

Progresión del cáncer de mama

Martín Carlos Abba

Introducción

Etiología del cáncer de mama

El cáncer de mama es una enfermedad multifactorial que puede desarrollarse como resultado de las interacciones entre factores genéticos, ambientales y hormonales (Figura 1). Entre los factores endocrinos y reproductivos asociados con un incremento del riesgo de desarrollo de cáncer de mama se pueden mencionar: menarca precoz, nuliparidad, edad madura del primer embarazo, y menopausia tardía. Los factores ambientales asociados con un riesgo incrementado de desarrollo de cáncer de mama incluyen: las dietas ricas en grasas, el consumo excesivo de alcohol, la exposición a radiación, y la exposición a químicos.

De esta manera, la constitución genética sumado a la influencia de los múltiples factores endógenos y exógenos, es lo que en última instancia determinará el riesgo específico de cada mujer de desarrollar cáncer de mama. La variabilidad genética (Ejemplo: SNPs, CNVs, etc.) encontrado en la población general explica en parte porque ciertos individuos desarrollan cáncer mientras que otros no, incluso

cuando son expuestos durante el mismo intervalo de tiempo a los mismos factores de riesgo.

A pesar de que el cáncer de mama, ha sido considerado un tipo de enfermedad neoplásica, la biología celular y molecular del mismo indicarían la existencia de heterogeneidad en cuanto a subtipos tumorales a la cual se puede arribar por causa de múltiples factores tales como alteraciones a niveles epigenéticos, genéticos y transcripcionales.

Progresión y clasificación del cáncer de mama

El cáncer de mama no es una única enfermedad en cuanto a su presentación. En cambio, es una colección de enfermedades de la mama que tienen diversas histopatologías, variaciones genéticas / genómicas y respuestas clínicas. Existen varios tipos de lesiones benignas en la mama humana, y sólo unas pocas parecen tener un potencial premaligno significativo. Las lesiones premalignas mejor caracterizadas son: hiperplasia ductal atípica (HDA), hiperplasia lobular atípica (HLA), carcinoma ductal in situ (CDIS), y carcinoma lobular in situ (CLIS). La HDA puede progresar para convertirse en CDIS preinvasivo, seguido por una posible progresión a carcinoma ductal invasivo (CDI) y la culminación con la enfermedad metastásica.

El CDIS es una lesión neoplásica local sin compromiso del tejido regional adyacente y precursora de los CDI. Estudios epidemiológicos establecieron que las pacientes con CDIS poseen un riesgo 10 veces mayor de desarrollar CDI que las mujeres sin antecedentes de este tipo de lesión. Se estima que un tercio de dichas lesiones premalignas, si no son tratadas, progresarían a CDI (Kerlikowske, 2010).

Se desconoce la razón por la cual solo algunos CDIS progresan al estadio invasivo, mientras que otros aparentemente no lo hacen. Por ello, la detección del subgrupo de CDIS con potencial de progresión no solo reduciría dramáticamente la incidencia de CDI facilitando

el desarrollo de nuevas estrategias terapéuticas, sino también impactaría positivamente disminuyendo el sobretratamiento de pacientes (Bleyer, 2012).

El análisis de perfiles de expresión génica ha sido extensamente aplicado al estudio de la progresión del cáncer de mama humano, permitiendo la identificación de nuevos subtipos tumorales. En un estudio pionero, Perou y col. (2000) demostraron que los carcinomas primarios infiltrantes de mama podían ser clasificados en 5 subtipos tumorales intrínsecos, en función de los patrones de expresión génica a escala genómica (Perou y otros, 2000). Estos grupos fueron denominados como: carcinomas del tipo Luminal A y Luminal B, ambos receptor de estrógenos (RE) y receptor de progesterona (RP) positivos; carcinomas del tipo basal/triple negativos (RE negativo, RP negativo, HER2 negativo); carcinomas HER2/NEU positivos y RE negativos; y carcinomas con perfiles de expresión similares a los de las células epiteliales normales. Es importante mencionar que cada uno de los subtipos tumorales intrínsecos posee valor pronóstico / predictivo, donde los subtipos luminales presentan la mejor curva de supervivencia en contraste con los subtipos basales con el peor valor pronóstico.

Estudios genómicos y sus aplicaciones

El nacimiento del Proyecto del Genoma humano dio lugar a la genómica estructural que tiene como principal objetivo determinar, mapear y anotar las secuencias nucleotídicas codificantes y no codificantes de los genomas completos en los distintos organismos (Quackenbush, 2011). El resultado de cualquier proyecto 'genoma', es la descripción digitalizada y estandarizada de la estructura del mismo, pero dista mucho de comprender su comportamiento. Cabe hacer la analogía con los cientos de años que costó pasar de una descripción detallada de la anatomía humana, a comprender la función de los

distintos órganos y tejidos. El conocimiento de la secuencia completa del genoma y de la localización de todos los genes, no es más que una descripción ‘anatómica’ del genoma.

