<211> 20	
<212> ADN	
<213> Đoạn mỗi	
<220> Nhân tạo	
<223>	
<220>	
<400> 776	
ATGCCTTGAC CGTCTTCCTG	20
<210> 777	
<211> 20	
<212> ADN	
<213> Đoạn mỗi	
<220> Nhân tạo	
<223>	
<220>	
<400> 777	
GCTTTGCCGC CCTTATGATG	20
<210> 778	
<211> 20	
<212> ADN	
<213> Đoạn mỗi	

<220> Nhân tạo
<223>

<220>
<400> 778
CACGTGTACG TGAGTGGACA

20

<210> 779
<211> 20
<212> ADN
<213> Đoạn mỗi
<220> Nhân tạo
<223>

<220>
<400> 779
AACCGCCAGA TTTGAATCGC

20

<210> 780
<211> 20
<212> ADN
<213> Đoạn mỗi
<220> Nhân tạo
<223>

<220>
<400> 780
GATGCGGTGG TCCTTGAGAA

20

<210> 781
<211> 20
<212> ADN
<213> Đoạn mỗi
<220> Nhân tạo
<223>

<220>
<400> 781
CCACTGAGAA CGAGCCAGAC

20

<210> 782
<211> 19
<212> ADN
<213> Đoạn mỗi
<220> Nhân tạo
<223>

<220>
<400> 782
CATCTGGCTC CCAGCCTTC

19

<210> 783
<211> 20
<212> ADN
<213> Đoạn mỗi
<220> Nhân tạo
<223>

<220>
<400> 783
CGGTTGCGCT TTCCTTTCTG

20

<210> 784

<211> 20
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 784
 TGGCTCTTTC TCTGTCCAGT

20

<210> 785
 <211> 20
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 785
 AGTCCCTGGC TCCTCTACTG

20

<210> 786
 <211> 20
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 786
 CGTGTGGAGA ACGTGACAGA

20

<210> 787
 <211> 20
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 787
 TCCCTCTGCA GTCTTTCCT

20

<210> 788
 <211> 18
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 788
 CGGGCCTCAC GGTTTCAG

18

<210> 789
 <211> 20
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 789

GACCTACGCC GAATATGCCA

20

<210> 790

<211> 20

<212> ADN

<213> Đoạn mỗi

<220> Nhân tạo

<223>

<220>

<400> 790

GATTTTGCCC CGGGTTTCAG

20

<210> 791

<211> 20

<212> ADN

<213> Đoạn mỗi

<220> Nhân tạo

<223>

<220>

<400> 791

GGGGCAATCC CGTACTGAAG

20

<210> 792

<211> 20

<212> ADN

<213> Đoạn mỗi

<220> Nhân tạo

<223>

<220>

<400> 792

CTCTGGCCCA GCACATAGTC

20

<210> 793

<211> 20

<212> ADN

<213> Đoạn mỗi

<220> Nhân tạo

<223>

<220>

<400> 793

CCCAGACTAC ATCCATGGGC

20

<210> 794

<211> 20

<212> ADN

<213> Đoạn mỗi

<220> Nhân tạo

<223>

<220>

<400> 794

ACATCCACCT GGACAAGCAG

20

<210> 795

<211> 19

<212> ADN

<213> Đoạn mỗi

<220> Nhân tạo

<223>

<220>
 <400> 795
 AAAGATCCCC AGCAGGCCC

19

<210> 796
 <211> 20
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 796
 TCCAGGCGAG GATCAATGTG

20

<210> 797
 <211> 19
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 797
 GGCGGTACCT GGAGACCTA

19

<210> 798
 <211> 21
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 798
 ACGAAGTCCT GCTCTAGCTT C

21

<210> 799
 <211> 21
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 799
 TGAAGTGAATG TCTGACCACC G

21

<210> 800
 <211> 20
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 800
 CCAAATGTGG TCAGGAGGGT

