

Modulector: una plataforma como servicio para el acceso a bases de datos de micro ARNs

Marraco, Agustín Daniel ¹; Camele, Genaro ^{3,4,#}; Hasperué, Waldo ^{3,5}; Menazzi, Sebastián ¹;

Abba, Martín ²; Butti, Matias ^{1,2,}*

*¹ Centro de Altos Estudios en Tecnología Informática (CAETI), Facultad de Tecnología Informática, Universidad Abierta Interamericana, CABA, Argentina, ² Centro de Investigaciones Inmunológicas Básicas y Aplicadas (CINIBA), Facultad de Ciencias Médicas, Universidad Nacional de La Plata, ³ Instituto de Investigación en Informática LIDI. Facultad de Informática. Universidad Nacional de La Plata, ⁴ Becario postgrado UNLP, ⁵ Investigador asociado - Comisión de Investigaciones Científicas (CIC-PBA), * matias.butti@uai.edu.ar; # gcamele@lidi.info.unlp.edu.ar*

Resumen. El notable crecimiento del volumen de datos genómicos y la enorme variedad de bases de datos que los almacenan, hacen indispensable disponer de mecanismos eficientes y eficaces de integración. En la actualidad se encuentran disponibles varias herramientas que ofrecen APIs (Interfaz de programación de aplicaciones) que permiten acceder a dicha información, que pueden ser utilizados tanto a través de lenguajes de programación como de navegadores a partir de servicios web. Sin embargo, en dominios específicos de la bioinformática como el caso de los micro ARN -pequeñas moléculas de ARN de gran interés por su capacidad de regular la actividad de otros genes- la mayoría de las soluciones recurren en problemas que dificultan su uso, incluyendo la falta de procesos que simplifiquen la actualización de sus bases de datos a medida que se publica nueva información, tiempos de respuesta inadecuados, dificultad para garantizar la escalabilidad, falta de consistencia en el formato de intercambio de datos, funcionalidad extremadamente limitada, errores por falta de mantenimiento, entre otros problemas frecuentes.

En el presente trabajo se presenta Modulector, una solución que integra información de bases de datos genómicas, con bases de datos de micro ARNs (microARNs), para simplificar

el acceso a las distintas dimensiones de información de los microARNs de interés (secuencias, fármacos y patologías asociadas, genes regulados, publicaciones científicas), poniendo especial énfasis en resolver las problemáticas técnicas comunes descritas anteriormente.

Moduletor brinda acceso a través de una API REST (API para la transferencia de estado representacional), garantiza tiempos de respuesta adecuados y escalabilidad, tiene capacidad de ordenamiento, filtro, búsqueda y paginado de resultados. La solución utiliza contenedores, simplificando el despliegue en cualquier servidor, lo que la hace adaptable para la mayoría de los casos de uso donde se quiere utilizar Moduletor de manera privada. Toda la información retornada por Moduletor se encuentra normalizada en formato JSON, haciéndola eficiente para su manipulación mediante cualquier herramienta de desarrollo. El código fuente de Moduletor está disponible en <https://github.com/omics-datascience/moduletor>.

Palabras clave: microARN; Regulación de la expresión génica; Bioinformática; Base de datos biomédica; Plataforma web

Moduletor: a platform-as-a-service for access to microRNA databases

Abstract. The remarkable growth in the volume of genomic data and the enormous variety of databases that store them make it essential to have efficient and effective integration mechanisms. Several tools are currently available that offer APIs (Application Programming Interfaces) that allow access to this information, which can be used both through programming languages and browsers from web services. However, in specific domains of bioinformatics such as the case of MicroRNAs -small RNA molecules of great interest due to their ability to regulate the activity of other genes- most of the solutions fall back on problems that make them difficult to use, including the lack of processes that simplify the updating of their databases as new information is published, inadequate response times, difficulty to guarantee scalability, lack of consistency in the data exchange format, extremely limited functionality, errors due to lack of maintenance, among other frequent problems.

This paper presents Moduletor, a solution that integrates information from genomic databases with microARN (miRNA) databases to simplify access to the different dimensions of microRNA information of interest (sequences, drugs and associated pathologies, regulated genes, scientific publications), with special emphasis on solving the common technical problems described above.

Moduletor provides access through a REST API (API Representational State Transfer), guarantees adequate response times and scalability, has sorting, filtering, searching, and pagination capabilities. The solution uses containers, simplifying deployment on any server, which makes it adaptable for most use cases where Moduletor is to be used privately. All information returned by Moduletor is normalized in JSON format, making it efficient for manipulation by any development tool. Moduletor source code is available at <https://github.com/omics-datascience/moduletor>.

