

Modelling Breast Adenocarcinomas In Situ with 3D Cellular Automaton: A Parallel Approach

A. Tomeu, A. Salguero, S. Zaldívar, and J. Aparicio

Abstract—Adenocarcinomas are solid tumors that begins in the duct architecture of the endocrine glands in human body, constituting some of the most frequent tumors (breast or prostate), with high morbidity and mortality, and treatment costs in constant growth for public health systems. This work starts from a mathematical model known and contrasted in the literature for breast adenocarcinoma in situ (DCIS), and aims to perform the implementation with a 3D cellular automata and parallel processing, to help a better understanding of the pathogenesis of the disease. We describe the biology of this class of tumors and the parallel implementation methodology used, which employs parallelism of data, locks on access to data shared between tasks, and dynamic management of the simulated tissue domain. The results obtained by running the proposed parallel simulation are discussed in terms of their consistency with the histological reality of the real tumor, with the kinetics of Gompertz's function for tumor growth, and with the statistical distribution of tumor cells in a mammary duct with disease in situ, with reasonable times and speedups. The conclusions establish the achievement of the proposed objective, compare the approach developed with other similar ones already published, and establish our future work.

Index Terms—Adenocarcinomas, Cellular automaton, DCIS, Duct, Mutual exclusion, Gland, Parallel programming, Speedup.

I. INTRODUCCIÓN

SE estima [8], [14], [28] que una de cada ocho mujeres sufrirán un cáncer de mama, de los cuales aproximadamente el 80% serán ductales, con origen en los conductos galactóforos del parénquima glandular. Igualmente, aproximadamente uno de cada nueve hombres [9] desarrollará un cáncer de próstata en las últimas décadas de su vida. Se constata pues, sin lugar a dudas, que la incidencia de los tumores glandulares (adenocarcinomas) en la población constituye un enorme problema de salud tanto en términos de morbi-mortalidad, como en términos de gasto sanitario. El despistaje y detección de estos tumores en sus etapas iniciales [8], [18], [20], cuando se desarrollan en el epitelio interno de los conductos glandulares pero aún no han infiltrado el parénquima glandular (DCIS, Fig. 1) es de enorme interés como estrategia de salud pública. Esto va a permitirnos curar la enfermedad con una cirugía limitada al conducto afectado, junto a un margen adicional de seguridad de tejido sano, con

tasas de éxito superiores [28] al 90%. Para el DCIS, el despistaje mediante técnicas de imagen ha demostrado ser una eficazísima estrategia de detección de la enfermedad in situ, mientras que, para el adenocarcinoma de próstata, las expectativas puestas en el antígeno prostático específico (PSA) como marcador sérico precoz de la enfermedad han sido, desgraciadamente, sobredimensionadas [9].

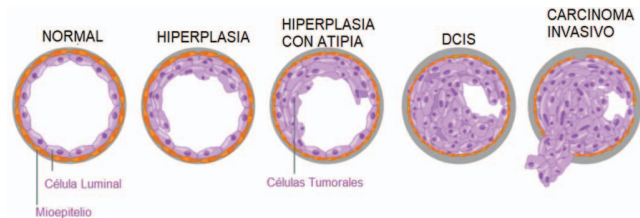


Fig. 1. Historia natural del adenocarcinoma de mama in situ (DCIS). La enfermedad parte de un conducto con estructura de células en doble capa, y tras originarse la enfermedad mediante mutaciones genómicas, pasa por fases intermedias de hiperplasia, para acabar en enfermedad neoplásica que invade la luz de conducto. Si no se trata en este punto, se infiltra el parénquima.

En ambos casos, el conocimiento y caracterización de la biología del DCIS en sus etapas iniciales, y de su historia natural (Fig. 1), se antoja de enorme interés para los especialistas en biología tumoral. Se hace necesario volver la mirada hacia la carcinogénesis, como fenómeno donde una o varias mutaciones sobre ciertos genes, permiten a las células normales de la doble capa que forman la sección transversal de los conductos de la mama, sobrevivir y reproducirse de manera incontrolada, bajo un proceso de selección que tras invadir los conductos glandulares, se transforma en infiltrante. El uso de modelos matemáticos para estudiar la carcinogénesis del DCIS ha sido amplio y está bien establecido [1], [2], [3], [6], [7], [10], [17], [21], [22]; el enfoque de modelado de tumores con autómatas celulares, también [5], [12], [13], [16], [17], [24]. El trabajo tiene como objetivo presentar el desarrollo de un modelo computacional paralelo tridimensional del DCIS, partiendo de un modelo matemático conocido y contrastado en la literatura [17] para dos dimensiones. Se pretende que el modelo sea de utilidad a los especialistas en biología tumoral, al permitirles simular de forma rigurosa y eficiente la carcinogénesis del DCIS desde un conducto sano a otro con DCIS, parametrizando en el mismo según necesiten los diferentes genes implicados que sufren mutaciones, la tasa de mutaciones que estos sufren, o la probabilidad de apoptosis para las células neoplásicas en presencia de fármacos citotóxicos. Esto puede redundar eventualmente, gracias a la simulación *in silico*, en una reducción de tiempos y costes de laboratorio.

A. J. Tomeu, A. G. Salguero and S. Zaldívar are with the Computer Science Department in the University of Cádiz (Spain). J. Aparicio is with the Pathology Department in the University of Cádiz (Spain); CA 11519, Spain. e-mail: {antonio.tomeu, alberto.salguero, szaldivar, jose.aparicio}@uca.es.

La estructura del trabajo es la siguiente: la sección II describe algunos elementos básicos de la biología tumoral del DCIS. El modelo computacional propuesto utiliza autómatas celulares, descritos en la sección III. La sección IV describe el modelo de DCIS con autómatas celulares tridimensionales que proponemos, y la sección V estudia la implementación paralela, identifica también las necesidades de sincronización y las resuelve. La sección VI describe los resultados, valida el modelo, y expone las medidas realizadas con una implementación del modelo en C++, sobre el clúster de procesadores de nuestra Universidad, derivando finalmente los tiempos y speedup medios, que ilustran una buena ganancia de rendimiento para la simulación paralela frente a la secuencial. La sección VII recoge las conclusiones.