Por otro lado, la genómica funcional se encarga de la recolección sistemática de información sobre la función e interacción entre genes y/o proteínas, mediante la aplicación de aproximaciones experimentales globales. Es una de las ciencias que permite comprender cómo funciona el genoma en su conjunto, mediante la expresión controlada de todos y cada uno de sus genes.

Existen distintas disciplinas dentro de la genómica funcional, en general nombradas con el sufijo “ómica”. La principal diferencia entre las ómicas es la especie de macromolécula objeto de estudio. De esta manera, la transcriptómica se encarga de la caracterización de transcriptomas, es decir el estudio de la expresión de todos los genes de un genoma en una muestra determinada. La proteómica persigue la caracterización de la totalidad de las proteínas expresadas en un grupo celular o tejido en particular.

De esta manera, la finalización del Proyecto Genoma Humano ha dado lugar al desarrollo de numerosas metodologías de análisis en genómica funcional, produciendo un gran impacto en el estudio de las bases genéticas de enfermedades complejas tales como el cáncer (Quackenbush, 2011). En general estas metodologías se basan en el análisis de datos genómicos, generados mediante el uso de diferentes tecnologías (Figura 2). Entre las tecnologías más populares se encuentran las plataformas basadas en la hibridación reversa sobre matrices sólidas (micro-arreglos de ADNc u oligonucleótidos) y las plataformas basadas en la re-secuenciación de genomas y transcriptomas (Exome-Seq, RNA-seq, etc.). Dichas tecnologías pueden emplearse para caracterizar cambios a nivel epigenético (como las modificaciones en histonas y los fenómenos de hipo/hipermetilación de regiones CpG), genómicas (mutaciones, variaciones en el número de copia de los genes, alteraciones numéricas y estructurales) y transcriptómicas (en genes codificantes y no codificantes).

En el ámbito de la genómica funcional, se destacan los análisis de perfiles de expresión génica; éstos tienen como objetivo principal la identificación de un grupo de genes, cuyo patrón de expresión se encuentren asociados a un fenotipo en particular, concepto conocido como 'gene expression signature' o 'signature'. Un objetivo particular de los 'signatures' en el ámbito de la medicina, es su utilidad como biomarcador diagnóstico, pronóstico o predictivo de una patología en estudio. Los biomarcadores con valor pronóstico permiten estimar la evolución de la enfermedad permitiendo mejor la estratificación de pacientes según su pronóstico de progresión. Por otro lado, biomarcadores con valor predictivo permiten predecir si un tratamiento tendrá o no efecto en un paciente, logrando evitar tratamientos en pacientes para los cuales se supone no tendrán la respuesta esperada.

Objetivos

Debido a que es esencial ampliar el conocimiento de los mecanismos moleculares responsables de la progresión de lesiones pre-malignas a malignas, el cual es actualmente limitado. El objetivo principal del presente trabajo fue el de identificar subgrupos de carcinomas ductales *in situ* con riesgos de progresión diferencial al estadio invasivo de la enfermedad mediante una aproximación en genómica funcional.

Materiales y Métodos

Se estudiaron 30 CDIS de alto grado y sus respectivas muestras mamarias normales adyacentes al tumor. Estas muestras fueron obtenidas del Biobanco de tumores del MDACC. Además se analizaron 10 muestras de tejido mamario normal obtenidas de mamoplastias cosméticas del Cooperative Human Tissue Network para su poste-