Keywords: Gene expression regulation; Bioinformatics; Biomedical database; Web platform

Novedad u originalidad local en el conocimiento

Moduletor surge como necesidad en el marco del proyecto Multiomix¹ del grupo de bioinformática en oncogenómica funcional del Centro de Altos Estudios en Tecnología Informática (CAETI) de Universidad Abierta Interamericana (UAI) y el Centro de Investigaciones Inmunológicas Básicas y Aplicadas (CINIBA) de la Facultad de Medicina de la Universidad Nacional de La Plata. El proyecto Multiomix tiene por objetivo desarrollar una plataforma para inferir eventos genómicos y epigenómicos asociados con la regulación de la expresión génica. Multiomix identifica genes

¹ <https://multiomix.org>

regulados por los distintos moduladores de la expresión génica, incluidos los microARNs. Ante un hallazgo de interés, Multiomix necesita mostrar información detallada sobre esa relación y sobre el microARN en particular. Sin embargo, no existe una solución que cumpla con las necesidades técnicas y funcionales. La mayoría de las alternativas existentes contienen bases de datos obsoletas, solo pueden ser utilizadas mediante software específico, los tiempos de respuesta de obtención de la información son más lentos que los umbrales aceptables para una web de baja latencia como Multiomix y/o no integran todas las dimensiones de información que se requieren. De estas problemáticas surge la necesidad de contar con una solución que unifique el acceso a las principales fuentes de datos disponibles relacionadas con microARN.

Modulector abastece a Multiomix de información molecular relevante y actualizada a partir de servicios web, optimizando los tiempos de respuesta. Además, puede ser fácilmente integrada programáticamente desde cualquier lenguaje de programación o utilizada desde cualquier navegador.

Grado de relevancia

Los micro ARN (microARNs) son pequeñas moléculas de ARN no codificantes (no se traducen a proteínas) con capacidad de modular la expresión de otros genes a nivel post-transcripcional. Por esta razón toman relevancia en estudios multi-ómicos que tienen por objetivo descifrar los mecanismos de modulación de expresión génica en

enfermedades complejas multigénicas y multifactoriales, como los que ocurren en la mayoría de las enfermedades crónicas metabólicas, neurodegenerativas y neoplásicas. Adicionalmente, los microARN son importantes en procesos de identificación de biomarcadores con poder diagnóstico, pronóstico/predictivo y como potenciales blancos de nuevas estrategias terapéuticas (Johora et al., 2019; Wu, 2010).

Grado de pertinencia

El desarrollo de Modulector permitió cumplir la necesidad original que era la integración con la plataforma Multiomix, para brindarle un único punto de acceso a toda la información de microARNs. La integración entre ambas plataformas no presentó dificultades. Modulector fue desplegado exitosamente y Multiomix lo utiliza en su versión productiva. La versión productiva de Modulector queda disponible para la comunidad científica de biólogos y bioinformáticos que necesiten acceder a información de microARNs.

Grado de demanda

Modulector permite a cualquier desarrollador o investigador familiarizado con la bioinformática hacer uso de sus servicios a través de Multiomix de manera gráfica y simple, con una puesta en producción pensada desde la experiencia del usuario. A su vez, aquellos usuarios que cuenten con los conocimientos técnicos requeridos

pueden realizar un despliegue en un servidor privado y consumir la información de la plataforma dentro de los márgenes de su propio espacio de trabajo.

Actualmente Modulector constituye un buen complemento a las herramientas y bases de datos existentes, ya que es una nueva alternativa rápida, consistente y mantenida.

Desarrollo del producto

La sección I presenta una revisión de los recursos bioinformáticos de microARN, mencionando características positivas y negativas de cada uno, que sirvieron como disparador para el desarrollo y la implementación de Modulector.

En la sección II se detallan las características más importantes de esta plataforma.

En la sección III se describen las herramientas que forman parte del repertorio de bases de datos que Modulector brinda al usuario. En la sección IV se brindan detalles sobre la implementación de la plataforma y los servicios que provee.

Si bien Modulector está disponible para ser consultado vía Internet, en la sección V se describen aspectos relacionados con el despliegue de esta plataforma en un ambiente privado. Para finalizar, en la sección de Conclusiones se discuten las ventajas conseguidas con la plataforma presentada.

I) Herramientas de consulta de microARN

Existen múltiples herramientas disponibles en Internet que permiten el uso de esta moduladores epigenéticos. Se pueden enumerar algunas populares como multiMiR (Ru et al., 2014; Licursi et al., 2018) que agrupa información de 14 bases de datos sobre interacciones microARN-gen con asociaciones a enfermedades y fármacos. Multimir ofrece una API consumible desde el lenguaje de programación R. Su estructura consta de cuatro servicios: interacciones de un microARN definido por el usuario con genes, interacción de dicho microARN pero limitando la búsqueda a un gen en particular, las enfermedades asociadas y por último los fármacos asociados a dicho regulador. La baja performance y el incorrecto manejo de errores dificultan su uso.