20

<210> 801
 <211> 20
 <212> ADN

<213> Đoạn mỗi
<220> Nhân tạo
<223>

<220>
<400> 801
CGGAGGCTGT TTGATGTCCT

20

<210> 802
<211> 20
<212> ADN
<213> Đoạn mỗi
<220> Nhân tạo
<223>

<220>
<400> 802
CCAGAATCCC TCCCCAGAGA

20

<210> 803
<211> 20
<212> ADN
<213> Đoạn mỗi
<220> Nhân tạo
<223>

<220>
<400> 803
AGGTATTCCC TGAGTCCCCC

20

<210> 804
<211> 20
<212> ADN
<213> Đoạn mỗi
<220> Nhân tạo
<223>

<220>
<400> 804
TTGGATCTTT GTCCTCGGGC

20

<210> 805
<211> 20
<212> ADN
<213> Đoạn mỗi
<220> Nhân tạo
<223>

<220>
<400> 805
CAGTAAGACC GAGGACCAGC

20

<210> 806
<211> 18
<212> ADN
<213> Đoạn mỗi
<220> Nhân tạo
<223>

<220>
<400> 806
GCCCGGTAGA CCTGCTTG

18

<210> 807
 <211> 23
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 807
 GTGCTGGAGA ATCTGGTAAA AGC

23

<210> 808
 <211> 25
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 808
 CCATTAAACC CATTACATG CAGGA

25

<210> 809
 <211> 20
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 809
 CCCCAGAGAAC CAGTTCAGAG

20

<210> 810
 <211> 21
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 810
 GGGAGGGAAG TCAAAGTCAG G

21

<210> 811
 <211> 20
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 811
 CAAGGCTCTG TGGGAGGATG