II. BIOLOGÍA DEL DCIS

Los DCIS se gestan en el interior de los túbulos que permiten a la glándula enviar su secreción al medio. Para el caso de la mama humana, la histología muestra que estos conductos, en su estado normal, están compuestos (Fig. 1) por dos capas de células: la capa más interna está formada por células luminales, y está envuelta por una segunda capa de células mioepiteliales la que, a su vez, está envuelta por una membrana basal [14], [18]. Las dos clases de células derivan de una única clase de células progenitoras stem (pueden diferenciarse en varios subtipos celulares), mediante un mecanismo de diferenciación bajo control de los genes implicados en el proceso (Fig. 2) que partiendo de ellas, permite la existencia de dos líneas germinales que concluyen en células diferenciadas bien mioepiteliales, bien luminales, formando la estructura de doble capa descrita que se observa con microscopía óptica (Fig. 3-a).

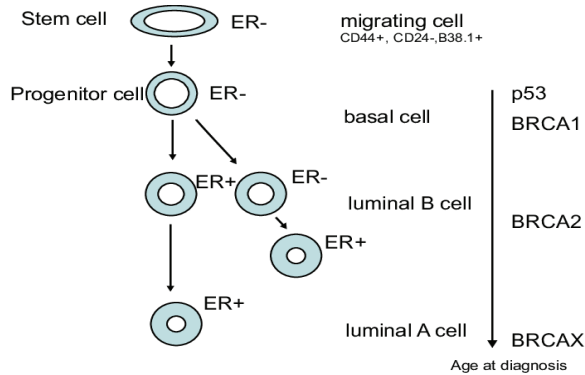


Fig. 2. Jerarquía de diferenciación de las células del conducto de una mama humana. Partiendo de células progenitoras stem, y mediante un proceso regulado por factores de crecimiento expresados por los genes que se indican, se llega a las células mioepiteliales y luminales que conforman un conducto de histología normal en doble capa de células.

Los DCIS comienzan [8], [14], [15], [18] generalmente con una (o varias mutaciones) en los genomas de estas células, aunque finalmente son las células luminales donde tiene lugar la transformación neoplásica, que inicialmente invaden la luz del ducto rompiendo la arquitectura normal de doble capa (hiperplasia), para luego ir llenando esa luz (Fig. 3-b), y finalmente traspasar la membrana basal e invadir el parénquima glandular; en ese momento la enfermedad adquiere carácter infiltrante, aparece la posibilidad de

procesos metastásicos por vías hematógica y/o linfática, y la morbi-mortalidad de las enfermas y los costes de tratamiento se elevan considerablemente [28]. La Fig. 2 ilustra la jerarquía de células que partiendo de las progenitoras stem, se diferencian en células mioepiteliales y luminales. Esta diferenciación es fruto de diferentes factores de crecimiento, en los que están implicados los genes que los expresan y que regulan el proceso [8], [20].

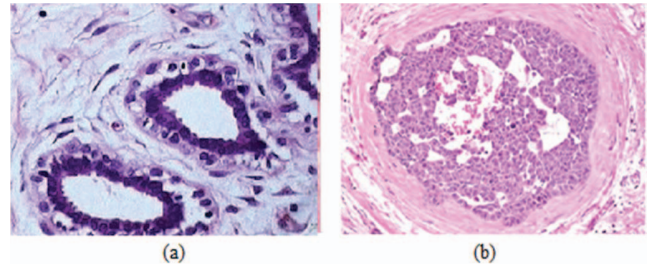


Fig. 3¹. Arquitectura histológica normal (a) de la mama humana. Se aprecian dos conductos en sección transversal. Las células luminales que forman el anillo interno, han captado mayor cantidad de contraste durante la tinción, y se ven más oscuras. Las células mioepiteliales, más claras, las rodean. La estructura en doble capa (comparar con la Fig. 1) es apreciable. En (b) se muestra un conducto con DCIS, con la luz invadida por células neoplásicas.

De entre todos los genes que constituyen el genoma de las células mamarias, la investigación [8], [14], [15], [18], [20] ha puesto de manifiesto que el DCIS está siempre ligado a alteraciones en uno o en varios de los genes *BRCA1*, *BRCA2*, *PTEN*, *TP53*, que son los implicados en los procesos de diferenciación y de mitosis de las células que forman los conductos, durante la cascada de etapas descritas por la Fig. 2. Son estos los genes que consideramos en la simulación. La Tabla I describe, para cada uno de estos genes, qué acción desarrollan. Cuando los genes sufren mutaciones acumuladas, los mecanismos de regulación celular (para la apoptosis, la mitosis o la migración quedan) seriamente afectados, y el DCIS se origina [8], [14], [20].

TABLA I
FUNCIONAMIENTO DE LOS GENES DEL DCIS

GEN	Funcionamiento Fisiológico
BRCA1	Controla el ciclo celular normal
BRCA2	Permite la división mitótica normal
PTEN	Activa la apoptosis si hay mutaciones malignas Inhibe la reproducción neoplásica si aparece
TP53	Apoptosis si se producen daños en protooncogenes Apoptosis si se altera la arquitectura de doble capa

III. AUTÓMATAS CELULARES

Existen múltiples definiciones del concepto de autómata celular en la literatura. Nosotros escogemos y adaptamos la establecida en [5], que ya hemos aplicado anteriormente a la simulación tumoral en [21], [24], [25] y en menor medida, en [26]. Definimos aquí un autómata celular (AC) como una estructura de 4-upla $(\zeta, \varepsilon, N^1, \rho)$ donde el significado de los elementos es el siguiente:

- ζ es un retículo discreto regular de células (nodos o celdas) junto con algún conjunto de condiciones de

¹ La Figura 3 (a y b) está tomada de la obra *Patología Estructural y Funcional*, de Kumar, V., Abbas, A. y Fausto, N., 7ª edición. Elsevier, 2006.

frontera,

- ε es un conjunto finito (habitualmente con estructura de anillo abeliano) de estados que las células del retículo pueden adoptar,
- N^I es el conjunto de células que definen la vecindad de células con las que interactúa una célula dada, y
- ρ es una función de transición que especifica cómo una célula del retículo cambia de estado en función del tiempo y del estado de la vecindad N^I .