MIENTURNET (Licursi et al., 2019) es otra herramienta popular que toma una lista de microARNs y/o genes y realiza una búsqueda en las bases de datos TargetScan (Agarwal et al., 2015) y mirTarBase (Huang et al., 2019, 2020). Otra herramienta es miRNet (Fan et al., 2016, 2018; Fan y Xia, 2018; Chang et al., 2020), que recopila la información de las bases de datos miR2Disease (Jiang et al., 2009), HMDD (Huang et al., 2019) y PhenomiR (Ruepp et al., 2010) con información sobre asociaciones relevantes con enfermedades de interés.

La herramienta miRGator (Seungyoon et al., 2007; Cho et al., 2010; Cho et al., 2012) exhibe perfiles de expresión, asociaciones con genes, correlaciones de expresión y otros recursos de apoyo. Otras herramientas como RNA22 (Miranda et al., 2006), mirPath (Vlachos et al., 2015), MirGeneDB (Fromm et al., 2015) las cuales poseen sus

respectivas bases de datos, ponen a disposición funcionalidades o información específicas de microARN como la identificación de sus sitios de unión, sus correspondientes heterodúplex, predicción de asociaciones, entre otras.

Aunque la cantidad puede sonar cuantiosa, la mayoría de las herramientas existentes contienen bases de datos con baja o nula frecuencia de actualización. Los hallazgos y las hipótesis nuevas son continuas en el área, y por esta razón la actualización es una característica fundamental en este ámbito. Aun cuando hay datasets disponibles que siguen vigentes, la manera de acceder a esta información es heterogénea o cuenta con una pobre interfaz de usuario.

Además, se ha observado que estas herramientas tienen un tiempo de respuesta de dos minutos promedio para resolver consultas complejas. Aun cuando la respuesta es generada en un tiempo aceptable, la estructura de los datos devueltos es inconsistente ya que poseen campos redundantes y tipos de datos nulos representados de diferentes maneras (algunas celdas contienen el valor null, otras una cadena vacía y otras la cadena "NA" (Not Available).

Por último, se encuentra la barrera del conocimiento técnico necesario para poder utilizarlas, dado que las interacciones sólo pueden darse a través de un lenguaje de programación específico como R, Java o Python. Estos lenguajes, si bien son populares en el ámbito tanto bioinformático como científico, requieren conocimiento técnico del usuario para acceder a la información almacenada.

En función de las limitaciones y debilidades que poseen las herramientas actuales se presenta el desarrollo de Modulector, una plataforma que, además de contar con bases de datos actualmente mantenidas por las respectivas entidades encargadas, agrupa la información de todas ellas en servicios eficientes y estructurados con funcionalidad extendida como paginación, filtros, búsqueda y ordenamiento.

II) Modulector

Modulector es una plataforma que integra la información de las principales bases de datos públicas de microARNs. Está disponible para ser consultada/integrada fácilmente desde cualquier plataforma/aplicación, independientemente del lenguaje de programación o entornos de desarrollo que utilice, manteniendo un aspecto simple y ágil. La información que ofrece Modulector incluye:

- Interacciones microARN-gen, es decir cuáles son los genes target (genes con los que interactúa para modular su actividad) para cada uno de los microARNs, y además un score que permite comparar las evidencias estadísticas/biológicas de diferentes interacciones.
- Fármacos con los que hayan sido asociados un microARN y el tipo de asociación. Si fue putativa o incluye comprobación experimental.
- Patologías con las que haya sido asociado un microARN y el tipo de asociación.
- Secuencia genética del microARN

- Alias del microARN
- Información general del microARN.
- Información general sobre genes.

Tanto los genes como los microARNs pueden ser accedidos mediante su id oficial o cualquiera de sus alias, razón por la cual la solución incorporó bases de datos bioinformáticas para soportar consultas a partir del nombre oficial o cualquiera de sus alias.

Moduletor ofrece funcionalidad a partir de una serie de servicios, desarrollados en lenguaje de programación Python, y los disponibiliza a partir del framework web Django que permite definir un modelo estructurado de la información que maneja. Esta información es almacenada en PostgreSQL que es una base de datos relacional de alta performance y código abierto. La figura 1 muestra la arquitectura y el circuito de consulta de Moduletor.

A su vez, se hace uso del meta framework Django REST Framework que simplifica la implementación de filtros, ordenamiento y búsquedas por campos específicos, como así también paginación parametrizada. De esta manera los servicios consumidos por los usuarios brindan una respuesta de manera rápida al tiempo que permiten una cómoda manipulación de la información obtenida.

Moduletor hace uso de la información de las principales bases de datos públicas de microARNs vigentes (detalladas en la sección III) que primero pasan por un proceso de limpieza, transformación, enriquecimiento y luego son cargadas en la

base de datos de Moduletor a través de un proceso de sincronización que será introducido en la sección IV.

