



Fundació
La Marató de TV3
21º SIMPOSIUM
Enfermedades del corazón



DIAGNÓSTICO GENÓMICO DE LA ENFERMEDAD CORONARIA

Josep Comín Colet

Institut d'Investigació Biomèdica de Bellvitge

Roderic Guigó Serra

Centre de Regulació Genòmica

1. Resumen

La enfermedad coronaria (EC) es una de las primeras causas de muerte y fuente de discapacidad en los países desarrollados. En Cataluña se producen más de 34.000 nuevos diagnósticos de EC al año. El coste sanitario de esta enfermedad representa casi un tercio de todo el presupuesto de la sanidad pública. Pese a los grandes avances en los tratamientos, debido al aumento de la edad en la población y a otros factores externos que aumentan el riesgo cardiovascular, cada año se ve incrementado el número de afectados. Muchos de los pacientes que presentan EC acabarán derivando en insuficiencia cardíaca. Esta contrariedad podría evitarse en muchos casos con un diagnóstico más rápido. Un diagnóstico precoz de EC permitiría realizar un manejo intensivo de la situación y evitar la transición a estadios más complicados y evolucionados de la enfermedad. La detección de EC con estrategias diagnósticas eficientes es esencial para alcanzar, entre otros, dos objetivos prioritarios: a) el diagnóstico del paciente con sospecha de EC pero sin ningún evento coronario agudo, y b) un rápido diagnóstico en poblaciones de pacientes con enfermedad arterial coronaria subclínica con factores de riesgo asociados. Una detección precoz resultará en una rápida intervención, permitiendo recortar significativamente el coste masivo que provoca esta enfermedad, tanto en vidas como en recursos. Por tanto, son necesarios avances en los métodos de diagnóstico de la EC.

Debido a su fácil detección en muestras de sangre total, la expresión de los perfiles de RNA se ha convertido en una técnica prometedora en la determinación de posibles biomarcadores. El conjunto de moléculas de RNA en una muestra biológica representa una fotografía dinámica del estado celular en ese momento. El análisis de biomarcadores sanguíneos se ha concentrado en una pequeña fracción de nuestra secuencia del genoma, el 2% que codifica por proteína. El 98% restante, con miles de especies de RNA no codificantes, está aún por explorar. Gracias a los avances en las nuevas tecnologías de secuenciación, en la actualidad es posible obtener el perfil de todo el genoma humano en una única medición. Este método se conoce como secuenciación de RNA (RNA-Seq), y detecta tanto RNA codificante como no codificante, lo que permite mejorar la firma genética y la candidatura de potenciales biomarcadores. Aprovechando esta técnica, el Proyecto GTEx (Genotipe Tissue Expression) ha trazado la expresión de RNA del genoma completo de centenares de pacientes con historial médico completo, incluyendo pacientes con EC.

El presente proyecto tiene como fin desarrollar una puntuación diagnóstica para evaluar y diagnosticar la presencia de EC, basada en un algoritmo que incorpore datos clínicos y de expresión genética del paciente. A tales efectos, se precisa la definición de un conjunto de biomarcadores de expresión genética que permita aportar información sobre la expresión diferencial entre casos con EC y controles sin ella. El desarrollo de estos modelos diagnósticos, según las recomendaciones internacionales, precisan de cohortes de pacientes, tanto para la derivación del modelo como para su validación. Las fases metodológicas del estudio son:

1. Identificación de potenciales biomarcadores de RNA asociados a EC en una cohorte retrospectiva:
 - a. Definición en esta primera población de estudio (primera cohorte retrospectiva del GTE_x) mediante análisis bioinformático de los datos procedentes del RNA-Seq, de un conjunto amplio de biomarcadores (*Biomarker Set #1*), que muestren la expresión diferencial entre casos y controles, y con asociación estadística.
 - b. Definición por criterios clínicos, metodológicos, estadísticos y basados en conocimiento previo, de un conjunto reducido de biomarcadores de RNA procedente de este primer análisis (*Biomarker Set #2*).
2. Confirmación del conjunto de biomarcadores de RNA (*Biomarker Set #2*) en una nueva cohorte de pacientes (cohorte para la confirmación y derivación del algoritmo diagnóstico).
3. Definición de un conjunto de biomarcadores a partir de estos datos, más concentrado y con mejor capacidad discriminatoria entre casos y controles (*Biomarker Set #3*), y con este conjunto definir un algoritmo diagnóstico que integre también los datos clínicos del paciente (expresado en forma de *score*).
4. Validación del algoritmo en nuevos conjuntos de pacientes evaluados de forma prospectiva.