Dado lo anterior, un espacio de células puede definirse como un retículo ζ que cubre de forma homogénea una porción del espacio euclídeo d -dimensional. Cada célula está etiquetada por su posición $r \in \zeta$. La disposición espacial de las células está especificada por las conexiones con sus vecinos más cercanos, y se obtienen uniendo pares de células en alguna disposición regular. Para cualquier coordenada espacial r el retículo de vecindad $N_b(r)$ es una lista de células vecinas definidas por

$$N_b(r) = \{r + c_i : c_i \in N_b, i = 1, \dots, b\} \quad (1)$$

Aquí b es el número de coordinación o, dicho de otra forma, el número de vecinos próximos en el retículo que interactúan con la célula ubicada en la coordenada r . Con $N_b(r)$ denotamos la plantilla de vecinos próximos con elementos $c_i \in \mathbb{Z}^d$, para $i = 1, \dots, b$. En el caso que nos ocupa, y para las secciones transversales del conducto, escogemos $d = 2$, mientras que para el modelo completo del conducto, tomamos $d = 3$, de manera que

$$\zeta = \{r : r = (r_1, r_2, r_3), r_j \in \mathbb{Z}^3, j = 1, 2, 3\} \quad (2)$$

El número total de células disponibles lo notaremos habitualmente por $|\zeta|$. En simulaciones de computador, los AC utilizan retículos que han de ser necesariamente finitos ($|\zeta| < \infty$), y deben imponerse condiciones de frontera que establezcan los vecinos de aquellas células situadas en las fronteras del retículo. Nosotros utilizaremos en general condiciones de frontera nula². El conjunto de células vecinas cuyo estado influye en una dada, queda definido mediante la vecindad de interacción $N_b^I(r)$ para una célula r dada, de acuerdo a la siguiente expresión:

$$N_b^I(r) = \{r + c_i : c_i \in N_b^I\} \quad (3)$$

Por otra parte, cada célula $r \in \zeta$ tiene asignado un estado $s(r) \in \varepsilon$. Los elementos del conjunto ε pueden ser números, letras o símbolos. Nosotros escogeremos un conjunto ε de cinco elementos para modelar todos los tipos celulares presentes en el DCIS. Una configuración global del autómata $c \in \varepsilon^{|\zeta|}$ queda determinada por el estado de todas las células del retículo y ofrece una instantánea del estado del tumor en

un instante dado del tiempo. Finalmente, la dinámica temporal del modelo viene determinada por la función de transición ρ , que especifica cómo una célula cambia de estado en función de la interacción con su vecindad:

$$\rho : \varepsilon^\mu \rightarrow \varepsilon \quad (4)$$

en la que $\mu = |N_b^I|$. La regla es espacialmente homogénea, y no depende por tanto en forma explícita de la posición r de una célula dada. Extensiones de la definición dada para incluir inhomogeneidades espaciales o temporales son factibles. Si el AC es determinista, la función de transición lleva a un único cambio de estado posible, mientras que si es estocástica, el nuevo estado de una célula depende de alguna distribución de probabilidad.

IV. MODELO DE DCIS CON AUTÓMATAS CELULARES

Para modelar un conducto de la mama humana, y con base en la información histológica conocida [7], [13], [14], [15], [18] y [20], y en el modelo bidimensional de [17], asumiremos una retícula ζ de $20 \times 20 \times 200$ nodos, construida a partir del modelo propuesto en [17], [25] y en menor medida en [26] añadiendo una dimensión adicional. Se han utilizado condiciones de frontera nula. Cada nodo contiene un número r de \mathbb{N} , que almacena información sobre el tipo de célula que ocupa el nodo, y sobre la carga genética de la célula. Si bien es cierto que una célula del conducto de una mama humana posee múltiples genes formados por millones de bases de ADN, nuestro modelo empleará un enfoque reduccionista, y se limitará a utilizar únicamente los cuatro genes para los que existe evidencia científica contrastada [8], [14], [20] de que participan en la patogénesis del DCIS: *BRCA1*, *BRCA2*, *PTEN*, *TP53*. Cada gen será codificado en el modelo por un entero de 4 bytes. La carga genética de una célula del modelo queda así definida por la 4-upla (*BRCA1*, *BRCA2*, *PTEN*, *TP53*), que es codificada en un único número natural utilizando la función de emparejamiento dada por la siguiente ecuación:

$$\langle x|y \rangle = (2^x + y) \neg 1 \quad (5)$$

Esta función³, que es una biyección, puede ser anidada para codificar de forma compacta y razonablemente eficiente el genoma g de una célula del modelo, como sigue:

$$g = \langle \langle BRCA1|BRCA2 \rangle | \langle PTEN|TP53 \rangle \rangle \quad (6)$$

El espacio de estados de las células del modelo es el conjunto $\varepsilon = \{\text{free, basal, luminal, myoepithelial, stem}\}$, que se representará numéricamente proyectando ε sobre el anillo residual abeliano \mathbb{Z}_5 . Ahora, la información (s, g) que representa el estado s de una célula de la retícula ζ y su carga genética g codificada numéricamente, vuelven a codificarse, para obtener el entero r que se sitúa en el nodo de la retícula. La decodificación del citado número para

² Se considera que las células situadas en la periferia del retículo únicamente tienen células vecinas en el interior del mismo.

³ La función \neg es la resta restringida; $x \neg y$ es igual a $x - y$ si $x \geq y$, y 0 en otro caso.

modificar el estado de una célula o su carga genética (cuando por ejemplo sufre una mutación en su genoma), es trivial. La función de codificación $\langle x|y \rangle$ propuesta, y también las de decodificación se calculan con muy bajo costo computacional [4]. La retícula ζ es inicializada por un algoritmo SetUp completamente determinista, descrito por el diagrama de flujo de la Fig. 4.

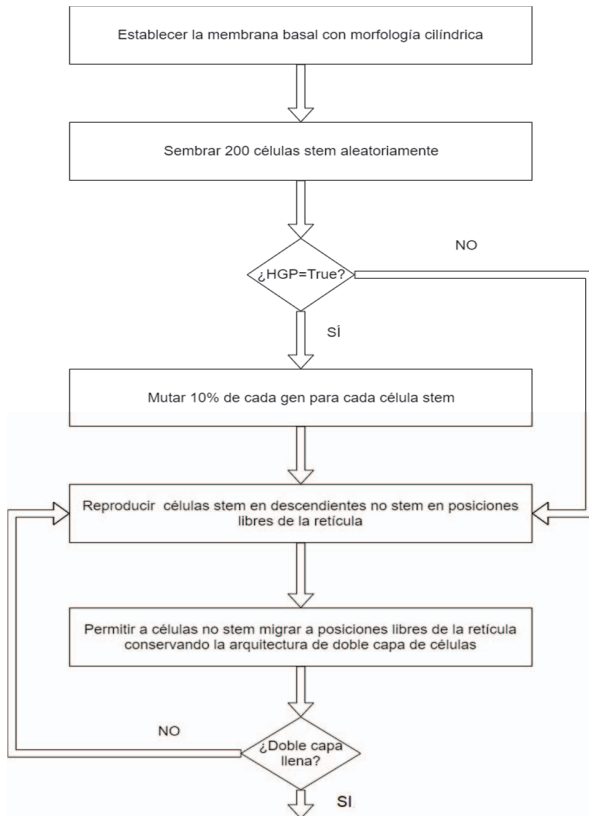


Fig. 4. Diagrama de flujo que ilustra las acciones desarrolladas por el algoritmo SetUp para configurar un conducto de estructura normal.