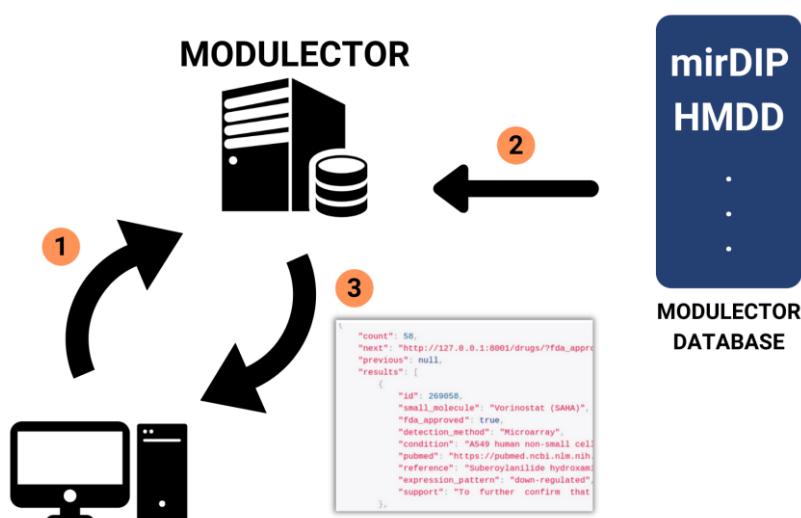


Figura 1. Circuito general de consulta de los servicios de Moduletor. En primer lugar el usuario, ya sea desde el sitio web de Multiomix o desde cualquier otra aplicación cliente (navegadores, herramientas bioinformáticas, entre otros), realiza una consulta con los filtros, búsqueda, ordenamiento y paginación requeridos (1). Luego Moduletor se encarga de obtener de manera eficiente la información almacenada en su propia base de datos (2). Por último dichos los resultados son devueltos en formato JSON al usuario (3).

Figure 1. General workflow of Moduletor services. First, the user, either from the Multiomix website or from any other client application (browsers, bioinformatics tools, among others), performs a query with the required filters, search, sorting and pagination (1). Then Moduletor efficiently retrieves the information stored in its database (2). Finally, the results are returned in JSON format to the user (3).

III) Herramientas y bases de datos integradas en Modulector

En esta sección se describen las herramientas y bases de datos que utiliza la plataforma Modulector.

mirDIP (Shirdel et al., 2011; Tokar et al., 2018) integra bases de datos de predicción gen-microARN con un score unificado que califica la asociación, permitiendo comparar distintas asociaciones. Como ocurre en las herramientas bioinformáticas en general, esta información no debe ser interpretada como determinante, pero es una gran guía para el investigador que la utiliza. Luego de evaluar diferentes herramientas que estiman puntajes, mirDIP fue la seleccionada para utilizar en Modulector.

HMDD (Huang et al., 2019), actualmente en su versión 3.0, recolecta de la literatura científica, asociaciones establecidas entre microARNs y patologías. De esta manera los usuarios de Modulector disponen de información sobre las enfermedades relacionadas a un microARN de interés. Una parte del pre-procesamiento que se tuvo que realizar con este dataset consistió en estandarizar las formas en que son nombrados los microARNs.

Sm2mir (Xinyi et al., 2013) es similar a HMDD pero aporta información sobre fármacos. Esta base de datos recopila asociaciones entre un microARN y un fármaco con la siguiente información: artículos científicos que avalan dicha relación, compuestos químicos relevantes en fármacos asociados identificado con el identificador de PubChem (CID), en caso de haber sido comprobado

experimentalmente cuál fue el método utilizado, tejidos o condiciones de detección, y si se encuentra aprobada por la Food and Drug Administration (FDA) estadounidense. Al igual que HMDD, se pre-procesó la información para transformar el modo en que se nombran los microARNs.

El **Comité de Nomenclatura HUGO Gene** (HGNC por sus siglas en inglés, HUGO Gene Nomenclature Committee) es la entidad responsable de aprobar y registrar los nombres oficiales de los genes. Utilizamos la lista de alias de genes que HUGO disponibiliza en la sección de descarga de su web. El incorporar esta información ofrece a los usuarios utilizar cualquier alias para consumir los servicios de Modulector y reducir al mínimo los falsos negativos que puedan ocasionarse debido a que algunas de las bases de datos integradas utilizan un alias distinto al que se está especificando en la consulta del usuario.

MirBase (Griffiths-Jones, 2004; Griffiths-Jones et al., 2006, 2008; Kozomara y Griffiths-Jones, 2011, 2014; Kozomara et al., 2019) es una base de datos de alias de microARNs, si bien tiene la misma finalidad que el conjunto de datos de HUGO (ofrecer una abstracción del alias de microARN utilizado por el usuario), también se tomaron de esta herramienta las secuencias de nucleótidos que conforman dicho microARN. Estos últimos datos son usados para un servicio de información extra sobre un microARN en particular.