20

<210> 812
 <211> 20
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>

<400> 812
ACTGGGCACA GTCAATCAGC

20

<210> 813
<211> 17
<212> ADN
<213> Đoạn mỗi
<220> Nhân tạo
<223>

<220>
<400> 813
CTGCTTCGCT GCCGTGT

17

<210> 814
<211> 22
<212> ADN
<213> Đoạn mỗi
<220> Nhân tạo
<223>

<220>
<400> 814
GGAAGTTGAC TTTGTCCACC TG

22

<210> 815
<211> 20
<212> ADN
<213> Đoạn mỗi
<220> Nhân tạo
<223>

<220>
<400> 815
ACTGCCATCA TCTTCGTGGT

20

<210> 796
<211> 20
<212> ADN
<213> Đoạn mỗi
<220> Nhân tạo
<223>

<220>
<400> 816
GGTTGGTCTG GTTGTCTCC

20

<210> 817
<211> 20
<212> ADN
<213> Đoạn mỗi
<220> Nhân tạo
<223>

<220>
<400> 817
CGCACCATCT CTGTGATCCT

20

<210> 818
<211> 20
<212> ADN
<213> Đoạn mỗi
<220> Nhân tạo

<223>

<220>

<400> 818

TCGATTTCCT AGCAAGGACT

20

<210> 819

<211> 20

<212> ADN

<213> Đoạn mỗi

<220> Nhân tạo

<223>

<220>

<400> 819

GCGTCACTAC TGCTACCCTC

20

<210> 820

<211> 20

<212> ADN

<213> Đoạn mỗi

<220> Nhân tạo

<223>

<220>

<400> 820

ATGATGTCAC GGCAGTCGTT

20

<210> 821

<211> 17

<212> ADN

<213> Đoạn mỗi

<220> Nhân tạo

<223>

<220>

<400> 821

CCGTGCTAGC CGTTGGG

17

<210> 822

<211> 20

<212> ADN

<213> Đoạn mỗi

<220> Nhân tạo

<223>

<220>

<400> 822

CGCTGGTTCC CTTTCCTTGT

20

<210> 823

<211> 20

<212> ADN

<213> Đoạn mỗi

<220> Nhân tạo

<223>

<220>

<400> 823

CTAGGTGACA GCGTACCACG

20

<210> 824

<211> 20

<212> ADN	
<213> Đoạn mỗi	
<220> Nhân tạo	
<223>	
<220>	
<400> 824	
CTTGTGGAGC CAGCCTTCTT	20
<210> 825	
<211> 20	
<212> ADN	
<213> Đoạn mỗi	
<220> Nhân tạo	
<223>	
<220>	
<400> 825	
ACTTCCTGCT GAAGAGCGAC	20
<210> 826	
<211> 20	
<212> ADN	
<213> Đoạn mỗi	
<220> Nhân tạo	
<223>	
<220>	
<400> 826	
GGGTAGAGTC TGATCAGGGG	20
<210> 827	
<211> 20	
<212> ADN	
<213> Đoạn mỗi	
<220> Nhân tạo	
<223>	
<220>	
<400> 827	
ATGCCAGCTG ATGAAGACCG	20
<210> 828	
<211> 20	
<212> ADN	
<213> Đoạn mỗi	
<220> Nhân tạo	
<223>	
<220>	
<400> 828	
ACGTGGAAGG TCCTCTCGAT	20
<210> 829	
<211> 20	
<212> ADN	
<213> Đoạn mỗi	
<220> Nhân tạo	
<223>	
<220>	
<400> 829	
CGAGGACCCC ATGGACTACA	20

<210> 830
 <211> 20
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 830
 CGCCACTTCC ATCTCCTCAG

20

<210> 831
 <211> 20
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 831
 TGGCAAAGTC ATCCTGGTGC

20

<210> 832
 <211> 21
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 832
 GCAATGATGA CTCCTTCCG C

21

<210> 833
 <211> 20
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 833
 CTGTTCTTCC ACCTGTCCCG

20

<210> 834
 <211> 20
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 834
 AGCCGAGACA ATCTCTGCAC

20

<210> 835
 <211> 20
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>		
<400> 835		
CTGTTCTTCC ACCTGTCCCG		20
<210> 836		
<211> 20		
<212> ADN		
<213> Đoạn mỗi		
<220> Nhân tạo		
<223>		
<220>		
<400> 836		
AGCCGAGACA ATCTCTGCAC		20
<210> 837		
<211> 20		
<212> ADN		
<213> Đoạn mỗi		
<220> Nhân tạo		
<223>		
<220>		
<400> 837		
TCACTGACTT TGGCCTCTGC		20
<210> 838		
<211> 20		
<212> ADN		
<213> Đoạn mỗi		
<220> Nhân tạo		
<223>		
<220>		
<400> 838		
GGGGTCCCAC AGAAGGTTTT		20
<210> 839		
<211> 20		
<212> ADN		
<213> Đoạn mỗi		
<220> Nhân tạo		
<223>		
<220>		
<400> 839		
GCCTCTTCGA GTCATCCTC		20
<210> 840		
<211> 20		
<212> ADN		
<213> Đoạn mỗi		
<220> Nhân tạo		
<223>		
<220>		
<400> 840		
TCCTTCTTAA GCAGCCCAGC		20
<210> 841		
<211> 19		
<212> ADN		
<213> Đoạn mỗi		