Posteriormente, se siembran dentro de la membrana basal células stem progenitoras con una topología cilíndrica, que se reproducen para dar lugar a un conducto con estructura de doble capa de células mioepiteliales y luminales. Durante la reproducción, las células hijas heredan una copia del genoma de las células padres, y ambos genomas son mutados con una probabilidad de 0.05, según ilustran los estudios *in vivo* e *in vitro* [14], [15], [20].

Si el componente de predisposición hereditaria [20] está presente en el algoritmo SetUp (HGP=true) entonces se mutan además [20] el 10% de todos los genes de todas las células stem de forma aleatoria y eventualmente, la transformación neoplásica se producirá antes. Al finalizar esta fase, cada sección transversal del conducto contiene aproximadamente de 45 a 50 células luminales y modela fielmente un conducto mamario normal [8], [14], [15], [18], [20]. En esta etapa de la simulación, todas las células que forman el conducto poseen un genoma que o bien aún no es patológico (representado por los cuatro bytes que codifican cada uno de los genes, con valores iguales a cero), o bien refleja mutaciones que ya se han producido a una tasa normal, o aumentada si HGP=true (quedan reflejadas en el

genoma cambiando el valor de uno o varios bits, en uno (o varios) genes, de cero a uno). En una segunda fase algorítmica que hemos denominado Evolve, los nodos de la retícula que modelan el conducto, son síncronamente actualizados nodo a nodo, de acuerdo con una función de transición ρ que considera el estado de las células vecinas y la probabilidad de mutación, todo lo cual se presenta mediante el diagrama de flujo de la Fig. 5.

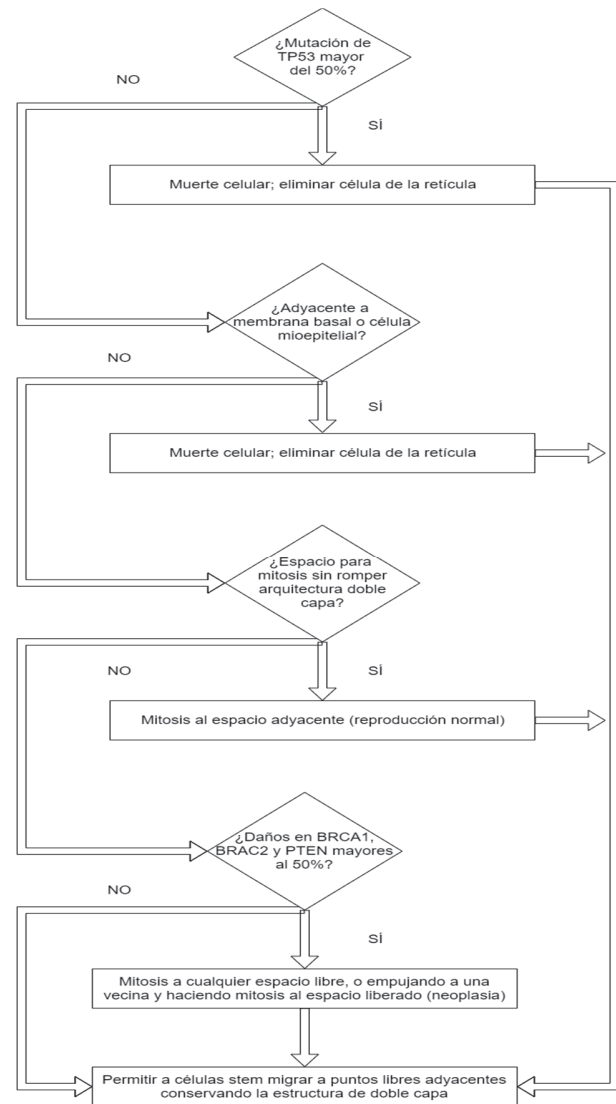


Fig. 5. Diagrama de flujo que ilustra las acciones desarrolladas por el algoritmo Evolve para modelar el comportamiento de las células de un conducto normal en función del tiempo. Cuando una célula efectúa una mitosis normal, su descendencia acumula errores en el genoma a una tasa del 0.05% por bit de genoma.

La selección y actualización de los 8×10^4 nodos de la retícula ζ configuran una generación, que como puede apreciarse fácilmente, supone una alta carga computacional que justifica el abordaje en paralelo del problema. El número de generaciones varía dependiendo de la longitud de la historia natural del tumor que se pretende simular, y que aumenta o disminuye el número de generaciones de la simulación. Para llegar desde un conducto normal a un punto donde la simulación llena la luz del conducto con

células neoplásicas sin romper la membrana basal son típicas simulaciones de al menos 15×10^3 generaciones. En el algoritmo Evolve se han tenido en cuenta tres situaciones principales:

- una célula dada muere cuando los mecanismos de control de la apoptosis funcionan normalmente frente a daños en el gen TP53, o bien la célula rompe la arquitectura normal del conducto.
- una célula efectúa una mitosis normal si algún punto adyacente de la retícula ζ está libre y no se rompe la estructura normal de doble capa del conducto; la célula hija recibe una copia del genoma de la célula madre, que puede estar dañado con una probabilidad de 0.05% por cada bit [14], según la información histológica *in vitro* e *in vivo* disponible [8], [14], [20].
- una célula efectúa mitosis anormal (neoplásica) si la célula tiene una tasa de mutaciones [14] en los genes BRCA1, BRAC2 y PTEN superior al 50% de las bases que lo forman; en ese caso, se escoge una de los posibles nodos libres de la retícula, y se sitúa a la célula hija, que recibe una copia del genoma mutado, sin tener en cuenta la estructura de doble capa normal del conducto, permitiendo la aparición en primer lugar de hiperplasia, y luego de neoplasia *in situ* (Fig. 1, y Fig. 3-b).

