

Tin Sinh học Bioinformatics

Chương 3. Các công cụ công nghệ sinh học phân tử

Bài thực hành phát hiện đột biến gen với công nghệ Sanger

TS. Nguyễn Hồng Quang
Khoa Kỹ thuật máy tính

Leader of Bioinformatics Group, BK.AI center
Trường Công nghệ thông tin và Truyền thông
Trường Đại học Bách Khoa Hà Nội

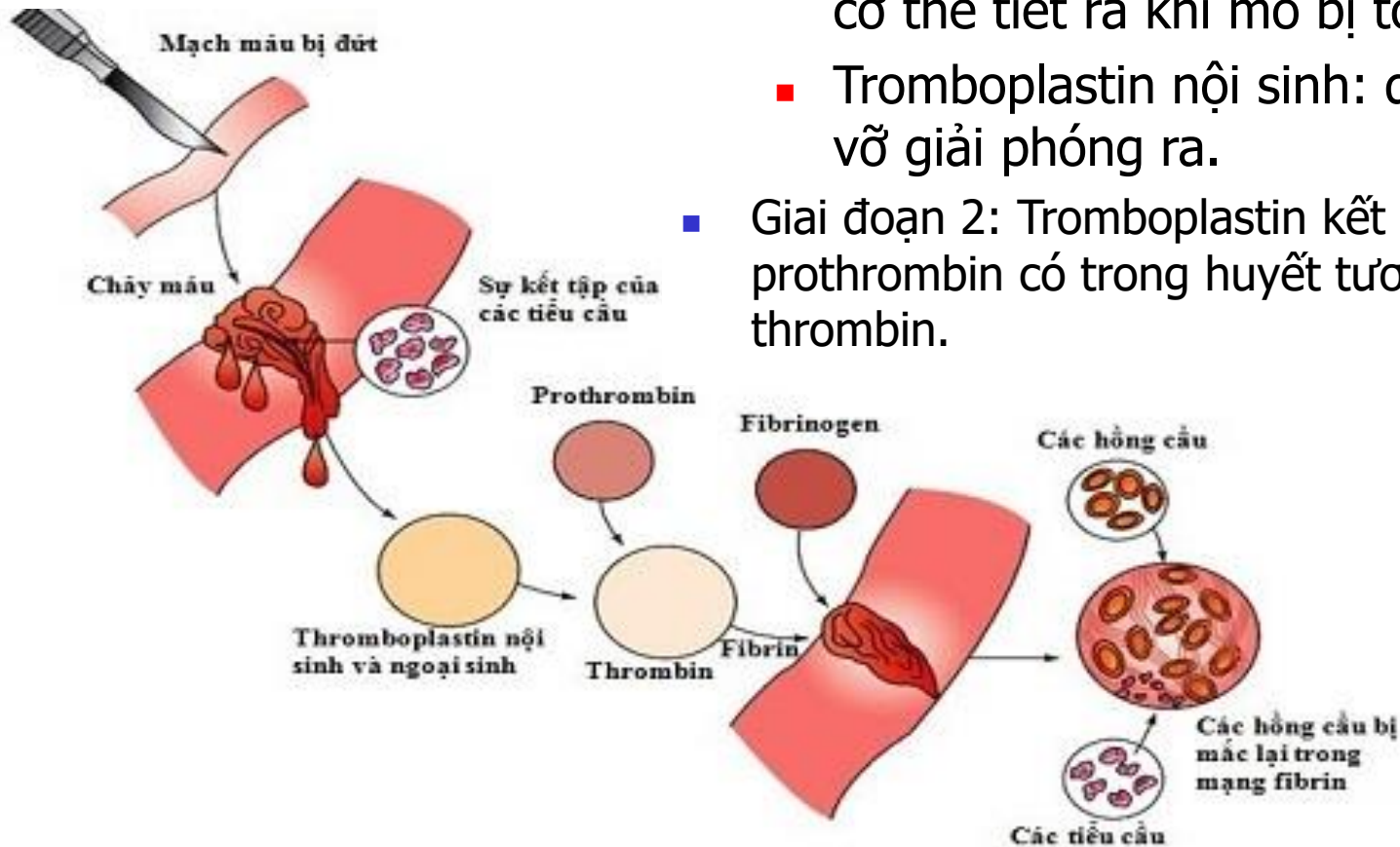


Tài liệu tham khảo

- Mạc Thanh Tùng, Nguyễn Thị Mai Ngọc, Đỗ Doãn Lợi, Kim Ngọc Thanh, Lê Thanh Tùng, Trương Thanh Hương, Đa hình gen COX-1 kháng Aspirin trên bệnh nhân có bệnh động mạch vành, TẠP CHÍ TIM MẠCH HỌC VIỆT NAM - số 94+95.2021

Tác nhân đông máu

- Giai đoạn 1. Sự hình thành và giải phóng tromboplastin
 - Tromboplastin ngoại sinh: do mô của cơ thể tiết ra khi mô bị tổn thương.
 - Tromboplastin nội sinh: do tiểu cầu bị vỡ giải phóng ra.
- Giai đoạn 2: Tromboplastin kết hợp với chất prothrombin có trong huyết tương thành thrombin.



