



**Fundació**  
La Marató de TV3

21è SIMPOSIUM  
Malalties del cor



## **DIAGNÒSTIC GENÒMIC DE LA MALALTIA CORONÀRIA**

**Josep Comín Colet**

Institut d'Investigació Biomèdica de Bellvitge

**Roderic Guigó Serra**

Centre de Regulació Genòmica

## 1. Resum

La malaltia coronària (MC) és una de les primeres causes de mort i font de discapacitats als països desenvolupats. A Catalunya es produeixen més de 34.000 nous diagnòstics d'MC l'any. Els costos sanitaris per aquesta malaltia representen gairebé un terç de tot el pressupost de la sanitat pública. Tot i els grans avenços en els tractaments, a causa de l'augment de l'edat en la població i d'altres factors externs que augmenten el risc cardiovascular, cada any s'incrementa el nombre d'afectats. Molts dels pacients que presenten MC acabaran derivant en insuficiència cardíaca. Això es podria evitar en molts casos amb un diagnòstic més ràpid. Un diagnòstic precoç d'MC permetria un maneig més intensiu i evitar la transició a estadis més complicats i evolucionats de la malaltia. La detecció de l'MC amb estratègies diagnòstiques eficients és essencial per aconseguir, entre d'altres, dos objectius prioritaris: a) el diagnòstic en pacients amb sospita d'MC però sense cap esdeveniment coronari agut, i b) un diagnòstic ràpid en poblacions de pacients amb malaltia arterial coronària subclínica amb factors de risc associats. Una detecció precoç resultarà en una intervenció ràpida, que ens permetria retallar significativament aquests costos massius que provoca la malaltia, tant en vides com en recursos. Per tant, calen avenços en els mètodes de diagnòstic de l'MC.

A causa de la detecció fàcil en mostres de sang total, l'expressió dels perfils d'RNA ha esdevingut una tècnica prometedora en l'establiment de possibles biomarcadors. El conjunt de molècules d'RNA en una mostra biològica representa una fotografia dinàmica de l'estat cel·lular en aquell moment precís. L'anàlisi de biomarcadors sanguinis s'ha concentrat en una petita fracció de la nostra seqüència del genoma, el 2% que codifica per proteïna. El 98% que resta, amb milers d'espècies d'RNA no codificants, està per explorar. Gràcies als avenços en les noves tecnologies de seqüenciació, actualment és possible obtenir el perfil de tot el genoma humà sencer en una única mesura. Aquest mètode es coneix com a seqüenciació d'RNA (RNA-Seq), i detecta tant RNA codificant com no codificant, fet que permet millorar la signatura genètica i la candidatura de potencials biomarcadors. Aprofitant aquesta tècnica, el Projecte GTEx (Genotipe Tissue Expression) ha traçat l'expressió de l'RNA del genoma complet en centenars de pacients amb historial mèdic complet, incloent-hi pacients amb MC.

Aquest projecte té com a objectiu desenvolupar una puntuació diagnòstica per avaluar i diagnosticar la presència d'MC, basada en un algoritme que incorpori dades tant clíniques com d'expressió genètica del pacient. Amb aquesta finalitat, cal definir un conjunt de biomarcadors d'expressió genètica que permeti aportar informació sobre l'expressió diferencial entre casos amb MC i controls sense MC. El desenvolupament d'aquests models diagnòstics, segons les recomanacions internacionals, exigeixen cohorts de pacients, tant per a la derivació del model com per validar-lo.

Les fases metodològiques de l'estudi són:

1. Identificació de potencials biomarcadors d'RNA associats a MC en una cohort retrospectiva:
  - a. Definició en aquesta primera població d'estudi (primera cohort retrospectiva del GTEx) mitjançant anàlisi bioinformàtica de les dades procedents de l'RNA-seq, d'un conjunt de biomarcadors ampli (*Biomarker Set #1*), que mostrin expressió diferencial entre casos i controls, i amb associació estadística.
  - b. Definició per criteris clínics, metodològics, estadístics i basats en coneixement previ, d'un conjunt reduït de biomarcadors d'RNA procedent d'aquesta primera anàlisi (*Biomarker Set #2*).
2. Confirmació del conjunt de biomarcadors d'RNA (*Biomarker Set #2*) en una nova cohort de pacients (cohort per a la confirmació i derivació de l'algoritme diagnòstic).
3. Definició d'un conjunt de biomarcadors a partir d'aquestes dades, més concentrat i amb millor capacitat discriminatòria entre casos i controls (*Biomarker Set #3*), i amb aquest conjunt definir un algoritme diagnòstic que integri també dades clíniques del pacient (expressat en forma d'*score*).
4. Validació de l'algoritme en nous conjunts de pacients avaluats de forma prospectiva.

