

# Implementación de algoritmo de múltiples escalas para la binarización de series de tiempo de expresión genética BASC <sup>1</sup>

Tatiana Andrea Higuera Munevar,  
Helber Giovanni Sissa Becerra

DIRECTOR,  
ING. WILMER GARZÓN

PROGRAMA DE INGENIERÍA DE SISTEMAS  
Escuela Colombiana de Ingeniería Julio Garavito  
Bogotá, Colombia

2 de agosto de 2017

<sup>1</sup>Binarization across Multiple Scales.



## **Agradecimientos**

El presente trabajo va dirigido con una expresión de gratitud a los profesores Wilmer Garzón y Jaime Seguel, por su acompañamiento y guía a lo largo del trabajo, agradecemos por la oportunidad de hacer parte de todo este proceso y brindarnos tantas enseñanzas, las cuales fueron motivo de una formación personal y profesional.

## Resumen

Por medio de este documento queremos constatar el desarrollo del trabajo realizado durante el proyecto de grado, el cual se centra en la implementación del algoritmo de BASC teniendo dos diferentes métodos llamados BASC A y BASC B. Los cuales son desarrollados específicamente para la binarización de los datos de expresión genética. Estos algoritmos de binarización incorporan los datos a diferentes escalas para producir umbrales adecuados y robustos, incluso para un pequeño número de puntos de datos. Las diferentes escalas se determinan por el número y la ubicación de las discontinuidades que se utilizan para la aproximación de la función escalonada, desarrolladas en dos diferentes enfoques de múltiples escalas para la binarización de series de tiempo de expresión genética (BASC A, BASC B). Y de esta manera lograr por medio de la implementación de dichos algoritmos, modelar y simular la construcción de la serie de tiempo de expresión genética y así tener un análisis mucho más sencillo.

# Índice general

<b>Agradecimientos</b>	<b>3</b>
<b>Resumen</b>	<b>4</b>
<b>1. Introducción</b>	<b>7</b>
<b>2. Descripción del proyecto</b>	<b>9</b>
2.1. Justificación . . . . .	9
2.2. Planteamiento del problema . . . . .	9
2.3. Alcance del proyecto . . . . .	10
2.4. Objetivos . . . . .	10
2.4.1. Objetivo general . . . . .	10
2.4.2. Objetivos específicos . . . . .	10
2.5. Metodología propuesta . . . . .	11
2.6. Área de aplicación del producto resultado del proyecto . . . . .	11
2.7. Resultados esperados . . . . .	11
2.7.1. PGR1 (2016-2) . . . . .	11
2.7.2. PGR2 (2017-1) . . . . .	11
2.8. Usuarios potenciales directos e indirectos de los resultados de la investigación . . . . .	11
2.9. Cronograma . . . . .	12
2.10. Herramientas de software utilizadas . . . . .	13
2.11. Criterios de terminación del trabajo . . . . .	13
2.12. Evaluadores . . . . .	13
<b>3. Marco Teorio</b>	<b>15</b>
3.1. Etapas computacionales BASC . . . . .	15
3.2. Transformada Wavelet . . . . .	15
3.3. Antecedentes . . . . .	17
3.3.1. BitrinA . . . . .	17
3.3.2. Funcionalidades del paquete BitrinA . . . . .	17
3.4. Definición de términos básicos . . . . .	18
3.4.1. Binarización . . . . .	18
3.4.2. Series de tiempo . . . . .	18

3.4.3. Redes booleanas . . . . .	19
3.4.4. Redes de regulación genéticas . . . . .	19
<b>4. Bibliografía</b>	<b>21</b>

# 1

## Introducción

En el contexto de análisis de datos biológicos, se ha prestado gran atención a la reconstrucción de las redes booleanas a partir de series de tiempo, datos biológicos como lo son perfiles de expresión genética. Seleccionando los candidatos más interesantes para su posterior análisis o para revelar datos aplicables para otros enfoques, tales como el modelado de redes booleanas [1], por lo tanto, es necesario binarizar los datos seleccionados, ya que estas redes modelan genes como variables binarias (ya sea .expresado.º "no expresado"), esto puede lograrse mediante la categorización de los datos en grupos distintos y asignar si esta categorización es válida, [2].