<https://medlatec.vn/tin-tuc/xet-nghiem-dong-mau--vai-tro-trong-phat-hien-chan-doan-va-xu-tri-cac-roi-loan-dong-cam-mau-s159-n11152>



Cơ chế đông máu

- Giai đoạn 1. Sự hình thành và giải phóng tromboplastin
 - Tromboplastin ngoại sinh: do mô của cơ thể tiết ra khi mô bị tổn thương.
 - Tromboplastin nội sinh: do tiểu cầu bị vỡ giải phóng ra.
- Giai đoạn 2: Tromboplastin kết hợp với chất prothrombin có trong huyết tương thành thrombin.
- Giai đoạn 3: Dưới tác dụng của trompin, chất fibrinogen ở dạng hòa tan liên kết lại thành các sợi mảnh fibrin. Những sợi này kết thành mạng lưới, chằng giữa các tế bào máu (hồng cầu) tạo thành cục máu đông.
- Thời gian đông máu ở người trưởng thành là 3-4 phút.

<https://www.vinmec.com/vi/tin-tuc/thong-tin-suc-khoe/suc-khoe-tong-quat/xet-nghiem-prothrombin/>



Quá trình cầm máu của tiểu cầu

- Quá trình giúp đông máu và dừng chảy máu của tiểu cầu trải qua 3 giai đoạn:
- Kết dính tiểu cầu: Khi phát hiện có tổn thương làm lộ lớp collagen nằm bên dưới tế bào nội mạc mạch máu, tiểu cầu sẽ tập trung và đến dính vào lớp collagen này.
- Tiểu cầu giải phóng các yếu tố hoạt động: tiểu cầu tiếp tục được hoạt hóa sau khi kết dính với collagen, tế bào được phình to, thò chân giả và giải phóng các chất với lượng lớn ADP, Thromboxane A₂.

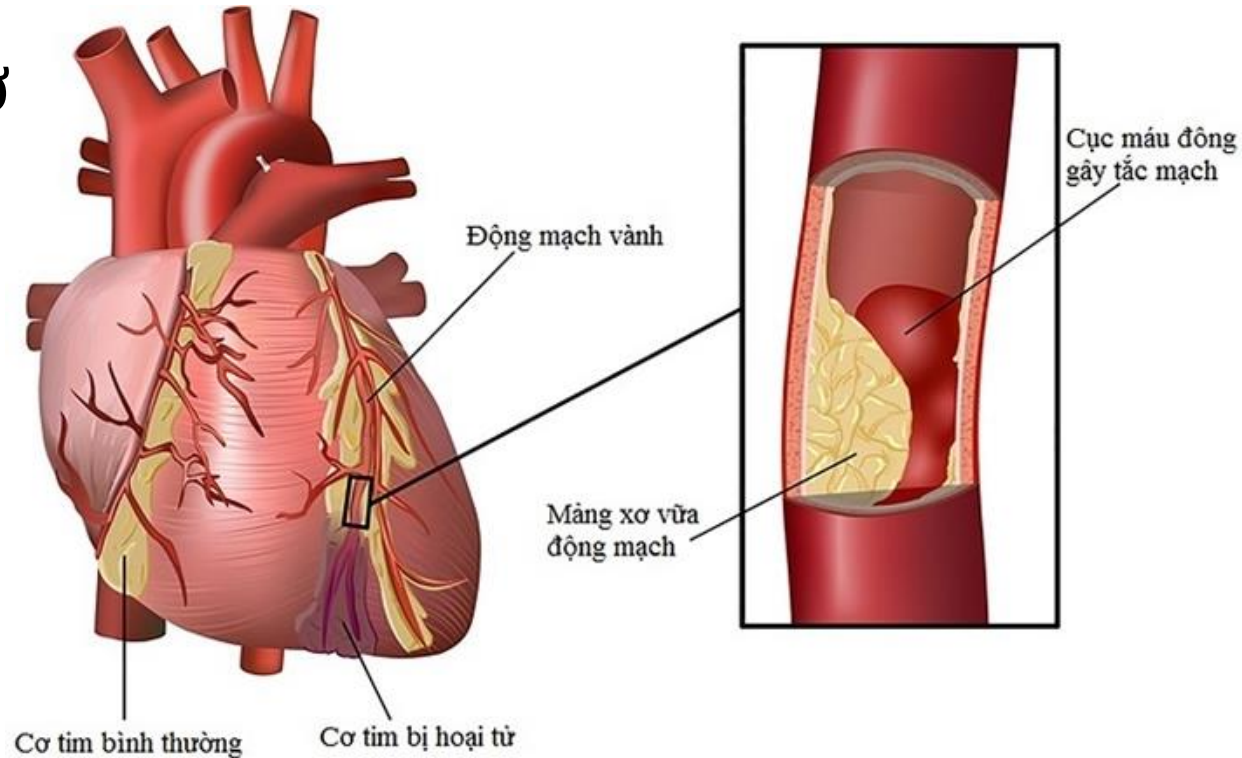


Ngưng tập tiểu cầu

- ADP và thromboxane A₂ hoạt hoá các tế bào tiểu cầu ở gần đó giúp chúng có khả năng dính vào lớp tiểu cầu ban đầu (ngưng tập tiểu cầu), quá trình này diễn ra liên tiếp, liên tục kết dính các lớp tiểu cầu đến gần tiếp theo đó để tạo nên các nút tiểu cầu.
- Đây chính là quá trình hình thành cục máu đông để cầm máu khi có vết thương.

Bệnh động mạch vành

- Giống như các cơ quan khác, tim cũng cần oxy và chất dinh dưỡng để hoạt động.
- Các động mạch vành chính là hệ thống thực hiện nhiệm vụ này.