<220> Nhân tạo
<223>

<220>
<400> 841
CAGCGATGCC AAGGAGGTC

19

<210> 842
<211> 19
<212> ADN
<213> Đoạn mỗi
<220> Nhân tạo
<223>

<220>
<400> 842
CTTCTGGACC ACGTCCTGC

19

<210> 843
<211> 20
<212> ADN
<213> Đoạn mỗi
<220> Nhân tạo
<223>

<220>
<400> 843
GTTCCCGGAT TTTTGTGGGC

20

<210> 844
<211> 20
<212> ADN
<213> Đoạn mỗi
<220> Nhân tạo
<223>

<220>
<400> 844
TCCGCCTTCC CTTAACCCTA

20

<210> 845
<211> 20
<212> ADN
<213> Đoạn mỗi
<220> Nhân tạo
<223>

<220>
<400> 845
CTCCCGAGGG TGATTGCTTC

20

<210> 846
<211> 20
<212> ADN
<213> Đoạn mỗi
<220> Nhân tạo
<223>

<220>
<400> 846
TGCTGGAGGT AGAGTGGTGA

20

<210> 847

<211> 20
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

 <220>
 <400> 847
 TGCAAAGCTA CTCCTGAGC 20

 <210> 848
 <211> 20
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

 <220>
 <400> 848
 AGGGTCTGAG TCTCTGTGCT 20

 <210> 849
 <211> 20
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

 <220>
 <400> 849
 GCTGCTTGAG GAAGTATAAG AATGA 25

 <210> 850
 <211> 20
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

 <220>
 <400> 850
 CTGGTTTCTC CGGTCCCAAT 20

 <210> 851
 <211> 19
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

 <220>
 <400> 851
 GCACAGACAT GAAGCTGCG 19

 <210> 852
 <211> 20
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

 <220>
 <400> 852

GTGGGCAGGT AGGTGAGTTC

20

<210> 853

<211> 20

<212> ADN

<213> Đoạn mỗi

<220> Nhân tạo

<223>

<220>

<400> 853

ACAACCAAGT GAGGCAGGTC

20

<210> 854

<211> 20

<212> ADN

<213> Đoạn mỗi

<220> Nhân tạo

<223>

<220>

<400> 854

AGGGGTGGTA TTGTTTCAGCG

20

<210> 855

<211> 20

<212> ADN

<213> Đoạn mỗi

<220> Nhân tạo

<223>

<220>

<400> 855

CAGCTCTGCT ACCAGGACAC

20

<210> 856

<211> 20

<212> ADN

<213> Đoạn mỗi

<220> Nhân tạo

<223>

<220>

<400> 856

CGAGAGCGGT TGGTGTCTAT

20

<210> 857

<211> 18

<212> ADN

<213> Đoạn mỗi

<220> Nhân tạo

<223>

<220>

<400> 857

CACTGTCTGT GCCGGTGG

18

<210> 858

<211> 19

<212> ADN

<213> Đoạn mỗi

<220> Nhân tạo

<223>

<220>
 <400> 858
 AGGCAGTCAG AGTGCTTGG

19

<210> 859
 <211> 20
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 859
 GCCTCCACTT CAACCACAGT

20

<210> 860
 <211> 20
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 860
 GGGATTGGGC ATGGACTCAA

20

<210> 861
 <211> 20
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 861
 TCTACGGACG TGGGATCCTG

20

<210> 862
 <211> 20
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 862
 GCTGTGTTCC ATCCTCTGCT

20

<210> 863
 <211> 20
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 863
 GGGCCTCTGG GATGATGATG

20

<210> 864
 <211> 19
 <212> ADN

<213> Đoạn mỗi
<220> Nhân tạo
<223>

<220>
<400> 864
CAACCCCTC CAGGACTGA

19

<210> 865
<211> 20
<212> ADN
<213> Đoạn mỗi
<220> Nhân tạo
<223>

<220>
<400> 865
GGCCTCTGG GATGATGATG

20

<210> 866
<211> 19
<212> ADN
<213> Đoạn mỗi
<220> Nhân tạo
<223>

<220>
<400> 866
CAACCCCTC CAGGACTGA

19

<210> 867
<211> 20
<212> ADN
<213> Đoạn mỗi
<220> Nhân tạo
<223>

<220>
<400> 867
GGTCCACCA TCAGCCATAG

20

<210> 868
<211> 20
<212> ADN
<213> Đoạn mỗi
<220> Nhân tạo
<223>

<220>
<400> 868
TGGAAGATGA GGGGCTCTGT

20

<210> 869
<211> 20
<212> ADN
<213> Đoạn mỗi
<220> Nhân tạo
<223>

<220>
<400> 869
GTCACCGATG AGGGGTCTC

20

<210> 870
 <211> 20
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 870
 GCGCTGGGGT GTAGAAGTAG

20

<210> 871
 <211> 20
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 871
 GGAGCTGAAC TTGGCCATCA

20

<210> 872
 <211> 19
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 872
 CTCTCGGGAG CACAGTTCC

19

<210> 873
 <211> 20
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 873
 GTCTCCTGGA ACAGGGCATC