Para simplificar, la binarización es el proceso de partición de datos en dos grupos y asignar un único valor para cada uno. Puede verse como un caso especial de la discretización de una señal continua en un vector de serie de tiempo medidas en un solo gen  $u = (u_1, \dots, u_K)$  con  $k$ -values siendo  $u(1) < \dots < u(k)$ . El enfoque más común es binarizar las mediciones de acuerdo con un umbral, como el valor medio, un cuartil o técnicas de estimación más elaboradas. También se puede obtener un umbral de información externa. Normalmente se asigna el valor 0 a los puntos de datos por debajo del umbral ('bajo') y 1 a los anteriores ('alto'). El objetivo de dicho método de binarización es encontrar el umbral más apropiado [3][4].

Como se dijo anteriormente la binarización es a menudo el preprocesamiento de datos más usado para realizar análisis de datos de genes. Por ejemplo, datos de expresión genética de las series de tiempo deben ser binarizados para la reconstrucción y simulación de redes de operadores booleanas. Sólo entonces pueden estos modelos de regulación genéticas poder ser la representación de cada gen como una variable binaria que ilustra los dos estados 'activo' y 'inactivo'. Además de ser un requisito previo importante para diversos análisis, también se pueden usar independientemente para extraer información. Puede, por ejemplo, servir como un método no supervisado para revelar patrones dentro de los datos. Zilliox y Irizarry (2007) mostraron que las señales binarizadas pueden ser utilizados para la clasificación

basada en el prototipo. Tuna y Niranjana (2010) demostraron que la capacidad de generalización de las máquinas de vectores de soporte (conjunto de algoritmos de aprendizaje supervisado) no es rechazada por perfiles de expresión génica binarizadas [1].

Se busca una alternativa a las desventajas en BASC, como lo es la programación dinámica ya que está limitada su eficiencia en el procesamiento de series de tiempo de tamaño real, buscamos por medio de la implementación de una nueva versión (BASC Wavelet) en la que no depende de la programación dinámica sino por el contrario realizar dichos cálculos por medio del uso de la 3.2 Transformada Wavelet .



## 2

# Descripción del proyecto

### 2.1. Justificación

La representación computacional de redes de regulación genética (Gene Regulatory Networks, 3.4.4 Redes de regulación genéticas) es el desafío actual más importante en el área de Sistemas Biológicos (Systems Biology). El constante crecimiento de bases de datos con series de tiempo de expresión genética, apoyado por el advenimiento de las tecnologías de secuenciación en paralelo de alta precisión conocidas como “Next Generation Sequencing”, ha estimulado el desarrollo de varios modelos de redes regulatorias, incluyendo entre ellos el de las redes booleanas.

Las redes booleanas presentan la red de regulación genética en términos de circuitos lógicos, lo cual las hace altamente descriptivas del sistema que modelan, y fáciles de usar en simulaciones. Un paso crucial en la construcción de una representación booleana es la binarización de las series de tiempo de expresión genética, obtenidas en experimentos en laboratorios. Esto es, determinar los instantes de tiempo en que un gen está expresado (activo) o no. BASC es uno de los algoritmos numéricos propuestos para ese fin. BASC usa programación dinámica lo cual limita su eficiencia en el procesamiento de series de tiempo de tamaño real. Además, de hacer la implementación de BASC, este proyecto busca una alternativa a esa desventaja, por medio de la implementación de una nueva versión (BASC Wavelet) en la que se busca no depender de la programación dinámica sino por el contrario realizar dichos cálculos por medio del uso de alguna transformada (wavelet).

### 2.2. Planteamiento del problema

La reconstrucción de una red booleana implica realizar una binarización de las series de tiempo de expresión genética, este procedimiento define en que instante de tiempo un gen está expresado (activo) o no.

### 2.3. Alcance del proyecto

Implementación del algoritmo de BASC en sus dos versiones y en una tercera versión la cual será la optimización del mismo por medio de las transformadas de Wavelet Haar, todo esto con una interfaz gráfica funcional y amigable, brindándole al usuario una interacción sencilla por la cual podrá hacer uso de los algoritmos de BASC y obtener de la manera más sencilla toda la información que brinda la reconstrucción booleana de redes de regulación genética.