Así, el presente estudio ha supuesto el análisis retrospectivo de la expresión genética diferencial entre pacientes con y sin EC en las cohortes del GTE_x, y el reclutamiento prospectivo de una cohorte de pacientes a nivel catalán que, por diferentes

indicaciones, eran sometidos a estudio de la presencia de EC de forma no invasiva con un TAC coronario. La idea original era, en primer lugar, realizar la derivación del *Biomarker Set #1* y *#2* de la cohorte del GTEx y, en segundo lugar, realizar una validación en la segunda cohorte (retrospectiva) del *Biomarker Set #2*, derivando un conjunto refinado de esta (*Biomarker Set #3*), que permitiese construir un *score* final que integre los datos genéticos y clínicos. Este *score* va a permitir efectuar una aproximación diagnóstica no invasiva de la presencia o ausencia de EC en otras poblaciones.

Con el presente trabajo se buscan por primera vez nuevos marcadores sanguíneos de RNA indicativos de EC, aprovechando los datos aportados por el Proyecto GTEx. La idea es generar un algoritmo para diagnosticar EC con mayor precisión que en la actualidad. Esta sería la base de una prueba asequible para el diagnóstico médico no invasivo para los pacientes con sospecha de EC.

2. Resultados obtenidos

A nivel clínico, se han monitorizado los datos clínicos correspondientes al seguimiento del estado vital de los pacientes reclutados. El proyecto completo ha permitido reclutar un total de 353 pacientes con datos tanto clínicos, como de imagen cardíaca y de laboratorio analizables. El trabajo clínico asociado a este aspecto del proyecto incluye no solo la recogida basal de la información, sino el seguimiento de la calidad de las muestras obtenidas, tanto a nivel clínico como de imagen.

Resultados clínicos (cohortes reclutadas)

La edad media de los pacientes fue de 60 ± 11 años, siendo un 54% mujeres. El origen de los pacientes era europeo en un 87,5%.

Las variables antropomórficas recogidas fueron las siguientes: la media del peso, $76,7 \pm 15$ kg; la altura, $162,7 \pm 20$ cm; perímetro de cintura, 96 ± 14 cm, y perímetro de cadera, 105 ± 10 cm. Las constantes en el momento de la prueba eran: FC, 71 ± 14 bpm; tensión arterial sistólica, 131 ± 23 mmHg, i tensión arterial diastólica, 74 ± 14 mmHg. La presencia de factores de riesgo, otras patologías relevantes y tratamiento médico en la población incluida, se resumen en la siguiente tabla.

Factores de riesgo y patologías relevantes	Frecuencia	%
Hipertensión arterial	193	55
Diabetes Mellitus	55	16
Dislipemia	181	51
Fumadores activos	54	15
Exfumadores	104	30
Antecedentes familiares de IC precoz	30	8,5
Tratamiento farmacológico	Frecuencia	%
Betabloqueadores	131	37
Diuréticos	80	23
IECAs /ARAII	147	42
Antagonistas Calcio	46	13
Insulina	14	4
Metformina	43	12
Estatinas	164	46
Ezetimibe	26	7
AAS	152	43
Clopidogrel	19	3

En la mayoría de los pacientes (74%) el TAC coronario se realizó por estudio de dolor torácico. El dolor torácico se clasificó como típico (16%), atípico (55%) o no coronario (14%). En un 82% de los pacientes se efectuó *score* de calcio, siendo la puntuación media de la población global de 114 ± 246 . En cuanto a los valores analíticos de la población reclutada, destacan los siguientes parámetros: Hb, $14,4 \pm 6$ g/dL; creatinina, $1,3 \pm 5,5$ mg/dL; Hb A1c, $5,8 \pm 2\%$; colesterol total/LDL/HDL fue de 186 ± 40 mg/dL/ 106 ± 44 mg/dL / 54 ± 16 mg/dL, y triglicéridos, 122 ± 66 ng/dL. En el análisis de la anatomía coronaria de los TAC, se observó que en la población seleccionada, un 75% de los pacientes no tenían EC. De los pacientes estudiados con EC, se observó que 49 (14%) tenían enfermedad significativa de un vaso, 29 (8%) tenían EC de dos vasos y 10 (3%) tenían EC de tres vasos.

