

Índice

1. Introducción	1
1.1. Planteamiento del Problema	1
1.2. Justificación	2
1.3. Delimitación	2
1.4. Hipótesis	2
1.5. Objetivos	2
1.5.1. Objetivo General	2
1.5.2. Objetivos Específicos	2
1.6. Aportaciones de la tesis	2
2. Estado del Arte	2
2.1. Tecnologías Genómicas	2
2.2. Microarreglos	3
2.3. Formato	5
2.3.1. Plataformas	6
2.4. Metilación	6
3. Propuesta de Solución	6
4. Metodología	7
4.1. Diseño de estudio	7
5. Cronograma	7
6. Referencias	7

1. Introducción

1.1. Planteamiento del Problema

El cáncer de mama es una enfermedad compleja y heterogénea con más de 1,300,000 casos y 450,000 muertes cada año en todo el mundo. Esta enfermedad se caracteriza por diferentes áptectos biológicos como desregulización de la expresión génica, alteraciones genómicas del ADN, etc. Todo esto da lugar al inicio y desarrollo del carcinoma de mama. En éstos últimos años el uso de datos ómicos, como los basados en microarreglos (microarrays) y secuenciación, esta en su pleno () en el campo de la biomedicina. Todos estos datos permiten estudiar enfermedades desde un punto de vista biomolecular. Con esto, se ofrecen grandes oportunidades para mejorar tanto la comprensión de la enfermedad, como el desarrollo de nuevos métodos para el diagnóstico y tratamiento del paciente, sin embargo, el análisis de estos datos producidos por estas tecnologías, es bastante complejo por lo que es necesario la aplicación de avanzadas técnicas de análisis y cálculos computacionales que permiten obtener la información biológica disponibles. Hasta el día de hoy todos estos datos óptenidos de diversos experimentos, datos de muy alta calidad, datos clinicamente bien anotados, y una gran cantidad de datos de canceres analizados con el fin de encontrar anomalías recurrentes que sean importantes de la enfermedad y estos datos se guardan en diversas plataformas que permiten tener uns gran cantidad de información , pero estos datos no son tan faciles de analizar, por ello, la bioinformática ayuda a manejar estructurar y organizarla para que sea mas fácil de

comprender. Circos plot es una de las herramientas que existen para la visualización de datos, ideal para explorar las relaciones entre objetos y posiciones. Esta herramienta es flexible, aunque originalmente fue diseñado para visualizar datos genómicos, se puede crear figuras a partir de datos en cualquier campo, desde la genómica hasta la visualización de la migración al arte matemático. Esta herramienta puede ser automatizada.

1.2. Justificación

Con el fin de mejorar el rendimiento y la cantidad de tiempo de los investigadores del Instituto Nacional de Medicina Genómica, es fundamental sistematizar este software para reducir el tiempo de programación y la investigación. La sistematización de circos plot, brindará la posibilidad de que el investigador ahorre en tiempo de programación o en dado caso que no se conozca nada del uso nativo de circos plot, leer todo el manual de uso de dicho software, para que ocupe su mayor tiempo en la investigación y solo tome varios minutos para diseñar su gráfica circular llamada circos plot.

1.3. Delimitación

1.4. Hipótesis

El desarrollo de una herramienta de Visualización basada en integral permitirá el análisis de la información de los datos derivados de múltiples plataformas.

1.5. Objetivos

1.5.1. Objetivo General

Implementar una herramienta de visualización que integre datos genómicos derivados de múltiples plataformas.

1.5.2. Objetivos Específicos

- Realizar el pretratamiento de los datos de transcriptoma (Microarreglos).
- Realizar el pretratamiento de los datos de Metilación.
- Desarrollar algoritmo computacional para hacer la búsqueda en la referencia del genoma humano con los datos de transcriptoma y metilación

1.6. Aportaciones de la tesis

2. Estado del Arte

2.1. Tecnologías Genómicas

Las tecnologías genómicas es el conjunto de herramientas orientadas al estudio integral del funcionamiento, contenido, evolución del genoma. Es una de las áreas más vanguardistas de la biología. La genómica usa conocimientos derivados de distintas ciencias como la biología molecular, la bioquímica, la informática, la estadística, las matemáticas y la física. Para entender un poco más de estas tecnologías

y de los datos que se obtiene de las antes mencionadas, hablaremos de las tecnologías genómicas que son: Microarreglos y Metilación.

2.2. Microarreglos

Un chip de ADN (del inglés DNA microarray) es una superficie sólida a la cual se une una colección de fragmentos de ADN. Las superficies empleadas para fijar el ADN son muy variables y pueden ser de vidrio, plástico e incluso de silicona. Los chips de ADN se usan para analizar la expresión diferencial de genes. Su funcionamiento consiste, básicamente, en medir el nivel de hibridación entre la sonda específica (probe, en inglés), y la molécula diana (target), y se indican generalmente mediante fluorescencia y a través de un análisis de imagen, lo cual indica el nivel de expresión del gen.