Finalmente, a todas las células no stem se les permite migrar de su posición actual en la retícula a otra diferente situada en su vecindad de *Moore* (definida como el cubo de células que rodean a una dada), siempre que exista espacio para ello. Dicha migración tendrá lugar respetando la estructura normal de doble capa del conducto, o invadiendo la luz del mismo si la célula presenta una alta carga de mutaciones en su genoma

V. IMPLEMENTACIÓN PARALELA

El modelo analizado en el párrafo anterior fue implementado utilizando el lenguaje de programación C++, y paralelizando la versión secuencial inicial. La paralelización empleó los principios del multiprocesamiento simétrico con un abordaje de división del dominio reticular ζ en la dimensión longitudinal (ver al respecto [11], [16], [17], [19], [21], [24], [25]) z de la misma en sub-retículas de topología cúbica (Fig. 6) que fueron procesadas por tareas diferentes, cada una ejecutándose sobre un core dedicado. Con el objetivo de aumentar el rendimiento de la implementación paralela, se utiliza una única vista de la retícula ζ . El prototipo de las tareas que implementan el código del algoritmo Evolve de forma paralela es el siguiente:

```
void parallelTask(int Inicio, int Fin);
```

Como se aprecia en la Fig. 6 con más detalle, cada tarea recibe los puntos de la dimensión z de la retícula ζ donde comienza (Inicio) y termina (Fin) la sub-retícula tridimensional que le corresponde procesar. A partir del prototipo indicado, las tareas fueron creadas mediante el uso de objetos de la clase `<thread>`, parametrizados [27] con la

función `parallelTask`. El enfoque planteado puede generar problemas de coherencia en la simulación [19], [21], cuando dos hebras diferentes procesando a células situadas a ambos lados de una frontera común a dos sub-retículas, necesitan que la célula que respectivamente procesan migre, o se reproduzca a un mismo lugar.

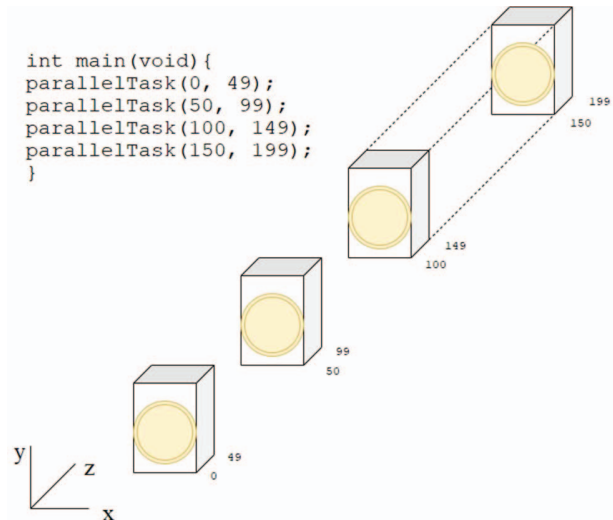


Fig. 6. División de la retícula tridimensional en sub-retículas procesadas por cuatro tareas paralelas diferentes. Cada sub-retícula se representa por un cubo, y es procesada a través de una tarea `parallelTask`. El programa principal lanza las cuatro tareas paralelas, y cada una procesa el sub-retículo de datos que le corresponde.

La Fig. 7-a ilustra esto para dos dimensiones, siendo la extensión al caso tridimensional obvia. Las hebras fueron sincronizadas utilizando una barrera cíclica, y procesadas a través de un ejecutor de pool de hebras de codificación propia, al cual se delegó su ciclo de vida.

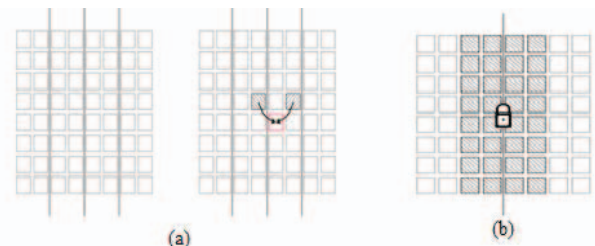


Fig. 7. (a) Situación (simplificada en dos dimensiones) donde dos tareas paralelas realizan procesos de migración o mitosis que intentan ocupar el mismo lugar de la retícula ζ , y que debe ser controlada adecuadamente. (b) Control con un cerrojo (b) de la región de la retícula donde los cambios afectan a dos tareas distintas, para garantizar la coherencia de la misma.

Para solucionar el problema de la coherencia de memoria, hemos empleado una estrategia de control bajo exclusión mutua que garantiza la coherencia de la información en puntos próximos a la frontera entre sub-retículas adyacentes. También hemos incorporado una estrategia de crecimiento dinámico del dominio tisular ζ simulado, para mejorar el rendimiento de la paralelización. Ambas estrategias se describen a continuación.

A. Sincronización en Zonas de Frontera

Un enfoque de control basado en el control de la exclusión mutua sobre la retícula ζ íntegra debe ser necesariamente

descartado de entrada, por ineficiente, ya que reduce el paralelismo real entre las tareas casi a cero. Un enfoque más adecuado estudiado en [21], [24], [26] es tomar exclusión mutua únicamente sobre las regiones fronterizas, que es donde verdaderamente reside el problema de la coherencia de datos, dejando que las tareas procesen libremente el resto de sus espacios de datos respectivos, donde no pueden darse conflictos. Definiremos una región de conflicto entre dos sub-retículas con frontera común, a las dos secciones transversales de cada sub-retícula más próximas a una frontera. Es decir, si dada la dimensión z de retícula ζ , existe una zona de frontera en la posición l , se define una región frontera como el conjunto de nodos $(i, j, l-2, l-1, l+1, l+2)$ de ζ . El esquema se ilustra más intuitivamente en la Fig. 7-b. para dos dimensiones, y proporciona un balance razonable entre la coherencia de la información de ζ , y un paralelismo entre tareas con las mínimas restricciones posibles. Ahora, basta establecer el acceso una región frontera bajo control de exclusión mutua mediante un cerrojo de clase `<lock>`, para garantizar la coherencia de la información. La decisión de una célula que decida reproducirse o migrar a un punto de una región frontera debe hacerse después de adquirir el cerrojo, momento en que tiene una visión actualizada del estado del conducto.

B. Crecimiento Dinámico del Dominio Tisular

Resulta obvio que tras aplicar el algoritmo `SetUp`, que construye el conducto con la topología cilíndrica de doble capa, no tiene sentido procesar a todos los nodos de la retícula ζ que lo modela, ya que la mayoría de ellas estarán vacías. Como estrategia de optimización, hemos introducido una subrutina adicional (que por propósitos de claridad no mostramos en el algoritmo `Evolve`), que en lugar de efectuar un procesamiento de todos los nodos de la retícula, procesa únicamente [24] los nodos de la retícula que tienen células activas (Fig. 8), y deja sin procesar a los nodos vacíos.