rior uso en la identificación de genes diferencialmente expresados y metilados respecto a los CDIS. El ADN y ARN se extrajo mediante el empleo de TRIZOL (Life Technologies) y el kit Rneasy y Dneasy Blood and Tissue Kit (Qiagen, Germantown, MD). Las concentración e integridad del ADN y ARN se determinó con el Bioanalyzer Agilent 2100 (Agilent Technologies; Santa Clara, CA). Las muestras de ARN con valores de integridad (RIN) superiores a 8.0 fueron consideradas para su subsecuente secuenciación mediante el protocolo de Illumina TruSeq RNA Sample Prep Kit. La secuenciación se realizó mediante la plataforma Illumina HiSeq2000. El control de calidad de los archivos FASTQ se realizó mediante el empleo del programa FASTQC. Se utilizó software TRIMMOMATIC cuando se requirió la edición de las mismas (ejemplo: remoción de los adaptadores, etc.). Posteriormente, las lecturas fueron alineadas contra el genoma de referencia UCSC Hg19 mediante el paquete de R/Bioconductor Rsubread. Luego se emplearon los paquetes GenomicFeatures y GenomicAlignments para el conteo de cada mapeo en el transcriptoma de referencia (TxDb.Hsapiens.UCSC.hg19.knownGene). El análisis estadístico de los perfiles de expresión génica se realizó mediante el empleo del paquete de R/Bioconductor edgeR para identificar los genes diferencialmente expresados entre grupos ($FC > \pm 2$; $FDR < 0.05$). Las visualizaciones y análisis de enriquecimiento funcional se realizaron con el programa MultiExperiment Viewer software (MeV v4.9) y las aplicaciones ClueGo y CluPedia en la plataforma Cytoscape. Los detalles metodológicos se pueden encontrar en la referencia de Abba y col 2015 (Abba y otros, 2015).

Resultados y Discusión

El análisis del exoma, transcriptoma y metiloma de CDIS nos permitió generar el primer catálogo completo de los eventos epigenéticos, genéticos y vías de señalización afectadas en el estadio pre-invasivo del

cáncer de mama. Dicho estudio nos permitió identificar dos grupos de CDIS, uno de alto riesgo de progresión (CDIS-AR) y otro de riesgo moderado de progresión (CDIS-MR) que se diferencian por los subtipos intrínsecos que comprenden, el índice proliferativo, la respuesta inmune y la actividad de determinadas vías de señalización.

Según Cáncer Genome Atlas Network, las mutaciones más prevalentes en genes 'conductores' de la progresión tumoral fueron en *PIK3CA* para los CDIS-MR y en *TP53* para los CDIS-AR de manera similar a lo previamente descrito para carcinomas ductales infiltrantes. Los genes más frecuentemente alterados por mecanismos de amplificación génica fueron en orden decreciente: *ERBB2*, *VEGFA*, *MYC*, *AURKA*, *MDM2*, *FGFR1* y *CCND1*.

A nivel transcriptómico, se identificaron 885 genes diferencialmente expresados entre CDIS-AR y CDIS-MR (FDR < 0.01; log₂FC > 1). El data mining de dichos genes desregulados permitió asociar a los CDIS-AR con vías de señalización altamente proliferativas, infiltrado linfocitario característico de anergia e inmunotolerancia. Mientras que los CDIS-MR expresan las vías de señalización asociadas a la respuesta del estradiol – RE.

Respecto de las alteraciones epigenéticas más relevantes se pudo identificar hipermetilación del gen *HOXA5*, factor de transcripción de *TP53*, en los CDIS-AR. Esta hipermetilación en la región proximal del promotor de *HOXA5* correlacionó con la pérdida de expresión cuando se compararon dichos tumores con el tejido normal de referencia. Otros genes muy relevantes en la biología de las células epiteliales de mama afectados por hipermetilación de sus regiones CpG fueron: *SOX10*, *SOX15* y *SOX17* los cuales fueron previamente caracterizados como genes supresores de tumores por ser antagonistas de la vía de señalización WNT/B-cateninas.

La tecnología actual para el estudio del transcriptoma y proteoma humano ha puesto en evidencia que aproximadamente el 75% del genoma se transcribe, aunque solo el 2% de dichos transcritos codifican proteínas. El 73% restante del transcriptoma corresponde a ARN

no codificantes (ARNnc) que se clasifican en dos grandes grupos: los ARNnc pequeños (snRNA, snoRNA, tRNAs, microRNAs, etc.) y los recientemente descritos ARNlnc. Los ARNlnc son definidos como ARN de más de 200 nucleótidos de longitud sin un marco de lectura abierto (Prensner, 2011). El consorcio GENCODE ha identificado más de 14.000 ARNlnc, los cuales pueden clasificarse según su localización como: ARNlnc génicos (ejemplo: ARNlnc antisentido) o intergénicos (ARNlnc LINC) (Derrien y otros, 2012) (Djebali y otros 2012). Estos exhiben una serie muy variada de funciones, pudiendo operar en cis como amplificadores de la transcripción de genes vecinos, o bien en trans reclutando complejos modificadores de la cromatina. A su vez, pueden actuar como esponjas de microRNAs, como correguladores de receptores nucleares u operar como andamios en procesos involucrados con la maduración y traducción de ARNm (Geisler, 2013). Aunque el conocimiento sobre la biología de los ARNlnc es muy limitado, los ARNlnc están comenzando a ser vinculados con procesos tales como el desarrollo normal (H19, Xist), la organogénesis, la proliferación, la diferenciación y la apoptosis celular (Geisler, 2013). Recientemente, dos estudios independientes señalan que los perfiles de expresión de los ARNlnc tendrían el potencial de estratificar a los pacientes con cáncer de mama en subgrupos con valor pronóstico y en asociación al subtipo intrínseco tumoral (Su, 2014) (Shen y otros 2015). Se han podido identificar ARNlnc específicos en el cáncer de mama con características oncogénicas, posibles biomarcadores de progresión tumoral y/o blancos terapéuticos (*HOTAIR*, *LSINCT5*, *SRA1*, *UCA1* y *MALAT1*); así como, ARNlnc que actúan como genes supresores de tumores (*HOTAIRM1*, *ZFAS1*, *PTENP1*, *MEG3* y *GAS5*). Estos estarían involucrados en la proliferación, motilidad, invasión y ciclo celular, así como en la diseminación tumoral metastásica, todos procesos celulares de importancia en el desarrollo neoplásico (Shore, 2014).