MirTarBase es una base de datos que identifica publicaciones científicas que postulan, confirman o reproducen una asociación microARN-gen. Las publicaciones

científicas están identificadas por su ID en [Pubmed](#) (repositorio de publicaciones científicas biomédicas). Sus datos se obtienen a partir de un algoritmo de minería de textos sobre los contenidos de los artículos de Pubmed, seguido de un proceso de curación manual. En Modulector utilizamos su información como complemento a los datos que ofrece mirDIP para aumentar la cantidad de referencias científicas disponibles.

IV) Implementación y servicios

La implementación de Modulector consiste en un mecanismo de sincronización para cada una de las fuentes de datos descritas en la sección anterior. Modulector extrae la información contenida en los archivos públicos de cada base de datos, conserva solo los datos de interés, los transforma, los enriquece y los almacena en una base de datos propia.

El procedimiento anterior solo se realiza una vez por base de datos, y se ejecuta nuevamente en forma manual en caso de que se detecte una nueva versión de las mismas con el fin de mantener la información de Modulector actualizada. Para el almacenamiento de esta información se utiliza PostgreSQL, el cual es un sistema de gestión de bases de datos relacional de alta performance y de código abierto.

Todos los servicios de Modulector se acceden mediante un navegador o cliente web (por ejemplo, Curl) y los parámetros pueden ser especificados en la URL. Por ejemplo, para consumir el servicio de fármacos ordenado por condición (patología)

y conservando sólo aquellos registros que hayan sido aprobados por la FDA, debe accederse mediante la URL:

http://<IP del servidor>/drugs/?ordering=condition&fda_approved=true

donde "*condition*" es el nombre del campo por el cual se aplica el ordenamiento de manera ascendente y "*fda_approved*" es el nombre campo por el cual se está realizando un filtrado, en este caso solo se conservan los registros que fueron aprobados por la FDA. El resultado es una respuesta paginada en formato JSON con la información filtrada y ordenada, según los criterios solicitados.

Si, además, se quisiera buscar por condición, fármaco o patrón de expresión y limitar el número de resultados a sólo las primeras dos asociaciones, podría obtenerse el resultado expuesto en la Figura 2 al acceder al siguiente recurso:

http://<IP del servidor>/drugs/?ordering=condition&fda_approved=true &search=Vorinostat&page_size=2

De esta forma, se ofrece al usuario una amplia variedad de filtros, ordenamiento, búsqueda y paginación, lo suficientemente flexibles como para cubrir una gran variedad de necesidades (Tabla 1). Todos los servicios disponibles junto con sus parámetros se especifican con detalle en el archivo README.md que se puede encontrar en el [repositorio de esta plataforma](https://github.com/omics-datascience/modulector)¹ con licencia MIT la cual habilita a cualquier usuario a usar, modificar, distribuir y vender versiones modificadas del código fuente de Modulector siempre y cuando se respete la propiedad intelectual del proyecto.

¹ <https://github.com/omics-datascience/modulector>

Con el fin de enriquecer la información de la base de datos integrada de Moduletor y de ofrecer al usuario los últimos hallazgos respecto a microARNs, se desarrolló un sistema de integración con el Centro Nacional de Información Biotecnológica (National Center for Biotechnology Information o NCBI).

```
{
  "count": 58,
  "next": "http://127.0.0.1:8001/drugs/?fda_approved=true&ordering=cc",
  "previous": null,
  "results": [
    {
      "id": 269058,
      "small_molecule": "Vorinostat (SAHA)",
      "fda_approved": true,
      "detection_method": "Microarray",
      "condition": "A549 human non-small cell lung cancer cells",
      "pubmed": "https://pubmed.ncbi.nlm.nih.gov/19513533",
      "reference": "Suberoylanilide hydroxamic acid (SAHA) change",
      "expression_pattern": "down-regulated",
      "support": "To further confirm that miRNA expression"
    },
    {
      "id": 268911,
      "small_molecule": "Vorinostat (SAHA)",
      "fda_approved": true,
      "detection_method": "Microarray",
      "condition": "A549 human non-small cell lung cancer cells",
      "pubmed": "https://pubmed.ncbi.nlm.nih.gov/19513533",
      "reference": "Suberoylanilide hydroxamic acid (SAHA) change",
      "expression_pattern": "up-regulated",
      "support": "To further confirm that miRNA expression"
    }
  ]
}
```

Figura 2. Resultado en formato JSON de la API Rest del servicio de fármacos aplicando una serie de parámetros de filtro, ordenamiento, búsqueda y paginación. El primer campo “count” indica la cantidad total de registros que cumplen con el criterio de búsqueda establecido, “next” y “previous” son enlaces a la próxima página y la anterior respectivamente. Por último, el campo “results” muestra los registros que corresponden a la página actual.