20

<210> 874
 <211> 20
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 874
 AAGTCGACGT TGGCCTCGTA

20

<210> 875
 <211> 20
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>

<400> 875
TCATGGTGGA CACGGACATC

20

<210> 876
<211> 20
<212> ADN
<213> Đoạn mỗi
<220> Nhân tạo
<223>

<220>
<400> 876
TCAGCCTCAG GCGTATTTTC

20

<210> 877
<211> 19
<212> ADN
<213> Đoạn mỗi
<220> Nhân tạo
<223>

<220>
<400> 877
CTGACGTGGT CAGTCACCC

19

<210> 878
<211> 20
<212> ADN
<213> Đoạn mỗi
<220> Nhân tạo
<223>

<220>
<400> 878
GATATCGAAG CTGAGCACGC

20

<210> 879
<211> 20
<212> ADN
<213> Đoạn mỗi
<220> Nhân tạo
<223>

<220>
<400> 879
TGTCATCCAG ATCTGCTCGC

20

<210> 880
<211> 20
<212> ADN
<213> Đoạn mỗi
<220> Nhân tạo
<223>

<220>
<400> 880
CCTTCTCGTA GCTCTGCACC

20

<210> 881
<211> 20
<212> ADN
<213> Đoạn mỗi
<220> Nhân tạo

<223>

<220>

<400> 881

CATCCGTATC ATGGACCCCG

20

<210> 882

<211> 20

<212> ADN

<213> Đoạn mỗi

<220> Nhân tạo

<223>

<220>

<400> 882

CCCAGAGAT GAGGTACGGA

20

<210> 883

<211> 20

<212> ADN

<213> Đoạn mỗi

<220> Nhân tạo

<223>

<220>

<400> 883

TTGTAACGGG AGTCGGGAGA

20

<210> 884

<211> 18

<212> ADN

<213> Đoạn mỗi

<220> Nhân tạo

<223>

<220>

<400> 884

CAGCGAGCTG TGGAGGAG

18

<210> 885

<211> 25

<212> ADN

<213> Đoạn mỗi

<220> Nhân tạo

<223>

<220>

<400> 885

GAGTTTGAAG AAACAGAAGA ACCTG

25

<210> 886

<211> 23

<212> ADN

<213> Đoạn mỗi

<220> Nhân tạo

<223>

<220>

<400> 886

TCTCCAAGT TAACCAAGCT CTC

23

<210> 887

<211> 20

<212> ADN	
<213> Đoạn mỗi	
<220> Nhân tạo	
<223>	
<220>	
<400> 887	
AGGAACTGTG GGGAAATCTGT	20
<210> 888	
<211> 20	
<212> ADN	
<213> Đoạn mỗi	
<220> Nhân tạo	
<223>	
<220>	
<400> 888	
AGCTCAGTAA AAGTGAACGA CA	22
<210> 889	
<211> 25	
<212> ADN	
<213> Đoạn mỗi	
<220> Nhân tạo	
<223>	
<220>	
<400> 889	
CCAGTACCAA AGTTGATAAT GCTAT	25
<210> 890	
<211> 22	
<212> ADN	
<213> Đoạn mỗi	
<220> Nhân tạo	
<223>	
<220>	
<400> 890	
TCCAATTTGC TGAAGAGTGC AA	22
<210> 891	
<211> 23	
<212> ADN	
<213> Đoạn mỗi	
<220> Nhân tạo	
<223>	
<220>	
<400> 891	
ACAAATGGAA GATGATCTGG TGA	23
<210> 892	
<211> 20	
<212> ADN	
<213> Đoạn mỗi	
<220> Nhân tạo	
<223>	
<220>	
<400> 892	
TGAGCAACAT GGGAGGTGAG	20

<210> 893
 <211> 24
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 893
 CATTAATGGT TCACCTCGAA CACC 24

<210> 894
 <211> 24
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 894
 TTCTAGTTGT TTTGCTATCC GTGC 24

<210> 895
 <211> 22
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 895
 TGAAACACAG AGAACACCAC GA 22