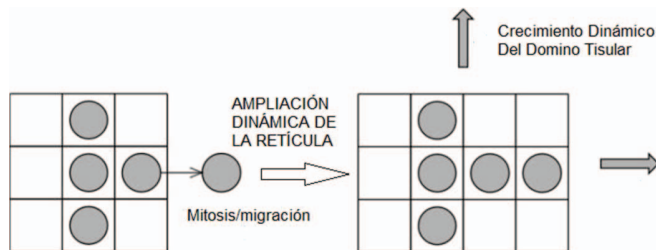


Fig. 8. Crecimiento dinámico del dominio tisular en dos dimensiones. El algoritmo `Evolve` utiliza esta estrategia generalizada a tres dimensiones, para evaluar únicamente aquellos nodos de la retícula ζ donde hay células activas.

Conforme el tumor se desarrolla, el marco de células procesadas va aumentando, aunque la retícula íntegra no se llega a procesar hasta las últimas generaciones de la simulación. La introducción de esta variante logra mejorar muy notablemente (15.72%) los tiempos respecto a la versión paralela previa que no la incorporaba.

VI. MEDIDAS Y DISCUSIÓN DE RESULTADOS

Las pruebas de ejecución se desarrollaron en cinco nodos diferentes del clúster de procesadores de nuestra Universidad.

Cada nodo tiene dos procesadores Intel © Xeon™ E5 a 2.6 GHz, que juntos rinden 20.8 GFlops, con 128 GB de memoria RAM sin hyperthreading. El nodo de entrada opera el HP Clúster Management Utility sobre Red Hat Enterprise Linux para HPC y los nodos de procesamiento la versión `Compute Node` del mismo sistema operativo. La versión del compilador GNU C++ fue 8.2 utilizada con el distintivo `-O3`.

A. Validación del Modelo

Para validar la implementación del modelo frente a la realidad biológica conocida [8], [14], [15], [18], [20] y modelos matemáticos establecidos previamente [7], [17] hemos tenido en cuenta los aspectos siguientes: la incidencia del DCIS en el modelo; con un daño inicial del genoma del 10% (HGP=true), es de esperar la presencia en el mismo de carcinogénesis temprana (la observamos, Fig. 9-inferior, a una media de 4703 ± 108 generaciones, frente a las 4536 ± 244 de [14]); para una simulación sin daño temprano (HGP=false) donde el cáncer aparece más tarde (lo observamos a una media de 6762 ± 70 generaciones, frente a las 6687 ± 45 de [14]), siendo el modelo en este aspecto, coherente con los resultados publicados; hemos considerado también la agresividad del DCIS: histológicamente, se mide con la tasa de actividad mitótica en %, que da lugar a un crecimiento más rápido del tumor a partir de las células luminales que predominan frente al resto de tipos celulares. Obtenemos en las etapas finales de la simulación tasas de actividad mitótica del 51.8%, frente al 48.93% de [14]. El histograma de la Fig. 9-derecha muestra cómo en la simulación esta estirpe de células (progenitoras luminales y luminales) predominan frente a las demás, tal y como [8], [18] y [23] demuestran en sus estudios; finalmente hemos estudiado con el modelo la cinética del DCIS; en [3], [8], [15], [20] se analiza esta cinética, que es de tipo gompertziano. La Fig. 9-superior muestra la curva del número de células en el conducto simulado, apreciándose cómo la simulación se ajusta a la cinética descrita.

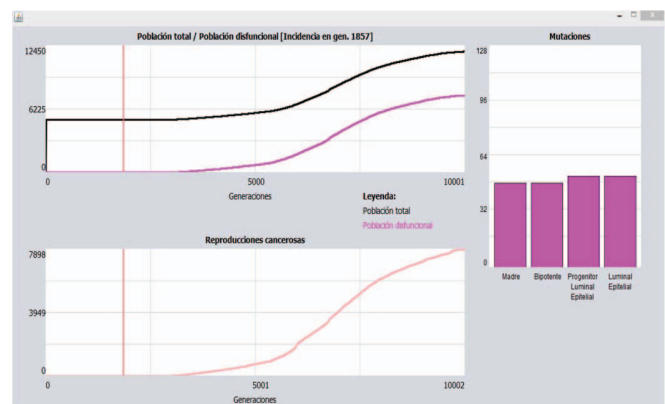


Fig. 9. Una parte de la interfaz gráfica de la simulación; superior, muestra la cinética tumoral en forma de curva de Gompertz; inferior, muestra el punto de aparición de carcinogénesis; derecha la población de células de cada uno de los tipos posibles en forma de histograma, a la derecha, con predominio de células luminales frente a las demás.

Por último, la Fig. 10 muestra el estado de cinco secciones del conducto (en horizontal) tomados en tres instantes de la

simulación ($10^2, 10^3, 10^4$ generaciones), y es posible apreciar cómo la simulación es estructuralmente coherente con la historia natural de la enfermedad (Fig. 1) y la información histológica (Fig. 3 a y b) conocida [8], [14], [18], [20].

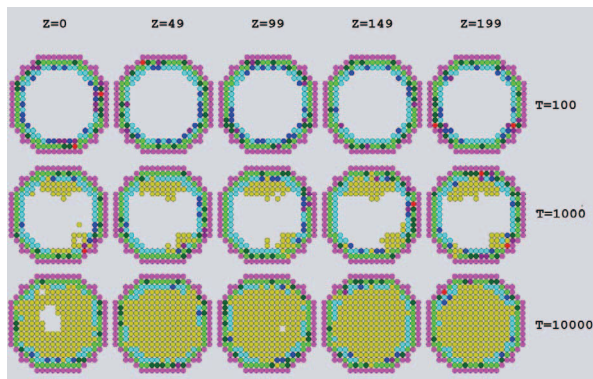


Fig. 10. Cinco secciones transversales diferentes del conducto en la dimensión z ($z = 0, 49, 99, 149, 199$) en horizontal. En vertical se muestra, para cada sección, el estado de la misma a las $10^2, 10^3, 10^4$ generaciones. Compárese con las Fig. 1 y Fig. 3-a-b.

B. Rendimiento de la Implementación Paralela

La Tabla II recoge un resumen estadístico (las Fig. 11 y 12 recogen la información completa) de los experimentos desarrollados para un número de tareas paralelas creciente, y un tamaño de la simulación constante de 15×10^3 generaciones; para cada número de tareas, la tabla ilustra los valores medios y las desviaciones estándares obtenidos tanto para el tiempo de ejecución como para el speedup. El tamaño de la muestra fue de $n = 16$ experimentos distintos (replicando cada experimento para obtener cierta fiabilidad estadística en cinco nodos distintos del clúster para un número de tareas creciente: 1, 2, 4, ..., 32), alcanzándose el tiempo mínimo y el speedup máximo respectivamente para un número de tareas paralelas igual diez.