Aquest estudi ha implicat, doncs, l'anàlisi retrospectiva de l'expressió genètica diferencial entre pacients amb i sense MC de les cohorts del GTEx, i el reclutament prospectiu d'una cohort de pacients a nivell català que, per diverses indicacions, eren sotmesos a estudi de la presència d'MC de manera no invasiva amb un TAC coronari.

La idea original era, primerament, fer la derivació del *Biomarker Set #1* i *#2* de la cohort del GTE*x* i, en segon lloc, fer una validació en la segona cohort (retrospectiva) del *Biomarker Set #2* i derivar-ne un refinament (*Biomarker Set #3*) que permetés construir un *score* final que integri dades genètiques i clíniques. Aquest *score* permetrà fer una aproximació diagnòstica no invasiva de la presència o absència d'MC en altres poblacions.

Amb aquest treball es volen cercar per primera vegada nous biomarcadors sanguinis d'RNA indicatius d'MC aprofitant les dades aportades pel Projecte GTE*x*. La idea és generar un algoritme per diagnosticar MC amb una major precisió que l'actual. Aquesta seria la base d'una prova assequible per al diagnòstic mèdic no invasiu per als pacients amb sospita d'MC.

## **2. Resultats obtinguts**

A nivell clínic, s'han monitorat les dades clíniques corresponents al seguiment d'estat vital de tots els pacients reclutats. El projecte complet ha permès reclutar un total de 353 pacients amb dades clíniques, d'imatge cardíaca i de laboratori analitzables. La feina clínica associada a aquest aspecte del projecte inclou no només la recollida basal de la informació, sinó el seguiment de la qualitat en les dades obtingudes, tant al nivell clínic com d'imatge.

### **Resultats clínics (cohorts reclutades)**

L'edat mitjana dels pacients va ser de  $60 \pm 11$  anys i un 54% eren dones. L'origen dels pacients era en un 87,5% europeu.

Les variables antropomòrfiques recollides van ser les següents: la mitjana de pes va ser de  $76,7 \pm 15$  kg, la talla va ser de  $162,7 \pm 20$  cm, perímetre de cintura de  $96 \pm 14$  cm, i perímetre de maluc de  $105 \pm 10$  cm. Les constants en el moment de fer la prova eren: FC,  $71 \pm 14$  bpm; tensió arterial sistòlica,  $131 \pm 23$  mmHg, i tensió arterial diastòlica,  $74 \pm 14$  mmHg. La presència de factors de risc, altres patologies rellevants i tractament mèdic en la població inclosa es resumeix a la taula següent.

Factors de risc i patologies rellevants	Freqüència	%
Hipertensió arterial	193	55
Diabetes Mellitus	55	16
Dislipèmia	181	51
Fumadors actius	54	15
Exfumadors	104	30
Antecedents familiars d' IC precoç	30	8.5
Tractament Farmacològic	Freqüència	%
Betablocadors	131	37
Diürètics	80	23
IECAs /ARAII	147	42
Antagonistes Calci	46	13
Insulina	14	4
Metformina	43	12
Estatines	164	46
Ezetimibe	26	7
AAS	152	43
Clopidogrel	19	3

En la majoria dels pacients (74%) el TAC coronari es va fer per estudi de dolor toràcic. El dolor toràcic es va classificar: típic (16%), atípic (55%) o no coronari (14%). En un 82% dels pacients es va fer score de calci, i la puntuació mitjana de la població global va ser de  $114 \pm 246$ . Pel que fa als valors analítics de la població reclutada, destaquem els paràmetres següents: Hb,  $14,4 \pm 6$  g/dL; creatinina,  $1,3 \pm 5,5$  mg/dL; Hb, A1c  $5,8 \pm 2\%$ ; el colesterol total/LDL/HDL va ser de  $186 \pm 40$  mg/dL/  $106 \pm 44$  mg/dL/  $54 \pm 16$  mg/dL, i triglicèrids  $122 \pm 66$  ng/dL.