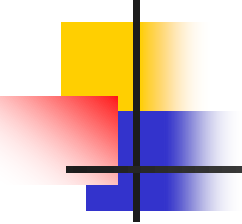


- Dẫn đến các cơn đau thắt ngực hay nặng hơn là nhồi máu cơ tim

Thuốc Aspirin



- Aspirin với tác dụng chống ngưng tập tiểu cầu (NTTC) có vai trò trung tâm trong điều trị cũng như dự phòng bệnh động mạch vành
- Tuy nhiên, các báo cáo gần đây về việc tiếp tục xuất hiện các biến cố thuyên tắc mạch mới trên bệnh nhân điều trị bằng aspirin đang làm dấy lên nhiều nghi ngờ về hiệu quả sử dụng của thuốc.
- Kháng aspirin được đưa ra như một lời giải thích hợp lý cho sự tái diễn các biến cố mới trên bệnh nhân đang điều trị.
- Tỷ lệ kháng aspirin dao động rất lớn từ 5-60% theo từng nghiên cứu

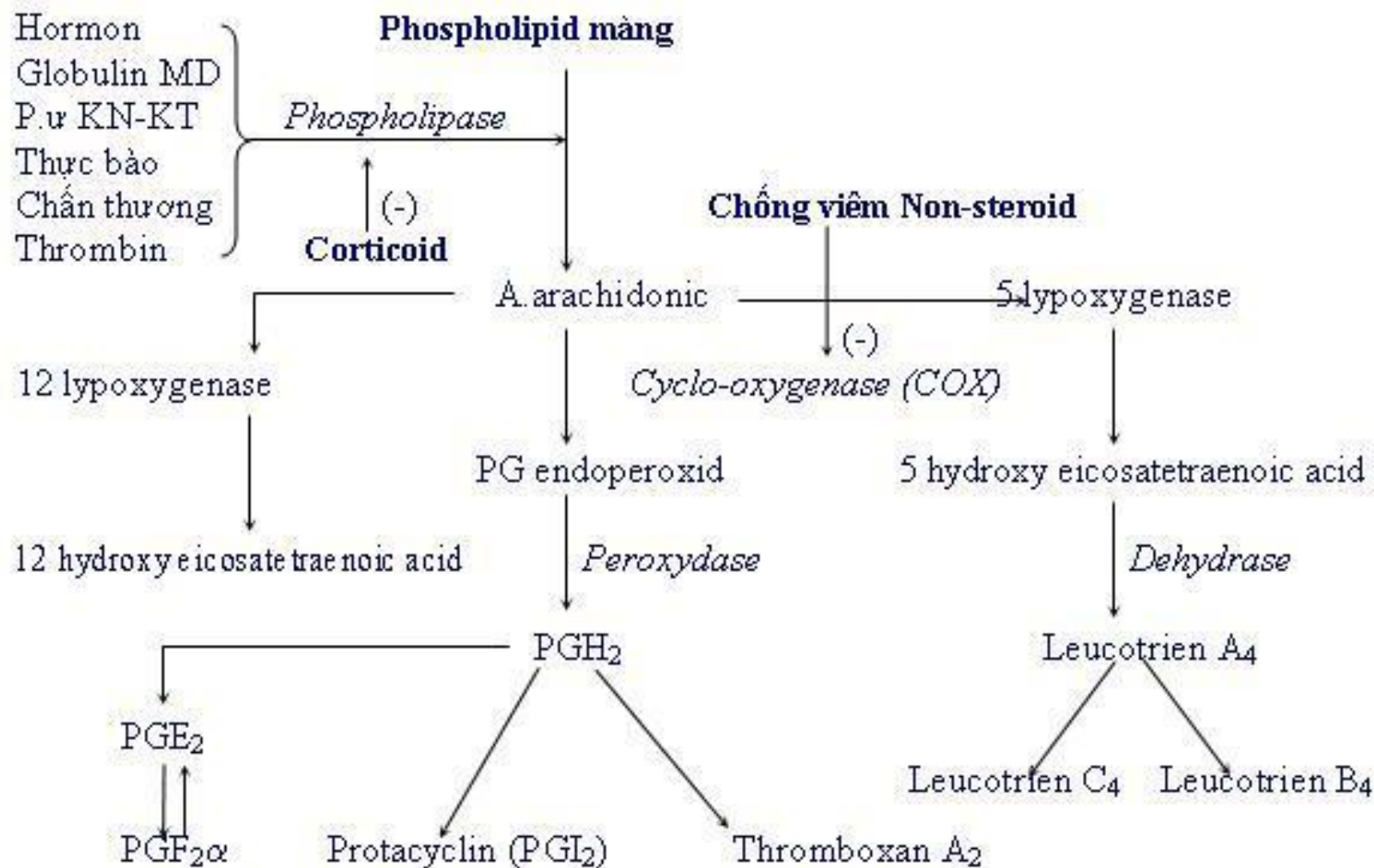
- 
- Có nhiều yếu tố làm giảm đáp ứng của aspirin đã được chứng minh như giới nữ, tuổi cao, các yếu tố nguy cơ của bệnh tim mạch do xơ vữa, một số đa hình gen chuyển hóa thuốc.
 - Trong đó, đa hình gen COX-1 đã chỉ ra khuynh hướng tiềm tàng cao tỷ lệ kháng aspirin trong nhiều nghiên cứu.



Gen PTGS-1

- Gen PTGS-1 được tìm thấy trên nhánh dài nhiễm sắc thể số 9 tại vị trí 9q32-q33,3, phát hiện bởi Yokoyama và Tanabe năm 1989
- Gen có chiều dài 22 kb bao gồm 11 exon
- Gen quy định tổng hợp enzym cyclooxygenase -1 (COX-1) có vai trò chính trong quá trình chuyển hóa axit arachidonic màng thành thromboxan A₂ qua đó gây kết tập tiểu cầu.