Figure 2 (previous page). Result in JSON format from the Rest API of the drug service applying a series of filtering, sorting, searching and pagination parameters. The first field "count" indicates the total number of records matching the set search criteria, "next" and "previous" are links to the next and previous page respectively. Finally, the "results" field shows the records that correspond to the current page.

Tabla 1. Parámetros de las URLs de los recursos ofrecidos por Modulector.

Table 1. URL parameters of the resources offered by Modulector.

Parámetro	Descripción
ordering	Indica los campos utilizados para el ordenamiento.
search	Indica el término de búsqueda que se aplicará a campos específicos de cada servicio.
page_size	Indica la cantidad de elementos que el servicio puede devolver como máximo
page_number	Número de página actual para poder iterar sobre los distintos bloques de resultados

Esta integración aporta dos funcionalidades principales:

1. Enriquecer los resultados de los servicios de consulta de asociación microARN-gen, a partir de una consulta en tiempo real al NCBI. Una vez obtenida la información se integra con la información de la base de datos de Modulector para generar el resultado. Cuando el tiempo de respuesta de los servidores de NCBI superan un umbral establecido, para no comprometer la performance de Modulector para con el usuario final, simplemente se cancela la petición interna entre NCBI y Modulector devuelve solo la información contenida en la base de datos propia.

2. Sistema de alertas para los usuarios, sobre nueva información de interés publicada. Un usuario se podrá registrar a una asociación gen-microARN de interés. Cuando se publique un artículo científico que apoye dicha asociación, el usuario será notificado por email. Este mecanismo consta de tres partes:

- Un servicio mediante el cual un usuario puede registrarse con su email y un objeto de interés (un microARN o un par microARN-Gen) para recibir las notificaciones vía email sobre el objeto en cuestión.
- Un servicio interno que se ejecuta semanalmente encargado de recorrer los registros del ítem anterior y consultar con las APIs públicas de NCBI para confeccionar los emails e informar a los usuarios pertinentes.
- Un servicio de desuscripción para que aquellos usuarios que no estén más interesados en las noticias de ciertos elementos puedan eliminar su registro de la plataforma.

V) Despliegue

Con el objetivo de mejorar la accesibilidad de Modulector, no solo se pone a disposición de cualquier desarrollador el código fuente en Github y las funcionalidades a través de Multiomix, sino que además se especifican una serie de instrucciones para poder realizar un despliegue de Modulector en cualquier servidor que disponga el usuario utilizando Docker. De esta manera se puede

desplegar la plataforma en un entorno privado, o en un servidor que escale frente a una demanda de performance más exigente.

Modulector se despliega en forma de servicio web a través de un servidor, su implementación está basada en tecnologías web robustas y populares, lo que hace a su implementación fácil, eficiente y mantenible. El acceso a la información se realiza a través de un navegador o un cliente web, evitando así la necesidad de instalar un software o lenguaje de programación específico para su uso por parte del usuario.

Por último, los datos sincronizados en Modulector se pueden descargar para ser importados directamente en una base de datos Postgres, a fin de ahorrarle al usuario que esté haciendo el despliegue en un entorno privado, todo el tiempo que requiere el proceso de sincronización inicial.

Es posible conseguir más información sobre cómo realizar el despliegue en un entorno privado en el siguiente link: <https://github.com/omics-datascience/modulector/blob/master/DEPLOYING.md>

Conclusiones

Los microARNs son moléculas de gran relevancia en estudios multi-ómicos que tienen por objetivo descifrar los mecanismos de desregulación de expresión génica en enfermedades complejas, multigénicas y multifactoriales, como es el caso de la mayoría de los tipos de cáncer. Adicionalmente, los microARN son de relevancia en

procesos de identificación de biomarcadores con poder diagnóstico, pronóstico/predictivo o para la identificación de potenciales nuevos blancos terapéuticos. Por esta razón, muchas herramientas y plataformas bioinformáticas necesitan acceder a información de microARNs de forma simple para enriquecer sus procesos. Modulector integra la información de las bases de datos públicas más relevantes de microARNs y las disponibiliza mediante una plataforma web cuyos pilares básicos se basan en la usabilidad, escalabilidad y mantenibilidad, simplificando el acceso a información actualizada.

Esta plataforma puede ser accedida vía web a través de Multiomix, desde otras herramientas/plataformas que la integren o realizar un despliegue en un entorno privado.

Con el fin de poder generar un servicio de fácil mantenimiento y actualización, se utilizaron tecnologías estables y en constante mantenimiento como Python, Pandas, PostgreSQL y Django. La utilización de este último junto con Django REST Framework permitió estandarizar la forma de trabajo.