### 2.4. Objetivos

#### 2.4.1. Objetivo general

Implementación del algoritmo BASC (Binarization across Multiple Scales) en sus dos versiones (BASCA, BASCB), para la binarización de series de tiempo de expresión genética, necesaria en la reconstrucción booleana de redes de regulación genética. Optimización del algoritmo por medio del uso de transformadas de Wavelet.

#### 2.4.2. Objetivos específicos

- Interfaz gráfica funcional amigable que permita la ejecución del algoritmo BASC en alguna de sus versiones.  
• ón eficiente del algoritmo BASC en sus dos versiones.
- Implementación del algoritmo BASC haciendo uso de la técnica de programación dinámica.
- Se podría hablar de los resultados de BASCPy vs BASC en R
- Realizar búsqueda literaria relacionada con las transformadas de Wavelet.
- Aplicar una transformada(wavelet) para el análisis de expresiones de tiempo.
- Implementación del algoritmo usando las transformadas de wavelet, con el fin de mejorar la eficiencia del algoritmo.

## **2.5. Metodología propuesta**

En el estudio propuesto se pretende seguir la siguiente metodología:

- Búsquedas bibliográficas.
- Selección de la bibliografía más importante.
- Análisis y posible implementación del algoritmo que lleve a la solución del problema.
- Pruebas del diseño e implementación.

## **2.6. Área de aplicación del producto resultado del proyecto**

Orientado a laboratorios científicos en el área de bioinformática.

## **2.7. Resultados esperados**

### **2.7.1. PGR1 (2016-2)**

- Implementación del algoritmo BASC A
- Implementación del algoritmo BASC B
- Pruebas con muestras simuladas tomando tiempos de ejecución y resultados.
- Interfaz gráfica funcional para las dos versiones del algoritmo BASC

### **2.7.2. PGR2 (2017-1)**

- Optimización del algoritmo BASC por medio de la transformada Wavelet.
- Creación y publicación del artículo académico en relación con el proyecto.
- Elaboración de documentos requeridos

## **2.8. Usuarios potenciales directos e indirectos de los resultados de la investigación**

- Personas del área de salud
- Científicos e investigadores en el área de ciencias de la vida para analizar los sistemas de regulación de genes a nivel molecular

## 2.9. Cronograma

El siguiente es el cronograma propuesto para los dos integrantes del proyecto, es decir por persona se busca que se emplee cantidad de tiempo propuesto en el cronograma:

	Task Mode	Task Name	Duration	Start	Finish	Predecessors
1		<b>PGR1 (2016-2)</b>	<b>91 days</b>	<b>Mon 8/08/16</b>	<b>Mon 12/12/16</b>	
2		Inducción a revisión y búsqueda bibliográfica	28 days	Mon 8/08/16	Wed 14/09/16	
3		Registro de los criterios de evaluación	2 days	Fri 12/08/16	Mon 15/08/16	
4		Registro de la formulación del Proyecto de Grado	2 days	Mon 15/08/16	Tue 16/08/16	
5		Implementación del algoritmo BASC A	35 days	Mon 5/09/16	Fri 21/10/16	
6		Implementación del algoritmo BASC B	26 days	Thu 20/10/16	Thu 24/11/16	
7		Pruebas con muestras simuladas tomando tiempos de ejecución y resultados	4 days	Thu 24/11/16	Tue 29/11/16	
8		Interfaz gráfica	14 days	Wed 16/11/16	Mon 5/12/16	
9		Entrega de productos finales PGR1	0 days	Mon 12/12/16	Mon 12/12/16	
<b>NTT CHART</b>						
	Task Mode	Task Name	Duration	Start	Finish	Predecessors
10		<b>PGR2 (2017-1)</b>	<b>91 days</b>	<b>Mon 16/01/17</b>	<b>Sat 20/05/17</b>	
11		Inducción a revisión y búsqueda bibliográfica	35 days	Mon 30/01/17	Fri 17/03/17	
12		Registro de los criterios de evaluación	2 days	Sat 4/02/17	Mon 6/02/17	
13		Registro de la formulación del Proyecto de Grado	2 days	Tue 17/01/17	Wed 18/01/17	
14		Optimización del algoritmo BASC por medio de la transformada Wavelet	58 days	Mon 13/02/17	Wed 3/05/17	
15		Registro de la versión la Cartelera técnica	12 days	Mon 3/04/17	Tue 18/04/17	
16		Entrega física de la Cartelera Técnica	1 day	Tue 2/05/17	Tue 2/05/17	
17		Vitrina Académica	1 day	Thu 4/05/17	Thu 4/05/17	
18		Creación del artículo académico en relación con el proyecto	31 days	Mon 20/03/17	Mon 1/05/17	
19		Registro del estado del arte	26 days	Fri 7/04/17	Fri 12/05/17	
20		Registro del artículo técnico	26 days	Fri 7/04/17	Fri 12/05/17	
21		Publicación de la evaluación del artículo técnico	15 days	Mon 1/05/17	Fri 19/05/17	
22		Entrega de productos finales PGR2	0 days	Fri 19/05/17	Fri 19/05/17	