PG được sinh tổng hợp ngay tại màng tế bào từ phospholipid



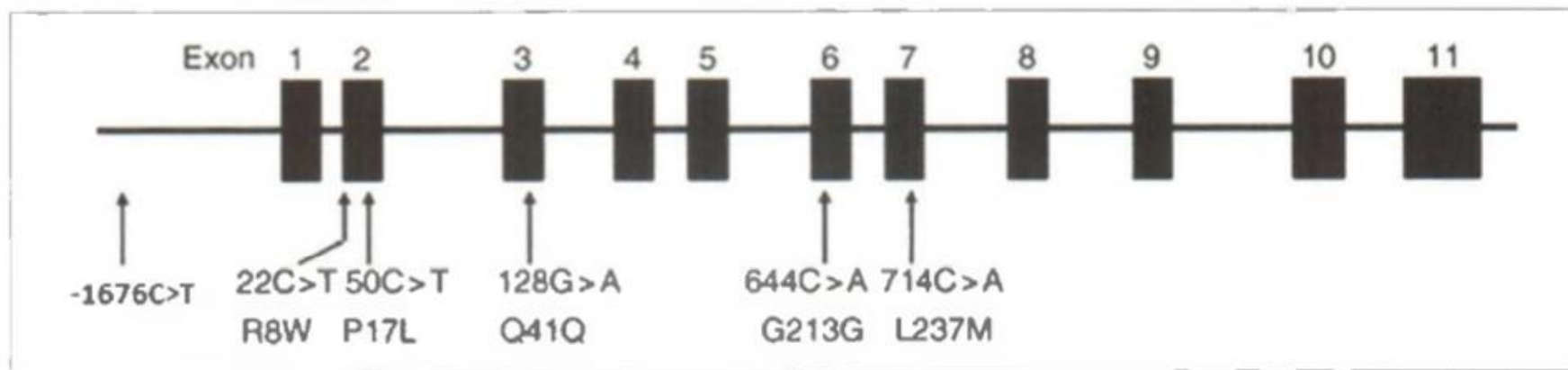


Vai trò của Aspirin

- Aspirin bất hoạt không khôi phục enzyme COX - 1 bằng cách gắn vào gốc Ser ở vị trí 529, ngăn chặn quá trình chuyển đổi acid arachidonic thành prostaglandin H₂ (PGH₂) rồi tạo thành thromboxan A₂, dẫn đến ức chế quá trình ngưng tập tiểu cầu.
- Vì tiểu cầu không có nhân, không thể tổng hợp thêm COX-1 mới nên sự ức chế này không thể đảo ngược, diễn ra trong suốt đời sống của tiểu cầu từ 7 đến 10 ngày

Vai trò của đột biến gen

- Tổng hợp từ các nghiên cứu, 5 đa hình của gen COX-1 gồm C22T (rs1236913), C50T (rs3842787), G128A (rs3842788), C644A (rs5788), C-1676T (rs1330344) thấy rõ sự phổ biến và nổi trội về tính kháng aspirin





Tiêu chuẩn lựa chọn bệnh nhân

- Bệnh nhân có bệnh động mạch vành đã được chụp động mạch vành qua da hoặc chụp cắt lớp vi tính đang sử dụng aspirin liều thấp 70-100 mg/ngày ít nhất trên 2 tuần
- 54 bệnh nhân



Kỹ thuật sử dụng

- Mẫu máu được lấy 2mL.
- DNA được khuếch đại bằng phản ứng PCR với 5 cặp mồi được khuếch đại 5 đa hình.
- Các cặp mồi được thiết kế theo NCBI và bằng phần mềm Primer 3 dựa theo vị trí gen trên NCBI.
- Sản phẩm PCR được tinh sạch bằng kit GeneJET PCR.
- Giải trình tự gen bằng hệ thống máy giải trình tự ABI 3500.
- Kết quả giải trình tự gen phân tích dựa trên phần mềm BioEdit.

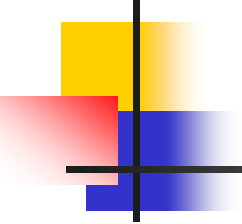
Tần suất phân bố một số đa hình gen COX-1

Kiểu gen		Kiểu gen	Nam (N=41)	Nữ (N=13)	Tổng cộng (N=54)
C-1676T	TT	BT	8 (14,8%)	2 (3,7%)	10 (18,5%)
	CT	DHT	24 (44,4%)	11 (20,4%)	35 (64,8%)
	CC	ĐHT	9 (16,7%)	0	9 (16,7%)

G128A	GG	BT	35 (64,8%)	11 (20,4%)	46 (85,2%)
	GA	DHT	4 (7,4%)	2 (3,7%)	6 (11,1%)
	AA	ĐHT	0	2 (3,7%)	2 (3,7%)
C644A	CC	BT	39 (72,2%)	13 (24,1%)	52 (96,3%)
	TA	DHT	2 (3,7%)	0	2 (3,7%)
C22T	AA	BT	13 (24,1%)	39 (72,2%)	52 (96,3%)
	CT	DHT	2 (3,7%)	0	2 (3,7%)
C50T	TT	BT	41 (75,9%)	13 (24,1%)	54 (100%)

$p > 0,05$

BT: Bình thường, DHT: Dị hợp tử, ĐHT: Đồng hợp tử

- 
- Đa hình C-1676T xuất hiện với tỷ lệ cao nhất 81,5% với 64,8% thể dị hợp tử và 16,7% thể đồng hợp tử.
 - Tiếp theo là đa hình G128C 14.8% trong đó đồng hợp tử là 3,7% và dị hợp tử là 11,1%.
 - Hai đa hình C644T và C22T đều xuất hiện với tỷ lệ 3,7% ở dạng dị hợp tử. Không tìm thấy đa hình C50T trong nhóm nghiên cứu.