En cuanto al seguimiento, la supervivencia de la población al año era de un 94%; 3 pacientes fueron *exitus*. Un 5% de los pacientes fueron tratados con revascularización en el seguimiento.

Proceso de análisis genómico y resultados

En nuestro proyecto, a partir de la base de datos del GTEx, hemos podido realizar una prospección del genoma completo que nos ha permitido definir la expresión genética diferencial entre casos de EC y controles sin las limitaciones de buscar solo lo que

previamente se había publicado en la bibliografía científica o en función de genes candidatos. Mediante un proceso de trabajo multidisciplinar integrado por clínicos, biólogos y bioinformáticos, hemos definido un conjunto inicial de 313 biomarcadores genéticos que se expresan de forma diferenciada entre estos sujetos con y sin EC. Este conjunto de biomarcadores constituye el conjunto de biomarcadores 1 (*Biomarker Set #1*). En base a la revisión de la bibliografía científica y a la interpretación clínica de los resultados iniciales, el equipo investigador ha decidido definir un conjunto más reducido de genes expresados, hasta un número total de 112, para su validación en siguientes cohortes. Este conjunto de genes constituye el conjunto de biomarcadores 2 (*Biomarker Set #2*).

Llegados a este punto, y antes de empezar cualquier tipo de validación del *Biomarker Set #2*, en las muestras que se estaban recogiendo en nuestros hospitales se decidió efectuar un control de calidad de los datos obtenidos en el análisis del GTEx comparándolo con la secuenciación de nuestras muestras directamente. Los argumentos a favor de esta validación en paralelo fueron:

1. La baja consistencia y reproducibilidad de los resultados en diversos estudios independientes para identificar paneles de genes asociados a cardiopatías.
2. Los recientes avances en las tecnologías de RNA-Seq en relación con la metodología utilizada en el GTEx.
3. La superposición entre los genes identificados en los diferentes estudios no es perfecta.
4. El sustrato genético y sociocultural de los individuos del proyecto GTEx es muy diferente al de la población de Cataluña a analizar.

Por tanto, se procedió a efectuar un RNA-Seq completo, siguiendo el mismo protocolo de extracción y manipulación de RNA llevado a cabo en el GTEx (Paxgene kit). Esta nueva secuenciación de un subgrupo de nuestra propia población recogido en hospitales catalanes nos permitiría:

1. Comparar los perfiles de expresión génicos de la población GTEx con nuestra población catalana.
2. Validar que los genes identificados en el conjunto de Biomarcadores #1 (extraídos de la población GTEx) sean extrapolables a la población catalana.
3. Afinar y restringir el conjunto de Biomarcadores #1 con aquellos genes que hayan sido reproducidos en ambas poblaciones o que aparezcan como marcadores de EC en la población catalana y se consideren más interesantes.

La selección de pacientes reclutados inicialmente en el momento de empezar este experimento fue realizada en función de las lesiones observadas en el TAC inicial. Se establecieron tres grupos:

GRUPO 1 - sin lesiones: el TAC no muestra ningún vaso afectado.

GRUPO 2 - lesiones intermedias: el TAC muestra lesiones de máximo 49% en uno o más vasos y/o segmentos.

GRUPO 3 - lesiones graves: el TAC muestra lesiones >49% en, como mínimo, uno o más vasos y/o segmentos.

Lamentablemente, tras dos tandas completas de secuenciación en 60 pacientes en total (con 30 pacientes por grupo) y añadir un paso extra de purificación de RNA, no se pudo llevar a cabo ningún análisis, ya que aparecieron problemas con la degradación de la muestra en el paso de depleción de hemoglobina y, por tanto, nunca se llegaba a obtener una muestra secuenciada óptima para su posterior análisis bioinformático.