## 2.10. Herramientas de software utilizadas

- Sistema operativo: Windows
- Servidor Windows server (Pruebas)
- Word (Documentos)
- Microsoft Project (Cronograma)
- JetBrains PyCharm Community Edition 2016.2.3 (Herramienta para codificar)

## 2.11. Criterios de terminación del trabajo

- Implementación de los algoritmos BASC A y BASC B.
- Interfase grafica para el uso de los algoritmos y el fácil manejo y entendimiento de cada uno de ellos.
- Implementación de los algoritmos BASC A y BASC B en términos de la transformada de wavelet (Haar).
- Creación y publicación del artículo académico en relación con el proyecto.
- Documentación Requerida (Estado del arte, Manual de usuario, Manual de instalación).

## 2.12. Evaluadores

- Director del proyecto: Wilmer Edicson Garzón Alfonso.
- Profesor Jaime Seguel, de la Universidad de Puerto Rico Mayagüez.
- Profesores de la decanatura.



# 3

## Marco Teorio

### 3.1. Etapas computacionales BASC

Los dos métodos BASC se pueden subdividir en tres pasos:

1. Calcular una serie de funciones escalonadas: Se obtiene una función de paso inicial reorganizando las mediciones originales de la serie temporal en orden creciente.
2. Se calculan las funciones escalonadas con menos discontinuidades. BASC A calcula estas funciones de paso de tal manera que cada una minimiza la distancia euclídea a la función de paso inicial. Utiliza la programación dinámica para calcular estas funciones de paso óptimas con  $n = 0 \dots N-2$  discontinuidades para minimizar la distancia euclídea a la función de paso inicial  $f$ . Además, en cada función de escalón, el algoritmo encuentra la discontinuidad más fuerte  $n'$  usando la métrica de puntuación que favorece el tamaño del salto grande (caracterizado por la diferencia entre el promedio de todos los  $f(i) > f(n)$  y el promedio de todos los  $f(i) < f(n)$ ) [5]. BASC B obtiene funciones escalonadas a partir de versiones suavizadas de la función de entrada en una forma de espacio de escala
3. Encontrar la discontinuidad más fuerte en cada función de paso: Una discontinuidad fuerte es un tamaño de salto alto (derivado) en combinación con un error de aproximación bajo. Estimación de la ubicación y variación de las discontinuidades más fuertes: Basándose en estas estimaciones, los valores de los datos pueden excluirse de otros análisis.

### 3.2. Transformada Wavelet

La transformada Wavelet y toda su teoría es una rama de la matemática relativamente nueva, la cual ha encontrado aplicaciones muy rápidamente en un gran número de disciplinas incluyendo la Física, Análisis Numérico,

Procesamiento de Señal, Probabilidad y Estadística.