Tần số alen trong các đa hình gen COX-1

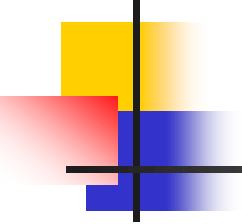
Đa hình	Tần số alen	
	Alen bình thường	Alen đa hình
C-1676T (rs 1330344)	C = 49,1%	T = 50,9%
G128A (rs3842788)	G = 90,7%	A = 9,3%
C644A (rs5788)	C = 98,1%	A = 1,9%
C22T (rs1236913)	C = 98,1%	T = 1,9%

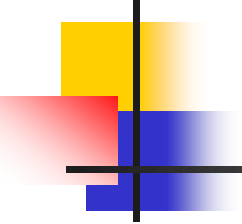
- Đa hình C-1676T có tần số alen đa hình cao hơn alen bình thường



Tỷ lệ tổ hợp các đa hình của gen COX-1

Tổ hợp đa hình		Tỷ lệ
Đồng hợp tử bình thường cả 5 đa hình khảo sát		12,9%
Mang 1 dị hợp tử đa hình	Dị hợp tử C-1676T	48,1%
	Dị hợp tử C22T	1,9%
	Dị hợp tử C644A	1,9%
	Dị hợp tử G128A	1,9%
Mang 2 dị hợp tử đa hình	Dị hợp tử C22T và C-1676T	1,9%
	Dị hợp tử C644T và C-1676T	1,9%
	Dị hợp tử G128A và C-1676T	9,2%
Mang đồng hợp tử đa hình	Đồng hợp tử G128A và dị hợp tử C-1676T	3,7%
	Đồng hợp tử C-1676T	16,6%
Tổng		100%

- 
- Số lượng đa hình đơn nucleotid (SNP) của gen COX-1 được cập nhật tại NCBI liên tục tăng và đạt 400 đa hình, tuy nhiên, chỉ có 72 trong số đó nằm trong vùng mã hóa của gen, còn lại nằm ở vùng 5' và 3' không mã hóa và vùng intron
 - Năm 2007, Lee và cộng sự đã xác định được 45 đa hình ở gen COX-1 khi nghiên cứu 92 người khỏe mạnh (24 người châu Phi, 24 người châu Á, 24 người châu Âu và 20 người ẩn danh)

- 
- Vẫn còn nhiều tranh cãi và chưa có một nghiên cứu nào khẳng định được vai trò chính xác của đa hình gen COX-1 nào liên quan tới sự kháng aspirin.
 - Nghiên cứu này tập trung vào 5 đa hình phổ biến sau khi tổng hợp từ nghiên cứu bao gồm C-1676T, G128A, C644T, C22T và C50T với tỷ lệ được tìm thấy lần lượt là 81,5%, 14,8%, 3,7% và 3,7%.
 - Đa hình C50T không quan sát thấy trong nghiên cứu.



Đa hình C-1676T

- Đa hình C-1676T được phát hiện trong nghiên cứu khá cao chiếm tới 81,5% với tỷ lệ 22,7% đồng hợp tử và 77,3% dị hợp tử
- Đa hình C-1676T xảy ra ở vùng điều hòa (promotor), vị trí gần một số yếu tố phiên mã (GATA-1, CdxA) của gen COX-1.
- Vì vậy, đa hình C-1676T ảnh hưởng tới mức độ biểu hiện của gen COX-1.
- Kết quả cho thấy bệnh nhân mang đa hình C-1676T làm giảm tác dụng của aspirin do làm tăng độ ngưng tập tiểu cầu



Đa hình G128A

- Đa hình G128A là đa hình xuất hiện phổ biến thứ 2 trong nghiên cứu này với 14,8%.
- Đa hình xảy ra tại exon thứ 3 của gen COX-1, đa hình không gây biến đổi axit amin (Q41Q).
- Câu trả lời cho tính kháng aspirin có thực sự liên quan đến đa hình G128A hay không thì vẫn còn nhiều câu trả lời trái chiều.



Đa hình C644A

- Đa hình C644A xảy ra ở exon thứ 6 gây biến đổi codon GGC \rightarrow GGA, kết quả không làm thay đổi axit amin (G213G).
- Là đa hình được tìm thấy phổ biến nhất trên quần thể người châu Phi với tỷ lệ 46% và ít gặp hơn ở người da trắng và người châu Á
- Trong nghiên cứu này, đa hình C633A tìm thấy với tỷ lệ 3,7% đều mang kiểu gen dị hợp tử. Kết quả nghiên cứu cũng không tìm thấy mối liên quan giữa bệnh nhân mang đa hình C644A với độ NTTC (ngưng tập tiểu cầu).



Đa hình C22T và C50T

- Đa hình C22T và C50T cùng xảy ra tại exon thứ 2 của gen COX-1.
- Hai đa hình này thay thế nucleotid C thành T ở vị trí 22 và 50 dẫn đến thay thế axit amin Arginine và Proline thành Tryptophan và Leucin
- Trong nghiên cứu này, tỷ lệ gặp hai loại đa hình này khá thấp với 3,7% đa hình C22T và không gặp đa hình C50T.
- Cũng không tìm thấy mối liên quan giữa đa hình C22T được tìm thấy với độ NTTC