Modulector expone distintas APIs que devuelven información de genes, microARNs y enfermedades, drogas y publicaciones relacionadas con las interacciones gene-microARN. Además, la documentación de Modulector forma parte integral del mismo, ya que uno de los objetivos es brindar a los usuarios con capacidades técnicas, la posibilidad de contribuir al proyecto.

Si bien Modulector surge a partir de una necesidad específica del proyecto Multiomix (www.multiomix.org) perteneciente al grupo de bioinformática en oncogenómica funcional de CAETI de UAI y al CINIBA de la Facultad de Medicina de la UNLP, queda disponible y será de gran utilidad para la comunidad científica trabajando en bioinformática que necesite acceder a esta información.

Referencias bibliográficas

-
- Agarwal, V., Bell, G.W., Nam, J. y Bartel, D.P. (2015). Predicting effective microRNA target sites in mammalian mRNAs. *eLife*, 4, e05005. <https://doi.org/10.7554/eLife.05005>
- Chang, L., Zhou, G., Soufan, O. y Xia, J. (2020). _miRNet 2.0 - network-based visual analytics for miRNA functional analysis and systems biology. *Nucleic Acids Research*, 48(W1), W244–W251. <https://doi.org/10.1093/nar/gkaa467>
- Cho, S., Jang, I., Jun, Y., Yoon, S., Ko, M., Kwon, Y. y Lee, S. (2012). MiRGator v3. 0: a microRNA portal for deep sequencing, expression profiling and mRNA targeting. *Nucleic Acids Research*, 41(D1), D252-D257. <https://doi.org/10.1093/nar/gks1168>
- Cho, S., Jun, Y., Lee, S., Choi, H. S., Jung, S., Jang, Y. y Kim, W. (2010). miRGator v2. 0: an integrated system for functional investigation of microRNAs. *Nucleic Acids Research*, 39(suppl_1), D158-D162. <https://doi.org/10.1093/nar/gkq1094>

- Fan, Y, Siklenka, K., Arora, SK., Ribeiro, P., Kimmins, S. y Xia, J. (2016). miRNet - dissecting miRNA-target interactions and functional associations through network-based visual analysis. *Nucleic Acids Research*, 44, W135-141. <https://doi.org/10.1093/nar/gkw288>
- Fan, Y. y Xia, J. (2018). miRNet: functional analysis and visual exploration of miRNA-target interactions in a network context. En: von Stechow, L., Santos Delgado, A. (eds) *Computational Cell Biology. Methods in Molecular Biology*, vol 1819. Humana Press, New York, NY. https://doi.org/10.1007/978-1-4939-8618-7_10
- Fan, Y., Habib, M. y Xia, J. (2018). Xeno-miRNet: a comprehensive database and analytics platform to explore xeno-miRNAs and their potential targets. *PeerJ*. 6, e5650. <https://doi.org/10.7717/peerj.5650>
- Fromm, B., Billipp, T., Peck, L. E., Johansen, M., Tarver, J. E., King, B. L. y Peterson, K. J. (2015). A uniform system for the annotation of vertebrate microRNA genes and the evolution of the human microRNAome. *Annual Review of Genetics*, 49, 213-242. <https://doi.org/10.1146/annurev-genet-120213-092023>
- Griffiths-Jones, S. (2004). The microRNA Registry. *Nucleic Acids Research*, 32(suppl_1), D109-D111. <https://doi.org/10.1093/nar/gkh023>
- Griffiths-Jones, S., Grocock R., Van Dongen S., Bateman A. y Enright A. (2006). miRBase: microRNA sequences, targets and gene nomenclature. *Nucleic Acids Research*, 34(suppl_1), D140-D144. <https://doi.org/10.1093/nar/gkj112>

- Griffiths-Jones, S., Kaur Saini, H., Van Dongen, S. y Enright A. (2008). miRBase: tools for microRNA genomics. *Nucleic Acids Research*, 36(suppl_1), D154–D158. <https://doi.org/10.1093/nar/gkm952>
- Huang, H.Y., Lin, Y.C., Li, J., Huang, K.Y., Shrestha, S., Hong, H.C., Tang, Y., Chen, Y.G., Jin, C.N., Yu, Y., Xu, J.T., Li, Y.M., Cai, X.X., Zhou, Z.Y., Chen, X.H., Pei, Y.Y., Hu, L., Su, J.J., Cui, S.D., ..., Huang, H.D. (2020). miRTarBase 2020: updates to the experimentally validated microRNA-target interaction database. *Nucleic Acids Research*, 48(D1), D148–D154. <https://doi.org/10.1093/nar/gkz896>
- Huang, Z., Shi, J., Gao, Y., Cui, C., Zhang, S., Li, J., Zhou, Y. y Cui, Q. (2019). HMDD v3.0: a database for experimentally supported human microRNA-disease associations. *Nucleic Acids Research*, 47(D1), D1013–D1017. <https://doi.org/10.1093/nar/gky1010>
- Jiang, Q., Wang, Y., Hao, Y., Juan, L., Teng, M., Zhang, X., Li, M., Wang, G. y Liu, Y. (2009). miR2Disease: a manually curated database for microRNA deregulation in human disease. *Nucleic Acids Research*, 37, D98-104. <https://doi.org/10.1093/nar/gkn714>
- Johora, H., Hossain, G.S. y Kocerha, J. (2019). The potential for microRNA therapeutics and clinical research. *Frontiers in Genetics*, 10, 478. <https://doi.org/10.3389/fgene.2019.00478>