Entre los hallazgos más relevante se pudo identificar un grupo de 193 ARNlnc diferencialmente expresados entre muestras normales de mama y CDIS ($p < 0.01$, $FDR < 0.01$), 127 de los cuales correspondieron a ARNlnc génicos antisentido y 66 a ARNlnc intergénicos (Figura 3). Aproximadamente el 90% de estos ARNlnc (168 de 193) se encontraron sobreexpresados en CDIS. Entre los ARNlnc desregulados en CDIS, cabe mencionar a *HOTAIR* (HOX transcript antisense RNA) y *HOTAIRM1* (HOXA transcript antisense RNA myeloid-specific 1). *HOTAIR* es un promotor de la diseminación metastásica del cáncer de mama, actuando en trans sobre un grupo de genes de la familia *HOXD* y se encontraría sobreexpresado en la mayoría de las neoplasias malignas humanas (Gupta y otros 2010) (Sorensen y otros 2013). Por otro lado, *HOTAIRM1* actúa en cis reprimiendo la expresión de los genes de la familia *HOXA*. Nuestros datos demuestran que *HOTAIR* se encuentra sobreexpresado en CDIS, mientras que *HOTAIRM1* es subexpresado en los estadios pre-invasivos del cáncer de mama. De modo que estos ARNlnc podrían modular conjuntamente la expresión de los genes de la familia *HOX* durante los estadios tempranos del cáncer de mama (Abba, 2015). Los otros ARNlnc identificados como desregulados en CDIS, constituyen un grupo novedoso de transcriptos cuya función celular es desconocida. El análisis de enriquecimiento funcional de los genes co-expresados con dichos ARNlnc permiten inferir funciones asociadas a la regulación del ciclo celular, y las vías de señalización de TP53 y diversas MAPKs mediante interacción con factores de transcripción (Figura 3).

Conclusiones

Del presente trabajo se desprenden las siguientes conclusiones:

- 1) Existen dos subgrupos de carcinomas ductales in situ que poseen alteraciones epigenómicas, genómicas y transcripcionales diferenciales.

2) Cada subgrupo de carcinomas ductales in situ posee diferentes riesgos de progresión al estadio invasivo.

3) Se identificaron biomarcadores con potencial valor pronóstico.

4) Las vías de señalización diferencialmente activadas entre subgrupos de carcinomas podrían ser utilizadas como nuevas dianas terapéuticas.

Referencias

1. Kerlikowske, K. (2010). Epidemiology of ductal carcinoma in situ. *Journal of the National Cancer Institute Monographs*, (41):139-41.
2. Allred, D.C. (2010). Ductal carcinoma in situ: terminology, classification, and natural history. *J Natl Cancer Inst Monogr*, 134-8.
3. Bleyer, A., Welch, H.G. (2012). Effect of three decades of screening mammography on breast-cancer incidence. *The New England Journal of Medicine*, 367:1998-2005.
4. Perou, C.M., Sørlie, T., Eisen, M.B. y otros. (2000). Molecular portraits of human breast tumours. *Nature*, 17; 406(6797):747-52.
5. Quackenbush, J. (2011). The Human Genome: Book of Essential Knowledge. Imagine!
6. Abba, M.C., Gong, T., Lu, Y. y otros. (2015) A molecular portrait of high-grade ductal carcinoma in situ (DCIS). *Cancer Res*, 75:3980-90.
7. Cancer Genome Atlas Network (2012) Comprehensive molecular portraits of human breast tumours. *Nature*, 490:61-70.
8. Prensner, J.R., Chinnaiyan, A.M. (2011). The emergence of lncRNAs in cancer biology. *Cancer Discov*, 1:391-407.
9. Derrien, T., Johnson, R., Bussotti, G. y otros. (2012). The GENCODE v7 catalog of human long noncoding RNAs: analysis of their gene structure, evolution, and expression. *Genome Res.*, 22:1775-1789.