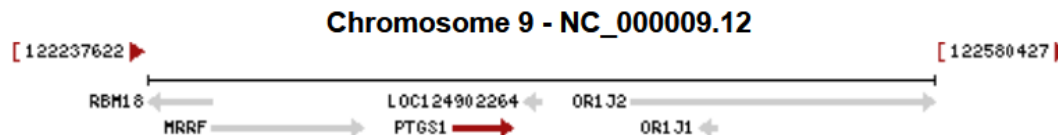
COX 1

Name/Gene ID	Description	Location	Aliases	MIM
<input type="checkbox"/> PTGS1 ID: 5742	prostaglandin- endoperoxide synthase 1 [<i>Homo sapiens</i> (human)]	Chromosome 9, NC_000009.12 (122370533..122395703)	COX1, COX3, PCOX1, PES-1, PGG/HS, PGHS-1, PGHS1, PHS1, PTGHS	176805
<input type="checkbox"/> MT-CO1 ID: 4512	mitochondrially encoded cytochrome c oxidase I [<i>Homo sapiens</i> (human)]	Chromosome MT, NC_012920.1 (5904..7445)	COI, MTCO1, COX1	
<input type="checkbox"/> COX1 ID: 6775083	cytochrome c oxidase subunit I [<i>Homo sapiens neanderthalensis</i> (Neandertal)]	Chromosome MT, NC_011137.1 (5899..7440)		
<input type="checkbox"/> COX1 ID: 8923218	cytochrome c oxidase subunit I [<i>Homo sapiens subsp. 'Denisova'</i> (Denisova hominin)]	Chromosome MT, NC_013993.1 (5906..7447)		

PTGS1 prostaglandin-endoperoxide synthase 1

- *Homo sapiens* (human)
- Location: 9q33.2
- Exon count: 14

Annotation release	Status	Assembly	Chr	Location
110	current	GRCh38.p14 (GCF_000001405.40)	9	NC_000009.12 (122370533..122395703)
110	current	T2T-CHM13v2.0 (GCF_009914755.1)	9	NC_060933.1 (134568002..134593163)
105.20220307	previous assembly	GRCh37.p13 (GCF_000001405.25)	9	NC_000009.11 (125132812..125157982)



<https://www.ncbi.nlm.nih.gov/gene/5742>

PTGS1 prostaglandin-endoperoxide synthase 1 [*Homo sapiens* (human)]

- <https://www.ncbi.nlm.nih.gov/gene/5742>

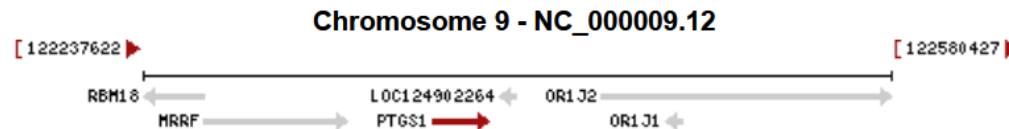
Genomic context

Location: 9q33.2

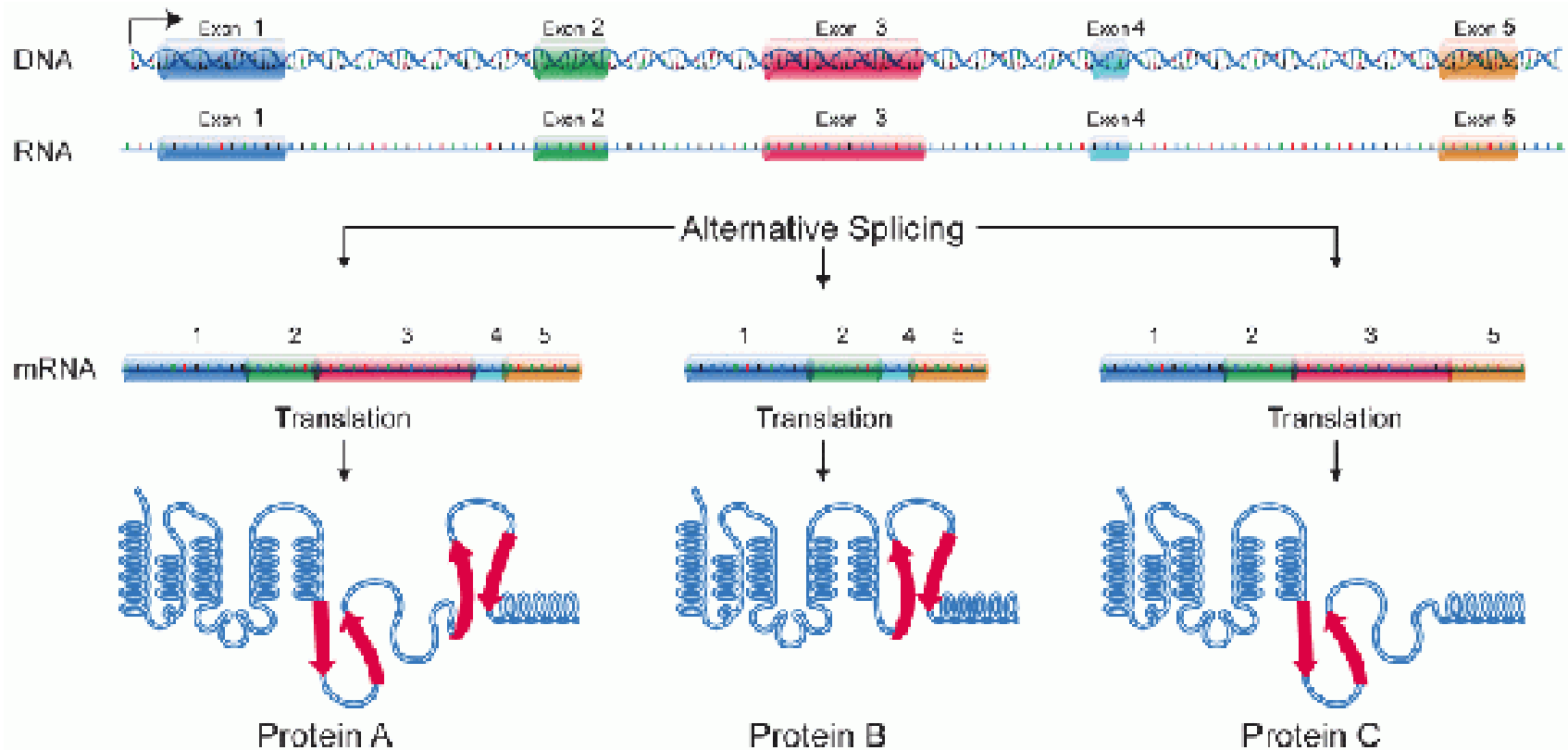
See PTGS1 in [Genome Data Viewer](#)