Tareas	Tiempo (s.)		Speedup	
	μ	σ	μ	σ
1	412.50	3.60	1	0
4	112.94	0.07	3.70	0.28
6	87.37	8.25	4.81	0.47
10	69.12	1.84	6.02	0.16
14	83.82	0.90	4.96	0.05
16	104.12	3.22	4.00	0.12
20	123.87	0.89	3.36	0.02
24	124.07	2.20	3.36	0.05
26	133.89	3.53	3.11	0.08
30	132.88	1.95	3.13	0.04

La Fig. 11 presenta los tiempos obtenidos como resultado de procesar una simulación utilizando un número creciente de tareas paralelas. La curva ilustra los tiempos medios para los 16 experimentos, mientras que los puntos representan para cada experimento los tiempos individuales obtenidos en cada uno de los cinco nodos de nuestro clúster de procesadores.

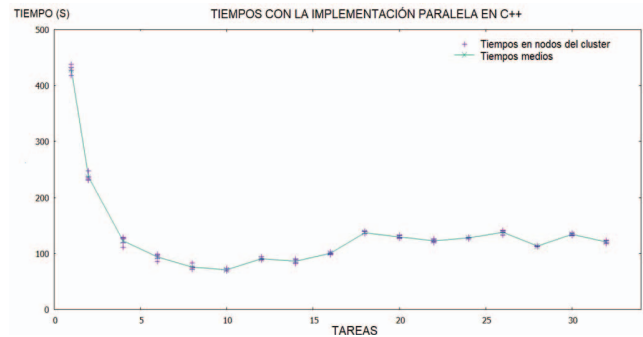


Fig. 11. Tiempos de ejecución de la simulación paralela con un número creciente de tareas y mínimo para diez tareas paralelas.

La Fig. 12 presenta el comportamiento del speedup para un número creciente de tareas, con igual tamaño de la simulación. Nuevamente representamos la curva de speedups medios y los puntos individuales con el speedup medido en cada uno de los cinco nodos. Puede apreciarse cómo se alcanza una aceleración máxima (y un tiempo de ejecución mínimo) de 6.02 para un número de tareas paralelas igual a diez. Incrementos posteriores en el número de tareas paralelas no aportan mejoras en los tiempos de ejecución ni en la aceleración; muy al contrario, se observa un ligero empeoramiento de los mismos, que se estabiliza posteriormente en una fase de meseta.

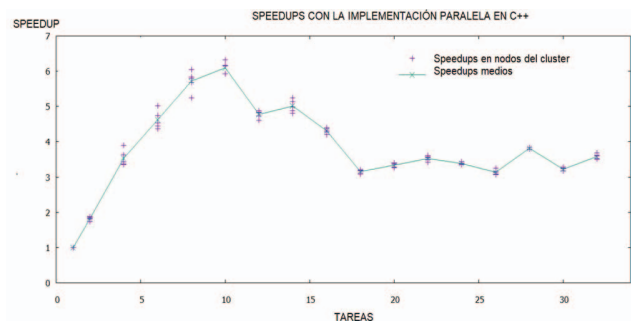


Fig. 12. Speedups de la simulación paralela con un número creciente de tareas, y máximo para diez tareas paralelas.

VII. CONCLUSIONES

En el trabajo hemos desarrollado un modelo computacional del DCIS a partir de un modelo matemático conocido [17] extendiéndolo a tres dimensiones, y analizado su validez y el impacto del procesamiento paralelo aplicado a la simulación de neoplasias glandulares in situ, con el objetivo de ofrecer una herramienta capaz de simular múltiples escenarios de carcinogénesis de forma rápida y eficiente. El modelo ha sido implementado en C++ con las capacidades concurrentes que el lenguaje ofrece desde la iteración de 2011 [27], y desarrollado una batería de experimentos en los nodos del clúster de procesadores de nuestra universidad. Los resultados obtenidos muestran coherencia de la simulación con el modelo matemático de partida [17] y con la realidad biológica de un conducto tridimensional [8], [14], [15], [18], [20] simulada en términos de la carcinogénesis, de la salida visual (Fig. 10) comparada con la histología (Fig. 3 a y b), de la cinética de crecimiento [8], [14] del tumor (Fig. 9, inferior), y también de

la tasa de actividad mitótica de las células presentes [8], [20] en el mismo (Fig. 9-derecha). La implementación paralela alcanza un speedup máximo de 6.02, que consideramos razonablemente bueno. La aceleración de las computaciones en esta clase de simulaciones resulta de especial interés, ya que permite a los especialistas en biología tumoral realizar una gran cantidad de experimentos *in silico* variando las condiciones a un coste despreciable, escogiendo solo los más prometedores para desarrollar *in vitro*, y reduciendo por tanto los costes globales de investigación. Nuestro trabajo futuro se centrará en desarrollar variaciones del modelo sobre otras neoplasias glandulares, y en el desarrollo de versiones de la paralelización desarrollada bajo arquitecturas GPU.