<210> 896
 <211> 20
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 896
 AACTGGAGTG TGTGGAGGAA 20

<210> 897
 <211> 25
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 897
 GACTGTTATG AACACTATCC AACAA 25

<210> 898
 <211> 25
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220> <400> 898 ATTTTCTGAA GGTGATCAC TTGCT	25
<210> 899 <211> 22 <212> ADN <213> Đoạn mỗi <220> Nhân tạo <223>	
<220> <400> 899 ACTGCACAGT GAATCCAAAA GA	22
<210> 900 <211> 20 <212> ADN <213> Đoạn mỗi <220> Nhân tạo <223>	
<220> <400> 900 TTTCGACACA ACCCTGTCCC	20
<210> 901 <211> 22 <212> ADN <213> Đoạn mỗi <220> Nhân tạo <223>	
<220> <400> 901 GCAGAAGGCA ACTTGACAAG AG	22
<210> 902 <211> 20 <212> ADN <213> Đoạn mỗi <220> Nhân tạo <223>	
<220> <400> 902 GAGCCATGCA AGGGATTCCA	20
<210> 903 <211> 19 <212> ADN <213> Đoạn mỗi <220> Nhân tạo <223>	
<220> <400> 903 GAGGCTTCCC CTCGTTTGG	19
<210> 904 <211> 20 <212> ADN <213> Đoạn mỗi	

<220> Nhân tạo
<223>

<220>
<400> 904
CAGAATGGGC GTCTCGCTTC

20

<210> 905
<211> 20
<212> ADN
<213> Đoạn mỗi
<220> Nhân tạo
<223>

<220>
<400> 905
CCCGTGAAGA TGTCCACTCC

20

<210> 906
<211> 20
<212> ADN
<213> Đoạn mỗi
<220> Nhân tạo
<223>

<220>
<400> 906
GCATCTGGTG CATGTTGTCC

20

<210> 907
<211> 20
<212> ADN
<213> Đoạn mỗi
<220> Nhân tạo
<223>

<220>
<400> 907
AGTCCATGCA TGAGAAGGGC

20

<210> 908
<211> 20
<212> ADN
<213> Đoạn mỗi
<220> Nhân tạo
<223>

<220>
<400> 908
GACCGCATCC CCATTCCTTT

20

<210> 909
<211> 20
<212> ADN
<213> Đoạn mỗi
<220> Nhân tạo
<223>

<220>
<400> 909
GCCTCTAGTC CAGTTCCAGC

20

<210> 910

<211> 20
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 910
 CCTGGCCAGC ATCTTGTAGG

20

<210> 911
 <211> 18
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 911
 AACATGCCTC CCCCAGGA

18

<210> 912
 <211> 19
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 912
 CTTTCAGGCCA GGGCTTGGG

19

<210> 913
 <211> 20
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 913
 GCACCCCTCA GAAGCTGATT

20

<210> 914
 <211> 20
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 914
 CTTCTGGATG GGGGTGATGC