En l'anàlisi de l'anatomia coronària dels TAC vam trobar que en la població seleccionada, un 75% dels pacients no tenien MC. Dels pacients estudiats amb MC, vam observar que 49 (14%) tenien malaltia significativa d'un vas, 29 (8%) tenien MC de dos vasos i 10 (3%) tenien MC de tres vasos.

Quant al seguiment, la supervivència de la població al cap d'un any era d'un 94%; 3 pacients van ser *exitus*. Un 5% dels pacients van ser tractats amb revascularització en el seguiment.

## **Procés d'anàlisi genòmica i resultats**

En el nostre projecte, a partir de la base de dades del GTEx hem pogut fer una prospecció de genoma complet que ens ha permès definir l'expressió genètica diferencial entre casos d'MC i controls sense les limitacions de cercar només el que prèviament s'havia publicat a la bibliografia científica o segons gens candidats. Mitjançant un procés de treball multidisciplinari integrant clínics, biòlegs i bioinformàtics, hem definit un conjunt inicial de 313 biomarcadors genètics que s'expressen de forma diferenciada entre aquests subjectes amb i sense MC. Aquest conjunt de biomarcadors constitueix el conjunt biomarcador 1 (*Biomarker Set #1*). A partir de la revisió de la bibliografia científica i la interpretació clínica dels resultats inicials, l'equip investigador ha decidit definir un conjunt més reduït de gens expressats, fins a un nombre de 112, per tal de ser validats en cohorts subsegüents. Aquest conjunt de gens constitueix el conjunt següent de biomarcadors (*Biomarker Set #2*).

Arribats a aquest punt, i abans de començar qualsevol tipus de validació del *Biomarker Set #2*, a les mostres que s'estaven recollint als nostres hospitals es va decidir fer un control de qualitat de les dades obtingudes de l'anàlisi del GTEx comparant-lo amb la seqüenciació de les nostres mostres directament. Els arguments a favor d'aquesta validació en paral·lel van ser:

1. La baixa consistència i reproductibilitat de les troballes en diversos estudis independents per identificar panells de gens associats amb cardiopaties.
2. Els recents avenços en les tecnologies RNA-Seq en relació amb la metodologia usada al GTEx.
3. La superposició entre els gens identificats en els diversos estudis no és perfecta.
4. El substrat genètic i sociocultural dels individus del projecte GTEx és molt diferent del de la població de Catalunya, objectiu de l'anàlisi.

Per tant, es va procedir a fer l'RNA-Seq complet seguint el mateix protocol d'extracció i manipulació d'RNA efectuat en el GTEx (Paxgene kit). Aquesta nova seqüenciació d'un subgrup de la nostra pròpia població recollida en hospitals catalans ens permetria:

1. Comparar els perfils d'expressió gènics de la població GTEx amb la nostra població catalana.
2. Validar que els gens identificats en el conjunt de Biomarcadors #1 (extrets de la població GTEx) siguin extrapolables a la població catalana.
3. Afinar i restringir el conjunt de Biomarcadors #1 amb aquells gens que hagin estat reproduïts en ambdues poblacions o que apareguin com a marcadors d'MC en la població catalana i que es considerin interessants.

La selecció de pacients reclutats inicialment en el moment de començar aquest experiment es va fer en funció de les lesions observades al TAC inicial. Es van establir tres grups:

**GRUP 1 - sense lesions:** el TAC no mostra cap vas afectat.

**GRUP 2 - lesions intermèdies:** el TAC mostra lesions de màxim 49% en un o més vasos i/o segments.

**GRUP 3 - lesions greus:** el TAC mostra lesions >49% en, com a mínim, un o més vasos i/o segments.

Malauradament, després de dues tandes completes de seqüenciació en 60 pacients en total (amb 30 pacients per grup), i d'afegir un pas extra de purificació de l'RNA, no es va poder arribar a fer cap anàlisi, ja que hi havia problemes de degradació de la mostra en el pas de la depleció d'hemoglobina i, per tant, mai s'arribava a tenir una mostra seqüenciada òptima per a l'anàlisi bioinformàtica posterior.