REFERENCIAS

- [1] Alarcon, T., Byrne, H., Owen, M., Webb, S. & Maini, P. *Modelling aspects of cancer dynamics: a review*. Philosophical Transactions of the Royal Society, vol 364, issue 1843, 2006.
- [2] Alrock, P., Liu, L. & Michor, F. *The mathematics of cancer: integrating quantitative models*. Nature Reviews Cancer vol. 15, pp. 730–745, 2015.
- [3] Byrne, H. *Dissecting cancer through mathematics: from the cell to the animal model*. Nature Reviews Cancer volume 10, pp. 221–230, 2010.
- [4] Martin Davis, Ron Sigal & Elaine Weyuker. *Computability, Complexity and Languages*. Academic Press, 1994.
- [5] Deutsch, A. & Dormann, S. *Cellular Automaton Modeling of Biological Pattern Formation*. Birkhäuser, 2005.
- [6] Edelman, L., Eddy, J. & Price, N. *In silico models of cancer*. Brugger, S.D. et al. Systems Biology and Medicine. Advanced Review, vol. 2, pp. 438-439, 2010.
- [7] Enderling, H., Chaplain, M., Anderson, A. and Vaidya, J. *A mathematical model of breast cancer development, local treatment and recurrence*. Journal of Theoretical Biology, vol. 246, issue 2, pp. 245-259, 2007.
- [8] Erbas, B., Provenzano, E., Armes, J. & Gertig, D. *The natural history of ductal carcinoma in situ of the breast: a review*. Breast Cancer Research and Treatment, vol. 97, issue 2, pp. 135-144, 2006.
- [9] European Association of Urology. *Policy Paper on PSA FOR PROSTATE CANCER*. Documento en PDF accesible en http://epad.uroweb.org/wp-content/uploads/EAU_policy-briefing_PSA.pdf, Consultado el día 4 de abril de 2019.
- [10] Gatenby, R. & Gawlinski, E. *A Reaction-Diffusion Model of Cancer Invasion*. Cancer Research, vol. 56, issue 24, pp. 5745-5753, 1996.
- [11] Giordano, A. et al. *Parallel Execution of Cellular Automata through Space Partitioning: The Landslide Simulation Sciddicas3-Hex Case Study*. Proceedings of 25th Euromicro International Conference on Parallel, Distributed and Network-based Processing (PDP), pp. 505-510, 2017.
- [12] Gevertz, J., Gillies, G. & Torquato, S. *Simulating tumor growth in confined heterogeneous environments*. Physical Biology, vol. 5, issue 3, 2008.
- [13] Kam, Y., Rejniak, K. & Anderson, A. *Cellular modeling of cancer invasion: Integration of in silico and in vitro approaches*. Journal of Cellular Physiology, vol. 227, issue 2, pp. 431-438, 2012.
- [14] Kumar, P. *Manual of Practical Pathology*, CBS Publisher & Distributors P Ltd, 2011.
- [15] Knudson, A. *Mutation and cancer*, Proc. Natl. Acad. Sci, 68, pp. 820-863, 1971.
- [16] Monteagudo, A. & Santos, J. *Studying the capability of different cancer hallmarks to initiate tumor growth using a cellular automaton simulation. Application in a cancer stem cell context*. Biosystems, vol. 115, pp. 46-58, 2014.
- [17] Norton, K., Winger, M., Bhanot, G., Ganesan, S., Barnard N. & Shinbrot, T. *A 2D mechanistic model of breast ductal carcinoma in situ (DCIS) morphology and progression*. Journal of Theoretical Biology, vol. 263, issue 4, pp. 393-406, 2009. *Rev. Microbiol.* 33, 183–209. 2007.
- [18] Pinder, S.. *Ductal carcinoma in situ (DCIS): pathological features, differential diagnosis, prognostic factors and specimen evaluation*. Modern Pathology, 23, S8-S13, 2010.
- [19] Pereira, Maelso and Junior, Josue and Ramos, Jorge, Azambuja Pagot, Christian, Nascimento, Tiago and Brito, Alisson. *Acceleration of Cellular Automata through Parallel Computing with OpenCL*. Proceedings of 8th Workshop on Applications for Multi-Core Architectures - WAMCA, At Campinas, Brazil, 2018.
- [20] Petridis, C. et al. *Genetic predisposition to ductal carcinoma in situ of the breast*. Breast Cancer Research, 2016. D.O.I.: <https://doi.org/10.1186/s13058-016-0675-7>
- [21] Salguero, A., Capel, M. and Tomeu, A.. *Parallel Cellular Automaton Tumor Growth Model*. Advances in Intelligent Systems and Computing book series (AISC, volume 803), 2018. D.O.I.: https://doi.org/10.1007/978-3-319-98702-6_21
- [22] Silva, A.; Gatenby, R., Gillies, R. & Yunesa, J. *A quantitative theoretical model for the development of malignancy in ductal carcinoma in situ*. Journal of Theoretical Biology, vol. 262, issue 4, pp. 601-613, 2009.
- [23] Sottoriva, A., Verhoeff, J., Borovski, T., McWeeney, S., Naumov, L., Medema, J., Slood, M. & Vermeulen, L. *Cancer Stem Cell Tumor Model Reveals Invasive Morphology and Increased Phenotypic Heterogeneity*. Cancer Research, vol. 70, issue1, pp 46-56, 2010.
- [24] Tomeu Hardasmal, A., Salguero Hidalgo, A. & Capel Tuñón, M. *Speeding Up Tumor Growth Simulations Using Parallel Programming and Cellular Automata*. IEEE Latin America Transactions, vol. 14, issue 11, pp.4603-4619, 2016. D.O.I.: [10.1109/TLA.2016.7795837](https://doi.org/10.1109/TLA.2016.7795837)
- [25] Tomeu Hardasmal, A., Salguero Hidalgo, A. & Capel Tuñón, M. *A Parallel Cellular Automaton Model For Adenocarcinomas in Situ with Java: Study of One Case*. To appear in Proceedings of 24th International European Conference on Parallel and Distributed Computing (EUROPAR 2018), 2018.
- [26] Tomeu Hardasmal, A. and Salguero Hidalgo, A. *A Lock Free Approach To Parallelize The Cellular Potts Model: Application To Ductal Carcinoma In Situ*. To appear in Journal of Integrative Bioinformatics, ISBN , 2020.
- [27] Williams, A. *C++ Concurrency in Action, Second Edition*. Manning Publications, 2017.
- [28] Winters, S, Martin, C., Murphy, D. and Shokar, N.K. *Breast Cancer Epidemiology, Prevention and Screening*. Progress in Molecular Biology and Traslational Science, vol. 151, pp. 1-32, 2017.



Antonio J. Tomeu received a Ph.D. in Mathematics from University of Cádiz (Spain), in 2002. He is associate professor of Computer Science in the University of Cádiz, coordinator of TECDIS (Red Iberoamericana de Investigación en Tecnologías Concurrentes, Distribuidas y Paralelas).



Alberto G. Salguero received Ph.D. in Computer Science from University of Cádiz (Spain), in 2013. He is associate professor of Computer Science in the University of Cádiz, member of TECDIS (Red Iberoamericana de Investigación en Tecnologías Concurrentes, Distribuidas y Paralelas). His work focus on ontologies.

Santiago Zaldívar Valle received a BEng. in Computer Science from the University of Cádiz (Spain) in 2016. He is assistant professor of Computer Science University of Cádiz and works in his doctoral dissertation.

José Aparicio Patino received his M.D. by the Faculty of Medicine of the University of Cádiz (Spain) in 1983 and his Ph.D. in Pathology in 1988 (University of Cádiz). He is full professor of Pathological Anatomy and Histology in the University of Cádiz. His research work focus in adenocarcinomas of prostate and breast.