20

<210> 915
 <211> 20
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 915

CGCACACCGA ATTCAGGAAC

20

<210> 916

<211> 20

<212> ADN

<213> Đoạn mỗi

<220> Nhân tạo

<223>

<220>

<400> 916

TGGAAGTTCA GCAGCCTGAG

20

<210> 917

<211> 20

<212> ADN

<213> Đoạn mỗi

<220> Nhân tạo

<223>

<220>

<400> 917

ACAGCCCTCA ATGCTAAGGC

20

<210> 918

<211> 20

<212> ADN

<213> Đoạn mỗi

<220> Nhân tạo

<223>

<220>

<400> 918

CTGCTGCTTC TCCAGCTTCT

20

<210> 919

<211> 20

<212> ADN

<213> Đoạn mỗi

<220> Nhân tạo

<223>

<220>

<400> 919

GCATTCTCCA GCATGCCAAG

20

<210> 920

<211> 20

<212> ADN

<213> Đoạn mỗi

<220> Nhân tạo

<223>

<220>

<400> 920

TCTTTCTTCT GCTCCCGCTC

20

<210> 921

<211> 20

<212> ADN

<213> Đoạn mỗi

<220> Nhân tạo

<223>

<220>		
<400> 921		
CGCAAGCTCA TCGACCAGAA		20
<210> 922		
<211> 20		
<212> ADN		
<213> Đoạn mỗi		
<220> Nhân tạo		
<223>		
<220>		
<400> 922		
CACCAGCTCC GTGAGGTTAG		20
<210> 923		
<211> 21		
<212> ADN		
<213> Đoạn mỗi		
<220> Nhân tạo		
<223>		
<220>		
<400> 923		
TGAGACTGGA CCTGAGGAGA C		21
<210> 924		
<211> 19		
<212> ADN		
<213> Đoạn mỗi		
<220> Nhân tạo		
<223>		
<220>		
<400> 924		
GGGTGTCGAA GAGTCGCAG		19
<210> 925		
<211> 20		
<212> ADN		
<213> Đoạn mỗi		
<220> Nhân tạo		
<223>		
<220>		
<400> 925		
GTTTGCTCTC CAAAGCTCGG		20
<210> 926		
<211> 20		
<212> ADN		
<213> Đoạn mỗi		
<220> Nhân tạo		
<223>		
<220>		
<400> 926		
CACTTGAGGA GTCTGTCGCA		20
<210> 927		
<211> 20		
<212> ADN		

<213> Đoạn mỗi
<220> Nhân tạo
<223>

<220>
<400> 927
AAGTGTGTGA AGAGGCTGGA

20

<210> 928
<211> 19
<212> ADN
<213> Đoạn mỗi
<220> Nhân tạo
<223>

<220>
<400> 928
CATCAACAGA GCCCGCCAA

19

<210> 929
<211> 20
<212> ADN
<213> Đoạn mỗi
<220> Nhân tạo
<223>

<220>
<400> 929
GCAGAACGCT TTGAGAGAAG

20

<210> 930
<211> 20
<212> ADN
<213> Đoạn mỗi
<220> Nhân tạo
<223>

<220>
<400> 930
GCCACGCAG AGAAATATCG

20

<210> 931
<211> 25
<212> ADN
<213> Đoạn mỗi
<220> Nhân tạo
<223>

<220>
<400> 931
GGAAGTTTAG CTGATGCTAT AAGTG

25

<210> 932
<211> 20
<212> ADN
<213> Đoạn mỗi
<220> Nhân tạo
<223>

<220>
<400> 932
CGGCCAACTT GCAAAAGGAG

20

<210> 933
 <211> 21
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 933
 TCGAACCTCA ATCCCAAATG C

21

<210> 934
 <211> 22
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 934
 ACATAACTTT GTTGATGCC CA

22

<210> 935
 <211> 23
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 935
 GTGATCTTGG GCATGTAACA AGG

23

<210> 936
 <211> 23
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 936
 TTTGCAAGAA AACGACTATC GCC

23

<210> 937
 <211> 20
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>
 <400> 937
 GCTTGCCCTC ACAGTGGTAT

20

<210> 938
 <211> 20
 <212> ADN
 <213> Đoạn mỗi
 <220> Nhân tạo
 <223>

<220>

<400> 938
GCACTTGTGG TATCCGAGGT

20

<210> 939
<211> 21
<212> ADN
<213> Đoạn mỗi
<220> Nhân tạo
<223>

<220>
<400> 939
TCCAGAGAGA AGACCTTCAG C

21

<210> 940
<211> 22
<212> ADN
<213> Đoạn mỗi
<220> Nhân tạo
<223>

<220>
<400> 940
ACTCTTTCTA GAAGCGGACA GC

22

<210> 941
<211> 20
<212> ADN
<213> Đoạn mỗi
<220> Nhân tạo
<223>

<220>
<400> 941
CTCCCCCTGA ACACTGTGAC

20

<210> 942
<211> 20
<212> ADN
<213> Đoạn mỗi
<220> Nhân tạo
<223>

<220>
<400> 942
TGCCTACAAA GTGCTCCCAG

20