Així doncs, en converses amb el laboratori del CRG, es va proposar una nova seqüenciació amb un protocol exclusiu dissenyat *in house* per resoldre els problemes de la degradació, ja que la qualitat de les mostres extretes abans de començar aquest punt del processament era òptima.

Per a la nova seqüenciació d'RNA-Seq amb un nou *kit* de depleció d'hemoglobina, es van triar 10 mostres, aquesta vegada; 4 del grup 1 (sense lesions) i 6 del grup 3 (lesions greus), per posar a punt el nou protocol. En aquesta tanda, les mostres no van presentar la degradació que presentaven les anteriors en aquest punt del procés. Per

tant, s'ha aconseguit posar a punt un protocol òptim de seqüenciació que ens permet el processament adequat, i una anàlisi bioinformàtica posterior de les mostres, per tal de continuar seguidament amb la validació. Les dades de seqüenciació finals les vam obtenir a la darrera del desembre del 2019. Una anàlisi bioinformàtica preliminar de les 10 mostres provinents dels dos grups (1 i 3) ens indica que hi ha més de 200 gens diferencialment expressats entre grups. La majoria presenten una sobreregulació de l'expressió en els pacients, tal com inicialment s'havia trobat a l'anàlisi de les dades GTEEx. La qual cosa ens indica que la nostra cohort local és parcialment comparable a les dades presents al GTEEx.

Actualment s'està pendent d'augmentar el nombre de mostres a seqüenciar amb el nou protocol de processament per RNA-Seq, tot augmentant el nombre a uns 30 pacients per cada grup d'estudi (grups 1, 2 i 3). Un cop seqüenciades i analitzades totes les mostres, ens permetrà definir el panel d'expressió d'RNA diferencial entre casos i controls. Un dels problemes a l'hora d'analitzar els resultats de l'RNA-Seq i les possibles diferències entre les poblacions és la gran variabilitat entre els pacients. Per disminuir aquesta variabilitat s'estan aplicant programes informàtics que permetin eliminar el possible efecte *batch*. La resta de les mostres que hem obtingut (un total de 353 pacients) estan emmagatzemades a -80 °C per a les futures aplicacions previstes en la fase final de l'estudi.

### **3. Rellevància i possibles implicacions futures**

Tot i no comptar amb els resultats finals derivats del projecte, el fet de seqüenciar mostres de pacients de la nostra població ens donarà un valor afegit a tot l'estudi, ja que podrem validar-los de forma més consistent, generar un algoritme diagnòstic més acurat i, a la vegada, ens permetrà comparar els resultats de la nostra població amb les dades de la primera anàlisi bioinformàtica del GTEEx.

La trobada de nous biomarcadors sanguinis d'MC i la possibilitat de definir un *score* diagnòstic permetria millorar i simplificar el procés diagnòstic de l'MC (cada any hi ha a Catalunya desenes de milers de nous casos). A la vegada, la detecció precoç significaria un tractament ràpid, i així es podria parar l'evolució de la malaltia, tot reduint els esdeveniments clínics i el cost sanitari associat a una malaltia tan prevalent.



D'altra banda, aquesta nova eina podria ajudar a fer un cribratge ràpid de pacients asimptomàtics d'MC, tot millorant l'estratificació del risc cardiovascular. A més, l'algoritme derivat del projecte podria incorporar-se a les guies clíniques, i així variar i millorar la pràctica mèdica diària. Finalment, ens permetria la descripció de nous mecanismes fisiopatològics, i per tant dissenyar noves estratègies terapèutiques per a l'MC.

En resum, aquest projecte té el potencial d'aconseguir avenços significatius en el diagnòstic clínic, tot facilitant una considerable millora en la vida de milers de persones a Catalunya que pateixen l'MC diagnosticada o no.

#### **4. Bibliografia científica generada**

A causa dels problemes i canvis de protocol anteriorment comentats, fins al moment no s'ha pogut generar cap article, tot i que actualment s'està treballant en l'elaboració del manuscrit de disseny de l'estudi amb les dades basals dels pacients. L'explotació dels resultats de l'estudi s'efectuarà un cop completada almenys la fase de descoberta de biomarcadors, i posteriorment es continuarà amb la validació i la possible generació de l'*score* diagnòstic.