Exon count: 14

Annotation release	Status	Assembly	Chr	Location
110	current	GRCh38.p14 (GCF_000001405.40)	9	NC_000009.12 (122370533..122395703)
110	current	T2T-CHM13v2.0 (GCF_009914755.1)	9	NC_060933.1 (134568002..134593163)
105.20220307	previous assembly	GRCh37.p13 (GCF_000001405.25)	9	NC_000009.11 (125132812..125157982)



Một gene có thể tạo thành nhiều protein: Alternative splicing



Có nhiều trường hợp trong tế bào của sinh vật nhân thực, cùng 1 gen được phiên mã tạo thành ARN nhưng lại tổng hợp ra nhiều loại protein khác nhau vì do trong quá trình cắt intron, có sự sắp xếp lại của các exon theo các cách khác nhau




Homo sapiens prostaglandin-endoperoxide synthase 1 (PTGS1) gene

```
source      1..26782
            /organism="Homo sapiens"
            /mol_type="genomic DNA"
            /db_xref="taxon:9606"
            /chromosome="9"
            /map="9q32"
gene        898..25644
            /gene="PTGS1"
mRNA        join(898..1039,1134..1220,7845..7961,8379..8519,
            8721..8864,11317..11498,11610..11693,13455..13701,
            16392..16678,20143..20290,22132..25644)
            /gene="PTGS1"
            /product="prostaglandin-endoperoxide synthase 1"
```

<https://www.ncbi.nlm.nih.gov/nucleotide/AF440204.1>

rs1236913

Organism	<i>Homo sapiens</i>	Clinical Significance	Not Reported in ClinVar
Position	chr9:122371200 (GRCh38.p13) ?	Gene : Consequence	PTGS1 : Missense Variant
Alleles	T>C / T>G	Publications	11 citations
Variation Type	SNV Single Nucleotide Variation		 21
Frequency	T=0.066020 (21647/327886, ALFA) T=0.047690 (12623/264690, TOPMED) T=0.073326 (18147/247484, GnomAD_exome) (+ 22 more)	Genomic View	See rs on genome

prostaglandin G/H synthase 1 isoform 2 precursor	NP_542158.1:p.Trp8Gly	W (Trp) > G (Gly)	Missense Variant
prostaglandin G/H synthase 1 isoform 3 precursor	NP_001258093.1:p.Trp8Arg	W (Trp) > R (Arg)	Missense Variant
prostaglandin G/H synthase 1 isoform 3 precursor	NP_001258093.1:p.Trp8Gly	W (Trp) > G (Gly)	Missense Variant
PTGS1 transcript variant 1	NM_000962.4:c.22T>C	W [TGG] > R [CGG]	Coding Sequence Variant
PTGS1 transcript variant 1	NM_000962.4:c.22T>G	W [TGG] > G [GGG]	Coding Sequence Variant

https://www.ncbi.nlm.nih.gov/snp/rs1236913#variant_details



TGG (Tryptophan - W) > CGG (Arginine - R)

```
1 agtgtgcgag gcgcacgcac aggagcctgc actctgcgtc ccgcacccca gcagccgcgc
61 catgagccgg agtctcttgc tctggtttctt gctgttcttg ctctgctcc cgccgctccc
121 cgtcctgctc gcggaccacag gggcgcccac gccagtgaat ccctgttggtt actatccatg
181 ccagcaccag ggcattctgt tccgcttcgg ccttgaccgc taccagtgtg actgcaccgg
```

Homo sapiens prostaglandin-endoperoxide synthase 1 (PTGS1), transcript variant 1, mRNA

NCBI Reference Sequence: NM_000962.4

[FASTA](#) [Graphics](#)

[Go to:](#) ☒

LOCUS	NM_000962	5020 bp	mRNA	linear	PRI 06-NOV-2022
DEFINITION	Homo sapiens prostaglandin-endoperoxide synthase 1 (PTGS1), transcript variant 1, mRNA.				
ACCESSION	NM_000962				
VERSION	NM_000962.4				
KEYWORDS	RefSeq; MANE Select.				
SOURCE	Homo sapiens (human)				
ORGANISM	Homo sapiens Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Haplorrhini; Catarrhini; Hominidae; Homo.				
REFERENCE	1 (bases 1 to 5020)				

https://www.ncbi.nlm.nih.gov/nucore/NM_000962.4/



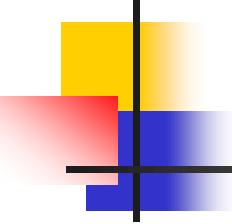
Các lỗi gặp phải khi giải trình tự bằng phương pháp Sanger