El tamaño de los arreglos es de 1.28 cm x 1.28 cm, hay 500,000 ubicaciones en cada matriz y por lo general tiene millones de cadenas de ADN construidas en cada ubicación, cada cadena contiene 25 pares bases (Figura 1).

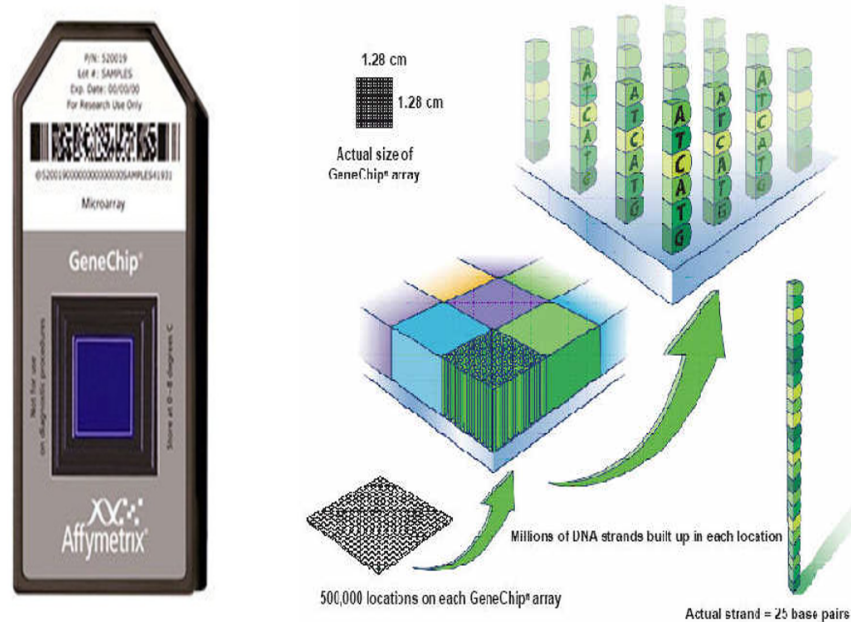


Figura 1: Chip de un Microarreglo.

La tecnología de microarreglos se basa en el dogma central de la biología molecular, mide los niveles transcripción en un determinado caso de estudio.

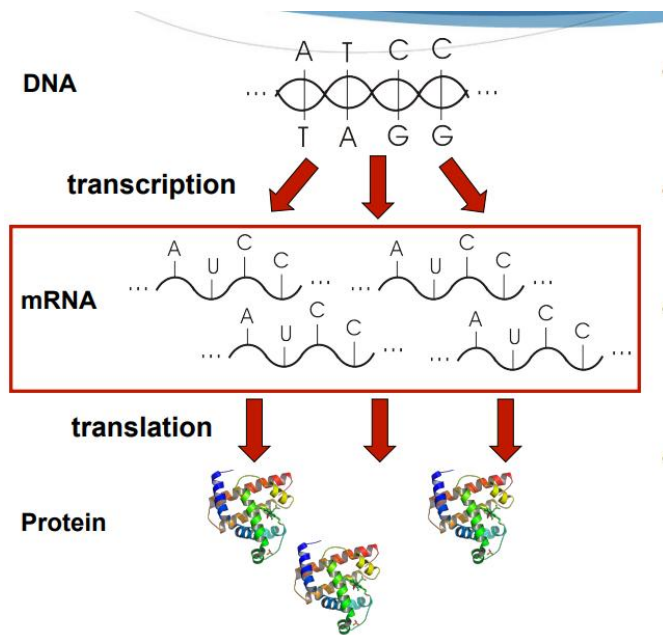


Figura 2: Dogma central de la biología

Con la tecnología de los microarreglos se imprimen las secuencias biológicas en un chip, de manera que se puede cuantificar el transcripción en una matriz numérica.