La utilidad de la transformada Wavelet es debida al hecho de que puede ser usada para aproximar funciones/señales de acuerdo con escala resolución usando un conjunto de funciones base llamadas wavelets. Los Wavelets permiten representaciones de funciones en las cual se retiene tanto la escala como la información espacial. Muchas funciones pueden ser aproximadas con gran exactitud usando sólo un pequeño número de coeficientes wavelet.

La Transformada Wavelet también puede ser usada para representar económicamente, características de interés localizadas en una señal, lo cual la convierte en un candidato ideal para la extracción de características en contextos de clasificación.

Transformadas Wavelets y Fourier La DWT es similar a la Transformada de Fourier en algunos aspectos, a diferencia de las funciones base seno y coseno de la Transformada de Fourier, los wavelets están localizados en el espacio así también como en la escala. A fin de discutir las ventajas de la DWT, es útil compararla con la Transformada de Fourier "Ventaneada.º de Corto Tiempo", ya que ha sido una de las técnicas clásicas más populares para preprocesamiento de datos con características localizadas. La Transformada de Fourier es un buen método para representar datos donde las características a pequeña escala (en este caso, alta frecuencia) representan el detalle o ruido causante en la señal o función, y las características a gran escala (baja frecuencia) representan las formas básicas.

Sin embargo, tiene la desventaja de que la información de frecuencia obtenida a partir de la Transformada de Fourier es global, debido a que sus funciones base son funciones seno y coseno. Esto no es satisfactorio cuando se requieren características de localización. El problema puede ser parcialmente superado usando la Transformada de Fourier Ventaneada o de Corto Tiempo, por medio de la cual la señal a ser analizada se multiplica por una función ventana antes de calcular su Transformada de Fourier. No obstante, este método tiene el problema que una ventana de tamaño fijo en el dominio original es acompañada por una ventana fija en el dominio de Fourier. Lo que realmente se necesita es: una ventana larga para analizar los componentes de escala grande y una ventana estrecha para detectar las características de pequeña escala. Esto es exactamente lo que provee la Transformada Wavelet [10].

Después de tener más en claro el significado de las Transformadas de Wavelet, se concluyó que una hipótesis que existía era el uso de las Transformadas de Haar Wavelet, para poder hacer la sustitución de la programación dinámica en el caso de BASCA, o la sustitución del cálculo de la ecuación que se realiza en BASCB, para lograr calcular las funciones escalonadas con menos discontinuidades, en una manera mucho más eficiente, obteniendo aun el mismo resultado en la binarización que con los procedimientos a los que se desea sustituir.



### 3.3. Antecedentes

#### 3.3.1. BitrinA

BitrinA es un paquete escrito en el lenguaje R, que integra diferentes algoritmos de múltiples escalas para binarización de datos unidimensionales con métodos para la evaluación de la calidad y la visualización de los resultados.

Mediante la identificación de las mediciones que muestran grandes variaciones en diferentes puntos de tiempo o condiciones, esta evaluación puede determinar la calidad de los candidatos que se relacionan con el entorno experimental específica [3][4].

#### 3.3.2. Funcionalidades del paquete BitrinA

BitrinA implementa tres algoritmos de binarización y tres algoritmos de trinarización, para los datos unidimensionales contiene pruebas integradas de binarización y para aquellos datos tridimensionales existen estadísticas. Este paquete proporciona funciones de visualización para analizar los resultados. Todos los algoritmos de binarización se basan en el umbral de una señal de valor real  $x$  de acuerdo con un umbral  $t$ .

El primer enfoque de binarización se basa en la agrupación de  $k$ -means (Hartigan, 1975) [8], que minimiza heurísticamente la suma sobre la distancia al cuadrado de los puntos de datos a su centro de agrupación respectivo. Puede ser visto como un modelo de mezcla gaussiana restringido a matrices de covarianza locales isotrópicas. El algoritmo  $k$ -means normalmente se inicializa al azar y se aplica con reinicios para evitar resultados subóptimos. Para la binarización, el número de clusters  $k$  se establece en dos. El umbral de binarización divide los puntos de datos de estos dos grupos. La prueba de inmersión de Hartigan para la unimodalidad (Hartigan y Hartigan, 1985, Maechler, 2013) [9] se utiliza para calificar la binarización de los datos de entrada. La binarizabilidad de la muestra (en realidad su multimodalidad) se estima por la distancia de la distribución de muestras  $F$  de la clase de distribuciones unimodales [8][9].