- **Quality Control**

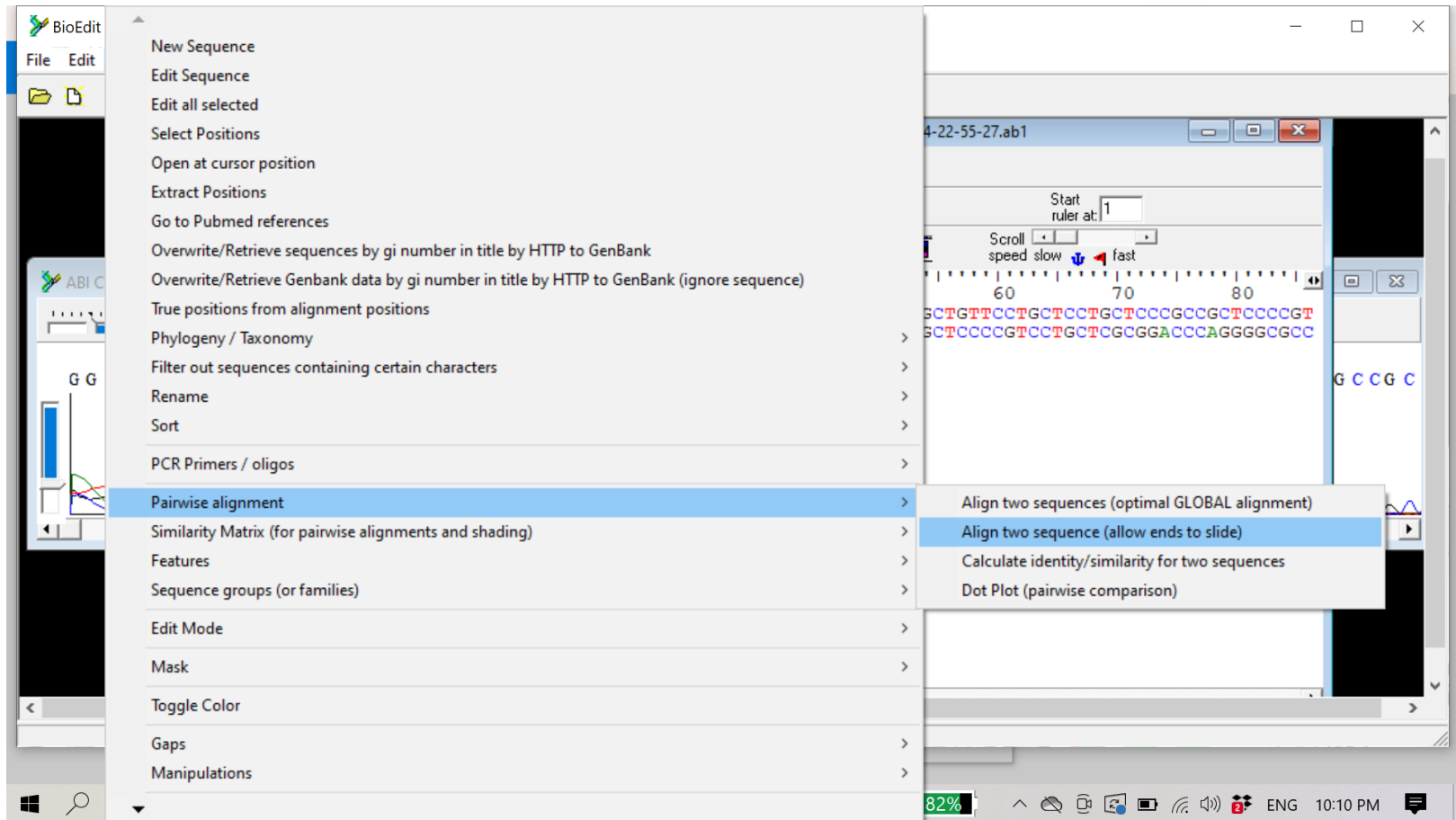
- It is essential that the quality of bases is manually assessed.
- The presence of unambiguous bases and overlapping peaks in the main body of the chromatogram and trace sequence requires application of the secondary genetic bases according to the International Union of Pure and Applied Chemistry (IUPAC) nucleotide code (Cornish-Bowden, 1985) (Table 1).

Table 1. IUPAC nucleotide single code recommendations (Cornish-Bowden, 1985).

<u>Symbol</u>	<u>Meaning</u>	<u>Origin of designation</u>
G	G	Guanine
A	A	Adenine
T	T	Thymine
C	C	Cytosine
R	G or A	puRine
Y	T or C	pYrimidine
M	A or C	aMino
K	G or T	Ketone
S	G or C	Strong interaction (3 H bonds)
W	A or T	Weak interaction (2 H bonds)
H	A or C or T	not-G, H follows G in the alphabet
B	G or T or C	not-A, B follows A
V	G or C or A	not-T (not-U), V follows U
D	G or A or T	not-C, D follows C
N	G or A or T or C	aNy

- 
-
- With Sanger sequencing chemistry, the quality of reads typically depletes towards the start and end of the read.
 - Therefore, these low-quality bases need to be removed in a process called trimming.

BioEdit



BioEdit

BioEdit Sequence Alignment Editor - [C:\BioEdit\Temp\~out.tmp]

File Edit Sequence Alignment View Accessory Application RNA World Wide Web Options Win

2 total sequences

Mode: Select / Slide Selection: 48 to 51 Position: Sequence Mask: None Numbering Mask: None

COX1-1 C22-5 GGGTGCAGMCCCTTCATMTCTCTCCTCTGCAG--GGAGTCTCTTTGCTCCGGGTTCCTTG
lc1|NM_00096-----ATG-AGCCGGAGTCTCTTTGCTCTGGGTTCCTTG