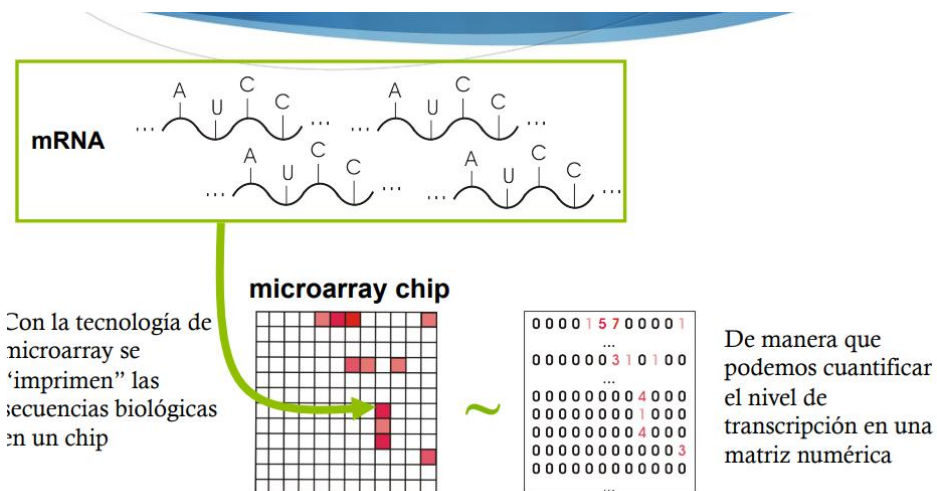


Figura 3: Matriz de expresión

En cada celda de un chip se pegan miles de copias de un segmento de mRNA (sonda), cada celda contiene secuencias de mRNA distintas, todas las celdas contienen el mismo numero de segmentos.

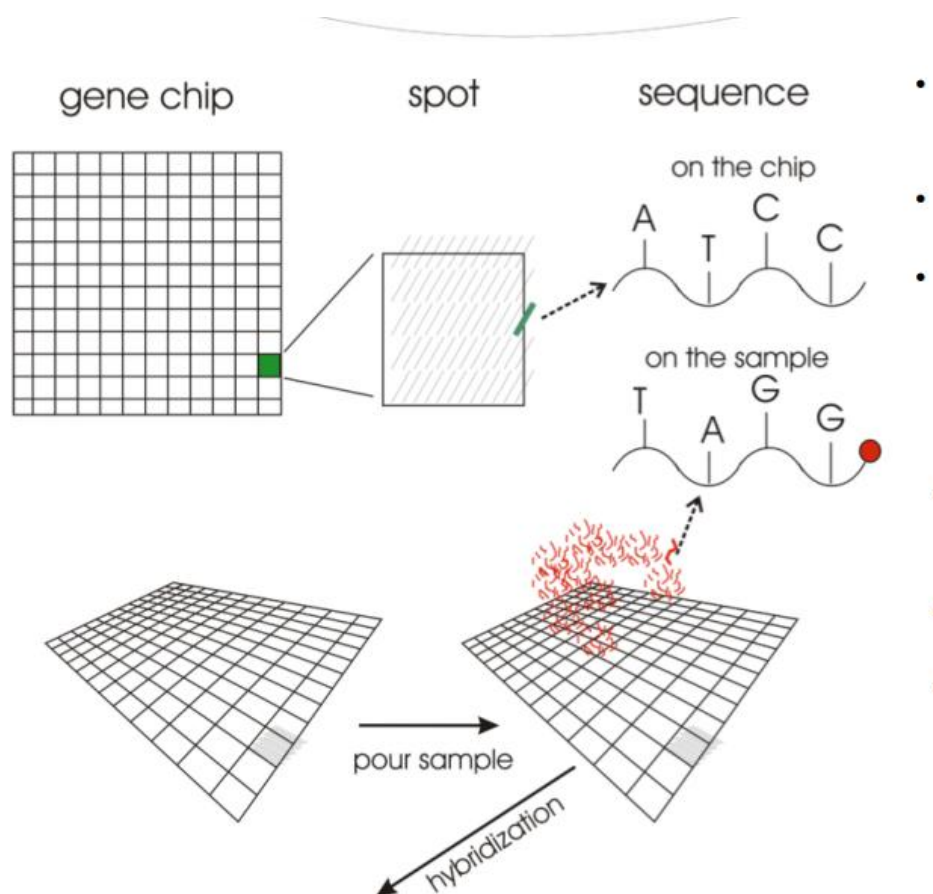


Figura 4: Ejemplo de un chip de microarreglo.

2.3. Formato

Los archivos están disponibles en un formato de valores separados por comas (CSV). Estos son archivos de texto sin formato con cada fila terminada por un carácter de nueva línea. Los datos en campos separados están entre comillas y separados por comas. Ninguno de los campos de datos contiene ninguno de estos caracteres: comillas, nueva línea, retorno de carro o tabulación.

Estos archivos se usan principalmente en aplicaciones de hojas de cálculo y programas de bases de datos (como bases de datos SQL). Los datos están formateados de tal manera que estos dos usos sean relativamente fáciles. Se tiene en cuenta que algunos de los archivos y los campos de datos en ellos son grandes.

La primera fila de cada archivo contiene los títulos de los campos que figuran en las filas siguientes.

Cada fila después de la primera fila contiene anotaciones para un solo conjunto de sondas. Todas las anotaciones para ese conjunto de sonda están contenidas en esa única fila. En algunos campos, como las anotaciones de dominio de proteínas, puede haber más de una anotación para un único conjunto de sondas. En este caso, los valores múltiples están separados por la cadena '///'.