Kozomara, A. y Griffiths-Jones, S. (2011). miRBase: integrating microRNA annotation and deep-sequencing data, *Nucleic Acids Research*, 39(suppl_1), D152–D157.

<https://doi.org/10.1093/nar/gkq1027>

Kozomara, A. y Griffiths-Jones, S. (2014). miRBase: annotating high confidence microRNAs using deep sequencing data. *Nucleic Acids Research*, 42(D1), D68–

D73. <https://doi.org/10.1093/nar/gkt1181>

Kozomara, A., Birgaoanu, M. y Griffiths-Jones, S. (2019). miRBase: from microRNA sequences to function. *Nucleic Acids Research*, 47(D1), D155–D162.

<https://doi.org/10.1093/nar/gky1141>

Licursi, V., Conte, F., Fiscon, G. y Paci, P. (2019). MIENTURNET: an interactive web tool for microRNA-target enrichment and network-based analysis. *BMC*

Bioinformatics, 20, 545. <https://doi.org/10.1186/s12859-019-3105-x>

Miranda, K.C., Huynh, T., Tay, Y., Ang, Y.S., Tam, W.L., Thomson, A.M. y Rigoutsos, I. (2006). A pattern-based method for the identification of MicroRNA binding sites and their corresponding heteroduplexes. *Cell*, 126(6), 1203-1217.

<https://doi.org/10.1016/j.cell.2006.07.031>

Ru, Y., Kechris, K., Tabakoff, B., Hoffman, P., Radcliffe, R., Bowler, R., Mahaffey, S., Rossi, S., Calin, G., Bemis, L. y Theodorescu, D. (2014). The multiMiR R package and database: integration of microRNA-target interactions along with their disease and drug associations. *Nucleic Acids Research*, 42(17), e133.

<https://doi.org/10.1093/nar/gku631>.

- Ruepp, A., Kowarsch, A., Schmidl, D., Buggenthin, F., Brauner, B., Dunger, I., Fobo, G., Frishman, G., Montrone, C. y Theis, F.J. (2010). PhenomiR: a knowledgebase for microRNA expression in diseases and biological processes. *Genome Biology*, 11(1), R6. <https://doi.org/10.1186/gb-2010-11-1-r6>
- Seungyoon, N., Bumjin, K., Seokmin, S. y Sanghyuk, L. (2007). miRGator: an integrated system for functional annotation of microRNAs. *Nucleic Acids Research*, 36(suppl_1), D159-D164. <https://doi.org/10.1093/nar/gkm829>
- Shirdel, E.A., Xie, W., Mak, T.W. y Jurisica, I. (2011). NAViGaTing the Micronome. Using Multiple MicroRNA Prediction Databases to identify signalling pathway-associated MicroRNAs. *PLoS ONE*, 6(2), e17429. <https://doi.org/10.1371/journal.pone.0017429>
- Tokar, T., Pastrello, C., Rossos, A.E.M., Abovsky, M., Hauschild, A.C., Tsay, M., Lu, R. y Jurisica, I. (2018). mirDIP 4.1-integrative database of human microRNA target predictions. *Nucleic Acids Research*, 46(D1), D360-D370. <https://doi.org/10.1093/nar/gkx1144>
- Vlachos, I.S., Zagganas, K., Paraskevopoulou, M.D., Georgakilas, G., Karagkouni, D., Vergoulis, T., Dalamagas, T. y Hatzigeorgiou, A. G. (2015). DIANA-miRPath v3.0: deciphering microRNA function with experimental support. *Nucleic Acids Research*, 43(W1), W460–W466. <https://doi.org/10.1093/nar/gkv403>

Wu, W. (2010). MicroRNA: potential targets for the development of novel drugs?.

Drugs in R & D, 10(1), 1-8. <https://doi.org/10.2165/11537800-000000000-00000>

Xinyi, L., Shuyuan, W., Fanlin, M., Jizhe, W., Yan, Z., Enyu, D., Xuexin, Y., Xia, L., Wei, J.

(2013). SM2miR: a database of the experimentally validated small molecules' effects on microRNA expression. *Bioinformatics*, 29(3), 409–411.

<https://doi.org/10.1093/bioinformatics/bts698>

DRAFT

Por Corina a las 17:54:02, 12 jul. 2021