10. Djebali, S., Davis, C.A., Merkel, A. y otros. (2012). Landscape of transcription in human cells. *Nature*, 489:101-8.
11. Geisler, S., Coller, J. (2013). RNA in *unexpected places: long noncoding RNA functions in diverse cellular contexts*. *Nat Rev Mol Cell Biol.*, 14:699–712.
12. Su, X., Malouf, G.G., Chen, Y. y otros. (2014). Comprehensive analysis of long non-coding RNAs in human breast cancer clinical subtypes. *Oncotarget*, 5:9864-76.
13. Shen, X., Xie, B., Ma, Z. y otros. (2015 Jun 10). Identification of novel long non-coding RNAs in triple-negative breast cancer. *Oncotarget*.
14. Shore, A.N., Rosen, J.M. (2014). Regulation of mammary epithelial cell homeostasis by lncRNAs. *Int J Biochem Cell Biol.*, 54:318-30.
15. Gupta, R.A. y otros. (2010). Long non-coding RNA HOTAIR reprograms chromatin state to promote cancer metastasis. *Nature*, 464:1071-6.
16. Sørensen, K.P. y otros. (2013). Long non-coding RNA HOTAIR is an independent prognostic marker of metastasis in estrogen receptor-positive primary breast cancer. *Breast Cancer Res Treat*, 142:529-36.

Leyendas de las figuras

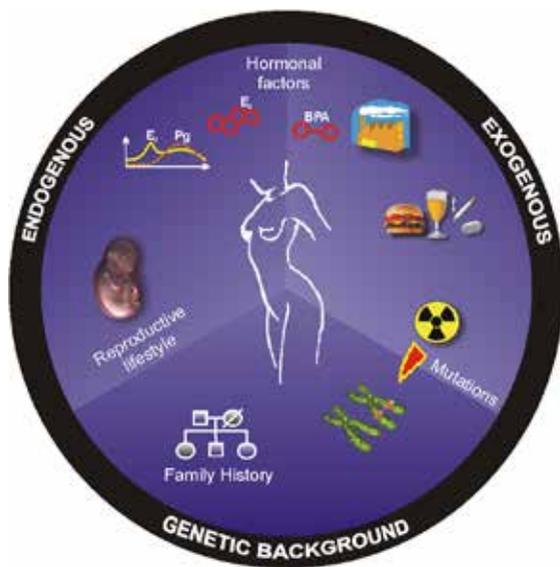


Figura 1. Esquema de los factores exógenos, endógenos y genéticos asociados al desarrollo del cáncer de mama.



Figura 2. Tecnologías genómicas disponibles para la caracterización de epigenomas, genomas y transcriptomas.

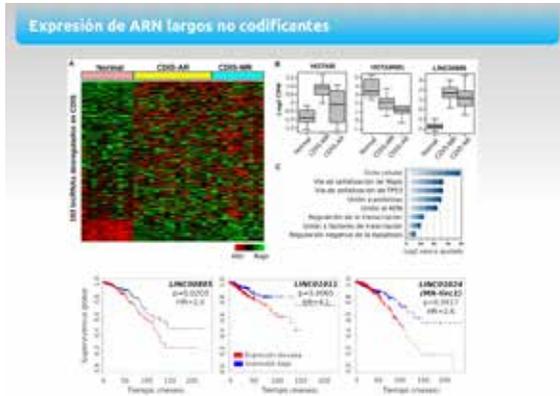


Figura 3. Identificación de ARNlnc desregulados en CDIS. **A.** Mapa de calor de los 193 ARNlnc diferencialmente expresados entre epitelio mamario normal y CDIS (FDR<0.01). **B.** Expresión relativa de los ARNlnc *HOTAIR* (p=2,63E-9), *HOTAIRM1* (p=1,65E-7) y *LINC00885* (p=4,09E-16) en tejidos normales y CDIS. **C.** Análisis de enriquecimiento funcional de los bioprocesos y vías de señalización asociados a los ARNlnc desregulados en CDIS identificados en función a los transcritos codificantes co-expresando en carcinomas mamarios. Análisis de supervivencia de pacientes con carcinomas infiltrantes de mama en función de la expresión de 3 ARNlnc del tipo intergénicos (LINC).