Así pues, en conversaciones con el laboratorio del CRG, se propuso una nueva secuenciación con un protocolo diseñado *in house* para tratar de solucionar los problemas de la degradación, ya que la calidad de las muestras extraídas antes de llegar a este punto del proceso era óptima.

Para la nueva secuenciación de RNA-Seq con un nuevo *kit* de depleción de hemoglobina, en esta ocasión se eligieron 10 muestras, 4 del grupo 1 (sin lesiones) y 6 del grupo 3 (lesiones graves), para poner a punto el nuevo protocolo. En esta tanda, las muestras no presentaron la degradación que aparecía en las anteriores ocasiones

en este punto del proceso. Por tanto, se ha logrado poner a punto un protocolo óptimo de secuenciación que nos permite el procesamiento adecuado, y el posterior análisis bioinformático de las muestras, continuando seguidamente con la fase de validación. Los datos de secuenciación definitivos fueron obtenidos a finales de diciembre de 2019. Un análisis bioinformático preliminar de las 10 muestras procedentes de los dos grupos (1 y 3) nos indica que hay más de 200 genes diferencialmente expresados entre grupos. La mayoría de ellos presentan una sobrerregulación de la expresión en los pacientes, tal y como inicialmente se había hallado en el análisis de los datos GTEx. Lo que nos indica que nuestra cohorte local es parcialmente comparable a los datos presentes en el GTEx.

En la actualidad está pendiente el aumento del número de muestras a secuenciar con el nuevo protocolo de RNA-Seq, aumentando a 30 por grupo el número de pacientes (grupos 1, 2 y 3). Una vez secuenciadas y analizadas todas las muestras, nos permitirá definir un panel de expresión de RNA diferencial entre casos y controles. Uno de los problemas a la hora de analizar los resultados del RNA-Seq y las posibles diferencias entre las poblaciones es la gran variabilidad entre los pacientes. Para disminuir dicha variabilidad se están aplicando programas informáticos que permitan eliminar el posible efecto *batch*. El resto de las muestras obtenidas (un total de 353 pacientes) se encuentran almacenadas a -80 °C para futuras aplicaciones previstas en la fase final del estudio.

3. Relevancia y posibles implicaciones futuras

Pese a no disponer de los resultados finales derivados del proyecto, secuenciar muestras de pacientes de nuestra propia población dará un valor añadido al estudio, ya que podremos validarlos de forma más robusta, generar un algoritmo diagnóstico más sensible y, a su vez, va a permitir comparar los resultados de nuestra población con los datos del primer análisis bioinformático del GTEx.

El hallazgo de nuevos biomarcadores sanguíneos de EC y la posibilidad de definir un *score* diagnóstico permitiría mejorar y simplificar el proceso diagnóstico de la EC (cada año hay en Cataluña decenas de miles de nuevos casos). Asimismo, la detección precoz conllevaría un rápido tratamiento, pudiéndose parar la evolución de la

enfermedad, al reducir los eventos clínicos y el coste sanitario asociado a una enfermedad tan prevalente.

Por otra parte, esta nueva herramienta podría ayudar a efectuar un rápido cribado de pacientes asintomáticos de EC, mejorando así la estratificación de riesgo cardiovascular. Además, el algoritmo derivado del proyecto podría incorporarse en las guías clínicas, variando y mejorando la práctica médica diaria. Finalmente, nos permitiría la descripción de nuevos mecanismos fisiopatológicos, y por tanto diseñar nuevas estrategias terapéuticas para la EC.

En resumen, este proyecto tiene como potencial alcanzar avances significativos en el diagnóstico clínico, dando lugar a una considerable mejora en la vida de miles de personas en Cataluña que sufren EC diagnosticada o no.

4. Bibliografía científica generada

Debido a los problemas y cambios de protocolo anteriormente comentados, hasta el momento no se ha podido generar ningún artículo, pese a que en la actualidad se está trabajando en la elaboración del manuscrito sobre el diseño del estudio con los datos basales de los pacientes. La explotación de los resultados del estudio se efectuará una vez completada al menos la fase de descubrimiento de biomarcadores, y posteriormente se seguirá con la validación de la misma y la posible generación del *score* diagnóstico.