El segundo enfoque son los algoritmos a multiescala (BASC A y BASC B; Hopfensitz et al. 2012). Estos algoritmos calculan una serie de funciones escalonadas para obtener un umbral de binarización. Una función inicial paso se obtiene reordenando los valores del vector de entrada en orden creciente. Entonces, las funciones escalonadas con menos discontinuidades se calculan secuencialmente. BASC A determina estas funciones escalonadas de tal manera que cada una minimiza la distancia euclídea a la función de paso inicial. Siendo

$$S_n \tag{3.1}$$

el conjunto de todas las funciones escalonadas con exactamente  $n$  discontinuidades. Una cuantificación óptima de

$$f \in (S_{n-1}) \quad (3.2)$$

por una función de paso con  $n$  pasos

$$(1 \leq n < N - 1) \quad (3.3)$$

BASC B construye funciones escalonadas utilizando un enfoque de escala espacial por medio de la función de Bessel (siendo la Función de Bessel modificada del segundo tipo de orden entero  $n$ ) con diferentes factores de alisamiento se aplican a la primera derivada de la función paso ( $x$ ) definido por (3.4) produciendo derivadas suavizadas.

$$\Delta_{\sigma}(x) = \sum_{k=-\infty}^{\infty} \frac{\sigma(x)}{\epsilon^{2\sigma}} \cdot I_{x-k}(2\sigma) \quad (3.4)$$

Aquí,  $I_n$  denota la función de Bessel modificada de orden  $n$ . Sobre la base de estos derivados, funciones escalonadas suavizadas se construyen y se seleccionan. Para cada función de paso, se determina la discontinuidad fuerte. Tanto para los algoritmos de múltiples escalas, el umbral de binarización se define como la media de estas discontinuidades fuertes. Una prueba de arranque se utiliza para evaluar la solidez de este umbral a través de la serie de funciones. Un significativo  $P$ -valor indica que la variación de la ubicación de la discontinuidad más fuerte a través de las funciones escalonadas es baja, es decir, el umbral de binarización no cambia en múltiples escalas y la binarización calculada es de buena calidad.

### 3.4. Definición de términos básicos

#### 3.4.1. Binarización

Conversión de datos en algo expresado o no expresado, en las imágenes digitales se habla de una imagen en dos valores, representados generalmente en blanco y negro. en las redes booleanas hablamos sobre determinar los instantes de tiempo en que un gen está expresado (activo) o no.

#### 3.4.2. Series de tiempo

Por serie de tiempo nos referimos a datos estadísticos que se recopilan, observan o registran en intervalos de tiempo regulares (diario, semanal, semestral, anual, entre otros). El término serie de tiempo se aplica por ejemplo a datos registrados en forma periódica

### 3.4.3. Redes booleanas

Una red booleana se compone de un conjunto de variables booleanas cuyo estado está determinado por otras variables en la red. Una red booleana es un sistema dinámico discreto introducido por Stuart Kauffman en 19691 para modelar redes de regulación génica. Las redes booleanas se transforman así en un modelo, o dicho en palabras más simples, en un juguete que permite predecir o reproducir parcialmente el comportamiento de una red de regulación génica. Para representar los genes se utilizan vértices los que se representan por un círculo, estos vértices pueden estar en estado activo (1) o inactivo (0). Las interacciones entre ellos se representan por funciones booleanas, que dependen de los valores de otros vértices de la red. Esta dependencia se representa por flechas que van desde un vértice a otro o a sí mismo. En una red booleana aleatoria las conexiones son decididas al azar y las salidas de los nodos están determinadas por funciones lógicas generadas al azar.