En muchos tipos de anotaciones, los subcampos están separados por '//' . Por ejemplo, una anotación para un "GO Biological Process" puede aparecer como "7155 // cell adhesion // predicted / computed".

En este caso, las secciones corresponden a "ID // Descripción // Evidencia", pero el significado de los subcampos varía entre los diferentes tipos de anotación, como se describe a continuación.

Los campos vacíos se indican con '- - -'. El hecho de utilizar una cadena de este tipo en lugar de dejar el campo vacío es que hace que la naturaleza columnar de los datos sea más visible en ciertos programas de hoja de cálculo. Algunas columnas en algunos archivos no contienen datos. Para ayudar a los usuarios a combinar datos de varios archivos, dichas columnas vacías no se eliminan. Por lo tanto, cada archivo tiene las mismas columnas en el mismo orden.

Algunos campos, como "Chip", contienen el mismo valor para cada conjunto de sonda en un archivo. Aunque estos datos son redundantes en cualquier archivo individual, son útiles para los usuarios que combinan datos de varios archivos.

2.3.1. Plataformas

2.4. Metilación

La metilación del ADN es un proceso epigenético que participa en la regularización de la expresión génica de dos maneras, directamente al impedir la unión de factores de transcripción, e indirectamente proporcionando la estructura "cerrada" de la cromatina.

3. Propuesta de Solución

El enfoque de solución ante la problemática planteada, consiste fundamentalmente en lo siguiente:

- Incorporar adecuadamente circos plot a un interface de usuario para su fácil uso.
- Los controles de manejo para el usuario sencillos y factibles con todas las funciones posibles de circos plot.
- Incorporar Circos plot a la plataforma de Galaxy.
- Instalar circos plot galaxy en Instituto Nacional de Medicina Genómica.

4. Metodología

4.1. Diseño de estudio

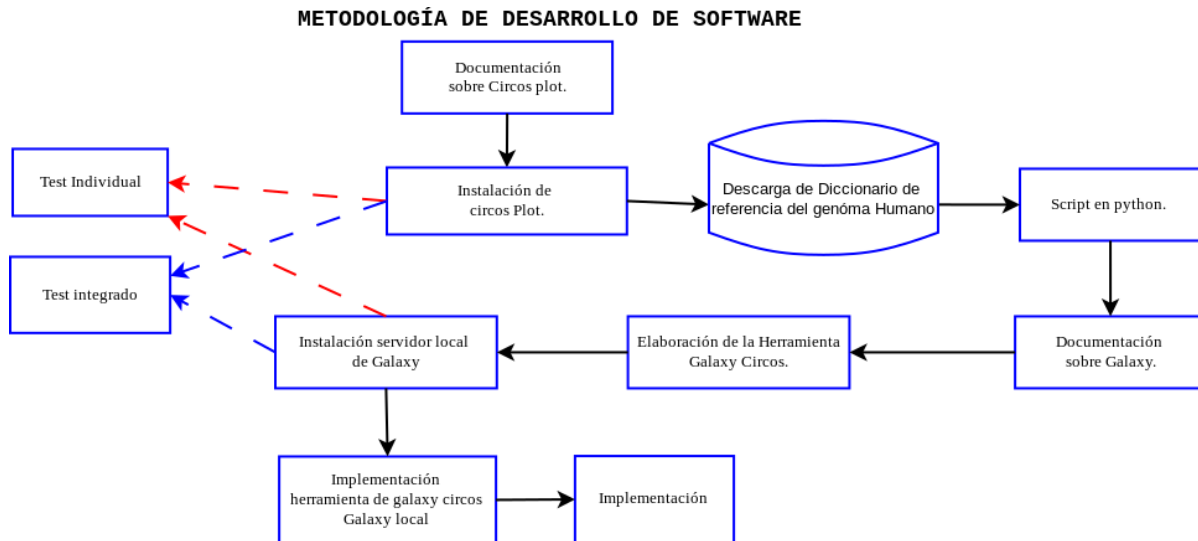


Figura 5: Metodología

5. Cronograma

6. Referencias

TUSHER V. G., TIBSHIRANI R., CHU G. Significance analysis of microarrays applied to the ionizing radiation response. Proceedings of the National Academy of Sciences of the United States of America, 98(9): 5116-21, 2001 Apr 24.

Metilación del ADN en cáncer de mama Diana Casandra Rodríguez-Ballesteros,* Sigfrid Leonardo García-Moreno-Mutio,* Joel Jaimes-Santoyo,* Rosa Elda Barbosa-Cobos,** Alberto de Montesinos Sampedro,*** Olga Beltrán-Ramírez*

Genómica, expresión génica y matrices de ADN David J. Lockhart 1 y Elizabeth A. Winzeler 1