### 3.4.4. Redes de regulación genéticas

Una red de regulación génica o red de regulación genética (GRN) es una colección de segmentos de ADN en una célula que interactúan entre sí (indirectamente a través de su ARN y productos de expresión de proteínas) y con otras sustancias en la célula, con lo que regulan las tasas a las que los genes de la red se transcriben en ARNm. En general, cada molécula de ARNm va a construir una proteína específica (o un conjunto de proteínas). La regulación genética es el proceso de activación y desactivación de los genes. En las etapas tempranas del desarrollo, las células comienzan a asumir funciones específicas. La regulación genética se asegura de que los genes apropiados se expresen en los momentos adecuados. La regulación genética también puede ayudar a un organismo a responder a su entorno. La regulación genética se lleva a cabo por una variedad de mecanismos, entre ellos la modificación química de los genes y la activación o desactivación de los mismos mediante su asociación con proteínas reguladoras.



## 4

# Bibliografía

- [1] C. Mussel, M. Hopfensitz, H. A. Kestler, `cran.r-project`, "24 June 2016. [Online]. Available: [https://cran.r-project.org/web/packages/BoolNet/vignettes/BoolNet\\_package\\_vignette.pdf](https://cran.r-project.org/web/packages/BoolNet/vignettes/BoolNet_package_vignette.pdf). [Accessed 15 August 2016].
- [2] M. Hopfensitz, C. Mussel, C. Wawra, M. Maucher, M. Kuhl, H. Neumann, and H. A. Kestler, «Multiscale Binarization of Gene Expression,» 23 December 2012. [En línea]. Available: <http://www.informatik.uni-ulm.de/ni/staff/HNeumann/publicationsYear/PDFs/JOURNALS/IEEE-ACM-ToCBB-HopfensitzEtAl12multiscaleBinarization.pdf>. [Último acceso: 8 August 2016].
- [3] C. Müssel, F. Schmid, T. J. Blätte, M. Hopfensitz, L. Lausser, H. A. Kestler, «International Society For Computational Biology,» OXFORD Academy , 14 October 2015. [En línea]. Available: <https://academic.oup.com/bioinformatics/article/32/3/465/1744073/BiTrinA-multiscale-binarization-and-trinarization>.
- [4] S. Mundus, C. Müssel, F. Schmid, L. Lausser, T. J. Blätte, M. Hopfensitz, H. A. Kestler, «BiTrinA: Binarization and Trinarization of One-Dimensional Data,» Hans Kestler <hans.kestler at uni-ulm.de, 06 February 2017. [En línea]. Available: <https://cran.r-project.org/web/packages/BiTrinA/index.html>. [Último acceso: 22 October 2016].
- [5] L. Nakhleh, N. Berestovsky, «An Evaluation of Methods for Inferring Boolean,» 21 June 2013. [Online]. Available: <http://journals.plos.org/plosone/article/file?id=10.1371/journal.pone.0066031type=printable>. [Accessed 15 September 2016].
- [6] S. D. Braudrick, «analyzing microarray time series data using a haar wavelet transformation,» University of Colorado at Denver, 2007. [En línea]. Available: [http://www.ucdenver.edu/academics/colleges/CLAS/Departments/math/graduateprograms/alumni0/alumni/Documents/Student%20Theses/Braudrick\\_MSThesis.pdf](http://www.ucdenver.edu/academics/colleges/CLAS/Departments/math/graduateprograms/alumni0/alumni/Documents/Student%20Theses/Braudrick_MSThesis.pdf).
- [7] F. E. Turkheimer, «CHROMOWAVE patterns of gene expression from microarray data: methodology, validation and clinical relevance in glio-

mas,» *BMC Bioinformatics*, 2006. [En línea]. Available: <https://www1.imperia.ac.uk/resources/ED80B430-39EE-464F-807F-CA051089A8B9/>.

[8] J. A. Hartigan and P. M. Hartigan, «The Dip Test of Unimodality,» *Institute of Mathematical Statistics*, Marzo 1995. [En línea]. Available: [http://www.jstor.org/stable/2241144?seq=1fndtn-page\\_scan\\_tab\\_contents](http://www.jstor.org/stable/2241144?seq=1fndtn-page_scan_tab_contents).

[9] J. A. Hartigan and P. M. Hartigan, «The Annals of Statistics,» *The Dip Test of Unimodality*, 1985. [En línea]. Available: <http://projecteuclid.org/euclid.aos/1176346577>.